The Evolution of the Immune System
Conservation and Diversification
Page left intentionally blank
The Evolution of the Immune System
Conservation and Diversification

Davide Malagoli
Department of Life Sciences
Biology Building, University of Modena
and Reggio Emilia, Modena, Italy
Dedication

To my teachers, my students, and all the people who taught me something.
Contents

Contributors xv
Preface xvii

1. Hematopoiesis and Hemocytes in Pancrustacean and Molluscan Models
Valerie J. Smith, Alice Accorsi, Davide Malagoli

1 Introduction 1
2 Hematopoiesis in Pancrustacean Models 2
2.1 Insect Hematopoiesis 2
2.2 Crustacean Hematopoiesis 5
3 Molluscan Hematopoiesis 6
3.1 Bivalves 6
3.2 Gastropods 7
3.3 Cephalopods 8
4 Hematopoiesis in Pancrustacean and Molluscan Models: Final Considerations 9
5 Hemocytes in Pancrustacean and Molluscan Models 9
5.1 Hemocytes and Their Lineages in Insects 10
5.2 Hemocytes and Their Lineages in Crustaceans 15
5.3 Hemocytes and Their Lineages in Mollusks 17
6 Concluding Remarks 20
Acknowledgments 21
References 21

2. Origin and Functions of Tunicate Hemocytes
Francesca Cima, Nicola Franchi, Loriano Ballarin

1 Introduction 29
2 Ascidian Circulation 30
3 Ascidian Hemocytes 31
3.1 Hematopoietic Tissues 31
3.2 Hemocyte Morphology 31
4 Immune Role of Tunicate Hemocytes 33
4.1 Phagocytosis 33
4.2 Encapsulation 34
4.3 Cytotoxicity 35
4.4 Inflammation 35
Yoichi Sutoh, Masanori Kasahara

1 Introduction 51
2 Overview of VLRs 52
   2.1 Structure of VLR Proteins and Gene Assembly 52
   2.2 Both Lampreys and Hagfish Have Three VLR Genes 54
   2.3 Crystal Structure of VLR Proteins and Antigen Recognition 55
   2.4 VLRA and VLRC Genes are Tightly Linked in the Lamprey Genome 56
3 Three Populations of Agnathan Lymphocytes 56
   3.1 VLRA+ Cells and VLRC+ Cells Resemble Gnathostome T Cells, Whereas VLRB+ Cells Resemble Gnathostome B Cells 56
   3.2 Antigen Recognition by Agnathan T-Like Cells is Still Full of Mystery 57
4 Lymphocyte Development in Jawless Vertebrates 59
   4.1 Development of VLRB+ Cells 59
   4.2 Development of VLRA+ Cells and VLRC+ Cells 59
   4.3 Cell Fate Determination of VLRA+ Cells and VLRC+ Cells 61
   4.4 Evidence for Selection in “Thymoids” 61
   4.5 Evolution of Lymphocytes in Vertebrates 61
5 Conclusions 64
Acknowledgments 64
References 64

4. The Evolution of Lymphocytes in Ectothermic Gnathostomata
Giuseppe Scapigliati, Francesco Buonocore

1 Introduction 69
2 Fishes 71
   2.1 B Cells 72
   2.2 T Cells 76
3 Conclusions 79
References 79
5. Vertebrate Cytokines and Their Evolution
Christopher J Secombes, Tiehui Wang, Steve Bird

1 Introduction 87
2 The IL-1 Cytokine Family 87
  2.1 Vertebrate IL-1F2 88
  2.2 Vertebrate IL-1F4 92
  2.3 Bird and Reptile IL-1F3 and IL-1F5 94
  2.4 Novel Fish IL-1F 96
3 The IL-2 Cytokine Family (γC Cytokines) 96
  3.1 IL-2 and IL-2L 96
  3.2 IL-4 and IL-4/13 97
  3.3 IL-7 98
  3.4 IL-9 98
  3.5 IL-15 and IL-15L 99
  3.6 IL-21 99
4 IL-6/IL-12 Superfamily 100
  4.1 IL-6 Family 101
  4.2 The IL-12 Family 104
5 IL-10 Family 109
  5.1 IL-22 109
  5.2 IL-26 110
  5.3 IL-10 110
  5.4 vIL-10 112
6 IL-17 Family 112
7 TNF Family 116
  7.1 Protostomian Invertebrates 116
  7.2 Primitive Chordates and Basal Vertebrates: Cephalochordates, Urochordates, and Cyclostomata 118
  7.3 Fish 119
  7.4 Birds, Reptiles, and Amphibians 126
References 128

6. The Evolution of Complement System Functions and Pathways in Vertebrates
Miki Nakao, Tomonori Somamoto

1 Introduction 151
2 Invertebrate Complement System Representing an Ancestral Architecture of Vertebrate System 153
3 Phylogeny of Complement Pathways 154
  3.1 Lectin Pathway 154
  3.2 Alternative Pathway 159
  3.3 Classical Pathway 160
  3.4 Lytic Pathway 161
4 Phylogeny of Complement Receptors 162
5 Phylogeny of the Regulatory Mechanism of Complement Activation 163
### 6. Evolutionary Significance of Multiple Isoforms in Teleost Complement Components

7. Concluding Remarks
Acknowledgments
References

7. **Antiviral Immunity: Origin and Evolution in Vertebrates**
*Jun Zou, Rosario Castro, Carolina Tafalla*

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Introduction</td>
<td>173</td>
</tr>
<tr>
<td>2 Virus-Sensing Pattern Recognition Receptors</td>
<td>174</td>
</tr>
<tr>
<td>2.1 Toll-Like Receptor Family</td>
<td>175</td>
</tr>
<tr>
<td>2.2 RIG-Like Receptor Family</td>
<td>178</td>
</tr>
<tr>
<td>2.3 NOD-Like Receptor Family</td>
<td>179</td>
</tr>
<tr>
<td>3 RNA Interference</td>
<td>180</td>
</tr>
<tr>
<td>4 Type I IFNs</td>
<td>181</td>
</tr>
<tr>
<td>4.1 Mammalian IFNs</td>
<td>181</td>
</tr>
<tr>
<td>4.2 Avian and Reptile Type I IFNs</td>
<td>184</td>
</tr>
<tr>
<td>4.3 Amphibian IFNs</td>
<td>184</td>
</tr>
<tr>
<td>4.4 Fish IFNs</td>
<td>185</td>
</tr>
<tr>
<td>5 IFN Receptors</td>
<td>187</td>
</tr>
<tr>
<td>6 Downstream Modulators of the Interferon Response</td>
<td>188</td>
</tr>
<tr>
<td>6.1 Mx Protein</td>
<td>189</td>
</tr>
<tr>
<td>6.2 Viperin</td>
<td>190</td>
</tr>
<tr>
<td>6.3 PKR</td>
<td>190</td>
</tr>
<tr>
<td>6.4 ISG15</td>
<td>191</td>
</tr>
<tr>
<td>6.5 TRIM</td>
<td>191</td>
</tr>
<tr>
<td>7 Conclusions</td>
<td>192</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>192</td>
</tr>
<tr>
<td>References</td>
<td>193</td>
</tr>
</tbody>
</table>

### 8. Lectins as Innate Immune Recognition Factors: Structural, Functional, and Evolutionary Aspects
*Gerardo R. Vasta*

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Introduction</td>
<td>205</td>
</tr>
<tr>
<td>2 Galectins: A Conserved Lectin Family with Multiple Roles in Development and Immunity</td>
<td>207</td>
</tr>
<tr>
<td>2.1 Molecular, Structural, and Evolutionary Aspects</td>
<td>208</td>
</tr>
<tr>
<td>2.2 Functional Aspects</td>
<td>210</td>
</tr>
<tr>
<td>3 F-type Lectins: A Structurally and Functionally Diversified Lectin Family</td>
<td>213</td>
</tr>
<tr>
<td>3.1 Molecular, Structural, and Evolutionary Aspects</td>
<td>213</td>
</tr>
<tr>
<td>3.2 Functional Aspects</td>
<td>217</td>
</tr>
<tr>
<td>4 Conclusions</td>
<td>218</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>219</td>
</tr>
<tr>
<td>References</td>
<td>219</td>
</tr>
</tbody>
</table>
Enzo Ottaviani

1 Introduction
1.1 Lymphocyte as a Neuroendocrine Cell
1.2 Immunocyte as an Immune–Neuroendocrine Cell

2 Role of Neuroendocrine Hormones, Opioid Peptides and Cytokines in the Invertebrate Immune and Neuroendocrine (Stress Response) Responses
2.1 Immune Responses
2.2 Cell-Shape Changes
2.3 Chemotaxis
2.4 Pagocytosis
2.5 Stress Response
2.6 The Immune–Mobile Brain

3 Concluding Remarks

References

10. The Immune-Related Roles and the Evolutionary History of Dscam in Arthropods
Sophie A.O. Armitage, Daniela Brites

1 The Down Syndrome Cell Adhesion Molecule
1.1 From One Gene to Thousands of Protein Isoforms
1.2 Structural Aspects of Dscam

2 Immune-Related Roles of Dscam-HV in Arthropods
2.1 What Is the Evidence That Dscam-hv is Involved in the Immune defence of Pancrustaceans?
2.2 How Has Dscam-hv Been Predicted to Function in Arthropod Immunity?

3 The Evolutionary History of Dscam-HV in Arthropods
3.1 The Origin of Dscam
3.2 Dscam Radiation in Arthropods
3.3 The Evolutionary History of Dscam-hv in Pancrustaceans
3.4 Dscam Clusters of Duplicated Exons
3.5 Are the Patterns of Dscam-hv Molecular Evolution Compatible with a Role in Immunity?

4 Experimental Methodology and Future Perspectives
4.1 Experimental Methodology: Choice of Experimental Host and Parasite/Antigen, and Other Experimental Parameters
4.2 Future Perspectives

5 Concluding Remarks

Acknowledgments

References
11. Structural and Functional Diversity of Fibrinogen-Related Domains
Russell F. Doolittle

1. Introduction 275
2. Structural Features 275
3. Phylogenetic Occurrence 276
4. FReDs and Neighbor Domains 278
5. Human FRePs 279
6. Evolution of Functional Diversity in Chordates 281
7. Rates of Sequence Change 284
8. FReDs in Protochordates 286
9. Extending the Deuterostome Time-Line 286
10. Nondeuterostome Animals 286
11. Sponges and Jellyfish 287
12. Bilaterians 288
   12.1 Phylum Mollusca 288
   12.2 Phylum Arthropoda 288
   12.3 Pancrustacea—Crustaceans 289
   12.4 Insects 289
13. FReD-Based Immunity and the Recognition Problem 289
14. Concluding Remarks 290
References 291

Matan Oren, Megan A. Barela Hudgell, Preethi Golconda, Cheng Man Lun, L. Courtney Smith

1. Introduction 295
2. The NBS-LRR Gene Family in Higher Plants 296
3. The 185/333 Gene Family in Echinoids 298
4. Gene Diversification 300
5. Clustering and Tandem Repeats 301
6. Speculations on Diversification Mechanisms of the Sp185/333 Genes 303
7. Conclusions 307
Acknowledgment 307
References 307

13. The Evolution of the Toll-Like Receptor System
Carlos G.P. Voogdt, Jos P.M. van Putten

1. Introduction 311
2. TLR Structure 312
   2.1 Extracellular Domain 312
2.2 Transmembrane Domain 313
2.3 Intracellular Domain 313

3 Evolution and Distribution of TLR Genes 314
3.1 Origin of TLRs 314
3.2 Structural Difference Between Protostomian and Deuterostomian TLRs 316
3.3 Toll and TLR Repertoire in Protostomes and Deuterostomes 316
3.4 Vertebrate TLR Phylogeny 317

4 TLR Function 318
4.1 Protostome Toll Function 319
4.2 Invertebrate Deuterostome TLR Function 320
4.3 Vertebrate TLR Function 320

5 Microbe-Driven TLR Evolution 322
5.1 Microbial Methods to Avoid Detection by TLRs 322
5.2 Purifying Selection on TLRs 323
5.3 Positive Selection on TLRs 323

6 Concluding Remarks 324
References 325

14. The Evolution of Major Histocompatibility Complex in Teleosts
Masaru Nonaka, Mayumi I. Nonaka

1 Genome Organization 331
2 Class I 334
3 Class II 339
4 PSMB8 341
5 Conclusions 345
References 345

Subject Index 351
Contributors

*Numbers in Parentheses indicate the pages on which the author’s contributions begin.*

Alice Accorsi (1), Stowers Institute for Medical Research, Kansas City; Howard Hughes Medical Institute, Stowers Institute for Medical Research, Kansas City, MO, United States

Sophie A.O. Armitage (241), Institute for Evolution and Biodiversity, University of Münster, Münster, Germany

Loriano Ballarin (29), Department of Biology, University of Padova, Padova, Italy

Megan A. Barela Hudgell (295), Department of Biological Sciences, The George Washington University, WA, United States

Steve Bird (87), Molecular Genetics, Department of Biological Sciences, University of Waikato, Hamilton, New Zealand

Daniela Brites (241), Swiss Tropical and Public Health Institute, Basel, Switzerland

Francesco Buonocore (69), Department for Innovative Biology, Agro-Industry and Forestry, University of Tuscia, Viterbo, Italy

Rosario Castro (173), Animal Health Research Center (CISA-INIA), Carretera de Algete a El Casar, Valdeolmos (Madrid), Spain

Francesca Cima (29), Department of Biology, University of Padova, Padova, Italy

Russell F. Doolittle (275), Departments of Chemistry & Biochemistry and Molecular Biology, University of California, San Diego, La Jolla, CA, United States

Nicola Franchi (29), Department of Biology, University of Padova, Padova, Italy

Preethi Golconda (295), Department of Biological Sciences, The George Washington University, WA, United States

Masanori Kasahara (51), Department of Pathology, Hokkaido University Graduate School of Medicine, Sapporo, Japan

Davide Malagoli (1), Department of Life Sciences, Biology Building, University of Modena and Reggio Emilia, Modena, Italy

Cheng Man Lun (295), Department of Biological Sciences, The George Washington University, WA, United States

Miki Nakao (151), Department of Bioscience and Biotechnology, Faculty of Agriculture, Kyushu University, Hakozaki, Fukuoka, Japan

Masaru Nonaka (331), Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan
Mayumi I. Nonaka (331), Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan

Matan Oren (295), Department of Biological Sciences, The George Washington University, WA, United States

Enzo Ottaviani (225), Department of Life Sciences, University of Modena and Reggio Emilia, Modena, Italy

Giuseppe Scapigliati (69), Department for Innovative Biology, Agro-Industry and Forestry, University of Tuscia, Viterbo, Italy

Christopher J Secombes (87), Scottish Fish Immunology Research Centre, University of Aberdeen, Zoology Building, Tillydrone Avenue, Aberdeen, United Kingdom

L. Courtney Smith (295), Department of Biological Sciences, The George Washington University, WA, United States

Valerie J. Smith (1), Scottish Oceans Institute, School of Biology, University of St Andrews, St Andrews, Fife, United Kingdom

Tomonori Somamoto (151), Department of Bioscience and Biotechnology, Faculty of Agriculture, Kyushu University, Hakozaki, Fukuoka, Japan

Yoichi Sutoh (51), Emory Vaccine Center and Department of Pathology and Laboratory Medicine, Emory University, Atlanta, GA, United States

Carolina Tafalla (173), Animal Health Research Center (CISA-INIA), Carretera de Algete a El Casar, Valdeolmos (Madrid), Spain

Jos P.M. van Putten (311), Department of Infectious Diseases and Immunology, Utrecht University, Utrecht, The Netherlands

Gerardo R. Vasta (205), Department of Microbiology and Immunology, University of Maryland School of Medicine, UMB, and Institute of Marine and Environmental Technology, Columbus Center, Baltimore, MD, United States

Carlos G.P. Voogdt (311), Department of Infectious Diseases and Immunology, Utrecht University, Utrecht, The Netherlands

Tiehui Wang (87), Scottish Fish Immunology Research Centre, University of Aberdeen, Zoology Building, Tillydrone Avenue, Aberdeen, United Kingdom

Jun Zou (173), Scottish Fish Immunology Research Centre, University of Aberdeen, Zoology Building, Aberdeen, United Kingdom
Preface

"Lord, we know what we are, but know not what we may be."

Shakespeare

“Evolution” is a fascinating term that may be used in several different contexts. In everyday life, evolution refers to what is in progress, and will change in a short time. In natural history and comparative anatomy, the word “evolution” still means change, but on a much longer period. Changes are continuous, and there is no need to be a paleontologist or a biologist to know that during geological eras several organisms got extinct, and new different species replaced them, profiting of the free ecological niches.

“Immunity,” is a well-known concept as well. The struggle to survive does not consist only of the ability to escape the predator and to catch the prey. It also requires contrasting potential pathogens, and limiting the expansion of microorganisms that have beneficial effects only as long as they are kept within the proper boundaries. Again, to be a scientist or a Doctor of Medicine is not necessary to understand that a good health implies a good functioning of the immune system, working 24/7 to prevent the pathogen widespread into the organism.

However, if the theory of evolution and the indispensable role of the immune system are obvious concepts for everyone, the evolution of the immune system is a less obvious topic. Scientists have no clues of the immune system of the extinct organisms because it does not leave traces in fossils. And since the largest part of immunologists is represented by individuals with a medical education, it is frequently ignored that apparently simple organisms, such as flies, snails, and sea urchins present an incredible ability to discriminate between numerous microorganisms and to elicit an efficacious immune response, without using antibodies.

This book will help the reader to gain a modern and open view of the evolution of the immune systems in multicellular organisms. Since the immune system is not leaving hints of its functioning in fossil records, it is fundamentally a comparative analysis of information from as much present-day models as possible. In these respects, the book gives the reader a complete overview of the most studied models and of the hot topics in comparative immunology. I would like to remark at this point that the outline of the book is conceived to allow a comparative analysis of different immune components, as it focuses on the several aspects shared by the immune systems of all the metazoans, highlighting
conserved components and remarking the important features that diversified along evolutionary lineages.

The book consists of three sections:

Section 1 is entitled The Players of Cell-Based Immunity in Metazoans and it contains an up-to-date description of the cells involved in the immune response in the principle nonmammalian models. Chapter: Hematopoiesis and Hemocytes in Pancrustacean and Molluscan Models highlights the similarities and the differences of the hemocytes described in Molluscs and Pancrustacea, and it gives a detailed description of the hematopoiesis and of the maturational niches observed in these vast, highly populated and extremely divergent taxa. Closer to humans, the Chordate phylum is considered in detail, starting from the sister-group of vertebrates (the Tunicates, chapter: Origin and Functions of Tunicate Hemocytes) and moving to lampreys (chapter: Lymphocyte Populations in Jawless Vertebrates: Insights Into the Origin and Evolution of Adaptive Immunity) and bony fish (chapter: The Evolution of Lymphocytes in Ectothermic Gnathostomata). In these chapters, the evolution of one of the most interesting cell type, that is, the lymphocyte, is described, and the functional convergence of the lymphocyte-like system of lampreys is also presented, in order to give the reader the opportunity to get in touch with the stunning evolutionary plasticity of the immune system.

Section 2 focuses on The Evolution of Immune-Related Soluble Factors. The section kicks off with a detailed and complete analysis of the cytokines observed in vertebrates to date (chapter: Vertebrate Cytokines and Their Evolution). Molecular details are provided, helping the reader in identifying which are the molecular components more conserved during vertebrate diversification. Chapter: Vertebrate Cytokines and Their Evolution also resumes the knowledge about potential homologs of vertebrate cytokines observed in nonvertebrate models. With a similar approach and detail, other basic and fundamental components of humoral immune response are described in chapters: The Evolution of Complement System Functions and Pathways in Vertebrates (the complement system) and Antiviral Immunity: Origin and Evolution in Vertebrates (antiviral molecules). Lectins, and more precisely galectins and F-type lectins, are the amazing protagonists of chapter: Lectins as Innate Immune Recognition Factors: Structural, Functional, and Evolutionary Aspects. Here, the numerous roles played by these molecules are presented, and the structural details are described to highlight the number of interactions that these molecules can entertain. The second section ends with the description of the origin and the evolution of the neuro-immune cross-talk, encompassing more than 30 years of studies in this field (chapter: Origin and Evolution of the Neuro-Immune Cross-Talk in Immunity).

Section 3 introduces the reader to The Evolution of Diversity in Immune System. Here, some of the most fascinating discoveries of the last decade are presented. Chapter: The Immune-Related Roles and the Evolutionary History of Dscam in Arthropods is centered on Dscam, a molecule representing an extreme
example of somatic diversification in Pancrustacea. Respecting in full the layout of the book, the chapter: The Immune-Related Roles and the Evolutionary History of Dscam in Arthropods reviews the involvement of Dscam in pancrustacean immunity. Dscam evolutionary history is also presented, together with potential developments for future research. Similarly, chapter: Structural and Functional Diversity of Fibrinogen-Related Domains describes the diversity developed by fibrinogen-related domains during the evolution of metazoans. Still on the topic of highly variable molecules, chapter: Genomic Instability and Shared Mechanisms for Gene Diversification in Two Distant Immune Gene Families: The Plant NBS-LRR Genes and the Echinoid 185/333 Genes presents the 185/333 protein family in sea urchins and highlights the mechanisms for gene diversification that 185/333 genes share with the evolutionary remote NBS-LRR genes of higher plants. The evolution of Toll-like receptors is reviewed in chapter: The Evolution of the Toll-Like Receptor System, and the selective microbial pressures that drive Toll-like receptors adaptation is described in detail. The book ends up with a focus on the major histocompatibility complex (MHC) of teleosts. This group of bony fish is the most diverse among vertebrates, and its MHC presents several unique features that are presented in the final chapter: The Evolution of Major Histocompatibility Complex in Teleosts.

This Book provides a new and original perspective about the evolution of the immune system, and it will help scientists from different educational paths and at different stages of their career to appreciate the conserved traits and the original innovations of immune systems.

Davide Malagoli
Department of Life Sciences, University of Modena and Reggio Emilia, Italy
Page left intentionally blank
Chapter 1

Hematopoiesis and Hemocytes in Pancrustacean and Molluscan Models

Valerie J. Smith*, Alice Accorsi**,†, Davide Malagoli‡
*Scottish Oceans Institute, School of Biology, University of St Andrews, St Andrews, Fife, United Kingdom; **Stowers Institute for Medical Research, Kansas City, MO, United States; †Howard Hughes Medical Institute, Stowers Institute for Medical Research, Kansas City, MO, United States; ‡Department of Life Sciences, Biology Building, University of Modena and Reggio Emilia, Modena, Italy

1 INTRODUCTION

All animals are at risk of damage from microorganisms, parasites, toxins, abrasive particles, or other threats that may kill or bring about somatic cell or tissue injury with longer-term detriment to health. The immune system is the key player in implementing surveillance and response to counter such threats, and the efficacy of its recognition and reactivity abilities is independent of the taxonomic or phylogenetic status of the species. This holds true for both long- and short-lived species, and both vertebrates and invertebrates.

Invertebrate immune systems are often divided into cellular and humoral components.1 Cellular components include the circulating blood cells, called hemocytes, as the dominant representatives. The renewal of hemocytes is usually referred to as hematopoiesis or hemopoiesis.2 The existing literature on invertebrate hematopoiesis is greatly exceeded by that on hemocyte morphology and functions. Notwithstanding, comparative studies on hematopoiesis are crucial for the understanding of the molecular basis of hemocyte proliferation, differentiation, and maturation, especially in models of economic relevance. For mammals and most other vertebrates, which have both innate and adaptive immune responses, the immune cells are derived from either a progenitor of a myeloid or a lymphoid lineage, both originating mainly from the bone marrow, but differentiating and maturing in secondary sites around the body.3–5 In protostome and deuterostome invertebrates, by contrast, the circulating blood cells that participate in host defense may come from a variety of tissues or organs and there is no clear distinction of myeloid or lymphoid cell lineages equivalent to that of vertebrates. In protostomes, to which
the Pancrustacea and Mollusca belong, the available evidence points to them originating in one or more dedicated parts of the connective tissues; for example, in the lymph gland in insects,\textsuperscript{6} in the cephalothorax in decapods,\textsuperscript{7,8} in the connective tissue at the base of the gills in bivalves,\textsuperscript{9,10} in the pericardial walls in gastropods,\textsuperscript{11} or in the white gland in cephalopods.\textsuperscript{12,13} The anatomies of the various species and their adaptations to different environments make almost impossible a homogenous description of hematopoiesis and hemocyte lineages in invertebrates. As a general concept, however, hemocytes are cells of mesodermal origin, arising from the coelom and referred to as hemocytes in pancrustaceans and mollusks because the coelom is reduced in these animals to the circulatory fluids contained within a cavity enclosing the heart and main body organs, known as the hemocoel.\textsuperscript{2} In this chapter, the hematopoietic organs and the different hemocytes present in relevant pancrustacean and molluscan models are described.

## 2 HEMATOPOIESIS IN PANCRUSTACEAN MODELS

Pancrustacea is a recently identified invertebrate clade that includes the two largest, most studied and representative groups of arthropods, namely, all hexapods and crustaceans.\textsuperscript{14–17} This clade comprises a monophyletic group that contains millions of species, all highly diverse in their development, anatomy, and physiology. They are united, however, through molecular data derived from comparisons of nuclear rRNA genes, mitochondrial rRNA genes, and certain protein-coding genes.\textsuperscript{14} Members of the clade are believed to have diversified from a common ancestor in the Silurian, 420–430 million years ago.\textsuperscript{18}

Here, we consider certain key model species that are scientifically and/or economically important, and thus well-studied. In both insects and crustaceans, the hemocytes are present in the hemocoel during both larval and adult life stages throughout life. Holometabolous insects (as represented by \textit{Drosophila melanogaster}) undergo a complete metamorphosis in the transition from larval to adult stage and they usually present a discontinuous hematopoiesis, limited to embryonic and larval life stages.\textsuperscript{19,20} This may be because in these insects the adult and final stage is principally devoted to reproduction,\textsuperscript{21} and it is relatively a short part of the lifespan.\textsuperscript{20} Conversely, decapods, the best-known crustacean clade, are generally iteroparous (ie, have multiple reproduction cycles) and grow after sexual maturity. Thus, hematopoiesis in these animals tends to be continuous.\textsuperscript{22}

### 2.1 Insect Hematopoiesis

Fruit flies, especially \textit{D. melanogaster}, are widely used models for understanding development and immunity, and, with respect to hematopoiesis, have been employed to study the process from early-stage embryos onwards. In this animal, hematopoiesis occurs in successive waves of cell proliferation.\textsuperscript{19} The first wave, where immature hemocytes (prohemocytes) form a loose mass of cells, occurs in the head mesoderm during early embryonic development.\textsuperscript{19,23} These
embryonic prohemocytes then migrate in a regulated way, along the dorsal blood vessel and the ventral neural chord, to permeate the entire embryo. In this phase, embryonic hemocytes acquire phagocytic capability. Embryonic hemocytes maintain their activity in larvae and persist even after metamorphosis into adult flies. In *Drosophila* larvae, immature hemocytes migrate to the hemolymph and populate specific niches. These niches are segmentally divided hematopoietic-microenvironments, in which the hemocytes may proliferate de novo during larval stages. Experimental manipulation of larvae promotes the dispersal of the hemocytes from the hematopoietic niches to the circulation, but they rapidly return to their home niche. The peripheral nervous system plays a fundamental role in regulating this hemocyte homing and the proliferation in the hematopoietic niches, indicating that the functional connection between immune and nervous systems may be crucial for hematopoiesis in the *Drosophila* larvae. In this insect, the main hematopoietic organ is the lymph gland that begins its formation during late embryogenesis. The lymph gland is hematopoietically active only in larvae. It arises from the cardiogenic mesoderm of the embryo, a feature reminiscent of hematopoiesis in mammals, where circulating cells derive from the hemangioblasts. The lymph gland accounts for most of the larval hematopoiesis in *Drosophila*, and it remains in close contact with the dorsal blood vessel, a circulatory organ with heart-like activity. Despite the lymph gland hosting hemocyte proliferation and differentiation, in the absence of immune challenges, the hemocytes are retained by the lymph gland and released immediately after pupation. The larval lymph gland has a lobular organization, and the anterior (primary) pair of lobes are the most important for the development of hemocytes, with the primary lobe having cortical zone containing maturing hemocytes. The inner layer, defined as the medullary lobe, contains immature hemocytes, usually described as prohemocytes. The primary lobe also includes the posterior signaling center (PSC), with specialized and unique hemocytes, which are not immediately recognizable either as the maturing hemocytes or as immature prohemocytes. The PSC exerts control over the proliferation and differentiation of the prohemocytes, and it also controls hemocyte maturation in the primary lobe. At the end of the larval period, the lymph gland reaches its maximum size, and after pupation it dissociates and releases all the hemocytes, irrespective of their level of maturation. The mediators and signaling pathways that regulate the release of hemocytes from the lymph gland are unknown at present. Transmission electron microscopy observations, plus analysis using hemocyte markers, indicate that the release of hemocytes from the lymph gland after pupation may occur as a consequence of localized degeneration of the gland basement membranes and through augmented hemocyte motility. This means that adult fruit flies, which can live for approximately 4 weeks, are able to survive after metamorphosis without a hematopoietic organ. Adults therefore rely on protection from the larval hemocytes that persist in the circulation and produce immune humoral factors, such as prophenoloxidase (proPO) activating components and antimicrobial peptides (AMPs).
The genetic tools available for *D. melanogaster* have allowed the detection of specific markers on maturing cells and of molecular mediators responsible for driving the hemocyte maturation.\textsuperscript{31} Recently, the *Drosophila* fibroblast growth factor receptor (FGFR), Heartless, has been shown to play a role in the maturation of larval hemocytes.\textsuperscript{31} Two FGF-8-like ligands, Thisbe and Pyramus,\textsuperscript{32,33} activate, through Heartless, a signaling pathway essential for hemocyte maturation and lymph gland development. However, several aspects of hemocyte maturation in the lymph gland remain to be elucidated, but it is worth noting that in *Drosophila*, the extracellular matrix is involved in the process.\textsuperscript{31} Trol, a *perlecan* homolog, is a heparan sulfate-containing proteoglycan that negatively regulates the availability of Thisbe and Pyramus. Acting on ligand availability, Trol intervenes in Heartless signaling within the lymph gland and regulating hemocyte maturation and lymph gland lobe differentiation.\textsuperscript{31}

There is less detailed information about hematopoiesis for other insect species, although in the silkworm, *Bombyx mori*, and the hornworm, *Manduca sexta* (Lepidoptera), embryonic hematopoiesis also involves the head mesoderm, as in *Drosophila*, and, in larvae, it takes place in the thoracic mesoderm-derived organ, the lymph gland, or the hematopoietic producing organ (HPO), positioned on imaginal disks of the wings.\textsuperscript{34} In *B. mori*, the lymph gland dissociates during metamorphosis.\textsuperscript{6,34,35} A further similarity to *D. melanogaster* is the presence of compact islets of prohemocytes surrounded by loose islets of differentiated hemocytes in *B. mori* larval HPO. The compact and loose islets of *B. mori* HPO structurally resemble the medullary and cortical zones of the *Drosophila* lymph gland, respectively.\textsuperscript{35} The PSC of *Drosophila* does not have an equivalent counterpart in the silkworm lymph gland, but it has been proposed that a specific group of hemocytes present in the silkworm HPO, the reticular cells, may play a role in controlling the hemocyte proliferation and differentiation.\textsuperscript{36} Studies on insects undergoing incomplete metamorphosis are much less frequent and provide less detailed data.\textsuperscript{6} The hematopoietic organs and the reticular cells have been investigated also in some orthopteran insect models, such as locusts and grasshoppers. The importance of reticular cells in the hematopoietic tissues of these insects is implied by the term, reticulo-hematopoietic organ.\textsuperscript{37,38} Whereas holometabolous insects such as *D. melanogaster* and *B. mori* possess distinct embryonic and larval sites for hematopoiesis and any adult hematopoietic organs, the hemimetabolan insects *Locusta migratoria* and *Euprepocnemis shirakii* have a functional lymph gland at adult stage that is tightly associated with the dorsal blood vessels.\textsuperscript{6} As in *D. melanogaster*, the lymph glands of *L. migratoria* and *E. shirakii* comprise two layers containing hemocytes at different degrees of maturation.\textsuperscript{6,38–40} Very recently, it has been observed that the hematopoietic tissue may not be the main source of new hemocytes in adult *L. migratoria*.\textsuperscript{41} A drop in the number of circulating hemocytes occurs immediately after injection of β-1,3-glucan (fungal membrane component) into the locusts. Hemocyte replenishment seemingly followed the mitotic activity of the circulating hemocytes rather than that of the hematopoietic tissue.\textsuperscript{41}
2.2 Crustacean Hematopoiesis

Among crustaceans, the hematopoiesis has been studied in detail only in decapods, principally, shrimp and crayfish. In these invertebrates, it takes place in a dedicated tissue located on the dorsolateral surface of the foregut, in the cephalothorax. In penaeid shrimp it is large and composed of lobules, full of small, undifferentiated, proliferative, immature cells (again, termed prohemocytes). The lobules may also extend into the bases of the pereiopods (walking legs) or along the ophthalmic and antennal arteries in the vicinity of the pericardial region. In other species, particularly crab, the tissue is small and diffuse, making it less conspicuous than the hematopoietic tissue in shrimp or crayfish.

The exact processes by which the hemocytes develop within the hematopoietic organ is not fully understood, but in the shrimp, *Penaeus monodon*, the smaller agranulated cells tend to occur at the periphery of the lobules, while the larger, more granulated cells are present in the center. This pattern of distribution is believed to reflect a gradient of maturation, with mature cells released from lobule center into the hemocoel. Previous authors, working on other decapod species, hold a different view, and have considered that maturation takes place in the circulation, as low numbers of dividing cells, resembling prohemocytes, may be present in the hemolymph. However, it is possible that under certain conditions—for example, following nonself challenge or injury-induced blood loss—hematopoiesis might be upregulated to repopulate the hemolymph, with new defense cells thus driving partially matured prohemocytes into the circulation prematurely. Certainly, the number of mitotic cells in the circulation rises approximately two- to fivefold (from <1%) after such treatments. Increased levels of mitosis have also been reported to occur within the hematopoietic tissue by immune stimulation.

Whatever the stimulus, a variety of signaling events directly or indirectly induces cell division within the hematopoietic tissue itself. Work on crayfish, *Pacifastacus leniusculus*, has shown that prokineticin-like cytokines, named astakines, and transglutaminases (TGases) play important roles in the process. Astakines cause upregulation of hematopoiesis, transiently increasing the total hemocyte count in a dose-dependent manner while concomitantly suppressing TGase activity. Similar events are likely to occur in other decapod species, as a gene encoding an astakine homolog has been found in *P. monodon*. Moreover, a further study on *P. monodon* has shown that astakine levels are influenced not only by TGase, but also by a crustin AMP. Depletion of these two proteins results in elevated levels of astakine and increased hemocyte proliferation, despite astakine mRNA levels remaining unaffected. Other factors reported to be involved in hematopoiesis are thymosins. Well-known to be associated with tissue repair and cell migration through interaction with ATP-synthase in vertebrates, two β-thymosins have been found to participate in a complex way to regulate hematopoietic stem cell proliferation and differentiation in *P. leniusculus*, and to induce expression of Astakine-1. Both β-thymosin-1 and β-thymosin-2
stimulate migration of cells in vitro, but β-thymosin-2 does so in combination with Astakine-1. 59 β-Thymosin-1 also causes a transitory increase in the number of circulatory semigranular cells, whereas both thymosins induce the expression of Astakine-1. 59 Interestingly, the two thymosin isoforms suppress production of reactive oxygen species (ROS) by the respiratory burst by enhancing superoxide production, 59 so it is interesting to speculate whether these findings are indicative of some kind of selection process within the hematopoietic tissue, with ROS used as a means of deleting superfluous or deleterious hemocytes.

3 MOLLUSCAN HEMATOPOIESIS

With more than 100,000 living species, the Mollusca is the second largest phylum of metazoans after the Arthropoda. The group contains species with wide variation in body plans, physiology, and anatomical specialization. They may be present in freshwater, the sea, or terrestrial environments 60 and show diversity in reproductive strategies and life histories. However, the immune systems of this group have received rather less research attention than that given to arthropods, with most directed at economically important bivalves and gastropods, and, to a lesser extent, cephalopods. These taxa are sometimes grouped together, along with scaphopods, in the monophyletic subphylum Conchifera. 60 To the best of our knowledge, almost nothing is known about the hemocytes or their production and maturation in members of the lesser-known groups of Solenogastres, Caudofoveata, Polyplacophora, or scaphopod. Research on the immune systems of Conchifera dates back to the early 1970s and has been done on a wide range of species. Unfortunately, there is no commonly adopted model (such as D. melanolobaster) that has received intensive research underpinned by full genome-sequence data. However, recent sequencing of the genomes from the oyster, Crassostrea gigas, 61 and the owl limpet, Lottia gigantea, 62 together with the developing of transcriptome databases, for example, for the mussel, Mytilus galloprovincialis, 63 and for the apple snail, Pomacea canaliculata, 64 should prompt new studies that will enable integration of morphological, physiological, and molecular findings.

3.1 Bivalves

As with pancrustaceans, bivalves and gastropods have open circulatory systems and possess populations of circulating hemocytes involved in host defense. Although there is a respectful body of literature on the morphology and functionality of these immune cells (see Section 5.3), hematopoiesis is very poorly characterized. It has been generally thought that hemocytes originate from the connective tissue and/or mantle, 9,65,66 and although some may mature before entering the circulation, 67–70 evidence is accumulating that mitosis may also occur after hemocyte release to the hemolymph, as demonstrated by the clam, Tapes philippinarum. 71 These clam cells are positive for the stem cell marker, CD34. 71 Also in oysters, a family of transcription factors Cg-tal, homologous to
Tal1/SCL, has been reported to be expressed during ontogeny of *C. gigas*.[72,73] This family of transcription factors is known to be involved in the hematopoietic process of vertebrates.[74] Challenge of oysters with *Vibrio* spp. has further been reported to induce the production of new hemocytes, and to upregulate the expression of certain genes, such as *Cg-runx1* and *Cg-BMP7*, which encode proteins involved in hemocyte proliferation.[75] Another recent study has provided some indication that at least some prohemocytes emerge from the basement of the gill epithelium in oysters,[10] which bears out the early proposition by Cuenot that bivalve hemocytes originate in the gills.[76] It is also noteworthy that molecular analyses of developing *C. gigas* juveniles have indicated that hematopoiesis-related factors are expressed in the hemocytes rather than in the connective tissue, suggestive that, as in pancrustacean models, the hematopoietic cells could derive from the vessel and/or artery endothelial cells.[72] In the scallop, *Chlamys farreri*, several molecules associated with the proliferation and differentiation of prohemocytes have been identified and characterized from the circulating hemocytes. These include Cf-Ets,[77] Cf-Runt, and Cf-CBFβ.[78] *Cf-runt* is similar to the mammalian *runt* gene, so it is possible that the encoded protein is necessary for hematopoiesis, and the recruitment of new hemocytes to the circulation after immune challenge.[78] Runt is a conserved family of transcription factors that has also been identified in crayfish.[53] A similar transcription factor, named *lozenge*, is present also in *D. melanogaster*, and is believed to control crystal cell commitment during embryogenesis.[6] These molecular studies are providing evidence that progenitors of fully differentiated hemocytes can occur in the circulation of bivalves, where presumably they continue to mature; but much remains unknown. This evidence also supports the hypothesis that in bivalves the proliferation of circulating hemocytes may occur as a consequence of an immune challenge, as already observed in insect larvae and crustaceans.[6,79] In the near future, information from genome sequencing of *C. gigas* or other bivalves should enable clarification of the contradictory views and permit a better understanding of the relationship between the different hemocyte lineages.

### 3.2 Gastropods

Hematopoiesis occurs in gastropods in specific and well-defined tissues positioned close to the reno-pericardial zone.[80,81] There is no evidence of proliferation in the circulating hemolymph. In the freshwater snail, *Biomphalaria glabrata*, the hematopoietic tissue, described as the amoebocyte-producing organ (APO), constitutes a cluster of cells between the pericardium and the posterior mantle epithelia.[82] The hematopoietic cells are contained in the stroma, composed of small fibroblasts and smooth muscle cells surrounded by blood sinuses.[82] Immune challenge with the trematode parasite, *Schistosoma mansoni*, induces mitotic activity in this tissue,[83] and transplantation of the hematopoietic tissue from a *S. mansoni*-resistant strain to a susceptible strain confers resistance to *S. mansoni* in the new host that otherwise would succumb to the parasite.[84] In the freshwater gastropod,
The Evolution of the Immune System

Planorbarius corneus, the hematopoietic tissue is located between the mantle cavity and the pericardium, with mature hemocytes transported to the circulation via the blood sinuses. In P. canaliculata, new hemocytes seem to originate close to the heart, as a highly conserved mitotic marker, phospho-Ser10 histone H3, has revealed that positive cells (arrowheads) around the vein on the side facing the pericardial cavity; (C) higher magnification of (B) showing proliferative cells (arrowheads), giving rise to new hemocytes; (D) negative control. Bars: (A) = 120 μm; (B) = 100 μm; (C, D) = 40 μm.

Planorbarius corneus, the hematopoietic tissue is located between the mantle cavity and the pericardium, with mature hemocytes transported to the circulation via the blood sinuses. In P. canaliculata, new hemocytes seem to originate close to the heart, as a highly conserved mitotic marker, phospho-Ser10 histone H3, has revealed that positive cells are localized in the fluid that fills the pericardial cavity. Mitotic figures have been noted as adhering to the external epithelium of the ctenidial/pulmonary vein that faces the pericardial cavity (Fig. 1.1). In other gastropods the hematopoietic tissue is less defined, but in the lymnaeids and the land slug, Incilaria fruhstorferi, the formation of new hemocytes seems to happen either in the connective tissues or the vascular system, but, as yet, the process has not been fully described or characterized.

3.3 Cephalopods

Unlike all other mollusks, cephalopods have a closed circulatory system with high systemic blood-pressure and a high rate of delivered oxygen. The circulating hemocytes (once called leucocytes) are produced in the white body,
which is localized in the orbital pits of the cranial cartilages, behind the eyes and close to the optic ganglia.\textsuperscript{91} This organ has been described for three cephalopod species, namely \textit{Octopus vulgaris}, \textit{Octopus briareus}, and the squid, \textit{Sepia officinalis}. It usually takes the form of internal strings of precursors of hemocytes, surrounded by external connective tissue.\textsuperscript{12,92} These hemocyte precursors are generated by proliferating cells known as leucoblasts or hemocytoblasts, and are thought to migrate from the strings into the vascular sinuses for differentiation and maturation.\textsuperscript{12,92} Light and electron microscopy studies suggest that hemoblasts proliferate, transform, and mature in the white body.\textsuperscript{92} During maturation, the young cells enlarge and modify their nuclear organization and cytoplasmic organelles before entering the circulatory system as mature hemocytes.\textsuperscript{12} In \textit{O. vulgaris}, proliferative activity in the white body has been demonstrated both in vitro\textsuperscript{93} and in vivo.\textsuperscript{90} The majority of young cells resident in the white body are either proliferating or maturing cells that show immunocompetence, but their phagocytic capacity is lower than that of the circulatory hemocytes.\textsuperscript{90}

\section*{4 HEMATOPOIESIS IN PANCRUSTACEAN AND MOLLUSCAN MODELS: FINAL CONSIDERATIONS}

From the accounts above, it is clear that there are differences in the location and structure of the hematopoietic tissues and in the hematopoietic process between members of the pancrustaceans and the mollusks. In general, hematopoiesis in pancrustaceans occurs in tissues/organs associated with the main blood vessels, which in embryos are little more than cellular masses. In larvae through to adults, it takes place in more defined tissues, often in the pericardial region of the hemocoel, which has, in older papers, been described as the so-called “lymphoid organ.”\textsuperscript{44} In these tissues there are layers of hemocytes at different stages of maturation that are released into the open circulatory system upon maturation. The release of hemocytes is not continuous, but dependent on either developmental or immunological cues. By contrast, in mollusks, the hematopoietic tissues tend to be more varied, with new hemocytes produced continuously throughout the animal’s lifespan, albeit with some upregulation following non-self stimulation. The best studied molluscan groups are bivalves and gastropods, and, in the latter, the APO or its equivalent occur in the pericardial region, a position comparable to that of the decapod hematopoietic tissue.

\section*{5 HEMOCYTES IN PANCRUSTACEAN AND MOLLUSCAN MODELS}

The finding of conserved elements between pancrustacean and molluscan model species is significantly harder when it comes to the description of hemocyte types. Insect, crustacean, and molluscan hemocyte types and morphologies have been reviewed several times,\textsuperscript{2,6,68,69,94–97} but differences in the naming and classification of the various hemocyte types are so numerous as to make it
almost impossible to present a generalized and widely accepted system. Here, a
general morphological description, considering the main functions of the cells,
will be presented with some consideration given to their likely lineages.

5.1 Hemocytes and Their Lineages in Insects

In 1981, Rowley and Ratcliffe\textsuperscript{2} tried to rationalize the nomenclature of insect
hemocyte types and identified six main types, although not all occur in every
species. The hemocyte types they defined are: prohemocytes, plasmatocytes,
granular cells, cystocytes, spherule cells, and oenocytoids\textsuperscript{2} (Table 1.1).

\textit{D. melanogaster} is generally regarded as having just three hemocyte types
both its larval and adult stages. According to Williams\textsuperscript{98} these are, in order of
decreasing abundance: plasmatocytes, crystal cells, and lamellocytes (Fig. 1.2). This
seems to be a derived condition because other dipterans, such as \textit{Calliphora}
vomitoria, also possess oenocytoids.\textsuperscript{99} In addition to these cells, some authors
have reported the presence of small and round cells, considered to be immature
hemocytes, and so are called prohemocytes.\textsuperscript{97,100} However, the presence of pro-
hemocytes in the circulation of adult \textit{Drosophila} spp. is at odds with the view
of other authors, in that hemocyte proliferation and differentiation occur only
during embryo and larval development.\textsuperscript{1,6} Plasmatocytes are the dominant cell
type in \textit{Drosophila} spp., but they may transform into lamellocytes upon para-
sitoid infection.\textsuperscript{101} \textit{Drosophila} plasmatocytes are agranular and have consider-
able phagocytic capacity, whereas the lamellocytes, which remain agranular,
are larger and nonphagocytic. Instead they adhere and contribute to encapsula-
tion of parasites that may gain entry to the larval body.\textsuperscript{1,6,100–102} By
contrast, crystal cells, which contain granules of the proPO activating system,
do not display significant morphological changes after immune challenge\textsuperscript{98} and
show no phagocytic activity.\textsuperscript{1,6,100} It is not clear if crystal cells can be consid-
ered to be derived from a different cell lineage from plasmatocytes.\textsuperscript{6,100–102}

In \textit{D. melanogaster} embryos, the hemocytes that move along fixed routes
and invade the developing larva are mainly plasmatocytes (approximately 95%),
along with a few crystal cells.\textsuperscript{96} Usually lamellocytes are absent in embryos.\textsuperscript{19} As
described in Section 2.1, in the absence of an immune challenge the circulating
hemocytes observed in late embryos are evident in the hemolymph of \textit{D. melanogaster}
larvae, along with a few lamellocytes.\textsuperscript{103} Immature cells occupy the medul-
lar zone of the primary lobes of the lymph gland in \textit{Drosophila}, while maturing
plasmatocytes occupy the cortical zone together with a minority of crystal cells, as
revealed by the expression of a number of specific cell markers.\textsuperscript{19} At present, there
is no evidence supporting the existence of two separate hemocyte lineages in the
lymph gland of \textit{Drosophila}.\textsuperscript{6,102,103} (Fig. 1.3A). It should be noted that larval hematopoiesis occurs also in specific niches along the larval body, producing plasmatocytes only.\textsuperscript{28} Data obtained on \textit{D. melanogaster} hemocytes and their development
are of great value for understanding hematopoiesis in fruit flies, but unfortunately
the findings do not necessarily apply to other flies, even of the same genus.\textsuperscript{97}
TABLE 1.1 Morphological and Functional Features of the Six Main Circulating Hemocyte Types Identified Among Insect Orders Following the Classification by Rowley and Ratcliffe (1981)²

<table>
<thead>
<tr>
<th></th>
<th>Prohemocyte</th>
<th>Plasmatocyte</th>
<th>Granular Cell</th>
<th>Cystocyte</th>
<th>Spherule Cell</th>
<th>Oenocytoid</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Size</strong></td>
<td>6–13 µm</td>
<td>10–15 µm</td>
<td>8–20 µm</td>
<td>8–15 µm</td>
<td>8–16 µm</td>
<td>12–30 µm</td>
</tr>
<tr>
<td><strong>Shape</strong></td>
<td>Round to oval, high N/C ratio</td>
<td>Round to spindle shaped</td>
<td>Round to oval</td>
<td>Round, high N/C ratio</td>
<td>Oval to spindle shaped</td>
<td>Round to oval</td>
</tr>
<tr>
<td><strong>Inclusions in the cytoplasm</strong></td>
<td>A few granule-like inclusions</td>
<td>None/moderate/large number of granules</td>
<td>Many granules with variable structure</td>
<td>Many granules with variable structure</td>
<td>Large and spherular inclusions</td>
<td>Homogeneous cytoplasm</td>
</tr>
<tr>
<td><strong>Percentage</strong></td>
<td>Less than 5%</td>
<td>30–60%</td>
<td>30–60%</td>
<td>30–60%</td>
<td>Less than 5%</td>
<td>1–2%</td>
</tr>
<tr>
<td><strong>Reaction in vitro</strong></td>
<td>Highly stable, formation of small protoplasmic extensions</td>
<td>Stable, formation of protoplasmic extensions and ameboid movement</td>
<td>Unstable, degranulation with formation of vacuoles</td>
<td>Unstable, quick degeneration, degranulation, increase of nucleus refractivity, formation of “islet of coagulation”</td>
<td>Highly stable, formation of rare protoplasmic extensions</td>
<td>Highly stable, no formation of protoplasmic extensions</td>
</tr>
<tr>
<td><strong>Function in vivo</strong></td>
<td>Mitotic division, direct differentiation to plasmatocytes</td>
<td>Phagocytosis, encapsulation, basement membrane formation, wound repair, nodule formation, metamorphosis</td>
<td>Phagocytosis, nodule formation, encapsulation, melanin formation, hemolymph coagulation, wound repair, hormone transport</td>
<td>Coagulation, encapsulation, nodule formation, and phagocytosis</td>
<td>Uncertain. In B. mori they are involved in silk production</td>
<td>Uncertain. They may be involved in melanization process and reactions to parasites</td>
</tr>
</tbody>
</table>

(Continued)
### Table 1.1 Morphological and Functional Features of the Six Main Circulating Hemocyte Types Identified Among Insect Orders Following the Classification by Rowley and Ratcliffe (1981)² (cont.)

<table>
<thead>
<tr>
<th>Orders in which the hemocyte type has been described</th>
<th>Prohemocyte</th>
<th>Plasmatocyte</th>
<th>Granular Cell</th>
<th>Cystocyte</th>
<th>Spherule Cell</th>
<th>Oenocytoid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blattoidea Coleoptera Collembola Diptera Hemiptera Hymenoptera Lepidoptera</td>
<td>Blattoidea Coleoptera Collembola Diptera Hemiptera Hymenoptera Lepidoptera Orthoptera</td>
<td>Blattoidea Coleoptera Collembola Diptera Hemiptera Hymenoptera Lepidoptera Orthoptera</td>
<td>Blattoidea Coleoptera Collembola Diptera Hemiptera Hymenoptera Lepidoptera Orthoptera</td>
<td>Blattoidea Coleoptera Collembola Diptera Hemiptera Orthoptera Phasmatodea</td>
<td>Coleoptera Collembola Diptera Hemiptera Lepidoptera Orthoptera</td>
<td>Coleoptera Collembola Diptera Hemiptera Hymenoptera Lepidoptera Orthoptera</td>
</tr>
<tr>
<td>Synonyms</td>
<td>Proleukocyte Proleukocytoid</td>
<td>Amoebocyte Leukocyte Macronucleocyte Macrophage Phagocyte</td>
<td>Granular leukocyte</td>
<td>Coagulocyte Explosive corpuscle</td>
<td>Adipohemocyte Coarsely granular cell Eruptive cell Eleocyte Spherulocyte</td>
<td>Crystal cells in dipterans may be a special type of oenocytoid Oenocytes</td>
</tr>
</tbody>
</table>
Lepidoptera is the most studied insect group after the Diptera, and the information on their hemocytes is extensive. In lepidopteran larvae, the most common cell type is the granulocyte, or granular hemocyte. These cells have high phagocytic ability, and so are functionally similar to plasmatocytes in Drosophila, but, unlike fruit fly plasmatocytes, lepidopteran granulocytes contain numerous cytoplasmic granules. Another difference between Drosophila and lepidopterans is that plasmatocytes in lepidopterans are agranular and seem to be functionally comparable to Drosophila lamellocytes, as their main function is encapsulation of parasitoids or pathogens too big for being phagocytosed. However, lepidopteran plasmatocytes are also able to engulf large particles in vitro, indicating that they might display some phagocytic activity also in vivo. A third type of hemocyte observed in lepidopteran larvae is the oenocytoid, a cell functionally equivalent to Drosophila crystal cells, as it contains granules of proPO. A fourth, and so far uncharacterized, hemocyte type present in lepidopterans is the spherulocyte (Table 1.1). Its function is thought to be associated with transport of cuticular components, but this has yet to be fully confirmed. Finally, mention needs to be made of reticular cells (or reticulocytes), whose role in insect hemocyte development has yet to be
The Evolution of the Immune System

Reticulocytes are found in several insect species, including lepidopteran or hemimetabolan insects, such as *L. migratoria*. Reticulocytes are not circulatory hemocytes and are found only in the lymph gland. Grigorian and Hartenstein have suggested that they may represent immature hemocytes able to mature into the other hemocyte types. Mandal et al. however, consider that reticulocytes may play a regulatory role in hemocyte proliferation and differentiation in hematopoietic organs, similar to the PSC cells in the *D. melanogaster* lymph gland.

Although the existence of separate hemocyte lineages in the *D. melanogaster* lymph gland has not been established, (Fig. 1.3A), lepidopteran
circulatory hemocytes have been proposed recently to develop from two separate cell lineages. By injecting prohemocytes obtained from a GFP-expressing *B. mori* lymph gland into wild-type larvae, three different fluorescently labelled cell types (granulocytes, plasmatocytes, and oenocytoids) can be observed at the end of the experiments. In vitro experiments demonstrate that, in the silkworm, mature plasmatocytes can develop into oenocytoids, which is not the case for mature granulocytes. These findings point to the existence of two hemocyte lineages in *B. mori*: the first one differentiating into granulocytes, and a second one differentiating into plasmatocytes and the plasmatocyte-derived oenocytoid (Fig. 1.3B). That both plasmatocytes and granulocytes in lepidopterans can divide in circulation has been shown by bromodeoxyuridine (BrdU) incorporation experiments and by the use of monoclonal antibodies. However, the results supporting the existence of two hemocyte lineages in silkworms are in disagreement with the observations of Beaulaton, who reported that hemocytes in *B. mori* larvae are all derived from the plasmatocytes.

The classification of hemocyte lineages in insects is further complicated by the transformation of one hemocyte type into another one (Fig. 1.3). The interconversion between cell types in insects has been reported several times, and further studies, focused on specific cell markers, are necessary to elucidate distinction of cell types and lineages. It has also been suggested that the nomenclature of the insect needs to be updated in order to avoid misunderstandings or misleading synonyms, for example, plasmatocytes in *Drosophila* are hemocytes with different morphology and functions than plasmatocytes in other insects (Table 1.1). This revision of hemocyte nomenclature would seem wise, but at present most authors adhere to the original names. On the basis of the descriptions presented in this chapter, larvae of flies and butterflies possess three main functional families of circulating hemocytes: one has high phagocytic activity, the second is able to encapsulate large pathogens, and the third is involved in ROS production and activation of the proPO cascade (Fig. 1.3). These three hemocyte groups are usually all present in the hemolymph and cooperate in the insect immune response, although their proportion may change in relation to the immune challenge considered. It remains to be established if three functional groups in other insects also stem from separated hemocyte lineages, or if they all derive from the same precursor through differentiated processes of maturation. The possible interconversion between cell types suggests that they may derive from the same embryonic/larval stem cell, though by different and species-specific pathways.

5.2 Hemocytes and Their Lineages in Crustaceans

Nearly all studies on crustacean immunity have been on decapods, particularly shrimp, crab, and lobster. Across these groups, a common pattern emerges of at least three main cell types, distinguishable morphologically by their degree of granulation (Fig. 1.3C). They can also be physically separated by density gradient
and distinguished by monoclonal antibody markers, flow cytometry, or activated cell sorting. The three main hemocyte types are termed hyaline cells, semigranular cells, and granular cells, with, in some species, a small population of prohemocytes also present. The prohemocytes are relatively rare (about 2% or fewer of the total number of hemocytes in the circulation), but are distinguishable by their small size, spherical shape, and large nuclear to cytoplasm ratio. Subdivisions of large and small granular cells have also been noted occasionally in the literature but it is unclear if these represent truly distinct functional categories. In shrimp, the dominant cells are the semigranular and granular cells, with the hyaline cells relatively uncommon in the circulation. A similar situation occurs in crayfish and lobster, and in all these decapods the semigranular cells are generally accepted as being those most responsible for phagocytosis. However, in crab the hyaline cells are dominant, and are the ones that are chiefly responsible for phagocytosis. Across all decapods studied, the granular hemocytes appear to have no phagocytic capability, but, together with the semigranular cells and hyaline cells (if present), they contribute to encapsulation reactions. The granular cells, and to a lesser extent the semigranular cells, are also the main repositories of important immune proteins, such as antimicrobial proteins or factors involved in activation of proPO. It is interesting to note that, at least in crab, only the hyaline cells and the semigranular cells exhibit the phenomenon of ETosis, where ET stands for extracellular traps. This is a relatively recently discovered cell-death pathway that aids entrapment and clearance of infective agents from the blood. The process was first reported for mammalian neutrophils, but is now known to be an ancient process that entails the expulsion of decondensed chromatin from the nucleus to the extracellular environment in a controlled and regulated way. The released nuclear material forms a mesh that becomes studded with antimicrobial proteins derived from the granules of the disrupted cell, which entraps bacteria or other microorganisms, preventing their spread around the body. More importantly, the entrapped microbes are brought into close contact with the antimicrobial proteins on the chromatin fibers and are killed. In crab, neither the granular cells nor the prohemocytes show the response, offering further evidence that two cell lineages exist in decapods. Whilst bivalve hemocytes are also able to undergo this cell death process, it is unclear whether cytosolic or nuclear proteins released from ETotic cells stimulate hematopoiesis to replace lost cells. In keeping with their immature status, crustacean prohemocytes in general tend to show a lack of immune reactivity, as they also appear to be incapable of phagocytosis or a respiratory burst, and do not seem to contain antibacterial proteins (Smith, VJ, unpublished observations). The prohemocytes are, however, able to undergo cell division in the circulation in vivo. Indeed, very low numbers of mitotic cells have been noted in the circulation of shrimp and other decapods, but it is still unclear whether these are prohemocytes, or a rare, previously unrecognized type of mature hemocyte.
The lineages to which the various hemocyte types belong remain uncertain; van de Braak et al.\textsuperscript{42} consider that two lines of cells are produced independently from the hematopoietic stem cells in \textit{P. monodon}. One line is proposed to develop into the granular cells through a sequence of intermediary semigranular cell stages, with the final maturation step taking place in the connective tissue.\textsuperscript{42} The other cell lineage is postulated to develop from hyaline cells into a separate semigranular cell type\textsuperscript{42} (Fig. 1.3C). Based on histological and ultrastructural analysis of the hematopoietic tissue in Chinese shrimp, \textit{Fenneropenaeus chinensis}, Zhang et al.\textsuperscript{46} concur with the notion that hemocytes originate from two separate lineages within the hematopoietic tissue. The finding that similarly sized prohemocytes, enriched in the hemolymph from crab, \textit{Hyas arenaeus}\textsuperscript{54} and \textit{Carcinus maenas} (Smith, VJ, unpublished observations), may be either granulated or agranular, lends some further credence to this idea.

5.3 Hemocytes and Their Lineages in Mollusks

Molluscan hemocytes have been described mainly in bivalves, with a few studies made on gastropods and cephalopods. However, the absence of available molecular databases means that classification of the cells remains based mainly on morphological, cytochemical, and functional characteristics, without the possibility to define separate cell lineages. Key features that define the cells include the presence or absence of cytoplasmic granules; acidic, basic, or neutral staining of the granules; phagocytic activity.

5.3.1 Bivalves

Bivalves generally possess three discernable hemocyte morphologies: hyaline cells (sometimes also described as hyalinocytes or agranulocytes), which lack conspicuous cytoplasmic granules; granular cells (also described as granulocytes)\textsuperscript{9,129,130}; and small agranular cells, possibly prohemocytes (also known as hemoblast-like cells).\textsuperscript{69,70,131,132} In the oysters, \textit{Crassostrea virginica} and \textit{Crassostrea rhizophorae}, and immediately in the mussel, \textit{Perna viridis}, prohemocytes may comprise 3–10\% of the total circulating cell number.\textsuperscript{69,70,131,132} In other species, these are either absent, or so rarely seen that they have been overlooked by researchers.\textsuperscript{67} The proportion of hyaline and granular cells varies from species to species, but wide fluctuations in hemocyte subpopulations may occur in response to different bacterial challenges.\textsuperscript{133–135} For example, in \textit{C. virginica} and \textit{C. rhizophorae}, the hyaline cells have been reported to comprise 70–80\% of the hemogram\textsuperscript{70,131} with high values also reported for the clam, \textit{Chamelea gallina}, the brown mussel, \textit{Perna perna}, and the scallop, \textit{Argopecten irradians}.\textsuperscript{67,69} The remaining proportion of the hemocyte pool is made up of granular cells irrespective of size or staining properties of the granules.\textsuperscript{67,69,70,131} In other bivalve species, such as in the blue mussel, \textit{Mytilus edulis}; the cockle, \textit{Cerastoderma edulis}; and the razor shell, \textit{Ensis siliqua}, the proportion of hyaline cells is much lower (6–16\%), with
basophilic and acidophilic granular cells dominant.\(^{134}\) *C. edulis* is unusual in also possessing large, vacuolated, spherical agranular cells.\(^ {134}\) Other bivalves with a high population of granular hemocytes include *M. galloprovincialis*, *Cerastoderma glaucum*, *P. viridis*, and the clams, *Ruditapes decussatus* and *T. philippinarum*.\(^ {67,69,136}\)

In some species phagocytic activity is usually carried out by the dominant cell type, whether it be the hyaline cells, as in oysters,\(^ {67,69,70,131}\) or the granular cells, as in blue mussel and other bivalves.\(^ {68,69,133,134}\) Thus, phagocytosis may not be executed by exclusively either hyaline or granulated hemocytes in bivalves, but it is species-specific. Rather, phagocytic activity seems to be associated more with the presence in the hemocytes of AMPs,\(^ {137}\) lysozyme, esterases, and/or an ability to undergo the respiratory burst.\(^ {68,69}\) Concerning *P. viridis*, *M. edulis*, and the giant clam, *Tridacna crocea*, authors\(^ {69}\) have further reported that hyaline hemocytes have no phagocytic ability, but they play a central role in coagulation, and could contribute to limiting blood loss and healing wounds.

With regard to the relationship of the different hemocyte types to each other in bivalves, much remains unclear. Few studies have identified cell-specific markers, although monoclonal antibodies have been developed to distinguish the various hemocyte populations of *M. edulis*.\(^ {138,139}\) Ottaviani et al.\(^ {140}\) and, subsequently, Rebelo and coworkers\(^ {70}\) suggest that bivalves possess only one cell lineage, with the different morphologies representing various maturational stages and/or degranulation events. However, other researchers\(^ {9,129}\) support the hypothesis that granulocytes and hyalinocytes are two separate cell lineages with different immunological roles. The general absence of cell-specific molecular markers confounds testing these conflicting views. Notwithstanding, Dyrynda et al.\(^ {139}\) were able to trace hemocytic changes during ontogenic development of *M. edulis*, using a panel of monoclonal antibodies. Interestingly, the antibodies raised against hemocytes from mature adult mussels also show reactivity with cells from both the trochophore and veliger larvae of *M. edulis*, indicating that epitopes on adult mussel hemocytes are also present at much earlier stages in the life.\(^ {139}\)

As of yet, it is still unknown whether *M. edulis* hyaline cells are immature hemocytes (perhaps equivalent to prohemocytes), or if they are fully and terminally differentiated cells. Further studies on the hematopoietic tissue, and prohemocytes capitalizing on new genome sequence data,\(^ {61,63}\) should reveal more about the origin and development of hemocyte stem cells, thus shedding light on how hemocytes in this hugely economically important group of mollusks mature and relate to each other.

### 5.3.2 Gastropods

Gastropods possess three types of hemocytes: agranular or hyaline cells (sometimes termed amoebocytes), granular hemocytes, and small spherical cells with big nuclei that may represent immature cells (ie, prohemocytes)\(^ {80,81}\) (Fig. 1.4). However, variations occur in functionality of different morphological types. For example, in the freshwater snails, *B. glabrata*, *Biomphalaria straminea*,
and abalone, *Haliotis tuberculata*, the hyaline cells are the most abundant in the hemolymph. In *Biomphalaria* spp. the different roles of the hyalino- cytes and granulocytes during the immune response are not completely clear. They both show acid phosphatase and peroxidase activities, but the main phagocytic activity seems to be played by the fixed phagocytes with a developed lysosomal apparatus in connective tissues in several organs. In abalones *H. tuberculata* and *Haliotis discus discus*, granulocytes are almost absent and the hyaline cells exhibit adhesion, aggregation, and phagocytic activities. In the pond snail, *Lymnaea stagnalis*, both hyaline cells and granular cells are present, and phagocytosis mainly involves the granular cells. Recently, light and electron microscopic studies, together with flow cytometry, has shown that the apple snail, *P. canaliculata*, possesses three distinct populations of hemocytes, namely, the small cells resembling the prohemocytes, the agranular cells, and the granular cells (Fig. 1.4). All, except the prohemocyte-like cells, are capable of phagocytosis. Lack of specific cell markers for these cells is the principal reason for the limited number of studies correlating the immune functions with hemocyte subpopulations, and limits our understanding of the lineages of gastropod hemocytes.

### 5.3.3 Cephalopods

The classification of hemocyte types in cephalopods is uncertain. Early studies considered that these invertebrates possess only one type of mature hemocyte, but a more recent study has identified two populations in octopus,
O. vulgaris. The dominant population is the large granulocyte, with a U-shaped nucleus and basophilic granules in the cytoplasm, whereas the remainder are hemocytes with a round nucleus, occupying almost the entire cell and few granules in the cytoplasm. Immune reactivity, such as phagocytosis and ROS and nitric oxide production, are more evident in the large granulocytes than in the smaller circulating cells. The morphology of the small agranular cell type, coupled with its apparent lack of immune reactivity, indicates that it might be immature cells, equivalent perhaps to the prohemocytes in other pan-crustacean and molluscan species, rather than a terminally differentiated and fully functional defense cell. The two hemocyte morphologies described in O. vulgaris are not evident in all the octopus species. In the octopus, Eledone cirrhosa, only granulocytes with phagocytic, bactericidal, and ROS production ability are present. Indeed, a study on S. officinalis has reported the existence of only one cell type that has phagocytic capacity. It has a variable number of eosinophilic granules, and some vesicles containing acid phosphatase, lysozyme, and proPO. As reported for bivalves, it remains to be established if in cephalopods the large granulocytes and the small agranulocytes derive from separate and postmitotic lineages.

With regard to functionality, early studies reported that cephalopod hemocytes in the circulation show relatively low levels of phagocytosis, at least compared to other molluscan species, and have been considered, by at least some authors, to have less of a role in inflammation than in wound healing. Certainly it has been known for some time that cephalopod hemocytes migrate to wounds and assume a fused shape, forming a barrier that protects the damaged area from infection. However, transcriptomic and proteomic research has established that circulating hemocytes in cephalopods do express a number of immune-related genes and proteins. Unfortunately, despite their economic importance, the immune system of cephalopods has received considerably less research interest than other mollusks.

6 CONCLUDING REMARKS

Although the main taxa discussed in this chapter are widely separated in evolutionary terms, and, as a consequence, have great diversity in form and life history, some trends are apparent. All possess hematopoietic tissue of mesodermal origin that is in close contact with the hemal cavity. These tissues generally comprise layers of undifferentiated, developing, and fully or semimature cells, although, sometimes, maturing cells may be present in the circulation, probably through internal or external stimuli. As yet, our understanding of the specific signals and pathways involved in the initiation and regulation of hematopoiesis remains poor. Notwithstanding, increasing availability of genomic, transcriptomic, and proteomic data should enable us to unlock these processes. Of particular importance is to understand if and how the generation, timing, and release of new hemocytes may be influenced...
by different types of stimuli. Does the phenotypic character of the emergent cells (eg, in terms of their repertoire of cell surface pattern-recognition receptors) vary among different types of pathogen-associated molecular patterns, in order to respond more efficiently to continued threats from invasive microorganisms?

Likewise, there is a dearth of information about the factors that determine what type of cell the stem and prohemocytes mature into. Certainly, members of the pancrustaceans and mollusks share a number of commonalities in the morphology and functionality of their mature hemocytes, although very little is known about the cell lineages, how they age, and the longevity of the different cell types in the hemocoel. An even more intriguing question is the pluripotency of their stem cells. To what degree can such cells generated in the hematopoietic tissue transform into other cells, for instance, neural cells? Such a scenario is not as implausible as once would have been thought, following the discovery that adult-born neurons in crayfish, *P. leniusculus*, can be derived from hemocytes. Invertebrate blood cells and immune systems continue to surprise us.

**ACKNOWLEDGMENTS**

Part of this work was funded by Grant A.006@FONDRESNOM@02BI-MALAGBIO (Department of Life Sciences, University of Modena and Reggio Emilia, Modena, Italy) and Fondazione di Vignola (Vignola, MO, Italy) grant to DM. Thanks goes to Dr Elisabeth Dyrynda (Heriot Watt University, Edinburgh, Scotland) for helpful discussions on the manuscript.

**REFERENCES**


125. Robb CT, Dyrynda EA, Gray RD, Rossi AG, Smith VJ. Invertebrate extracellular phagocyte traps show that chromatin is an ancient defence weapon. *Nat Commun* 2014;5:4627.


Chapter 2

Origin and Functions of Tunicate Hemocytes

Francesca Cima, Nicola Franchi, Loriano Ballarin

Department of Biology, University of Padova, Padova, Italy

1 INTRODUCTION

Tunicates, or urochordates, are a subphylum of the phylum Chordata, sharing with other members of the phylum: (1) a permanent or temporary notochord, in the form of a dorsal rod; (2) a central nervous system, in the form of a dorsal tube; (3) a pharynx provided with gill slits or pharyngeal pouches, and a ventral gland secreting iodoproteins (endostyle or thyroid); and (4) a muscular tail.

Tunicates are considered the sister group of vertebrates, forming with the latter the clade Olfactoria. Recently, it has been proposed to classify them as a phylum within the superphylum Chordata. They are traditionally subdivided in three classes: (1) Ascidiacea (benthic and sessile), (2) Thaliacea, and (3) Larvacea or Appendicularia (pelagic).

Ascidians have a free-swimming, tadpole-like larva, an adult sac-like body with two siphons that allow water flux, and a large branchial basket, provided with a ventral endostyle that secretes the mucous net required for filtration. They comprise two orders: Enterogona (including the suborders Phlebobranchia and Aplousobranchia) and Pleurogona (with the suborder Stolidobranchia). Thaliaceans include three orders: the colonial Pyrosomida, and the solitary/colonial Doliolida and Salpida. They have a barrel-like adult body, and, with the exception of Doliolida, are devoid of larval stages. Larvaceans or appendicularians resemble the ascidian larvae and use the tail to create the water current for filtration; filters are included in the gelatinous house secreted by the animals themselves. Most of the recent authors consider larvaceans as a sister group of the other tunicates, and thaliaceans as a sister group of Enterogona (Fig. 2.1).

Ascidians include about 2300 species, and most of the information on tunicate hemocytes comes from studies on this group of organisms. This review will, then, focus mainly on ascidian hemocytes and will discuss their role in immunity. Where possible, information on circulating cells of pelagic tunicates will be added.
2 ASCIDIAN CIRCULATION

Ascidians have an open circulatory system with a colorless hemolymph that is forced inside blood sinuses and lacunae within the body tissues by the contractions of a tubular hearth. The direction of the peristaltic waves reverses periodically with the consequent reversal of the hemolymph flow.\textsuperscript{11,12} 

The colorless hemolymph is isotonic with seawater and contains low concentrations of proteins, which are mainly secreted by the hemocytes: they exert important roles in humoral-defense reactions.\textsuperscript{12–14}
3 ASCIDIAN HEMOCYTES

Ascidian hemocytes are involved in a variety of functions, such as: (1) storage and transport of nutrients and catabolites, (2) asexual reproduction, (3) tunic synthesis, (4) allorecognition, and (5) defense reactions.\textsuperscript{15–17}

3.1 Hematopoietic Tissues

Hematopoietic sites are mainly located in peripharyngeal regions, clearly evident in solitary species as “lymphatic nodules,” or clusters of stem cells surrounded by hemocytes at various developmental stages.\textsuperscript{12,18,19} In colonial species of the genus \textit{Botryllus}, mitotic figures have been occasionally observed in circulating hemocytes.\textsuperscript{20} In the same species, stem cell niches have been identified close to the adult endostyle.\textsuperscript{21–23} From this location, they colonize the cell islands, that is, aggregates of hemocytes located along the sides of the endostyle, from which they enter the circulation and contribute to various zooid tissues, germline and hemolymph included.\textsuperscript{23}

3.2 Hemocyte Morphology

Many investigators studied the morphology of living and fixed hemocytes and their ultrastructure.\textsuperscript{16,17} Various authors have also proposed unifying classification frameworks,\textsuperscript{11,12,24} but a certain degree of uncertainty still persists on the terminology and differentiation pathways of ascidian hemocytes. In our opinion, they can be grouped in the following categories: undifferentiated cells, immunocytes (i.e., cells involved in immune defense), and storage cells (Fig. 2.2).

Undifferentiated cells, known as hemoblasts, feature a high nucleus/cytoplasmic ratio, a well-defined nucleolus, and a basophilic cytoplasm (Fig. 2.2A). The now obsolete term “lymphocyte” has also been used in the literature, both as a synonymous of the term “hemoblast” and to indicate a young hemocyte having entered a differentiation pathway.\textsuperscript{18,25,26} In \textit{Botryloides leachii} and \textit{Botryllus schlosseri}, hemoblasts show positivity to antibodies against mammalian CD-34.\textsuperscript{16,23} In \textit{B. schlosseri}, new hemoblasts cyclically enter the circulation to replace the hemocytes that undergo apoptosis at the generation change.\textsuperscript{27}

Circulating immunocytes include both phagocytes and cytotoxic cells. Phagocytes can assume both a spreading and a round morphology. Spreading phagocytes are also known as hyaline amoebocytes: they are multifunctional cells that can actively move toward foreign cells or particles and ingest them. Their cytoplasm contains fine cytoplasmic granules, unresolvable under the light microscope, and shows positivity for lysosomal enzyme activities\textsuperscript{15,17,28} (Fig. 2.2B). Upon the ingestion of foreign material, they withdraw their projections, assuming a round morphology (in this phase, they are also called macrophages or macrophage-like cells). Round phagocytes are large cells, with one or more phagosomes (Fig. 2.2C). A detailed study of the properties of round
phagocytes in the colonial ascidians *B. leachii* and *B. schlosseri* confirmed the presence of hydrolytic enzymes, as well as of lipids and lipofuscins inside their phagosomes. In *B. schlosseri*, the presence of a static and a mobile population of phagocytes have been described, the former homing in ventral islands, on both sides of the endostyle.

**Cytotoxic cells** are vacuolated cells containing the enzyme phenoloxidase (PO). They constitute one of the most abundant (sometimes more than 50%) circulating hemocyte-types and are represented, in most cases, by morula cells (MCs), so-called for the berry-like morphology they assume after fixation (Fig. 2.2D–E). In the solitary ascidian *Ciona intestinalis*, PO-containing cells are represented by univacuolar refractile granulocytes, whereas in *Phallosia mammillata*, compartment cells and granular amoebocytes contain the enzyme. MCs have a large diameter (10–15 μm) and cytoplasm filled by many vacuoles, uniform in size (around 2 μm in diameter) where the enzyme PO, probably stored as an inactive precursor (proPO), resides. In botryllid ascidians, granular amoebocytes are considered to be the direct precursors of MCs, because they share with the latter similar cytochemical properties and common enzymatic content inside their granules.
Storage cells, that is, circulating hemocytes that do not exert immune-related functions, consist of vanadocytes, nephrocytes, pigment cells, and trophocytes. Vanadocytes, present in Enterogona, can accumulate vanadium (V) up to a concentration of 350 mM (the mean V concentration of seawater is 35 nM) inside their vacuoles, complexed with V-binding proteins. The function of such high V concentration is still unexplained.

Nephrocytes, 10–15 μm in diameter, accumulate urate crystals in Brownian motion inside large vacuoles (Fig. 2.2F). They represent a sort of “circulating kidney,” as tunicates lack a defined excretory system. Nephrocytes can leave the circulation and contribute to the ascidian pigmentation.

Pigment cells are large hemocytes (up to 40 μm in diameter) containing blue, orange, and/or reddish pigment granules, either in the form of cytoplasmic globules or as crystals in Brownian motion inside a few large vacuoles (Fig. 2.2G). Their frequency tends to increase with the age of the organism/colony.

Trophocytes represent the most abundant circulating cells of many colonial ascidians. They are large (10–15 μm in diameter), round cells with the cytoplasm filled with many granules, which store nutrients. They sustain bud morphogenesis, especially in species such as Polyandrocarpa misakiensis, where buds detach very early from the parental organism.

4 IMMUNE ROLE OF TUNICATE HEMOCYTES

4.1 Phagocytosis

Ascidians rely only on innate immunity for their defense, and phagocytes constitute a fraction of approximately 20% of the total immunocytes number. The ingestion of foreign materials occurs either in liquid phase (macropinocytosis) or through classical phagocytosis, which is a “zipper-like” process involving integrins. The recognition of nonself molecules by phagocytes requires the presence of surface receptors. Two typical Toll-like receptor (TLR) genes, composed of TIR, transmembrane, and LRR domains, have been identified in the solitary species C. intestinalis; they are expressed in both plasma membrane and the endosomes of phagocytes. Also in B. schlosseri, preliminary data indicate that TLRs are involved in phagocytosis.

Nonself recognition by phagocytes triggers evolutionary-conserved signal-transduction pathways, such as those mediated by trimeric G-proteins and the protein kinases PKA and PKC. In addition, active PI3K, MAPKs, NF-kB, and a transient increase of cytosolic calcium are required. In addition, genes putatively encoding for MyD88, IRAK, TRAF, NF-kB, and IkB, involved in the signal transduction triggered by activated TIR domains, have been identified in the Ciona genome, as well as a gene for CD36, which is a scavenger receptor able to recognize oxidized lipids and surface microbial molecules that are required for phagocytosis of apoptotic cells.
The ingestion of foreign materials is usually associated with the activation of both a membrane oxidase, with the consequent increase in oxygen consumption, and nitric oxide synthase (NOS); this leads to the production of reactive oxygen species (ROS) and reactive nitrogen species (RNS), with microbicidal function.\textsuperscript{55,56}

In \textit{B. schlosseri}, as a consequence of nonself recognition, phagocytes produce and release a rhamnose-binding lectin (BsRBL), able to coat the microbial surfaces and act as opsonin.\textsuperscript{57} The same lectin, when added to hemocyte cultures, can trigger the phagocyte respiratory burst and exert chemotactic activity toward phagocytes that assume a spreading morphology. Both the transcription of the BsRBL gene and the synthesis of the protein significantly increase during the colonial generation change, when a consistent fraction of circulating hemocytes (\(~30\%\)) undergoes cell death by apoptosis and is cleared by phagocytes.\textsuperscript{58}

In \textit{B. schlosseri}, phagocytosis of apoptotic cells occurs massively during the cyclical generation change\textsuperscript{59–61}; the recognition of dying cells is mediated by the exposure of phosphatidylserine and sialic acid, and involves molecules recognized by a polyclonal anti-CD36 antibody.\textsuperscript{61,62} The correct clearance of apoptotic cells by phagocytes is required for the completion of the generation change and progression of bud development to adulthood.\textsuperscript{63}

Moreover, in \textit{B. schlosseri}, a population of phagocytes is exposed to seawater and adheres to the internal tunic of the oral and atrial siphons. Such sentinel cells are active in immunosurveillance, controlling the entrance to branchial and atrial chambers.\textsuperscript{64}

Some tunic cells, probably derived from hemocytes, have phagocytic activity and contribute to protect organisms from pathogens or parasites.\textsuperscript{65–68}

### 4.2 Encapsulation

Circulating phagocytes usually encapsulate any foreign material too large to be ingested by a single cell. The formation of capsule around parasitic crustaceans has been reported in \textit{Ascidiella aspersa},\textsuperscript{69} \textit{Microcosmus savignyi},\textsuperscript{70} and \textit{Styela gibbsii}.\textsuperscript{71} In \textit{Molgula manhattensis}, both phagocytes and MCs are recruited in the infectious area during capsule formation.\textsuperscript{72} In \textit{C. intestinalis}, the injection of mammalian erythrocytes, BSA, or LPS in the tunic leads to capsule formation following the massive recruitment of hemocytes to the inoculum site.\textsuperscript{73–75}

\textit{Botryllus scalaris} is the only botryllid species reported so far in which capsule formation by circulating phagocytes is involved in allorecognition between contacting, incompatible colonies. In this case, after the fusion of the ampullae and the initial blood-exchange between contacting colonies, circulating phagocytes crowd inside the fused vessels and stimulate the aggregation of hemocytes into large clusters, which are encapsulated by other phagocytes. In this way, the hemocyte-mass plugs the lumen of the fused ampullae and the blood flow is interrupted within a few minutes.\textsuperscript{76}
4.3 Cytotoxicity

When hemocytes of *C. intestinalis* and *Styela plicata* are incubated in the presence of mammalian erythrocytes or tumor cells, a rapid cytotoxic reaction can be observed. Cytotoxic reactions require the contact between hemocytes and target cells, are Ca$^{2+}$-dependent, and are inhibited by sphingomyelin.\(^{77-80}\) Cytotoxic cells are represented, in both cases, by PO-containing hemocytes\(^{30,80}\); their cytotoxicity is related to either the release of active PO in the medium, as in *Styela*\(^{81}\) and *Ciona*,\(^{82}\) or to the activity of the enzyme phospholipase A2, modulated by lectins with specificity for galactosides, as in *Ciona*.\(^{83}\)

In *B. schlosseri*, MC degranulation and the induction of cytotoxicity may be reproduced *in vitro* by exposing hemocytes to the cell-free hemolymph of incompatible colonies\(^{84-86}\) and can be prevented by PO inhibitors.\(^{87}\) Cytotoxicity is related to the induction of oxidative stress consequent to the production of ROS by PO, as indicated by the observation that ROS scavengers can suppress the induction of cytotoxicity both in vitro and in vivo, in colony allorecognition, although they have no effect on MC degranulation or PO activity.\(^{87,88}\) The role of nitric oxide (NO) in the induction of cell death is suggested by the in vitro production of nitrite, when hemocytes are exposed to nonself molecules, and by the decrease of in vitro cytotoxicity in the presence of the NOS inhibitor N\(^\text{ω}\)-nitro-L-arginine methyl ester.\(^{89}\) The production of NO by hemocytes, after their exposure to either LPS or zymosan, has been reported also in *Phallusia nigra*.\(^{90}\)

Cytotoxicity is related to the oxidation of polyphenol substrata, likely represented by tunichromes,\(^{91-93}\) contained inside MCs, by PO, causing the production of ROS and quinones that can damage various kinds of macromolecules.\(^{17,81,88}\)

4.4 Inflammation

Inflammation in ascidians is characterized by recruitment of circulating cells, extravasation, cell degranulation, and induction of cytotoxicity to the infection area.

*S. plicata* granulocytes contain histamine, which, when released upon the recognition of nonself, induces an inflammatory reaction by recruiting hemocytes, which, in turn, prompt tunic vessel-contraction and inhibit phagocytosis.\(^{94}\)

The injection of foreign material (rabbit erythrocytes, BSA, and LPS) into the tunic of *C. intestinalis* leads to an inflammatory reaction because of hemocyte recruitment to the injection area,\(^{75,95,96}\) hemocyte degranulation, and, eventually, capsule formation\(^{74}\) and tissue damage.\(^{97}\) Circulating granulocytes in treated animals increase the transcription of a gene for a member of the CAP-superfamily of proteins.\(^{98}\)

Tissue transplantation is another cause of inflammation: in solitary species, a higher recruitment of hemocytes has been observed in the case of allografts.
with respect to autografts; the ultimate rejection of allografts is more rapid in previously challenged animals. \(^{99–101}\) Graft rejection in \(M.\ manhattensis\) involves the recruitment of MCs and induction of cytotoxicity.\(^ {19,72}\) Analogously, in \(S.\ plicata\), MCs are massively recruited in the graft region and contribute to the induction of cytotoxicity.\(^ {102}\)

In colonial ascidians, the induction of inflammation is observable during the nonfusion reaction among genetically incompatible colonies. The phenomenon has been particularly studied in botryllid ascidians. In \(Botryllus\ primigenus\) and \(B.\ schlosseri\), it is controlled by a highly polymorphic \(Fu/HC\) (fusibility/histocompatibility) allorecognition gene with codominant alleles.\(^ {103,104}\) Contacting colonies can fuse into a single chimeric colony, with anastomosis of the tunic and the tunic vasculature, if they share at least one allele at the \(Fu/HC\) locus.\(^ {105,106}\) Conversely, a rejection reaction occurs, characterized by the appearance, along the contact border, of a series of cytotoxic foci, called points of rejection (POR), which contain melanin deposits due to PO activity.\(^ {85}\)

In \(B.\ schlosseri\), MCs are the first cells to sense nonself molecules,\(^ {107}\) and, as a consequence of the recognition, they produce and release molecules able to activate phagocytes, stimulate the release of BsRBL, and enhance their recruitment to the site of infection (see further).

The contact among allogeneic colonies leads to partial fusion of the facing tunics with the local disappearance of the limiting cuticles, leakage of humoral factors from one colony to the other, and the selective recruitment of MCs inside the blind endings of the tunic vessels (ampullae) facing the alien colony. Then, MCs migrate into the tunic where they degranulate and release their vacuolar content, in particular, the enzyme PO and its polyphenol substrata and quinones, which are responsible for the induction of cytotoxicity along the contact border.\(^ {85,88,89}\) The involvement of MCs in a nonfusion reaction has been demonstrated also in other botryllid ascidians, such as \(Botrylloides\ simodensis\), \(Botrylloides\ fuscus\), \(Botrylloides\ violaceus\), and \(B.\ leachii\).\(^ {106,108–112}\)

Genes involved in allorecognition and their receptors\(^ {113,114}\) are also expressed by hemocytes, although some uncertainties still persist as to the identity of the allorecognition gene.\(^ {113,115}\)

When incompatible colonies are artificially brought into contact at their cut surfaces, an intense inflammatory reaction is observed in ovoviviparous botryllid ascidians,\(^ {106,112,116–118}\) whereas fusion of tunics and blood vessels always occurs in the case of the viviparous species studied so far, suggesting that the hemocytes of these species have lost their ability for allorecognition. This can be related to the necessity to prevent the immune system from attacking the brooded embryos, which share only one allele at the \(Fu/HC\) locus with the mother colony, and might undergo rejection/resorption, as they are exposed to the circulation for more than a week.\(^ {119}\) In partial confirmation of the above statement, the PO activity of the hemolysate of viviparous species is much lower than that of ovoviviparous ones.\(^ {120,121}\)
5 HUMORAL FACTORS PRODUCED BY HEMOCYTES

5.1 Phenoloxidase

Chaga was the first to describe a DOPA-oxidase activity, ascribable to PO, in ascidians. Today, the presence of PO activity in ascidian hemocytes is widely reported in both solitary and colonial species. PO is assumed to be stored as an inactive proenzyme (proPO) inside the vacuoles of PO-containing hemocytes, and, analogous to what is known in arthropods, is activated by plasmatic serine proteases once released outside the cells.

In solitary species, PO is involved in cytotoxic responses triggered by foreign material; in colonial ascidians, the enzyme contributes to the formation of the POR along the contact border among genetically incompatible colonies.

Ascidian POs have a molecular weight ranging between 60 and 90 kDa, in agreement with molecular data predicting a molecular weight of 92 and 87 kDa for Ciona intestinalis POs. Electrophoretic data indicate that ascidian PO monomers interact to form dimers.

The analysis of nucleotide and deduced amino-acid sequences of PO from C. intestinalis, B. schlosseri, and P. misakiensis confirmed their high similarity with arthropod hemocyanins.

5.2 Lectins

Hemocytes are generally considered responsible for the synthesis and release of humoral lectins, the presence of which, in the ascidian hemolymph, has been described in various species. Only in a few cases, a definite role in immunity has been ascertained.

In the solitary species Halocynthia roretzi, a 120-kDa lectin, produced by MCs, recognizes LPS and other constituents of the microbial walls and also agglutinates bacteria; it can act as opsonin, facilitating the phagocytosis of microbes.

In B. schlosseri, BsRBL can agglutinate yeast, mammalian erythrocytes, and some bacterial strain. At high concentration, the lectin can trigger MC degranulation.

The genes for two galectins of 32 and 37 kDa have been identified in C. intestinalis. They are overexpressed after LPS injection in the body wall; in situ hybridization shows that unilocular and multilocular granulocytes contain the specific mRNAs.

5.3 Cytokines

Various reports indicate the presence of molecules with immunomodulatory activity in ascidians, and it is generally believed that they are produced by immunocytes.
In *B. schlosseri*, MCs are the main source of molecules recognized by antibodies raised against mammalian IL1α and TNFα; they are synthesized and released upon the recognition of foreign molecules.\(^{86,107}\) The same molecules can induce the synthesis and the release in the medium of BsRBL by phagocytes\(^ {133}\); conversely, BsRBL can stimulate MCs to release molecules immunopositive to anti-IL1α and anti-TNFα antibodies,\(^{58}\) revealing the existence of a cross-talk between immunocytes.

In *B. schlosseri*, *B. leachi*, and *B. simodensis*, molecules immunopositive to anti-IL1α and anti-TNFα antibodies are produced and released by MCs during the inflammatory reaction due to contact among genetically incompatible colonies;\(^ {89,112,134}\) the same antibodies inhibit both the in vitro cytotoxicity, observed when hemocytes are exposed to cell-free hemolymph from incompatible colonies,\(^ {89}\) and the MC chemotaxis caused by the same hemolymph.\(^ {86}\)

Anti-IL1α and anti-TNFα antibodies can prevent, in vitro, the increase of yeast phagocytosis occurring when *B. schlosseri* hemocytes and yeast cells are incubated in a conditioned medium, that is, the supernatant of hemocyte cultures matched with yeast cells. Fractionation of hemocytes through discontinuous density-gradient centrifugation indicates that the increase in phagocytosis occurs when conditioned media come from hemocyte fractions rich in MCs.\(^ {135}\) The above observations suggest that MCs are the main sentinel cells in the organism, which, upon the recognition of nonself, release cytokines able to recruit other MCs and phagocytes, and stimulate the release of BsRBL by phagocytes, which, in turn, strengthen MC activity and favor phagocytosis of foreign material, allowing for its disposal.\(^ {58}\)

A similar cooperation between MCs and phagocytes has been reported in *C. intestinalis*.\(^ {136}\) In the same species, the synthesis of an inducible lectin with galectin features increases during the inflammation, following intratunic LPS injection. Its opsonizing and hemagglutinating activities are inhibited by anti-human IL1α antibodies.\(^ {137}\) In the *Ciona* genome, putative genes for IL-1 and TNF receptors are present,\(^ {54,138}\) and the analysis of a series of ESTs from *Ciona* hemocytes revealed the presence of transcripts for a putative protein similar to a mammalian IL17 receptor.\(^ {138}\) In addition, a gene for a TNFα homolog has been identified and cloned;\(^ {54,139}\) its transcription increases after LPS injection in the body wall, and its mRNA is located in hemocytes, in particular, granulocytes.\(^ {139}\) It probably exerts its role in recruiting hemocytes to the inflamed area; the same gene is active also during larval development, likely in relation to the migration of mesenchyme cells.\(^ {140}\)

### 5.4 Complement

Chapter 6: The Evolution of Complement System Functions and Pathways in Vertebrates treats in great detail the activation of a complement system, especially in vertebrates. In ascidians, both the alternative and lectin complement
activation-pathways are present. In *H. roretzi*, transcripts for MASP-like proteins, C3, and Bf have been identified. In the same species, integrin α and β subunits are expressed on the surface of hemocytes, probably forming a C3 receptor.

In *C. intestinalis*, two genes for C3 (CiC3-1 and CiC3-2) have been described. In situ hybridization indicates that transcription occurs in hemocytes (hyaline and granular amoebocytes, and compartment cells) and in ciliated cells bordering branchial stigmata. The transcription of CiC3-1 by hemocytes and the quantity of the fragment CiC3-1a detected by immunohistochemistry increase after LPS injection in the tunic. CiC3-1a exerts a chemotactic role toward hemocytes, recruiting them to the inflammation site through its binding to a 100-kDa, G protein-coupled receptor, constitutively expressed in hyaline and granular amoebocytes. In the same species, three genes (CiBf-1, CiBf-2, and CiBf-3) for complement Bf were described. Moreover, a transcript for a collectins showing similarity for mammalian MBL, located in circulating granulocytes, has been reported: its expression increases after LPS injection in the tunic. As described in more detail in chapter 11: Functions and Diversity of Fibrinogen-Related Domains, transcripts for MASP-like proteins and a ficolin were also identified.

In *S. plicata*, a protein recognized by anti-C3 antibodies is secreted by phagocytes upon the exposure of hemocytes to molecules of bacterial, fungal, and algal origin. In this species, a C-type lectin, similar to mammalian collectins, has been identified. Recently, in *B. schlosseri*, genes for C3 and Bf have been described: their transcription occurs in MCs, and the opsonic role of C3 was indirectly demonstrated.

### 5.5 Antibacterial Activity

The involvement of hemocytes in the production of antimicrobial molecules has been demonstrated in *H. roretzi*, where two peptides, halocyamines A and B, are synthesized by MCs. Their cytotoxic activity is probably related to the presence of diphenol rings, which can represent the natural substrate for PO, analogously to tunichromes. In *S. clava*, MCs produce four histidine-rich, α-helix peptides, clavanins A–D, which are 23 amino acids in size and isolated from hemocyte lysates. In the same species, the transcripts for five putative cationic antimicrobial peptides, called styelins, later isolated from hemocyte lysates, have been identified. In *C. intestinalis*, the presence of antibacterial activity in hemocyte lysates was described. Transcriptome analysis allowed the identification of two families of α-helix antimicrobial peptides synthesized by hemocytes; their transcription is enhanced after the inoculum of nonself material in the body wall, and their microbicidal activity is related to their ability to permeabilize plasma membranes of target cells.
6 ASCIDIAN HEMOCYTES AND OXIDATIVE STRESS RESPONSE

Phagocytes contain active antioxidant enzymes, such as glutathione S-transferase and glutathione peroxidase, protecting the cells from the ROS that they produce.\textsuperscript{165}

In \textit{C. intestinalis}, hemocytes are the site of synthesis of mRNAs coding for detoxifying molecules, such as metallothionein CiMT1, glutathione synthase, \(\gamma\)-glutamyl-cysteine ligase, and phytochelatin synthase. The extent of transcription increases after exposure to Cd.\textsuperscript{166–168} \textit{Ciona} hemocytes also represent the sole source, together with ovarian follicle cells, of mRNAs for the superoxide dismutases Ci-SODa and Ci-SODb.\textsuperscript{169}

7 HEMOCYTES OF PELAGIC TUNICATES

Few data are currently available on the hemocytes of pelagic tunicates. Larvaceans lack hemocytes,\textsuperscript{170,171} whereas some authors, in the past, have described the morphology of thaliacean hemocytes.\textsuperscript{171} Recently, Cima et al.\textsuperscript{171} have reported a morpho-functional and ultrastructural characterization of the circulating hemocytes of \textit{Thalia democratica} oozooids. Four hemocyte types have been described: undifferentiated cells, phagocytes, granular cells, and storage cells. Phagocytes include amoeboid cells, with hyaline cytoplasm, and cells containing large vacuoles. Both phagocyte types share a similar content of hydrolytic enzymes, can migrate into the tunic and colonize it, and can be found outside of the animal body, adhering to the tunic that embeds the internal lumen of the oral siphon; it probably exerts an immunosurveillance role, similar to the one already described for immunocytes/phagocytic cells in \textit{B. schlosseri}. Granular cells resemble vertebrate mast-cells in having heparin and histamine inside their granules; they do not show any similarity with ascidian cytotoxic cells. Storage cells are represented by nephrocytes, which, like in ascidians, have uric-acid crystals inside their vacuoles.

8 CONCLUDING REMARKS

Hemocytes represent an important constituent of tunicate hemolymph and play fundamental roles in the clearance of nitrogenous catabolites, the detoxification of toxic compounds, the storage of nutrients, and immunosurveillance, assuring the animal survival by coping with foreign, potentially pathogenic microbes entering the organism.

Today, the new sequencing and bioinformatic tools for genome and transcriptome analysis are opening new, interesting, and previously unimaginable perspectives on hemocyte research. The increasing amount of available sequences are shedding new light on hemocyte functions, their activation pathways, mutual cross-talk, and ontogenesis. The possibility of synthesizing (using chemical synthesis of recombinant technology) the gene products once the gene
sequence is known, is rendering available a quantity of bioactive molecules produced by hemocytes, such as new antimicrobial peptides or anticancer compounds that could find interesting applications in biomedical and sanitary fields. In addition, future research on hemocytes will surely contribute to clarify unresolved aspects of hemocyte ontogeny and differentiation pathways, and their role in tunicate biology.

REFERENCES


34. Michibata H. New aspects of accumulation and reduction of vanadium ions in ascidians, based on concerted investigation from both a chemical and biological viewpoint. Zoolog Sci 1989;6:639–47.


The Evolution of the Immune System


70. Monniot C. *Kystodelphysdrachi* n. g.n.sp., copepode enkystedans une branchie d’ascidie. *Vie Milieu* 1963;14:263–73.


Page left intentionally blank
Chapter 3


Yoichi Sutoh*, Masanori Kasahara**
*Emory Vaccine Center and Department of Pathology and Laboratory Medicine, Emory University, Atlanta, GA, United States; **Department of Pathology, Hokkaido University Graduate School of Medicine, Sapporo, Japan

1 INTRODUCTION

The human body is thought to have tens of millions of lymphocyte clones, each expressing antigen receptors with distinct specificities. When infected with pathogens, lymphocyte clones expressing specific receptors undergo proliferation and differentiate into effector lymphocytes. Most of these effector lymphocytes die by apoptosis shortly after the elimination of pathogens. Some lymphocytes, however, survive for a long time to build an immunological memory. These lymphocytes—known as memory lymphocytes—enable the host to mount a more prompt and vigorous immune response upon reexposure to the same pathogen. Although the adaptive immune system (AIS) is a complex biological system, its backbone is formed by highly diverse antigen receptors clonally expressed on lymphocytes.

The origin and evolution of the AIS and lymphocytes has attracted the interest of immunologists for a long time. Animals that have received particular attention in this regard are the jawless vertebrates represented by lampreys and hagfish. Studies conducted in the 1960s and 1970s showed that both lampreys and hagfish are capable of producing specific agglutinins against particulate antigens and rejecting skin allografts with immunological memory,1-8 suggesting that the origin of adaptive immunity can be traced back to the emergence of jawless vertebrates. Consistent with this, jawless vertebrates have blood cells morphologically indistinguishable from mammalian lymphocytes. In lampreys, naive lymphocyte-like cells measure about 10 μm in diameter and have a round shape, with thin cytosol and a large nucleus.9,10 In electron microscopy, thin
heterochromatin is found on the inside rim of the nuclear membrane.\textsuperscript{9} After stimulation with particulate antigens such as sheep blood-cells and anthrax spore-coats, the lymphocyte-like cells undergo proliferation, differentiating into plasma cell-like cells with enlarged cytosol that contain well-developed rough endoplasmic reticulum.\textsuperscript{11} Transcriptome analysis revealed that lamprey and hagfish lymphocyte-like cells express many genes whose mammalian counterparts are expressed by lymphocytes, such as the genes coding for a homolog of Spi, B-cell receptor-associated protein (BCAP), GATA2/3, CXC chemokine receptor (CXCR) 4, and CD98.\textsuperscript{12–14} However, the transcripts coding for major histocompatibility complex (MHC) molecules, T-cell receptors (TCRs), B-cell receptors (BCRs), or the RAG (recombination activating gene) enzymes were never identified from jawless vertebrates. This was puzzling because, at first glance, it contradicted the earlier studies that clearly demonstrated the ability of jawless vertebrates to reject skin allografts and to produce agglutinins specific for particulate antigens. Ultimately, this puzzle was resolved by the discovery that, instead of TCRs and BCRs, jawless vertebrates use a unique antigen receptor now known as variable lymphocyte receptors (VLR).\textsuperscript{15–20}

2 OVERVIEW OF VLRs

2.1 Structure of VLR Proteins and Gene Assembly

VLR was discovered in lampreys as a gene coding for proteins with highly diverse sequences, through the analysis of a cDNA library, enriched for transcripts upregulated in antigen-stimulated lymphocyte-like cells.\textsuperscript{21} Structurally, VLRs are a member of the leucine-rich repeat (LRR) family of proteins composed of a signal peptide (SP), an N-terminal LRR cassette (LRRNT), an 18-residue N-terminal LRR cassette (LRR1), a variable number of 24-residue LRR cassettes (LRRV), a 24-residue end LRRV cassette (LRRVe), a 13-residue truncated LRR cassette [also called the connecting peptide (CP)], a C-terminal LRR cassette (LRRCT), and an invariant domain containing a stalk region (Fig. 3.1A). The sequences of the regions ranging from the 3’-half of LRRNT (3’-LRRNT) to the 5’-half of LRRCT (5’-LRRCT) are highly diverse. By contrast, the sequences of the remaining regions—the SP, the 5’-half of LRRNT, the 3’-half of LRRCT, and the stalk region—are invariant.

Remarkably, the VLR gene has an incomplete structure incapable of encoding any protein in the genome of nonlymphoid cells, such as erythrocytes; this germline gene encodes only the invariant region of VLR proteins and lacks the sequences coding for the diversity region (Fig. 3.1B). Cassettes such as 3’-LRRNT, LRR1, LRRV, LRRVe, SP, and 5’-LRRCT, constituting the diversity region, are located in multiple copies (sometimes in several hundred copies) in the vicinity of the germline VLR gene. During lymphocyte development, the invariant intervening sequence of the
FIGURE 3.1 Domain organization, gene assembly, and tertiary structure of VLRs. (A) Domain organization of mature VLR protein. The diversity region is composed of the following LRR cassettes: the N-terminal LRR cap (LRRNT), LRR1, a variable number of LRRVs, LRRVe, a connecting peptide (CP), and the C-terminal LRR cap (LRRCT). SP, signal peptide. (B) Assembly of VLR genes. The germline VLR gene has an invariant intervening sequence instead of highly diverse LRR cassettes and is unable to encode functional proteins (top). During lymphocyte development, LRR cassettes, which are scattered around the germline VLR, replace the intervening sequence in a stepwise manner from either the 5'- or 3'-end (middle and bottom). This process, which resembles gene conversion, is thought to be mediated by cytidine deaminases of the AID-APOBEC family. (C) Tertiary structure of VLRs. VLR proteins have a horseshoe-shaped structure characteristic of LRR family proteins. The majority of variable amino acid residues are located on the β-sheet, facing the concave surface (shared by all three receptors), and the hypervariable loop in LRRCT (present only in VLRA and VLRB receptors; indicated by red arrows). Both ends of the β-sheet are capped by LRRNT and LRRCT (indicated by blue and red, respectively). Unlike VLRA or VLRB, lamprey VLRC has an invariable loop in LRRNT (indicated by a blue arrow). Like TCRs and BCRs, all three VLRs lack a signaling domain in the cytoplasmic region. Therefore, it is assumed that VLRs associate with signal transducers yet to be identified.
The germline VLR gene is replaced by a gene conversion-like mechanism in a stepwise manner, beginning either from its 5′- or 3′-end, by adding flanking cassettes, and eventually forming a completely assembled VLR gene.22–25 Short stretches of nucleotide homology (10–30 bps) are found between donor and acceptor sequences. 24,26,27 Therefore, the sequence located at the ends of the most newly copied cassette presumably determines which flanking LRR cassettes should be copied into the germline VLR gene in the next step. The sequences of LRR cassettes are highly variable and the number of copied LRRV cassettes is also variable. This enables assembled VLR genes to acquire sequence diversity comparable to that of TCRs and BCRs.22,24 The assembly of VLR genes is mediated by cytidine deaminases (CDA) of the activation-induced cytidine deaminase (AID)-apolipoprotein B mRNA editing enzyme (APOBEC) family.24 In lampreys, two CDAs named CDA1 and CDA2 have been identified.24

2.2 Both Lampreys and Hagfish Have Three VLR Genes

Initially, only one VLR gene was identified in lampreys.21 Subsequent studies revealed that lampreys have two more VLR genes.24,28 The three VLR genes are now known as VLRA, VLRB, and VLRC. Hagfish also have three VLR genes named VLRA, VLRB, and VLRC, thought to be orthologous to lamprey VLRA, VLRB, and VLRC genes,29,30 indicating that these three genes already existed in a common ancestor of lampreys and hagfish.

Functional studies in lampreys demonstrated that the products of the three VLR genes are expressed on three distinct populations of lymphocyte-like cells31,32 (Fig. 3.2). VLRA and VLRC, which are more closely related to each other in sequence than they are to VLRB, are membrane-bound receptors with no secretory form. VLRA+ cells and VLRC+ cells are T-cell-like and develop in the “thymoid,” an organ assumed to be the equivalent of the gnathostome thymus.33 On the other hand, VLRB is a glycosylphosphatidylinositol-anchored protein that also occurs in a secretory form. Secreted VLRB molecules form pentamers or tetramers of dimers and have 8–10 antigen binding sites.34 This type of subunit organization resembles the subunit arrangement of gnathostome IgM antibodies and accounts for the strong agglutinating activities of VLRB antibodies. Interestingly, VLRB+ cells are B-cell-like, develop in hematopoietic organs, undergo clonal proliferation in response to antigen stimulation, and differentiate into plasma cell-like cells secreting VLRB antibodies.11,22,31 Furthermore, similar to the allelic exclusion in TCR and BCR genes, each lymphocyte expresses only a single functional VLR allele, indicating that VLR assembly occurs in a monoallelic manner.21,28,31 Collectively, these observations provided convincing evidence that lymphocyte-like cells in jawless vertebrates resemble gnathostome lymphocytes, not only morphologically, but also functionally, and hence should be regarded as authentic lymphocytes.
Crystallographic analysis revealed that VLR proteins adopt a horseshoe-like solenoid structure characteristic of the LRR protein family, capped by LRRNT and LRRCT cassettes at N- and C-terminal ends, respectively (Fig. 3.1C). The inner concave surface of the solenoid structure is formed from multiple β-strands (derived from LRRNT, LRR1, LRRV, LRRVe, and CP), which assemble into a continuous β-sheet. In VLRA and VLRB receptors, this β-sheet and a protruding loop formed in the 5′-LRRCT cap contain the majority of variable residues. Structural analysis of VLRB receptors in complex with antigens such as Type O blood antigen (H-trisaccharide) and hen egg-white lysozyme (HEL) demonstrated that VLRB binds antigens via the β-sheet and protruding loop in 5′-LRRCT. Interestingly, the protrusion in 5′-LRRCT penetrated deeply into the catalytic cleft of HEL in the VLRB-HEL complex. Immunoglobulins made up of VH and VL chains normally recognize epitopes exposed on the surface of molecules. The ability of VLRs to bind to residues hidden in the cleft is reminiscent of camel and shark VH antibodies that preferentially target clefts.
In one study, lampreys immunized with HEL produced not only specific VLRBs, but also specific VLRA with binding affinity comparable to that of IgG. The crystal structure of lamprey VLRA in a complex with HEL revealed that VLRA interacts with HEL through its concave surface and LRRCT.

Unlike VLRA or VLRB receptors, VLRC lacks a protrusion in its LRRCT region (Fig. 3.1C), and its LRRCT is semiinvariant. Also, unlike VLRA or VLRB receptors, the LRRNT of lamprey VLRC has a loop that protrudes into the concave surface. This loop shows little sequence variation. These structural features of VLRC suggest that VLRC might recognize antigens in a unique manner. It is possible that VLRC interacts with putative antigen-presenting molecules through its invariant or semiinvariant LRRCT region and LRRNT loop.

2.4 **VLRA and VLRC Genes are Tightly Linked in the Lamprey Genome**

VLRA and VLRC loci are located close to each other in the lamprey genome. Similar to the gnathostome TCRA/TCRD locus that shares some V segments, lamprey VLRA and VLRC genes often use the same LRR cassettes for their assembly. Such sharing of LRR cassettes is also observed in assembled hagfish VLRA and VLRC genes, suggesting that VLRA and VLRC genes are also tightly linked in the hagfish genome. In contrast, neither lamprey nor hagfish VLRB receptors share identical LRR cassettes with the VLRA or VLRC receptors of respective species, suggesting that the germline VLRB gene is not situated close to the VLRA/C locus. Consistent with this, previous fluorescence in situ hybridization analysis showed that hagfish VLRB and VLRC (then known as VLRA) genes are physically well-separated, although they are on the same chromosome. It appears that VLR genes increased their copy number by tandem duplication; subsequently, chromosomal inversion or intrachromosomal translocation presumably separated the VLRB gene from the VLRA/C genes, facilitating the functional specialization of the receptor genes.

3 **THREE POPULATIONS OF AGNATHAN LYMPHOCYTES**

3.1 **VLRA** Cells and **VLRC** Cells Resemble Gnathostome T Cells, Whereas **VLRB** Cells Resemble Gnathostome B Cells

VLRA cells and VLRC cells resemble gnathostome T cells in that they undergo proliferation in response to a T-cell mitogen such as phytohemagglutinin (PHA). They also resemble T cells in terms of gene expression profiles (Fig. 3.2). For instance, they express transcription factors involved in T-cell development, such as GATA2/3, B-cell lymphoma 11B (BCL11b), and T-cell factor 1 (TCF1), signaling molecules such as LAT (linker for activation of T cells), cell surface molecules such as NOTCH1, CD45, and aryl hydrocarbon receptor (AHR), cytokines such as IL-17, and cytokine receptors...
such as chemokine receptor 9/7 (CCR9/7), which regulate the migration of T-cell progenitors to the thymus.\textsuperscript{31,32} VLRC\textsuperscript{+} cells generally have a gene expression profile similar to that of VLRA\textsuperscript{+} cells, but differ from VLRA\textsuperscript{+} cells in that they express genes characteristically expressed in gnathostome γδ T cells, such as sex-determining region Y-box 13 (SOX13) coding for a fate-determining factor for the γδ T cell lineage, and Toll-like receptor 3 (TLR3).\textsuperscript{32} VLRC\textsuperscript{+} cells also express an integrin family of adhesion molecules involved in the epithelial localization of γδ T cells, such as integrins αL (ITGAL), α4 (ITGA4), and β1 (ITGB1).

On the other hand, VLRB\textsuperscript{+} cells resemble gnathostome B cells, in that they differentiate into plasma cells and secrete specific VLRB molecules as antibodies, when challenged with particulate antigens.\textsuperscript{11} Also, the gene expression profile of VLRB\textsuperscript{+} cells is similar to that of gnathostome B cells; they express transcription factors such as B-lymphocyte-induced maturation protein 1 (BLIMP-1), B-cell CLL/lymphoma 6 (BCL6), and paired box protein 5 (PAX5), signal-transducing molecules such as spleen tyrosine kinase (Syk) and B-cell adaptor protein (BCAP), and Toll-like receptors such as TLR2a-c, TLR7, and TLR10.

Interestingly, T-cell-like VLRA\textsuperscript{+} cells express the IL-8 receptor and IL-17, whereas B-cell-like VLRB\textsuperscript{+} cells express IL-8 and the IL-17 receptor. This observation suggests that VLRA\textsuperscript{+} and VLRB\textsuperscript{+} cells may have a cross-talk via cytokines and possibly interact in a manner analogous to T–B cell collaboration\textsuperscript{32} (Fig. 3.3).

In rodents, γδ T cells are a minor population of T lymphocytes in blood and peripheral lymphoid tissues; however, they occupy the majority of lymphocytes in epithelial layers of tissues such as skin, intestine, tongue, and the reproductive tract. These intraepithelial γδ T cells express TCRs with limited variability. Most notably, γδ T cells residing in the epidermis, known as dendritic epidermal T cells (DETC), express an invariant Vγ5Vδ1 (alternate nomenclature Vγ3Vδ1) TCR, lacking junctional diversity.\textsuperscript{47,48} Interestingly, lamprey VLRC\textsuperscript{+} cells are located predominantly in epithelial tissues and express restricted antigen receptor diversity, showing intriguing similarity to mouse DETCs and intraepithelial lymphocytes (Fig. 3.2).\textsuperscript{49} When stimulated with poly I:C, a synthetic viral double-strand RNA (dsRNA) analog, VLRC\textsuperscript{+} cells respond via their TLR3 and upregulate IL-16 expression. IL-16 is a chemoattractant recruiting CD4\textsuperscript{+} leukocytes, such as T cells, monocytes, and eosinophils to the site of infection.\textsuperscript{50} Therefore, similar to DETCs, epithelial VLRC\textsuperscript{+} cells may perform a sentinel function in the epithelium.

### 3.2 Antigen Recognition by Agnathan T-Like Cells is Still Full of Mystery

When lampreys were immunized with anthrax spores, they generated VLRB\textsuperscript{+} cells capable of binding to the spores, but VLRA\textsuperscript{+} cells with such ability were
not detectable,\textsuperscript{29} suggesting that VLRA receptors do not bind native bacterial surface epitopes and might recognize processed antigens in vivo similar to gnathostome $\alpha\beta$ T cells. However, screening of a yeast display library resulted in the identification of lamprey VLRA molecules that directly bound to HEL.\textsuperscript{40} Although this observation demonstrates that VLRA molecules are capable of binding unprocessed antigens, it is in contrast to the fact that $\alpha\beta$ T cells, the presumed counterpart of agnathan VLRA$^+$ cells, recognize only processed antigens bound to MHC class I or class II molecules. At present, it remains unknown whether VLRA receptors always recognize antigens without any requirement for antigen processing, or direct antigen recognition occurs only in some or exceptional cases. Also unknown is whether direct recognition occurs in vivo.

\textbf{FIGURE 3.3} The presumed role of three populations of agnathan lymphocytes in eliminating pathogens. Invading pathogens are recognized by VLRs with specific binding specificities. VLRC$^+$ cells also express TLR3 that recognizes viral double-stranded RNAs. Activated VLRA$^+$ cells and VLRC$^+$ cells secrete pro-inflammatory cytokines, such as IL-17, MIF, and IL-16, which presumably induce the migration of various immune cells to the site of infection. IL-17 released from VLRA$^+$ cells may activate VLRB$^+$ cells that express the IL-17 receptor and induce the secretion of VLRB antibodies into the serum. Secreted VLRB antibodies form an antigen–antibody complex that activates the complement pathway via binding to C1q-like protein and mannose-binding lectin-associated serine protease (MASP). VLRB$^+$ cells might use IL-8 for cross-talk with VLRA$^+$ cells and VLRC$^+$ cells that express the IL-8 receptor. It is still uncertain whether jawless vertebrates have antigen-presenting molecules with functions equivalent to MHC molecules. Also unknown is whether VLRA and VLRC require antigen-presenting molecules for antigen recognition. This figure was modified from Sutoh and Kasahara.\textsuperscript{46}
Currently, no information is available as to how VLRC$^+$ cells recognize antigens. If VLRC$^+$ cells are $\gamma$6 T-cell-like, as discussed previously, they might recognize antigens directly in a manner similar to $\gamma$6 T cells. This is, however, in apparent conflict with the speculation that lamprey VLRC receptors might interact with putative antigen-presenting molecules through their invariant loop in LRRNT and semiinvariant LRRCT. Much remains to be learned about antigen recognition by agnathan T-cell-like lymphocytes.

4 LYMPHOCYTE DEVELOPMENT IN JAWLESS VERTEBRATES

4.1 Development of VLRB$^+$ Cells

B cells develop in hematopoietic organs such as the bone marrow in mammals, the bursa of Fabricius in birds, the pronephros in bony fishes, and the spleen and spiral valves in cartilaginous fishes.\textsuperscript{51–53} In adult lampreys, VLRB$^+$ cells are abundant in kidneys and blood. On the other hand, in lamprey larvae, VLRB$^+$ cells are abundant in the typhlosole, an invaginated spiral valve spanning the length of the intestine. Recent flow cytometric analysis showed that 26 and 41\% of lymphocytes are VLRB$^+$ cells in typhlosoles and kidneys, respectively.\textsuperscript{29} In both organs, cell proliferation was activated within 28 days after immunization by anthrax spore-coats. Moreover, in situ hybridization analysis detected CDA2 expression in VLRB$^+$ cells, in typhlosoles and kidneys supporting the idea that VLRB gene assembly occurs in these organs.\textsuperscript{33} Although rigorous analysis with molecular markers remains to be conducted, available evidence indicates strongly that VLRB$^+$ cells develop in hematopoietic organs such as typhlosoles and kidneys, consistent with the idea that VLRB$^+$ cells represent B-lineage cells (Fig. 3.4).

4.2 Development of VLRA$^+$ Cells and VLRC$^+$ Cells

In jawed vertebrates, T-lineage progenitors generated in hematopoietic tissues migrate to the thymus. After successful V(D)J recombination of TCR genes, T cells are positively and negatively selected in the thymic cortex and medulla, respectively. As a result of thymic selection, only T cells expressing self-tolerant, MHC-restricted TCRs are allowed to survive and exit to the periphery. The thymus was histologically and/or functionally identified in all jawed vertebrates ranging from cartilaginous fish to mammals, consistent with the fact that they have MHC molecules and TCRs. In contrast, jawless vertebrates lack anatomical structures with discrete cortical and medullary regions resembling the thymus. Therefore, it was generally assumed until recently that the thymus emerged in a common ancestor of jawed vertebrates, concomitant with the emergence of TCR, BCR, MHC, and RAG molecules.\textsuperscript{54,55} However, the discovery that VLRA$^+$ cells are T-cell-like,\textsuperscript{31} and that jawless vertebrates have a genetic network involved in thymopoiesis,\textsuperscript{54} initiated a renewed search for a thymus equivalent in lampreys, which led to the identification of the “thymoid”\textsuperscript{33} (Fig. 3.4). “Thymoids” are located at the tips of the gill filaments in the gill basket, thus occurring, not
The Evolution of the Immune System

as a single organ, but as a constellation of specialized lymphoid tissues with no obvious corticomediullary differentiation. In “thymoids,” VLRA⁺ or VLRC⁺ lymphocytes expressing CDA1 occur in close proximity to pharyngeal epithelial cells expressing FOXN1, a marker of the thymopoietic microenvironment in jawed vertebrates. Also, sequencing of VLRA and VLRC genes isolated from gill-tip tissues contained nonfunctionally assembled sequences over fourfold more frequently than in peripheral blood cells. Furthermore, VLRA⁺ cells and VLRC⁺ cells in peripheral blood, but not in “thymoids” underwent proliferation in response to PHA. These results suggest that the “thymoid” is a primary, rather than a secondary, lymphoid organ. The thymus and the gill are both derived from the pharyngeal arch, and the thymus is located in the vicinity of the gill in cartilaginous and bony fishes. Therefore, the thymus and the “thymoid” appear to be homologous rather than analogous organs.

![Diagram of Lymphocyte Development](image)
4.3 Cell Fate Determination of VLRA$^+$ Cells and VLRC$^+$ Cells

VLRB$^+$ cells never express the VLRA gene nor the VLRC gene. Conversely, neither VLRA$^+$ nor VLRC$^+$ cells express the VLRB gene, consistent with the observation that the assembly of VLRA/C and VLRB genes occur in different primary lymphoid organs. In contrast, VLRA$^+$ cells sometimes express incompletely assembled, nonfunctional VLRC transcripts. Likewise, a fraction of VLRC$^+$ cells expresses incompletely assembled, nonfunctional VLRA transcripts, indicating that the assembly of VLRA and VLRC genes occurs concomitantly or sequentially. When analyzed by genomic PCR, the assembly of the VLRA gene was hardly found in VLRC$^+$ cells. In contrast, the assembly of the VLRC gene was observed in the majority of VLRA$^+$ cells, although 79% of assembly was nonproductive. The higher frequency of nonfunctional VLRA assembly in VLRA$^+$ cells than nonfunctional VLRC assembly in VLRC$^+$ cells suggests that the assembly of the VLRC gene precedes that of the VLRA gene, and that the germline VLRA gene undergoes assembly when both copies of the germline VLRC gene fail to undergo successful assembly.

4.4 Evidence for Selection in “Thymoids”

Accumulating evidence suggests that selection, somewhat similar to thymic selection, may be operating on VLRA$^+$ cells and VLRC$^+$ cells. On average, VLRB receptors have about 1.5 LRRV cassettes (excluding LRRVe), with the copy number of LRRV cassettes showing the binomial distribution. In contrast, the average number of LRRV cassettes is about 3 in VLRA/VLRC receptors, and the copy number distribution of LRRV cassettes deviates from the binomial distribution, mainly because transcripts with two or fewer LRRV cassettes occur infrequently, thus suggesting that the VLRA/VLRC receptors with two or fewer LRRV cassettes may be selected against. Indeed, nonfunctional VLRC transcripts in VLRA$^+$ cells shows greater variation in the copy number of LRRV cassettes than functional VLRC transcripts in VLRC$^+$ cells. Furthermore, the copy number of LRRV cassettes shows greater variation in VLRC transcripts isolated from the “thymoid” than in those isolated from the peripheral blood. These observations suggest that VLRA/VLRC receptors expressed on the surface of lymphocytes are selected in the “thymoid”.

4.5 Evolution of Lymphocytes in Vertebrates

Jawless vertebrates have two major lineages of lymphocytes: one resembling gnathostome T cells and the other resembling gnathostome B cells. Therefore, it is
very likely that a common ancestor of jawed and jawless vertebrates had two lymphocyte lineages: one specialized for cellular immunity, and the other endowed with phagocytic activity but also oriented toward humoral immunity. It appears that jawless and jawed vertebrates then coopted different antigen receptors within the context of such lymphocyte lineages (Fig. 3.5). Because there is no evidence that urochordates (ascidians and other tunicates, described in chapter: Origin and Functions of Tunicate Hemocytes) or cephalochordates (amphioxus or lancelets) have lymphocytes, an ancestral lymphocyte appears to have emerged in the vertebrate lineage after it diverged from the invertebrate lineages of chordates. (For further discussion about the evolution of vertebrate lymphocytes, see chapter: The Evolution of Lymphocytes in Ectothermic Gnathostomata.)

Like jawed vertebrates, lampreys have two lineages of T-cell-like cells, with VLRA+ cells and VLRC+ cells resembling αβ and γδ T cells, respectively, in terms of gene expression profiles and tissue distribution. Therefore, not only the separation of lymphocyte lineages into T-cell-like and B-cell-like cells, but also the separation of T-cell-like cells into αβ T-like and γδ T-like cells appears to have taken place in a common ancestor of vertebrates.

Searches for immunoglobulin superfamily proteins in jawless vertebrates identified potential evolutionary precursors of TCR/BCR, such as the lamprey “TCR-like gene” coding for an immunoreceptor tyrosine-based inhibition motif-bearing membrane protein, with one V-type and one C2-type immunoglobulin-like domain, and “agnathan paired receptors resembling antigen receptors” (APAR) of hagfish, having a single extracellular V-type immunoglobulin-like domain with a canonical J segment. The occurrence of these receptors in jawless vertebrates indicates that a common ancestor of jawed and jawless vertebrates had V-type immunoglobulin-like domains that could evolve into those of TCR/BCR. On the other hand, VLR is assumed to have emerged from a glycoprotein Ibα (GPIbα)-like protein, a component of the platelet glycoprotein-receptor complex conserved in all vertebrates. Therefore, it is likely that a common ancestor of jawed and jawless vertebrates possessed building blocks for both VLR-like and TCR/BCR-like receptors. Hence, it is reasonable to assume that a common ancestor of vertebrates had the potential for developing both VLR-based and TCR/BCR-based adaptive immunity. Then, what kind of receptors did a common ancestor of vertebrates use for antigen recognition?

A key observation in addressing this question is that both jawed and jawless vertebrates have cytidine deaminases of the AID/APOPEC family. In jawless vertebrates, two cytidine deaminases named CDA1 and CDA2 generate diversity of VLRs by a gene-conversion-like mechanism. In jawed vertebrates, AID is involved in gene conversion, class switch recombination, and somatic hypermutation, thus having functions overlapping with those of CDA1 and CDA2. Furthermore, in some animals, such as rabbits, sheep, and chickens, AID plays a major role in the diversification of antibody repertoire. In contrast, the RAG enzymes involved in the V(D)J recombination of TCR/BCR genes are of transposon origin and are present only in jawed vertebrates. These observations
suggest that a common ancestor of jawed and jawless vertebrates probably used VLR-like receptors and employed the members of the AID/APOPEC family to generate the diversity of their antigen receptors. Presumably, VLRs were superseded by TCRs and BCRs in the jawed-vertebrate lineage because the acquisition of RAG transposons enabled the development of a more efficient and powerful antigen-recognition system.\textsuperscript{16,65} The RAG transposon is thought to have been inserted into an ancestor of TCR/BCR genes. Therefore, the functional specialization of a gnathostome antigen-receptor into TCRs and BCRs, and then of the former into \(\alpha\beta\)TCRs and \(\gamma\delta\)TCRs, must have occurred under the strong functional constraint of the three lineages of specialized lymphocytes.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image.png}
\caption{Origin and evolution of lymphocytes in vertebrates. Available evidence indicates that three lineages of lymphocytes (two T-cell-like and one B-cell-like lineage) emerged in a common ancestor of jawed and jawless vertebrates. This common ancestor most likely possessed cytidine deaminases (CDA) and VLR-like antigen receptors because CDA and GPIb\(\alpha\) are present in both jawed and jawless vertebrate lineages. This ancestor also appears to have possessed potential precursors of TCRs/BCRs that became the targets of RAG insertion. Because the RAG enzymes are present only in the jawed vertebrate lineage, this insertion appears to have occurred only in a common ancestor of jawed vertebrates. Therefore, it is likely that a common ancestor of jawed and jawless vertebrates had an AIS based on VLRs and CDAs, and that this AIS was superseded by a more efficient TCR/BCR-based AIS in the jawed vertebrate lineage after the acquisition of RAG enzymes.}
\end{figure}
5 CONCLUSIONS

Despite the fact that jawed and jawless vertebrates use structurally unrelated receptors for antigen recognition, their AISs have much in common. Particularly striking is the conservation of lymphocyte lineages. Extant invertebrates, either protostomes or deuterostomes, lack lymphocytes. Therefore, it appears that T- and B-lymphocyte lineages specialized for cellular and humoral arms of adaptive immunity emerged in a common ancestor of jawed and jawless vertebrates.

Jawless vertebrates have neither MHC class I nor MHC class II molecules. An important issue that remains unanswered is whether jawless vertebrates have antigen-presenting molecules with functions equivalent to gnathostome MHC molecules. The observation that VLRC$^+$ cells (and presumably VLRA$^+$ cells as well) undergo selection in the “thymoid” suggests that jawless vertebrates have antigen-presenting molecules with functions equivalent to those of MHC molecules. The basic design of the AIS is well-conserved between jawed and jawless vertebrates, most likely because the lymphocyte lineages had been established in their common ancestor. Given the overall similarity of the gnathostome and agnathan AISs, it seems likely that jawless vertebrates have antigen-presenting molecules that function as substitutes of gnathostome MHC molecules.

ACKNOWLEDGMENTS

This work has been supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. We thank Dr. Yukiko Miyatake for her kind help with the preparation of figures.

REFERENCES


Chapter 4

The Evolution of Lymphocytes in Ectothermic Gnathostomata

Giuseppe Scapigliati, Francesco Buonocore
Department for Innovative Biology, Agro-Industry and Forestry, University of Tuscia, Viterbo, Italy

1 INTRODUCTION

Lymphocytes of vertebrates could be regarded as the direct successors of invertebrate immunocytes, in which the evolution shaped new functional defense activities coded by sets of new genes. This hypothesis can be drawn by observing the documented events that accompanied the origin of vertebrates. A first whole-genome duplication occurred in chordates about 550 million years ago and originated agnathan vertebrates that were provided with activation-induced cytidine deaminase (CDA) enzymes, which were able to promote ordered shuffling of gene segments, known as somatic recombination, and also clonal cells that produce recombined antigen-recognizing molecules, or lymphocytes. The secreted or cell surface-associated antigen-recognizing molecules (antibodies) of agnathans are made of recombined peptides having leucine-rich repeats (LRR) motifs.

A subsequent genome duplication event around 300 million years ago produced vertebrates with jaws, or gnathostomes, in which, by means of paralogous evolutionary mechanisms, the nonsself recognition machinery is organized in a diverse assembly. Antigen and nonsself sensing is based on immunocytes armed with receptors made of LRR motifs (Toll-like receptors, TLR), whereas effector lymphocytes use antibodies constituted by immunoglobulin (Ig)-based molecular motifs. These events have been reviewed more in detail in recent work, and summarized in Fig. 4.1.

In gnathostomes, representing approximately 2% of extant metazoan species, immune defenses are based on innate-plus acquired systems, and, on the basis of available knowledge, it is henceforth conceivable to speculate on a difference between the “relative amounts” of innate and acquired responses in aquatic and terrestrial gnathostome species (GSP). Indeed, by analyzing cellular and molecular activities involved in antigen recognition, elimination, and memory, it
appears that fishes rely on a predominance of innate responses, whereas avian and mammals have developed more finely tuned acquired-responses. A possible explanation of this observation might be found in the diverse exposure of aquatic and terrestrial GSP to their surrounding microbiomes because in a water environment the diffusion of pathogens can be easier and faster than in an aerial (terrestrial) environment. As a consequence of this, fishes are armed with a mucosa-associated innate system capable of fast antigen-recognition and elimination. In addition, recent evidence has shown that some fish species lack of fundamental component of acquired responses such as IgM genes, specific IgM production, or an entire subset of T-helper responses or MHC\textsubscript{II} genes.

A “big event” in the evolutionary history of vertebrates was the acquisition of jaws, which permitted in early GSP the swallowing of large-sized food, with possible consequent injuries to the intestinal tract. It can be speculated that these possible injuries could have permitted the entry of pathogens through the intestinal mucosa, and that, in turn, gut-associated immunocytes faced this new situation with novel Gnathostome-associated genes. Among these genes, the recombination-activating genes (RAG) coded for enzymes were able to shuffle Ig-based antigen receptors, and thus, continuing the speculation, RAG-containing immunocytes in intestine could have become precursors of intra-epithelial lymphocytes (IEL). On the other hand, immunocytes armed with LRR-based
receptors maintained the ancient function of antigen sensing and phagocytosis. Subsequently, the lymphocytes acquired specialization either by retaining the Ig-based antigen receptor on their surface (T cells), or shedding it in body fluids (B cells). These speculations (reassumed in Fig. 4.1) could be useful to discuss the presence of innate-type lymphocytes originating in the intestine, not requiring MHC-driven antigen presentation, and of MHC-instructed T cells originated in the thymus, as previously postulated.

In support of this hypothesis, it should be remembered that invertebrates have an intestinal recognition-system that may discriminate between pathogenic and beneficial microbes; this feature has been conserved in GSP in innate-type lymphocyte subpopulations.

A GSP lymphocyte could be defined in the simplest way as an immunocyte expressing Ig domain-based antigen receptors and having RAG enzymes. The Ig domain is as old as unicellular metazoans, and has been rearranged during evolution as a recognition mechanism in multicellular species. The GSP Ig domain-based antigen receptors are B-cell receptors (BR), or Ig, and T-cell receptors (TR). The BR and BR coreceptors are inserted into the outer membrane of B lymphocytes, and thus function as an antigen–receptor complex; alternatively the BR can be secreted as an antibody-effector molecule in body fluids by plasma B cells. The TR and TR coreceptors are permanently inserted into the plasma membrane of T cells, and the cells must have a physical contact with the antigen to perform their functions. Each lymphocytic cell carries one type of somatically rearranged BR or TR, and thus may originate a clone of cells upon induction.

2 FISHES

Fishes are a taxonomically heterogeneous group of chordates in the subphylum of Vertebrata, included in the supraclasses of Agnatha and Gnathostomata. Fish Gnathostomata classes are Condrychthyes (cartilaginous fish; eg, sharks/skates), Actinopterygii (ray-finned fish; eg, trout/tuna), and Sarcopterygii (lobe-finned fish; eg, coelacanths), and constitute more than 50% of the extant vertebrate species.

Fishes are the most ancient vertebrate group in which “classical” features of acquired immune responses have originated, namely, RAG enzymes driving somatic recombination, cell-bound and soluble BR and TR, and long-lasting memory for antigens. Lymphocytes of fish are cells showing the main morphological and functional adaptations that have been conserved during the evolution of jawed vertebrates, and that are specialized with diverse adaptation strategies in the various clades, as shown in Fig. 4.2. For the similarities among the basic features of the immune system in vertebrates, the importance of studying the fish immune-system is clear for evolutionary and applicative studies. Indeed, some fish species, and in particular the zebrafish *Danio rerio*, are actually employed as models for vertebrate physiology studies, including animal and human health.
After early speculations, more than 60 years of investigation on fish immunology has established that jawed fish species are endowed with a lymphocyte configuration that is conserved in the following orders, and include B-cell co-receptors and accessory molecules, TRαβ/γδ with accessory molecules, and a complete set of T-cell subpopulations.  

2.1 B Cells

B lymphocytes are characterized by two physiological stages: a resting state, where the cell surface-bound Ig serves as a receptor antigen and can receive input from other cells; and an activated state, where B cells secrete Ig of various classes. Activated B cells expand clonally, undergo dramatic morphological changes, and start to produce and secrete soluble antigen-specific Ig in body fluids. This feature had allowed for the early detection of antigen-specific antibodies in the blood of trout about 70 years ago.

2.1.1 Cartilaginous Fish

The most ancient extant gnathostomes are cartilaginous fish, and much information is available on the molecular immunobiology of skates, and, principally, sharks. The B lymphocytes of sharks were deduced by the presence of a clear antibody response against bovine albumin in 1967, and subsequently, by the
first quantitation in skates of Ig-bearing cells with an antiserum. Lymphocytes had been first described microscopically in the blood and intestines of sharks for their morphological similarity to their mammalian counterparts, but the possible presence of B-cell subpopulations was only deduced by the identification of diverse Ig-heavy chain genes. Cartilaginous fish express genes coding for IgM (monomeric and pentameric), IgW, and IgNAR. The IgNAR is particular to cartilaginous species, distinct from Ig- and TR-rearranging gene systems, and it is diversified by somatic mutation at a higher frequency than IgM and IgW. However, it is intriguing to observe that, despite the availability of much literature on the molecular and genetic features of cartilaginous fish Ig, there is a lack of information on immuno-morphology features of B cells and of their in vitro activities, with the exception of a work describing IgM-mediated cytotoxicity in a shark species.

In summary, little information is available on the cell biology properties of B lymphocytes of cartilaginous fish species, whereas much knowledge is accumulated on the B cells of bony fish.

2.1.2 Bony Fish

The possible presence of antibody-secreting cells in bony fish was discovered around 70 years ago, when a salmonid species was immunized against an antigen. Following that, antigen-induced antibody production with biochemical features similar to mammalian IgM was reported in 1966. The identification of a bony fish IgM-type antibody was done in sera and mucus, and was followed much later by the discovery of genes coding for IgD in 1997, and for IgZ/IgT in 2005. The immunobiology of B cells and antibody production in response to pathogens has been the subject of much research, boosted by the increasing demand of information for fish vaccination in aquaculture.

The IgM-secreting B cells are the predominant class of B lymphocytes, and the kinetic of IgM production in response to antigens appears to be conserved in all vertebrates, with the exception of coelacanth fish, the only vertebrate species where IgM genes are missing.

By using specific antibodies, the number of IgM-bearing lymphocytes have been determined in a number of fish species; their percentages range from 10 to 20% of leukocytes in nonmucosal tissues (head, kidney, spleen, and blood), to 1–5% in mucosal tissues (intestine, gills, and skin). This striking difference reflects the antigen route inside an organism: uptake through mucosa, recognition of nonself nature, molecule processing and triggering of responses, and delivery of information to B cells for systemic immunity and long-term effects.

The capability of IgM-B cells to secrete antigen-specific antibodies has been employed for many years to evaluate B-cell activity, by stimulating “in vitro” leukocytes from an immunized fish with an antigen and measuring the number of IgM-producing cells by ELISPOT, or the amount of IgM produced by memory B cells. The ELISPOT technology has reached an exquisite sensitivity, with the possibility of detecting single IgM-producing B cells. Molecular-biology
data have elucidated many aspects of immunogenetics of Ig heavy- and light-chain rearrangement, and recent investigations have described the extreme diversity between the Ig classes and B-cell repertoire of clonal responses in fish, and complex IgM and IgT responses during viral infections in trout. However, despite the amount of information describing molecular features of Ig genes, the knowledge of cell biology features of fish B lymphocytes is limited and mostly related to IgM-producing cells. The distribution of B cells in fish is prominently related to the tissue because mucosal surfaces have a low number of IgM B cells, whereas internal lymphoid tissues such as spleen and kidney show higher percentages. On the other hand, mucosa-associated IgT-B cells can be found in the intestine, in the skin, and in nasal epithelium. Regarding the mucosal epithelium of gills, a substantial expression of the IgD gene and a low expression of IgM and IgT has been recently observed in trout, suggesting a role for IgD-producing B cells in this tissue. The teleost-related molecule CCR7, which defined the presence of B lymphocytes that produce IgM, IgD, and IgT, has been preliminarily shown by ISH in mandarin fish.

From available data, it can be speculated that in teleost fish the amount of expressed Ig genes may follow a gradient IgM > IgT > IgD, but it is not clear whether this gradient is related to a similar number of B cells producing these Ig.

A striking feature of fish B cells is their phagocytic capability reported in trout, cod, and salmon; this capability led us to speculate about the evolution of this feature. By assuming that B cells inherited from invertebrate immunocytes the capability of phagocytosing particles also for food, then, by using novel vertebrate-associated features (eg, proteasome-associated genes), they can use phagocytosis also for reprocessing the nonself in the form of membrane-exposed antigens to perform antigen-presenting cell functions. Indeed, to support this hypothesis it should be remembered that lymphocytes from amphibians and reptiles also have phagocytic activity. Human B cells are not phagocytic (with the exclusion of B cell lymphomas), but they may act as antigen-presenting cells. Recently, an interesting work showed the possible presence of innate-type B lymphocytes in zebrafish with similarities to mammalian B-1 and B7 B-cell subpopulations.

Some knowledge is available on the ontogenesis and development of B cells in fish. This aspect was investigated in more detail in carp and sea bass, and what appears evident in all species investigated is the late appearance of B cells with respect to T cells.

2.1.3 Amphibians

Amphibians are terrestrial animals characterized by having a physiological dependence on the water environment, mainly related to reproduction and development. For this feature, amphibians are subjected to environmental stressors deriving from either terrestrial or water origin, and these stressors may induce immunodepression, which is considered to be a condition linked to severe global amphibian decline. Amphibians, and, in particular, the clawed frog *Xenopus*
laevis, have been the subject of investigations that had added impressive knowledge to vertebrate lymphocyte biology from 50 years ago. The first reports on antibody activity and the presence of plasma cells from amphibians were in 1968,\textsuperscript{54,55} followed years later by papers on lymphocytes.\textsuperscript{56–58} Since then, the distribution of IgM in splenic lymphocytes,\textsuperscript{59} the identification of two different types of B cell populations,\textsuperscript{60} and the first preparation of mAb against B cells able to separate these cells from other leukocytes were described.\textsuperscript{61} Subsequently, the preparation and use of mAb allowed for the first identification in Axolotl of different B-cell populations\textsuperscript{62} and of different Ig heavy-chain isotypes.\textsuperscript{63} As in fish, however, most of the work on amphibian B cells has been done by analyzing data from molecular biology experiments that showed the presence of four diverse Ig light chains\textsuperscript{64}, the sequencing of the Xenopus genome allowed for the definition of Ig heavy-chain classes of amphibians, namely, IgM, IgX, IgY, and IgE.\textsuperscript{65} The IgY appears to be present in mucosal tissues during larval development, then substituted by IgX,\textsuperscript{66} and the IgY gene is considered to be the ancestor of mammalian IgG and IgE.\textsuperscript{67,68} The data on characterization of B cells are scarce and mostly restricted to IgM-producing cells, where it was shown the in vitro proliferation induced by anti-IgM antibodies,\textsuperscript{69,70} and the distribution of IgM-positive cells in tadpoles and adult Xenopus. More recently, ELISA data employing antibodies specific for IgX showed the modulation of this Ig mucosal class after oral immunization of Xenopus.\textsuperscript{71} With respect to B cells, a more extensive literature on amphibian T cells is available (see further).

2.1.4 Reptiles

Among the Vertebrata classes, the reptile immunology is the least investigated for unclear reasons, but likely due to reptiles minor use as farmed or domestic animals, or as animal models in general. However, some knowledge is available on gene sequences of Ig classes present in reptiles (IgM, IgD, IgY, IgA),\textsuperscript{72} and the genome cloning of a lizard species (Anolis carolinensis) sheds light on the immunogenetics of reptiles.\textsuperscript{73} At present, little information is available in the literature describing lymphocytes in reptiles, although their presence can be deduced by a discrete amount of knowledge of the molecular biology of immunoglobulin genes.\textsuperscript{74–76} First data on the reptile acquired immune-system date back to 1963, with the determination of serological antibody responses in a turtle.\textsuperscript{77} Then, immune responses in the Tuatara were reported, without a characterization of effector cells.\textsuperscript{78} Following that, investigations were reported on Ig structure\textsuperscript{79} and Ig classes.\textsuperscript{80} The presence of cell-surface-associated IgM was demonstrated in a snake species,\textsuperscript{81} followed by revealing the percentages of B cells in tissues of a lizard (Calotes versicolor), employing rabbit antisera against IgM and IgY. These data showed B-cell percentages in lizard tissues in a range similar to that of other vertebrates. The presence of a cell-mediated antibody response, obtained through cooperation of B and T cells, was shown in reptiles by immunizing turtles against a proteic antigen, and then measuring specific antibody responses and in vitro proliferation of lymphocytes.\textsuperscript{82}
The phagocytic activity of B cells has also been shown in nonadherent PBL leukocytes from the turtle *Trachemys scripta*, thus showing the conservation of this “ancient” lymphocyte feature in reptiles.

### 2.2 T Cells

T cells of vertebrate species are responsible for cell-mediated immunity and for regulation of lymphocyte responses; they perform these activities by expressing in their surface TR and associated coreceptors for nonself recognition, and by producing families of cytokines that act as regulators of immune responses. T lymphocytes originate in the thymus—although there is a debate on the possible origin of some subpopulations in the intestine—and are considered to have preceded B cells in evolution, as supposed by functional and developmental data. Diverse subpopulations of T cells have been identified in mammals and classified on the basis of associated physiological activities, namely cytotoxic, helper, regulators, and IL-17-producing; all the genes associated with these activities have been found in ectothermic vertebrates. A common molecular feature grouping T-cell subpopulations is the presence of a CD3 coreceptor, despite these subpopulations very diverse behavior with respect to nonself recognition because some T cells that are mostly concentrated in mucosal tissues are spontaneously cytotoxic without MHC-mediated antigen presentation. In addition, it is interesting to observe that in some species the evolution allowed for unconventional assets of T lymphocyte physiology, such as the lacking of expressed MHC genes in pipefish or the lacking of CD4 and MHC genes in the genome of cod.

#### 2.2.1 Cartilaginous Fish

Extant cartilaginous fish can be considered the oldest representatives of jawed vertebrates having TR gene systems, and T lymphocytes were supposed to be present in cartilaginous fish because of the proliferation induced by lectins in lemon shark. Interestingly, first speculations on TR of cartilaginous species reported that the Ig molecule was the TR of sharks, until complete TR sequences had been identified sometime later, with typical signatures of TR genes present in jawed vertebrates. Subsequently, RAG1 and RAG2 genes were discovered in sharks, demonstrating that cartilaginous species may somatically rearrange Ig and TR, and that the physiological plan of acquired immune-responses present in all jawed vertebrates is assumed to be as old as sharks and skates.

As reported previously, most of the knowledge of cartilaginous-fish lymphocytes has been achieved on molecular assets of Ig and TR genes and on their expression; moreover, a gap exists with regard to functional studies on “in vivo” and “in vitro” lymphocyte responses, as these studies are still at an initial stage.

The anatomical localization of T cells in the shark thymus has been investigated in situ by using riboprobes for TRαβ, TRγδ, RAG-1, and TdT, and results showed a T-cell topography generally similar to that of other vertebrates,
but with a higher number of thymocytes expressing TRγδ. By northern blotting analysis, the expression of TRγδ resulted similarly with TRαβ in examined tissues, suggesting a possible contribution of innate immune lymphocytes to lymphocyte activities in sharks. However, putative functional features of T lymphocytes in sharks and skates can only be deduced by the presence of functional TR transcripts, and thus much work is required to investigate in vitro and in vivo immunophysiology of T cells.

Recently, the sequencing of elephant shark genomes provided interesting insights into the evolution of cartilaginous fish T-cell responses. This work provided molecular evidence for the presence of CD8- and Th1-related master gene regulators, and the lacking of Th2-related master gene regulators, thus hypothesizing the presence of complete sets of T lymphocytes, including CD4/Th cells from Osteichthyes onward, as shown in a recent study.

### 2.2.2 Bony Fish

First investigations on T-cell activities in bony fish reported transplant rejection by T cells, and B–T-cell cooperation by the hapten–carrier effect in secondary in vitro stimulation. However, these activities were only deduced in the absence of specific lymphocyte markers. The use of monoclonal antibodies and the obtainment of IgM-purified cells tested in vitro with LPS and lectins, established the presence of lymphocyte subpopulations homologous to mammalian B and T cells. Subsequently, the presence of an immunological memory and B–T-cell cooperation was assessed by the analysis of an in vitro secondary antibody-response from an antigen-sensitive precursor cell pool. In contrast to B lymphocytes, specific markers for T cells of teleosts have been obtained in 1995 (mAb DLT15, which recognizes pan-T cells in sea bass), and in 1998 (mAb WCL38, which recognizes carp mucosal T cells). Descriptions of these T-cell markers have been previously reported in detail. Historically, the first direct in vivo observation of a functional T-cell activity in fish, namely an allograft rejection, was reported in sea bass, using the DLT15 mAb. More recently, specific mAbs for salmonids CD3ε and CD8α, CD8α and CD4 for gibel carp, have been obtained.

An interesting model to investigate in vitro cellular activities of T cells has been achieved through the development of T-cell lines from carp. These T-cell clones express both CD8- and CD4-related genes, may proliferate in vitro, and represent an interesting model to investigate functional in vitro properties of cyprinid T cells.

The obtainment of molecular and cellular markers for T-cell molecules boosted the knowledge of functional in vivo and in vitro activities in bony fish. In rainbow trout, it was shown the typical T-cell recombination profiles already known in mammals by spectra-typing analysis of somatic recombination of TRβ and of TRγ CDR3 sequences, induced in vitro and in vivo by viral stimulants. The presence of cytotoxic T cells (CTL) was shown in trout and carp through the expression of TRβ and CD8α, and in sea bass by
in situ hybridization with TRβ mRNA riboprobes in effector T cells involved in allograft rejection.\textsuperscript{113} The activities of T cells against viruses are a hallmark in vertebrate immunity, and there is extensive literature on antiviral activities in fish; of particular mention are studies on the in vitro production of ginbuna carp CTL clones specific for viral antigens,\textsuperscript{114,107,108} and studies on T-cell proliferation induced by nodavirus infection in Atlantic halibut.\textsuperscript{115}

Bony fish express all the sets of genes coding for T-cell activities as they are known in mammals, and a recent review summarized in detail the knowledge of functional activities of CD8- and CD4-bearing T cells.\textsuperscript{7} In the sea bass, a recent study showed that the gills express all master genes for cytotoxic and helper activities of T cells, leukocytes proliferated in response to lectins, and proliferating cells contained T cells.\textsuperscript{17} Interestingly, and in contrast to the intestine, the gills showed little expression of RAG genes, suggesting different pathways of T-cell differentiation and activities between the intestine and gills. The antigen-induced differentiation of lymphocytes and the maintenance of a memory are pathways in which dendritic cells (DC) play a fundamental role; a recent study described in zebrafish conserved immunophenotypes for DC, with cell activities showing antigen-specific CD4(+) T-cell activation similar to mammalian counterparts.

\subsection{2.2.3 Amphibians}

As for B cells, most of the knowledge regarding amphibian T lymphocytes has been achieved in \textit{Xenopus}. Early reports on identification of T cells in \textit{Rana} by mitogen-induced proliferation date back to 1978,\textsuperscript{116} and by using agglutination and rosetting technologies the presence of T- and B-lymphocytes was later confirmed.\textsuperscript{117} Cytotoxic and helper T cells were identified in \textit{Xenopus} through involvement of IL-2 stimulation.\textsuperscript{118} First mAbs against thymocytes and T cells were reported soon after,\textsuperscript{119} whereas a mAb for the T cell-specific marker molecule CD3 was only available much later.\textsuperscript{120} The production of IL-2 by T cells stimulated with mitogens was later confirmed,\textsuperscript{121} together with the expression of MHC\textsubscript{II} antigens on T cells,\textsuperscript{122} and MHC-restrictions of T-cell proliferative responses.\textsuperscript{123} The metamorphosis in amphibians, largely investigated in \textit{Xenopus}, is a process involving extensive remodeling of lymphoid tissues and of the thymus, with the persistence of thymus-originated T cells in adult frogs,\textsuperscript{124} the in vitro culturing of CD4/CD8 thymocytes,\textsuperscript{125} and the development of T cells and natural killer-cells in thymectomized animals.\textsuperscript{126} Regarding functional activities of T cells in \textit{Xenopus}, the presence of primary- and memory-CD8 responses against viral infections has been reported,\textsuperscript{127} along with the restricted diversity of TCR delta chains in Axolotl.\textsuperscript{128}

\subsection{2.2.4 Reptiles}

Although the information on reptilian T cells is scarce, for the possible reasons already outlined in Section 2.4, it can be easily argued that T cells have been
found in all of the reptile species investigated. An early study on a lizard species reported on the possible presence of cytotoxic and regulatory T cells; subsequent investigations on T cells were reported for a lizard, and a T cell-mediated allograft rejection was described in a skink species. Most available information on T-cell responses in reptiles is on mitogenic responses related to seasonal variations in lizards. Basic T-cell functions such as mitogen-stimulated proliferation and mixed leucocyte reaction (MLR) have been investigated in Tuatara, the only extant species of an ancient order, and found similar to other vertebrates. By employing a commercially available rabbit anti-human CD3 marker immunoreactive against a green turtle after an in-vivo injection of PHA, it has been demonstrated that these injections had activated T lymphocytes in reptiles in a manner similar to mammals.

3 CONCLUSIONS

Studies on the developmental appearance of lymphocytes in ectothermic vertebrates revealed the precedence of T lymphocytes with respect to B lymphocytes, and this could reflect the origin of these cells, with T cells evolutionary older than B cells. Also, the sites of lymphopoiesis are more evident for T cells, being the thymus and/or the intestine, whereas the sites for B-cell lymphopoiesis are less clear. In cartilaginous fish, the spleen, Leydig’s organ, and the spiral valve may be a site for B-cell development. In teleosts, the kidney is the principal source of B-cell development, and amphibians appear to use several different sites (spleen, bone marrow, and/or kidney).

Importantly, emerging functional and morphological evidence suggests the presence of innate-type lymphocytes in fish that display spontaneous antibacterial activity (CD4+ T cells) and spontaneous killing without antigen presentation (TRγδ/CD8α-T cells). In addition, TRγδ T cells are important effectors in mucosal tissues, and they have been shown to be functionally active in fish, whereas in mammals they have been considered to be a bridge between the innate and adaptive immune-systems.

Last but not least, an important but poorly investigated feature in the physiology of vertebrate lymphocytes is the relationship between the nervous system and immune system. It should be remembered that the intestine of all vertebrate classes contains high percentages of lymphocytes, which are mainly T cells, and studies in fish showed the expression of receptors for serotonin (5-HT3), which has been found to be implicated in T-cell proliferation.

REFERENCES


51. Zhu LY, Lin AF, Shao T, Nie L, Dong WR, Xiang LX, Shao JZ. B cells in teleost fish act as pivotal initiating APCs in priming adaptive immunity: an evolutionary perspective on the origin of the B-1 cell subset and B7 molecules. *J Immunol* 2014;192:2699–714.


Chapter 5

Vertebrate Cytokines and Their Evolution

Christopher J Secombes*, Tiehui Wang*, Steve Bird**
*Scottish Fish Immunology Research Centre, University of Aberdeen, Zoology Building, Tillydrone Avenue, Aberdeen, United Kingdom; **Molecular Genetics, Department of Biological Sciences, University of Waikato, Hamilton, New Zealand

1 INTRODUCTION

This review will outline what we know about cytokines and their evolution within vertebrates. It will focus on the interleukins, as key cytokines involved in the regulation of adaptive immunity, but will also include the TNF family of cytokines. It will not cover the evolution of interferons, which is described in chapter: Antiviral Immunity: Origin and Evolution in Vertebrates, or the chemokines, which have had many recent reviews (eg, Refs. [1–6]). It will build upon our recent reviews of fish cytokines,7–10 to update on the latest fish discoveries and discuss, in the context of cytokine evolution, within the other—mostly nonmammalian—vertebrate classes.

2 THE IL-1 CYTOKINE FAMILY

Interleukin-1β was the first family member discovered within mice and humans, and has been characterized within a wide variety of mammalian species.11 The interleukin-1 (IL-1) family of cytokines has four main members: IL-1α (IL-1F1), IL-1β (IL-1F2), IL-1RN (IL-1F3), and IL-18 (IL-1F4), with a number of molecules showing clear homology to this group: FIL-1δ or IL-36RN (IL-1F5), FIL-1ε or IL-36α (IL-1F6), IL-1H4 or IL-37 (IL-1F7), IL-1H2 or IL-36β (IL-1F8), IL-1ε or IL-36γ (IL-1F9), IL-1Hy2 or IL-38 (IL-1F10), and IL-33 (IL-1F11). These are encoded at three separate loci within the human genome, with IL-1F4 and IL-1F11 lying on chromosomes 11 and 9, respectively, and the remaining nine present at a single locus on chromosome 2.12 All members of the IL-1 family belong to the β-trefoil superfamily, as β-sheets are present within their tertiary structure, which fold to form a trefoil-like structure.13,14 In addition, each member contains an IL-1 family signature pattern or motif, taken from a selected conserved region in the C-terminal section, with the following consensus pattern: [FC]-x-S-[ASLV]-x(2)-P-x(2)-[FYLV]-[LI]-[SCA]-T-x(7)-[LIVM],
as found in the PROSITE database. To date, only IL-1F2, IL-1F3, IL-1F4, and IL-1F5 homologs have been discovered outside of mammals, which are discussed later in the chapter.

2.1 Vertebrate IL-1F2

In mammals, IL-1 is a major mediator of inflammation, and in general initiates and/or increases a wide variety of nonstructural, function-associated genes that are characteristically expressed during inflammation, particularly other cytokines. It is one of the key mediators of the body’s response to microbial invasion, inflammation, immunological reactions, and tissue injury. Both in vivo and in vitro experiments have shown that IL-1α and β have similar, if not identical, multiple biological effects which have been well-cataloged, with both forms affecting nearly every cell type and sharing a common receptor on target cells. Work, however, has shown that the endogenous roles of IL-1α and IL-1β are different where IL-1β, but not IL-1α, is a potent activator of the humoral immune response, and has been shown more recently to induce local inflammation in delayed-type hypersensitivity responses. IL-1α and IL-1β are produced by a wide range of cell types, including neutrophils, natural killer-cells, B-lymphocytes, T-lymphocytes, and cells of the central nervous system, but are made in the largest amounts by blood monocytes and tissue macrophages. These cells act as an important source because of their strategic locations, ability to synthesize large amounts of IL-1, and ability to process the IL-1 precursor more effectively than other cells.

IL-1α and IL-1β also differ substantially in relation to localization, maturation, and secretion. Both forms are translated into pro-IL-1α and pro-IL-1β (both of which are 31 kDa), which have no signal peptide, and thus remain in the cytosol and do not accumulate in any organelles. Enzymes are then involved which cleave them, to produce mature forms, which then augments the activity of these peptides. Pro-IL-1α, but not pro-IL-1β, is cleaved by calcium-dependent membrane proteases called calpains, to give a 17 kDa mature peptide. IL-1α is rarely found in the circulation or extracellular biological fluids, and is generally associated with the plasma membrane of the producing cells, acting locally. Pro-IL-1α is just as active as the mature form, and appears to remain intracellular and act at this level. Pro-IL-1β remains primarily cytosolic and is only active once proteolytically cleaved and transported out of the cell; several mechanisms control this. A cysteine protease, IL-1β converting enzyme (ICE) or caspase-1, is the enzyme responsible for this cleavage of pro-IL-1β, and requires an exogenous source of adenosine 5’-triphosphate (ATP) to trigger it. Both ICE and pro-IL-1β can coexist without mature IL-1β being formed, suggesting ICE activity must be regulated (Schonbeck et al., 1997). In humans, to generate mature IL-1β, ICE specifically cleaves the bond between the aspartate amino acid (aa) at position 116 and alanine aa at position 117. Cleavage of pro-IL-1β seems an obligatory step for IL-1β release, as treatment with ICE inhibitors fails to cause detectable levels of extracellular pro-IL-1β,
although some ICE inhibitors cause pro-IL-1β to be secreted instead of the mature form. In addition, there are alternative mechanisms present for the processing of the pro-IL-1β, and several sites in pro-IL-1β have been shown to be vulnerable to cleavage by enzymes in the vicinity of alanine 117. Proteolytic enzymes, including chymase, elastase, cathepsin G, collagenase, and matrix metalloproteinases as well as neutrophil- and macrophage-derived serine proteases, have been shown to cleave pro-IL-1β into a form similar in size and specific activity to the ICE-processed IL-1β.

Due to the importance of IL-1 in the immune system of mammals, and evidence suggesting its biological activity in the immune system of nonmammalian vertebrates, work has focused on isolating this gene in selected species. To date, no homolog of IL-1F1 has been found in nonmammalian vertebrates, however, IL-1F2 has been characterized within a wide range of organisms.

2.1.1 Nonmammalian IL-1F2

IL-1β has been cloned from selected species within each nonmammalian vertebrate group, which include birds, reptiles, amphibians, bony fishes, including representatives from Salmoniformes, Cypriniformes, Gadiformes, Percomorpharia, Elopomorpha, Carangimorphariae, Osmeriformes, Scombriformes, and cartilaginous fish. Overall, aa identities of these molecules with their mammalian homologs are generally low, however, early studies showed that features such as the IL-1 family signature are reasonably well-conserved, and that highest aa homology was generally found where β-sheets were predicted to form, providing evidence that each protein had a similar folding pattern indicative of a β-trefoil family member. Subsequently, the crystal structure of chicken IL-1β and the 3D modeling of selected fish IL-1β has supported this conclusion. Interestingly, although there is conservation of a 12 β-sheets structure, when the chicken and fish ligands are modeled while interacting with their respective receptors, there is a high level of variability of positions involved in receptor binding. However, the mode of binding and overall shape of the ligand–receptor complex appears to be maintained, implying that each species has evolved its own unique interleukin-1 signaling system through ligand–receptor coevolution.

The involvement of IL-1β in the nonmammalian vertebrate immune response is supported by expression studies in chicken, Chinese soft-shelled turtle, Xenopus, numerous bony fish, and cartilaginous fish. In chicken, IL-1β is quickly induced in blood monocyte derived macrophages, reaching optimal levels within 1 h after LPS treatment, and is upregulated in Sertoli cells exposed to LPS. The Chinese soft-shelled turtle IL-1β was significantly increased within the spleen and intestine that had undergone acute cold stress after Aeromonas hydrophilia infection. Within Xenopus, the IL-1β transcript was inducible in vivo, either following injection with LPS, or after exposure to immunotoxins. For bony fish IL-1β, many studies have investigated its expression in vitro and in vivo, where it has been shown to be significantly enhanced in immune tissues, primary cultures of head kidney (HK) cells, isolated macrophage cultures, or
available fish cell-lines, in response to immunostimulants and immune-response modifiers, such as LPS, PMA, and imiquimod\textsuperscript{36,41,43,44,61–64} or disease-causing agents, such as viruses, bacteria, or parasites\textsuperscript{45,47,49,65,66} IL-1\(\beta\) expression has also been shown to be modulated in fish in response to tissue damage, UV exposure, herbicides, and diet.\textsuperscript{67–71} In addition, incomplete spliced variants of the IL-1\(\beta\) gene have been identified in a number of fish, which are generally expressed at lower levels, relative to the mature transcript, and are not universally present in all tissues that express IL-1\(\beta\).\textsuperscript{61,63,72} It is thought that these incompletely spliced forms may be involved in the regulation of the mature transcript, as has been seen in mammals in certain situations.\textsuperscript{73} Lastly, expression studies in \textit{Scyliorhinus caniculus} have also shown that the IL-1\(\beta\) molecule is biologically relevant within cartilaginous fish immune responses.\textsuperscript{51} IL-1\(\beta\) could be induced using LPS in vivo in the spleen and testes, and in vitro in splenocytes.

\subsection*{2.1.2 Multiple IL-1\(\beta\)2 Genes in Fish}

Within trout, carp, and the channel catfish, more than one IL-1\(\beta\) gene was found to be present,\textsuperscript{39,74,75} with each copy sharing a high level of aa identity. In trout, the second gene (IL-1\(\beta\)2) has a similar intron–exon organization to the first IL-1\(\beta\) gene (IL-1\(\beta\)1) sequenced,\textsuperscript{36} except for small differences in the sizes of the introns and exons. Induction of IL-1\(\beta\)2 expression has also been observed, where a clear dose-dependent induction of the trout IL-1\(\beta\)2 is seen in cultured trout leucocytes in response to LPS. In carp, along with IL-1\(\beta\)1, there were two different IL-1\(\beta\)2 sequences identified, and IL-1\(\beta\)2 mRNA expression was upregulated in HK cells in vivo, similar to carp IL-1\(\beta\)1, after infection with \textit{Trypanoplasma borreli} and in vitro by stimulation with LPS.\textsuperscript{39} At the time of their discovery, the existence of two IL-1\(\beta\) copies in the trout and carp genome came as no surprise, due to the tetraploid nature of these species.

More recently, additional copies of the IL-1\(\beta\) gene have been discovered in trout and salmon, called IL-1\(\beta\)3,\textsuperscript{76} and in diploid species, such as the Japanese flounder, called nIL1\(\beta\)-L1,\textsuperscript{77} and in gilt-head seabream, called IL-1Fm2.\textsuperscript{78} Although clearly related to IL-1\(\beta\) genes already discovered in fish, these copies share low aa identity to them, which is clearly shown within phylogenetic analysis, where they form their own distinct group. Furthermore, where gene organization has been determined, this is different between each IL-1\(\beta\) gene. From this, it has been clearly shown that fish have two very clear types of IL-1\(\beta\) that can be found within a wide variety of fish species, using available genomes and sequence databases.\textsuperscript{76,78} In addition, investigating where each of these genes are found within genomes has shown that they are located in different regions; however, there is conservation of gene synteny around them, which suggests that they may have arisen from a fish-specific genome duplication event.\textsuperscript{79} Although a lot remains to be done, in order to understand the role of each type of IL-1\(\beta\) gene in fish, preliminary investigations have been carried out. In flounder and trout, clear differences in the constitutive expression of the two groups of IL-1\(\beta\) genes in tissues, such as the gills and spleen, can be seen. In trout, IL-1\(\beta\)3 was upregulated in primary macrophage cultures, in response to pathogen-associated molecular
patterns (PAMPs), proinflammatory cytokines, and viral infection, suggesting that it also has a role in inflammation and host defense. In sea bream, although IL-1Fm2 could be slightly upregulated in a variety of tissues following exposure to PAMPs or bacterial infection, the combination of PAMPs with the recombinant IL-1Fm2 was able to significantly upregulate expression in macrophages and granulocytes. Lastly, recombinant IL-1Fm2 could also activate the respiratory burst of sea-bream phagocytes, and synergistically induce the expression of IL-1β, TNF-α, IL-8, and IL-10 when combined with PAMPs.

2.1.3 Processing of Nonmammalian IL-1F2

Functional studies with the recombinant (r)IL-1 protein have been carried out in nonmammalian vertebrates, based on predictions of a possible start of the mature peptide from alignments with mammalian species. It is clear from the proteins that have been produced for chicken and a wide range of bony fish, that they are active, and even peptides from the mature peptide region can elicit some responses. In chicken, the rIL-1β-induced expression of a CXC chemokine in chicken fibroblasts and plasma cortisol levels were significantly upregulated upon injection with IL-1β, in a similar fashion as that seen in mice and rats. In fish, early studies in trout showed that IL-1β induces the expression of itself, COX2, and MHC II in macrophages in vitro, and when injected into the peritoneal cavity of fish it significantly enhances peritoneal leucocyte phagocytosis and resistance to challenge with the bacterium Aeromonas salmonicida. More recent studies in the grass carp have found that grass carp rIL-1β can simultaneously upregulate grass carp IL-1β and TGF-β1 expression via NF-κB and MAPK6, signaling in grass carp HK leukocytes; moreover, its use as an adjuvant within flounder can enhance the antibody titer against the antigens, bovine serum albumin, or green fluorescent protein.

Despite proven activity of these recombinants, how nonmammalian IL-1β is processed into its mature form and the location of the actual cut site, is still in question. Analysis of nonmammalian vertebrate IL-1β genes shows that no apparent signal peptide is present, indicating that these molecules are secreted through a nonclassical pathway not involving the golgi/endoplasmic reticulum route. In fish, evidence exists showing that IL-1 is made as a precursor that is subsequently cleaved, as seen in mammals, but the mechanism of cleavage is less clear. In mammals, ICE cleaves immediately after an aspartic acid, however, analysis of this region in known nonmammalian vertebrates typically reveals no clear ICE cut site. Recent investigations within sea bass have provided evidence that IL-1β is cleaved by ICE at a phylogenetically conserved aspartic acid present in all known sequences. Interestingly, in chicken a different phylogenetically conserved aspartic acid appears to be the preferred cleavage site, which could mean processing may be class- and/or species-specific. Further evidence of the involvement of ICE in the processing of mature IL-1β has been shown in studies using the zebrafish model. Primary leucocytes from adult zebrafish display caspase-like activity that results in processing of IL-1β during Francisella noatunensis infection, which can be considerably reduced with the use of caspase-1.
or pancaspase inhibitors. Similarly, tissue injury of larvae induces leucocyte expression of IL-1β, which was shown to activate leucocytes throughout the embryo,\(^{92}\) with the response drastically reduced with ICE or pancaspase inhibitors.

It is important to note that mammalian studies have also shown that the precursor IL-1β can be cleaved at different aa residues and by different cleavage enzymes to obtain a biologically active molecule. These include proteases that can be found in inflammatory fluids, at sites of neutrophil, lymphocyte, or macrophage infiltration, and can therefore participate in the generation of active IL-1β. Examples include granzyme A, elastase, trypsin, and chymotrypsin, producing biologically active IL-1β.\(^{93,94}\) In addition, a mast-cell chymase and various extracellular matrix metalloproteinases (MMP) have also been shown to produce biologically active IL-1β.\(^{31,95,96}\) Evidence of ICE-independent mechanisms have also been shown to exist in fish, where the stimulation or infection of sea-bream macrophages could induce ICE-independent processing and release of IL-1β.\(^{78}\)

### 2.2 Vertebrate IL-1F4

IL-1F4, also known as IFN-γ inducing factor (IGIF) or IL-18, was initially cloned in mice from a partial aa sequence, deduced from the purified protein\(^{97}\) and then later in humans.\(^{98}\) Analysis of the primary aa sequence of both human and mouse IL-18 indicated the presence of an IL-1 signature-like sequence, and it was proposed that IL-18 contained 12 strands of β-sheets, forming the β-trefoil fold, typical of the IL-1 family.\(^{99}\) IL-18 exists as an inactive pro-IL-18 molecule until it is cleaved by ICE to generate the mature active form of IL-18.\(^{100,101}\) It appears that the correct folding of IL-18 occurs prior to processing with ICE, at the level of pro-IL-18.\(^{102}\) The structure and processing of this molecule suggest that IL-18 and IL-1 originate from a common ancestor, although the target specificities and the receptor systems are completely different. It has also been reported that the IL-1 receptor-related protein is the functional component of the IL-18 receptor.\(^{103}\) In contrast to what is observed with IL-1β, constitutive gene expression of IL-18 has been seen within various cell types, including unstimulated, freshly isolated human PBMC and murine splenocytes,\(^{104}\) and hematopoietic cell lines of myelomonocytic origin.\(^{105}\) It has been shown that the promoter for IL-18 does not contain a TATA box, and that promoter activity upstream of exon-2 acts constitutively.\(^{106}\) It is also found that the 3′−UTR of human IL-18 lacks the AUUUA destabilization sequence, allowing for more sustained levels of the polyadenylated species and translation into protein. The presence of preformed pro-IL-18 in the cell would allow for a very rapid production of the mature, active molecule once the cell is activated.

Currently, within nonmammalian vertebrates, IL-18 has only been cloned within birds and bony fish. It was first isolated from chicken,\(^{107}\) but has also been discovered in turkey,\(^{108}\) duck,\(^{109}\) trout, and fugu.\(^{110}\) In addition, an analysis of available genomes and sequences available within the databases has identified that this gene can be found within each major vertebrate group (Fig. 5.1A). Highest
### FIGURE 5.1

Multiple alignment of selected vertebrate IL-18 (A) and IL-1ra and IL-1F5 (B) amino-acid sequences, generated using ClustalX. Highlighted in red are the IL-1 family signatures. Above the human IL-18 sequences are the regions that contain the 12 β-sheets and the potential ICE cut site is indicated (↓). Accession numbers for IL-18 are: human, AAK95950.1; anole, FG762485.1; chicken, NP989939.1; Xenopus, XP00294250.2; coelacanth, AGH59335.1; trout, CADB3552.2; sea bass, FM014619.1; dogfish, CX789574.1. Accession numbers for IL-1F3 are: human, CAA36262.1; chicken, CCD83122.1. Accession numbers for IL-1F5 are: human, NP775262.1; chicken, CCE26532.1. Anole IL-1RN/IL-1F5-like sequences are all predicted from a genomic scaffold sequence, accession number GL343238.1.
identity is seen within the areas encoding the 12 β-strands, involved in the β-trefoil structure, and there is good conservation of a variation on the IL-1 signature, F-x(12)-F-x-S-x(6)-F-L. However, unlike IL-1F2, sequence comparison does reveal a possible site for cleavage by ICE, at an aspartic acid, indicating that it may be processed in a similar way as its mammalian counterpart. Using this predicted cut site, a chicken and duck mature rIL-18 has been produced that stimulates the synthesis of IFN-γ in cultured primary chicken spleen-cells and proliferation of duck lymphocytes. More investigations into the bioactivity of chicken IL-18 have also revealed it as a major growth factor for CD4+ T-cells, and can stimulate IFN-γ release in these cells, indicating the conservation of a Th1-like response in this vertebrate group. This has prompted many investigations looking at its use as an adjuvant alongside viral vaccines toward infectious bronchitis virus, infectious laryngotracheitis virus, fowlpox virus, infectious bursal disease virus, avian influenza, and Newcastle disease virus. In many cases, beneficial effects have been seen, which include higher protection, and an enhancement in the immune response in the form of significantly higher levels of antibodies and lymphocyte proliferation.

Although no functional studies exist for fish IL-18, expression studies in trout do show high constitutive expression in immune and nonimmune tissue, but no modulation of expression in HK cells stimulated with LPS, polyI:C, or trout rIL-1β, similar to what is seen in mammalian studies, where IL-18 is stored as an inactive precursor in the cytoplasm of cells and its processing by ICE stimulates release. Lastly, in trout, an alternatively spliced form of IL-18 mRNA was also identified, with a 17 aa deletion in the precursor region, which is also widely expressed, but at much lower levels than the unspliced form. Interestingly, expression is upregulated by LPS and polyI:C in HK cells, which suggests alternative splicing could be regulating IL-18 activities in trout.

2.3 Bird and Reptile IL-1F3 and IL-1F5

Other than the family members previously discussed, the question of which IL-1F homologs exist outside of mammals requires investigation. Recent work in chickens has uncovered the existence of gene sequences that encode for homologs of IL-1F3 and IL-1F5. IL-1F3, also known as IL-1 receptor antagonist (IL-1RN), is a specific antagonist of IL-1, binding to the IL-1 receptor without transmitting any signal into the cell. Unlike IL-1α/β, IL-1RN has only one binding site for IL-1RI, which explains the absence of signal transmission. The IL-1RN cDNA encodes a 177 aa protein containing a 25 aa signal peptide, allowing the protein to be secreted into the extracellular environment, named secretory IL-1RN (sIL-1RN). Upon LPS stimulation, human blood monocytes initially express this gene, but after 24 h the primary transcript in these cells is an intracellular form of IL-1RN (icIL-1ra), which arises due to alternative splicing of the sIL-1RN mRNA, which modifies the exon encoding for the signal peptide. A second form of icIL-1RN is also
generated by alternative splicing of the mRNA, and has been termed icIL-1RN type 2, and these two intracellular isoforms may represent a reservoir of IL-1RN, released upon cell death, whose function is to limit the proinflammatory action of cell debris. Within chicken, two novel chicken IL-1 family sequences have been identified, representing sIL-1ra and icIL-1ra structural variants of the IL-1RN gene. In addition, both structural variants had two further putative splice variants (SVs), which is different in humans, where alternative splicing of icIL-1RN gives three different transcripts, and no known SVs for sIL-1RN exist. In functional studies using the full length sIL-1RN and icIL-1RN, inhibition of chicken IL-1F2 activity was shown within a macrophage cell line, however the four SVs did not show inhibition.

IL-1F5, also known as IL-36 receptor antagonist (IL-36RN), shows a strong relationship to IL-1RN, sharing 44% aa identity. It is a receptor antagonist of the receptor named IL-1 receptor related protein (IL-1Rrp2), and inhibits the multiple stimulatory effects induced by the agonists, IL-1F6, IL-1F8, and IL-1F9. Both IL-1F2- and LPS-induced inflammation has been shown to be antagonized by IL-1F5 in the brain of mice and rats, leading to an antiinflammatory response. Within chicken, a sequence representing a putative IL-1F5 homolog has been identified with high similarity to both chicken IL-1F2 and IL-1F3. Although it exhibits low homology with its mammalian orthologs, high conservation is observed in the areas that are predicted to encode the β-strands, important for the β-trefoil structure. Similar to what is found in humans, neither a prodomain nor a signal peptide appears to exist in the N-terminal of the predicted protein sequence; however, a putative nuclear export sequence (NES) that is involved in the controlled transport of proteins across a nuclear membrane, is present, which is absent in mammalian IL-1F5. Unlike the human gene, this gene showed high constitutive expression in all tissues and cell subsets examined. Expression was shown to be significantly upregulated following bacterial infection, but significantly downregulated by infectious bursal disease virus (IBDV) in a line of birds which are susceptible to this disease. Functional activity of the protein was attempted, and inhibition of LPS-mediated upregulation of IL-1F2 and iNOS expression in a chicken macrophage cell line was shown. However, it is not known whether homologs of IL-1F6, IL-1F8, and IL-1F9 exist in chickens, so the mechanism of action remains unclear.

How each of the chicken IL-1F members are arranged in the genome is of great interest because in humans IL-1F3 and IL-1F5 are grouped with a number of other members, including IL-1F2. It is known in chickens that neither IL-1F3 nor IL-1F5 are found near IL-1F2 in the genome, however, it is unclear if both of these newly discovered family members are close to each other, as the region where IL-1F3 is located has not yet been sequenced, and IL-1F5 is located on unfinished contigs. However, evidence from the reptile genome suggests that these two family members may exist together in the genome, as within a scaffold (Accession no. GL343238.1) up to six individual sequences with strong
homology to mammalian IL-1F3 and IL-1F5 can be identified (Fig. 5.1B), although their exact identity is uncertain.

2.4 Novel Fish IL-1F

Although only IL-1F2 and IL-4 have been discovered in fish, candidate genes have been characterized in a few fish species that are related to the already characterized fish IL-1F genes, but that have no direct homolog in mammals. This was first discovered in trout, where the unique gene organization, together with its location in the genome and low homology to known family members, suggested that this was a novel IL-1F (nIL-1F) member. However, the predicted aa sequence was found to contain an IL-1F signature region, a potential IL-1 converting enzyme cut site, and by building a 3D model, predicted to have a β-trefoil structure and be able to bind to the IL-1RI receptor protein. In addition, expression levels were increased either following infection or stimulation of macrophages with LPS or trout rIL-1β. Interestingly, although not a direct homolog of mammalian IL-1F3, the functional activity of the recombinant protein could antagonize the effects of trout recombinant IL-1β by changing inflammatory gene expression in a trout macrophage cell line, suggesting it could be a novel antagonist of fish IL-1β. More recently, a homolog to this novel gene has been isolated from Japanese flounder, where its expression can be increased in PBL following LPS stimulation, and in grass carp, where the nIL-1F can bind the type I IL-1β receptor and attenuate rIL-1β activity, confirming that it is a receptor antagonist.

3 THE IL-2 CYTOKINE FAMILY (γC CYTOKINES)

The IL-2 cytokine family (also known as γ-chain (γC) cytokines) are those that signal through a receptor containing the γC (CD132). This family consists of IL-2, IL-4, IL-7, IL-9, IL-15, and IL-21, which have been well-studied in mammals, in addition to IL-2-Like (IL-2L), IL-15L, and IL-4/13 molecules, which have been identified recently.

3.1 IL-2 and IL-2L

IL-2 genes have been described in mammals, birds, and recently in teleost and cartilaginous fish. The homology of IL-2 protein sequences between lineages is low, for example, avian IL-2 shares 18.4–27.3% aa identity to mammalian IL-2, and the trout IL-2 shares 17.8–23.2% aa identity with avian and mammalian IL-2. The identity of IL-2 in different lineages is supported by the conserved synteny and gene organization. IL-2 and IL-21 are closely related cytokines that are next to each other, not only in mammals, but also in birds, teleost fish, and the elephant shark.

Mammalian IL-2 is a pleiotropic cytokine produced after antigen activation. It drives T-cell growth, augments NK cytolytic activity, induces the differentiation
of regulatory T cells, and mediates activation-induced cell death. Bird IL-2 has effects comparable to mammals, including lymphocyte proliferation, activation of NK cells, and clearance of intracellular pathogens. The bioactivity of fish rIL-2 has only been described in rainbow trout. The trout rIL-2 protein increases the expression of at least two transcription factors, STAT5 and Blimp-1, as well as IFN-γ, γIP, and IL-2 itself.

An IL-2-Like gene (IL-2L) has also been discovered in several teleost fish genomes, for example, fugu, tetraodon, and stickleback. The fugu IL-2 and IL-2L molecules share only 20.5% aa identity. The fish IL-2 and IL-2L are located next to each other and may have arisen from a local gene duplication.

No functional studies have been performed to date.

3.2 IL-4 and IL-4/13

IL-4 was initially identified as a T cell-derived B cell growth factor and an immunoglobulin switch factor for IgG1 and IgE, and was later found to also act on T cells, mast cells, and other cells. Although they only share low aa sequence identity to IL-4 (eg, 23% in humans and 22% in cows), IL-13 was found to mediate virtually all of the IL-4 actions on nonhematopoietic cells, and to some degree on hematopoietic cells. IL-4 and IL-13 are present in mammals and birds, where they are gene neighbors linked to KIF3A and RAD50, respectively. This locus contains several other cytokines, including IL-3, IL-5, and granulocyte-macrophage colony-stimulating factor, and forms the type 2 cytokine gene-cluster. Mammalian IL-4 and IL-13 are Th2-type cytokines with pleiotropic functions in immunity. They mediate resistance to many gastrointestinal parasites, but also promote allergic inflammation, asthma, and fibrosis.

In chicken, coadministration of the DNA vaccine against Newcastle disease with the IL-4 gene resulted in highest IgY levels and increased protection.

IL-4 and IL-13 can bind to a common receptor composed of the IL-4Rα and IL-13Rα1 subunits (type II receptor), although they can individually bind to the type I receptor composed of the IL-4Rα and γC subunits (IL-4), or to the IL-13Rα2 receptor (IL-13). All of the receptor subunits have been cloned in fish, with two copies of each in salmonids, due to the 4R whole genome duplication (WGD). However, clear orthologs of IL-4 or IL-13 are missing in other vertebrates. The first IL-4/IL-13 related gene, IL-4/13A, was identified in the teleost pufferfish (Tetraodon nigroviridis) genome, next to RAD50. A second IL-4/IL-13-like gene, IL-4/13B, was later discovered at a different locus, next to KIF3A. More recently a single IL-4/13 gene was identified in the genome between KIF3A and RAD50 in the 2R bony fish, spotted gar Lepisosteus oculatus, and at least two IL-4/13 genes between KIF3A and RAD50 in the 2R cartilaginous fish, elephant shark Callorhinichus milii. Three IL-4/13 genes, IL-4/13A, IL-4/13B1, and IL-4/13B2, have been characterized in salmonids, with the last two having arisen from the salmonid 4R WGD.

It seems likely that a single IL-4/IL-13 gene existed in ancestral gnathostomes,
which has been duplicated in different lineages by WGD and/or tandem gene-duplication events. Functional studies of IL-4/13A have been performed in zebrafish and rainbow trout. Zebrafish injected with rIL-4/13A showed increased numbers of DC-SIGN+ (CD209) cells (a possible dendritic cell-marker) in PBL 5 days postinjection and increased numbers of IgZ-2+ B cells in PBL 2 days postinjection. In salmonids, the IL-4/13A was highly and broadly expressed constitutively, but the IL-4/13B paralogs are highly inducible by the T-cell mitogen PHA and the cytokines IL-2 and IL-21. Recombinant rainbow trout IL-4/13A and B isoforms showed overlapping but also distinct functions. They both increase IgM secretion by B cells, but cannot induce B-cell proliferation in vitro. They induce an early expression of acute-response proteins and antimicrobial peptide genes. They also modulate the expression of receptors for IL-4/13, IFN-γ, and IL-6 family cytokines, cellular markers for T cells, macrophages, dendritic cells, and pro- and antiinflammatory genes, suggesting a pleiotropic role of these molecules in fish immunity.

3.3 IL-7

Mammalian IL-7 is a nonhematopoietic cell-derived cytokine with a central role in the adaptive immune system. It promotes lymphocyte development in the thymus and maintains survival of naive and memory T-cell homeostasis in the periphery. The IL-7 gene is present in mammals and birds (chicken IL-7, ENSGALG000000022798). A fish IL-7 gene that shares limited homology to mammalian IL-7 has been described in fugu. The organization of the fugu IL-7 gene is the same as in the mouse, and consists of five exons and four introns, but differs from the human gene, where six exons and five introns are present. Comparison of the fugu and human genomes shows that some synteny exists around the IL-7 gene, with the presence of both the protein kinase inhibitor-a and chromosome 8 ORF 70 (C8orf70) genes, with IL-7 and C8orf70 having the same transcriptional orientation. IL-7 is also present in the elephant shark genome.

3.4 IL-9

IL-9 is a pleiotropic cytokine that has several effects on numerous hematopoietic cells. It stimulates the proliferation of activated T cells, promotes the proliferation and differentiation of mast cells, and increases production of IgE by B cells. It also promotes expression of mast cell proteases, upregulates the high-affinity IgE receptor, and induces IL-6 production. One source of IL-9 production is Th9 cells that might contribute to both protective immunity and immunopathological disease. IL-9 has only been described in mammals, but sequences are present in the database for birds (eg, chicken IL-9, ENSGALG00000006329) and amphibians (eg, Xenopus tropicalis IL-9, DQ221744). No IL-9 has been identified in the genomes of teleost and cartilaginous fish.
3.5 IL-15 and IL-15L

IL-15 has been described in mammals, birds, teleosts, and cartilaginous fish. IL-15 is also present in the genome of amphibians (e.g., *Xenopus IL-15, XP_012812039*) and reptiles (e.g., lizard IL-15, ENSACAG00000028882). Two genes with homology to IL-15 have been discovered in fish. One of the genes shows a similar gene organization and gene synteny with mammalian and chicken IL-15, and has been also termed IL-15. The second gene has a four-exon structure and is in a different genome location, and has been termed IL-15-like (IL-15L). IL-15L genes have also been found recently in cattle, horses, sheep, pigs, and rabbits, but is a pseudogene in humans and mice. IL-15L shares only ~21% aa identity with IL-15, however, IL-15L possesses many of the IL-15 residues important for binding to IL-15Rα; moreover, bovine rIL-15L can interact with IL-15Rα. IL-2 and IL-15 are closely related γC cytokines that share an additional receptor chain IL-2Rβ with their private, but related, IL-2Rα and IL-15Rα chains. It seems that three similar cytokines, IL-2, IL-15, and IL-15L were present in early vertebrates, and IL-15L had been lost in several, but not all, groups of mammals.

Mammalian IL-15 is a proinflammatory cytokine involved in the development, survival, proliferation, and activation of multiple lymphocyte lineages. It is not expressed in T cells, but is abundantly produced by a large variety of tissues and cell types: (1) tissues include the placenta, skeletal muscle, kidney, lung, and heart tissue; and (2) cell types include epithelial cells, fibroblasts, keratinocytes, nerve cells, monocytes, macrophages, and dendritic cells. In agreement with this, there is no increase in IL-15 expression in HK leucocytes after treatment with the T-cell mitogen PHA, or with PMA, CI, or a combination of the latter two, whereas the expression of other γC cytokines, including IL-2 and IL-21, is highly induced. The expression of the trout IL-15 gene is induced by rIFN-γ and polyI:C, but is refractory to LPS in cell lines (RTS-11 and RTG-2). IL-15 signaling is well-known to evoke a Th1-type immune response by inducing release of IFN-γ and TNF-α. Trout rIL-15 can also upregulate IFN-γ expression in splenic leucocytes. To date there are no functional reports on IL-15L.

3.6 IL-21

IL-21 genes have been described in mammals, birds, amphibians, and recently in teleost and cartilaginous fish. Although the gene synteny is clearly conserved at this locus in mammals, birds, and teleost fish, the IL-21 gene organization shows some interesting features. In mammals and birds the gene has 5 exons/4 introns, as does the fugu gene; however, the other fish genes show a 6 exon organization. It appears that all of the fish genes have an extra intron that splits exon 3 of mammals.

Mammalian IL-21 is predominantly produced by T follicular helper (Tfh) cells, Th17 cells, and NKT cells. Chicken IL-21 is expressed in most lymphoid
tissues, predominantly by CD4<sup>+</sup> TCR<sub>αβ</sub> T cells. The IL-21 gene in fugu and tetraodon shows low constitutive expression but can be induced by stimulation in vivo with LPS or polyI:C. The expression of trout IL-21 is relatively high in gills, intestine, skin, HK, and spleen, and is induced in vivo by bacterial and viral infection. In HK cells, trout IL-21 expression is rapidly induced by the T-cell mitogen PHA, suggesting that trout IL-21 may be produced in activated T cells, as seen in mammals.

Mammalian IL-21 has potent effects on all lymphocytes. IL-21 can contribute to the generation of murine Th2, Th17, and Thf cells, whereas it impedes development of Th1 and Treg cells. It synergizes with other γC cytokines (eg, IL-15 and IL-7) to induce proliferation and expression of effector molecules (eg, IFN-γ, granzyme, and perforin) in CD8<sup>+</sup> T cells, and the production of IFN-γ and IL-10 in NK and NKT cells. IL-21 also induces B cells to differentiate into memory and plasma cells, hallmarks of long-lived humoral immunity. Chicken IL-21 synergistically enhances T cell proliferation and inhibits maturation of dendritic cells as in mammals. Fish IL-21 bioactivity has been reported in rainbow trout. Trout rIL-21 rapidly induces the expression of IFN-γ, IL-10, IL-17A/F1 and 3, IL-22 and IL-4/13B paralogs in HK cells, and marker genes of Th1, Th2, and Th17 cells. It also increases the expression of the Th cell markers CD4, T-bet, and GATA3, and maintains the expression of CD8α, CD8β, and IgM at a late stage of stimulation. The activation of JAK/STAT3, Akt1/2, and PI3K pathways were shown to be responsible for trout rIL-21 action. Intraperitoneal injection of trout rIL-21 increased the expression of IFN-γ, IL-10, IL-21, IL-22, CD8, and IgM. These results suggest that IL-21 is likely a key regulator of T and B cell function in fish.

4 IL-6/IL-12 SUPERFAMILY

The IL-6 family in mammals includes IL-6, IL-11, leukemia inhibitory factor (LIF), oncostatin M (OSM), ciliary neurotrophic factor (CNTF), cardiotrophin-1 (CT1), cardiotrophin-2 (CT2; also known as neuropoietin, NP) and cardiotrophin-like cytokine (CLC). These cytokines signal through receptor complexes consisting of the common signal-transducing receptor protein, gp130, which is combined with another signal-transducing β receptor (LIFRβ or OSMRβ), or with a nonsignaling α receptor (IL-6Rα, IL-11Rα, or CNTFRα). A recent addition to the IL-6 family is IL-31 that signals through a unique receptor complex, composed of the IL-31Rα and OSMR without gp130. Human IL-31Rα and gp130 share 28% aa identity and are located head-to-head on chromosome 5q11.2, suggesting that these two genes are likely the product of a gene duplication, with IL-31Rα replacing gp130 in the IL-31 receptor complex. IL-31 plays an important role in the functioning of skin and of airway and intestinal epithelia, but to date it has been reported only in mammals.

The mammalian IL-12 cytokine family consists of four heterodimeric cytokines, IL-12, IL-23, IL-27, and IL-35, that signal through unique pairings of five
cytokine receptor chains: IL-12Rβ1, IL-12Rβ2, IL-27R (WSX-1), IL-23R, and gp130. Each cytokine consists of an α-chain (p19, p28, and p35) and a β-chain (p40 and Ebi3). The α subunits are structurally homologous to IL-6 family cytokines that are characterized by a unique up-up-down-down four-helix bundle conformation. The β subunits are homologous to the extracellular domains of the α receptors for IL-6 cytokines (eg, IL-6Rα and CNTFR), but lack a transmembrane domain, and are therefore secreted as soluble α/β heterodimers. Thus, the IL-6 and IL-12 families are structurally related, forming the IL-6/IL-12 superfamily.179,180

4.1 IL-6 Family

4.1.1 IL-6

Mammalian IL-6 is a pleiotropic cytokine produced by various cells to regulate hematopoiesis, inflammation, immune responses, and bone homeostasis.181 Mammalian IL-6 can signal by binding to the membrane-bound IL-6Rα that subsequently associates with membrane-bound gp130, and initiates intracellular signaling (classic signaling). IL-6 can also bind to the soluble (s)IL-6Rα that can be generated by proteolytic cleavage or alternative splicing. The complex of IL-6 and the sIL-6Rα, which is structurally similar to IL-12 cytokines, binds to membrane-bound gp130, which initiates signaling (trans-signaling).182 IL-6 genes are present in mammals, birds, reptiles, amphibians, teleost, and cartilaginous fish (Fig. 5.2).

In the mammalian paradigm, the classic signaling of IL-6 is important for the protection against bacterial infection, and the proinflammatory activities of IL-6 are mainly mediated by the trans-signaling mechanism. Chicken rIL-6 induces proliferation of the IL-6-dependent murine hybridoma cell-line 7TD1 and increases serum corticosterone levels in vivo.187 Trout rIL-6 induced the expression of the antimicrobial peptide genes cathelicidin-2 and hepcidin in macrophages.181 Fish rIL-6 also promotes antibody production, as seen in mammals.188,189

4.1.2 IL-11

Mammalian IL-11 exerts pleiotropic activities by stimulating hemopoiesis and thrombopoiesis, regulating macrophage differentiation, and conferring mucosal protection in the intestine.186 In addition to mammals, IL-11 genes are present in reptiles and amphibians, but they are missing in the current chicken genome. However, an IL-11Rα has been reported.190

The first nonmammalian IL-11 gene was discovered in fish by EST-type analysis in rainbow trout.191 Further studies in other species, for which genomes are available, have now discovered that in fact two IL-11 genes exist in 3R fish, termed IL-11a and IL-11b, due to the fish-wide 3R WGD192,193; (Fig. 5.2). Trout IL-11a is highly upregulated in spleen, HK, and liver by bacterial infection in vivo and by LPS, polyI:C, and rIL-1β in vitro in the macrophage-like RTS-11 cell line.191 No bioactivity analysis has been reported for fish IL-11 to date.
FIGURE 5.2  Phylogenetic tree analysis of the alpha chains of the IL-6/IL-12 superfamily. Sequences were chosen from representative species of mammals, birds, reptiles, amphibians, and fish. The phylogenetic tree was constructed using amino-acid multiple alignments and the neighbor-joining method within the MEGA6 program. Node values represent percent bootstrap confidence derived from 10,000 replications. The accession number for each sequence is given after the common species/molecule names. The Atlantic salmon molecules are from Wang et al., Husain et al., and Jiang et al. The bootstrap values of highly supported molecular types are indicated by red circles, and that of the paralogs due to the 3R (fish) WGD are indicated by blue circles.
4.1.3 *LIF*, *OSM*, and *M17*

In mammals, *LIF* and *OSM* are closest in the IL-6 family, located on the same genomic locus in the same direction, and can signal via a receptor complex of LIFR and gp130 that is used by three other cytokines (CNTF, CT1, and CLC). OSM can also signal via gp130 and OSMR, where the OSMR sits tail-to-tail with LIFR on the same chromosome in mammals. Both *LIF* and *OSM* have a very broad range of overlapping activities on almost all organ systems, including the hemopoietic, bone remodeling, and nervous systems. However, *LIF* knockout mice have a rather restricted set of developmental defects, including loss of female fertility and defects in some neurons and glial populations, whereas *OSM* knockout mice have reduced thymocytes and neurons, and abnormal thymus structure (Richards, 2015).

*LIF* genes are present in birds, reptiles, and amphibians (Fig. 5.2) in the syntenic genomic loci of mammals that contain *LIF* and *OSM*, with *OSM* missing in these lineages. A *LIF/OSM*-related gene, named *M17*, has also been reported in a few teleosts, which is also present in a syntenic region and has the same gene organization as mammalian *LIF* and *OSM*. A *LIF* gene has also been identified in the shark genome. It is likely that an ancestral *LIF/M17* gene had been present in early vertebrates and then was expanded to *LIF* and *OSM* in mammals by local gene duplication.

Trout *M17* is highly expressed in immune tissues, including gills, spleen, and HK, as well as in the brain. Furthermore, trout *M17* expression is upregulated by stimulation of RTS-11 cells by LPS, polyI:C, and rIL-1β, and by bacterial infection in vivo. Zebrasfish *M17* (LIF) is upregulated after nerve injury, and the *M17* knockdown delayed functional recovery after optic nerve injury in vivo. Goldfish r*M17* induces the production of nitric oxide in goldfish macrophages, and stimulates the proliferation of macrophage progenitor cells when added in combination with cell-conditioned medium. These results suggest roles of *M17* in the fish immune and nervous systems.

4.1.4 *CNTF* and *CNTF*-Like

*CNTF* is a pluripotent neurotrophic factor originally isolated from chick-embryonic ciliary neurons. It has potent effects on the development and maintenance of the nervous system, as well as on cardiomyocytes, osteoblasts, immune cells, adipocytes, and skeletal muscle cells. The *CNTF* genes are present in tetrapods, including coelacanth, amphibians, birds, reptiles, and mammals (Fig. 5.2). Two *CNTF*-like genes (CNTFa and CNTFb) are present in several teleost fish genomes, likely due to the 3R WGD. The fish *CNTF* share limited aα identities (21–24%) with tetrapod *CNTF* molecules, and have a three-exon/two-intron structure, whereas the tetrapod *CNTF* gene has a two-exon gene organization. Furthermore, the genomic loci of fish *CNTF*-like genes lack conserved synteny to tetrapod *CNTF* loci. These data suggest that the fish *CNTF*-like genes are not orthologs of tetrapod *CNTF*.

The trout *CNTFa* gene is highly expressed in brain and muscle, and does not respond positively to immune stimulation and infection. The medaka
The CNTFβ gene is located in the Y-chromosome-specific region containing the sex-determining gene, and its expression is male-biased prior to the onset of sexual maturity. The comparative analysis of CNTF paralogs has yet to be performed.

### 4.1.5 CLC, CT1, and CT2

The cardiotrophin-like cytokine (CLC) (also known as neurotrophin-1/B cell-stimulating factor-3) has been demonstrated to have potent neurotrophic effects and B-cell stimulatory properties, and to act as a modulator of pituitary corticotroph POMC gene expression and ACTH secretion. CLC contains a putative signal peptide and so should enter the classical secretory pathway; however, it is retained within the cell by a mechanism that is unknown as of yet. For cellular secretion of CLC, heteromeric complex formation with other factors (eg, cytokine receptor-like factor-1 or soluble CNTFRα) seems to be required. In addition to mammals, CLC genes are predicted in birds, reptiles, amphibians, and teleost fish (Fig. 5.2). No functional analysis has been reported in lower vertebrates.

Mammalian CT1 and CT2 are closely located in the genome, with a head-to-head transcriptional orientation, and have a conserved three-exon gene structure, suggesting that they arose from a gene duplication event. Both genes are active in mammals, with the exception of humans, where the CT2 gene has been pseudogenized. A CT2-like sequence is present in amphibians and reptiles, but is apparently missing in birds and fish (Fig. 5.2).

### 4.2 The IL-12 Family

#### 4.2.1 IL-12

IL-12, the prototypical heterodimeric cytokine, was purified as an NK cell stimulatory factor (NKSF). When added to human PBL, NKSF induces IFN-γ production, augments NK cell-mediated cytotoxicity, and enhances the mitogenic response of T cells. NKSF activity was associated with a 70-kDa anionic glycoprotein that could be reduced to two peptides with a molecular mass of 40 (p40) and 35 (p35) kDa, suggesting that this cytokine is a heterodimer. IL-12 cDNA cloning revealed that the p35 and p40 subunits in human and mouse are encoded by two independent genes located on different chromosomes.

#### 4.2.1.1 p35, p35a, and p35b

A single p35 gene has been reported in mammals and birds. Fish p35 was discovered, along with p40, by analysis of the fugu genome. The cloning of p35 in other fish species revealed that two p35, p35a, and p35b are present in teleost fish, due to the 3R WGD. Furthermore, the salmonids possess four p35 loci, with three active p35 genes—p35a1, p35a2, and p35b—due to the 4R WGD in this lineage. Lineage-specific gene organizations of p35 genes are apparent. Whereas the exons of all p35 genes are separated by a phase I intron in the signal peptides and phase 0 introns in the mature peptide regions, exon duplication and intron insertion events appear to have happened in different lineages. Human
and mouse p35 genes have 7 exons, with exons 4 and 5 the same size. However, only one equivalent exon is present in birds and teleost fish, suggesting a mammalian-specific exon duplication. Fish p35b genes have a 7 exon/6 intron structure, with the sum of exons 5 and 6 equivalent to the second-to-last exon in salmonid p35a and tetrapod p35 genes, suggesting an intron-insertion event early after the 3R WGD. P35 is also found in the elephant shark genome.

4.2.1.2 p40, p40a, p40b, and p40c

A single p40 gene has been reported in tetrapods, including mammals, birds, reptiles, amphibians, and coelacanth. The fish p40 chain was first found in the fugu genome along with p35. Remarkably, multiple isoforms of this gene are present in fish. Thus, in carp, zebrafish, and pufferfish three isoforms are known, with p40a and p40b more similar, and p40c more distantly related, in phylogenetic tree and homology analysis. For example, in carp, p40a and p40b have 32% aa identity, but 24–26% to p40c. The salmonids also have three p40 paralogs (p40b1, p40b2, and p40c), with p40a missing, and p40b1 and p40b2 derived from the 4R WGD. All the p40 genes have the 8 exon/7 intron structure, except for salmonid p40b genes, which have an extra intron in the 3′-UTR, and tilapia p40b that has an intron insertion in exon 3. One difference of note, however, is that the last intron of fish p40c genes is phase 0, in contrast to phase II in fish p40a and p40b, and mouse p40. Partial sequences of two p40-like genes have also been predicted in the elephant shark genome.

Teleost p40c molecules are divergent from p40a, p40b, and mammalian p40 in terms of identities/similarities, disulphide bond potential, and gene organization. They branch from the root of the p40 group in phylogenetic tree analysis, where teleost p40a and p40b group first to the tetrapod p40 clade. The loci of teleost p40a, p40b, and mammalian p40 show considerable gene synteny that is largely missing in the teleost p40c loci. Such evidence suggests that teleost p40a and p40b molecules are indeed coorthologs of tetrapod p40, likely due to the 3R WGD, whereas teleost p40c perhaps emerged in a different way.

In general, fish p40 isoforms are highly expressed in most tissues compared to that of p35. The lower level of constitutive expression of p35 paralogs relative to p40 paralogs in most tissues in many different fish species suggests that the fish p35 subunit may be a rate-limiting factor for heterodimeric IL-12 production in fish. Although differentially expressed in most tissues, the comparable and high-level expression of p35a1, p35a2, and p40c in salmonid blood suggests that the IL-12 isoforms, p35a1/p40c and p35a2/p40c, may be constitutively produced and have a homeostatic role. In salmonids, the p40c expression is dominant in the major immune organs (ie, spleen, HK, thymus, gills, and blood), whereas the p40b1 and/or p40b2 are highly expressed in integumentary tissues (eg, intestine, gills, tail fins, and scales). This could indicate a differential tissue-specific distribution of
different IL-12 isoforms that could have some bearing on their differential roles in fish immunity.\textsuperscript{180,206}

### 4.2.1.3 Fish IL-12 Isoforms

It seems that IL-12 has been present from early vertebrates (cartilaginous and teleost fish) to mammals. The presence of multiple divergent p40 and p35 subunits means multiple IL-12 isoforms may exist in teleost fish that could have different functions. Thus, up to nine IL-12 isoforms could be produced in salmonids, and up to six IL-12 isoforms in other 3R teleosts (Table 5.1). The evaluation of the bioactivities of each of these putative fish IL-12 isoforms will be a challenge because of the dimeric nature of IL-12 and the potential presence of other isoforms. Two isoforms of trout rIL-12 containing the same p35a1 and different p40 (p40b1 or p40c) has been produced in CHO cells as single chains. Both isoforms induce IFN-\(\gamma\) expression, but only one isoform upregulates IL-10 expression, suggesting functional differentiation of different isoforms.\textsuperscript{180} A single-chain rIL-12 has also been described in Grouper (\textit{Epinephelus coioides}) that had a modest impact on gene expression (eg, less than 2.5-fold for TNF-\(\alpha\)) and stimulated PBL proliferation at high concentrations (>125 pM, equivalent to 8 \(\mu\)g/ml).\textsuperscript{209} The bioactivity of p40c has also been investigated in rock bream (\textit{Oplegnathus fasciatus}) and has been shown to stimulate respiratory burst activity of PBL at high concentrations (>10 \(\mu\)M, equivalent to 400 \(\mu\)g/ml).\textsuperscript{210} The biological relevance of these studies remains to be clarified, as the concentrations used are unlikely to be within the physiological range.

### 4.2.2 IL-23

IL-23 is comprised of the unique p19 \(\alpha\)-chain linked by disulphide bonding to the p40 \(\beta\)-chain that is shared by IL-12. Human and mouse p19 genes were
discovered by a computational approach, and have no biological activity unless combined with p40 to form IL-23. IL-23 activates STAT4 in PHA-activated T-cell blasts and induces strong proliferation of mouse memory (CD4 + CD45Rb<sup>low</sup>) T cells, a unique activity of IL-23 versus IL-12, which has no effect on this cell population. IL-23 is also a central cytokine controlling Th17 development, but can contribute to autoimmune diseases.

Using a synteny approach, fish p19 genes were first cloned in zebrafish, and later in salmonids. Only one p19 gene is present in zebrafish, but two (p19a and p19b) in salmonids, with trout p19b pseudogenized. The salmonid p19 translations share moderate identities (23–30%) to zebrafish and mammalian p19 molecules, but their identity is supported by structural features, a conserved 4 exon/3 intron gene organization, and phylogenetic tree analysis. Besides mammals and teleost fish, no p19 genes have been reported in other vertebrates to date.

The salmonid p19 genes are highly expressed in blood and gonad. Trout p19 gene expression is induced by bacterial and viral infection in vivo, by PAMPs (polyI:C and peptidoglycan), and by the proinflammatory cytokine IL-1β in primary HK macrophages. Zebrafish p19 expression is also induced in vivo after infection with Mycobacterium marinum and in vitro by LPS. These data may indicate functional roles of IL-23 in regulating the immune response in fish.

Since IL-23 is a heterodimer of a p40 subunit and a p19 subunit, and three p40 paralogs are present in fish, it is possible that teleost fish have up to three IL-23 isoforms in 3R teleost fish and up to six IL-23 isoforms in salmonids (Table 5.1). Since the trout IL-12 isoforms containing a single p35 paired with different p40 can have distinct bioactivity, it is also possible that multiple IL-23 isoforms may be present in teleosts, with distinct functions to fine-tune a unique fish-Th17-like response with multiple IL-17A/F isoforms (see Section 6). The bioactivity of any fish IL-23 remains to be determined.

4.2.3 IL-27
IL-27 consists of the two subunits p28/IL-30 and EBI3 that signal via a receptor complex of gp130 and IL-27R (also known as WSX-1/TCCR). Being a heterodimeric cytokine that signals through gp130, IL-27 is either grouped into the IL-6 or the IL-12 family of cytokines. IL-27 is a multifaceted heterodimeric cytokine with pronounced pro- and antiinflammatory, as well as immunoregulatory, functions.

The EBI3 gene was first identified in B lymphocytes as a gene induced by EBV infection. It encodes a 34 kDa soluble type-1 cytokine receptor that is homologous to the p40 subunit of IL-12. EBI3 is able to associate with the p28 subunit to form IL-27, or with the p35 subunit to form IL-35. EBI3 genes are present in tetrapods, including mammals, birds, reptiles, amphibians, and coelacanth, as well as in cartilaginous and teleost fish.

Mammalian EBI3 is expressed in tonsil, spleen, and placental trophoblast, and is upregulated by activation of myeloid cells and regulatory T and B
cells. In half-smooth tongue sole (Cynoglossus semilaevis), EBI3 expression was high in immune organs (eg, blood, spleen, and HK) but less abundant in liver (Li et al., 2013). However, the constitutive expression of EBI3 in Atlantic salmon (and in rainbow trout, unpublished data) is low in immune organs (HK, spleen, blood, gills, and thymus), but is two orders higher in liver and one order higher in caudal kidney, compared to spleen. The different expression patterns may suggest different roles of EBI3 in different species/lineages. Tongue sole rEBI3, at a nonphysiological concentration (10 μg/ml), induced the expression of proinflammatory cytokines (IL-1 and IL-8) and chemokines in PBL.

Human and mouse p28 genes were also discovered by a computational approach. Outside of mammals, p28 has only been characterized in fish. Two p28 paralogs (p28a and p28b) sharing 72% aa identity have been cloned in Atlantic salmon, and a single p28 gene has been found in other teleost fish. A potential p28 gene also appears to exist in cartilaginous fish. Mammalian p28 is expressed in cells of the myeloid lineage (mainly monocytes and activated dendritic cells) and plasmacytoid dendritic cells. Both salmon p28 genes are highly expressed in immune relevant tissues, such as thymus, gills, spleen, and HK. This is in contrast to salmon EBI3 that is highly expressed in liver and caudal kidney. The expression of p28 is induced in HK cells by PAMPs and recombinant cytokines in vitro, and in spleen after injection of polyI:C in vivo.

One IL-27 (p28/EBI3) isoform is present in mammals and potentially in 3R teleost fish. However, two IL-27 isoforms may exist in the 4R salmonids (Table 5.1). No bioactivity analysis has been reported for fish IL-27 to date.

4.2.4 **IL-35**

The broad presence of p35 and EBI3 from cartilaginous and teleost fish to coelacanth, amphibians, birds, reptiles, and mammals, described previously, suggests that IL-35 is likely present in all vertebrates. Due to the extra WGD, two IL-35 isoforms may exist in 3R teleost fish and three in 4R salmonids (Table 5.1).

The p35/EBI3 heterodimer was discovered in 1997 in mammals. However, its function and physiological relevance were unknown until 2007, when the name IL-35 and function were given to this novel IL-12 family member. IL-35 is produced mainly by regulatory T cells in contrast to other IL-12 family cytokines, which are mainly produced by antigen-presenting cells. It can also be produced by regulatory B cells and regulatory CD8+ T cells. IL-35 can signal via four receptor complexes: gp130/IL-12Rβ2, gp130/gp130, IL-12Rβ2/IL-12Rβ2, and IL-27Rα/IL-12Rβ2. IL-35 exhibits its suppressive activities in a range of autoimmune diseases and cancer models by inhibition of proliferation, and induction of an IL-35-producing induced regulatory T-cell population, referred to as iTr35. No IL-35 bioactivity has been reported outside of mammals to date.
5 IL-10 FAMILY

The IL-10 family in mammals consists of IL-10, IL-19, IL-20, IL-22, IL-24, IL-26, and the interferons,\(^{221}\) although IL-26 is absent in mice (see later in the chapter), showing that in particular species one or more of these genes may not be expressed. These genes are found at two loci: in humans on Chr 1q32 that contains IL-10, IL-19, IL-20, and IL-24, and on human Chr 12q15 that contains IL-22 and IL-26 (as well as IFN-\(\gamma\)). They have a common 5 exon/4 intron gene organization, suggesting that they have arisen by gene duplication from a common ancestor. The locus containing the IL-22 and IL-26 genes is well-conserved throughout gnathostome (jawed) vertebrate evolution, and these two genes have been reported in birds (eg, chicken\(^{222,223}\); turkey\(^{224}\)), amphibians,\(^{225}\) and bony fish.\(^{226,227}\) However, in cartilaginous fish only IL-22 has been found at this locus.\(^{160}\) Curiously, in mice it appears that the IL-22 and IL-26 genes have been duplicated locally, and whereas one of the IL-22 genes is still functional (IL-22\(^{a}\)), the second IL-22 gene (IL-22\(^{b}\)), and both IL-26 genes, are pseudogenes. However, it is not certain whether a full-length IL-26 gene paralog might exist in the genome of “wild” mice. Whereas the function of IL-22 has been investigated in some detail in different species, IL-26 is understudied.

5.1 IL-22

In fish, relatively high levels of IL-22 expression have been reported in mucosal tissues such as gills\(^{226,228}\) and intestine.\(^{226,229}\) Following pathogen challenge of immunized fish, a significantly increased expression is seen relative to control (unvaccinated) fish.\(^{228}\) Increased expression of IL-22 is also seen in immune tissues, including gills of mullet, following challenge with \textit{Streptococcus dysgalactiae},\(^{230}\) and in liver, spleen, and HK of turbot challenged with \textit{A. salmonicida}.\(^{229}\) Expression of IL-22 in whole larvae is also enhanced following exposure to \textit{Yersinia ruckeri}, where it is thought to play a role in disease protection during early developmental stages.\(^{231}\) In turbot, IL-22 expression is also highly induced by stimulation with PMA, as seen with HK, spleen, or intestinal cells.\(^{229}\) In \textit{rag1}\(^{-}\) \textit{zebrafish}, IL-22 expression is \(\sim\) two-fold higher than in wild-type fish, with high expression in the HK, especially in macrophages.\(^{232}\) In \textit{Xenopus}, no constitutive expression is found, but IL-22 can be induced in spleen and intestine after LPS stimulation.\(^{225}\) In chicken, constitutive expression was seen in thymus, spleen, and small intestine, but, curiously, LPS stimulation of lymphocytes decreased IL-22 expression.\(^{223}\)

Only a few studies have looked at the bioactivity of IL-22 in nonmammalian species. In rainbow trout, rIL-22 enhanced the expression of antimicrobial peptides such as LEAP2, hepcidin, and \(\beta\)-defensins in splenocytes, showing that it promotes host innate immunity against microbes, as seen in mammals.\(^{233}\) In turbot, rIL-22 induced the expression of hepcidin, IL-1\(\beta\), IL-8, and TNF-\(\alpha\) in HK cells, in a JAK/STAT-dependent manner,\(^{232}\) and in zebrafish
induced myeloperoxidase expression in enterocytes, as visualized in mpx:GFP/lysC:DsRED2 transgenic fish. Fish injected with rIL-22 and infected with the bacterium A. salmonicida were found to have higher survival rates relative to control fish. The importance of IL-22 for disease resistance was also shown in zebrafish embryos, where knockdown of IL-22 with morpholinos decreased survival to infection with A. hydrophila, and even control fish given the morpholinos (ie, without infection) had a decreased survival. In So-iny mullet, in vivo injection of rIL-22 increased the expression of β-defensin in liver, spleen, kidney, and gut tissue, and survival following infection with Streptococcus dysgalactiae. Lastly, in chickens, rIL-22 had little effect on immune cells but increased expression of IL-8 and IL-10 in epithelial cells (chicken embryo kidney-cells), and at low doses (0.01 μg/ml) enhanced the stimulatory effect of LPS on expression of IL-1β, IL-6, IL-8, SAA and the antimicrobial peptide gallinacin-2 in these cells. In addition, chicken rIL-22 enhanced IL-1β, IL-8, SAA, and LEAP2 expression in hepatocytes, which was further enhanced in the presence of LPS. rIL-22 can also induce IL-10 expression in chicken kidney cells.

Modeling of the So-iny mullet IL-22 suggests that it has six alpha-helices, as seen in the human monomer, although helix A, C, and E are shorter, whereas the AB and DE loops are longer. However, the cysteine residues do not form clear disulphide bonds. Recently, the crystal structure of zebrafish IL-22 has been determined. The zebrafish molecule was most similar to the human IL-22 structure, with some differences located in the N-terminus, followed by human IL-20, then IL-19.

5.2 IL-26

Typically IL-26 shows either low or no constitutive expression in tissues. IL-26 was induced in intestine of fugu following polyI:C stimulation and in spleen of Xenopus following polyI:C stimulation, suggesting a possible role in antiviral defense.

5.3 IL-10

The IL-10 locus is also well-conserved, with IL-10 readily identifiable in all gnathostome vertebrates. However, the number of related genes and their homology to mammalian IL-10 family members varies. In the elephant shark, two additional IL-10 family genes are present next to IL-10, but show no clear homology to IL-19/20/24. In bony fish, one additional gene is present that branches in phylogenetic tree analysis to the base of the IL-19/IL-20/IL-24 clades, suggesting that it is related to an ancestral gene that has subsequently duplicated in evolution. Similarly, in Xenopus and chicken, a single additional IL-10 family gene is present next to IL-10 at this locus, referred to as IL-20 and IL-19, respectively.

In bony fish, IL-10 was first discovered in the fugu genome and has since been found in a wide range of fish species. In trout two genes are present
that differ in expression level and inducibility. Constitutive expression is typically seen in immune tissues, as in carp/grass carp, where it is detectable in HK, gills, intestine, and spleen. Expression can be upregulated in vitro by exposure to a variety of stimulants, such as LPS, PHA, or IL-21, and in vivo by injection with (live or dead) bacteria. Relatively few studies of IL-10 bioactivity have been performed in fish to date. In goldfish, rIL-10 pretreatment of monocytes can reduce the production of reactive oxygen species (ROS) in response to A. salmonicida or IFN-γ stimulation, and this is associated with phosphorylation and nuclear translocation of STAT3, and an increase of SOCS3 expression. In carp, rIL-10 downregulated the production of ROS and nitrogen radicals (NOS) from neutrophils and macrophages, with the latter being more sensitive, and the expression of proinflammatory molecules such as IL-6, p35, and MHC class I/II in response to LPS stimulation, again associated with STAT3 phosphorylation and induced SOCS3 expression. Effects on carp T and B cell responses were also apparent. rIL-10 increased the proliferation of a subset of memory T cells (primed by Trypanosoma borreli infection in vivo) in response to antigen exposure (T. borreli lysate) in vitro, and increased the proliferation (whether from immunized or naive fish) and antibody secretion (total and antigen-specific) of immunoglobulin (Ig)M+ B cells.

In Xenopus, IL-10 has highest constitutive expression in spleen, kidney, and intestine, and was induced by LPS (ip) injection, whereas in chicken, IL-10 is constitutively expressed in thymus, bursa of Fabricius, cecal tonsils, liver, and lung, and is induced in thymocytes by PHA stimulation and in macrophages by LPS stimulation. Similar expression profiles are seen with turkey and duck IL-10, although in the latter case two splice variants are present: one lacking exon 5 and one with a spliced 3′UTR (with the last 400 nt from a downstream exon 6). IL-10 has also been shown to be induced by infection in birds, as seen with Eimeria-infected chickens, where higher levels of IL-10 expression appear to correlate with disease susceptibility. Interestingly, intercrossing of different chicken lines has revealed an association between single nucleotide polymorphism (SNP) in the IL-10 (and MAPKAPK2) gene with Salmonella enteritidis burden, with potential for use in marker-assisted selection. rIL-10 bioactivity has also been studied in these three species. Chicken and turkey rIL-10 has been shown to inhibit mitogen-induced IFN-γ synthesis in splenocytes, whereas duck rIL-10 can inhibit mitogen-induced IL-2 expression in peripheral blood mononuclear-cells.

In relation to the second gene found at the IL-10 locus in euteleostomi (bony fish and tetrapods), in trout the IL-20L gene is highly expressed following stimulation of RTS-11 cells (a trout macrophage cell line) with LPS, and following infection with the bacterial pathogen Y. ruckeri. No studies on the expression of the IL-20 gene in Xenopus have yet been reported, but in chicken the IL-19 gene is highly expressed in spleen, heart, lung, and skin, and can be induced in the ileum by infection with Ascaridia galli and by LPS stimulation of
monocytes.\textsuperscript{255} rIL-19 was able to induce the expression of IL-1β, IL-6, IL-8, and iNOS/nitric oxide, and to increase apoptosis in monocytes, whereas in splenocytes it increased IL-4, IL-13, and IFN-γ expression, and this effect was enhanced by preincubation with Con A.\textsuperscript{255}

### 5.4 vIL-10

It should be noted that IL-10-like genes have been discovered in some viruses. Whereas they are well-known in mammalian viruses such as Herpesviridae and Poxviridae, more recently they have also been found in fish viruses (Alloherpesviridae).\textsuperscript{256,257} The biological properties have yet to be studied in most cases, but in mammals they have broad immunosuppressive properties, despite the sequence identity usually being quite low relative to the host molecule. For example, the IL-10 molecule in eel (Anguilla anguilla) and the Anguillid herpesvirus-1 ORF25 have 33.6\% identity, whereas the IL-10 molecule in carp with Cyprinid herpesvirus-3 (CyHV-3) ORF 134 have 26.9\% identity. Injection of zebrafish with the CyHV-3 IL-10 mRNA increased the number of lysozyme-positive cells, as seen with zebrafish IL-10, and this effect could be inhibited by injection of a morpholino to one of the IL-10 receptor chains (IL-10R1). Although these transcripts are highly expressed during infection, in carp they do not appear to be essential for viral replication or virulence.\textsuperscript{258}

### 6 IL-17 FAMILY

The first member of the IL-17 family of cytokines was discovered in rodents as CTLA-8 in 1993.\textsuperscript{259} Since then it has become apparent that six isoforms are present in mammals, designated IL-17A-F. They share 20–50\% identity, and, with the exception of IL-17A and IL-17F that are in tandem on Chr 6 (Fig. 5.3), they are found at different loci (eg, IL-17B on Chr 5, IL-17C on Chr 16, IL-17D on Chr 13, and IL-17E on Chr 14). Studies into the evolution of this cytokine family have unexpectedly found that they are an ancient lineage with homologs within the invertebrates. Multiple genes can be present within non-vertebrate species, as seen in Pacific oyster,\textsuperscript{260} and they form a clade separate from the vertebrate molecules in phylogenetic tree analysis. Within vertebrates the most ancient isoform appears to be IL-17D, where agnathan species such as the Japanese lamprey \textit{Lethenteron japonicum} possess a single isoform that clusters with IL-17D from other vertebrates.\textsuperscript{46} It is constitutively expressed by most tissues, and is upregulated by LPS in cultured skin-cells. Interestingly, IL-17 has also been found to be produced by variable lymphocyte receptor (VLR) A\textsuperscript{+} lymphocytes in sea lamprey \textit{Petromyzon marinus}, in response to the T-cell mitogen, PHA, in an analogous way to the response in T cells in gnathostome vertebrates.\textsuperscript{261}

Within gnathostomes it seems likely that four genes were present in early species, equivalent to IL-17A/F, IL-17B, IL-17C, and IL-17D. In the case of
FIGURE 5.3  Synteny analysis of the IL-17A/F loci from 2R tetrapods, 2R and 3R teleost species. Schematic diagram shows the syntenically conserved genes in the IL-17A/F loci in 2R tetrapods (human chromosome Chr6, Chicken Chr4, Chinese turtle Chr JH212851, frog Chr GL172699); 2R fish (coelacanth Chr JH128580 and spotted gar Chr LG1), and 3R fish (tilapia Chr GL831133 and GL831143, and zebrafish Chr 11 and Chr20). An arrow indicates transcriptional direction. The synteny analysis was carried out using the Genomicus browser (http://www.genomicus.biologie.ens.fr/genomicus-74.01).
IL-17A/F, this molecule appears to have subsequently expanded in different lineages. Thus, the IL-17A/F genes have expanded to IL-17A/F1, -2, and -3 in teleost fish, or to IL-17A and IL-17F in reptiles, birds, and mammals. It should be noted that in the case of IL-17C, it is possible that this molecule subsequently diverged into IL-17C and IL-17E in early mammals, whereas in some teleost fish, two IL-17C genes are present. This explanation of IL-17 evolution has yet to be verified, and in the case of IL-17A/F evolution in particular, other scenarios are possible, such as gene conversion (Fig. 5.4).

Thus, in cartilaginous fish, five IL-17-like genes were discovered in the elephant shark genome. Two have homology to IL-17A and IL-17F, and are potentially two IL-17A/F genes. Two genes related to IL-17B and IL-17D are present that are potentially the true homologs of these molecules, and, lastly, one gene that is described as equivalent to IL-17C is present. Similarly, extensive

FIGURE 5.4 Evolution of IL-17A/F genes in vertebrates. Two IL-17A/F genes emerged via local gene duplication in the ancestral vertebrates. These two genes (IL-17A/F1 and IL-17A/F2) were preserved during the evolution of the 2R teleost fish and amphibians, but expanded by genome duplication in 3R fish that generated IL-17A/F1, 2, 3, and N; and in 4R salmonids that produced IL-17A/F1a, 1b, 2a, 2b, 3, and N. One of the two ancestral IL-17A/F genes might have been deleted, but the remaining gene was duplicated locally during evolution that produced the IL-17A and IL-17F genes in reptiles, birds, and mammals. Alternatively, the two ancestral IL-17A/F genes might have been converted to IL-17A and IL-17F in reptiles, birds, and mammals.
analysis of teleost fish genomes, initially in zebrafish, has revealed the presence of three IL-17A/F genes at two different loci (with IL-17A/F1 and IL-17A/F2 in tandem on chr 17), IL-17C, and IL-17D. Subsequent analysis of the fugu genome revealed a second IL-17C locus, and the presence of what appeared to be a novel IL-17 family member that was termed IL-17N, giving seven IL-17 family members in this species. In salmonids, the situation appears to be even more complex in that the 4R WGD that occurred in this lineage has given rise to further paralogs. Thus, rainbow trout and salmon possess two IL-17A/F1, two IL-17A/F2, and one IL-17A/F3, in addition to two IL-17C, one IL-17D, and one IL-17N, giving nine family members. Most recently, it has become apparent that IL-17B is also present in at least some teleost fish groups, as seen in catfish and eels (Wang, unpublished). With the additional IL-17A/F sequence information from further teleost species, it is now proposed that IL-17N is in fact an IL-17A/F, and it is likely that the IL-17A/F locus containing IL-17A/F1 and IL-17A/F2 was duplicated in teleosts to generate IL-17A/F3 (with high homology to IL-17A/F/1), and IL-17N (with homology to IL-17A/F2).

Functional studies in teleost fish have shown complex patterns of expression that vary in a time-, tissue-, and paralog-dependent manner following infection or vaccination. Interestingly, in carp infected with Trypanosoma carassii, induced p19 and p40(c) expression is seen, the necessary components of IL-23, and is correlated with IL-17A/F/2 expression, suggestive of a Th17-like response in this species. In salmon, high levels of IL-17A/F2 (together with IL-17AR and TGF-β1) expression, but not IFN-γ or GATA3, correlate with the severity of side effects seen following injection with oil-adjuvanted vaccines, also suggesting that Th17-type responses are involved. Lastly, in zebrafish overexpression of FoxP3 (a transcription factor involved in Treg differentiation) in developing embryos led to downregulation of IL-17A/F levels, whereas injection with a morpholino to FoxP3 enhanced IL-17A/F expression, hinting that IL-17 responses are tightly regulated at this level of phylogeny, potentially to prevent autoimmunity, as seen in mammals.

Relatively few studies have reported the bioactivity of the IL-17 molecules in fish. In grass carp, rIL-17A/F1 can increase IL-1β, IL-6, IL-8, and TNF-α expression in HK leucocytes, and in trout, rIL-17A/F/2a increases the expression of IL-6, IL-8, and an antimicrobial peptide (β-defensin 3) in splenocytes. In the case of grass carp rIL-17D, it can also increase IL-1β, IL-8, and TNF-α expression in HK cells, but does not increase IL-6 expression, hinting at subtle differences in the ability of different isoforms to induce proinflammatory genes.

No functional studies have been reported on the amphibian IL-17 family members, with IL-17B, IL-17C, and IL-17D all clearly present in Xenopus, as well as genes likely equivalent to IL-17A/F1 and IL-17A/F2 (Fig. 5.3). In the case of Xenopus, in fact three IL-17A/F1 genes (a, b, c) can be found in tandem next to IL-17A/F2. However, in the amniotes, the IL-17A and IL-17F locus appears highly conserved, with both genes in the same orientation, suggesting that they are true orthologs of each other. Although no functional studies have been
reported in reptiles, in chickens IL-17A and IL-17F are highly expressed in the gut, and the liver in the case of IL-17F, and in the CU205 (reticuloendotheliosis virus-transformed) chicken lymphoblast cell-line.\textsuperscript{273,274} They can be induced in splenocytes by stimulation with ConA, LPS, and polyI:C, although with differences in kinetics and magnitude of induction. Similarly, duck IL-17A is induced by ConA stimulation of splenocytes.\textsuperscript{275} Both chicken rIL-17A and rIL-17F have been shown to induce the expression of the proinflammatory cytokines IL-1β, IL-6, and IL-8 in chicken embryonic fibroblasts.\textsuperscript{273,274} IL-17A has also been shown to play a role in the immunopathology caused by \textit{Eimeria tenella} (intracellular apicomplexan parasite) infection, with rIL-17 administration increasing fecal oocyst shedding and cecal lesion score, but decreasing body weight, whereas anti-IL-17 treatment gave the converse effects.\textsuperscript{276} However, comparative expression of IL-17A and IL-17F in intestinal tissue during \textit{E. tenella} (cecum) or \textit{E. maxima} (duodenum) infection has shown that IL-17F appears to be the main isoform upregulated.\textsuperscript{273} Similarly, during experimental infection of chickens with the intracellular protozoal parasite \textit{Cryptosporidium baileyi}, IL-17A expression was found increased in trachea and spleen, suggesting that it contributes to the inflammation seen during infection.\textsuperscript{277} Curiously, both chicken and duck IL-17A have homology to ORF13 of \textit{Herpesvirus saimiri} (HVS 13),\textsuperscript{275,274} and perhaps this hints at the potential for immune evasion strategies targeting this pathway during herpesvirus infection in birds (as seen for IL-10 in other species).

7 TNF FAMILY

In mammals, the tumor necrosis factor superfamily (TNFSF) currently has 19 members\textsuperscript{278} (Table 5.1), each of which had been identified under a different synonym until the current nomenclature assigned by the HUGO Gene Nomenclature Committee (HGNC).\textsuperscript{279} Each member is a type II transmembrane molecule that contains a TNF homology domain and is thought to be active primarily in trimeric form, either on the cell surface or soluble after extracellular cleavage.\textsuperscript{278} They are expressed predominately by activated immune cells, and elicit their biological effects by binding to type I transmembrane TNF superfamily receptors. They mediate a wide range of important functions, including embryogenesis, cell differentiation, homeostasis, immune system development, regulation of immune responses, inflammation, and anticancer defense.\textsuperscript{278} With the increasing numbers of genomes available and the addition of transcriptomic data from a wide variety of species, a number of TNFSF members outside of mammals have now been identified, and it is possible to study how this family has evolved within a wide range of living organisms (Table 5.2).

7.1 Protostomian Invertebrates

The TNFSF is one of the few families in which ortholog genes from this group can be found within protostomian invertebrates. The initial discovery was
## TABLE 5.2 Summary of the Currently Characterized TNFSF Members Within Vertebrates

<table>
<thead>
<tr>
<th>Other Synonyms</th>
<th>CD Number</th>
<th>Birds</th>
<th>Reptiles</th>
<th>Amphibians</th>
<th>Bony Fish</th>
<th>Cartilaginous Fish</th>
<th>Agnathans</th>
</tr>
</thead>
<tbody>
<tr>
<td>TNFSF1 LT-a, TNF-b, TNFSF1B</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNFSF2 TNF-a, DIF, Necrosin, Cytotoxin, TNFSF1A</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNFSF3 LT-b, TNF-c</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNFSF4 gp34, OX40L, CD134L, TXGP1</td>
<td>252</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNFSF5 CD40L, TRAP, Gp39, T-BAM</td>
<td>154</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNFSF6 FASL, APTL, APT1LG1, CD95L</td>
<td>178</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>TNFSF7 CD27L</td>
<td>70</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNFSF8 CD30L</td>
<td>153</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNFSF9 CD137L, 4-1BBL</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNFSF10 TRAIL, APO-2L</td>
<td>253</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNFSF11 OPGL, RANKL, TRANCE, ODF, SOFA</td>
<td>254</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNFSF12 TWEAK, APO-3L, DR3L</td>
<td>253</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNFSF13 APRIL, TALL-2,TRDL-1</td>
<td>256</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNFSF13B BlyS, BAFF, TALL-1, THANK, TNFSF20,</td>
<td>257</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNFSF14 LIGHT, HVEML</td>
<td>258</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNFSF15 TL1, TL-1A, VEGI</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNFSF18 GITRL, TL-6, AITRL</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>TNFSF19 DEDD</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Squares colored yellow indicate where additional family members have recently been identified, using available genomes and sequence databases for analysis in this chapter. 

P, where a sequence has been predicted from the genome. 

M, where multiple sequences exist for this TNSF member.
within *Drosophila melanogaster*\(^{280,281}\), however, subsequent discoveries have been made within the molluscs *Haliothis discus discus*\(^{282,283}\) and *Crassostrea gigas*,\(^{157}\) and the crustacean *Marsupenaeus japonicus*\(^{284}\) Investigations in *Drosophila* have shown Eiger to have highest homology with the TNF family member Ectodysplasin A (EDA) and shown to be involved in a variety of cellular and tissue processes, such as cell death, cell proliferation, tissue growth regulation, host defense, and pain sensitization.\(^{281}\) A homolog to Eiger has also been identified in *M. japonicus*\(^{284}\) where expression was highest in healthy tissues, such as the muscle, stomach, brain, and gill. Peptidoglycan (PG) or polyI:C stimulation of cells taken from the “lymphoid organ,” that has been shown to be important in the elimination of microorganisms, increases the expression of this gene. All other TNFSF members identified within other invertebrates show varying homology to members of the mammalian TNFSF. The gene identified from *H. discus discus*\(^{282}\) has a TNFSF homology domain and has been described as a homolog of TNFSF2. These investigations have shown that this gene can be upregulated in vivo, in response to lipopolysaccharide (LPS), pathogenic bacteria, or virus; moreover, by using in situ hybridization both tissue and circulating hemocytes can be stained positive for this gene. A more recent study has identified that it is possible for invertebrates to contain a number of TNFSF members, as 23 individual TNFSF genes can be found in the genome of *C. gigas*,\(^{157}\) many of which have come from the duplication of genes. Functional diversity of each of these genes is predicted, due to differences in their expression in a variety of tissues. The role each of these TNFSF members have in the immune responses of these organisms remains unclear, and functional studies are required. In addition, caution should be taken in attempting to link sequences from invertebrates to homologs in vertebrates. Although it cannot be disputed that TNFSF members exist in invertebrates, any relation to a vertebrate gene may be purely coincidental, especially when the largest identity between the genes is mainly within the TNF homology domain, which is found in all TNFSF sequences.

### 7.2 Primitive Chordates and Basal Vertebrates: Cephalochordates, Urochordates, and Cyclostomata

Investigations have looked for the presence of TNFSF members within the most primitive chordates alive. Cephalochordates is a chordate subphylum that is represented by the Amphioxiformes (lancelets, also known as amphioxus), which are small, fish-like, filter-feeding animals found in shallow water, and are defined by the presence of a notochord that persists throughout their life. As described in more detail in chapter: Origin and Functions of Tunicate Hemocytes, urochordates are now considered the sister group of vertebrates. The Cyclostomata is a group of primitive vertebrates that comprises the jawless fishes, whose only living representatives are the lampreys and hagfishes. A draft amphioxus genome is available\(^{285,286}\) and offers insights into vertebrate origins and the evolution of immune system components, as cephalochordates and vertebrates evolved from
a common ancestor over 520 million years ago. In addition, a genome exists for the sea lamprey, and a transcriptome has been created from leucocytes isolated from peripheral blood, collected from the inshore lamprey. Using the amphioxus genome, the information on the tunicates Ciona intestinalis and Ciona savignyi, and the available lamprey transcriptome, a preliminary analysis of the immune gene repertoire has revealed that these groups contain members that belong to the TNFSF. Two were identified within the lamprey, whereas amphioxus had around 24, many of which were clustered in one region of the genome. Of these, only five sequences were cloned from Branchiostoma japonicum, and expression analysis revealed differences in their expression patterns in a variety of tissues. Attempts were made to identify whether any of the amphioxus and agnathan TNFSF sequences were homologous to existing vertebrate TNFSF members. What can be concluded is that the amphioxus genes form three very separate groups, with one of these containing a large number that is thought to have some homology to vertebrate TNFSF10. Interestingly, genes from the artic lamprey and inshore hagfish do group well with other vertebrate TNFSF10 sequences.

7.3 Fish

Using available fish genomes and sequences that have been deposited into databases, a number of studies have already been carried out in an attempt to characterize the TNFSF members that exist in bony fish. It is clear from these investigations that the diversification of the TNFSF occurred early in evolution, as fish have had an extensive number of homologs from this family identified. In addition, there may also be members of the TNFSF that are novel to this group of vertebrates.

7.3.1 TNFSF1-3

TNFSF2, also known as TNF-α, is a key cytokine in mammals that induces cell survival, apoptosis, and necrosis, and contributes to both physiological and pathological processes. It plays an important role in systemic inflammation, and is one of the cytokines released during the acute-phase reaction (Locksley et al., 2001). This was one of the first cytokines characterized within teleosts, and had been initially isolated from Japanese flounder and trout. Since then, it has been identified in a wide range of bony fish, which includes zebrafish, gilt-head seabream, common carp, channel catfish, fugu, tilapia, Atlantic salmon, mandarinfish, sea bass, ayu, turbot, goldfish, large yellow croaker, bluefin tuna, striped trumpeter, rock bream, orange-spotted grouper, grass carp, and crucian carp. Each of these sequences were shown to contain a transmembrane domain, a tumor necrosis factor-alpha converting enzyme (TACE) cut site, and the TNF ligand family signature, similar to their
human homolog. In addition, expression has also been looked at, and is upregulated in HK or macrophage primary cell-cultures after treatment with immunostimulants such as LPS, ConA, or PMA, or is upregulated in various tissues such as HK, spleen, and blood from fish infected with bacterial or viral pathogens.

Of interest in teleosts is the presence of multiple isoforms of TNFSF2 that have been discovered within a variety of species. Two copies were initially found in rainbow trout\textsuperscript{312} and four within the common carp,\textsuperscript{297,313,314} which share high sequence identity, but was not unexpected, as both species are tetraploid.\textsuperscript{315,316} Although the role of each of these genes was unclear, early investigations into their study patterns of expression showed that there were clear differences. However, with the discovery of at least two TNFSF2 genes within nontetraploid fish, such as bluefin tuna,\textsuperscript{307} orange-spotted grouper,\textsuperscript{309} zebrafish, and medaka,\textsuperscript{317} along with an additional gene within rainbow trout,\textsuperscript{318} it has become clear that teleosts have two very clear groups of TNFSF2 genes. Within the genomes of zebrafish and medaka,\textsuperscript{317} it has been shown that members from group I and group II can be found on a different chromosome (Fig. 5.5), with conservation of similar genes around it, indicating that this was due to one of the duplication events that has occurred within bony fish.\textsuperscript{315} Interestingly, in the bluefin tuna, zebrafish, and the trout there is a difference in expression of each group member in fish, or in cells stimulated with LPS. An analysis of sequences found within available fish genomes and databases shows that both groups of TNFSF2 molecules do exist in a variety of fish species. However, the role that each of these genes plays within the immune response of teleosts remains to be determined.

Some bioactivity studies have been carried out within selected species, with the early investigations using group I TNFSF2 proteins. A variety of activities were shown, including induction of proinflammatory factors in cells from immune tissues, recruitment of phagocytic cells, enhancement of phagocytic activity, enhancement of nitric oxide (NO) production, and regulation of nonspecific cytotoxic cells.\textsuperscript{300,304,305,307,308,319,320} However, as more studies were carried out within different fish species, some inconsistencies with certain activities were seen, which were initially highlighted in common carp and sea bream,\textsuperscript{323,324} especially with regard to the ability of TNFSF2 to activate phagocytes. These studies showed that TNFSF2 could act on phagocytes indirectly, by activating endothelial cells to produce stimulatory factors. To date, not all fish TNFSF2 molecules have been explored, and some of these differences could be attributed to the presence of multiple forms of this gene. Investigations have been carried out on the activities of members of the fish TNFSF2 group II. In bluefin tuna, similar activities were seen with regard to enhancement of phagocytic activity of peripheral blood leukocytes,\textsuperscript{308} and, in trout, primary macrophages, induction of pro- and anti-inflammatory cytokines, antimicrobial peptides, and the macrophage growth factor, IL-34, were observed.\textsuperscript{318} Apart from its role in the immune response
FIGURE 5.5  Synteny analysis of the locus containing the TNFSF2 gene from human, reptile, bird, amphibia, and fish. GeneID\textsuperscript{121} and GENSCAN\textsuperscript{122} were used to predict gene-coding regions of nonmammalian species to discover the gene order.
of fish, the activities of TNFSF2 in trout on other physiological pathways have also been investigated. A role within skeletal muscle metabolism was shown, where glucose uptake within trout muscle cells was stimulated,\textsuperscript{325} and some involvement in ovarian function was found, having a stimulatory effect on oocyte maturation.\textsuperscript{326}

To date, what still remains in question is the existence of fish homologs to TNFSF1 and TNFSF3, which in mammals are located, along with TNFSF2, next to each other on chromosome 6 in humans, within the major histocompatibility complex (MHC) class III region.\textsuperscript{327} TNFSF1, also known as TNF-\(\beta\) or LT-\(\alpha\), and TNFSF3, also known as LT-\(\beta\), mediate a large variety of inflammatory and immunostimulatory responses, are involved in the formation of secondary lymphoid organs during development, and also play a role in apoptosis.\textsuperscript{328} Initially in fugu and zebrafish,\textsuperscript{299} a gene was discovered next to the already characterized TNFSF2, within the genomes of these species, which has been called TNF-new (TNF-n). Subsequent investigations identified homologs of this gene within medaka, zebrafish,\textsuperscript{317} and trout,\textsuperscript{329} where two genes were identified (TNF-n1 and TNF-n2). Although not functional evidence, phylogenetic analysis of these sequences with known vertebrate TNFSF1-3 sequences, has shown a closer relationship to TNFSF3 than to TNFSF1 and TNFSF2. This becomes clearer with the inclusion of \textit{Xenopus} sequences, which are available for all three members\textsuperscript{329} and is strengthened with the addition of a recently isolated sequence from the African lungfish, where good identity to TNFSF3 is found.\textsuperscript{330} Preliminary investigations have looked at the expression of TNF-n in the previously mentioned species. Constitutive expression of TNF-n is highest within the gill, liver, and intestine of zebrafish, and shows little change in HK cells stimulated with LPS.\textsuperscript{331} In trout, it was shown that TNF-n1 had high constitutive expression in the gills and intestine, along with the spleen and HK, whereas there was little or no expression of TNF-n2 in any tissues.\textsuperscript{329} However, no expression of either gene was found in unstimulated and stimulated macrophages, or in fibroblasts. The characterization of this gene from more fish species and functional studies are required, before its exact identity can be confirmed.

7.3.2 TNFSF13 and 13B

TNFSF13, also known as apoptosis-inducing ligand (APRIL), and TNFSF13B, also known as B-cell activating factor belonging to the TNF family (BAFF), are two related members of the TNF ligand superfamily that share receptors, however, their functions are not redundant.\textsuperscript{332} TNFSF13B is a critical survival/maturation and proliferation factor for peripheral B cells, whereas APRIL appears to play a role in T-independent type II antigen responses and T-cell survival, but can also induce proliferation/survival of nonlymphoid cells.\textsuperscript{333–336} To date, TNFSF13 has only been characterized in two cyprinid species, grass carp and zebrafish,\textsuperscript{337,338} both of which contain a transmembrane domain, a TNF family signature, and a furin protease cleavage site. Searching the available sequence
databases provides evidence to show that TNFSF13 can be found in a number of fish species, and is not restricted to this group of fish (Bird, unpublished). In grass carp, their expression was highest in the skin, spleen, and HK, and their involvement within the immune response was shown as expression was upregulated in these tissues as well as in gill and liver, in response to bacterial and viral pathogens. Functional activity of this molecule was observed in the zebrafish, with the recombinant protein shown to bind the surface of spleen lymphocytes and help prolong the survival of these lymphocytes.

TNFSF13B has been characterized in a larger number of fish species, which includes zebrafish, \textsuperscript{339} mefugu, \textsuperscript{340} Japanese sea perch, \textsuperscript{341} grass carp, \textsuperscript{342} yellow grouper, \textsuperscript{343} miuuy croaker, \textsuperscript{344} and tongue sole.\textsuperscript{345} Similar to their human homolog, each of these sequences are shown to contain a transmembrane domain, a furin protease cleavage site, the TNF ligand family signature, and a conserved D-E loop (known as the “Flap”), which is unique to TNFSF13B and found in no other TNFSF. Comparison of the predicted 3D structure for the miuuy croaker and mefugu TNFSF13B with human revealed that they had a very similar folding pattern.\textsuperscript{340,344} Where expression of this gene was studied in healthy tissues, it was consistently shown that it could be detected within immune-relevant tissues, where highest expression was in the spleen, followed by the HK. Exceptions to this were seen in the grass carp and miuuy croaker, where highest expression was in the skin, followed by the spleen and HK. Modulation of this expression was seen in response to bacterial and viral pathogens. In tongue sole, an increase in expression of TNFSF13B was seen in both the spleen and kidney, 6, 12, and 24 h after infection with \textit{E. tarda}.\textsuperscript{345} TNFSF13B expression was also significantly upregulated within the gill, liver, kidney, spleen, and skin of grass carp after \textit{A. hydrophila} or \textit{Aquareovirus} infection.\textsuperscript{342} Recently highlighted, and similar to that found for TNFSF2 in fish, was the presence of more than one TNFSF13B gene in the miuuy croaker\textsuperscript{344} and fugu,\textsuperscript{346} which was termed TNFSF13B-like. An extensive search of the available fish genomes and sequences available within databases has revealed that many fish species have more than one TNFSF13B gene present, allowing two distinct groups to be observed. The relationship of each gene, with what is known in humans, has been investigated, and a prediction of the group II TNFSF13B gene 3D structure in the miuuy croaker shows that it has a very similar structure to the group I TNFSF13B.\textsuperscript{344} In addition, mapping the position of each TNFSF13B gene in the genomes of selected fish species has shown that group I TNFSF13B genes are located in a region that shares conserved synteny with the human genome, whereas those in group II are located in a different region, where no synteny is observed.\textsuperscript{346} However, this TNFSF13B-like gene is located next to another known TNF family member, ectodysplasin A (EDA). A preliminary investigation has looked at the role that these genes may have within the immune response of miuuy croaker, where the group I TNFSF13B is expressed at higher levels in the liver and kidney than the group II TNFSF13B during bacterial infection; however, this is reversed within the spleen. Limited functional
studies have been carried out, however, group I TNFSF13B genes from selected fish species have been investigated using recombinant proteins. TNFSF13B has been shown to bind to the surface of lymphocytes and to promote survival and proliferation of splenic B cells or lymphocytes in vitro. In addition, overexpression of TNFSF13B in vivo enhanced macrophage activation and reduced bacterial infection, whereas knocking down the expression led to fish being more susceptible.

As well as bony fish, TNFSF13B has also been characterized within a number of cartilaginous fish, such as the bamboo shark, spiny dogfish, and the small-spotted catshark. All proteins were predicted to contain a transmembrane domain, TNF family signature, and furin protease cleavage site. Modeling of the protein structure in bamboo shark also revealed the presence of the long D-E loop that is specific to all other characterized TNFSF13B proteins. Expression in healthy tissues from all three species indicates expression is high in immune tissues, especially the spleen; however, in the spiny dogfish the pancreas also showed high constitutive expression. Stimulation of peripheral blood leucocytes with various mitogens showed that PWM, which is mitogenic for both T and B cells, had the greatest effect on TNFSF13 expression in the dogfish and catfish, with polyI:C, LPS, ConA, PHA, and PMA having limited effects. In bamboo shark, functional activity of the protein was investigated and shown to stimulate the proliferation of mammalian B cells, however, its effects on cartilaginous fish cells remains to be seen. Interestingly, using sequences obtained from the available genomes and databases, it can be seen that, like bony fish, cartilaginous fish may also have two groups of TNFSF13B sequences, where the currently characterized bamboo shark belongs to group I and the dogfish and catshark belongs to group II, with sequences found in the elephant shark genome represented in both (Bird, unpublished).

7.3.3 Remaining TNFSF Members

Previous investigations using the available genomes and sequence databases has given good insight into which TNFSF members may exist in bony fish. Along with the already characterized molecules mentioned previously, homologs for TNFSF5, TNFSF6, TNFSF10, TNFSF11, TNFSF12, TNFSF14, and TNFSF18 have been identified and have subsequently been characterized in a variety of fish species, where a transmembrane domain and a TNF homology domain was conserved in all cases. TNFSF5, also known as CD40 ligand (CD40L), is important in the promotion of v dependent B cell responses, where CD40L is expressed by activated T cells that engage CD40 expressed by B cells, inducing B cell proliferation, differentiation, Ig production, Ig isotype class-switching, and promotion of germinal center (GC) formation. Initially this was identified within salmon, but has also been discovered within fugu. In fugu, highest constitutive expression was seen in the HK, liver, intestine, and heart, whereas in salmon it was observed in the HK, spleen, and gills, and was increased in HK leucocytes stimulated with PHA or ConA. In addition, overexpression of CD40L in a salmon cell-line,
which was cocultured with HK leucocytes, induced the rapid and long-lasting upregulation of key costimulatory molecules and cytokines that are important in T-helper cell responses.

TNFSF6, also known as Fas ligand (FasL), is predominantly expressed on activated T cells and natural killer-cells, and plays a crucial role in modulating immune responses by inducing cell-specific apoptosis. Investigations in Japanese flounder and fugu have identified this family member. Both predicted amino-acid sequences were shown to contain a proline-rich N-terminal sequence, which is essential for protein expression on the cell surface, and FasL is the only member of the TNF family with this feature. Constitutive expression was high in the fugu liver and heart, whereas in flounder it was the kidney and thymus; moreover, cytotoxic activity of the recombinant flounder FasL protein was shown against a flounder cell-line.

TNFSF10, also known as TNF-related apoptosis-inducing ligand (TRAIL), can induce apoptosis in various tumor cell-lines but not in normal cells, and administration in vivo can induce apoptosis in tumors with no cytotoxicity in normal organs or tissues. TNFSF10 has been identified in a wide range of fish, including zebrafish, grass carp, mandarin fish, and fugu. A predicted 3D model using the fugu sequence showed that it was similar to the human and contained all of the important structural amino acids conserved in the same positions. Constitutive mRNA expression was found in all tissues tested in each species, with the protein also widely expressed in the mandarin fish. Functional activity has also been investigated, where overexpression of the protein in HeLa cells, or the effect of the recombinant protein on Jurkat or HeLa, induces apoptosis in these cell lines. Interestingly in fugu, three genes have been identified with homology to TNFSF10, and multiple genes have also been shown to exist in other species, including fugu. Despite their relatedness, each gene in fugu has a very different gene organization, however, their expression profiles are similar in healthy tissues and almost the same in HK cells stimulated with LPS or polyI:C. Phylogenetic analysis has shown that each of these TNFSF10 genes in fish forms distinct groups, which is similar to our analysis showing that three very distinct groups of TNFSF10 can be seen.

The remaining TNFSF members identified in fish have had limited investigations carried out. TNFSF12, also known as TNF-like weak inducer of apoptosis (TWEAK), has multiple biological activities, including stimulation of cell growth and angiogenesis, induction of inflammatory cytokines, and stimulation of apoptosis. It has been characterized within the grass carp and fugu, where skin shows highest expression in both species, followed by spleen, HK, gill, and brain in the carp. In addition, bacterial and viral expression significantly upregulated TNFSF12 expression in carp spleen, HK, gill, and brain. TNFSF14, also known as LT-related Inducible ligand that competes for Glycoprotein D binding to Herpesvirus entry mediator on T cells (LIGHT), functions as a costimulatory molecule for T cells and promotes the activation
and expansion of T-cell populations. Homologs of this gene have been investigated in zebrafish, where highest constitutive expression was seen in the spleen, followed by HK, and in fugu where expression was highest in the HK and liver. A recombinant protein was also shown to bind to the surface of zebrafish lymphocytes. TNFSF18, also known as glucocorticoid-induced TNF receptor ligand (GITRL), participates in the immune response against tumors and infectious agents by influencing the activity of effector and regulatory T cells. Only one functional study exists, where a homolog of this gene is found in the zebrafish. Its activity has been looked at during development, where its knockdown results in disruption of embryogenesis. Lastly, within the databases, a sequence can be found within a number of bony fish with homology to TNFSF19 (also known as DEDD), which is a new member to the TNFSF. It is also a regulator and executor of the programmed cell-death pathway, with its overexpression shown to induce weak apoptosis. This sequence groups well with TNFSF19 found within other vertebrate groups (Bird, unpublished).

Not much has been done to characterize these remaining TNFSF members in cartilaginous fish. Recently, a TNFSF5 homolog has been characterized in the small-spotted catshark. In this study, expression was found to be highest within the gut, spleen, and gill, which correlated directly with a T-cell marker, TCRα, and its expression could be upregulated in PBLs, following stimulation with PHA, PMA, and PWM. A search of the elephant shark genome and sequences available in databases has revealed that a number of other TNFSF members may exist. Phylogenetic analysis of these with representative sequences from each major vertebrate group shows that, along with TNFSF5, homologs to TNFSF6, TNFSF10, and TNFSF15 can also be identified, with multiple sequences available for TNFSF6 and TNFSF10. However, a number of sequences do exist in the elephant shark genome that belong to the TNFSF, but are difficult to assign to an existing homolog in humans.

### 7.4 Birds, Reptiles, and Amphibians

An interrogation of both the chicken and turkey genomes has allowed the identification of a number of TNFSF members that have been characterized within mammals. However, TNFSF1, TNFSF2, TNFSF3, TNFSF7, TNFSF9, TNFSF12, TNFSF13, TNFSF14, and TNFSF19 remained unidentified within birds, and for some it was speculated that they may not be present within this vertebrate group. The absence of TNFSF1 and TNFSF3 was not unexpected, as these genes play an important role in lymph node development in mammals, which is lacking in birds. Interestingly, the presence of TNFSF2 was expected, as TNF-α-like activity had been reported in chickens. An examination of all available bird genomes has revealed a partial sequence with homology to TNFSF2 within the Tibetan ground-tit, *Pseudopodoces humilis*, on scaffold 880 (Accession no. NW_005087991.1). Using this sequence within a phylogenetic analysis does
group it with other characterized TNFSF2 sequences, and prediction of genes on this scaffold does reveal that genes GPANK1 and CSNK2B can be determined, which are also found within regions of the human genome that contains TNFSF2, showing some conservation of synteny (Fig. 5.5). In addition to what has previously been characterized, an examination of all available bird sequence-data has also revealed the existence of a TNFSF14 and TNFSF19 homolog (Bird, unpublished).

Other than the discovery of gene sequences using bird genomes, limited information exists on the activity of these identified TNFSF genes in birds. The best characterized is TNFSF13B, where the gene has been isolated from chicken, goose, duck, quail, dove, and ostrich. In each of these studies, expression was highest within the spleen and bursa of fabricius, where lymphocyte concentrations are high, and the recombinant protein was shown to play an important role in the survival and proliferation of B cells in each species. The activity of recombinant chicken TNFSF5 has also been investigated, and shown to maintain proliferation of B cells in culture for up to 3 weeks and to maintain direct differentiation of cells toward a plasma-cell phenotype. In addition, antigen-specific IgM and IgG was secreted by TNFSF5-activated B cells, taken from immunized birds. Lastly, TNFSF8, also known as CD30L, which plays an inhibitory role in modulating Ig class-switching in B cells along with TNFSF10 have also been loosely characterized in chicken. Both genes were shown to have high levels of expression in the spleen, in bursa of Fabricius, and in a chicken leukemia cell-line, IN24.

Almost no studies have been carried out looking at the TNFSF members that are present within reptiles. One recent study exists that has isolated and characterized BAFF within the Yangtze alligator. Real-time PCR indicated that the highest expression was within the spleen, and the recombinant protein was able to promote the survival of alligator spleen lymphocytes. In addition, this protein also showed some cross-reactivity in mammals, helping to promote the survival of mouse spleen cells in culture. To determine the other TNFSF members that exist within reptiles, a search of the green anole genome was carried out, along with the available reptile sequences within the databases. Many of the TNSF members that had been found in birds had a homolog within this vertebrate group (Table 5.2). Only TNFSF18 needed to be predicted from the genome, and was found on Scaffold chrUn0590 (Accession No. GL343782.1), along with a related sequence from the green sea turtle (Accession no. XP_007072858.1). The only main differences found were within A. carolinensis, where a homolog for TNFSF13 was identified (Accession no. XP_008120421.1), and a distinct region was discovered in the genome that contained a TNFSF2 homolog, which shared some synteny with the human and amphibian genomes (Fig. 5.5). Interestingly, two TNFSF2 genes are predicted within this region, with no sequence similarity to TNFSF1 and TNFSF3.

Amphibians have had some TNFSF members identified, which have been predicted using the X. tropicalis genome, with a select few having been sequenced and functionally characterized. TNFSF2 was one of the first to be
studied\textsuperscript{379} in *Xenopus laevis*. In amphibians, it has been suggested that TNFSF2 may play an important role in tadpole metamorphosis, as TNFSF2 expression was prominent in the blood at prometamorphosis, and the recombinant protein found to suppress thyroid hormone induced apoptosis in a vascular endothelial cell-line.\textsuperscript{379} More recently, homologs of *X. laevis* TNFSF10 and TNFSF13B have also been identified. Two homologs of TNFSF10 exist in *X. laevis*\textsuperscript{380} and have been shown to play a role in erythroid maturation, which is similar to human studies.\textsuperscript{381} TRAIL1 and TRAIL2 both show high expression within the liver and red blood cells during metamorphosis; in tadpoles, TRAIL1 was shown to accelerate hemoglobin switching and induce apoptosis in larval, but not adult, red blood cells. TNSF13B was isolated from *X. laevis*, and, similar to what was found in reptiles and birds, the expression was highest in the spleen, and the recombinant protein was able to promote survival and proliferation of lymphocytes from both *Xenopus* and mouse.\textsuperscript{382} Lastly, using the *X. tropicalis* genome\textsuperscript{378} and available databases, a search of the remaining TNFSF that had not yet been characterized was undertaken. This search revealed the presence of TNFSF4, TNFSF8, TNFSF9, TNFSF13, TNFSF14, TNFSF15, and TNFSF19 (Table 5.2). Of particular interest is the presence of TNFSF1 and TNFSF3, which have currently no homologs in reptiles or birds, and were first characterized in a study looking at a TNFSF member discovered within trout.\textsuperscript{329} These two genes were identified due to the conservation of synteny that this species has with the human genome around the TNFSF2 gene. An analysis of Scaffold\_752 (Accession no. GL173388.1) identified that TNFSF1, TNFSF2, and TNFSF3 were located next to each other, similar to what is found in the human genome, with conservation of other genes around them (Fig. 5.5).

REFERENCES

37. Ingerslev HC, Cunningham C, Wergeland HI. Cloning and expression of TNF-alpha IL-1 beta and COX-2 in an anadromous and landlocked strain of Atlantic salmon (Salmo salar L.) during the smolting period. Fish Shellfish Immunol 2006;20:450–61.
45. Covello JM, Bird S, Morrison RN, Battaglene SC, Secombes CJ, Nowak BF. Cloning and expression analysis of three striped trumpeter (Latris lineata) pro-inflammatory cytokines, TNF-alpha, IL-1beta and IL-8, in response to infection by the ectoparasitic, Chondracanthus goldsmidi. Fish Shellfish Immunol 2009;26:773–86.

49. Lu XJ, Chen J, He YQ, Shi YH. Molecular characterization of an IL-1β gene from ayu (Plecoglossus altivelis). Fish Shellfish Immunol 2013;34:1253–9.


64. Øvergård AC, Nepstad I, Nerland AH, Patel S. Characterisation and expression analysis of the Atlantic halibut (Hippoglossus hippoglossus L.) cytokines: IL-1β, IL-6, IL-11 IL-12β and IFNγ. Mol Biol Rep 2012;39:2201–13.


128. Gibson MS, Fife M, Bird S, Salmon N, Kaiser P. Identification, cloning, and functional characterization of the IL-1 receptor antagonist in the chicken reveal important differences between the chicken and mammals. *J Immunol* 2012;189:539–50.


135. Towne JE, Garka KE, Renshaw BR, Virca GD, Sims JE. Interleukin (IL)-1F6, IL-1F8, and IL-1F9 signal through IL-1Rrp2 and IL-1RAcP to activate the pathway leading to NF-kappaB and MAPKs. *J Biol Chem* 2004;279:13677–88.


150. Wang T, Huang W, Costa MM, Martin SA, Secombes CJ. Two copies of the genes encoding the subunits of putative interleukin (IL)-4/IL-13 receptors, IL-4Rα, IL-13Rα1 and IL-13Rα2, have been identified in rainbow trout (Oncorhynchus mykiss) and have complex patterns of expression and modulation. *Immunogenetics* 2011;63:235–53.


226. Igawa D, Sakai M, Savan R. An unexpected discovery of two interferon gamma-like genes along with interleukin(IL)-22 and -26 from teleost: IL-22 and -26 genes have been described for the first time outside mammals. *Mol Immunol* 2006;43:999–1009.


The Evolution of the Immune System


262. Wang T, Martin SAM, Secombe C. Two interleukin-17C-like genes exist in rainbow trout Oncorhynchus mykiss that are differentially expressed and modulated. Dev Comp Immunol 2010;34:491–500.


309. Lam FW, Wu SY, Lin SJ, Lin CC, Chen YM, Wang HC, Chen TY, Lin HT, Lin JH. The expression of two novel orange-spotted grouper (*Epinephelus coioides*) TNF genes...


The Evolution of Complement System Functions and Pathways in Vertebrates

Miki Nakao, Tomonori Somamoto
Department of Bioscience and Biotechnology, Faculty of Agriculture, Kyushu University, Hakozaki, Fukuoka, Japan

1 INTRODUCTION

The complement is a major humoral system of innate immunity. The mammalian system, the best characterized to date, is composed of more than 30 distinct plasma proteins and membrane-bound proteins, and mediates various immune effector functions, including elimination of invading pathogens, promotion of inflammatory responses, and clearance of apoptotic cell and necrotic cell debris, as well as modulation of humoral and cellular responses of adaptive immunity.\(^1,2\)

The mammalian system is equipped with three well-defined activation pathways—the classical, alternative, and lectin pathways—to proteolytically convert C3, the central complement component, into its activated fragments C3a and C3b, followed by activation of the terminal lytic pathway, which damages the target-cell membrane. (Fig. 6.1)

The classical activation pathway is triggered upon recognition of an antigen-bound antibody by C1, followed by the proteolytic cascade involving C4 and C2, which form a C4bC2a complex or the C3-convertase. The lectin pathway is activated by various microbial carbohydrates, defined as the pathogen-associated molecular patterns (PAMPs). PAMPs are recognized by two distinct serum lectins, mannose-binding lectin (MBL) and ficolin (FCN), which are oligomeric proteins coupling with the MBL-associated serine proteases (MASPs) for the proteolytic activation of C4 and C2, as in the classical activation pathway.\(^3\) Since C1 and MBL/FCN-MASPs require Ca\(^{2+}\) ion to maintain their quaternary structure, and C2 needs Mg\(^{2+}\) ion for its proper conformation, these two pathways proceed in the presence of both Ca\(^{2+}\) and Mg\(^{2+}\) at a millimolar level.\(^4\) The initiation of the third activation pathway, the alternative pathway, does not rely on any clear recognition molecules. Instead, certain exogenous activators such
The Evolution of the Immune System

The Evolution of the Immune System

as bacterial lipopolysaccharide and fungal β-glucans lead to spontaneous C3 activation or C3-tickover, thanks to the formation of a C3-convertase, C3bBb, involving C3, factor B, and factor D. In this case, the normal suppression operated by various regulators of complement activation is overcome.

Upon proteolytic activation by the C3-convertases, the nascent C3b can bind covalently to target-cell surface using its intrachain thioester bond, allowing further formation of surface-bound C3bBb; thus the alternative pathway can amplify C3-activation on the target surface. At the same time, C3b can bind to C4bC2a or C3bBb to form trimolecular complexes, C4bC2aC3b, or C3bBbC3b, respectively, which can cleave C5 into its activated fragments C5a and C5b. The latter initiates molecular gathering of the terminal components C6 through C9 to form the membrane-attack complex (MAC) in the lytic pathway.

In addition to the cytotoxicity against the target cell, the activated fragments of C3 and C5 mediate various innate immune-effector functions, including opsonization of target cells and promotion of inflammatory responses. Namely, C3b and its further degradation product iC3b are recognized by complement receptors type 1 and type 3 (CR1 and CR3), respectively, expressed on phagocytes. On the other hand, C3a and C5a stimulate various leukocytes through interaction with their specific receptors C3aR and C5aR, respectively, inducing degranulation and chemotaxis.

The potent physiological activities of the complement are tightly controlled at various steps of the activation cascades to prevent excess or misdirected

FIGURE 6.1 Activation pathways, receptors, and regulators of mammalian complement system. Complement components for activation are in gothic letters, receptors are in italic, and negative regulators are in roman fonts. Components with proteolytic activity are shown in red. Abbreviations: MBL, mannose-binding lectin; FCN, ficolin; MASP, MBL-associated serine protease; P, properdin; C1INH, C1-inhibitor; CR, complement receptor; C4bp, C4-binding protein; DAF, decay-accelerating factor; MCP, membrane-cofactor protein; If, factor I; CP-N, carboxypeptidase N. Thick arrows indicate proteolytic actions and block arrows mean receptor-ligand recognitions.
activation, which may damage host tissues in a proximity of the complement activation sites. C1-inhibitor (C1INH) inactivates MASPs, C1r, and C1s; C4-binding protein and factor H enhance degradation of C4b and C3b by factor I protease (If); membrane-cofactor protein (MCP) and decay-accelerating factor (DAF) disturb the C3-convertase formation in an If-dependent and independent manner, respectively; serum carboxypeptidase N attenuates anaphylatoxic activities of C3a and C5a by removing their C-terminal Arg; CD59 inhibits the MAC formation on the host cell.5

Taxonomically, vertebrate is a subphylum of the Chordata phylum, composed of two superclasses, Agnatha (jawless cyclostome species), such as hagfish and lamprey, and Gnathostomata (jawed species), which includes cartilaginous and bony fishes, amphibians, reptiles, birds, and mammals.6 As an evolution mechanism of vertebrates, the 2R hypothesis was proposed by Ohno (1970)7 and is now well-accepted by a number of phylogenetic and genomic evidence.8 According to the 2R hypothesis, a vertebrate ancestor (?) should have undergone two rounds of whole-genome duplication (WGD), which was probably tetraploidization of chromosomes. The first round of WGD is considered to occur prior to the appearance of the first jawless vertebrates, and the second before the emergence of the jawed vertebrates. In the immune system, this whole genome duplication is believed to have contributed to creation of the adaptive immunity made up of immunoglobulins (Igs), T-cell receptors (TCRs), and the major histocompatibility complex (MHC). Interestingly, a large (and the most common) bony fish group, teleost, has experienced an additional WGD (designated 3R) at a stem stage of the lineage, and this further large-scale gene duplication is believed to have impacted a teleost-specific diversification both in morphology and physiology.9

2 INVERTEBRATE COMPLEMENT SYSTEM REPRESENTING AN ANCESTRAL ARCHITECTURE OF VERTEBRATE SYSTEM

Comprehensive searches of complement-related components by molecular cloning, as well as genomic and transcriptomic database surveys from a number of invertebrate animal phyla, including Cnidaria,10 Arthropoda,11 and Mollusca,12 have illustrated an original architecture of the ancestral complement system. The ancestral complement was probably an opsonic system composed of C3, Bf, and MASP. In this system, MASP could be activated upon sensing of PAMPs by some pattern-recognition molecule, resulting in activation of Bf, which in turn proteolytically activates C3 into C3a and C3b fragments to mediate inflammatory and phagocytic cellular responses, respectively.13

Although the complement system had lost in some protostome invertebrate lineages such as nematodes and insects, the basic architecture as an opsonic system has been well-maintained in deuterostome invertebrate phyla such as the Echinodermata and the Chordata subphyla Cephalochordata and Urochordata.13 In chapter: Origin and Functions of Tunicate Hemocytes, Cima et al.
briefly describe the complement elements discovered in Urochordates. The architecture of a prototypic complement system in the chordates seems to be a combination of the lectin and the alternative pathways, composed of MBL-like lectin as a PAMPs sensor, Bf as a C3-activating enzyme, C3 as an opsonin, and C3b-receptors on phagocytes (Fig. 6.2).

In the chapter, we deal with phylogeny of the complement pathways, complement receptors, and regulatory factors in vertebrates, from the molecular and functional points of view, with a reference to the contribution of the WGD, postulated by the 2R hypothesis. Evolutionary impact of additional WGD occurring in a stem lineage of teleost on the complement system will also be discussed.

3 PHYLOGENY OF COMPLEMENT PATHWAYS

Genomic and transcriptomic approaches in recent years have accelerated identification of homologues of the mammalian complement components in lower vertebrates and invertebrates. Table 6.1 is an update of the phylogenetic distribution of components, pathways, and physiological activities of a complement system, focusing more on the vertebrate system, modified from the preceding information in the literature.14

3.1 Lectin Pathway

In mammals, pattern-recognition molecules of the lectin pathway are two collectins, MBL and CL-K1, and three ficolins, namely FCN1, FCN2, and FCN3.3,15
| Table 6.1 Complement Components, Receptors, and Regulatory Factors Identified in Representative Vertebrate Classes

<table>
<thead>
<tr>
<th>Invertebrate</th>
<th>Agnatha</th>
<th>Gnathostome</th>
<th>Gnathostome</th>
<th>Tetrapods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cephalo-chordate/urochordate</td>
<td>Jawless cyclostome</td>
<td>Cartilaginous fish</td>
<td>Bony fish (teleost)</td>
<td>Amphibia</td>
</tr>
<tr>
<td>MBL</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>FCN</td>
<td>x</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>MASP1</td>
<td>o</td>
<td>o</td>
<td>x</td>
<td>o</td>
</tr>
<tr>
<td>MASP2</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>MSAP3</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>MAp19</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>MAp44</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>C1q</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>C1r</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>C1s</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>C3</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>C4</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>C5</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Bf</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
</tbody>
</table>

(Continued)
### TABLE 6.1 Complement Components, Receptors, and Regulatory Factors Identified in Representative Vertebrate Classes (cont.)

<table>
<thead>
<tr>
<th>Invertebrate</th>
<th>Cephalo-chordate/urochordate</th>
<th>Cephalo-chordate/urochordate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jawless cyclostome</td>
<td>Jawless cyclostome</td>
</tr>
<tr>
<td></td>
<td>Cartilaginous fish</td>
<td>Cartilaginous fish</td>
</tr>
<tr>
<td></td>
<td>Bony fish (teleost)</td>
<td>Bony fish (teleost)</td>
</tr>
<tr>
<td></td>
<td>Amphibia</td>
<td>Amphibia</td>
</tr>
<tr>
<td></td>
<td>Birds</td>
<td>Birds</td>
</tr>
<tr>
<td></td>
<td>Mammals</td>
<td>Mammals</td>
</tr>
<tr>
<td>C2</td>
<td>o?</td>
<td>o?</td>
</tr>
<tr>
<td>Df</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>C6</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>C7</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>C8α</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>C8β</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>C8γ</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>C9</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Hf</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>DAF</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>MCP</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>CR1</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>CR2</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Factor</td>
<td>Agnatha</td>
<td>Cephalochordate</td>
</tr>
<tr>
<td>------------</td>
<td>---------</td>
<td>----------------</td>
</tr>
<tr>
<td>C4bp</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Lf</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>C1INH</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>CR3</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>C1qR</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>C3aR</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>C5aR</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>CPaseN</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>CD59</td>
<td>o</td>
<td></td>
</tr>
</tbody>
</table>

*The marks “o” and “x” indicate the presence and the absence, respectively, of the factor supported by evidence in the literature or database.*
While MBL and FCN partly share binding specificity toward monosaccharide ligands such as GlcNAc, these two lectins recognize distinct PAMPs such as bacterial peptidoglycans and fungal β-1,3-glucans, respectively. In nonmammalian vertebrates, genes encoding MBL, CL-K1 (COLLEC11), and an FCN have been identified in chickens; those encoding three FCN homologues have been reported for *Xenopus*; three types of collectins—MBL, GalBL, and CL-K1—have been found in teleost species, and an MBL homologue has been isolated from lamprey.

Interestingly, teleosts probably lack any homologue of FCN defined as a collagenous lectin with a C-terminal fibrinogen-like domain, but possess GalBL, an MBL-like collectin with galactose-directed binding specificity. Therefore GalBL may be a functional homologue of FCN in teleost, but this hypothesis needs verification by functional assay at the protein level.

Another interesting point in evolution of vertebrate collectins is a conservation of MBL and CL-K1 from teleosts up to mammals. These two collectins are classified into two distant groups, designated “classical” collectins, including MBL, SP-A, and SP-D, and “novel” collectins, including CL-K1, CL-P1, and CL-L1, although they all share the same domain organization, namely N-terminal Cys-rich region, collagenous stalk, neck region, and C-terminal C-type lectin domain. The existence of the classical and novel collectins in teleosts, chickens, and mammals suggest their ancient evolutionary origin and functional importance in the recognition of pathogen and other danger-associated molecular patterns.

A group of homologous protease, termed MBL-associated serine proteases (MASPs), are complexed with the collectins and FCNs, and play an essential role in triggering a proteolytic cascade for C3- and C4-activation. In human, five MASP-like proteins with and without proteolytic activity are produced from two distinct genes, MASP1/3 gene and MASP2 gene. The MASP1/3 gene produces two proenzymes, MASP1 and MASP3, and a truncated form without catalytic domain, MAp44 or MAP-1, through alternative polyadenylation and splicing of mRNA. On the other hand, the MASP2 gene produces a proenzyme MASP2 and a noncatalytic truncated form, MAp19 (sMAP), as splicing variants.

In the mammalian system, MASP2 is responsible for proteolytic activation of C4 and C2, leading to a formation of the C3-convertase, C4bC2a. MASP1 can cleave proenzymes MASP2 and MASP3 into its active form and also profactor D into active factor D, in addition to many other noncomplement substrates, such as coagulation factor XIII, prothrombin, and kinogen. Natural substrate of MASP3 has long remained unclear, but a recent report demonstrated that MASP3, as well as MASP1, converts pro-factor D into active factor D, implying its crucial role in triggering the alternative pathway activation. Roles of MAp19 and MAp44 are still to be analyzed, though negative regulatory effects on lectin pathway activation by competing with MASP1, MASP2, and MASP3 have been speculated.

An origin of the MASP1/3 gene can be traced back to an ancestral invertebrate chordate based on identification in cephalochordates, urochordates, agnathan
Evolution and Pathways in Vertebrates

Chapter | 6 | 159

(lamprey), and jawed vertebrates (shark, teleost, and frog). On the other hand, the MASP2 gene seems to have occurred after emergence of agnathans, but before that of teleosts, and to have been retained in all the jawed vertebrates.

It is intriguing that MASP1 gene encoding a protein with MASP1-specific serine protease domain is probably missing in teleosts and chickens, even after a comprehensive survey of genomic-sequence data. Mammalian MASP1 is believed to play various roles in not only the complement lectin pathway, but also in coagulation and inflammation. Thus, the absence of an MASP1 gene raises an interesting issue on the evolution of MASP1-mediated functional linkage between innate immunity and other physiological pathways.

It should also be noted that MAp44 also has an ancient origin. It was first described in a teleost species, the common carp, as a truncated isoform of MASP3, designated MRP, with domain organization of CUB1-EGF-CUB2-SCR1. After identification of its mammalian homolog, named MAp44, generated by exactly the same splicing patterns as those of carp, a similar homologue has been found in urochordates, in which MAp44 is encoded by a distinct gene, but not as a splicing variant. Based on such strong conservation, MAp44 may be an important regulator of lectin pathway activation, rather than MAp19, which only appears in mammals.

3.2 Alternative Pathway

The basic architecture of the alternative pathway, composed of C3 and factor B, seems to be tightly conserved in any group of vertebrates, as supported by identification of the two core components from all the vertebrate classes. This pathway is believed to be important to amplify the generation of C3a and C3b, which are responsible for most of the complement-mediated defense and inflammatory functions.

There is a divergence in the polypeptide structure in C3 between lampreys and the more recent jawed species. Namely, whereas C3 of jawed species has a two-chain structure composed of α- and β-chains linked by a disulfide bond, lamprey C3 has a mammalian C4-like three-chain structure (α-, β-, and γ-chains), caused by insertion of a posttranslational processing site to cleave into α- and γ-chains. It is interesting that C3 of hagfish, another jawless species, keeps the two-chain protein in spite of the presence of a similar processing site.

There is also a slight but significant structural difference in Bf between jawless and jawed vertebrates. While the serine protease domain shows a trypsin-like substrate specificity cleaving at C-terminal Arg residue corresponding to C-terminus of C3a region, an Asp residue that defines the specificity is located at different positions in their primary structures. In lamprey, as in invertebrate species, the Asp is six residues N-terminal from the active Ser. Conversely, the Asp is found at 39–42 residues C-terminal side of the active Ser in all the jawed vertebrate species analyzed so far. Thus, it is suggested that a structural specialization of the Bf serine protease domain has occurred in a common ancestor of Gnathostomes.
Another intriguing issue on phylogeny of the alternative pathway is the occurrence of factor D. To date, a serine protease clearly identified as factor D has been found only in jawed vertebrates, except for cartilaginous fish. Since proteolytic conversion of Bf, a proenzyme, into an active Bb fragment seems essential for the alternative pathway activation, it is very interesting to identify a protease responsible for the Bf cleavage. So far, even comprehensive transcriptomic approaches have not detected any Df-like molecule in jawless species.

3.3 Classical Pathway

The classical pathway, as detected by antibody-dependent and Ca\(^{2+}\)- and Mg\(^{2+}\)-dependent hemolytic activity, has long been demonstrated in the jawed vertebrate species, including sharks, teleosts, and tetrapods. This antibody dependency is attributable to a C1q subcomponent of C1, which recognizes immunoglobulins complexed with an antigen. The divalent cation-requirement reflects that C1 is a Ca\(^{2+}\)-dependent complex of C1q, C1r, and C1s, and that C2 needs coordination of Mg\(^{2+}\) for its normal conformation, as does Bf.

C1q, a member of the C1q family, is composed of many proteins with an N-terminal collagenous region and a C-terminal C1q-like globular domain. In mammals, C1q is a hetero 18-mer, made up with six chains each of C1q-A, C1q-B, and C1q-C polypeptides. As a subcomponent of C1, C1q is responsible for recognition of antigen-bound Ig Fc-region and other natural ligands. In the serum, C1q is complexed with two molecules each of C1r and C1s, which are serine proteases responsible of C4 activation. In human chromosome 1, C1q-A, C1q-C, and C1q-B genes are located in tandem in this order. Conservation of the heterooligomeric composition of C1q, with the A, B, and C-chains, and their tandem gene arrangement has been reported in zebrafish. Since C1q binds to a range of nonantibody ligands such as LPS and pentraxins, the classical pathway may be triggered by various PAMPs in an antibody-independent manner in vivo.

Complement C1q-like molecules have also been identified as a N-acetylglucosamine (GlcNAc)-binding serum protein from two agnathan species, lamprey and hagfish. Lamprey C1q is a homo-oligomer coupled with MASP-A, which is a MASP orthologue that activates C3. It was recently reported that lamprey C1q recognizes the variable lymphocyte receptor B (VLRB), a functional alternative of immunoglobulins, bound to target antigen cells, and may possibly lead to its cytolyis. Hagfish GlcNAc-binding C1q showed an affinity to agarobiose and recognized a wide spectrum of PAMPs and microbes.

C4, a C3 homologue composed of three disulfide-linked polypeptides, α-, β-, and γ-chains, covalently binds to the target surface through the thioester site, and provides a platform for binding of C2 to construct the classical pathway C3-convertase. Thus, emergence of C4 would have played a key role for establishment of the classical pathway. C4-encoding cDNAs or C4 proteins have been isolated from all the jawed vertebrates as low as cartilaginous and bony fish, but not from jawless species.
In the mammalian classical pathway, C2 plays a role homologous to that of Bf in the alternative pathway, providing a catalytic subunit of the C3-convertase, C4bC2a, for proteolytic C3 activation.\(^5\) From a jawless species, lamprey, two Bf/C2-like sequences have been cloned and identified as ancestral components yet to diverge into Bf and C2, functioning in the alternative pathway C3-convertase.\(^33\) In cartilaginous and bony fish, two Bf/C2-like sequences as divergent as that between mammalian Bf and C2 have been cloned from a single species. Phylogenetic analyses of these Bf/C2-like molecules, however, did not show clear assignment to either Bf or C2, making it unclear whether cartilaginous and bony fish have functional C2.\(^{42}\) More data on their functions, such as ability of C3-convertase formation with C4b or C3b, are still needed to answer this question. Interestingly, it has been reported that depletion of a Bf-like protein from trout serum resulted in total loss of its hemolytic activity through both the alternative and classical pathway, suggesting that teleost Bf also functions as C2 in the classical activation cascade.\(^{43}\) Thus, presence or absence of a functional C2 molecule in fish is a key issue to better understand evolution of the classical pathway, from a functional point of view, in the lower vertebrates. The C2 molecule clearly identified as a classical pathway component seems to have evolved before the divergence of Amphibia from the tetrapod lineage, as evidenced by phylogenetic sequence analysis and functional assay.\(^{14}\)

### 3.4 Lytic Pathway

Lytic activity clearly attributable to the complement system has been described only in jawed vertebrates, in which a group of terminal components, including C6, C7, C8, and C9 have been identified at both DNA and protein levels.\(^{42,44–47}\) On the other hand, a natural hemolytic activity of lamprey serum is caused by a single 25-kDa protein and shows no relation to the lamprey complement system, which is only opsonic.\(^{29}\) Agreeing well, comprehensive genomic and transcriptomic searches have found no sequences similar to C6, C7, C8, or C9 from lamprey or hagfish databases.\(^{33}\)

In a recent publication, Nonaka classified the terminal complement components into two subgroups, the C6-like (C6 and C7) and the C9-like (C8α, C8β, and C9), based on the presence and absence of C-terminal FIMAC domains.\(^{13}\) This classification is functionally relevant because the C6-like components interact with C5b through the FIMAC domains to start MAC formation, whereas C9-like components show no direct binding to C5b due to the lack of FIMAC domains.\(^{48}\) Although C9-like molecules have been found in cephalochordates and urochordates, it seems unlikely that they are functionally connected, with their opsonic-complement cascade even lacking C5, according to the structure–function relationship of the terminal components.\(^{18}\) Nevertheless, that invertebrate C9-like molecule might represent an ancestor of the terminal components seen in the jawed vertebrates.

Among the terminal components, C8 has a novel quaternary structure in that it is composed of three polypeptides, which are C8α and C8β, homologous
to each other, and C8γ, belonging to a remote protein family, lipocalin. C8γ has been cloned from teleost and mammals, and the mammalian C8-like three-chain structure has also been confirmed at the protein level in teleosts. In cartilaginous fish, the presence of C8γ remains to be proven.

In the functional point of view, C9 play a key role to complete the formation of a barrel-shaped membrane–attack complex, but is optional for lysis of particular target cells, as reported for hemolysis of rabbit erythrocytes, which can be lysed without C9. Although molecular mechanisms of the complement-mediated cytolysis without C9 are still to be clarified, it should be noted that the C9-independent rabbit-cell lysis has also been described in the bony fish complement, suggesting tight conservation of the membrane-attack mechanism by the terminal components.

4 PHYLOGENY OF COMPLEMENT RECEPTORS

Complement receptors play crucial roles in mediating biodefense responses of the complement system by sensing various active fragments of complement components. Mammalian complement receptors include opsonic receptors such as C1q-receptor, CR1, CR3, and CR4, anaphylatoxin receptors such as C3a-receptor and C5a receptor, and immunomodulating receptors such as CR2. Opsonic function of the complement system has been described in bony fish and higher vertebrates, in which C3b and iC3b are candidates of opsonic complement ligand. At the molecular level, CR3 subunits, or integrin CD18 and CD11, have been cloned from bony fish and mammals. Although none of the CR3 subunits have been found in the jawless species, CD18/CD11 homologues have been shown to mediate opsonic function of the complement in a urochordate species. Therefore, CR3-mediated opsonization is considered to be a well-conserved mechanism for pathogen elimination in all the vertebrate species. Other opsonic receptors homologous to C1q-receptor and CR1 are not clearly recognized outside of mammals.

Complement fragments corresponding to mammalian C3a, C4a, and C5a have been described in bony fish. Similarly, jawless fish C3 can also release a corresponding C3a fragment, although its function is yet to be analyzed. C3a- and C5a-receptors have been identified also in bony fish, and chemotactic and anaphylatoxic response induced by C3a and C5a have been described in cartilaginous and bony fish and higher vertebrates. In jawless species, chemotactic response of leukocytes seemingly attributable to complement anaphylatoxins have been reported in hagfish, but the ligand and receptors that mediated the chemotaxis have not been identified, yet.

Mammalian CR2 expressed on B cells modulates antibody response by recognizing C3d, a terminal activation fragment of C3, bound to a foreign antigen. While generation of C3d-like fragments is evident in bony fish, no direct CR2 homologue has been found in this group. It would be particularly interesting if the bony fish complement has such immunomodulatory connection to the adaptive
Evolution and Pathways in Vertebrates

Chapter 6

immunity, not only from an evolutionary aspect, but also from a practical point of view. This would indeed represent the basis for the development of a molecular adjuvant utilizing fish C3d fragments for disease prevention in aquaculture.

5 PHYLGENY OF THE REGULATORY MECHANISM OF COMPLEMENT ACTIVATION

Although potent physiological activities of the complement system are crucial for effective pathogen elimination, they can potentially damage host cells in case of excessive or misdirected activation, resulting in an allergic or auto-immune pathology. Therefore, a complement system must be equipped with appropriate regulators to prevent such side effects. Controlling excessive complement activation on nonself targets and complement attack misdirected to the host cell needs both soluble and membrane-bound regulators of complement activation. In mammals, as soluble regulators, C1-inhibitor (C1INH) inactivates C1r/C1s and MASP, C4-binding protein (C4bp) and factor H (Hf) promote proteolytic degradation of C4b and C3b, respectively, and clusterin interferes with fluid-phase formation of the membrane-attack complex. Mammalian membrane-bound regulators are membrane-cofactor protein (MCP or CD46), CR1 (CD35), and decay-accelerating factor (DAF or CD55), which attenuate C3-convertase formation, and CD59, which inhibits MAC formation.

Recent phylogenetic analysis of C1INH has depicted that the presence of C1INH in bony fish, reptiles, birds, and mammals, but absence in lamprey and amphibians, based on genomic database survey, suggests that C1INH has occurred in a common ancestor of jawed vertebrates. C4bp, Hf, MCP, and CR1 act as cofactors of factor I protease (If) that catalyze degradation of C4b and C3b fragments into their inactive forms, whereas DAF enhances dissociation of the C3-convertases, C3bBb and C4bC2a. Factor I and its potential cofactors homologous to Hf have been identified from both jawed and jawless vertebrates. Therefore, proteolytic C3b-degradation dependent on Hf-like and If-like regulators seems to be an ancestral regulatory mechanism of C3 activation. Membrane-bound cofactors of If, which are CD46-like, have been identified from bony fish, amphibians, and mammals, suggesting its ancient origin, whereas the membrane-bound RCA from jawless species is yet to be identified. On the other hand, any RCA protein with DAF-like activity has not been found in the lower vertebrates, probably due to the lack of a functional assay method to evaluate the decay-acceleration of the C3-convertase complex.

In mammalian genomes, genes encoding RCA proteins are located in two tight clusters present in the same chromosome, group 1 cluster contains several factor H-like soluble proteins (Hf and Hf-related proteins), and group 2 cluster contains C4bp, DAF, CR2, CR1, and MCP-like soluble and membrane-bound proteins. Similar gene clusters have been reported from chickens, Xenopus, and zebrafish, indicating that the gene clustering has been well-conserved while varying the number of the members in the clusters.
It is interesting that zebrafish possess a group 1 RCA gene cluster with a comparable number of the Hf-like genes to that of humans, but with much fewer members of group 2 genes, despite a high degree of isotypic diversity of C3 and C4, which would be recognized by these RCA molecules. CD59 has been cloned from bony fish, hagfish, and even some invertebrate parasites such as *Schistosoma mansoni*, suggesting its “prevertebrate” ancient origin. However, the parasite CD59 showed no complement inhibition, and the hagfish complement even lacks lytic activity. There is also no direct evidence for CD59 to inhibit MAC formation in bony fish, which do possess the lytic pathway. A recent study reported that recombinant CD59 from zebrafish-bound bacteria, and their PAMPs such as LPS and LTA, showed a slight antibacterial activity. It is interesting data, but could be evaluated if such activities are immune-relevant in vivo.

6 EVOLUTIONARY SIGNIFICANCE OF MULTIPLE ISOFORMS IN TELEOST COMPLEMENT COMPONENTS

Bony fish is divided into two classes, Actinopteryii, or the ray-finned fishes, and Sarcopterygii, or the lobe-finned fishes, such as coelacanth. Teleost is a major infraclass of Actinopteryii, believed to have arisen in the Triassic period, and contains a very diverse extant species more than 26,000. An important phylogenetic feature of teleost is that a whole-genome duplication occurred in its stem lineage, as evidence by a number of researches. This duplication is believed to have provided genetic raw materials for further diversification, as seen in their morphology, physiology, and behavior.

A striking feature of the teleost complement system is an isotypic diversity of several components. Since multiple isoforms of rainbow trout C3 have been first described at the protein level, a number of studies, including protein isolation and molecular cloning, have uncovered the presence of multiple isoforms of complement components, receptors, and regulatory factors.

Whereas several isotypes are limited to pseudotetraploid species such as carp and trout, indicating that they arose from the lineage-specific tetraploidization, some duplicates such as those of C3, C4, and C7 are commonly seen in various species, regardless of the chromosome ploidy. Thus, the whole-genome duplication probably played a crucial role to increase the number of isoforms of these complement components, although it is difficult to trace the exact gene-duplication scenario to generate the isotypes.

From a functional point of view, diversity of C3 and C4 has drawn our best attention because the sequence divergence between the isoatypes includes a functionally important amino acid substitution at the position that determines binding specificity of these components. In most species other than teleost, C3 is encoded by a single-copy gene and has a His residue at about 100 residues C-terminal from the thioester site. Thus, we designate it His-type C3. This His residue plays a catalytic role in cleavage and covalent binding of the thioester, enabling nascent His-type
C3b to bind to hydroxyl-groups by transacylation on the target surface or to surrounding water.\textsuperscript{75} In vitro mutagenesis studies revealed that substitution of this catalytic His with other less nucleophilic residues, such as Ser and Asp, dramatically changes the binding specificity of the thioester toward amino-groups, which can directly cleave the thioester.\textsuperscript{75} Very interestingly, teleost species analyzed so far, such as trout, carp, zebrafish, and medaka fish, possess both the His-type C3 and C3 carrying a Ser, Gln, Asp, or Ala at the catalytic position, designated non-His type C3, suggesting that the two C3 types show different functions, such as binding specificities against targets. At the protein level, C3 isoforms purified from trout and sea bream have been reported to show different binding specificities to various complement-activating model targets, such as sheep and rabbit erythrocytes, bacteria, and the yeast cell-wall.\textsuperscript{76} We have tried to correlate the binding specificity of the thioester to hydroxyl- or amino-groups with binding spectra of His- and non-His-types of C3 against natural complement-activating targets, using monoclonal antibodies specific to His-type (C3-H1) and non-His-types (C3-S) of carp complement. The results showed that C3-S showed a wider range of binding spectrum than that of C3-H1, suggesting that non-His-type C3 unique to teleost play more prominent roles in complement activation and pathogen elimination (Table 6.2).

### Table 6.2 Distinct Binding Spectra of His-Type and Non-His-Type C3 Isotypes of Carp Complement Determined by ELISA or Flow Cytometer

<table>
<thead>
<tr>
<th>Targets</th>
<th>Assay</th>
<th>Binding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>C3-H1 (His-type)</td>
</tr>
<tr>
<td>Poly-L-lysine</td>
<td>ELISA\textsuperscript{a}</td>
<td>–</td>
</tr>
<tr>
<td>Glycogen (oyster)</td>
<td>ELISA</td>
<td>+</td>
</tr>
<tr>
<td>LPS (\textit{E. coli} O44:B5)\textsuperscript{b}</td>
<td>ELISA</td>
<td>+</td>
</tr>
<tr>
<td>LTA (\textit{S. aureus})\textsuperscript{c}</td>
<td>ELISA</td>
<td>–</td>
</tr>
<tr>
<td>IC\textsuperscript{d}</td>
<td>ELISA</td>
<td>+</td>
</tr>
<tr>
<td>Erythrocyte (rabbit)</td>
<td>FCM\textsuperscript{e}</td>
<td>+</td>
</tr>
<tr>
<td>Erythrocyte (sheep)</td>
<td>FCM</td>
<td>–</td>
</tr>
<tr>
<td>Baker’s yeast</td>
<td>FCM</td>
<td>+</td>
</tr>
<tr>
<td>\textit{Staphylococcus aureus}</td>
<td>FCM</td>
<td>–</td>
</tr>
<tr>
<td>\textit{Escherichia coli} DH5α</td>
<td>FCM</td>
<td>+</td>
</tr>
<tr>
<td>\textit{Aeromonas hydrophila}</td>
<td>FCM</td>
<td>+</td>
</tr>
<tr>
<td>\textit{Aeromonas salmonicida}</td>
<td>FCM</td>
<td>–</td>
</tr>
<tr>
<td>\textit{Edwardsiella tarda}</td>
<td>FCM</td>
<td>–</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Enzyme-linked immunosorbent assay.  
\textsuperscript{b}Lipopolysaccharide.  
\textsuperscript{c}Lipoteichoic acid.  
\textsuperscript{d}Immune complex of hen egg lysozyme and antilysozyme carp IgM.  
\textsuperscript{e}Flow cytometry.
However, their target-binding specificities do not clearly correlate with the preference of hydroxyl- or amino-group by carp C3-H1 and C3-S, respectively.\textsuperscript{77}

Other complement proteins also likely show some different functional activities between the isoforms, based on their amino-acid sequence diversity, especially at functionally important sites.\textsuperscript{42,73} More functional analyses are needed, to gain comprehensive and convincing conclusions on the biological impacts of diversification of complement-component isoforms and their significance to the evolution of a complement system.

\section*{7 CONCLUDING REMARKS}

It is reasonable to hypothesize that a prototype of the vertebrate complement system is well-represented by the system of cephalochordate and urochordate, that is, the invertebrate species of the chordate phylum, where MBL-like collectin, MASP, Bf, and C3 constitute an opsonic system with

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig6.3.png}
\caption{Hypothetical diversifications of MASP/C1r/C1s, Bf/C2, and C3/C4/C5 families driven by two rounds (1R and 2R) of WGD. Invertebrate chordates include urochordates (ascidian) and cephalochordates (amphioxus), and jawless vertebrates contain lamprey and hagfish. Genes in gray color are postulated to be lost in the extant species, but might be retained in a possible lineage ancestral to the derived class. C3 ($\alpha/\beta$) and C3 ($\alpha/\gamma/\beta$) represent C3 composed of $\alpha$- and $\beta$-chains and C3 having C4-like $\alpha$, $\gamma$, and $\beta$-chain structure, respectively. Evolutionary correspondences between MASP1/MAST2 genes of jawless species and MASP1/3, MASP2, and C1r/C1s of jawed vertebrates are unclear. Bf/C2-A and Bf/C2-B identified in cartilaginous and bony fish may correspond to Bf and C2, respectively, of higher vertebrates, but the assignments are not convincing enough on phylogenetic analyses based on the sequence data available so far. C4-1 lineage is found only in some cartilaginous and bony fish species, and might be lost in the descendent classes, whereas C4-2 seems to be maintained in all the lineages of jawed vertebrates.} 
\end{figure}
a pattern-recognition ability. An evolutionary scenario of the complement system, starting from the prototype to the system observed in jawed vertebrates, can be summarized as follows: (1) The complement system of jawless species remains opsonic with no cytolytic activity. An important development is recruitment of C1q, in addition to MBL, as a pattern-recognition molecule to activate C3. C1q may possibly connect VLR-dependent antigen recognition and the complement system in an analogous way to the classical pathway in jawed vertebrates.

(2) The classical and lytic pathways emerged, seemingly together, in a common ancestor of jawed vertebrates. C4, C2, and C1r/C1s in the classical pathway were most likely generated by gene duplication from C3, Bf, and MASP, respectively. For the establishment of the lytic pathway, more gene duplication events should have been needed, such as generations of C5 from C3, and that of C6, C7, C8, and C9 from their C9-like common ancestor, which is missing in the extant jawless vertebrates. Here, we attempt to fit this scenario to the 2R hypothesis of vertebrate evolution by postulating that some of the duplicated gene has been lost in the extant jawless species after the first round GWD, while they kept in a lineage directly ancestral to the jawed vertebrates until the second round GWD happened, as shown in Fig. 6.3. Overall, the 2R hypothesis reasonably explains the genomic mechanism to acquire the classical and lytic pathways in the vertebrate complement system.

ACKNOWLEDGMENTS

This work was supported in part by a grant in aid from Ministry of Education, Science, Culture and Sports, Japan (25292127 to MN).

REFERENCES


Page left intentionally blank
Antiviral Immunity: Origin and Evolution in Vertebrates

1 INTRODUCTION

Viruses are infectious agents of ancestral origin with particular characteristics that place them at the edge of living forms. They are intracellular pathogens that lack a cellular structure and need the host cellular resources for their replication. Despite the fact that viruses have no complete autonomy to exist, they can reproduce themselves through self-assembly and are subjected to natural selection. These pathogens can infect all types of cells throughout the living kingdoms, although most of them have a limited range of hosts and are species-specific. Consequently, cellular mechanisms aimed at blocking viral entry and replication have evolved in both unicellular organisms and complex multicellular species. The innate arm of the immune system is an ancient form of host defense against pathogens, with evolutionary conserved mechanisms in both invertebrate and vertebrate lineages. These mechanisms are triggered through a diversified system of receptors called pattern recognition receptors (PRRs) that recognize particular conserved microbial structures, including proteins, lipids, lipoproteins, glycans, and nucleic acids of viruses, bacteria, fungi, or parasites (pathogenic or nonpathogenic), generally referred to as pathogen-associated molecular patterns (PAMPs). PRRs, which are germline encoded, include large gene families consisting of closely related but polymorphic members, and in vertebrates include membrane-bound receptors such as Toll-like receptors (TLRs), and cytosolic sensors such as retinoic acid-inducible gene-I (RIG-I)-like receptors (RLRs) and nucleotide-binding oligomerisation domain (NOD)-like receptors (NLRs).1,2 Due to their critical role and their efficiency, these innate defense tools have been largely conserved through evolution, even if there is a significant degree of polymorphism and diversity intra- and inter-species, likely due to evolutionary pressures to match rapidly evolving ligands (pathogens). Viral nucleic acids,
including single stranded RNA (ssRNA), double stranded RNA (dsRNA), or DNA, are one of the common viral features suitable for innate immune recognition by PRRs. Upon PAMP recognition, PRRs activate a signaling cascade that leads to an upregulation of cytokine genes as messengers of an inflammatory response that mediates the recruitment of cells to the site of intrusion, activates antimicrobial effector systems, and stimulates adaptive immunity (in vertebrates), all of them ultimately focused on the clearance of the intruder. The interferon (IFN) system is the key cytokine network released after viral sensing. In fact, the name “interferon” was given by Isaacs and Lindenmann, based on the ability of these molecules to “interfere” with viral progression. IFNs are pleiotropic cytokines that function as key regulators of the antiviral response. Three groups of IFNs have been described in mammals, based on their biological and structural properties, including type I, type II, and type III IFN. Type I and type III IFNs are directly induced by virus infection, promoting the transcription of antiviral genes through a single pathway. The type II IFN, on the other hand, functions as a regulatory cytokine of innate and adaptive immunity through a slightly different pathway. Since type III IFN has only been recently identified in mammals and there is almost no information available in other vertebrate groups, in this review we will focus on type I IFN. Once an infected cell has secreted type I IFN in response to viral stimuli, this IFN signals in the producer and neighboring cells through type I IFN receptors that are present on nearly every nucleated cell. This signaling ultimately leads to the induction of IFN stimulated genes (ISGs) with different roles in a large number of immune pathways, many of which are capable of blocking intracellular viral replication at different levels.

In the current chapter, we provide an overview of IFN-mediated antiviral immunity in vertebrates, from early viral recognition to the antiviral activity of ISGs. The purpose of the chapter is to provide a general overview of this complex system, focusing on the most relevant functions and molecules, and specifically highlighting commonalities and unique features among vertebrate groups to provide an evolutionary perspective.

2 VIRUS-SENSING PATTERN RECOGNITION RECEPTORS

Virus detection can be accomplished either at the extracellular surface of cells or intracellularly, for viruses that have either been internalized by membrane-enclosed structures (ie, derived from phagocytosis of microbes or infected apoptotic cells) or for viruses that actively enter the cytoplasm. In general, the induction of antiviral immunity initiates with the recognition of viral nucleic acids by TLRs, RLRs, and NLRs, and/or recognition of viral glycoproteins by TLRs (Fig. 7.1). Signaling through these receptors leads to the activation of immune effector molecules, mainly the induction of the IFN system that mediates the primary early response. Although all nucleated cells express some kind of PRRs, innate immune cells such as dendritic cells, macrophages, and lymphocytes are specially equipped to sense pathogens and initiate an immune response.
2.1 Toll-Like Receptor Family

The first identified and best characterized PRR family is that of the membrane-bound TLRs. TLRs were first described in the fruit fly *Drosophila melanogaster* as receptors with a developmental function. They were later identified in humans and other species, and the nature of their immune-related function was subsequently revealed in both vertebrates and invertebrates. Similar structures are also found in bacteria and plants, revealing the ancient origin of these receptors. TLRs are type I transmembrane proteins that form homo- or heterodimers. Each TLR molecule consists of an N-terminal domain, responsible for the detection of PAMPs, containing numerous variable leucine-rich repeat (LRR) facing toward the extracellular space or the endosomal lumen; a transmembrane...
domain; and a cytosolic Toll-interleukin 1 (IL-1) receptor (TIR) domain that is involved in downstream signal transduction and the recruitment of adaptor molecules with TIR domains through homotypic TIR/TIR interactions.\textsuperscript{5,12} In contrast to the high variability in the LRR motifs, the cytoplasmic TIR domain is highly conserved among TLRs and across species.\textsuperscript{13} In both agnathans and gnathostomes, a TLR system has been characterized at the molecular level, and phylogenetic studies suggest a true orthology of this system among vertebrates.\textsuperscript{14,15} The TLR2 subfamily, TLR3, TLR5, TLR7/8, as well as the fish-type TLR21/22, are essentially conserved in the lamprey \textit{(Lethenteron japonicum)} and teleosts, suggesting that lampreys and jawed vertebrates share the same ancestral TLR repertoire composed of hybrid mammalian- and fish-type TLRs, which may represent the origin of the TLR repertoire in vertebrates.\textsuperscript{16,17} Because an extensive review of the evolution of the TLR system can be found in chapter: The Evolution of Toll-like receptor system in Teleosts, in this chapter we will focus exclusively on viral sensing TLRs.

In mammals, viral nucleic acid recognition is achieved by the endosomal TLRs, namely TLR3, 7, 8, and 9, whereas in fish extracellular TLR22\textsuperscript{18} or avian endosomal TLR21\textsuperscript{19} are also involved in viral recognition. Endosomal TLR3 is the only mammalian TLR that signals solely via the TIR domain-containing adaptor, inducing IFN-β (TRIF), whereas the other TLRs, including endosomal TLR7, 8, and 9, depend on myeloid differentiation factor-88 (MyD88) for signaling. Fish TLR22 recruits TRIF for signaling, whereas avian TLR21 signals via the MyD88 adaptor. Both signaling adaptors, TRIF or MyD88, subsequently activate two main pathways. One is the activation of the \( \text{IkB} \) kinase (IKK) complex that leads to nuclear translocation of the transcription factor nuclear factor (NF)-\( \kappa B \) to upregulate the expression of inflammatory cytokines and chemokines. Additionally, TLRs also activate the IFN regulatory factor (IRF) transcription factors 3/7 pathways that are crucial for the induction of type I IFNs\textsuperscript{2,20,21} (Fig. 7.1). Although TLR families show distinct features among vertebrates, several components of the downstream TLR signaling pathways have been identified across vertebrates and are structurally and functionally well-conserved.\textsuperscript{22,23} TLR3 specifically recognizes dsRNAs, such as the genomes of dsRNA viruses or dsRNAs generated as intermediate products during the replication of ssRNA viruses, or DNA viruses such as herpesviruses, as a by-product of symmetrical transcription.\textsuperscript{3} TLR7 and TLR8 are known to sense guanosine/uridine (GU)-rich ssRNA of RNA viruses. Endosomal TLR9 is a DNA-sensing TLR and specifically binds unmethylated cytosine-guanosine (CpG)-containing oligodeoxynucleotides (CpG-ODN) present in bacteria and DNA viruses, which are rare in the vertebrate genome.\textsuperscript{3}

True orthologues of mammalian TLR3 exist in lower vertebrates, and there is structural and functional evidence of a similar involvement in virus detection. For example, the organization of the exon/intron boundaries is highly conserved between the fish and human \textit{tlr3} gene, and fish TLR3 conserves all key residues for binding to dsRNA.\textsuperscript{24–26} Reports on the transcriptional upregulation
of tlr3 in response to different viral stimuli have provided indirect evidence of a conserved role for fish tlr3. These include evidence for an upregulation of tlr3 after poly I:C stimulation in lamprey or after exposure to the dsRNA virus in species such as grass carp (Ctenopharyngodon idella) and rare minnow (Gobio cyprisrarus) or to the ssRNA virus in species such as zebrafish (Danio rerio) and rainbow trout (Oncorhynchus mykiss). More direct evidence of an equivalent antiviral role includes the increased IFN response of a trout fibroblastic cell-line transfected with tlr3 following stimulation with poly I:C. A role of avian TLR3 in ssRNA virus-recognition has also been described during infection with a highly pathogenic avian influenza virus (HPAIV). All of these findings indicate that the common ancestor of mammals, teleosts, and agnathans likely contained a dsRNA recognition system involving TLR3 capable of detecting diverse viral families.

While TLR3 is an endosomal dsRNA receptor, fish TLR22 is a TLR3 analogue, sensing the presence of dsRNA outside the cells. TLR22 is widely conserved among teleosts and amphibians, but it is likely absent in avian or mammalian genomes, and it has been proposed as a TLR required for aquatic vertebrates that evolves under positive selection. Thus, teleosts possess two dsRNA-sensing TLRs, TLR3 at the endosomal compartment and TLR22 at the cell surface, with a functional difference regarding location and dsRNA discrimination by length. TLR3 recognizes preferentially short dsRNAs (<1 kbp), whereas TLR22 recognizes longer fragments (>1 kbp).

TLR7, 8, and 9 form an evolutionary cluster, present in both mammals and teleosts, all localized in intracellular vesicles such as endosomes and the endoplasmic reticulum (ER). TLR7 and 8 are present in lamprey and were mapped to the TLR7/8 cluster of jawed vertebrates. TLR7 and 8 sense viral-derived ssRNA, whereas TLR9 senses DNA CpG-ODN motifs. Interestingly, although a tlr7 ortholog has been found in chickens, chicken tlr8 is a pseudogene disrupted by several introns, and tlr9 seems to have been deleted from avian genomes over evolutionary time. Although the chicken genome does not contain tlr9, avian TLR21 has an immune function similar to mammalian TLR9 as a CpG-ODN receptor. This nonmammalian receptor is an endosomal TLR and is also present in Xenopus tropicalis, teleost fish, and lamprey. Phylogenetic analysis indicates that TLR21 is a member of the TLR11 subfamily, together with mouse TLR12 and 13 and teleost TLR20. Although avian TLR21 is known to be involved in CpG-ODN sensing, the functionality of fish TLR21 remains to be determined. Furthermore, a number of fish TLR genes may have experienced insertion of additional introns after the divergence from tetrapods. Among these, tlr9 genes have different numbers of exons in fish species than in mammals.

Mammalian TLR2 and TLR4, present on the cellular plasma membrane, have also been shown to recognize viral components, including envelope proteins and surface hemagglutinin proteins, such as the hemagglutinin of varicella virus, herpes simplex virus-1, human cytomegalovirus, and mouse mammary
tumor virus. Little information is available today on the recognition of viral pathogens by TLR2 or TLR4 in other vertebrates. TLR4 is thought to be absent in several fish, as in the salmonids, but has been identified in cyprinid species and catfish (*Ictalurus punctatus*). An induction of *tlr4* has been reported after viral infection in cyprinids, but the fact that the two coreceptors needed by mammalian TLR4 to recognize viral structures, namely myeloid differentiation protein-2 (MD-2) and CD14, are absent from the fish genomes means that the function of fish TLR4 signaling is still to be characterized. On the other hand, the G proteins of novirhabdoviruses are known to trigger the synthesis of IFN in rainbow-trout cells, suggesting that the fish immune system conserves the capacity to recognize viral glycoproteins.

### 2.2 RIG-Like Receptor Family

The other main group of PRRs acting in viral RNA detection are the RLRs, which are Asp-Glu-x-Asp/His (DExD/H)-box helicases. RLRs are mobilized to sense viral RNA species in the cytosol during intracellular viral invasion and replication, leading to a potent induction of type I IFNs in most cell types (in contrast to the more immune cell-restricted expression pattern of TLRs). The RLR family is expressed ubiquitously in the cytoplasm and belongs to the IFN-stimulated gene (ISG) family, and, consequently, is subject to positive feedback regulation by IFN. It consists of three members: RIG-I, melanoma differentiation-associated gene 5 (MDA5), and laboratory of genetics and physiology 2 (LGP2).

RIG-I and MDA5 share a similar domain structure, composed of two N-terminal caspase activation and recruitment domains (CARDs) required for downstream signaling; a central DExD/H-box RNA helicase domain with the capacity to hydrolyze ATP; and a C-terminal domain (CTD). Once activated, RIG-I and MDA5 are subsequently recruited to the mitochondrial antiviral signaling (MAVS) adaptor, followed by homotypic CARD-CARD interaction with the adaptor and initiation of downstream signaling via MAVS oligomerization, leading to mobilization of NF-kB and IRFs for cytokine gene-induction (IFN and pro-inflammatory cytokines) (Fig. 7.1). In contrast to TLRs that are restricted to the endosomal compartment, RLRs are in close contact with host RNAs in the cytoplasm, thus suggesting the existence of a precise mechanism for the discrimination of self versus nonself. Indeed, RIG-I is activated by short, uncapped 5′-triphosphate (5′ppp) ssRNA motifs juxtaposed to a short region of dsRNA, generated by viral RNA polymerases. In contrast, eukaryotic cells do not produce dsRNA, and mRNA transcripts are blocked at their 5′-ends by a 7-methyl-guanosine group. Interestingly, several viruses have evolved a mechanism to remove their own 5′ppp groups to avoid detection by RIG-I. Although both RIG-I and MDA5 are able to sense dsRNA, RIG-I recognizes short dsRNA, whereas MDA5 discriminates long dsRNA, generated during the course of virus infection. Consequently, MDA5, but not RIG-I, responds to
poly I:C stimulation, leading to IFN production. Furthermore, RIG-I can also detect DNA from either bacteria or DNA viruses through the DNA-dependent RNA polymerase III.44

The third member of the RIG-like receptor family, LGP2, lacks the signaling CARDs. Since embryonic fibroblasts from mice, with a targeted deletion of LGP2, produce increased amounts of IFN-β and active NF-κB in response to transfected poly-I:C relative to wild-type cells, LGP2 has been proposed as a negative regulator of RIG-I signaling.45 However, more recent studies have demonstrated that LGP2 is a positive upstream regulator of RIG-I and MDA5 signaling, facilitating viral RNA recognition by RIG-I and MDA5 through its ATPase domain.46

Using a combination of gene modeling, phylogenetics, structural analysis, population genetics, and molecular-functional characterization of ancestral and extant RLRs, Mukherjee and coworkers found that the RLR-based system had arisen with the emergence of multicellularity.47 Early functional differentiation of RLRs occurred in metazoan evolution through a series of gene duplication events followed by modifications of the RNA-binding domain, suggesting a course of multiple adaptive changes in sensitivity and specificity of these receptors throughout metazoan evolution, in parallel with viral evolution.47 MDA5 and LGP2 appear to be common to all fish genomes, whereas RIG-I had been lost from some fish species that belong to the Acanthopterygii class.48 To date, RIG-I orthologs have been found in primitive fish species such as those in the classes Ostariophysii, Protacanthopterygii, and Paracanthopterygii, as well as in Agnatha (lamprey) and in the Sarcopterygii (coelacanth), suggesting that RIG-I could have been lost after divergence of the Acanthopterygii from the Paracanthopterygii.14,48,49 Similarly, the chicken genome, unlike that of ducks, geese, and finches, also lacks rig-I but contains mda5 and lgp2, which functionally compensate for the absence of RIG-I. For example, while duck RIG-I is the cytosolic recognition element for HPAIV, chicken cells sense the virus through MDA5.19

Several studies have reported the induction of RLR mRNA expression in different fish tissues after viral infection, IFN treatment, or poly I:C treatment, as observed in mammals.14,50 Concerning specific antiviral effects, overexpression of trout MDA5 and LGP2 resulted in a significant accumulation of Myxovirus (Mx)-resistance transcripts, which correlated with an enhancement of protection against viral infection.14,50 Likewise, a recent study performed in Japanese flounder also revealed that MDA5 knockdown reduced the antiviral and immune effects of poly I:C.51

2.3 NOD-Like Receptor Family

TLRs and RLRs are important for the production of type I IFNs and several other cytokines, whereas NLRs play a major role in the regulation of IL-1β (Fig. 7.1). NLRs primarily recognize microbial molecules of bacterial origin.52
However, in mammals, NLRs sense several different RNA or DNA viruses, and either positively or negatively regulate the innate antiviral response. The human NLR family is composed of 23 members, whereas there are at least 34 members in the mouse NLR family. Several NLR family proteins regulate innate immunity by forming large molecular complexes called inflammasomes, which lead to the secretion of IL-1β and IL-18. Both cytokines activate NF-κB via binding their cognate receptors, leading to a caspase-dependent inflammatory cell-death program, known as pyroptosis. NLRs are composed of three major domains: C-terminal leucine-rich repeats (LRR) responsible for microbe recognition, a central nucleotide-binding oligomerization (NOD) domain (also known as the NACHT cassette) that mediates self-oligomerization during activation, and a variable N-terminal protein–protein interaction domain involved in signal transduction. Based on this N-terminal effector domain, the NLRs are subdivided into four subfamilies that include: the NLRA subfamily, containing an acidic transactivating domain; the NLRB subfamily, containing a baculovirus inhibitor of apoptosis protein repeat (BIR) domain; the NLRC subfamily, containing a caspase recruitment domain (CARD); and the NLRP subfamily, containing a pyrin domain (PYD). According to phylogenetic relationships, NLRs can also be categorized into three subfamilies named NODs, NLRPs (or NALPs), and IPAF.

Complex NLR repertoires evolved very early in animal evolution, and NLR genes with a putative immune function are present already in basal metazoans. Interestingly, three distinct NLR subfamilies were identified in the genome of various nonmammalian vertebrates, including fish, amphibians, and birds. The first subfamily (NLR-A) resembles mammalian NODs, the second (NLR-B) resembles mammalian NALPs, whereas the third (NLR-C) appears to be unique to teleost fish. In zebrafish, several hundred genes belong to this unique subfamily, which is thought to have evolved from a NOD3-like molecule. However, the role for these avian, amphibian, and fish NOD-like receptors in antiviral immunity is still largely unknown.

Other recently discovered IFN-inducible proteins involved in the mammalian antiviral immune system, such as the cytoplasmic dsDNA sensors DAI, AIM2, and IFI6-16, have not yet been investigated in other vertebrates. AIM2 and IFI6-16 form a new family termed AIM2-like receptors (ALRs), and are both involved in inflammasome formation and apoptosis. Future investigation is needed to provide clues as to the origin and evolution of these proteins.

### 3 RNA INTERFERENCE

RNA interference (RNAi) is a defense mechanism operating against RNA viruses, discovered recently, and identified as the primary and most robust antiviral response in plants and invertebrates. The mechanism consists of the cleavage of viral dsRNAs into small (20 bp) RNA fragments by an RNAase III nuclease (dicer). The small interfering RNAs (siRNAs) are then recruited into
a biochemical pathway, the RNA-induced silencing complex (RISC), which unwinds the dsRNA into two ssRNA strands to be silenced and degraded. 62–65 Although it was believed at first that this RNA-based antiviral RNAi pathway had been supplanted in vertebrates with a protein-based antiviral IFN response, recent findings demonstrated the existence of a functional RNAi pathway in mammals, suggesting that both RNA and protein-based antiviral mechanisms operate simultaneously in mammalian cells. 66,67 siRNAs have been shown to inhibit viral replication in fish cells, 68,69 however, the relevance of this mechanism in the overall vertebrate antiviral response still needs to be determined.

4 TYPE I IFNs

4.1 Mammalian IFNs

The *ifn* genes have been well-characterized in mammals. Mammalian type I IFNs are encoded by genes with no introns, and can be divided into several phylogenetic groups, 70,71 consisting of IFN-α, -β, -ω/τ, -δ/limitin, -ε, and -κ (Table 7.1). They have a similar α-helical structure and all of them belong to the IL-10 cytokine family. Multiple IFN-α subtypes sharing high sequence homology are apparent within a single species, while IFN-β, -ω/τ, -ε, and -κ are encoded by a single copy gene. All the type I IFNs in mammals bind to the same receptor complex and exert antiviral functions via the JAK/STAT signaling pathway. The genes encoding type I IFNs are clustered in the genome; in chromosome (chro) 9 of humans and chro_4 in mouse. Interestingly, some IFN subgroups are lineage specific, and, for example, IFN-δ and limitin have been described in some species but not in humans. 72,73 On the other hand, at least 10 copies of *ifn*-δ genes have been reported in porcine species. 73 IFN-α is predominantly expressed by activated dendritic cells and primarily participates in immune regulation by bridging innate and adaptive immunity. Conversely, IFN-β is ubiquitously produced by viral-infected cells and is the key player in promoting an antiviral state in the host. IFN-ε is constitutively produced by cells of the reproductive organs and protects the host from viral and bacterial infection. 74,75 Consequently, IFN-ε deficient mice develop severe epidermal lesions in the reproductive tissue, accompanied by high viral titer after a herpes simplex virus 2 challenge. 74 IFN-τ is limited to ruminant species, including sheep and cattle, 76 where multiple *ifn*-τ genes are found to be clustered with other type I *ifn* genes in chro_2 in sheep and in chro_8 in cattle. 77,78 Although a putative homologue has also been reported in humans, 79 it could not be confirmed in the sequenced genome by other studies. IFN-τ is closely related to IFN-ω 70 and is transiently expressed by the conceptus during early pregnancy. It is suggested to play an important role for the fetus maintenance and development, in addition to its primary antiviral function. 80,81 IFN-κ is produced by keratinocytes and certain lymphocyte subsets, and has been shown to modulate cytokine expression. 82,83
### TABLE 7.1 Vertebrate IFNs and IFN Receptors

<table>
<thead>
<tr>
<th>Vertebrates</th>
<th>IFN</th>
<th>Exon/Intron</th>
<th>Disulfide bond</th>
<th>Expression</th>
<th>Receptor 1</th>
<th>Receptor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mammals</td>
<td>IFN-α</td>
<td>1/0</td>
<td>2</td>
<td>Leukocytes, dendritic cells</td>
<td>IFNAR1</td>
<td>IFNAR2</td>
</tr>
<tr>
<td></td>
<td>IFN-β</td>
<td>1/0</td>
<td>1</td>
<td>Ubiquitous expression</td>
<td>IFNAR1</td>
<td>IFNAR2</td>
</tr>
<tr>
<td></td>
<td>IFN-δ</td>
<td>1/0</td>
<td>2</td>
<td>Reproductive and conceptus tissues</td>
<td>IFNAR1</td>
<td>IFNAR2</td>
</tr>
<tr>
<td></td>
<td>IFN-ε</td>
<td>1/0</td>
<td>1</td>
<td>Reproductive and mucosal tissues</td>
<td>IFNAR1</td>
<td>IFNAR2</td>
</tr>
<tr>
<td></td>
<td>IFN-κ</td>
<td>1/0</td>
<td>2</td>
<td>Keratinocytes, lymphocytes</td>
<td>IFNAR1</td>
<td>IFNAR2</td>
</tr>
<tr>
<td></td>
<td>IFN-ω/τ</td>
<td>1/0</td>
<td>2</td>
<td>Conceptus tissues, leukocytes</td>
<td>IFNAR1</td>
<td>IFNAR2</td>
</tr>
<tr>
<td>Birds</td>
<td>IFN-1 (IFN-α)</td>
<td>1/0</td>
<td>2</td>
<td>Lymphoid tissues, leukocytes</td>
<td>IFNAR1</td>
<td>IFNAR2</td>
</tr>
<tr>
<td></td>
<td>IFN-2 (IFN-β)</td>
<td>1/0</td>
<td>2</td>
<td>Ubiquitous expression</td>
<td>IFNAR1</td>
<td>IFNAR2</td>
</tr>
<tr>
<td>Reptiles</td>
<td>IFN</td>
<td>1/0</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Amphibians</td>
<td>IFN-1-5</td>
<td>5/4</td>
<td>2</td>
<td>Most tissues, lungs, bone marrow, kidney cells, splenocytes</td>
<td>IFNAR1</td>
<td>IFNAR2</td>
</tr>
<tr>
<td></td>
<td>IFN</td>
<td>Enhancements</td>
<td>Receptors</td>
<td>Expression</td>
<td>Expression of</td>
<td>Receptors</td>
</tr>
<tr>
<td>--------------</td>
<td>-------</td>
<td>--------------</td>
<td>-----------</td>
<td>------------</td>
<td>---------------</td>
<td>------------</td>
</tr>
<tr>
<td>Coelacanth</td>
<td>IFN-1-5</td>
<td>5/4</td>
<td>2</td>
<td>NA</td>
<td>IFNAR1</td>
<td>IFNAR2</td>
</tr>
<tr>
<td>Teleost fish</td>
<td>IFN-a</td>
<td>5/4</td>
<td>1</td>
<td>Ubiquitous</td>
<td>CRFB5, CRFB5a-c</td>
<td>CRFB1, CRFB1a</td>
</tr>
<tr>
<td></td>
<td>IFN-b</td>
<td>5/4</td>
<td>2</td>
<td>Lymphoid tissues, leukocytes</td>
<td>CRFB5x</td>
<td>CRFB2</td>
</tr>
<tr>
<td></td>
<td>IFN-c</td>
<td>5/4</td>
<td>2</td>
<td>Lymphoid tissues, leukocytes</td>
<td>CRFB5, CRFB5a, CRFB5c</td>
<td>CRFB2</td>
</tr>
<tr>
<td></td>
<td>IFN-d</td>
<td>5/4</td>
<td>1</td>
<td>Ubiquitous expression</td>
<td>CRFB5</td>
<td>CRFB1</td>
</tr>
<tr>
<td></td>
<td>IFN-e</td>
<td>5/4</td>
<td>1</td>
<td>Kidney, spleen</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>IFN-f</td>
<td>5/4</td>
<td>2</td>
<td>Kidney, spleen, fibroblasts, monocytes/macrophages</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Cartilaginous fish</td>
<td>IFN</td>
<td>4/3</td>
<td>2</td>
<td>NA</td>
<td>IFNAR1</td>
<td>IFNAR2a, IFNAR2b</td>
</tr>
</tbody>
</table>

NA, information not available. See references in the text.
4.2 Avian and Reptile Type I IFNs

Avian type I IFNs have been studied mostly in chickens, where two distinct groups of type I IFNs with limited sequence identity (57%) have been found (Table 7.1), designated ChIFN1 and ChIFN2. They have also been termed as IFN-α and IFN-β, mainly based on the comparison of gene copies and expression responsiveness to immune stimulants between mammalian and chicken type I IFN members. Despite this, phylogenetic analysis reveals that ChIFN1 and ChIFN2 are not 1:1 orthologues to mammalian IFN-α and IFN-β. ChIFN1 exists as multiple copies sharing high sequence identities. The genes encoding these IFN molecules are located in the short arm of the sex chromosome, which also harbors the single copy of ChIFN2. Both ChIFN1 and ChIFN2 contain 4 cysteines in the mature peptide, a characteristic pattern seen in mammalian IFN-αs and fish group-II IFNs.

ChIFN1 and ChIFN2 are induced in response to viral infection and stimulation with viral mimics. Potent inducers include Newcastle disease virus and influenza virus. Curiously, the expression patterns of the ChIFN1 and ChIFN2 resemble that of mammalian IFN-α and IFN-β. The ChIFN1 is predominantly produced in lymphoid organs after administration of imidanoquinolines, but only the ChIFN2 and not ChIFN1 can be activated by nonviral nucleotide PAMPs. Analysis of the promoter region of the chifn2 gene revealed the presence of NF-kB binding sites, partly responsible for the observed difference in gene expression.

Reptilian IFN activities were detected three decades ago, exhibiting similar chemical and antiviral properties to their avian and mammalian counterparts. However, the gene and protein sequences were only confirmed recently, when multiple genes (GenBank accession number: XM_008103493-9) were predicted in the genome of the green anole (*Anolis carolinensis*). Like birds and mammals, the green anole *ifn* genes lack introns (Ref. [87], unpublished data).

4.3 Amphibian IFNs

Type I *ifn* genes were first reported in the *X. tropicalis* genome, where five were predicted in the same scaffold, all consisting of five exons and four introns (Table 7.1). The translated proteins possess four conserved cysteine residues in the mature peptides and belong to the vertebrate four cysteine containing IFN group (group II). Phylogenetic tree analysis reveals two apparent groups where the IFN1 and IFN2 are related to the teleost IFNs and the IFN3–5 are related to the amniote IFNs, which are believed to have diversified from a bona-fide gene shared by ray-finned fish and tetrapods. The two groups share low sequence-identities at the protein level, ranging from 15.5% to 23.7%. A homologue to the *X. tropicalis* IFN3–5 has also been sequenced in another amphibian species, *X. laevis*, and has been shown to trigger a conserved antiviral response similar to that observed in fish and amniote IFNs, protecting cells and tadpoles against ranavirus infection.
4.4 Fish IFNs

Fish \textit{ifn} genes were first identified in zebrafish,\textsuperscript{95} Atlantic salmon (\textit{Salmo salar}),\textsuperscript{96} and channel catfish.\textsuperscript{97} Later on, taking advantage of the completion of fish genomes and the vast amount of transcriptomic sequences available, significant advances have been made in discovery of \textit{ifn} genes in other fish species. At the molecular level, the teleost type I \textit{ifn} genes share the same genetic organization as their amphibian and sarcopterygian (coelacanth) counterparts, and are encoded by genes with a genomic structure of 5 exons and 4 introns.\textsuperscript{93,98,99} Multiple copies are common, and, as in other vertebrates, are clustered in the genome. However, accumulating data suggest that the IFN locus may be segregated in teleosts, due to further whole genome duplications (WGD) after the second round of WGD leading to the appearance of jawed vertebrates. In the zebrafish genome, four \textit{ifn} genes have been found, with three residing in chro_3 and one in chro_12. Interestingly, the IFN locus in chro_3 is linked with the growth hormone gene 1, a synteny also seen in the genome of elephant shark (\textit{Callorhinchus milii}),\textsuperscript{100} Atlantic salmon,\textsuperscript{101} and Chinook salmon (\textit{Oncorhynchus tshawytscha}).\textsuperscript{102} On the other hand, several other genes, including CD79B and SCN4A, seen in the elephant shark IFN locus are also present in the IFN locus in zebrafish chro_12. It is probable that the IFN locus has been segregated into two in teleosts by the third-round WGD, which is believed to have occurred before the radiation of teleosts.

The teleost \textit{ifn} genes are divergent in their primary sequences, and can be classified into two groups that contain two (group I) and four (group II) conserved cysteine residues in the mature peptide, and can be further divided into six phylogenetic subfamilies (IFN-a-f). In the West Indian Ocean coelacanth (\textit{Latimeria chalumnae}), one of the two extant species of the oldest representatives of sarcopterygian lineage, all of the five IFNs identified belong to group II IFNs, and have a closer phylogenetic relationship with their counterparts from tetrapods than from teleosts.\textsuperscript{103} Fish IFNs do not have 1:1 orthologous relationships with their counterparts in birds and mammals. Group I IFNs, consisting of IFN-a, -d and -e, are ubiquitously present in teleosts and are not the equivalent to the mammalian two-cysteine-containing IFN-\(\beta\) and -\(\epsilon\),\textsuperscript{104} because the cysteines are paired at different positions. Group II IFNs (IFN-b, -c, and -f) are limited to some species, as of yet the cyprinids and salmonids, and contain a relatively well-conserved motif (CAWE) near the C-terminus.\textsuperscript{87,101,104} It should be noted that the two-cysteine-containing IFNs have only been found in teleosts and mammals, whereas the four cysteine containing IFNs are present in all the jawed-vertebrate lineages.\textsuperscript{99} Despite the limited sequence homology between fish group I and II IFNs, their overall protein structure remains similar. Hamming et al.\textsuperscript{105} recently determined the crystal structure of zebrafish IFN-\(\text{phi1}\) (IFN-a, group I) and IFN\(\text{phi2}\) (IFN-c, group II), both of which display a typical structure of class II cytokines containing six \(\alpha\)-helices. This study also provides structural evidence confirming the homologous relationship of fish IFNs to type I rather than type III IFNs in mammals.
The copy numbers of fish IFNs vary significantly among species. Salmonids have all six phylogenetic subfamilies and the largest number of gene copies in the genome. Thus in rainbow trout, at least 26 genes/pseudogenes have been identified,\textsuperscript{104} whereas cyprinids have only three subgroups (IFN-a, -d and -c).\textsuperscript{106} Multiple copies have also been found in some acanthopterygian species, and, in this case, all belong to one subfamily (IFN-d).\textsuperscript{104,107–109}

Fish IFNs are produced by different cells in response to viral stimuli, suggesting distinct roles in eliciting immune responses to fend off invading pathogens during infection. The \textit{ifn-a} genes are widely induced in different cell types, including fibroblasts, monocytes, macrophages, and lymphocytes, resembling the expression patterns of IFN-\textit{b} in humans.\textsuperscript{87,110} On the other hand, IFN-b and -c are shown to be produced primarily by limited leucocyte populations, and can be drastically upregulated in response to PAMP stimulation and viral infection.\textsuperscript{97,110} There are contradicting reports regarding whether IFN-d is inducible by virus or by viral PAMPs in salmonids.\textsuperscript{106,110,111} Nevertheless, acanthopterygian IFN-d is activated after viral infection.\textsuperscript{108} Information regarding the expression of IFN-e and -f is scarce because they have only recently been described in salmonid species. Although inducible, IFN-e and -f are not upregulated to the magnitude seen for IFN-a, -b and -c during viral infection in brown trout (\textit{Salmo trutta}).\textsuperscript{104} All of these data demonstrate that the fish IFN response is likely to be modulated at multiple levels and may differ from species to species.

The direct antiviral activities of type I IFNs appear to be well-conserved in vertebrates. The recombinant mature peptides produced in either bacteria or mammalian cells are potent inducers of IFN-stimulated antiviral genes, including Mx, PKR/PKZ, and viperin in fish cells, and protect fish against viral infection when administered in vivo.\textsuperscript{87,95,106,112,113} These antiviral activities of fish IFNs are in line with the observations in birds and mammals, but with some exceptions. For example, administration of recombinant IFN-a at the time of infection or postinfection does not inhibit virus replication in Atlantic salmon cells.\textsuperscript{112} Furthermore, recombinant IFN-d from salmonids and zebrafish weakly enhances Mx/viperin gene expression, even though it is induced by both virus and viral PAMPs,\textsuperscript{106,110,111} suggesting it may not be directly involved in activation of the cellular antiviral response. However, this observation is in contrast to other studies, where the IFN-d molecules from some acanthopterygian species are shown to trigger antiviral gene expression and protect the host against viral infection.\textsuperscript{99,114}

IFNs are produced by viral infected cells, and activate the target cells to establish an antiviral state via cell-surface receptors in an autocrine or paracrine manner.\textsuperscript{70} The majority of fish IFNs have a predicted hydrophobic leader peptide at the N-terminus, suggesting that their mode of action resembles that of IFNs in mammals.\textsuperscript{98,99} However, several groups have described transcript variants generated by RNA alternative splicing in some species, including Atlantic salmon,\textsuperscript{115} catfish\textsuperscript{97} and rainbow trout,\textsuperscript{116} indicating that this might represent a common mechanism to increase the diversity of IFN function in teleosts. A later
study has confirmed that some IFN transcript variants translate into functional proteins that are not secreted by the cells, and interact with intracellular receptors to induce expression of antiviral genes.\textsuperscript{117} This intracellular IFN system has only been reported in rainbow trout and may be important for the defense of virus-infected cells at the very early stage. It would be interesting to see whether this is a common IFN feature in other fish species.

5 IFN RECEPTORS

In humans, a sole heterodimeric receptor complex consisting of IFNAR1 and IFNAR2 mediates the cellular responses triggered by all type I IFN family members, despite their low sequence homology.\textsuperscript{70} Upon activation by the IFN ligands, IFNAR1 and IFNAR2 undergo a conformational change to facilitate a cascade of downstream signaling events leading to phosphorylation of the tyrosine kinases Tyk2 and JAK1, as well as the signal transducers and activators of transcription STAT1 and STAT2, with the subsequent formation of the IFN-stimulated gene factor 3 (ISGF3) complex, consisting of phosphorylated STAT1 and STAT2 and IFN-regulated factor 9 (IRF9). The ISGF3 complex translocates to the nucleus and binds to a specific sequence motif in the promoter region of target genes to initiate expression. To activate phosphorylation of cellular proteins, IFNs bind to the high-affinity receptor IFNAR2 and are then transported to IFNAR1, which has a weak ligand-binding affinity.\textsuperscript{118} However, this model has been challenged by a recent study, where binding of human IFN-\(\beta\) to the IFNAR1 in the absence of IFNAR2 can also augment gene expression in target cells via a signaling pathway independent from the conventional JAK/STAT pathway.\textsuperscript{119} Interestingly, Chang et al. also demonstrated that cotransfection of trout intracellular IFN-\(\alpha\) with IFNAR1, but not IFNAR2 in human HEK293 cells, resulted in phosphorylation of STAT1.\textsuperscript{117}

The genes encoding IFNAR1 and IFNAR2 are clustered in a single chromosome that harbors IL-10R2 and IFN\(\gamma\)R2 in birds and mammals. The corresponding locus is also identifiable in elephant shark\textsuperscript{100} and coelacanth,\textsuperscript{103} and in some teleost species such as pufferfish, zebrafish, and Atlantic salmon.\textsuperscript{107,120,121} However, recent studies have shown that the respective locus has been duplicated in several cyprinid and salmonid species. In zebrafish, IFNAR1, IFNAR2, and IL-10R2 are placed in chro_9, whereas IFN\(\gamma\)R2 is in chro_5.\textsuperscript{121} In contrast, four IFNAR1 genes and five IFNAR2 genes are found in two loci in chro_21 and chro_25 in the Atlantic salmon genome.\textsuperscript{121} It must be noted that the IFNAR2 gene has expanded in all the teleost species as a result of random local-gene duplications or WGD duplications. Considering that the elephant shark genome also has duplicated copies of IFNAR2 homologues, it is difficult to ascertain whether the IFNAR2 genes have been duplicated prior to gnathostomes, or have experienced gene loss in the coelacanth/tetrapod lineage. In humans and mice, multiple forms of IFNAR2 also exist, but arise from a single gene by alternative splicing.\textsuperscript{122,123}
Both IFNAR1 and IFNAR2 have been characterised functionally in teleost fish, including Atlantic salmon, zebrfish, and rainbow trout. In contrast with the fact that all type I IFNs interact with a common receptor complex in mammals, emerging evidence suggests that teleost group I and II IFNs have biased receptor usage. In zebrafish, two receptor complexes, involving IFNAR1(CRFB5)/IFNAR2a(CRFB1) and IFNAR1(CRFB5)/IFNAR2b(CRFB2), were proposed (Table 7.1). The former is believed to be preferentially activated by group I IFNs, while the latter by group II IFNs. In Atlantic salmon, the receptor usage seems likely to be even more complex, as there are five IFNAR2s (CRFB1–3) and four IFNAR1s (CRFB5a–c and CRFB5x).

Studies in mammals demonstrate that the fibronectin (FN)-like domains in the extracellular region of IFNAR1 and IFNAR2 are indispensable for the ligand and receptor interaction, and dictate the differential roles elicited by the ligands and the species specificity seen. Human IFNAR1 comprises four tandemly assembled FNIII-like subdomains (SD1–SD4), of which SD1–SD3 have been shown to be important for IFN binding. The four SD domain arrangement is also present in known IFNAR1 proteins sequenced in sarcopterygians. However, teleost IFNAR1 proteins possess only two extracellular SD domains, suggesting that the interaction between IFN ligands and receptors may differ from that in humans. The two tandem SD structure is also present in the recently identified IFNAR1 of elephant shark. These findings support the notion that the four SDs seen in sarcopterygian IFNAR1 have been duplicated in the ancestral molecule shared by coelacanths and tetrapods. In the case of IFNAR2, the two extracellular FNIII-like domains (D1 and D2) are conserved across the Vertebrata phyla, with D1 being the main region interacting with the IFN ligands in humans.

6 DOWNSTREAM MODULATORS OF THE INTERFERON RESPONSE

IFN signaling activates a cascade of intracellular events that leads to the formation of a multimeric complex STAT1/STAT2/IRF9, initially designated as ISGF3, that translocates into the nucleus and binds specific motifs located in IFN promoters, as well as promoters of ISGs. The elements involved in this cascade, including stat1, stat2, and irf9 genes, have been reported in many fish species (reviewed in Ref. [130]), suggesting a conservation of the intracellular IFN signaling pathway in all vertebrates. Many of these ISGs (between 50 and 100) are believed to be directly responsible for the antiviral effects of IFNs, but only a few of them have been fully characterized in mammals. These include Myxovirus-resistance (Mx) proteins, 2,5-oligoadenylate synthase (OAS), protein kinase R (PKR), viperin, and the ubiquitin-like Interferon Stimulated Gene 15 kDa protein (ISG15). Although the number of ISGs seems to be lower in fish, and also some ISGs such as OAS are missing...
from fish genomes, it seems that some essential ISGs appeared early in vertebrate history. Actually, Langevin et al. identified two different groups of ISGs: a family that differentiated in parallel in tetrapods and fish, which includes genes such as Mx; and a family of genes that had already diversified in the common ancestor to tetrapods and fish, which mainly includes regulatory factors and signal-transduction components such as the STATs.

### 6.1 Mx Protein

Mx proteins are IFN-induced, dynamin-like, large GTPases, first identified in association with influenza virus-resistance in mice. Since then, Mx homologs have been identified in most vertebrates examined, and even in invertebrates such as the abalone (*Haliotis discus discus*); however, their cellular location, their level of activity, and the viruses they inhibit vary widely. Mx genes have been reported in amphibians and reptiles, although little information regarding their functionality is available in these species. While in most species at least two Mx isoforms are present, in birds only one cytoplasmic Mx isoform exists. Furthermore, initial descriptions of duck and chicken Mx proteins reported a lack of antiviral activity. However, this issue is controversial since later reports suggested that certain polymorphic variants of chicken Mx were able to confer some antiviral activity, whereas other studies with these same variants found no significant differences in viral resistance. Furthermore, recent analyses have confirmed that some of these chicken Mx variants with suspected antiviral activity lacked GTPAse activity, which is an essential requirement for antiviral activity. However, the high degree of polymorphisms in this gene observed in birds does suggest a role in antiviral immunity that is yet to be clarified.

In fish, *mx* genes have been identified in many different species, in which usually two or three isoforms with different subcellular locations are present. Because in most fish species Mx genes were identified much earlier than IFN genes, the induction of Mx transcription has been commonly used as a marker of IFN induction. Consequently, fish Mx proteins are known to be transcriptionally modulated in response to treatment with viral mimics such as poly I:C, diverse viral pathogens, or antiviral vaccines, including DNA vaccines, virus-like particles (VLPs), or inactivated viruses, demonstrating that viral replication is not required for induction of Mx. In the case of intramuscular DNA vaccination against fish rhabdovirus, Mx transcription levels induced in trout seemed to correlate with protection levels observed at early times postvaccination until adaptive immunity takes over. Despite this, no antiviral activity has ever been reported for the trout Mx isoforms, and the bioactivity of Mx proteins in fish has only been demonstrated in a few species. In sea bream (*Sparus aurata*), a marine species naturally resistant to many viral pathogens, all three Mx isoforms showed antiviral activity against several RNA and DNA viruses. Antiviral activity
has also been reported for Mx proteins in species such as Atlantic salmon,\textsuperscript{171,172} rare minnow,\textsuperscript{173} grass carp,\textsuperscript{174} and Senegalese sole (\textit{Solea senegalensis}).\textsuperscript{175} In barramundi (\textit{Lates calcarifer}), Mx was shown to interact with the betanodavirus RNA-dependent RNA polymerase (RdRp), thus limiting viral RNA synthesis, providing for the first time in fish some insights into the possible mechanisms of action of Mx proteins.\textsuperscript{176}

### 6.2 Viperin

Viperin, initially designated as vig-1 [viral hemorrhagic septicemia virus (VHSV)-induced gene 1], was identified by mRNA differential display methodology in rainbow-trout leukocytes as a major responder gene to VHSV.\textsuperscript{177} Subsequent studies demonstrated that its transcription could also be induced by IFN in fish\textsuperscript{178} and mammals,\textsuperscript{134} where it was designated as viperin. Recently, a \textit{viperin} gene has been reported in birds\textsuperscript{179} and some other fish species.\textsuperscript{180,181} In mammals, the antiviral activity of viperin has been revealed in several viral models when expressed in mitochondria,\textsuperscript{134} when associated to the endoplasmatic reticulum\textsuperscript{182} or through its interaction with lipid rafts.\textsuperscript{183} In a respiratory syncytial virus-infection model, viperin did not affect viral protein expression or inclusion of body formation, but was associated with the inhibition of virus filament formation and reduced cell-to-cell virus transmission.\textsuperscript{184} In contrast, equine viperin restricted equine infectious anemia virus replication at multiple steps of the replication process, including virus entry into the cells.\textsuperscript{185} On the other hand, in humans, viperin is needed for the changes in lipid metabolism triggered during cytomegalovirus infection.\textsuperscript{186} Although the mechanism of action against nonmammalian viral pathogens has never been reported, the antiviral activity of crucian carp (\textit{Carassius auratus}) and rock bream (\textit{Oplegnathus fasciatus})\textsuperscript{180} viperin has been revealed recently.\textsuperscript{181}

### 6.3 PKR

PKR is a serine-threonine kinase that inhibits protein synthesis by phosphorylation of the eukaryotic translation initiation factor 2-alpha (eIF2\textalpha). It has also been shown to activate NF\kappaB by inducing NF\kappaB-inducing kinase and I\kappaB (inhibitor of NF\kappaB) kinase.\textsuperscript{187} Although it has been mainly proposed as an antiviral agent in mammals that is known to limit viral translation in infections with viruses such as vesicular stomatitis virus (VSV)\textsuperscript{188} or vaccinia virus,\textsuperscript{189} some viruses like hepatitis C virus have developed strategies to interfere with the action of PKR, and consequently use it as a pro-infection agent.\textsuperscript{187} The \textit{pkr} gene has been identified in amphibians,\textsuperscript{190} birds,\textsuperscript{191} and in several fish species such as zebrafish,\textsuperscript{192} Atlantic salmon,\textsuperscript{193} rare minnow,\textsuperscript{194} and grass carp.\textsuperscript{195} Although these fish PKR-like proteins have Z-DNA-binding domains instead of dsRNA-binding domains in the regulatory domain, they were designated as Z-DNA-binding protein kinases (PKZs).\textsuperscript{192} Although it was first thought that \textit{pkz} genes
were the only orthologues of mammalian pkr genes in fish, true pkr genes were later identified in several fish species, including zebrafish, Japanese flounder (Paralichthys olivaceus), fugu, and rock bream. Fish pkr genes are closer to fish pkz than to mammalian pkr, suggesting that they are paralogues that derive from an ancestral kinase gene duplicated after the divergence from the tetrapod lineage as an adaptation to specific fish viruses. The salmon PKZ and both fugu PKRs were capable of inhibiting protein translation and activating NFκB, demonstrating equivalent roles for fish and mammalian PKRs. In Japanese flounder, PKR was also shown to phosphorylate eIF2α and to inhibit the replication of Scophthalmus maximus rhabdovirus (SMRV) in flounder embryonic cells, also revealing a conservation of its antiviral capacities in fish.

6.4 ISG15

ISG15 is one of the most abundant transcripts after type I IFN treatment in mammals. The protein has an important sequence homology to ubiquitin, and similar to ubiquitin, ISG15 is conjugated to target proteins via a pathway named “ISGylation.” Both ISGylation and ubiquitination provoke posttranslational modifications that modify the stability, the function, or the localization of the targeted proteins. Within the cells, ISG15 can be found either free or conjugated to target proteins. In macrophages, B and T lymphocytes and epithelial cells, an additional free form of ISG15 that functions as a cytokine is also secreted into the medium, where it is known to activate IFN-γ production in T lymphocytes. ISG15 has been shown to limit viral replication in different mammalian viral models, such as HIV-1, Sindbis virus, or Ebola. Whereas isg15 genes are missing in birds and have not been reported in reptiles or amphibians, one or several isg15 genes are expressed in all fish species, in which a search for these genes has been performed. These include rainbow trout, goldfish, Atlantic salmon, crucian carp, Atlantic cod (Gadus morhua), and Japanese flounder; all of these genes are upregulated in response to viral stimuli. In Atlantic cod, three different isoforms are present, and even though the three of them were induced in response to Poly I:C, only one of them seemed to conjugate with intracellular proteins, suggesting different functionalities for each isoform. Even though the antiviral capacity of fish ISG15 has not been widely explored, results showing that Atlantic salmon ISG15 can bind an infectious salmon anemia virus (ISAV) protein as well as different cellular proteins, and the fact that ISG15 is transcriptionally regulated in response to viral stimuli in all fish species examined, suggest a conservation of IFN-induced ISG15-mediated antiviral properties in fish.

6.5 TRIM

The superfamily of tripartite motif-containing (TRIM) proteins is conserved throughout the metazoan kingdom and has expanded rapidly during vertebrate
evolution through gene expansion, positive selection, and alternative splicing.\textsuperscript{214} This group of proteins is defined by the N-terminal TRIM or RBCC motif, which include in this order: a RING domain, one or two B-box domains that determine the partners of the protein and in which pathway they are involved, and an associated coiled-coil domain.\textsuperscript{215} Interestingly, the RING domain has an E3 ubiquitin ligase activity that has the capacity to bind proteins either with ubiquitin or ubiquitin-like molecules such as ISG15,\textsuperscript{130} suggesting a cooperative action of different ISGs. Many different mammalian \textit{trim} genes are known to be induced by type I IFN or viral infection.\textsuperscript{216} Additionally, many of them were shown to have direct antiviral effects through diverse mechanisms (reviewed in\textsuperscript{214}). Several \textit{trim} genes have been identified in birds,\textsuperscript{217} amphibians,\textsuperscript{218} fish, and even nematodes.\textsuperscript{219} In teleost fish, a group of \textit{trim} genes with variable members depending on the species was identified after viral infection.\textsuperscript{220} Since these genes do not seem to have true orthologues in tetrapods, they were designated as finTRIMs for “fish novel TRIMs”. On the other hand, while orthologues of some mammalian \textit{trim} genes such as \textit{trim5}, \textit{trim22}, or \textit{trim19} are clearly missing from fish genomes, what appear to be true counterparts of mammalian \textit{trim} genes have also been reported in fish.\textsuperscript{221} While the diversity of this family in fish implies an important role in antiviral immunity, the bioactivity of these proteins remains unexplored.

7 CONCLUSIONS

All vertebrates are able to mount strong and effective antiviral responses in response to pathogenic viral exposure. The first line of antiviral defense is that mediated by innate immunity through the IFN system, including recognition of viral features, induction of IFN, and translation of ISGs that directly interfere with viral replication at the cellular level. This system, conserved throughout vertebrate evolution, seems particularly important to vertebrates in which an adaptive immune system is not fully developed and/or is dependent on environmental parameters such as temperature. The variability and diversity of IFNs and ISGs illustrate the diversity of cellular mechanisms involved in antiviral responses, as well as the long term coevolutionary history of viruses and vertebrate cells. Consequently, comparing the IFN systems in the different vertebrate groups provides an evolutionary perspective to understand the basis and complexity of this remarkable system.

ACKNOWLEDGMENTS

This work was supported by Starting Grant 2011 (Project No. 280469) from the European Research Council, and the European Commission under the 7th Framework Programme for Research and Technological Development (FP7) of the European Union (Grant Agreement 311993 TARGETFISH). The authors want to thank Professor Chris Secombes for critically reviewing this chapter.
REFERENCES


75. Hermant P, Francius C, Clotman F, Michiels T. IFN-epsilon is constitutively expressed by cells of the reproductive tract and is inefficiently secreted by fibroblasts and cell lines. PLoS ONE 2013;8:e71320.


121. Sun B, Greiner-Tollersrud L, Koop BF, Robertse B. Atlantic salmon possesses two clusters of type I interferon receptor genes on different chromosomes, which allows for a larger repertoire of interferon receptors than in zebrafish and mammals. *Dev Comp Immunol* 2014;47:275–86.


176. Wu YC, Lu YF, Chi SC. Anti-viral mechanism of barramundi Mx against betanodavirus involves the inhibition of viral RNA synthesis through the interference of RdRp. *Fish Shellfish Immunol* 2010;**28**:467–75.


Chapter 8

Lectins as Innate Immune Recognition Factors: Structural, Functional, and Evolutionary Aspects

Gerardo R. Vasta
Department of Microbiology and Immunology, University of Maryland School of Medicine, UMB, and Institute of Marine and Environmental Technology, Columbus Center, Baltimore, MD, United States

1 INTRODUCTION

Invertebrates and protochordates rely mostly on innate immunity for defense against microbial infection. Moreover, it has become now well-established that in vertebrates, innate immunity not only carries a substantial burden of the defense functions against infectious diseases, but it is also critical for the development of an effective adaptive immune response. Although immunoglobulin superfamily members such as hemolin, FREP, and Dscam appear to mediate immune responses in selected mollusk and insect species, true immunoglobulins, which are the typical recognition/effecter components of the adaptive immunity of vertebrates, are lacking in invertebrates. Consequently, the identification and structural/functional characterization of the other recognition factors, such as lectins, toll and toll-like receptors, and other “nonself” recognition and effector mechanisms that in invertebrates may be responsible for defense against infectious disease, has generated substantial interest. Additionally, the early realization that many of these factors/mechanisms have been conserved along the vertebrate lineages leading to the mammals has expanded the interest on these studies even further.

It is currently firmly established that soluble and cell-associated lectins play critical roles as recognition/effecter factors in the immune responses of both invertebrates and vertebrates. Unlike immunoglobulins, most lectins do not generate diversity in recognition by genetic recombination. Therefore, attention has focused on the potential germline-encoded diversity of the lectin repertoires, including allelic variation; the presence of multiple carbohydrate
recognition domains (CRDs) resulting from tandem gene duplications; the expression of chimeric structures housing multiple recognition and effector domains, resulting from exon shuffling; and other mechanisms, such as alternative splicing and somatic mutation that expand the lectin’s structural diversity and their ligand-recognition spectrum. Further, the structural basis for the specificity and potential “plasticity” of the carbohydrate-binding sites for recognition of topologically related ligands by any given lectin has recently generated great interest.

In most lectin families, the polypeptide subunits are organized as oligomeric structures that result in increased recognition avidity for multivalent ligands displayed on soluble and cell-surface glycans. The density of the surface ligands and their scaffolding (as glycoproteins, glycolipids, or polysaccharides) modulates affinity of the CRD–ligand interaction, via negative cooperativity. Based on their structural folds and the presence of unique sequence motifs in the CRDs, lectins have been classified within several families that considerably differ in their structural and functional aspects, as well as their evolutionary history. Among the lectin families that mediate innate immune recognition in both invertebrate and vertebrate species, the C- and F-type lectins, rhamnose-binding lectins, pentraxins, and galectins (formerly S-type lectins) have been characterized in considerable detail.

Like the Toll and Toll-like receptors, those lectins that recognize microbial components have been considered as “pattern recognition receptors” (PRRs), and their binding targets, such as lipopolysaccharide, lipoteichoic acid, or peptidoglycan, have been designated as pathogen- or microbe-associated molecular patterns (PAMPs or MAMPs, respectively). The binding of a lectin to multiple nonreducing terminal carbohydrate ligands on the microbial surface, which are not readily exposed in the host, leads to agglutination and immobilization, and potentially to the opsonization of the potential pathogen. Among these lectins, the mannose-binding lectin (MBL), a member of the C-type lectin family and the collectin subfamily, has been described as the prototypical PRR.

Collectins, which also include the ficolins, conglutinin, and pulmonary surfactants, are C-type lectins with a collagenous region that, upon the CRD–ligand interaction, can recruit a serine protease (MBL-associated serine protease; MASP), and the MBL(ligand)-MASP complex can activate the complement cascade. Therefore, MBL can function not only as a recognition molecule for microbes, but also as an effector factor that triggers complement-mediated opsonization or lysis, and killing of the potential pathogen. As the collectins and complement components are present in extant invertebrate and protochordate taxa that lack an adaptive immune system, it has been proposed that both lectins and complement played a pivotal role in both innate immune recognition and complement activation long before the emergence of adaptive immunity in vertebrates.

The structural analysis and taxonomic distribution of selected lectins and their families have revealed key information about their evolutionary history.
For example, among the lectin families present in both invertebrate and vertebrate species, it has been concluded that galectins have been highly conserved through evolution,\textsuperscript{12,28} whereas the F- and C-type lectins, which are largely heterogeneous from the structural standpoint, are considered as evolutionary diversified lectin families.\textsuperscript{12,21,22,29} A comparative analysis of the structural and evolutionary aspects of galectins and F-type lectins will be discussed in this review to illustrate their functional diversification, with particular focus on their roles in innate immune recognition.

2 GALECTINS: A CONSERVED LECTIN FAMILY WITH MULTIPLE ROLES IN DEVELOPMENT AND IMMUNITY

Galectins are nonglycosylated proteins, characterized by a conserved sequence motif in their CRDs and an affinity for $\beta$-galactosides. Some galectins, however, show a preference for $N$-acetyl-lactosamine (LacNAc; Gal$\beta$1,4GlcNAc) and related disaccharides.\textsuperscript{28,30,31} (Fig. 8.1A). Most galectins are soluble proteins, although a few exceptions have transmembrane domains.\textsuperscript{32,33} Galectins are synthesized in the cytoplasm; some family members, such as galectin 3, can be translocated into the nucleus, where they can form part of the spliceosome.\textsuperscript{12,34} Galectins can also be secreted to the extracellular space by nonclassical mechanisms—as they lack a typical signal peptide—possibly by direct translocation across the plasma membrane.\textsuperscript{35,36} It has been proposed that the release of galectins to the extracellular space under noninfectious stressful conditions can be perceived as a “danger-associated molecular pattern” signal that can trigger inflammatory responses.\textsuperscript{37} In the extracellular space, galectins can not only bind to cell-surface glycans and to the extracellular matrix, but can also recognize potential pathogens and parasites.\textsuperscript{38–44}

FIGURE 8.1 Structure of the galectin binding site in complex with LacNAc. Detail of the binding cleft, indicating the amino-acid residues that interact with the disaccharide.
2.1 Molecular, Structural, and Evolutionary Aspects

Galectins are characterized by their extensive taxonomic distribution and striking evolutionary conservation of primary structures, gene organization, and structural fold\textsuperscript{12,28} (Fig. 8.2A, B). The identification of galectin-like proteins in the fungus \textit{Coprinopsis cinerea} and in the sponge \textit{Geodia cydonium}, and a protein sharing the galectin fold in the protozoan parasite \textit{Toxoplasma gondii} revealed the early emergence and structural conservation of galectins in eukaryotic evolution.\textsuperscript{45–47} As indicated previously, the C- and F-type lectin families are structurally diversified, and comprise numerous members that display chimeric polypeptide subunits, that, in addition to the lectin CRDs, display a variety of functionally diverse domains.\textsuperscript{12,21–23} In contrast to C- and F-type lectins, the galectin polypeptides exhibit a relatively simple domain organization, housing one, two, or four galectin CRDs.\textsuperscript{12,48} The only chimeric galectin described so far, is the galectin-related protein (GREP), identified in the freshwater snail \textit{Biomphalaria glabrata}, in which a C-terminal galectin domain (GLECT) is joined via a short interceding region (ICR) to two immunoglobulin superfamily domains (IgSF1 and IgSF2), separated by a small connecting region (SCR), and with the signal peptide (SP)\textsuperscript{49} (Fig. 8.2A). Based on the CRD organization of the polypeptide monomer, mammalian galectins (and, by extension, galectins in vertebrates) have been classified into three types: “proto,” “chimera,” and “tandem-repeat” (TR)\textsuperscript{50} (Fig. 8.2B). Proto-type galectins, such as galectin-1, contain one CRD per subunit and are noncovalently linked homodimers. The chimera-type galectins (galectin 3) have a C-terminal CRD and an N-terminal...
domain rich in proline and glycine. The dimerization of prototype galectins is critical for their function in mediating cell–cell or cell–ECM interactions,\(^ {51,52}\) and similar interactions via the N-terminus domain mediate oligomerization (mostly trimers and pentamers) of the chimera galectins.\(^ {53}\) In TR galectins such as galectin-9, two CRDs are joined by a functional linker peptide.\(^ {50}\) In invertebrate species, galectins that display one, two, or four tandemly arrayed CRDs have been described.\(^ {48,54–56}\) As in the mammalian tandem-repeat galectins, in those invertebrate galectins that carry multiple CRDs, these are similar but not identical. Based on the structure of their binding sites, they are likely to differ in their fine specificity and recognize distinct ligands.\(^ {48,57}\)

Analyses of the gene organization and primary structures of galectins from invertebrates and vertebrates have provided insight into their evolutionary history.\(^ {48,54–57}\) Among the arthropods, although the typical galectin sequence motif is present in the galectin MjGal from the shrimp Marsupenaeus japonicus, the primary structural organization does not fit the canonical features of either the proto- or the chimera-type galectins from vertebrates. A phylogenetic analysis indicated that MjGal clustered with galectins from invertebrates and a hemichordate, and was clearly distinct from the vertebrate galectins.\(^ {56}\)

Within the mollusks, the phylogenetic analysis of the full-length sequences of the 2- and 4-CRD galectins revealed that these were already present in the most recent common ancestor of both bivalves and gastropods, and that the individual CRDs in both the 2-CRD and 4-CRD galectins have been maintained in the same arrangement/sequence without domain shuffling since the divergence of these two major clades. This remarkable structural conservation of the biologically active carbohydrate-binding domains in the evolution of the mollusk taxa supports the key roles of galectins in essential biological functions related to \(\beta\)-galactoside recognition.\(^ {46}\) How the multiple CRD galectins from invertebrates relate to the vertebrate TR galectins remains to be fully understood, but a preliminary phylogenetic analysis of the oyster galectin CvGal1 revealed that the individual CRDs cluster with the mammalian single-CRD galectins rather than with the TR galectins, suggesting that the CvGal1 gene is the product of two consecutive gene duplications of a single-CRD galectin gene.\(^ {54}\)

When considering the evolution of galectins along the vertebrate taxa, either prior to or early on in chordate evolution, the duplication of a mono-CRD galectin gene would have led to a bi-CRD galectin gene, in which the N- and C-terminal CRDs subsequently diverged into two different subtypes, defined by exon–intron structure (F4-CRD and F3-CRD). All vertebrate single-CRD galectins belong to either the F3- (eg, gal-1, -2, -3, -5) or F4- (eg, gal-7, -10, -13, -14) subtype, whereas the mammalian TR galectins such as gal-4, -6, -8, -9, and -12 contain both F4 and F3 subtypes.\(^ {28,57}\) Galectins have also been subject to evolutionary cooption. Examples of this process are the galectin-like proteins, such as the eye-lens crystallin protein GRIFIN (galectin related interfiber protein) that lack carbohydrate-binding activity in mammals, but in teleost fish display the typical galactoside-binding activity.\(^ {58,59}\)
The amino acid residues of the galectin 1 CRD that participate in protein–carbohydrate interactions and the hydroxyl groups recognized on the ligands have been rigorously identified by the resolution of the structure of the protein cocrystallized with LacNAc.\textsuperscript{30,31,60} The galectin 1 subunit is a typical \(\beta\)-sandwich, which contains one carbohydrate-binding site that is formed by three continuous concave strands (\(\beta_4–\beta_6\)), which in turn contains all residues involved in direct interactions with LacNAc; these include histidine 44, asparagine 46, arginine 48, histidine 52, asparagine 61, tryptophan 68, glutamic acid 71, and arginine 73.\textsuperscript{30} The higher affinity for LacNAc over Lac can be rationalized by the water-mediated interactions between His52, Asp54, and Arg73 in the galectin 1 CRD with the nitrogen of the NAc group. In galectin-3, the carbohydrate-binding site is shaped like a cleft open at both ends, exposing the GlcNAc of the LacNAc to the solvent.\textsuperscript{61} This extended binding site in galectin 3 results in increased affinity for polylactosamines and for ABH blood-group oligosaccharides [Fuc\(\alpha_1,2\)GalNAc\(\alpha_1,3\)(Fuc\(\alpha_1,2\)); and Gal\(\alpha_1,3\)(Fuc\(\alpha_1,2\)). In general, glycans that display LacNAc and polylactosamine chains [(Gal\(\beta_1,4\)GlcNAc)n], such as laminin, fibronectin, and mucins are the preferred endogenous ligands for galectins.\textsuperscript{62–67}

Although from the structural standpoint galectins are substantially conserved, as observed for other lectin families, the galectin repertoire is constituted by multiple galectin types, subtypes, and isoforms.\textsuperscript{62–68} Proto- and TR-type galectins comprise several distinct subtypes: Galectins-1, -2, -5, -7, -10, -11, -13, -14, and -15 are proto-type. Galectin-3 is the only chimera-type. Galectins-4, -6, -8, -9, and -12 are TR-type. In addition, any galectin subtype may exhibit multiple isoforms in a single individual.\textsuperscript{62} Further, because galectin types and subtypes exhibit notable differences in carbohydrate specificity, and bind a broad range of glycans that display the requisite topologies, the galectin repertoire displays considerable diversity in recognition properties that suggests functional diversification.\textsuperscript{62–68}

### 2.2 Functional Aspects

The experimental evidence obtained soon after the initial identification of galectins during the early 1980s suggested that their expression was developmentally regulated, and that their functions related to embryogenesis and early development. Later studies revealed their roles in neoplastic transformation and progression, and metastasis (reviewed in Vasta and Ahmed\textsuperscript{12}). Glycans that contain polylactosamine chains [(Gal\(\beta_1,4\)GlcNAc)n], such as laminin, fibronectin, lysosome-associated membrane proteins, and mucins, are the preferred endogenous ligands for mammalian, bird, and amphibian galectins.\textsuperscript{61–68} Chicken galectins have been proposed to participate in myoblast fusion, whereas murine galectin-1 and galectin-3 would have roles in notochord development, somitogenesis, and development of muscle tissue and the central nervous system.\textsuperscript{69–72} More recently, other genetically tractable model organisms endowed with a less
diversified galectin repertoire, such as *Drosophila* and zebrafish, have become attractive alternatives for functional studies.\(^{73,74}\)

Since the 1990s, the roles of galectins as regulators of both innate and adaptive immune homeostasis have firmly established and characterized in detail.\(^{75}\)

Galectins are ubiquitously expressed and distributed in mammalian tissues, including most cells of the innate (dendritic cells, macrophages, mast cells, natural killer-cells, gamma/delta T cells, and B-1 cells) and adaptive (activated B and T cells) immune system, and in other cell types.\(^{76,77}\)

Endogenous glycans recognized by galectins include \(\beta\)-integrins, CD45, GM1, CD44, Tim3, MUC1, podoplanin, CD166, ABH-type oligosaccharides CD43, CD45, CD7, CD71, CD44, TIM3, CTLA4, MUC1, MUC16, and MerTK.\(^{43,78–83}\)

Although it has been proposed that a certain degree of functional redundancy exists among the members of the galectin repertoire, as the subtle aspects of their binding properties and natural ligands are characterized, and their biological roles are elucidated in increasing detail, it has become clear that this is not the case.\(^{75}\)

In recent years, it has become clear that galectins can also recognize “non-self” carbohydrate moieties on the surface of microbial pathogens and parasites, and function as PRRs.\(^{38,42}\) These ligands on foreign cells can be similar to those displayed on host cells such as ABH or Le blood-group oligosaccharides,\(^{42}\) and LacNAc present in viral and bacterial glycans, or structurally different and absent from the host glycome, such as \(\alpha\)-1-2-mannans in *Candida*\(^{43}\) and LacdiNAc in *Schistosoma*.\(^{84}\)

Whereas the first scenario can be rationalized as molecular mimicry by the microbial pathogens and parasites, understanding the molecular basis of galectin binding to distinct self and nonself glycans via the same CRD requires additional considerations. In this regard, galectins with tandemly arrayed CRDs such as the TR galectins from vertebrates, and the 4-CRD galectins from invertebrates, are intriguing both in their binding properties and functional aspects. Vertebrate TR galectins such as galectins-4, -8, and -9 differ from the proto- and chimera types, in that they display two tandemly arrayed CRDs (N- and C-CRDs) that are similar but not identical, suggesting that they have distinct recognition properties.\(^{85}\)

The structures of the TR galectins-4, -8, and -9 have been partially resolved, either by crystallization of NMR analysis of their isolated N- or C-CRDs and revealed differences in their binding specificity, or affinity for oligosaccharides or their scaffolding as glycolipids or glycoproteins.\(^{86–88}\)

The structure of the N-CRD of the mouse galectin-4 revealed binding sites for lactose with different affinities, whereas the galectin-8 binds preferentially to larger glycans, such as glycosphingolipids.\(^{86–88}\)

The capacity of TR galectins to crosslink cells with different synthetic glycoconjugates\(^{88,89}\) strongly suggests significant differences in the binding properties of their N- and C-CRDs.

From the functional standpoint, galectins can function as opsonins,\(^{56}\) inhibit viral adhesion to the host cell,\(^{90}\) or directly kill bacteria.\(^{42}\)

The expression of the galectin MjGal from the kuruma shrimp *M. japonicus* is upregulated hemocytes and hepatopancreas upon bacterial challenge, and can bind to both Gram-positive
and Gram-negative bacteria through the recognition of lipoteichoic acid or lipopolysaccharide, respectively. By also binding to the shrimp hemocyte surface, MjGal functions as an opsonin for microbial pathogens, promoting their phagocytosis and clearance from circulation.\textsuperscript{56} In the zebrafish (\textit{Danio rerio}), the extracellular proto-type galectin-1 (DrGal1-L2) and a chimera galectin-3 (DrGal3-L1) interact directly and in a carbohydrate-dependent manner with the glycosylated envelope of the infectious hematopoietic necrosis virus (IHNV), and also with the glycans on the epithelial cell surface, significantly reducing viral adhesion.\textsuperscript{90} In mammals, the TR galectins-4 and -8 can recognize and kill \textit{Escherichia coli} strains that display B-blood-group oligosaccharides (BGB+ \textit{E. coli}). Mutation of key residues in either CRD revealed that the C-CRD mediates recognition of the BGB+ \textit{E. coli}, but does not affect its viability, whereas the N-CRD was not affected, suggesting that N-CRD might be endowed with killing activity.\textsuperscript{42} Taken together, the results of these studies indicate that galectins can function not only in immune recognition but also as effector factors.

In recent years, mounting experimental evidence has shown that some pathogens and parasites can “subvert” the roles of host or vector galectins as PRRs, to attach to or gain entry into their cells. The participation of galectin interactions in the infection mechanisms of HIV has been reported.\textsuperscript{91,92} Galectin-1, which is abundant in organs that represent major reservoirs for HIV-1, such as the thymus and lymph nodes, promotes infection by HIV-1 by facilitating viral attachment to the CD4 receptor, and increasing infection efficiency. It is noteworthy that HIV also uses recognition by DC-SIGN, a C-type lectin, to enter dendritic cells, thereby underscoring the multiple adaptations of the viral glycome for host infection.\textsuperscript{91,92} \textit{Leishmania} species, which spend part of their lifecycle in phlebotomine sandflies that constitute vectors for transmission to the vertebrate hosts, attach to the insect midgut epithelium via the sandfly galectin PpGalec, to prevent their excretion along with the digested blood meal, and differentiate into free-swimming infective metacyclics.\textsuperscript{93} The galectin is only expressed by epithelial midgut cells, is upregulated in the blood-feeding females, and binds to the Gal(β1-3) side chains on the \textit{Leishmania} LPG.\textsuperscript{93} The protozoan parasite \textit{Perkinsus marinus}, a facultative intracellular parasite that causes “Dermo” disease in the eastern oyster \textit{Crassostrea virginica},\textsuperscript{94} is recognized via two 4-CRD galectins (CvGal1 and CvGal2) that are expressed by the phagocytic hemocytes.\textsuperscript{54,95,96} The parasite is phagocytosed by the oyster hemocytes, where it inhibits respiratory burst and proliferates, eventually causing systemic infection and death of the oyster host. Thus, \textit{P. marinus} may have evolved to adapt its glycocalyx to be selectively recognized by the oyster hemocytes CvGal1 and CvGal2, thereby subverting the oyster’s innate immune/feeding recognition mechanism to gain entry into the host cells.\textsuperscript{38,48,54,95,96} A recent study identified galectin-1 as the receptor for the protozoan parasite \textit{Trichomonas vaginalis},\textsuperscript{97} the causative agent of the most prevalent nonviral sexually transmitted human infection in both women and men. As an obligate extracellular parasite, establishment and persistence of \textit{T. vaginalis} infection requires adherence to
the host epithelial cell surface. Like *Leishmania* spp, *T. vaginalis* displays a surface LPG rich in galactose and *N*-acetyl glucosamine, which is recognized in a carbohydrate-dependent manner by galectin-1 expressed by the epithelial cells in the cervical linings, as well as by placenta, prostate, endometrial, and decidual tissue, also colonized by the parasite.  

Recent studies in a murine model for influenza (influenza A virus, IAV) and pneumococcal pneumonia (*Streptococcus pneumoniae*) revealed that the murine lung expresses a diverse galectin repertoire, from which selected galectins, including galectin 1 (Gal1) and galectin 3 (Gal3), are released to the bronchoalveolar space upon the viral infection. In vitro studies on the human airway epithelial cell line A549 were consistent with the observations made in the mouse model, and further revealed that both Gal1 and Gal3 bind strongly to IAV and *S. pneumoniae*, and that exposure of the cells to viral neuraminidase or influenza infection increased galectin-mediated *S. pneumoniae* adhesion to the cell surface, suggesting that upon influenza infection, pneumococcal adhesion to the airway epithelial surface is enhanced by an interplay among the host galectins and viral and pneumococcal neuraminidases. The combined activity of microbial neuraminidases, and the secreted Gal1 and Gal3 at the epithelial cell surface, also modulate the expression of SOCS1 and RIG1, and activation of ERK, AKT, or JAK/STAT1 signaling pathways, leading to a disregulated expression and release of proinflammatory cytokines. These studies suggest that galectins released to the lung bronchoalveolar environment “fine-tune” the inflammatory response to viral and bacterial challenge.

### 3  F-TYPE LECTINS: A STRUCTURALLY AND FUNCTIONALLY DIVERSIFIED LECTIN FAMILY

In contrast with the galectins, which are considered as substantially conserved in evolution, the F-type lectins, together with the C-type lectins, are characterized by their highly diversified structural organization. The F-type lectin family is the most recent to be identified, and it is characterized by a unique structural fold and a canonical sequence motif in the F-type sugar-binding domain. In this family, the F-type lectin domain (FTLD) can be presented either as a single CRD, as in the European eel agglutinin (*Anguilla anguilla* agglutinin, AAA), or as tandemly arranged F-type CRD repeats, in some examples combined with unrelated domains, yielding mosaic or chimeric polypeptide subunits of variable sizes and function(s) within a single species.

#### 3.1 Molecular, Structural, and Evolutionary Aspects

The structure of the AAA/L-Fuc complex consists of a β-jellyroll sandwich composed of three- and five-stranded β-sheets (Fig. 8.3A), which interacts with the α-Fuc ligand through hydrogen bonds, established between side chains from a trio of basic amino-acid residues (Ne of His and the guanidinium groups of Arg)
and Arg\textsuperscript{103}, located in a shallow pocket, and the ring O5, and the equatorial 3- and axial 4-OH groups of the sugar (Fig. 8.3B). A van der Waals contact is established between a unique disulfide bridge, formed by contiguous cysteines (Cys\textsuperscript{78} and Cys\textsuperscript{79}) and the bond between ring atoms C1 and C2 of the monosaccharide, and the C6, which docks loosely in a hydrophobic pocket, stacking against the aromatic rings of two residues His\textsuperscript{16} and Phe\textsuperscript{44} together with the Leu\textsuperscript{24} and Tyr\textsuperscript{91} residues.\textsuperscript{105} As for most animal lectins, the specificity of AAA for α-Fuc is not absolute, but rather nominal, as carbohydrates (eg, 3-O-methyl-d-galactose and 3-O-methyl-d-fucose) that share critical configurational features of α-Fuc (ie, axial hydroxyl and hydrophobic moiety) also behave as ligands for AAA (Fig. 8.3B). Unlike C-type lectins, the role that calcium appears to play in AAA is structural stabilization, rather than participating in direct cation–saccharide interactions.\textsuperscript{101} (Fig. 8.3A). F-lectins specifically recognize selected oligosaccharides via interactions with amino-acid residues, located in what is known as an “extended binding site.” For example, the recognition of H type-1 and Le\textsuperscript{a} oligosaccharides by AAA requires additional interactions of residues in the loops (CDRs 1–5) that surround the binding pocket with subterminal sugars of the tyrisaccharides Fuca1-2 Galb1-3GlcNAcβ1-3Galβ1-4Glc (blood-group H type-1) and Galb1-3[Fuca1-4] GlcNAcβ1-3Galβ1-4Glc (Le\textsuperscript{a}). F-lectins with a shorter CDR1 (Fig. 8.3A), such as MsFBP32 from the striped bass \textit{Morone saxatilis} (see later in the chapter), would have a broader specificity for Le oligosaccharides.\textsuperscript{102}
Although F-lectins such as AAA and MsFBP32 can specifically recognize a limited number of oligosaccharides, a considerable diversity in carbohydrate recognition is achieved by the presence of multiple isoforms with amino-acid substitutions at positions revealed by structural analyses.\textsuperscript{102,104} This diversity in recognition is a critical feature for proteins that mediate immune defense. Variability of critical residues in the binding pocket and surrounding loops in the multiple isoforms, as expressed in the Japanese eel,\textsuperscript{101,102,104} suggests that alternative interactions with terminal and subterminal sugar-units may expand the range of diverse oligosaccharides recognized by the lectin isoform repertoire.\textsuperscript{105} Like collectins, the native structure of AAA is a homotrimer, which suggests that ligand binding is enhanced through cooperative binding to multivalent glycans. Further, the threefold cyclic symmetry of the native AAA, very similar to that observed in collectins,\textsuperscript{106} would optimize the spacing and orientation of binding sites for recognition of glycoconjugates displayed on microbial surfaces. The distances between binding sites in the AAA trimer (26 Å) and those in MBL (45 Å) suggest that they bind to differently arrayed surface glycans on the surface of microbes. Therefore, although F- and C-type lectins may recognize the same monosaccharide (MBL also binds fucose), they may bind to different microorganisms, thereby expanding the immune recognition spectrum in those species that are endowed with both lectin types.

In contrast to AAA, the F-lectin MsFBP32 consists of two tandemly arrayed F-type CRDs. The structure of the complex of MsFBP32 with L-fucose revealed a “tail-to-tail” arrangement of three subunits.\textsuperscript{102} Thus, the native MsBP32 trimer of approximately cylindrical shape (81-Å-long and 60-Å-wide) is divided into opposing globular structures—one containing the three N-CRDs, and the other, the three C-CRDs.\textsuperscript{102} The resulting binding surfaces at the opposite ends of the cylindrical trimer resemble the collectin-like “bouquet” that CRD displays, and have the potential to crosslink cell surface or humoral carbohydrate ligands. The overall structure of the N-CRD is highly similar to that of the C-CRD, but significant differences between the binding sites of the MsFBP32 N- and C-CRD domains strongly suggest that the N-CRD recognizes more complex fucosylated oligosaccharides, and with a relatively higher avidity than the C-CRD.\textsuperscript{102}

F-type lectins have been identified in a variety of taxa, from prokaryotes to amphibians (Fig. 8.4A, B). However, the F-type lectin sequence motif appears to be absent from protozoa, fungi, nematodes, ascidians, and higher vertebrates such as reptiles, birds, and mammals.\textsuperscript{100} Despite the fact that AAA and MsFBP32 possess one and two FTLDs, respectively, substantially diverse domain topologies, in some cases lineage-related, were identified based on sequence alignments of the F-type lectin motif.\textsuperscript{100} For example, most teleost F-type lectins contain either duplicate or quadruplicate tandem domains, whereas in \textit{Xenopus} spp, these lectins are composed of either triplicate or quintuple tandem F-type domains. Clearly, the F-type fold with its joined N- and C-terminals favors
The formation of concatenated CRD topologies in numbers that appear lineage-related. These tandem arrays may yield mosaic proteins by including pentraxin (*Xenopus laevis*) or C-type domains (*D. melanogaster* CG9095, malarial mosquito, and honey bee). The F-type sequence motif is also present in lophotrochozoan (ie, mollusks and planaria) and ecdysozoan protostomes (ie, horseshoe crabs and insects), invertebrate deuterostomes (ie, echinoderm), elasmobranchs (ie, skate), lobe- and ray-finned teleost fish, and amphibians (ie, *X. laevis* and salamander).

The binary FBPLs have diversified through lineage-dependent gene duplications and speciation events, producing a combination of paralogous relationships unique to teleosts. In contrast, in *Xenopus* spp., frogs are not only single-domain F-type lectins expressed, but also combinations of two, three, four, and chimeric protein containing five tandem F-type domains adjacent to a pentraxin domain. Clearly, *Xenopus* spp. exhibit a greater diversity of F-type lectins than the pufferfish. Despite the diversity evident in this early tetrapod, no homologs are detectable in genomes of higher vertebrates, including reptile and avian representatives. This observation begs the question of whether this lectin family is uniquely restricted to invertebrates and cold-blooded vertebrates, and had been subsequently lost, as such, above the level of the amphibians. Specifically, the F-lectins may have either became truly extinct or have been coopted into other biological roles, which in the course of evolution may
have imposed structural constraints, such as those proposed for the C-1 and C-2 domains of the coagulation factors V and VIII.\textsuperscript{100–102,107}

The F-type lectin sequence motif is not restricted to eukaryotes; homologs were identified in both a Gram-positive (eg, \textit{S. pneumoniae})\textsuperscript{108} and a Gram-negative (ie, \textit{Microbulbifer degradans})\textsuperscript{109} eubacteria. The absence of the F-type lectin sequence motif in protozoa, fungi, nematodes, ascidians, and higher vertebrates suggests that it may have been selectively lost, even in relatively closely related lineages.\textsuperscript{100,107} The paucity of bacteria possessing F-type CRDs suggests that it may have either been acquired through horizontal transfer from metazoans, or less likely, that most prokaryote lineages lost this CRD. Even the multiple duplicate tandem homologs present within modern teleost orders appear to be the product of independent duplications. The spotty phylogenetic distribution, diverse temporospatial expression, and varied domain architecture of the F-type family members point to a functionally plastic CRD, which has been specifically tailored in each lineage, and has apparently lost its fitness value in some taxa.\textsuperscript{100,107} The absence of the F-type CRD in higher vertebrates is an evolutionary quandary that correlates with the appearance of cleidoic egg, and the colonization of land by vertebrates.\textsuperscript{100,107}

3.2 Functional Aspects

The opposite orientation of the binding surfaces of the trimeric MsFBP32 supports the notion that the function of this lectin in circulation is to crosslink fucosylated glycoconjugates displayed on different cells, with an epitope separation of 25 Å on the cell surface.\textsuperscript{102} Modeling of MsFBP32 complexed with fucosylated glycans that are widely distributed in prokaryotes and eukaryotes rationalizes the observation that binary tandem CRD F-type lectins can function as opsonins. This would take place by crosslinking “nonself” carbohydrate ligands and “self” carbohydrate ligands, such as sugar structures, displayed by microbial pathogens and glycans on the surface of phagocytic cells from the host.\textsuperscript{102} For example, MsFBP32 may crosslink Le\textsuperscript{a}-containing glycans on the phagocytic cell surface via the N-CRD, with glycans on the microbial surface via the C-CRD, such as those containing \(\alpha\)-linked L-Fuc, 2-acetoamido L-Fuc, 3-deoxy-L-fucose (colitose), or L-Rha (6-deoxy-L-mannose, present in \textit{E. coli} glycans) as nonreducing terminal residues.\textsuperscript{102} Preexposure of \textit{E. coli} to a binary tandem F-lectin from sea bream significantly increases their phagocytosis by peritoneal macrophages relative to the unexposed bacteria,\textsuperscript{110} confirming that F-lectins with tandemly arrayed CRDs such as MsFBP32 function as opsonins that mediate innate immune responses against microbial pathogens.

The immune-recognition functions of F-type lectins that we have identified in teleost fish are not always shared by those F-lectins expressed in other taxa. For example, the sperm “bindins” from the Japanese oyster (\textit{Crassostrea gigas}), recently identified as F-type lectins, are polymorphic gamete-recognition proteins stored in the acrosomal rings that bind sperm to egg during fertilization.\textsuperscript{111}
In contrast to most lectins, by mechanisms of positive selection, recombination, and alternative splicing, the single-copy bindin gene produces highly diversified transcripts, both in sequence and domain organization, within and among individuals among this species. However, only one or two polymorphic molecular species housing between one and five tandemly arrayed F-lectin domains are translated in each individual male oyster. The unusually high intraspecific diversity in sequence and domain organization of the oyster bindin F-lectins could represent coevolution of sperm-gamete recognition mechanisms to “catch-up” with the high diversification of egg receptors aimed at avoiding polyspermatism. This diversification of the bindin F-lectins in gamete recognition resembles the eel F-lectin isoforms involved in host–pathogen interactions described previously.

In other examples, the recognition properties of F-type lectins have been even more drastically modified or coopted. In this regard, a structure-based search identified seven nonlectin proteins with negligible sequence similarity to the F-type lectin motif; however, they shared the same jellyroll fold with AAA. These are the C domains of human blood coagulation factor V and VIII, the C-terminal domain of a bacterial sialidase, the NH2-terminal domain of a fungal galactose oxidase, a subunit of the human APC10/DOC1 ubiquitin ligase, the N-terminal domain of the XRCC1 single-strand DNA repair complex, and a yeast allantoicase. Curiously, the FA58C domain of identical structure is present next to the F-type domain in the M. bulbifer protein previously described. In most of these proteins the AAA-like domain appears to mediate binding. However, allantoicase is the first reported analog to exhibit intrinsic enzyme activity.

4 CONCLUSIONS

The structural and functional analysis of two representative families of animal lectins, such as the galectins and F-type lectins presented in this review, reveals that, in spite of their distinct evolutionary history and their structural conservation or diversification, from the functional standpoint both lectin families are mostly pleiotropic, that is, they can orchestrate an array of functions not only in innate and adaptive immunity, but also in development and homeostatic regulation of various physiological aspects. Furthermore, in recent years evidence has accumulated to support the notion that selected members of both families have been coopted to carry out other functions that, in several cases, are not dependent on their carbohydrate-binding sites, a property that is key to their definition as lectins, and appears to have been lost in the evolutionary process.

With regard to their roles in immune recognition, recent studies have firmly established that both F-type lectins and galectins can recognize self and nonself glycans. Because F-lectins and TR galectins display tandemly arrayed CRDs of similar but distinct specificity in a single polypeptide monomer, the binding and crosslinking of endogenous and exogenous glycans can be rationalized by the distinct properties of their binding sites. For other lectins, such as the single CRD F-lectins, and the proto- and chimera-type galectins that display a single
binding-site per monomer, their capacity to recognize both endogenous and exogenous glycans through the same binding site can be explained by taking into consideration the multiple factors pertaining to the local lectin concentrations and oligomerization, the geometry of the presentation of the multivalent carbohydrate ligands on the host or microbial cell surface, and the properties of the microenvironment in which interactions take place.

The recent availability of genomic databases for numerous animal species has enabled greater insight into the structural complexity and functional diversity, and of lectin repertoires in invertebrates, protochordates, and ectothermic vertebrates. The identification in these taxa of members of the lectin families typical of mammals such as galectins, has resulted in the discovery of novel structural features, most likely revealing functional adaptations along the lineages leading to the higher vertebrate taxa. Further, the identification of novel lectin families such as the F-type lectins, underscores the fact that more research in nonmammalian model organisms will provide new information on all of the structural, functional, and evolutionary aspects of lectin repertoires that may not be as obvious in mouse or man. For example, structural analysis of the eel multiple isoforms as mechanisms that generate substantial diversity in oligosaccharide binding, provide the structural basis for a tantalizing novel mechanism for generating diversity for nonself recognition in innate immunity, that resembles those operative through adaptive immunity in higher vertebrates. Similarly, analysis of the genetic mechanisms that are operative in the diversification of the bindin transcripts in the Pacific oyster,\textsuperscript{113} has contributed conceptually transformative evidence for the processes through which lectins can generate structural (and possibly, functional) diversity. The ongoing genome, transcriptome, and proteome projects on additional model organisms representative of nonmammalian taxa will reveal not only the extent of their full lectin repertoires, but, coupled to the structural analysis of selected components, has the potential to uncover novel structural features, on which a rigorous experimental assessment of their biological roles may be supported. In turn, these studies will provide greater insight into the evolutionary history of the various lectin families, from prokaryotes to the mammals.

**ACKNOWLEDGMENTS**

The author’s research reviewed herein was supported by Grants IOS 1050518, IOB-0618409, and IOS-0822257 from the National Science Foundation, and Grant R01GM070589 from the National Institutes of Health.

**REFERENCES**


Chapter 9

Origin and Evolution of the Neuro-Immune Cross-Talk in Immunity

Enzo Ottaviani
Department of Life Sciences, University of Modena and Reggio Emilia, Modena, Italy

1 INTRODUCTION

The intercellular communication is mediated mainly by chemical signal molecules. During the course of evolution, those organisms who better developed this form of communication have increased their fitness, and thus have passed this trait on to their descendants. A plausible explanation for the emergence of a ligand-based intercellular communication can be attributed to its inherent level of sophistication, as noted by not only synaptic molecules that can enter into intercellular communication, but hormonal ones as well.¹

Central nervous system (CNS) and immune system “talk to each other” in order to maintain the body homeostasis. In this context, lymphocyte migration, circulation, and traffic are under the influence of the CNS.² Studies³,⁴ performed on mammals demonstrated a bidirectional communication between immune and neuroendocrine systems, and the levels of integration can be summarized as follows: (1) hormones and neurotransmitters bind to specific receptors on immune cells and modulate their activity; (2) products of the immune system, that is, cytokines, can act on cells of the neuroendocrine system, modifying their functions; (3) immune stimuli and hypothalamic releasing-factors induce the immune cells to synthesize neuropeptides, which, in turn, may influence the activity of the neuroendocrine system; and (4) cytokines able to modulate the activity of immune cells are produced by cells of the nervous system.

In higher organisms, communication between immune and neuroendocrine system is highly complex. Mammals certainly represent the most sophisticated level of complexity, but their roots must be traced back to defense mechanisms, which are also present in the simpler forms of life, such as invertebrates. In this context, the innate immune response and the neuroendocrine cross-talk of vertebrates resembles a mosaic of different invertebrate immune mechanisms toward pathogens.⁵
1.1 Lymphocyte as a Neuroendocrine Cell

Evidence in favor of a neuroendocrine role of the lymphocyte is well-documented, and illustrated in Fig. 9.1. Human lymphocytes are immunoreactive to neuroendocrine hormones and opioid peptides. The opioid peptides may be classified into three families: the proopiomelanocortin (POMC), the proenkephalin (PENK), and the prodynorphin (PENK B). The peptides of the POMC family derive from a single macromolecular precursor, and from the cleavage of this molecule derive hormones, such as the adrenocorticotropic hormone (ACTH) and melanocyte-stimulating hormone (MSH), as well as β-endorphin. The proteolytic cleavage of PENK produces enkephalin peptides, such as Met-enkephalin and Leu-enkephalin.

ACTH and its receptors are reported in anuran amphibian, reptile, bird, and mouse lymphocytes. ACTH and endorphins are encoded by the POMC gene. The POMC gene is expressed by a variety of different cell types, which posttranslationally process the precursor protein into entirely distinct sets of peptide products. In the pituitary gland, POMC mRNA transcripts are generated from the three exons of the mammalian POMC gene, but shorter transcripts are also found for extrapituitary sites of POMC gene expression. The peptides derived from the proteolytic fragmentation of POMC undergo further modifications of utmost importance for their biological activity. Additional posttranslation modifications may include glycosylation, phosphorylation, amidation, sulfation, and acetylation, thus affecting the secretion and function of these products.

It is interesting to emphasize that ACTH and endorphins found in lymphocytes are identical to those produced by the pituitary. Receptors for ACTH and β-endorphin have been shown in human peripheral blood lymphocytes.

With regard to opioid peptides, PENK mRNA and PENK-derived peptides were detected in many cells of the mammalian immune system, including monocytes and lymphocytes, as well as Met-enkephalin and/or syntenkephalin-containing peptides. Furthermore, opioid peptides may specifically bind to at least four types of opioid receptors, such as μ, δ, κ, and ε, present on the immune cells.
Furthermore, a variety of peptidic hormones, such as corticotropin-releasing hormone (CRH), thyrotropin (TSH), growth hormone (GH), vasoactive intestinal peptide (VIP), somatostatin, vasopressin, oxytocin, and so on, were also found. Lastly, another group of molecules, the cytokines, play an important role in the interactions between the immune and neuroendocrine systems.

1.2 Immunocyte as an Immune–Neuroendocrine Cell

Two major research groups, one in Italy (Modena) and the other in the USA (New York) found in invertebrates a correlation between immune and neuroendocrine systems similar to that observed in vertebrates. These findings emerged mainly from studies in molluscan immunocytes. Using different techniques, the presence of factors similar to POMC-derived products has been found, such as ACTH-, β-endorphin-, and α-melanocyte-stimulating hormone (α-MSH)-like molecules, in the immunocytes of different molluscan species. The immunocytochemical results revealed that immunocytes were the only cell-type positive for immunocytochemical reactions for POMC-derived products (Fig. 9.2). Other methods, such as flow cytometry, revealed the presence of ACTH- and β-endorphin-like molecules on the surface of immunocytes (Fig. 9.3), and by radioimmunoassay (RIA) tests we have quantified these molecules both in immunocytes and in the serum. In particular, the concentrations for ACTH-like molecules were 47 pg/10^3 cells in cell homogenates and 47 ± 11 pg/ml in the serum, while for β-endorphin were 39 pg/10^3 cells and 69 ± 7 pg/ml, respectively. Moreover, also quantified were the concentrations of corticotropin-releasing hormone (CRH)-like molecules, that is, 21 pg/10^3 cells and 23 ± 2.5 pg/ml, respectively.

The previous data were also confirmed by molecular biology studies. Using a probe for the human POMC gene by in situ hybridization, we showed that molluscan immunocytes were positive for POMC mRNA.

FIGURE 9.2 Immunocytochemical staining with anti-ACTH polyclonal antibody on the immunocytes of *Viviparus ater*. Nuclei were counterstained with hematoxylin. Scale bar = 10 μm.
Overall, these findings support the hypothesis that the melanocortin system in vertebrates is a complex that had arisen prior to the emergence of jawless vertebrates, over 500 million of years. In this context, it should be reminded that the protozoan *Tetrahymena pyriformis* is probably the most primitive organism in which ACTH- and β-endorphin-like molecules have been detected. Furthermore, ACTH-receptor-like messenger RNA was found in molluscan immunocytes.

Study performed by Stefano and coworkers have demonstrated the presence of opioid-like molecules and the related opioid receptors in the hemocytes of invertebrates such as the marine mussel *Mytilus edulis*, the insect *Leucophaea Maderae*, and the leech *Theromyzon tessulatum*. The existence of bioactive peptides (BAPs) that function both as neurotransmitters and hormones is now accepted, and a variety of these peptides has been identified in molluscan hemocytes. In this context, another important BAP such as CRH was found in hemocytes of *Planorbarius corneus*, *Viviparus connectus*, *Lymnaea stagnalis* and *M. galloprovincialis*. In this last species it has also found the expression of molecules homologous to human mRNAs of the two receptor subtypes (CRH-R1 and CRH-R2).

Cytokine- and growth factor-like molecules were detected in various tissues of different invertebrates, such as mollusks, insects, annelids, echinoderms, and tunicates, but immunocytes are the main source. With regard to mollusks, the presence of IL-1α-, IL-1β-, IL-2-, IL-6- and TNF-α-like molecules was detected in immunocytes and hemolymph. Moreover, IL-1-, IL-6- and TNF-like
molecules have been found in neurons and in a small population of glial cells in the pedal ganglia of *M. edulis*.\textsuperscript{55–57} In the insect *Calliphora vomitoria*, the presence of TNF-\(\alpha\)-like molecules in both plasmatocytes and granular cells has been found, and these molecules are induced in activated plasmatocytes.\textsuperscript{58} Induction of TNF-\(\alpha\)-like molecules has also been observed in *M. edulis* immunocytes by lipopolysaccharides (LPS).\textsuperscript{51} IL-1\(\alpha\)- and TNF-\(\alpha\)-like molecules have been found in earthworm celomocytes of *Eisenia foetida*,\textsuperscript{59} whereas IL-1\(\alpha\)-like molecules were also detected in some small cells of the brain and in the eleocytes, a subpopulation of celomocytes of the marine worm *Nereis diversicolor*.\textsuperscript{57} In the shrimp *Litopenaeus vannamei*, the TNF superfamily gene was isolated and characterized, and the TNF-\(\alpha\) factor gene was LPS-induced.\textsuperscript{60} Moreover, the TNF-receptor-like system could cooperate with the LPS-induced TNF-\(\alpha\) factor pathway, the Toll and IMD pathways, and the JAK/STAT pathway in shrimp immune functions. Cytokine-like substances have been reported in the celomic fluid and celomocytes of the echinoderms *Pisaster ochraceus* (IL-1\(\alpha\))\textsuperscript{61} and *Asterias forbesi* (IL-1, IL-6).\textsuperscript{62,63} Studies performed in tunicates such as *Botryllus schlosserii*, *Styela plicata*, *Molugula occidentalis*, and *Ciona intestinalis* have detected IL-1-like molecules.\textsuperscript{64} An IL-1-like fraction, designed tunicate IL-1\(\beta\), has been isolated from the hemolymph of *Styela clava*, and this fraction is able to stimulate the proliferation of tunicate cells in vitro.\textsuperscript{65}

Regarding the presence of growth factors, the majority of the available data refers to insects and mollusks. Large DNA sequence similarity has been found between the genes that encode for proteins involved in *Drosophila* embryo-genesis and vertebrate growth factor, such as epidermal growth factor (EGF) and transforming growth factor (TGF)-\(\beta\).\textsuperscript{66–70} Growth-promoting factors have been found in the hemolymph of *Bombyx mori*.\textsuperscript{71–72} *Samia cynthia*\textsuperscript{73} and a hemolymph trophic factor has been found in *Manduca sexta*.\textsuperscript{74} Insulin-like peptides were identified in invertebrates, such as mollusks, insects, and worms.\textsuperscript{75–77} Blumenthal (2010)\textsuperscript{77} surmized that the genes that encoding vertebrate insulins and insulin-like growth factors (IGFs), and invertebrate insulin-like molecules evolved from a common ancestral gene, and from the concept of an insulin superfamily of growth-promoting peptides. Moreover, the presence of platelet-derived growth factor (PDGF)-AB and TGF-\(\beta1\)-like molecules in immunocytes of different mollusks (*P. corneus*, *V. ater*, *V. contectus*, *L. stagnalis*, and *M. galloprovincialis*) has been detected.\textsuperscript{78}

## 2 ROLE OF NEUROENDOCRINE HORMONES, OPIOID PEPTIDES AND CYTOKINES IN THE INVERTEBRATE IMMUNE AND NEUROENDOCRINE (STRESS RESPONSE) RESPONSES

### 2.1 Immune Responses

Cell-shape changes (ie, the expression of cell motility), chemotaxis (ie, the expression of cell migration), and phagocytosis are the main ancestral mechanisms
used by all organisms to eliminate nonself material. These defense reactions are triggered by regulatory molecules, including POMC-products, CRH, cytokines, and growth factors.

2.2 Cell-Shape Changes

ACTH, PDGF, TGF-β, CRH, and IL-8 provoke changes in cellular shape affecting the locomotor activity of molluscan immunocytes. Using computer-assisted microscopic image analysis, it has been found that ACTH induces cell-shape changes via the adenylate cyclase/AMP/protein kinase A pathway, and the protein kinase C. The PDGF and TGF-β extracellular signals are transduced via the phosphoinositide signaling pathway. IL-8 provokes the cell-shape changes via protein kinase A and C pathways. Moreover, PKA, PKC, and PKB are involved in CRH-induced cell-shape changes in immunocytes, but the synergistic effect of two second messengers, that is, cAMP and inositol 1,4,5-triphosphate, is crucial.

2.3 Chemotaxis

The chamber for the study of chemotaxis is a tool that is used to measure the ability of cells to capture chemotactic stimuli. The assay was performed in 48-well microchemotaxis chambers (Nucleopore, Pleasanton, CA, USA), in which the upper and lower compartments were separated by a 5-µm pore, polycarbonate polyvinylpyrrolidone-free filter, allowing the cells to migrate actively through the pores. The cellular suspension was placed in the upper compartment and the substance to test in the lower one.

CRH and ACTH fragments (1–24), (1–4), (4–9), (1–13), (1–17), and (11–24) significantly stimulate molluscan immunocyte migration, while the whole sequence (1–39) and the fragment (4–11) have an inhibitory effect. Differences among species were found with respect to the response to individual fragments. Furthermore, the whole sequence of β-endorphin, its N- and C-terminal fragments, the sequence (2–17), which lacks both N- and C-terminals, and the N-acetylated derivative are able to influence migration of molluscan immunocytes. The stimulatory effect of β-endorphin and its fragments is only partially inhibited by the antagonist naloxone. Moreover, a chemotactic activity has been demonstrated against molluscan immunocytes, even by growth factors such as PDGF and TGF-β1, and interleukins such as IL-1α, TNF-α, and IL-8.

These observations probably have a general relevance, in view of the evidence that opioid neuropeptides exert stimulatory effects on locomotor activity and conformational changes of the mollusk M. edulis and the insect L. maderae. Immunocompetent immunocytes showed flattening elongation and formation of pseudopodia in the presence of opioids.
2.4 Pagocytosis

When the chemoattractant is represented by particulate material such as bacteria, chemotaxis may be followed by engulfment and phagocytosis. All these phenomena are of pivotal importance for nutrition and defense, and are present throughout the animal kingdom. The incubation of immunocytes of P. corneus with bacteria (Staphylococcus aureus) revealed that only cells with the morphology of macrophages were able to engulf bacteria.\(^\text{88}\) The addition of CRH, ACTH (1–4), (1–24), (4–10), \(^\text{89}\) IL-1\(\alpha\), IL-2, TNF-\(\alpha\), \(^\text{90}\) IL-8, \(^\text{82}\) and PDGF-AB and TGF-\(\beta\)\(^1\)\(^\text{84}\) increase the phagocytic activity. ACTH fragments have different effects, depending on the concentration and the species. In particular, chemotactic and phagocytic effects are not directly correlated because peptides that influence cell migration do not always affect phagocytosis. In this context, ACTH (4–9) and (1–17), \(\beta\)-endorphin, and its related fragments influence the chemotactic activity, but not phagocytosis.\(^\text{89}\) Furthermore, the mode of action of an individual peptide or growth factor could be species-specific and dose-dependent.\(^\text{84,90}\)

Moreover, with regard to POMC-products, another important role was designed\(^\text{91}\) in the biological system, consisting of the nematode Schistosoma mansoni (responsible for Schistosomiasis) by polymorphonuclear leukocytes in humans (definitive host of the parasite), and the immunocytes of the mollusk Biomphalaria glabrata (intermediate host). From these investigations emerged the fact that the coincubation of these cells with adult worms of S. mansoni gave rise to the appearance in the culture medium of \(\alpha\)-MSH, an effect due to the conversion of the parasite ACTH by the enzyme-neutral endopeptidase 24.11, present in the molluscan immunocytes or human polymorphonuclear cells. Thanks to the inhibitory role exerted by \(\alpha\)-MSH, these molecules are able to interfere with the immunological functions of the intermediate or final host, and favor the completion of the biological cycle of the parasite.

2.5 Stress Response

The presence of neuropeptides, and, in particular, of ACTH in immunocytes, was at first sight difficult to explain. This molecule has usually been associated with stress, that is, the complex series of responses that the body sets when its balance and internal composition are threatened. In mammals, the stress is expressed fundamentally in the production, by the paraventricular nuclei, of the hypothalamus of CRH, which induces the production of ACTH by the anterior pituitary cells. In turn, ACTH stimulates the synthesis and release of glucocorticoid hormones by the cells of the adrenal cortex. These hormones regulate the activity of the enzyme that catalyzes the final step in the biosynthesis of adrenaline from the adrenal medulla, a phenomenon that is basically under the control of the sympathetic nervous system.\(^\text{92,93}\) What meaning could the presence of ACTH-like molecules in animals have, molecules that do not possess any of the sophisticated organs, such as the hypothalamus, pituitary, and adrenal glands that are the
target points of these hormones? The first point to make in this regard is that, as previously reported, these and other molecules are found—from invertebrates to humans—in cells with phagocytic activity. Lymphocytes, though endowed with POMC-derived molecules, can be retrieved only in vertebrates, and their evolution cannot be traced back to invertebrates. These immunocytes are traceable only in lymphocytes of tetrapods, whereas in teleosts ACTH-responsive corticosteroidogenic cells have been retrieved in other immune-related tissues, such as the head kidney. The biological meaning of these observations is unknown, but potentially very exciting, as it may reflect, at the cellular level, the integration of the neuroendocrine response with adaptive immunity, in addition to the interconnection with the ancestral innate immune responses. The integration between neuroendocrine functions and adaptive immunity is present in vertebrates, but it finds its maximum expression in mammals, where lymphocytes are both immune and neuroendocrine cells. The occurrence of ACTH-like molecules in lymphocytes of anurans may not be by chance. In addition to immunoglobulin class M (IgM) already present in fish, in the amphibians a new class of antibody, IgY, is present; this is considered to be the precursor of mammalian IgG and IgE. The presence of ACTH-like molecules in anuran lymphocytes may represent an example of how a particular gene is expressed in a new cell line, in conjunction with the emergence of a more sophisticated function.

Returning to the specific topic of the mechanisms of the stress response, it was observed that antigenic stimuli causes a release of ACTH, suggesting the existence of an interconnection between stress and immune responses, even in invertebrates. In this respect, initially assessed is the presence of enzymes, such as tyrosine hydroxylase and dopamine β-hydroxylase, responsible for the synthesis of biogenic amines, the last mediators of the stress response. Later, the hemolymph of mollusks was incubated with antigens, such as ACTH and CRH, and the values of biogenic amines were determined by HPCL, both in serum and in immunocytes, in a short period of time, ranging from 0 to 45 minutes. An increase was observed of biogenic amines in the serum, as well as a concomitant decrease in immunocytes, both in the presence of either ACTH or CRH. This rapid phenomenon is probably due to the release of biogenic amines by immunocytes. Given that the incubation of the hemolymph with CRH increases the concentration of ACTH in immunocytes, it can be assumed that the release of biogenic amines follows the sequence CRH-ACTH-biogenic amines. However, the preincubation of the hemolymph with an anti-ACTH antibody does not abolish the phenomenon of the release of biogenic amines, suggesting that CRH could induce a direct release not mediated by ACTH, even if modest. The role of ACTH in the release of biogenic amines seems to be specific because the addition to the hemolymph of other peptides, for example, the opioid β-endorphin, causes no release of amines by immunocytes.
As already described in mammals, the cytokines are involved in the stress response in invertebrates as well. The incubation of molluscan hemolymph with different cytokines, such as IL-1α and -1β, IL-2, and TNF-α and -β, provoked the release of biogenic amines from immunocytes. It is interesting to note that the preincubation of the immunocytes, with all the cytokines tested, exerted an inhibitory effect on the CRH-induced release of biogenic amines. Immunocytochemical and cytofluorimetric studies have demonstrated that the cytokines and CRH probably bind to the same receptor. The competition between CRH and a large number of cytokines supports the idea that invertebrate cytokine receptors show a certain degree of promiscuity. In this context, it should be stressed that the structure of mammalian cytokine receptors has been described as multisubunits, with the same subunit shared by different cytokine receptors. Also, the growth factors PDGF-AB and TGF-β1 have been found capable of regulating biogenic amine production: the former inhibits, whereas the latter stimulates the release of these molecules via the CRH-ACTH-biogenic amine axis.

2.6 The Immune–Mobile Brain

The data reported in the chapter show the surprising capacity of invertebrate immunocytes to combine the typical properties of both immune and neuroendocrine cells. This combination of functions and copresence in the most ancestral cell, that is, the immunocyte, suggests that we use for this cell the term “immune-mobile brain,” coined for the human lymphocyte. The immunocytes from the outset, and consistently over the course of evolution, flank other nervous and endocrine cells because they form a complex network responsible for maintaining homeostasis. Moreover, these considerations allow a great unifying hypothesis between stressors and antigens, assuming that the antigenic challenge is equivalent to a stress, and that the difference between antigen and stressor is often only quantitative and semantic, as illustrated in Fig. 9.4.

![FIGURE 9.4 Immune-neuroendocrine integration in response to stressful agents.](image-url)
3 CONCLUDING REMARKS

On the basis of the reported data, it is possible to make the following assumptions:

1. Invertebrates, despite their apparent simplicity, are capable of very refined performances, both from an immunological and neuroendocrine point of view, such as the ability to distinguish between self and nonself, and the ability to recognize extraneous stimuli and provide complex responses.

2. These animals survive in environments where there are many agents and potentially harmful substances. In this context, they show very efficient forms of defense. Indeed, a complex network of responses has been observed, comprising, among others: chemotaxis, phagocytosis, and the release of biogenic amines and other bioactive mediators. In vertebrates, the response to individual stress factors/antigens becomes more specialized and specific, but the network of responses that is activated remains essentially the same.

3. Most of the molecules that are used to perform this activity and to mediate the necessary cell–cell interactions are apparently present since ancient times, having also been described in unicellular organisms. It can be assumed that many of these molecules are considered in invertebrates to be “defensive molecules,” because they are used to neutralize, directly or indirectly, agents that perturb the body homeostasis. A typical example are the ACTH-like molecules.

4. Same, or relatively similar “defensive molecules” are found in vertebrates. Even in these more advanced life forms, their function remains essentially the same. However, nature seems to have made a new use of these ancient molecules, functions, and organs while evolving toward more complex and centralized systems.

5. The neuroendocrine and immune functions in invertebrates show a partial overlap with those observed in vertebrates being assigned to the same cell type, such as the invertebrate immunocyte. This probably explains why in higher forms of life a specialized cell like the lymphocyte has retained the ability to both produce and respond to neuroendocrine signals.

REFERENCES


Chapter 10

The Immune-Related Roles and the Evolutionary History of Dscam in Arthropods

Sophie A.O. Armitage*, Daniela Brites**
*Institute for Evolution and Biodiversity, University of Münster, Münster, Germany;
**Swiss Tropical and Public Health Institute, Basel, Switzerland

1 THE DOWN SYNDROME CELL ADHESION MOLECULE

The Down syndrome cell adhesion molecule (Dscam) gene was first described in humans associated with defects in the nervous system.1 Subsequently, several members of the Dscam family were discovered in other metazoans, in which its main described function is related to the development of the nervous system.2–6 Both vertebrates and insects have Dscam members that have resulted from whole gene duplications, for example, DSCAM and DSCAM-like in humans,1,5 and Dscam, Dscam2, Dscam3, and Dscam4 in Drosophila melanogaster.7 These proteins are typically cell-surface receptors composed of 9(Ig)-4(FN)-1(Ig)-2(FN), where Ig stands for immunoglobulin, and FN for fibronectin type III domain (Fig. 10.1). The 16 extracellular domains are usually followed by a transmembrane domain and a cytoplasmic tail. In insects and crustaceans (pancrustaceans), one Dscam paralog is the most remarkable example known of protein diversification by duplication and alternative splicing.2 We will refer to this gene hereafter as Dscam-hv (Dscam hypervariable).8 Other synonymous notations have been used by different authors, that is, Dscam2 and the most recent D. melanogaster notation Dscam1.6,9 Furthermore, species-specific notations have also been used, for example, Anopheles gambiae Dscam has been named AgDscam,10 etc.

In the following sections, we will introduce the organization of the Dscam-hv gene and how protein diversity is generated from this single locus. We will also introduce what is known about the Dscam-hv protein structure and discuss the hypothetical implications of isoform diversity for immunity. Dscam-hv is part of a large gene family, which in arthropods has diversified independently through different molecular mechanisms. To better understand the origin and the importance of Dscam-hv diversity in arthropods, we will briefly review the
inferences gained on the evolutionary history of the molecule from comparative analyses.

### 1.1 From One Gene to Thousands of Protein Isoforms

In Dscam-hv, half of the second and third Ig domains (Ig2 and Ig3) and the complete Ig7 domain are encoded by exon duplicates arranged in three clusters in the Dscam locus (Fig. 10.1). This is made possible by a refined mechanism of mutually exclusive alternative splicing that ensures that in the mature mRNA only one exon per cluster is present (Fig. 10.1). Furthermore, in some species, the diversity of Dscam-hv Ig2, Ig3, and Ig7 domains can be coupled with diversity generated by mutually exclusive alternative splicing of two exons encoding a transmembrane domain, and by regular splicing of different cytoplasmic tails. In this way, one single Dscam-hv can produce thousands of protein isoforms within a single individual. The number of isoforms generated varies, depending on the species, ranging from around 10,000 in *Daphnia* species to around 100,000 in Drosophilids.

Dscam-hv plays an essential role during the development of the nervous system wiring in *D. melanogaster*. The basis of that role is the homophilic binding between identical Dscam-hv isoforms, allowing nervous cells to recognize each other, and leading to a self-avoidance behavior. Interestingly, the large Dscam isoform diversity has been shown to be essential for the correct development of the nervous system in flies, suggesting that isoforms are not functionally redundant. Homophilic binding between identical isoforms has been demonstrated in vitro, indicating a degree of binding specificity, in which 95% of all isoforms will bind only to other identical isoforms.

Dscam-hv has also been shown to be involved in the immune defense of insects and crustaceans. However, its role in immunity remains controversial, because, despite several lines of evidence placing Dscam-hv as a receptor contributing to immune recognition, many important questions remain open. Evidence (or lack thereof) for a role of Dscam in pancrustacean immunity will be discussed in detail in Section 2.

### 1.2 Structural Aspects of Dscam

The protein structure of the first eight Ig domains of Dscam has been elucidated. The first four Ig domains adopt a so-called horseshoe conformation. The horseshoe conformation has a similar shape to other cell-adhesion molecules involved both in the nervous system (eg, axonin) and in the immune system (eg, hemolin). The remaining four Ig domains (Ig5 to Ig8) provide Dscam with a serpentine shape (S-shape) (Fig. 10.1). Homophilic binding between identical isoforms occurs through the formation of Dscam dimers. Interestingly, the dimer-binding regions are segments of Ig2, Ig3, and Ig7 domains coded by the alternative exons. In this way the genetic diversification caused by
FIGURE 10.1  Dscam-hv in D. melanogaster. (A) Dscam-hv protein and mRNA; (i) protein domains. Ig stands for Immunoglobulin and FNIII stands for fibronectin domains, respectively. Colors denote part of the molecule diversified by mutually exclusive alternative splicing; (ii) mRNA, by mutually alternative splicing only one exon of array of exons 4, 6, 9, and transmembrane domain is present; (iii) arrays of duplicated exons present in the locus Dscam-hv of D. melanogaster; (iv) number of Dscam-hv isoforms of Dscam-hv in D. melanogaster, which results both from mutually alternative splicing and regular alternative splicing; (B) example of one mRNA produced by mutually exclusive alternative splicing; (C) model based on Dscam 1-8 crystal structure for the conformation of the first seven Ig domains in Dscam-hv; (v) examples of Dscam-hv monomers; the first four Ig domains form a horseshoe structure, whereas the remaining domains remain flexible; (vi) upon homophilic binding between identical isoforms, mediated by the regions of Ig2, Ig3, and Ig7 encoded by the alternative exons, the Dscam-hv dimer acquires an S shape. (Source: Adapted from Hattori et al.11)
the duplications, coupled with the strong specificity of Dscam’s homophilic binding, provide a highly diverse “lock and key” system, which nerve cells exploit extensively.\(^1^1\) The implications of these structural features for immunity have not been experimentally tested. Nevertheless, it has been suggested that certain variable regions within the horseshoe structure that are not involved in the formation of dimers could be responsible for recognizing pathogen-associated molecular patterns.

2 IMMUNE-RELATED ROLES OF DSCAM-HV IN ARTHROPODS

Despite lymphocytes being present only in vertebrates (see chapters: Lymphocyte Populations in Jawless Vertebrates: Insights Into the Origin and Evolution of Adaptive Immunity; The Evolution of Lymphocytes in Ectothermic Gnathostomata), there is mounting evidence that some species of pancrustaceans exhibit immune phenomena thought to be unique to vertebrates, that is, increased protection against pathogens upon a secondary encounter with the same pathogen\(^2^6–^2^9\) (reviewed recently).\(^3^0\) In some cases, the responses found had revealed a high degree of specificity, implying the ability to distinguish between Gram-positive and Gram-negative bacteria, between different species of the same genus of parasite,\(^2^7\) or even between strains of the same parasite.\(^2^6,^2^9\) A comprehensive view of the immune functions underlying such responses is lacking, but there is evidence in different taxa for an involvement of phagocytosis.\(^2^8,^2^9\) Dscam’s hypervariability, combined with the first demonstrations that it is involved in insect immunity,\(^1^0,^1^9\) has sparked a great deal of interest and discussion\(^3^1–^3^4\) about whether the variable isoforms produced by this molecule might provide antigen-receptor diversity, hypothetically allowing a host to discriminate between different parasites and pathogens. The same commentaries speculated that Dscam-hv could be a convergent insect example of the diversity produced by vertebrate antigen-receptors,\(^3^1–^3^4\) even though it produces diversity of many orders of magnitude lower than vertebrate antibodies, and is mechanistically unrelated. At present this exciting idea lacks empirical support,\(^2^0,^3^5\) and as we shall see, we are still far from understanding whether that could in fact happen.

2.1 What Is the Evidence That Dscam-hv is Involved in the Immune defence of Pancrustaceans?

There are two recent reviews on pancrustacean Dscam-hv in immunity\(^2^0,^2^1\) and we try not to cover the same ground in detail. Similarly to Ng et al.,\(^2^1\) we have taken a parasite class-based approach, although we present this information in the form of a table (Table 10.1), splitting the evidence into that which has been found on the RNA and on the protein level. We briefly discuss herein the general trends that emerge from the data in Table 10.1.
<table>
<thead>
<tr>
<th>Parasite/Stimulant</th>
<th>Host</th>
<th>Experimental Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Viruses</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Drosophila</em> C Virus</td>
<td>I</td>
<td><em>Drosophila melanogaster</em> Adu</td>
</tr>
<tr>
<td><em>P. interp.</em> g. v.</td>
<td>I</td>
<td><em>D. melanogaster</em> Adu</td>
</tr>
<tr>
<td>White spot syndrome virus (WSSV)</td>
<td>C</td>
<td><em>Litopenaeus vannamei</em> NI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adu</td>
</tr>
</tbody>
</table>
**TABLE 10.1** Overview of Studies Examining Dscam-hv in Relation to Immunity, Sorted by Parasite or Immune Stimulant Used (cont.)

<table>
<thead>
<tr>
<th>Parasite/Stimulant</th>
<th>Host</th>
<th>Experimental Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constitutive Exon mRNA Expression</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Species</td>
<td>b</td>
<td>Species</td>
</tr>
<tr>
<td>a</td>
<td>Species P</td>
<td>b</td>
</tr>
<tr>
<td>C</td>
<td>P. leniusculus</td>
<td>Int</td>
</tr>
<tr>
<td>WSSV (Ev)</td>
<td>C</td>
<td>L. vannamei</td>
</tr>
</tbody>
</table>

**Bacteria**

| G− Al | E. coli | I | Anopheles gambiae | Adu | RNAi knockdown; after infection ↓ adult survival; Exon 4.8 knockdown ↓ survival compared to Exon 4.1 (Inj) | | |
| | | | | | | | |
| I | D. melanogaster | NA | | | | Recombinantly expressed Dscam-hv bound to live E. coli | 18 |
| | | Adu | | | | | |

Whole body (Adu: Inj: 6 and 30 h) or haemolymph, head, remainder of body (Adu: Sep: 6 h); Illumina RNA sequencing of Ig2 and Ig3; no change | | | | 95 |
<table>
<thead>
<tr>
<th>Parasite/Stimulant</th>
<th>Host</th>
<th>Experimental Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E. coli</strong> RFP</td>
<td>Adu and Cel</td>
<td>No change; (Adu: Sep: 18 h; Cel: 12 and 18 h) Whole-body minus head; PacBio RNA sequencing of Ig2, Ig3 and Ig7: no change; (Adu: Sep: 18 h; Cel: 12 and 18 h)</td>
</tr>
<tr>
<td><strong>D. melanogaster</strong> Emb</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Serratia symbiotica</strong> I</td>
<td>Acyrthosiphon pisum Juv</td>
<td>Whole body; transcriptome; no changes. (stably infected aphid line) Haemocytes of Dscam loss of function embryos bound E. coli similarly to wild type (Inj)</td>
</tr>
<tr>
<td><strong>Vibrio harveyi</strong> C</td>
<td>L. vannamei NI</td>
<td>Haemocyte; 1 h and 36 and 48 h. Cytoplasmic tail variants regulation (Inj: 0, 1, 3, 12, 18, 24, 36, 48, 72 h) Haemocyte; Sequencing cDNA of Ig2 and Ig3: Association between sequence of isoforms and bacterial challenge (Inj: 36 h)</td>
</tr>
<tr>
<td>G— De</td>
<td>E. coli I</td>
<td>A. gambiae Adu and Cel</td>
</tr>
</tbody>
</table>

(Continued)
### TABLE 10.1 Overview of Studies Examining Dscam-hv in Relation to Immunity, Sorted by Parasite or Immune Stimulant Used (cont.)

<table>
<thead>
<tr>
<th>Parasite/Stimulant</th>
<th>Host</th>
<th>Experimental Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E. coli</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alexa Fluor® 595</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>D. melanogaster</strong></td>
<td>Adu and Cel</td>
<td>Dscam-hv RNAi and loss of function larvae had reduced phagocytosis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S2 cells preincubated with Dscam-hv antibody had reduced phagocytosis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>E. coli</strong> Fluoroscein conjugate</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dscam-hv knockdown in Sua5B cells resulted in lower phagocytosis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pseudomonas veronii</strong></td>
<td>A. gambiae</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cel</td>
<td>RTqPCR of Ig2 variants showed they were differentially expressed compared to a naïve control (12 h)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNAi of exon 4.8 binding of Dscam-hv to bacteria compared to exon 4.1 and GFP control</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bacteria had affinity to membrane-bound and secreted Dscam-hv produced by cells previously challenged with the same bacteria</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Vibrio harveyi</strong></td>
<td>L. vannamei</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>1 of 2 V. harveyi-induced RE isoforms bound more strongly to V. harveyi than 2 control-induced RE isoforms. All isoforms bound (to differing degrees) to WSSV, E. coli and Staphylococcus aureus</td>
</tr>
<tr>
<td>G−Mi</td>
<td>LPS E. coli</td>
<td>C</td>
</tr>
<tr>
<td>------</td>
<td>------------</td>
<td>-------</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>P. leniusculus</td>
</tr>
<tr>
<td></td>
<td>LPS P aeruginosa</td>
<td>I</td>
</tr>
<tr>
<td>G+Al</td>
<td><em>Micrococcus luteus</em></td>
<td>I</td>
</tr>
<tr>
<td></td>
<td><em>Pasteuria ramosa</em></td>
<td>C</td>
</tr>
<tr>
<td></td>
<td><em>P. ramosa: 2 strains</em></td>
<td>C</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Parasite/Stimulant</th>
<th>Host</th>
<th>LHS</th>
<th>Constitutive Exon mRNA Expression</th>
<th>Alternatively Spliced Exon mRNA Expression</th>
<th>RNAi/Mutagenesis</th>
<th>Cellular/Binding Assays</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>a Staphylococcus aureus</td>
<td>I A. gambiae</td>
<td>Adu</td>
<td>RNAi ↓ adult survival after infection. Knockdown of exon 4.1 ↓ survival after infection compared to 4.8 knockdown (Inj)</td>
<td></td>
<td></td>
<td>Dscam-hv was highly concentrated at site of interaction with S. aureus in challenged cells</td>
<td>10</td>
</tr>
<tr>
<td>b S. aureus</td>
<td>I A. gambiae</td>
<td>Adu and Cel</td>
<td>RTqPCR analysis of Ig2 variants in adults and cell line: differentially expressed compared to control. Cells: splicing patterns at 12 and 18 h correlated with each other (Adu: NI: 12 h; Cel: 12 and 18 h)</td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>G+ De S. aureus Fluoroscein conjugate</td>
<td>I A. gambiae</td>
<td>Cel</td>
<td>↓ Phagocytosis in after knockdown in Sua5B cells</td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>G+Mi PGN from S. aureus</td>
<td>I A. gambiae</td>
<td>Cel</td>
<td>RTqPCR of Ig2 variants: differentially expressed compared to control (12 h)</td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>I</td>
<td>Host</td>
<td>Parasite/Stimulant</td>
<td>Experimental Evidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td><em>E. sinensis</em></td>
<td>Adu</td>
<td>↑6 h; ↓24 h (Inj: 2, 6, 12, 24 h)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td><em>P. lenisculus</em></td>
<td>Int</td>
<td>Haemocytes; no change (Inj: 6, 12, 24 h)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G− and G+ Al</td>
<td><em>E. coli</em> and <em>Pilibacter termitis</em></td>
<td>I</td>
<td>Coptotermes formosanus</td>
<td>Whole bodies; no change (Sep: 24 h)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G− and G+ Mi and ?</td>
<td>Mixed</td>
<td>I</td>
<td><em>Manduca sexta</em></td>
<td>Juv</td>
<td>Fat body and haemocytes; whole genome transcriptome ↑ (Inj: 24 h)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>Opportunistic bacteria</td>
<td>I</td>
<td><em>A. gambiae</em></td>
<td>Adu</td>
<td>↑ In opportunistic bacteria within the mosquito haemolymph 3 days after Dscam-hv RNAi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td><em>Anopheles stephensi</em></td>
<td>Adu</td>
<td>↓ in bacteria in the midgut after overexpressing splice variants from <em>A. gambiae</em> that are increased in expression toward bacteria and <em>Plasmodium</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Parasite/Stimulant</th>
<th>Host</th>
<th>Experimental Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fungi</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al</td>
<td><em>Beauveria bassiana</em></td>
<td>I</td>
</tr>
<tr>
<td>De</td>
<td><em>B. bassiana</em></td>
<td>I</td>
</tr>
<tr>
<td></td>
<td><em>S. cerevisiae</em></td>
<td>I</td>
</tr>
<tr>
<td>Mi</td>
<td>β-1,3-glucan</td>
<td>C</td>
</tr>
<tr>
<td>Sp</td>
<td><em>Nosema ceranae</em></td>
<td>I</td>
</tr>
<tr>
<td>Protozoa</td>
<td>Al</td>
<td>Crithidia mellificae</td>
</tr>
<tr>
<td>----------</td>
<td>----</td>
<td>---------------------</td>
</tr>
<tr>
<td>Al</td>
<td>C. mellificae and N. ceranae</td>
<td>I</td>
</tr>
<tr>
<td>Al</td>
<td>Plasmodium berghei</td>
<td>I</td>
</tr>
<tr>
<td>I</td>
<td>A. stephensi</td>
<td>Adu</td>
</tr>
<tr>
<td>Plasmodium falciparum</td>
<td>I</td>
<td>A. gambiae</td>
</tr>
</tbody>
</table>

(Continued)
### TABLE 10.1 Overview of Studies Examining Dscam-hv in Relation to Immunity, Sorted by Parasite or Immune Stimulant Used (cont.)

<table>
<thead>
<tr>
<th>Parasite/Stimulant</th>
<th>Host</th>
<th>Experimental Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>a Species P</td>
<td>b Species</td>
<td>Constitutive Exon mRNA Expression</td>
</tr>
<tr>
<td></td>
<td>LHS</td>
<td>Dscam-hv deficient mosquitoes: number of oocysts in midgut ↑ cf. control injected mosquitoes. Knockdown of exon 4.1 ↑ oocyst load, but silencing <em>P. berghei</em> infection-responsive exon 4.14 did not affect susceptibility to <em>P. falciparum</em> (Ora)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Whole body cDNA sequencing: combination of Ig2 and Ig3 splice variants more diverse after infection; Multiple <em>P. falciparum</em> genotype infections meant more diverse splice variants compared to a single genotype. (Ora: 24 h)</td>
</tr>
</tbody>
</table>

---

**TABLE 10.1** Overview of Studies Examining Dscam-hv in Relation to Immunity, Sorted by Parasite or Immune Stimulant Used (cont.)
| I  | A. stephensi | Adu | Infection intensity in midgut ↓ in transgenic mosquitoes overexpressing two splice variants that in A. gambiae are induced after infection with P. falciparum (Ora) |
| Plasmodium gallinaceum | Aedes aegypti | Adu | Fat body; genome wide microarray; ↑two Dscam genes (Ora: 24 h) | Antibody staining showed ↑ association of AgDscam-hv with P. falciparum ookinetes in transgenic mosquitoes expressing splice variants that in A. gambiae are induced after infection with P. falciparum (Ora) |

The nature of the experimental evidence is split into four categories.

a: Phylogenetic information on parasite or immune stimulant is shown in bold. If eukaryotic or prokaryotic parasites were used, they were administered: Al, alive; De, dead; Mi, microbe associated molecular pattern (MAMP); Sp, spores; and if bacteria whether: G-, Gram negative or G+, Gram positive. †, Information regarding whether the immune stimulants used were alive or dead was not explicitly given; NA, not applicable.
Species P: The species of parasite or immune stimulant used. Abbreviations include: Ev, recombinantly expressed envelope proteins; Mixed, Curdlan (β-1,3-glucan) from Alcaligenes faecalis, E. coli and M. luteus; PGN, peptidoglycan; P. interp. g. v., Plodia interpunctella granulosis virus; P. aeruginosa, Pseudomonas aeruginosa; RFP, expressing red fluorescent protein; S. cerevisiae, Saccharomyces cerevisiae.
b: The subphyla of arthropod host used: C, crustacean; I, insect.
LHS, Life history stage; Adu, adult; Cel, cell culture; Emb, embryo; Int, intermoul; Juv, juvenile; NA, not applicable; NI, no information.
Experimental evidence: in parentheses at the end of comments, information about the way in which the parasites were administered: Inj, injection; Ora, oral infection; Sep, septic wound. Also stated, time post infection at which Dscam-hv mRNA was assayed; NI, no information. † Indicates increase and ↓ indicates decrease. R, references.
Modified from Ref. [20].
When considering functional categories of the parasite, it is only microparasites, that is, those up to a few hundred micrometers in size, including viruses, bacteria, fungi, and protozoa, which have been tested in relation to Dscam-hv and immunity. The majority of the tested species are bacteria, which have been introduced into the host in either a live or a dead state, and mainly through the cuticle via injection or with a needle covered in a bacterial solution. Moving to the hosts, a handful of crustacean and insect species have been tested to date, with particular emphasis on the whiteleg shrimp, Litopenaeus vannamei, and the mosquito, A. gambiae. Furthermore, it is largely cell lines or the adult, rather than the juvenile, life-history stage that has been tested. It is not always straightforward to make generalizations on the behavior of Dscam-hv from the available data (Table 10.1); for example, if one considers whether general Dscam-hv mRNA expression changes after exposure to a parasite or immune stimulant, one observes that of the 18 examples cited (Table 10.1), 7 show no modulation (increased or decreased transcription rate) of Dscam-hv at any of the time points studied. Studies that have reported upregulation in mRNA transcripts of Dscam-hv upon pathogen exposure, have shown relatively small (<10-fold), or sometimes stronger (10–100-fold), changes in relative expression, in terms of effect size, when compared to control groups. Furthermore, of the species where expression is changed, this can be early (eg, 2 h), or relatively late (eg, 5 days), after exposure. Clearly, interpretations are limited to the time points studied; in the majority of cases, a time-span between 2 h and 5 days has been used.

It might be the case that in addition to/instead of quantitative changes in Dscam-hv transcription after infection, that there are qualitative changes in the expression of exon repertoires. To date, A. gambiae has shown the most promising evidence to support pathogen-specific splice-variant expression after exposure to parasites. Dead bacteria introduced into the hemocoel, as well as live-ingested Plasmodium species, elicited variability in the expression of different splice variants from Ig2.10 Staying with A. gambiae, multiple genotype infections of live P. falciparum resulted in more diverse Ig2 and Ig3 variants, compared to infection with a single parasite genotype.43 However, RNA sequencing of splice variants after exposure of D. melanogaster or Daphnia magna to live bacteria did not result in strong variation in the expression of exons from Ig2, Ig3, or Ig7.43,44

Dscam-hv knockdown via RNAi or mutagenesis has mainly been used in D. melanogaster and A. gambiae: in both species it has helped to elucidate the fact that Dscam-hv is involved in phagocytosis of dead bacteria.10,19 However, it is interesting to note that D. melanogaster Dscam-hv loss-of-function embryos showed no impairment in binding live E. coli compared to wild-type embryos.45 Once again, A. gambiae showed strong support for parasite-mediated responses in Dscam-hv, and indeed the highest level of specificity/involvement of particular exons to certain parasites: RNAi of A. gambiae exon 4.8 resulted
in reduced adult survival after infection with live E. coli, but not after infection with S. aureus, and vice-versa results were found for exon 4.1. This is the only published study that we are aware of to-date to have examined whether Dscam-hv is important for survival in the face of pathogen challenge. The same research group more recently inserted Dscam-hv splice variants from A. gambiae, which were induced after infection with P. falciparum, into A. stephensi: this resulted in a reduction of infection intensity in A. stephensi midgut after P. falciparum infection. These results and similar earlier experiments on A. gambiae are also the only data to-date indicating a direct role for Dscam-hv in parasite resistance.

Most experiments described in Table 10.1 have attempted to quantify expression of Dscam mRNA. However, it is becoming increasingly evident that the correlation between mRNA and protein abundance varies widely across genes. Only a limited number of studies have addressed Dscam-hv on the protein level (Table 10.1), which we will discuss in more detail in the following section. The ways in which Dscam-hv interacts with other known molecules in the immune system has been reviewed in Ref. [21] so it will not be covered here.

### 2.2 How Has Dscam-hv Been Predicted to Function in Arthropod Immunity?

The initial findings from D. melanogaster suggested that Dscam-hv is used during phagocytosis of pathogens, and it was proposed that it might be a signaling receptor or coreceptor during this process. The same authors showed that when combinations of three variable Ig domains were recombinantly expressed that this was sufficient to bind live E. coli, which led to the hypothesis that secreted isoforms might opsonize pathogens. The experiments by Dong et al. led them to the conclusion that Dscam-hv in A. gambiae acts as a hypervariable pattern-recognition receptor, and among many other results, they found that when confronted with a pathogen, Dscam-hv responds by producing distinctive splice variants for the different pathogens tested. In the following we discuss some of the ideas that have been stimulated from the aforementioned two studies, and discuss other more recent observations and empirical data regarding how Dscam-hv could and does function mechanistically in an immunity context.

#### 2.3.1 Is There an Amplification of Specific Dscam-hv Isoforms in Response to an Immune Challenge?

Let us now make the assumption that Dscam-hv isoforms bind parasites with some degree of specificity. What concentrations of each isoform would be necessary to elicit an efficient immune response? The number of isoforms expressed by hemocytes and by the fat body compared to nervous system tissues
is considerably reduced in *D. melanogaster* and in *Da. magna*. For instance, populations of *D. melanogaster* hemocytes can express the vast majority of alternative variants from exons 4 and 6, but unlike, for example, the larval brain, only five exons from the exon 9 cluster are highly expressed. Our rough calculations based on the numbers of *D. melanogaster* exons from the three extracellular Ig domains that were found to be expressed by hemocytes and fat bodies in microarray studies, suggest that a subset of just over half of the estimated 18,496 isoforms would actually be found in these tissues. Individual *D. melanogaster* hemocytes subjected to RT-PCR followed by a microarray, showed that each of the limited number of hemocytes tested expressed at least two to fourteen different isoforms of Dscam-hv. A restricted isoform repertoire of hemocytes could result in higher concentrations of each isoform, thus increasing their functional capacity. Yet, would a limited number of hemocytes expressing a few different isoforms have a meaningful immune action? How would this signal be amplified to reach a “critical amount”? It has been suggested that there could be proliferation of hemocytes expressing the “appropriate” Dscam-hv variants related to a particular antigen. There is evidence that in *D. melanogaster* adults, hemocytes can proliferate after infection, but it is unknown whether these cells would have specific properties enhanced by the elicitor. If hemocytes were to collectively change transcription, how would this be signaled? Although there is some evidence that whole-gene Dscam-hv transcription increases after immune challenge (Table 10.1), and that in mosquitoes there is amplification of specific Dscam-hv Ig2 variants, amplification of specific Ig3 and Ig7 variants has not been demonstrated. Furthermore, how amplification of specific splice variants might occur remains an open question.

### 2.3.2 Soluble Forms of Dscam-hv Occur in the Hemolymph as the Result of an Active “Shedding” Process, and These Forms Act as Opsonins

Watson et al. observed a Dscam-hv protein of reduced molecular weight compared to the whole protein in *D. melanogaster* S2 cell conditioned-medium and in hemolymph serum, although the sizes of the proteins indicate something smaller than what might be predicted, if only the cytoplasmic tail and the transmembrane domain had been left behind after proteolytic cleavage of the extracellular domain. The shrimp species, *L. vannamei* and *Penaeus monodon*, and the Chinese mitten crab, *E. sinensis*, have Dscam-hv variants that lack the transmembrane domain and cytoplasmic tail, respectively. Rather than undergoing proteolytic cleavage, they are therefore suggested to be expressed directly in a secreted form. Interestingly, *L. vannamei* has been shown to upregulate expression of the tailless Dscam-hv form after infection with WSSV (Table 10.1). In their model for how Dscam-hv may trigger phagocytosis, Ng et al. thus postulated that a pathogen might stimulate proteolytic cleavage, or stimulate the direct secretion of tailless Dscam.
It is unknown whether the secreted Dscam-hv acts as an opsonin, that is, binds to an antigen and in this way boosts phagocytosis, although some recombinantly expressed Dscam-hv isoforms have been shown to bind to bacteria.

2.3.3 Membrane-Bound Dscam-hv Acts as a Receptor for Phagocytosis

Dscam-hv RNAi, or incubating cells with a Dscam-hv antibody, have been shown to reduce phagocytosis of dead *E. coli* in *D. melanogaster* larvae and S2 cells (Table 10.1); furthermore, mosquito Sua5B cells showed reduced phagocytosis of dead *E. coli* and *S. aureus* (Table 10.1). Dscam-hv has also been implicated in phagocytosis of the same two bacteria species as in *P. leniusculus*. Dscam-hv would presumably be predicted to be involved in the phagocytosis of live and nonfluorescently labeled parasites in addition to dead and labeled bacteria, although this has not yet been tested. It has been proposed that membrane-bound Dscam-hv acts as a receptor for phagocytosis, either by directly interacting with a pathogen or by interacting with a *Dscam*-hv protein that is already bound to a pathogen.

2.3.4 How Could Dscam-hv Bind to Antigens?

The first four Ig domains of Dscam-hv have been referred to as having a horseshoe conformation (Section 1.2). In the cell-adhesion molecule hemolin, the horseshoe structure has been demonstrated to create a binding site to bacterial lipopolysaccharides. In Dscam-hv, parts of the Ig2 and Ig3 domains encoded by the exons 4 and 6 form two surface epitopes on either side of the horseshoe structure. One of those epitopes (epitope I) is essential for the formation of Dscam-hv dimers, that is, for homophilic binding, which mediates cell-to-cell recognition in the nervous system. The other epitope (epitope II) is oriented toward the external environment of the Dscam-hv dimer. When comparing individual alternatively spliced exons between *Drosophila* species (orthologous exons), the amino-acid sequence of the region involved in homophilic binding, that is, epitope I, is conserved, yet the sequences of epitope II show little conservation, suggesting faster sequence evolution, perhaps as a result of the interaction with antigens. In contrast, comparisons within *D. melanogaster* duplicated exons (paralogous exons) within each cluster show that the amino-acid sequence of alternatively spliced isoforms is diverse at both epitopes I and II. These two findings suggest that both epitopes I and II diverged extensively before speciation, but at some point during the evolution of Dscam-hv exons stopped diverging, with the exception of the regions encoding epitope II. If parasite-induced selection were the main reason for the sequence divergence found in epitope II regions, signatures of molecular evolution compatible with positive selection would be expected, and have indeed been found for other genes involved in insect immunity. Such signatures have been inferred for several *Drosophila* species and for the crustaceans *Daphnia pulex* and *Da. magna* for
both epitopes I and II. The results did not unequivocally suggest that positive selection was the main force that maintained diversity of epitope II, yet the number of available sequences was low, precluding robust tests. Intraspecific comparisons denoted an excess of nonsynonymous polymorphisms in epitope II regions encoded by exon 6, which would be compatible with selection mediated by parasite recognition. However, those polymorphisms were mostly present at low frequencies in the populations, and could therefore be interpreted as the result of a relaxed selection acting on epitope II coding sequences.

To date, the above elegant solution for how Dscam-hv might simultaneously function in heterophilic and homophilic binding, lacks empirical evidence showing that epitope II is involved in heterophilic interactions. In the nervous system, only the molecule Netrin engaging in heterophilic interactions with Dscam-hv has been characterized, but the exact binding sites are not known.

3 THE EVOLUTIONARY HISTORY OF DSCAM-HV IN ARTHROPODS

In this section, we will briefly review how comparative evolutionary analysis of Dscam genes across different organisms has contributed to our understanding of how Dscam-hv has evolved, and to our understanding of the parts of the molecule that seem to be under stronger functional constraints, or alternatively, evolve faster. Such insights may provide us with information regarding which parts of Dscam-hv are more likely to be important in the interaction with parasites.

3.1 The Origin of Dscam

Cell adhesion molecules were needed early in evolution for intercellular cohesion and communication in multicellular organisms. Throughout the evolution of metazoans, cell adhesion molecules (CAMs) have been recruited for many different cellular functions, such as cell proliferation and differentiation, apoptosis, migration, and parasite recognition. Many members of this family are at least in part built from Ig domains, and several show relatively high molecular diversity that is also associated with alternative splicing. Similar to Dscam, the tertiary conformation of the first four Ig domains of several CAMs is that of a horseshoe shape (Fig. 10.1), which creates, as mentioned previously, singular adhesive properties by allowing homophilic and heterophilic adhesion to similar and different proteins, respectively. The appearance of this structure might have allowed for the expansion of a subfamily of CAMs used by nervous system cells, such as, among others, axonin, roundabout, contactin, and Dscam, and by immune system cells such as hemolin and Dscam. Precursors of Dscam could have been present before the evolution of the Bilateria, for example, although lacking a canonical Dscam organization, some genes in the cnidarian Nemastostella vectensis (Fig. 10.2) and in the sponge Amphimedon
queenslandica have Ig domains with strong amino-acid conservation when compared to Dscam. In addition, cytoplasmic motifs such as SH2, ITIM, and polyproline motifs are conserved between the NV 1 protein of N. vectensis and the human Dscams, indicating that they use similar signaling pathways. This suggests that some of the Dscam features characteristic of complex groups such as vertebrates might have already evolved in early metazoans. The phylogenetic reconstruction of available Dscam molecules from the main animal groups suggests that the ancestral forms of Dscam were used by the nervous system.
system and that its usage by immune cells is a derived state.\textsuperscript{67,69} The usage by immune system cells seems to have evolved concurrently with mechanisms of molecular diversification, such as duplication and mutually exclusive alternative splicing in Dscam. However, care must be taken because this association does not necessarily imply causality. Hemocytes have several important functions unrelated to immunity throughout the embryonic development of insects.\textsuperscript{70} Dscam could therefore have an ancestral role of mediating cellular recognition. Light will only be shed on the ancestral functions of Dscam by performing more studies addressing its role in hemocytes or hemocyte-related cells.

\section{Dscam Radiation in Arthropods}

Arthropods are the most diverse group of animal phyla, with an enormous variety of morphologies and ecological niches.\textsuperscript{71} This diversity is grouped into four major extant groups: insects and crustaceans, which are considered sister groups (pancrustaceans),\textsuperscript{72} myriapods and chelicerates.\textsuperscript{73,74} The completion of genome-sequencing projects of the centipede \textit{S. maritima}, a myriapod, and several chelicerate species\textsuperscript{75,76} have revealed that Dscam has also diversified in these groups. Strikingly, in the species of myriapod and chelicerate studied to date, Dscam diversification has occurred through a different mechanism than that of pancrustaceans, that is, via dozens of whole-gene duplications of Dscam (Fig. 10.2).

By superimposing the genealogy of the Dscam gene family in arthropods, on the consensual reconstructions of the phylogenetic relationships between the main arthropods groups, it becomes evident that the most basal arthropod Dscams are the ones of the chelicerates.\textsuperscript{67,77} In the tick, \textit{I. scapularis}, there are around 27 genes with strong similarity to Dscam genes. The \textit{I. scapularis} genes do not have a canonical domain organization but are clearly Dscam homologs, that is, they share a common ancestor with other metazoan Dscams.\textsuperscript{67} Furthermore, the genome of \textit{I. scapularis} has incomplete Dscam proteins containing duplicated exons coding for Ig7 and Ig8 (Fig. 10.2). Five additional chelicerate species have since been predicted to have between 4 and 35 Dscam gene copies.\textsuperscript{76} In the myriapod \textit{S. maritima}, 60–80 duplications of whole Dscam genes have occurred (Fig. 10.2).\textsuperscript{67} Unlike what has been found in \textit{I. scapularis}, a high number of these Dscam paralogs have a canonical domain organization.\textsuperscript{67} In addition to whole-gene duplications, approximately half of the \textit{S. maritima} paralogs have internal duplications of the exons coding Ig7, but not of exons coding for Ig2 and Ig3 Dscam domains (Figs. 10.2 and 10.3).\textsuperscript{67} What is the relationship between Dscam-hv and the remaining paralogs of the pancrustaceans, and the Dscams of the myriapods and chelicerates? These Dscam sensu lato sequences have diverged extensively and confident phylogenetic inferences of the relationships among Dscam members of pancrustaceans and myriapods and chelicerates are difficult to infer.\textsuperscript{67,69,77} However, some patterns can be drawn: (1) the extensive whole-gene duplications of \textit{I. scapularis} and \textit{S. maritima} probably originated
independently, or they were inherited from a common ancestor and diverged so extensively that a common origin can no longer be ascertained; (2) the domain organization of the Dscam genes of myriapods are more similar to the ones of pancrustaceans than to the chelicerates, in agreement with the view that myriapods are a basal group to pancrustaceans, sharing a common ancestor with the latter, which is not shared by chelicerates; (3) Dscam Ig7 domains were the first domains to engage in duplications (it is not clear whether Ig7 and Ig8 belong to functional Dscam genes in *I. scapularis*); (4) the three clusters of duplicated
exons coding for half of Ig2 and Ig3 domains are a derived character which appeared for the first time in pancrustaceans; (5) despite generally low conservation between the cytoplasmic tails of pancrustacean Dscam-hv paralogs, all pancrustacean Dscam-hv’s and Dscam genes of *S. maritima* shown to be expressed by hemocytes, share motifs CC1 (involved in axon guidance signaling pathways and in leukocyte mobility control), which are not in other Dscam paralogs. What about the origins of the Dscam mechanism of mutually exclusive alternative splicing? The Ig7 exon duplications of *S. maritima* Dscam undergo mutually exclusive alternative splicing, indicating that this character evolved before the evolution of Dscam-hv of pancrustaceans. In *I. scapularis* Dscam expression has not been studied, but *I. scapularis* lacks several genes that operate as splicing factors in *D. melanogaster* Dscam. The in silico reconstruction of the Dscam genes found in *I. scapularis* and *S. maritima* suggest that many of these genes are functional. What are the functions of this myriad of Dscam genes? In *S. maritima*, Dscam expression has been observed both in organs of the nervous systems and in hemocytes, consistent with a role both in the nervous and the immune systems (but also generally in development, as referred to previously). Other interesting aspects from the in silico reconstructions are (1) that a high diversity of Dscam forms had already evolved in myriapods and chelicerates; (2) in the tick *I. scapularis* Dscam molecules with alternative transmembrane domains such as in insects were found; (3) Dscam putative soluble forms were found, as is seen in decapod crustaceans. The genetic diversification of Dscam in arthropods might have meant that Dscam acquired new functions in these animals, but detailed functional studies in basal arthropods are needed to address this question. Based on the data available, it seems plausible that the extensive Dscam duplications as seen in chelicerates and myriapods were the raw material from which Dscam-hv evolved in the ancestors of the pancrustaceans.

### 3.3 The Evolutionary History of Dscam-hv in Pancrustaceans

In pancrustaceans the number of Dscam paralogs is much lower than in myriapods and chelicerates, ranging from four to six, depending on the species. Some of these paralogs have been predicted in silico, but expression and functional data also exists for other paralogs. The majority of paralogs have been described for insects; in crustaceans they have only been described for *D. pulex* (but no other full crustacean genome sequence is available). Of all pancrustacean Dscams, only the hypervariable gene Dscam-hv and another member called Dscam2 (AbSCAM in *Apis mellifera*), which has two exon duplications encoded in Ig7, have duplications of alternative exons. Both Dscam-hv and Dscam2 are essential for the correct development of the nervous system in flies. These are also the only two Dscam paralogs for which a common ancestry to all pancrustacean species with available genomic sequences can be found. Given the situation described for myriapods, additional Dscam genes were most likely also present in the ancestor of the pancrustaceans, but
these were lost and/or diverged extensively in different pancrustacean groups, maybe because their functions are not as essential as the ones of Dscam-hv and Dscam2. The loss of Dscam paralogs in the ancestor of pancrustaceans, and the appearance of Dscam-hv seems to reflect the evolution toward a more efficient mechanism of regulating the expression of Dscam isoform diversity, that is, expression of a high repertoire of isoforms from a single gene.35

In Dscam-hv across pancrustaceans, the constitutive domains and the domains that result from alternatively spliced exons display considerably different levels of sequence conservation. A comparison between crustaceans and insects reveals high amino-acid conservation of the constitutive domains (around 60% identity over the complete extracellular domains), despite the fact that these groups are predicted to have last shared a common ancestor around 420 million years ago72 The domains that result from exons that undergo alternative splicing are, on the contrary, considerably more divergent than constitutive exons when comparing crustaceans and insects, and even when comparing within certain insect groups.67,77,88 The constitutive and alternative domains of Dscam-hv thus have had different evolutionary histories: whereas the former have evolved mainly under purifying selection probably related to the essentiality of Dscam-hv, the exons encoding the alternative domains have, at least during a certain period, diversified and expanded independently in different pancrustacean groups. It remains unclear if that diversification had been driven by positive selection (ie, was adaptive).56,89 Alternatively, spliced exons could have diverged under a period of relative selective relaxation (“relative” because key residues remained conserved among paralogs), perhaps due to the redundancy of the duplicated sequences.56 A recent study inferred positive selection acting to promote the divergence of Dscam-hv Ig7 orthologous domains in the social insects *A. mellifera*, *Apis florea*, *Bombus impatiens*, and *Bombus terrestris*.88 The test was, however, based inadvertently on alignments of paralogous sequences of Ig7 coding sequences, which have diverged too extensively to be aligned reliably.

Regardless of the context in which diversity among paralogs was generated, Dscam-hv diversity became essential for survival, as referred to already. Accordingly, in extant pancrustacean species, selection acts to preserve Dscam diversity both by keeping the number and the sequences of the alternative exons unchanged, by preventing gene conversion to homogenize exon sequences.8,56,67 However, the extent of the action of purifying selection seems to differ in the three clusters of alternative exons, as we shall discuss next.

### 3.4 Dscam Clusters of Duplicated Exons

The three alternative exon clusters in Dscam-hv have arisen by reiterative exon duplication and deletion.88 For the sake of simplicity, we will use the exon nomenclature of the *D. melanogaster* Dscam-hv, that is, exons 4, exons 6, and exons 9. Generally, the exon duplications of Dscam-hv are believed to be the outcome of homologous recombination among neighboring exons with similar sequence composition.88 A feature common to the exons of the three clusters is
that exons resulting from duplication (paralogs) diverged extensively, whereas exons resulting from speciation (orthologs) are conserved. As an example, the *Drosophila* species or *Daphnia* species that have been examined have their own “set” of alternative exons, indicating that alternative exons evolved independently in different pancrustacean groups. When looking within each taxonomic genus, duplicated exons (paralogs) are quite diverse, but the vast majority of paralogs have an orthologous exon in the other con-generic species (with some exceptions for clusters of exons 6 and 9). This suggests that most exons duplicated and diverged by accumulating mutations in the ancestors of each genus. The high conservation of amino-acid sequences of orthologous exons indicates that they had not diverged much since the split of the extant species from their most recent common ancestor, indicating that selection acts to preserve the ancient diversity that had been created. An exception to this is amino-acid sequence variation in the orthologous regions of exons 4 and 6 that encode epitope II, as discussed previously. Despite this general pattern of sequence evolution, the three clusters of exons seem to have undergone different patterns of exon radiation. *Fig. 10.3* The number of exons in cluster 4 tends to be more conserved among species of the same genus, whereas exons of clusters 6 and 9 seem to have higher duplication rates. This could reflect that the different clusters specialized in different functions prior to species divergence, as suggested by Crayton and coworkers, which is an attractive hypothesis, given the dual role of Dscam-hv in the nervous and immune systems. Alternatively, it could also simply reflect protein structural constraints.

### 3.5 Are the Patterns of Dscam-hv Molecular Evolution Compatible with a Role in Immunity?

Genes involved in the immune system typically present signatures of diversifying selection, for example, the MHC locus presents hallmarks of balancing selection, such as elevated polymorphisms and deeply diverged transpecific alleles. In Dscam-hv, strong signatures of parasite-imposed selection currently acting in populations have not been found. However, perhaps such signatures will not be very apparent within a population because the variability of Dscam-hv is obtained via somatic combinatorial arrangements of variable exons and is thus not inherited. As referred to previously, selection seems to act to preserve ancient Dscam-hv diversity. It is entirely possible that such diversity has been advantageous for the nervous system, yet why would the usage of Dscam-hv in the nervous system select for different “sets” of exons in different animal groups? Why would selection imposed by the nervous system select for soluble forms of Dscam? It is now around 10 years since the discovery that Dscam-hv is involved in insect immunity, and although much progress has been made in deciphering what it might do in the face of pathogen challenge, there are abundant unanswered questions. In the following section we discuss some problems, open questions, speculations/ideas, and potential avenues for future research on Dscam-hv in immunity.
4 EXPERIMENTAL METHODOLOGY AND FUTURE PERSPECTIVES

4.1 Experimental Methodology: Choice of Experimental Host and Parasite/Antigen, and Other Experimental Parameters

Which host species should be used? This will to some degree relate to the question at hand: for example, a large crustacean would be a suitable model for testing whether specific proliferation of hemocytes occurs, and for examining whether regulation of alternative splicing occurs during an immune response since large amounts of hemolymph can be withdrawn. What is the origin of the experimental animals? It will be more controlled to use minimum laboratory F1 generation rather than directly using wild-caught animals with unknown previous exposure to parasites and pathogens. Furthermore, how do we choose the antigenic challenge? For example, does one use an off-the-shelf bacterium, or a microbial derivative that is highly controllable and has been used by others, potentially allowing for comparisons across studies? This approach relies on the assumption that Dscam-hv proteins could hypothetically react against any pathogen. Alternatively, does one use a pathogen that naturally occurs with the host, with the assumption that these pathogens might be ecologically relevant? Or does one use a pathogen that is coevolving with the host species? In this case, one might predict that other host immune-defense genes have been selected to respond to adapting parasite virulence genes. How would Dscam-hv respond in this case? Does Dscam-hv provide a certain degree of specificity toward parasites? If that is the case, then one would expect the Dscam-hv isoform repertoire to be deterministic, and therefore to be predictable in repeated experiments. If the repertoire is arbitrary then it would suggest that Dscam-hv diversity, but not the exact composition of the variable regions, is important. Conversely, if isoform variant expression is related to the infecting parasite, then it would imply that the amino-acid composition of the variable regions is important. If the latter were the case, then site-directed mutagenesis could be employed to test whether binding to the parasite is affected. Further considerations are: which infection doses should be used? This will be linked to the question at hand: to test survival after knockdown, some level of mortality would be desirable. However, when examining mRNA expression, is it more appropriate to have a nonlethal immune challenge that stimulates the immune system, or a strong challenge that causes significant mortality? Should the pathogen be administered alive or dead, and in what state would a host come across the pathogens in the wild? As touched upon previously (Section 2.1), which time point should be used postinfection to study, for example, mRNA expression, and how would Dscam-hv behave after a second challenge with the same parasite? Ideally, one would have a response profile covering a frequent and robust duration after infection, but this is often time- and financially demanding.
## 4.2 Future Perspectives

In Table 10.2, we have briefly summarized a few, of what we consider to be, outstanding questions relating to Dscam-hv’s role in immunity.

<table>
<thead>
<tr>
<th>Question</th>
<th>Relevant Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Does isoform regulation occur at translation and/or secretion into the hemolymph?</td>
<td>The majority of studies have been done on RNA. Only one study on Dscam-hv peptide sequencing and this was not in relation to parasites.(^19)</td>
</tr>
<tr>
<td>2. Which Dscam-hv epitopes interact with parasites?</td>
<td></td>
</tr>
<tr>
<td>3. What is the Dscam-hv protein half-life?</td>
<td></td>
</tr>
<tr>
<td>4. Why should Dscam-hv mediated cellular interactions be different in hemocytes and fat body compared to nervous-system cells?</td>
<td>Nervous system Dscam-hv functions in cell–cell signaling.(^6) Dscam-hv could theoretically be involved in hemocyte–hemocyte, hemocyte–fat body or other cell interactions. Does knockdown affect cell–cell interactions, which secondarily affects immunity?</td>
</tr>
<tr>
<td>5. Does Dscam-hv play a role in maintaining / controlling hemolymph and gut bacterial fauna?</td>
<td>Dscam-hv RNAi resulted in increase in opportunistic bacteria in mosquito hemolymph.(^10)</td>
</tr>
<tr>
<td>6. Do myriapod or chelicerate Dscam’s play a role in immunity?</td>
<td>Dscam-hv transgenic mosquitoes had reduced microbiota in gut.(^46)</td>
</tr>
<tr>
<td>7. Do insect Dscam-hv paralogs play a role in immunity?</td>
<td>(S. maritima) and (I. scapularis) have multiple whole gene duplications of Dscam, but it is unknown if they have an immune role.(^57) Dscam alternative splicing of Ig7 encoding exons is expressed by hemocytes of (S. maritima).</td>
</tr>
<tr>
<td>8. Are there specific signaling pathways activated in Dscam-hv related to immunity?</td>
<td>Insects have multiple paralogs.(^7,8,77) These have two, or no, alternatively spliced exons. Therefore, if a role in immunity exists, it would not be as a diverse pattern recognition receptor.</td>
</tr>
<tr>
<td></td>
<td>Dscam-hv has different cytoplasmic tails, which are used in different stages of nervous system morphogenesis. Furthermore, its cytoplasmic tails are rich in signaling motives, which could be related to role in immunity.(^35)</td>
</tr>
</tbody>
</table>
5 CONCLUDING REMARKS

The immune system of arthropods is, similar to that of other invertebrates, “not homogeneous, not simple, not well understood.” The accumulation of genome projects from nonmodel species consistently reveals how different examples of molecular diversification have been selected for in nature. The cells expressing diversity are in many cases not only committed to immune functions, and an understanding of how diversity is used is lacking, which complicates mechanistic inferences. Dissecting the function and evolution of the Dscam gene family in arthropods will be a challenging endeavor. However, it might be rewarded by improving considerably our understanding of the nervous and immune systems of these animals, and our understanding of how evolution has built this extremely complex solution to serve these two systems.

ACKNOWLEDGMENTS

We acknowledge Prof Louis Du Pasquier for fruitful discussions. We would like to thank Seth Barribeau and Louis du Plessis for allowing us access to their Apis and Bombus Dscam alignments and analysis.

REFERENCES


The Evolution of the Immune System


54. Chou PH, Chang HS, Chen IT, Lee CW, Hung HY, Han-Ching Wang KC. Penaeus monodon Dscam (PmDscam) has a highly diverse cytoplasmic tail and is the first membrane-bound shrimp Dscam to be reported. Fish Shellfish Immunol 2011;30(4–5):1109–23.


95. Smith P. Dscam gene expression in invertebrate immunity: alternative splicing in response to diverse pathogens PhD; [2012]
Chapter 11

Structural and Functional Diversity of Fibrinogen-Related Domains

Russell F. Doolittle
Departments of Chemistry & Biochemistry and Molecular Biology, University of California, San Diego, La Jolla, CA, United States

1 INTRODUCTION

Fibrinogen-related domains (FReDs) were named and defined for the globular portions of vertebrate fibrinogen, the precursor molecule of fibrin blood clots. As will happen, the first instance of a protein domain being found may not be representative of its general function, and the vast majority of FReDs in the animal world have nothing to do with blood clotting. Rather, these domains are found in numerous extracellular proteins with widely different functions, a substantial number of which appear to be associated with innate immunity.

In this chapter, an attempt is made to weave together phylogenetic occurrence, structural features, rates of change, and function. In keeping with the general theme of this book, the emphasis will be on the diversity of some putative immune-related FRePs (FReD-containing proteins), eschewing, except for brief mention, those others with more specialized functions like the clotting of blood, or the stimulation of neurite development, or blood vessel growth, and so on, although these latter are included in a “time-line” of when various FReDs first appear in the chordate lineage.

2 STRUCTURAL FEATURES

FReDs actually consist of three subdomains, usually denoted A, B, and P (Fig. 11.1). The amino-terminal subdomain (“A domain”) consists of about 50 amino-acid residues and a single disulfide bond. The B and P subdomains are closely associated, the middle strand of the central β sheet in the B subdomain being provided by a segment that originates after the third subdomain. The B and P subdomains together amount to 150–200 residues.
The carboxy-terminal P domain invariably contains a disulfide bond, almost always with 12 residues between the two cysteines. The peptide bond between the second cysteine and its amino-terminal neighbor is in the cis configuration, a feature that is important to the architecture of this region, and especially in maintaining a cavity that is, in most instances, the binding site by which FReDs interact with other entities. Remarkably, the same natural binding cavity is used to bind a sugar derivative to a horseshoe-crab lectin as is used to bind amino-terminal peptide “knobs” on fibrinogen molecules during fibrin formation (Fig. 11.2).

A bound calcium is also an integral feature of the P domain. In line with the fact that FReDs, with rare exception, occur in extracellular environments, many are glycosylated, carbohydrate clusters occurring sporadically on any of the three subdomains.

As it happens, the A (amino-terminal) subdomain often occurs in many other settings, in which cases it is usually referred to as a NEC domain (an acronym based on the neuro- and collagenous environments in which they are found). In FReDs, these subdomains are more or less accessory in nature and do not participate directly in ligand binding. Because genuine FReDs consist of all three subdomains, NEC occurrences will not be considered further, except to discount irrelevant observations.

3 PHYLOGENETIC OCCURRENCE

We recently conducted a comprehensive bioinformatics survey of the occurrence of FReD genes in a wide assortment of animal genomes—from very early diverging phyletic groups like sponge and sea anemone, to crown protostomes
and deuterostomes like insects and mammals. In line with using past experimental data together with sequence searching regimens, those data have served as the raw material for proposals and inferences made in this chapter. The survey was sufficiently broad that the first appearance—in a phylogenetic sense—of various types of FReP can be gauged.

In this regard, FReDs are very much restricted to animals, making their first evolutionary appearance in sponges (Fig. 11.3). Contrary to what is often reported, no gene for a full-length FReD can be found in the genome of the choanoflagellate Monosiga brevicollis, generally regarded as the closest extant organism sharing common ancestry with animals. The mistaken reports are the unfortunate result of the National Center for Biotechnology Information (NCBI) referring to the 50-residue NEC domains—which do occur in several choanoflagellate proteins—as FReDs. In our 2012 survey, a FReD was found in a single bacterium, Bacteriovorax marinus, and an updated search now reveals one more such occurrence, this time in the proteobacterium Cupriavidus metallidurans. Both of these instances appear to be the result of horizontal transfers from an animal genome to the bacterium.

As a reminder, phylogenetic trees have two important features: branching order and branch lengths. The branching order depicts the temporal order in which gene duplications or speciation events occur; the branch lengths reflect the amount of change along each interval. A long branch-length indicates faster change for a particular interval than would a short one. As we shall see, occasionally variations in rates of change can lead to mistaken branching orders. Matters can sometimes be resolved by a consideration of unique indels (insertion/deletion events).
The vocabulary of phylogenies and gene duplication is important. Gene or protein sequences from different organisms that play the same roles as a result of simple parent to progeny passage are said to be orthologs, whereas homologous proteins that are the result of gene duplications are said to be paralogs. One of the purposes of tracking the evolution of protein domains like FReDs that are shuffled around in genomes is to sort out the orthologs from the paralogs. All 23 FReDs in Fig. 11.4 are paralogs, because they occur in the same organism (human) and must be the result of gene duplications.

4 FReDs AND NEIGHBOR DOMAINS

FReDs often, but not always, occur in the accompaniment of other domains, like immunoglobulin (Ig) or fibronectin type III (Fn3) or collagen or coiled-coil domains; when they do, the FReD is invariably at the carboxy-terminal domain of the cluster. The nature of these accompanying domains can bear heavily on the function of the FReP, especially with regard to the valence of ligand binding.

For example, the tight triple-helix nature of an adjacent collagen segment insures that three terminal FReDs are packed together as a symmetrical homotrimer, a feature with obvious advantage for binding to surfaces with numerous, uniformly distributed ligands, as might be expected to occur on the cell walls of bacteria. In some other FRePs, two-stranded coiled-coils are important.
for orienting identical pairs of FReDs side-by-side so that they can bind to and bring together diffusible entities on cell surfaces. In contrast, the FReDs in tenascins are only loosely associated, operating more or less independently at the ends of long, flexible tethers composed of Fn3 and/or EGF domains. These examples illustrate how identifying neighbor domains during a bioinformatic analysis can provide critical information for the inference of function.

Importantly, FReDs can also occur by themselves with no neighbor domains, in which case we will refer to them as “stand-alone FReDs.” It must be clear, however, that these solitary FReDs can associate with each other to form dimers, trimers, and higher oligomers, without the aid of neighbor domains, thereby achieving the benefits of polyvalence that can aid pattern recognition.

5 HUMAN FRePs

Genes for 23 different FReDs (in 23 different FRePs) were identified in the human genome, including four for tenascins, three for ficolins, nine for angiopoietin or angiopoietin-like proteins, three that give rise to fibrinogen, and four others: fibroleukin, hepassocin, microfibril-associated protein (MFAP), and the FIBCD1 protein (Fig. 11.4).
Although FRePs in humans are arguably the most studied, the functions of the 23 kinds are not all known with the same degree of certainty. On the secure side, tenascins have been researched intensely for several decades; it is well established that they are part of the fabric of the extracellular matrix (ECM) and have a variety of regulatory and developmental functions. They attract and interact with numerous cell-surface proteins, including those of leucocytes and neurites.\textsuperscript{8,9}

Another case where function is not in question involves the FReDs in the $\beta$ and $\gamma$ chains of fibrinogen, the protein for which FReDs were named. The “holes” in these FReDs unequivocally bind “knobs” from other fibrinogen molecules, causing the polymerization that is at the heart of fibrin blood clots. In contrast, the closely related FReD that occurs in an alternative form of the fibrinogen $\alpha$ chain does not seem to play a role in polymerization events, and remains mysterious.\textsuperscript{10} And although the liver-specific FReP called hepassocin is clearly a mitogen,\textsuperscript{11} the evidence for fibroleukin being an immuno-procoagulant with prothrombinase activity remains tenuous.\textsuperscript{12} Indeed, early claims about possible immunofunctions for fibroleukins were dealt a serious blow by a study with knockout mice in which the fibroleukin gene was deleted but no differences in any immune-related phenomena could be detected.\textsuperscript{13}

Angiopoietins are well-characterized proteins that stimulate blood vessel growth;\textsuperscript{14} the related angioarrestins are paralogs that act antagonistically.\textsuperscript{15} On another front, the FReP called “fibrinogen C domain connecting 1” (FIBCD1) occurs in the intestinal brush border, as well as some airway and salivary gland cells, and is well known to bind chitin, an entity that is indigestible in most animals.\textsuperscript{16} The mature protein contains a transmembrane segment at its amino-terminus that holds it bound to the component cell membrane. The function of the closely related—in a phylogenetic sense—microfibril(lar)-associated proteins (MFAPs) remains vague, however.\textsuperscript{17} MFAPs are “stand-alone” FRePs without accompanying domains, and it is not impossible that their binding to microfibrils accords them a kind of polyvalence that would aid in recognizing repetitive patterns on the surfaces of pathogens. The association of a serious developmental syndrome in persons with a deleted gene for MFAP4 argues for something more profound, however.\textsuperscript{18,19}

Finally, the human genome encodes three different ficolins,\textsuperscript{20–22} designated M-, L- and H- (also called ficolin-1, -2 and -3). In addition to having different amino-acid sequences, the three types differ with regard to the number of collagen triplets (GXX-) in their amino-terminal segments, leading to slightly different molecular weights. These are well-studied proteins with similar but different capabilities for binding ligands that have a number of attributes associated with immunity, and we will need to revisit them during our discussion of FReDs with apparent immune-related functions. Interestingly, recombinant forms in which the collagen segments are omitted still form trimers.\textsuperscript{22}

Homologs for the 23 human FRePs are found in most mammals, but strict orthologs, as we will now try to demonstrate, appear at distinctly different stages during chordate evolution. Moreover, the survey of FReD occurrences...
revealed a surprising amount of gain and loss, with closely related groups having different numbers of FReDs, both in general and of particular kinds. As an example, although human and chicken genomes both have 23 FReP genes, they are not the same 23. Gains and losses were sustained along both lineages on the way to birds and mammals.

In explanation, defense mechanisms in animals tend to be highly redundant and genetically “backed up,” even in the face of the evolutionary dictum “use it or lose it.” In the course of the never-ending war against pathogens, the challenge of expecting the unexpected from an ever-changing enemy may require numerous copies of a defensive agent. But such genes are readily lost, perhaps accounting for why frogs have four MFAPs and lizards and humans only one, and chickens none at all. Another possibility is that on occasion a particular FReD may develop the capability for binding a broader array of ligands, thus lowering the number of genomic versions needed for responding to potential adversaries. As an example, human L-ficolin has been reported to bind a much more diverse collection of ligands than M-ficolins or H-ficolin.23

6 EVOLUTION OF FUNCTIONAL DIVERSITY IN CHORDATES

Because FReDs occur in a very wide range of proteins with very different functions, we need to inquire: when and under what circumstances during evolution did these diverse functionalities arise? The newly engendered FReDs need to adapt to two quite different needs depending on whether they are recognizing endogenous entities, like proteins of the extracellular matrix, or components of developmental networks, and so on, or whether the function at hand is to fend off exogenous, alien materials, like microorganisms or indigestible foreign substances. The first category would include tenascins, angiopoietins, hepassocin, and so on; the second should include ficolins and FIBCD1, although it should be noted that it has been suggested that some ficolins may also recognize endogenous materials from disrupted cells.23

Let us begin with tenascins, the evolution of which has been thoroughly reviewed in the past.8,9 If the presence of an Fn3 neighbor domain is regarded as diagnostic, then it is generally accepted that tenascins occur in all chordates, making their first appearance in protochordates such as amphioxus.24 Next, angiopoietin and angiopoietin-like proteins, which, as we noted, are involved in the development of the vascular system, make their first appearance in jawless fish like the lamprey,3 coincident with the development of a closed vascular system (Fig. 11.5). We can feel confident about this identification because genuine angiopoietins have a unique signature in the way of an additional disulfide bond. There is no sign of these entities among those protochordates whose genome sequences are available.

The intestinal chitin-binding protein known as FIBCD1 protein is more problematic. The NCBI lists two entries from the protochordate Ciona intestinalis labeled as “FIBCD1-like,” but neither has the obligatory membrane-spanning
The sequence needed for localizing in the intestinal wall. In fact their sequences are much more “ficolin-like” than they are FIBCD1-like. Nor has an authentic FIBCD1 yet been identified in the jawless vertebrates. The problem in this case is that a high-quality genome assembly for neither the lamprey nor the hagfish (the only two genera of jawless fish extant) is yet available. It is possible that one or both may have such a gene, but it has not yet been found. In our recent review based on a fragmentary finding, we thought the lamprey might have such a gene, but that conclusion may have been premature.

Genes for fibroleukin and the liver mitogen hepassocin, like the three genes encoding FReDs that occur in fibrinogen, are found in all vertebrates. This includes jawless fish. Indeed, a gene for fibroleukin is present in the protochordate *Ciona intestinalis*. More remarkably, genes for a protofibrinogen have also been found in those creatures, even though the hemolymph of protochordates does not clot.

Authentic MFAPs, of the type designated MFAP4 in humans, make their first appearance in amphibians. There are four in the frog, and one each in lizard, opossum, and human. None was found in the chicken or in a draft genome of the basal mammalian platypus. It has been reported that asparagine-linked carbohydrate in MFAP plays a role in their binding to integrins, and indeed a specific glycosylation site is found in these proteins in all of the sequences available, as well as in the intestinal chitin-binding protein, consistent with the close relationship of these two FReDs (Fig. 11.6).
Ficolins of the kind found in mammals with collagen domains adjacent to FReDs also make their first appearance among the Amphibia (Fig. 11.5). Indeed, the African clawed frog has 25 verifiable ficolins. Four of these do not have collagen segments, but phylogenetic trees show that they have lost those segments, as opposed to never having had them. The number of collagen-triplets in the remaining 21 ficolins ranges from 2 to 36. These fluctuations are in accord with the old maxim that “gene duplication begets more gene duplication,” the consequence of the increasing opportunity for mismatching of similar DNA sequences during meiotic divisions.

Although jawed fish have an abundance of proteins labeled as MFAP-like by the NCBI, these are not orthologs of the MFAPs found in tetrapods. They are apparently absent from jawless ones like lamprey. We suggested that in fish these proteins are playing the same role as ficolins, and indeed, the notion was promptly confirmed when it was found that in a fish called the sculpin (Trachidermus fasciatus) a protein, the amino-acid sequence of which clearly shows it to be “MFAP-like” (Fig. 11.6), is expressed in response to challenge by the bacterial parasite Vibrio anguillarum. Even more convincing, a recombinant version of the same protein agglutinated the bacteria. We propose that
The Evolution of the Immune System

all these fish proteins might better be called “collagen-free” ficolins, or, perhaps more reasonably, “protoficolins” (Fig. 11.7).

The NCBI probably denoted these fish proteins as “MFAP-like” because simple amino-acid sequence comparisons show them to be more similar to tetrapod MFAP sequences than to tetrapod ficolins. The average similarity for zebrafish entries with frog MFAPs is about 50%, but the same proteins are only about 45% identical to frog ficolins. Nonetheless, as we will show here, there is good reason to think that the fish proteins are really orthologs of tetrapod ficolins and not of frog or human MFAPs.

Pointedly, the FIBC proteins from zebrafish, frog, and human and the MFAP proteins from frog and human all have a unique 5-amino-acid insert that is lacking in other FreDs (Fig. 11.6), the presence of which indicates that tetrapod MFAPs diverged from FIBCD1 after the separation of fish from other vertebrates. The discrepant relationship of the MFAP-like and MFAPs is the combined result of very rapid change among the ficolins and very slow change for the FIBC–MFAP group, making the distances between paralogs seem less than between orthologs. Evidence for such confounding comparisons has been noted in the past; the phenomenon is the flip-side of well-known systematic errors attributable to “long branch attraction.”

7 RATES OF SEQUENCE CHANGE

Perhaps not unexpectedly, it was found that FRSeDs with different functions have changed at different rates, with ficolins being by far the fastest-changing group. Significantly, the greatest amount of change in ficolin FRSeDs occurs in
the third subdomain (“P domain” in Fig. 11.1), the very opposite of the situation in FReDs with either singular or more specialized functions like blood clotting or angiogenesis, in which cases the P domain sequences are locked into more conserved modes. The different patterns are readily codified as the ratio of percent identities of the A and B domains relative to the percent identities of P domains (Table 11.1). Ficolins, for example, have ratios greater than 1.0, whereas tenascins, angiopoietins, and other FReDs have ratios less than 1.0.

Not coincidentally, similar patterns of unbalanced sequence change occur among the fish proteins mistakenly labeled “MFAP-like,” and, as we shall see, the unbalanced style of change is also observed in FReDs from nonchordates,
the rapid change in the P domain being observed in many FRePs that have been implicated in innate immunity. All indications are that the significantly different rates of change in the subdomains of FReDs with putative immune functions are the result of positive selection.

8 FReDs IN PROTOCHORDATES

At this point, two protochordates have had their whole genomes sequenced. One is a urochordate called the sea squirt (Ciona intestinalis), and the other is the earlier diverging cephalochordate amphioxus (Branchiostoma floridae). Both genomes encode numerous FReDs, and experiments on closely related species have implicated some of them with immune functions. In the amphioxus, B. belcheri, bacteriolytic activity has been reported for a FReD described as a multivalent pattern-recognition receptor, and a novel ficolin has been cloned from the urochordate, Halocynthia roretzi. Sequence-based phylogenetic trees showed this ficolin to be the result of a genetic joining of a collagen domain and a FReD independent of the one that had led to ficolins in tetrapods.

Beyond that, the genome of the sea squirt contains a remarkable set of evolutionary innovations involving FReDs, including self-incompatibility factors, a nonclottable protofibrinogen, and a putative transcription factor. None of these appear to have a counterpart in the amphioxus genome.

9 EXTENDING THE DEUTEROSTOME TIME-LINE

Echinoderms are early diverging deuterostomes and a reasonable place to search for ancestors of chordate FRePs. The common sea urchin (Strongylocentrotus purpuratus) has 35 identified FReDs, several of which tend to cluster with human ficolins and tenascins, but none of them have collagen segments or any hint of domains found in combination with vertebrate tenascins. So far it has not been possible to place any of these FReDs as orthologs with chordate counterparts. Recently, it was reported that a FReP in a different echinoderm, the sea cucumber Apostichopus japonicus, is involved in tissue regeneration as determined by expression levels in various tissues. This may be so, but it seems a remarkable coincidence that several decades ago the same gene product had been cloned from another sea cucumber species as part of a general search for nonvertebrate FReDs, simply on the basis of its abundance. These are “stand-alone” proteins without any accompanying domain, and, as a result, that attribute cannot be used as an aid in defining function.

10 NONDEUTEROSTOME ANIMALS

Although FReDs from nondeuterostome animals are frequently called “ficolin-like,” “tenascin-like,” and so on, mostly these are not genuine ficolins (no collagen domains), nor are they authentic tenascins (no Fn3 domains), nor any of the other types of FReD found in deuterostomes. Nonetheless, the
conventions of current nomenclature are such that many FRePs from nondeuterostome animals have been named “angiopoietin-like” and some other ones “FReD-like” based on the kind of sequence they tend most to resemble.

In the past we have cautioned that it would be unwise to infer any commonality of function from these names. Even so, I am now going to suggest, cautiously, that the majority of FReDs found in early diverging organisms like sponges and jellyfish may actually be involved with a kind of immune response similar to that reported for chordate ficolins, probably antibacterial in nature. As in the case of a group of fish FReDs discussed previously, I will use the term protoficolins for these collagen-free entities.

### 11 SPONGES AND JELLYFISH

Sponges (phylum Porifera) and jellyfish (phylum Cnidaria) are the earliest diverging animals (Fig. 11.3) and where one might logically look for new physiological functions. Immune phenomena in sponges have long been a matter of discussion, especially with intrinsic activity directed against bacteria. Although the molecular details of most of these interactions remain to be studied, there is one case of a particular sponge, *Suberites domuncula*, where a FReD has been implicated as part of a system for fending off fungi.

Given that limited experimental background, the completion of the genome DNA sequence for the sponge, *Amphimedon queenslandica*, presents an opportunity for testing a strict bioinformatics approach for assigning functions to FReDs for which no experimental data are yet in hand. Partial sequences aside, 92 full-length FReDs were identified in the sponge genome, the majority of which have short uncharacterized sequences on their amino-terminal sides (< 60 residues between the signal sequence and the FReD). None have collagen-like sequences nor any of the domains associated with FRePs found in chordates. However, 41 of the 106 had an approximately 50-residue motif containing five or six cysteines, consistent with a tightly folded, disulfide-bonded domain. The motif was not found in any other proteins in the NCBI protein database.

When included in phylogenetic trees with FReDs from other phyla, sponge entries invariably cluster separately, indicating an independent diaspora that has been occurring since the time of the last common ancestor of sponges and other animals. That said, and neighbor domains aside, sponge FReD sequences tend to most resemble chordate ficolins and tenascins. The question is, are any, all, or most of these FReDs involved with binding to bacteria or other pathogens?

Indeed, a case for sponge FReDs being involved in some kind of immune function can be made on the following counts: (1) the number of related FReD genes is extensive; (2) among these, the interactive P subdomain changes at a much faster rate than the A and B subdomains, just as it occurs in chordate ficolins and protoficolins (Table 11.1); and (3) the key residues known to be involved in binding to N-acetylated small molecules in ficolins, FIBCD1,
The Evolution of the Immune System

and an arthropod lectin, vary in a similar way against a background of a conserved structural core (Fig. 11.8).

A similar case can be made for the jellyfish Nemastostella vectensis (more formally, the sea anemone, a cnidarian), with the added observation that among its 25 full-length FReDs there is one with an adjacent collagen segment, its position high in the phylogenetic tree, which is clear evidence of an independently evolved ficolin, but true to the original acronym based on FIbrinogen and COLlagen (Fig. 11.3).

12 BILATERIANS

12.1 Phylum Mollusca

Inquiries into possible immune responses for molluscs go way back, and as a result of these longstanding interests, FReDs were long ago found in the snail Biomphalaria glabrata. That discovery has been greatly exploited in the intervening years, and we will not dwell on this work beyond noting that a convincing case has been made for snail FRePs having a role in immunity, particularly with regard to resistance to infection by trematode parasites. Neighboring domains in these FRePs include immunoglobulin domains, and, occasionally, EGF domains.

FReDs with putative immune-related properties have also been identified in other mollusks, including the scallop, Argopecten irradians, in which studies have demonstrated the ability of these FRePs to agglutinate or otherwise immobilize specific bacteria. Moreover, in the case of the slug, Limax flavus, a FReP has been thoroughly characterized and shown to bind sialic acid.

12.2 Phylum Arthropoda

Defense mechanisms in the horseshoe crab Tachypleus tridentatus have been subjected to intense scrutiny over the past several decades, during the course
of which a series of lectins was isolated that bind to and agglutinate bacteria and other foreign materials. Quite unexpectedly, two of these, tachylectin 5A and 5B, were found to be FRePs. Remarkably, an X-ray structure showed that these proteins bound N-acetylated small molecule ligands at a site coincident with the peptide-binding holes on vertebrate fibrinogen (Fig. 11.2). Even though the horseshoe crab tachylectins 5A and 5B are “stand-alone” proteins, evidence was provided that showed that the solitary proteins tend to form hexamers and octamers, demonstrating that their binding capabilities could be polyvalent and well-adapted for recognizing patterned repeats. As such, it was suggested that these proteins were the functional equivalent of ficolins found in vertebrates.

12.3 Pancrustacea—Crustaceans

A situation similar to that described in the horseshoe crab was found more recently in a crustacean, the kuruma shrimp, Marsupenaeus japonicus, where a stand-alone FReD has been found to agglutinate the bacterium Vibrio anguil- larum.

12.4 Insects

Insect FRePs have also been the subject of numerous inquiries in the past. The possibility of FRePs playing a role in fending off infection by malarial parasites as well as bacteria has spurred these studies in mosquitoes especially. The point must be underscored that, like the situation in chordates, there is great variation in numbers of FReDs from one organismic group to another, ranging from upwards of 30 in some mosquitoes to only 2 in the honeybee, Apis mellifera. Even within a single genus there is great diversity, as a dozen different fruit-fly species, for example, having anywhere from 14 to 43 FReDs.

In the case of the honeybee, both FRePs appear to be orthologs of proteins found in other insects with specific functions, one being the well-known scabrous protein originally found in Drosophila melanogaster. It is significant that once a specific nonimmune function is acquired by one of these genes, the rate and pattern of sequence change becomes greatly moderated.

13 FReD-BASED IMMUNITY AND THE RECOGNITION PROBLEM

Given that an assortment of FReDs from across the entire animal spectrum has been found to bind bacteria, it is tempting to think that their initial appearance in sponges heralded that phenomenon. We have already pointed out those features of sponge FReDs in accordance with that notion, particularly the rapid and unbalanced nature of amino-acid sequence change. The obvious center of interest is the binding cavity on the P subdomain. The fact that so many extant FReDs
bind to acetylated small molecules argues that the original cavity likely had that capability also. The still unanswered question is, how did the binding site on the P-domain originate?

The binding sites of human ficolins and FIBCD1, which bind similar sets of acetylated entities, have been scrutinized with great care, and several important observations have emerged. For one, it is surprising how many different small molecules, sugars (or not), and acetylated (or not), are bound to various different FReDs. For another, it is now clear that the same ligand can be bound in slightly different ways in different FReDs. Unexpectedly, also, the binding site can be expanded in some instances, as occurs in human L-ficolin, where a set of subsites has been found. With regard to fine-tuning, adjustments to how the cavities in a single FReP cluster may be arrayed have been attributed to small shifts in the P-domain, relative to the adjacent B-domain.

Despite all the attention to structural detail, several questions linger about the nature of recognition and about the rapid rate of change, as well as the multiplicity of FReDs in some organisms compared to a paucity in others. Why does a frog have 25 ficolins and a chicken only one? Perhaps this reflects two different strategies of defense, in the one case, maintaining a diverse inventory of FReDs, each with different binding potential, and, in the second case, having a specialized FReD that can somehow recognize a diverse set of pathogens.

There are other problems to contemplate, beginning with the fact that N-acetylated small molecules are ubiquitous. How does a FReD distinguish between friend and foe? The glib explanation is that alien ligands often come in regularized patterns, like the peptido-glycans found in bacteria capsules. In these cases an intrinsically weak binding becomes amplified (exponentially) by having more interactive sites (polyvalence). But pattern recognition, often mentioned but seldom spelled out, has its limits. Bacterial mimicry of animal surfaces is another circumstance where distinguishing self from nonself remains an issue. The simple observation that a FReD can agglutinate a bacterium always deserves further contemplation.

Finally, we need to inquire where those original sponge FReDs with putative magical binding-sites came from. It seems unlikely that they were fashioned de novo in a primitive sponge. More likely, a gene for the more widely occurring NEC domain found itself adjacent to one resembling a B–P subdomain combination, the origins of which remain unknown. At this point, it is an open question.

14 CONCLUDING REMARKS

In conclusion, FReDs are widely distributed in the animal kingdom, almost entirely as extracellular, diffusible proteins, the interactive sites of which typically involve the carboxyl-terminal subdomain. The broad diversity of functions exhibited by these structures range from the binding of dangerous pathogens to endogenous cell signaling and fibrin gel formation. I have tried to make the case that
it is the potential for binding and neutralizing bacteria that is reflected by a very rapid change in the P subdomain—clearly manifested in vertebrate ficolins and protoficolins—as well as numerous and as yet uncharacterized FReDs in sponges where the system had its start. This unusual and uneven kind of variability has to be the result of positive selection. During the course of evolution, advantage was taken of gene duplications, sometimes with the accompaniment of new neighbor domains, which led to new capabilities, derived from a deformable and easily tweaked binding site, to adopt new and singular nonimmune functions.

REFERENCES


Chapter 12

Genomic Instability and Shared Mechanisms for Gene Diversification in Two Distant Immune Gene Families: The Plant NBS-LRR Genes and the Echinoid 185/333 Genes

Matan Oren, Megan A. Barela Hudgell, Preethi Golconda, Cheng Man Lun, L. Courtney Smith
Department of Biological Sciences, The George Washington University, WA, United States

1 INTRODUCTION

One of the major challenges faced by immune systems is to generate a protein repertoire that is broad and competent enough to recognize the ever-diversifying array of pathogenic nonself. Eukaryotes have numerous strategies to achieve this. Innate immune systems consist of large families of pattern recognition receptors (PRRs) that identify different pathogen associated molecular patterns (PAMPs) with high specificity. Examples include (1) Toll-like receptors (TLRs) found in most animals from Porifera to humans, with PAMP recognition function demonstrated in some species, including human, mouse, and fruit fly; (2) fibrinogen-related proteins (FREPs) with antiparasite activities in mollusks; (3) Down syndrome cell adhesion molecule (Dscam) in insects and crustaceans with opsonin function; and (4) variable domain-containing chitin binding proteins (VCBPs) in protochordates that respond to gut microbes. The adaptive immune system in jawed vertebrates uses somatic recombination of gene segments to create enormous diversity of T cell and B cell receptors. Alternatively, the adaptive immune system in the jawless vertebrates relies on a copy-choice mechanism to assemble sections of leucine-rich repeat (LRR) cassettes into a germline gene to create similar diversity of variable lymphocyte receptors. On the other hand, innate immune systems have been suggested
to lack the flexibility of adaptive immunity to identify and respond to novel PAMPs that have either newly appeared, or have been newly introduced into a population, due to either environmental changes or as a result of the arms race with host immunity. Therefore, it is necessary for innate immune systems in eukaryotes to employ other types of swift genomic diversification mechanisms either within the lifespan of the host or between generations, to stay even in the arms race with the pathogens. Here, we discuss different aspects of genome diversification in two very distinct innate immune gene-families: the nucleotide binding site leucine rich repeat (NBS-LRR) genes in plants, and the 185/333 genes in echinoids. The first is abundant in many species of plants, is a subset of the resistance (R) genes, and appeared early in the plant lineage more than 500 million years ago whereas the second is restricted to the echinoid lineage of echinoderms, and the extant genes are estimated to be only 2.7–10 million years old. Although there are many differences between these two gene families, there are some striking similarities in the genomic structure and the gene diversity among and within species, which will be the focus of this review.

2 THE NBS-LRR GENE FAMILY IN HIGHER PLANTS

The immune response in plants consists of two arms: PAMP-triggered immunity (PTI) and effector-triggered immunity (ETI). PTI relies on cell-surface-membrane mounted PRRs that extend into the apoplast and recognize and respond to microbial molecules. ETI functions most often in the plant cell cytoplasm, either acting directly by detecting pathogen virulence-factors called effectors, or acting indirectly by monitoring host proteins that have been altered by effector activity. The guard hypothesis suggests that the indirect detection of effector activity is facilitated through a cytoplasmic complex of an R protein that functions as a guard for a host guardee protein. In normal conditions the guard/guardee complex is stable, but upon injection of effectors into the plant cell by a pathogen, the effectors alter the guardee, which is detected by the R protein guard, and induces a signaling pathway to activate the ETI response. The indirect ETI response to changes in the guardee proteins maximizes the capacity of the plant host to detect the activity of a large variety of pathogens with a much smaller number of R proteins. The key players in the ETI response are a diverse group of mostly intracellular R proteins that are encoded by a few to hundreds of R genes that are present typically in clusters in every plant genome, with an expanded repertoire in flowering plants. Most R proteins, although not all, are characterized by the presence of a nucleotide binding site (NBS) domain, a linker region, and a variable number of LRRs. There are 151 NBS-LRR proteins in the mouse-ear or thale cress, Arabidopsis thaliana, 458 in rice, 459 in wine grape, but only two in the much more primitive plant, the spike moss. The NBS-LRR type of R proteins are divided into two major structurally distinct sub-groups, defined by the N-terminal domain, which is either a
Toll/interleukin-1 receptor (TIR) domain in the TIR-NBS-LRR (TNL) type, or a coiled-coil (CC) domain in the CC-NBS-LRR (CNL) type (Fig. 12.1B, C) (reviewed in[21]). Binding of the LRRs of the TNL and CNL proteins to effector molecules, or to altered guardee proteins, triggers different downstream signaling cascades that lead to the hypersensitivity response (HR) in plants. HR is a rapid apoptotic reaction in infected cells, and those nearby, which functions to remove the availability of cytoplasmic nutrients to pathogens, and thereby restricts their growth and spread. [24]
THE 185/333 GENE FAMILY IN ECHINODIMS

The 185/333 gene family encodes a diversified repertoire of immune-response proteins in sea urchins. To date, the 185/333 gene families have only been identified in two species of sea urchins, Strongylocentrotus purpuratus (the California purple sea urchin)\(^{15,16}\) and Heliocidaris erythrogramma (the Australian purple sea urchin).\(^{29}\) However, these genes are likely present in most echinoids, as they have been identified in the genome sequences of Strongylocentrotus franciscanus and Allocentrotus fragilis\(^{16}\) and Lytechinus pictus (K. Buckley, University of Toronto, personal communication). Among those, the most studied is the Sp185/333 gene family in the California purple sea urchin, \(S. \) purpuratus, which was first identified because it showed significant up-regulation in response to immune challenge with heat-killed bacteria and PAMPs, including lipopolysaccharides (LPS), peptidoglycans (PGN), and \(\beta\)-1,3-glucan.\(^{30–33}\) The family consists of up to 60 members; however, the gene number may vary among individuals\(^{16}\) and among different species (K. Buckley, personal communication). The Sp185/333 genes range in size from 1.2 to 2 kb and have only two exons separated by a small intron (380–413 nucleotides).\(^{32,34}\) The first exon (51–54 nucleotides) encodes the hydrophobic leader, whereas the second encodes the mature protein that shows significant sequence diversity. Optimal alignments of genes and transcripts require the insertion of artificial gaps, which define the presence and absence of short blocks of sequence, known as elements (Fig. 12.2A). The combinations of different elements result in recognizable mosaics of elements, called element patterns.\(^{31,32,34}\) This gene structure is persistent among sea urchin species studied to date, although the elements in 185/333 genes from different sea urchin species are not the same.\(^{15,29}\) The predicted structure of the Sp185/333 proteins is a signal peptide at the N-terminus, a glycine-rich region with an arginine–glycine–aspartic acid (RGD) motif (suggestive of integrin binding), a histidine-rich region, and a C-terminal region (Fig. 12.2B). No secondary structure can be predicted based on the amino-acid sequence for any of the proteins deduced from the cDNA or gene sequences.\(^{31–33,35}\) The 185/333 genes are expressed in specific subpopulations of sea urchin coelomocytes, and the encoded proteins appear to be localized internally in perinuclear vesicles in some phagocytes, and on the cell surface of small phagocytes.\(^{29,36–38}\) In \(S. \) purpuratus, single phagocytes from immune-challenged sea urchins express a single Sp185/333 message, inferring complex regulation of gene expression from the family and the production of a single Sp185/333 protein per cell.\(^{38}\) It should be noted that although a genome sequence exists for an individual California purple sea urchin, the Sp185/333 gene family is artificially underrepresented within this genome, likely due to computational assembly-contraction problems resulting from the variety of repeat sequences that are present between and within the genes (Fig. 12.2A). The size and organization of the Sp185/333 gene family is currently known, based on gene and message
**Figure 12.2** Sp185/333 gene cluster, repeat-based alignment, element patterns and protein structure. (A) Repeat-based alignment of the Sp185/333 genes shown in (C). The alignment optimizes correspondence between repeats and elements whenever possible.\(^{34}\) Optimal alignments require artificial gaps (horizontal black lines) that delineate individual elements shown as different-colored rectangles. The consensus of all possible elements are numbered across the top of the alignment. Each gene is composed of two exons; the first encodes the leader (L) and the second encodes the mature protein. Almost all genes have a single intron (int) of \(\sim 400\) nt (not to scale). The mosaic combinations of presence or absence of different elements in the second exon defines the element pattern (E2, B8, D1, and A2).

Elements that correlate with each of the six types of repeats are shown in different-colored rectangles at the bottom (type 1, red; type 2, blue; type 3, yellow; type 4, green; type 5, pink; type 6, dark gray); the brackets under the type 2 to type 6 repeats indicate the two duplicated regions. (B) The deduced Sp185/333 protein structure. The protein size and regions of the protein are correlated with the gene structure in (A). (C) Six Sp185/333 genes in a BAC insert (GenBank accession number BK007096) are closely linked. Genes are indicated by element pattern and color; A2 (red), B8 (orange), three D1 (yellow, green, blue), and E2 (purple). The genes are located near the 3’ end of the BAC insert within 34 kb. Gene orientations are indicated and spacing is relative to the scale. GA microsatellites flank each gene and GAT microsatellites flank segmental duplications within which are positioned three D1 genes. (Sources: Part A modified from [34]; part B modified from [39].)
sequences, and on the assembled insert for one BAC clone (GenBank accession number BK007096), which contains six tightly clustered genes (Fig. 12.2C).  

4 GENE DIVERSIFICATION

There is ample evidence for rapid evolution in the NBS-LRR gene family in higher plants and in the 185/333 gene family in sea urchins. One has only to evaluate the variability of both the gene numbers among and within species, and the sequence diversity of the genes, to obtain a general understanding of the pace of diversification. The NBS-LRR gene family is one of the largest and most variable gene families in plants. Although the common ancestor for the plant NBS-LRR genes is predicted to be much older than the common ancestor of the Sp185/333 genes, the NBS-LRR family has continued to expand and diversify. Many of its members exhibit allelic polymorphism, and for some NBS-LRR loci, polymorphism within populations is as great as that characterized for the major histocompatibility complex in vertebrates. The NBS-LRR genes show two general types of models for gene evolution: the majority are type I genes that show diversifying selection with a rapid rate of evolution and high sequence exchange among genes, and the rest are type II genes that show a slower rate of diversification correlating with less frequent exchanges. These two models of gene evolution are not mutually exclusive, and NBS-LRR genes positioned within the same cluster can show signatures of both diversification rates. It is noteworthy that the TNL class tends to show significantly higher evolution rates than the non-TNL genes, including the CNL class. Within the TNL genes, sequences that encode the solvent-exposed regions of the LRRs (Fig. 12.1B) seem to be under the highest positive selection and show the highest levels of genetic diversification. These regions show elevated ratios of the nonsynonymous versus synonymous substitutions (dN/dS). This is likely driven by the shared sequences among the LRRs, together with selection based on the function of the LRRs in pathogen-associated recognition. In contrast, the region encoding the NBS domain undergoes purifying selection and is highly conserved, which is likely based on its functions in nucleotide binding, which is crucial for R protein function to initiate signaling in order to activate the protective HR.

Similar to the NBS-LRR gene family, the 185/333 genes show exceptional diversity both among animals and among sea urchin species. An unrooted phylogenetic tree of 185/333 sequences from H. erythrogramma and S. purpuratus shows a complete separation of sequences from the two species into different clades. The recognizable element patterns in the second exon of the Sp185/333 genes are composed of a mosaic of 25–27 different possible elements (depending on the alignment) that range in size from 12 to 258 nucleotides (Fig. 12.2A) and generate 51 different patterns that have been identified to date. Similarly, the He185/333 genes have 26 elements and 31 element patterns, based on the first report on this gene family.
patterns of the different Sp185/333 genes impart high sequence diversity, but paradoxically, because they share element sequences, they are up to 88% identical. Furthermore, although element sequences are shared among genes, identical sequences of full-length genes are not shared among individual sea urchins. This is because (1) only subsets of elements are shared among genes and among animals, (2) there is sequence diversity within different versions of the same element, and (3) there are sequence variations among intron from different genes. The 185/333 genes from both species show many nonsynonymous substitutions with respect to synonymous substitutions (dN/dS ratio) for some element sequences, indicating diversifying selection for these regions, whereas for other elements, a low dN/dS ratio, suggesting purifying selection, has been noted. Furthermore, when Sp185/333 gene sequences are compared, the level of diversity among the elements shows significant differences. In general, the 185/333 and the NBS-LRR gene families portray sequence diversity patterns with exceptionally fast diversification rates and high dN/dS ratios for some regions within the genes, and slow diversification rates and low dN/dS ratios for other regions. For both families, a conserved basic structure of the genes that encode the functional regions of the proteins is maintained.

5 CLUSTERING AND TANDEM REPEATS

The NBS-LRR genes are unevenly distributed in the genome, and tend to be present in clusters that vary in size from 2 to 23 genes, with possibly more in single clusters (Fig. 12.1A). For example, the rice Xa21 gene cluster has seven paralogs within 230 kb, the tomato I2 cluster has seven paralogs within 90 kb, and the RPW8 cluster in Arabidopsis has five paralogs within 13 kb. NBS-LRR clusters can be homogeneous, with all members showing similar structure of either TNL or CNL genes (Fig. 12.1B, C), or can be heterogeneous with TNL and CNL genes mixed together. Homogeneous NBS-LRR clusters that contain tandemly repeated genes are very common in many plant genomes. For example, ~40 homogeneous clusters are present in the Arabidopsis genome, compared to ~10 clusters that are heterogeneous. There is evidence that the clustering of NBS-LRR genes is a major factor in the sequence diversification among the members of the family. The cluster size and gene copy number is positively correlated with sequence-exchange frequency among members of the cluster. Furthermore, there are greater dN/dS ratios for paralogs in clusters compared to isolated paralogs. The published Sp185/333 cluster consists of six closely linked Sp185/333 genes within 34 kb. Five of the genes are tightly clustered within 20 kb and are 3.2 kb apart, whereas a peripheral sixth gene is located at a distance of 14 kb (Fig. 12.2C). The peripheral genes are oriented in the same direction, whereas the four internal genes are oriented in the opposite direction. The cluster is composed of a mixture of homogeneous and heterogeneous genes based on the element patterns of the second exon (Fig. 12.2A). The three central genes all
have a $D1$ element pattern, and are positioned within three tandem segmental duplications of $\sim 4.5$ kb that show 99.7% sequence identity and are flanked by GAT microsatellites (Fig. 12.2C).\textsuperscript{39} The near-identity among the $D1$ genes and their flanking regions suggest very recent duplication events.\textsuperscript{16,39}

Both the $R$ and $185/333$ gene families contain several types of repeats. The $NBS-LRR$ genes contain exons that encode LRRs of 20–29 amino acids with a consensus sequence of LxxLxLxxNxL(T/S)GxIPxxLGxLxx, in which “L” is Leu, Ile, Val, or Phe, T/S is Thr or Ser, and “x” is any amino acid.\textsuperscript{56–58} The number of LRRs can vary among $NBS-LRR$ genes, ranging from 4 to 50 repeats.\textsuperscript{27} For example, in Arabidopsis, the number of LRRs ranges from 8 to 25\textsuperscript{28} and the Resistance Gene Candidate 2 ($RGC2$) genes in lettuce have 40 to 48 LRRs.\textsuperscript{27} Although the LRRs have an established function for interaction with PAMPs or pathogen elicitors (reviewed in\textsuperscript{17,59}), they also serve as an important component in creating genomic instability due to their repetitive nature, which leads to gene-family diversification. Evidence for the participation of LRRs in gene diversification processes lies within the differences in the LRRs among quickly diversifying type I $R$ genes, compared to more slowly diversifying type II genes. The sequence identity of introns within type I genes vary between the 5’ region and the 3’ region of the gene (Fig. 12.1B). Introns within the LRR region have high sequence-identity when compared to each other, which may reflect higher rates of sequence exchanges within the LRR region. Introns within slowly evolving type II genes have low sequence identity, reflecting their lower rates of sequence-exchange events. TNLs have additional introns within the LRR coding regions that are absent from most known CNLs (Fig. 12.1B), which may be indicative of differences in the evolutionary history of the two gene types.\textsuperscript{27,28} The greater number of introns within TNL genes versus CNL genes may indicate that TNL genes originated from a fusion of independent genes and are younger than CNL genes, which have few to no introns.\textsuperscript{48} It is noteworthy that, although the CNL genes have lost their modular gene structure over time, the encoded proteins may maintain modular functions.

The repeats within the second exon of the $Sp185/333$ genes allow two different alignments that are equally optimal.\textsuperscript{34} The initial alignment is based on the cDNA sequences, and did not take into account the positions of the internal repeats.\textsuperscript{32,33} The second alignment is repeat-based that optimized the correspondence of elements and repeats.\textsuperscript{34} There are six types of imperfect repeats in the second exon that are both tandem and interspersed (Fig. 12.2A).\textsuperscript{31,32,34,39} Depending on the gene, there are two to four type-1 repeats at the 5’ end of the exon, plus multiple copies of type 2–6 repeats that are present in two duplications of the interspersed repeats, in addition to an extra type-3 repeat (Fig. 12.2A). In addition, there are GA microsatellites positioned on either side of each gene within the intergenic regions, and are located about 430 bp from the 5’ end of each gene, and 300–700 bp from the 3’ end (Fig. 12.2C).\textsuperscript{39} The GAT microsatellites are positioned at the edges of three $\sim 4.5$ kb tandem segmental duplications that include three $D1$ genes (Fig. 12.2C). Based on their
positions at the edges of the duplicated regions, they may act as mediators of the duplication process. Moreover, pairwise sequence comparisons among the clustered genes identified in the BAC insert show that the sequences between the ends of the coding regions and the nearby flanking GA microsatellite are much more conserved than the regions outside of the GA repeats. This suggests that the microsatellites surrounding the Sp185/333 genes and those surrounding the segmental duplications may promote diversification of the family through regional instability, including sequence duplication and limiting sequence homogenization from gene conversion. Taken together, both the 185/333 and R gene families are characterized by clustering, repeats, and duplications. These features are found abundantly within the genomic structure for each of these innate immune gene-families, and are likely crucial for the processes that lead to gene diversification.

6 SPECULATIONS ON DIVERSIFICATION MECHANISMS OF THE Sp185/333 GENES

The regions of the genome in which the NBS-LRR and the Sp185/333 gene families are located, are very likely prone to genomic instability, which leads to gene sequence diversification. Gene diversification is initiated by mechanisms that regulate changes in the gene-copy number and organization of the whole family in which entire genes are duplicated, transferred to another location, deleted, or incur changes within the gene sequences (Fig. 12.3). The arms race between host and pathogen drives changes in host immune gene-sequence, which in turn drives functional adaptations in genes encoding effector proteins in pathogens, as demonstrated for the regions of the plant R genes that encode the LRRs. The variety of repetitive sequences in the NBS-LRR and Sp185/333 gene families promote genomic instability and nucleotide mismatches that may take place when homologous chromosomes interact either during meiosis or DNA repair processes. Meiotic recombination and homologous DNA repair may be regarded as special events in which homologous chromosomes interact and promote sequence exchange. Several mechanisms that directly and indirectly lead to gene sequence rearrangements have been suggested for the NBS-LRR gene family. Based on the structural and diversification similarities of these two immune gene families, we speculate that these mechanisms apply to the Sp185/333 genes as well. NBS-LRR genes are diversified by recombination between alleles and similar family members that result in new R genes with altered sequences. This spontaneous allele recombination is combined with selective pressures to detect PAMPs or elicitors, and results in gene variants with altered binding specificity. For example, individual L genes in flax that are derived from intragenic crossing-over show distinct phenotypes with regard to pathogen recognition. Recombination in the Sp185/333 genes has been detected computationally and is evident, not only between, but within elements and within the intron, suggesting that recombination events can occur at
FIGURE 12.3  Genomic modifications that potentially lead to changes in the size of gene families, changes in the organization of clusters, and alterations to gene sequences. Genes are represented as polygons (white and striped genes are in nonallelic clusters), with the pointed end indicating gene orientation. The genomic DNA in which the genes are located is shown as a solid or dashed horizontal line representing nonallelic regions. The generation of diversity within clusters and sequence diversity within genes is illustrated. (A) An unequal crossing-over in an intergenic region between nonallelic clusters can alter the sizes of the clusters, and result in heterogeneous clusters. (B) Unequal crossing-over within genes in nonallelic clusters can generate recombinant genes, alter cluster sizes, and result in heterogeneous clusters. (C) Gene conversion results when the sequences of one gene are copied into a nonallelic gene of similar sequence. (D) Inversion changes the orientation of a gene within a cluster, whereas the tandem duplication of genes or sets of genes increases the size of a cluster. (E) A duplicated gene can be inserted into an ectopic location, generating a heterogeneous cluster. (F) Meiotic mispairing occurs when chromatids misalign in regions of allelic clusters of highly similar genes, with the outcome of more genes in one cluster and fewer in the allelic cluster. The recombination event is shown between genes, but can occur within genes, as in (B). The processes shown in (A), (B), (E), and (F) can increase and decrease gene-copy numbers in clusters and in gene families.
any point throughout the entire gene sequence and are not focused in hotspots. For example, there is no correlation between the patterns and numbers of the tandem type I repeats in the 5’ end of the second exon and the patterns of the interspersed repeats located towards the 3’ end of the exon (Fig. 12.1A). It has been suggested that highly similar sequences between duplicated genes within homologous clusters drive further diversification through processes such as unequal crossing-over (Fig. 12.3A, B),\(^{61}\) resulting in unequal numbers and mispaired linked genes in the progeny, followed by processes that drive further diversification.\(^{55}\) In both families, shared sequences among paralogs leads to a swift rate of recombination among the genes.

Gene conversion occurs either during meiosis or as a result of DNA repair processes when homologous sites show mismatches in base pairing. These mismatches are recognized and corrected by the DNA repair machinery to convert the sequence of one allele to the sequence of its homologous counterpart (Fig. 12.3C). Gene conversion is an important diversification mechanism in TNL genes that undergo rapid sequence diversification followed by pathogen-driven selection for function (reviewed in\(^ {22}\)). The \(RGC2\) type I genes in lettuce undergo rapid rates of gene conversion and recombination within the 3’ end that encodes the LRRs, which have resulted in a large variety of \(RGC2\) genes.\(^ {27}\) Bioinformatic analysis of the \(Arabidopsis\) genome shows that gene-conversion events are driven by genes in clusters with sequence similarity.\(^ {61}\) The \(Arabidopsis\) gene-conversion events take place most commonly between genes that share 60–70% sequence identity, with most conversion events spanning 60–528 bp.\(^ {27,61,62}\) A greater tendency for gene conversion occurs when genes are proximal to each other, and is rarely found in genes dispersed farther away in the genome. It is not known whether gene conversion is a key mechanism for diversification in the \(185/333\) gene family. The structural components necessary for promoting gene conversion exist in the family, particularly given the significant sequence identity that is shared among the \(Sp185/333\) genes, which is based on shared element sequences.\(^ {34,39}\) Within a cluster, the presence of the microsatellites may initiate gene conversion, and then may limit the size of converted regions to block homogenization of the entire cluster.\(^ {39}\) Although it would be expected that higher sequence similarity would be present among tightly linked \(Sp185/333\) genes based on the likelihood of conversion occurring among proximal genes, comparisons among 121 genes of unknown linkage relationships from three \(S. purpuratus\) genomes show the same level of sequence similarity as genes of known linkage.\(^ {39}\) This lack of significant differences in the sequence diversity among clustered \(Sp185/333\) genes and 121 unique unlinked genes suggests that gene conversion may occur within the family among both local and more distant genes,\(^ {16}\) and that it occurs relatively swiftly within the family.

Both gene conversion and unequal crossing-over can drive gene duplication.\(^ {20}\) The most frequent duplication of whole \(NBS-LRR\) genes are tandem duplications, resulting in two similar genes in close proximity, which leads to the formation of a homogeneous gene cluster (Fig. 12.3D).\(^ {14,49,61}\) Gene duplication
and ectopic insertion of either a small set of genes or single genes to a distant location on the same or on a different chromosome (Fig. 12.3E), may also contribute to the family sequence diversity, which includes the formation of heterogeneous clusters. Chromosomal segmental duplication can affect large portions of plant genomes, and is involved in the expansion of NBS-LRR gene families. Small segmental duplications in the Sp185/333 gene cluster appears to be the source of the D1 gene duplication. Furthermore, duplications of the tandem type 1 repeats in the Sp185/333 family (Fig. 12.2A) may have been derived from ancestral sequences through duplications of the repeats, in addition to recombination and deletions, based on a computational estimation of the evolutionary history of this region of the genes. Finally, similar to the NBS-LRR gene families, meiotic mispairing (Fig. 12.3F), based on the close proximity of the Sp185/333 genes within the cluster, in addition to the sequence similarities among the genes, has been speculated to drive changes in the size of the Sp185/333 gene family.

Transposable elements may also contribute to genomic instability, which may drive diversity in both single genes and gene clusters. It has been shown that some NBS-LRR genes are associated with transposable elements. For example, the rice Xa21 gene family contains a large number of transposable elements, including LTR-retrotransposons and miniature inverted repeat transposable elements (MITEs). Fragments of transposable elements are also present within the Sp185/333 gene cluster. A portion of a Gypsy 10 long terminal repeat (LTR) S element is positioned near the 3’ end of the A2 gene in association with the flanking GA microsatellite. In addition, three tandem, incomplete Tc1-N1-SP DNA transposon fragments are positioned at the 5’ end of the E2 gene in association with the GA microsatellite. It is not known whether transposable elements contribute to the diversification of the NBS-LRR and the Sp185/333 gene families. However, we speculate that the transposable elements may contribute to the instability of the genomic regions harboring the gene families, through unequal crossing-over promoted by the duplication of transposable elements in the vicinity of members of the families.

Gene fragments and pseudogenes are commonly found in tightly linked clusters of paralogous genes, including 25% of the sea urchin SpTLR genes. It is thought that this is a result of duplication and recombination among similar genes that also promotes sequence diversification. The levels of NBS-LRR pseudogenes vary from one species of plant to another, but are generally abundant. In Arabadopisis, 8.05% of the NBS-LRR genes are pseudogenes, whereas 51.3% of the NBS-LRR gene family in two rice subspecies are pseudogenes. Contrary to the NBS-LRR family and the sea urchin SpTLR family, only one pseudogene of 171 sequenced genes has been identified in the Sp185/333 gene family. The pseudogene had no intron, and had a deletion in part of the coding region in the second exon that introduces a frame shift. Curiously for a gene family with significant levels of shared sequence within and surrounding the genes, no gene fragments have been found in the genome. The unexpectedly
low level of pseudogenes may be the result of rapid gene conversion (see previous sections) that may correct pseudogenes using sequences from nearby (and perhaps distant) genes, or alternatively by an unknown diversification regulation mechanism.

7 CONCLUSIONS

The 185/333 and the NBS-LRR gene families share several structural features, including inter- and intra-genic sequence repeats, duplicated genes, clustering, gene conversion, and diversifying selection in response to pathogens. These features are well established in the NBS-LRR gene family as components that are necessary for the initiation of a variety of diversification mechanisms. We find that the use of a comparative approach, even between echinoderms and higher plants, can be useful in understanding the biology of immune gene families, or for establishing hypotheses for how innate immune systems diversify and how potentially common mechanisms may function similarly in distantly related eukaryotes.

ACKNOWLEDGMENT

Funding to support the writing of this review was awarded by the United States National Science Foundation (IOS-1146124) to LCS.

REFERENCES


44. Krasileva KV, Dahlbeck D, Staskawicz BJ. Activation of an Arabidopsis resistance protein is specified by the in planta association of its leucine-rich repeat domain with the cognate oomycete effector. Plant Cell Online 2010;22(7):2444–58.


Chapter 13

The Evolution of the Toll-Like Receptor System

Carlos G.P. Voogdt, Jos P.M. van Putten
Department of Infectious Diseases and Immunology, Utrecht University, Utrecht, The Netherlands

1 INTRODUCTION

Ever since their emergence, multicellular hosts evolved strategies to survive in optimal symbiosis with parasitizing microorganisms. In turn, microbes continuously developed to evade the defensive barricades put up by their hosts. This ongoing evolutionary arms race has led to the development by the host of a sophisticated, germline-encoded immune system, commonly referred to as the innate immune system. The innate immune system distinguishes beneficial and harmful microbes and responds to environmental threats through an extensive arsenal of so-called pattern recognition receptors (PRR). Throughout host–microbe coevolution, these PRRs have evolved to recognize highly conserved microbe associated molecular patterns (MAMP). These cell-wall or nucleic-acid structures are essential for microbial survival, and hence difficult for the microbe to modify. Detection of these MAMPs enables the recognition of diverse microbes with a minimum set of receptors.

The best-studied family of PRRs are the Toll-like receptors (TLRs). TLRs are type I membrane-spanning glycoproteins that are typically composed of an extracellular domain (ECD), a transmembrane domain, and an intracellular signaling domain. Although the presence of TLR genes is conserved across the animal kingdom,\(^1\) TLR structure and function have diversified in response to the changing habitat and environmental challenges. The discovery of TLRs started with the identification of a protein called Toll in the fruitfly *Drosophila melanogaster*.\(^2\) Toll was identified as a regulator during embryonic development. Later, it was found that during an infection with fungi, the Toll protein was activated by its endogenous ligand, Spätzle. This activation initiated the production of antimicrobial peptides, thereby conferring immunity to fungi in *D. melanogaster*.\(^3\) A search for proteins resembling Toll in other species resulted in the discovery of a murine Toll-like receptor (TLR4). TLR4 proved to be essential for the innate recognition of bacterial lipopolysaccharide (LPS).\(^4\)
Since the discovery of TLR4 as the LPS receptor, many more TLRs with their respective microbial ligands have been identified and characterized in many different animals. The revolutionary advances in whole genome sequencing now allow studies on TLR evolution across diverse phyla. Here we will review current knowledge of the evolution of the TLR structure, diversity, and distribution across the animal kingdom, and the functions of TLRs in invertebrate and vertebrate animals, also in the context of ongoing host–microbe coevolution.

2 TLR STRUCTURE

2.1 Extracellular Domain

The extracellular domain of TLRs is composed of multiple consecutive leucine-rich repeats (LRR) that each consist of 22 to 29 amino acids with specifically placed hydrophobic residues. Upon folding, the LRRs form an arch-shaped structure in which closely packed β-sheets form the concave surface of the arch. The consecutive LRR motifs are typically flanked by an N- and C-terminal LRR (LRRNT and LRRCT, respectively) that often contain characteristically spaced cysteine residues. As also discussed in chapters: Lymphocyte Populations in Jawless Vertebrates: Insights Into the Origin and Evolution of Adaptive Immunity; Genomic Instability and Shared Mechanisms for Gene Diversification in Two Distant Immune Gene Families: The Plant NBS-LRR Genes and the Echinoid 185/333 Genes, the presence of LRR motifs in numerous proteins of animals, plants, fungi, bacteria, and viruses indicates that these motifs are of ancient origin and have remained important for protein–protein and receptor–ligand interactions throughout evolution. In 2005, the crystal structure of the first TLR ECD (human TLR3) was resolved, and, since then, additional ECD structures of different vertebrate TLRs and of \textit{D. melanogaster} Toll have been determined. Results indicate that in all TLRs the extracellular LRR motifs form an arch-shaped structure that directly interacts with a particular microbial ligand, in contrast to \textit{D. melanogaster} Toll that interacts with the endogenous cytokine Spätzle rather than directly with microbial ligands.

Throughout evolution, continuous diversification in the number and sequence of LRR motifs has resulted in an extensive family of structurally distinct TLRs. Due to the structural diversity, the TLR family is able to recognize a large array of microbial ligands, ranging from lipids to proteins and nucleic-acid motifs. Upon interaction with a ligand, the LRRCT of two monomeric TLR molecules come into close proximity, with the LRRNTs spaced far apart. This results in the formation of a homo- or heterodimeric TLR complex that obtains a somewhat “M”-shaped configuration (Fig. 13.1).

One striking example of structure-based ligand specificity is the interaction of TLR2/1 and TLR2/6 heterodimers with bacterial lipoproteins. After binding triacylated lipoproteins, TLR2 forms a heterodimer with TLR1, whereas binding of diacylated lipoproteins leads to heterodimerization with TLR6. The
The structural difference between TLR1 and TLR6 that allows for the discrimination of lipoproteins is the blocking of a lipid channel in TLR6 by only two phenylalanine residues. Substitution of these residues opens the lipid channel in TLR6 and makes the receptor receptive for binding triacylated lipoproteins, just like TLR1. This shows that minute structural differences in the TLR ECD allows TLR family members to recognize ligands with great specificity.

2.2 Transmembrane Domain

TLRs are embedded in the membrane via a single-membrane spanning region of approximately 20 amino acids. The TLR family members that recognize lipid or protein ligands are generally positioned at the cell surface, whereas TLRs that bind nucleic acid motifs are located in endosomes. Some TLRs appear in a soluble form. The soluble TLRs originate from enzymatic cleavage of the full-length receptor (TLR2), alternative splicing of the TLR gene (TLR4), or from a separate gene (TLR5). Both soluble TLR2 and TLR4 reduce the response of their membrane-bound form, and thus may act as decoy receptors that prevent an excessive response to their TLR ligands. Soluble TLR5, however, enhances the reactivity of membrane-bound TLR5 and might therefore aid in microbial detection. Soluble TLRs thus provide an additional form of structural TLR diversity.

2.3 Intracellular Domain

On the cytoplasmic side, TLRs contain a TIR domain, named after its structural and functional homology with Drosophila Toll and the Interleukin-1 receptor.
The evolutionary importance of the TIR domain is evidenced by its presence in multiple proteins of animals, plants, and even bacteria. The TIR domain of TLRs is structurally composed of five alternating β-sheets and α-helices, connected by short loops that fold into a core of β-sheets surrounded by the α-helices. Its function is to initiate downstream signaling upon ligand-induced receptor dimerization. Dimerization of TLRs brings their TIR domains in close proximity, creating a docking site for recruited adaptor proteins. The adaptor proteins also contain TIR domains and associate with TLRs through TIR–TIR domain interactions. In mammals, TLRs interact with five adaptor proteins: MyD88 (myeloid differentiation primary response protein 88), MAL or TIRAP (myeloid differentiation factor-88 adaptor-like protein), TRIF (TIR domain-containing adaptor protein inducing interferon β), TRAM (TIR domain-containing adaptor protein inducing interferon β-related adaptor molecule), and SARM (sterile α- and armadillo motif-containing protein). Based on the interaction with two major adaptor proteins, TLR signaling can be divided into two signaling routes. The MyD88-dependent route, used by all TLRs except for TLR3, results in early activation of the NF-κB (nuclear factor-κB) transcription factor that drives the transcription of proinflammatory cytokines. The TRIF pathway used by TLR3 and (often) TLR4, activates IRF (interferon regulatory factor) transcription factors that stimulate transcription of type I interferon cytokines. The MAL and TRAM adaptors facilitate the interaction between MyD88 and TRIF and TLRs, whereas SARM functions as a negative regulator of the TRIF pathway. Although considerably different in sequence and receptor makeup, both animals and plants use TIR domains and also LRR motifs in receptors involved in microbial recognition, implying that these structures may have originated before the divergence of the plant and animal kingdom.

3 EVOLUTION AND DISTRIBUTION OF TLR GENES

3.1 Origin of TLRs

Bioinformatics analysis on whole genome data indicates that prokaryotes and fungi lack TLR orthologs. Within the kingdom of plants, receptors composed of LRR motifs attached to various signaling domains (so-called Receptor-like kinases or Nucleotide-binding site LRRs) are present, but show only low sequence similarity to TLRs. Functional studies indicate that these plant receptors respond to different microbial motifs and exploit fundamentally different signaling networks compared to animal TLRs. This indicates that the LRR-containing plant receptors are not ancient TLR orthologs, but rather form separate types of plant-specific receptors that have adapted a similar function as TLRs through the process of convergent evolution. The origin of TLRs therefore lies in the animal kingdom (Metazoa).

At the root of the metazoan evolutionary tree is the phylum of sponges (Porifera) (Fig. 13.2).
The sponge species *Suberites domuncula* and *Amphimedon queenslandica* do not contain typical TLRs, but do carry TLR-related genes. The predicted proteins contain a TIR and transmembrane domain, but have a very short ECD without canonical LRRs (*S. domuncula*), or instead an extracellular immunoglobulin domain (*A. queenslandica*). TLR-related proteins are also present in species of Cnidaria (jellyfish, sea anemones, corals, and Hydra), a sister phylum of Porifera. *Hydra magnipapillata* recognizes microbial ligands through interaction of two membrane proteins, one carrying a cytoplasmic TIR domain and the other carrying extracellular LRR motifs. However, in the genome of a different cnidian, *Nematostella vectensis* (sea anemone), a typical TLR gene (ie, with TIR and LRR domains), is present. Typical TLR genes are also present in many
species within the superphylum of Bilateria. The origin of TLRs therefore dates back to the eumetazoan (all animals except Porifera) ancestor, before the separation of Cnidaria and Bilateria, approximately 600 million years ago\(^1\) (Fig. 13.2).

### 3.2 Structural Difference Between Protostomian and Deuterostomian TLRs

TLRs have been extensively conserved in bilaterian animals. The superphylum of Bilateria is divided in the protostomes (animals of which the embryonic blastophore forms the mouth) and in the deuterostomes (animals of which the embryonic blastophore forms the anus). Protostomes (including Ecdysozoa and Lophotrochozoa) carry genes encoding Toll proteins, which are structurally distinct from the deuterostomian TLRs. The protostomian Toll proteins generally contain two or more cysteine-rich clusters in the midst of their LRR motifs, whereas the deuterostome TLRs generally contain only one or two cysteine-rich clusters capping the LRR motifs (the LRRNT and LRRCT).\(^27\) Exceptions are the *D. melanogaster* and *Anopheles gambiae* (Hexapoda) Toll-9, in which LRR motifs are more similar to deuterostomian TLRs than to other Toll proteins.\(^27\) Conversely, some TLRs of the invertebrate deuterostome *Strongylocentrotus purpuratus* (sea urchin) are more similar to protostomian Tolls than to other TLRs.\(^28\) Since the gene in the sea anemone *N. vectensis* (Cnidaria) shows a higher structural similarity to protostomian Toll than to deuterostomian TLR, it is likely that the structure of Toll represents the ancestral form, whereas in deuterostomes this ancestral form independently evolved to the TLR form.\(^1,27\)

### 3.3 Toll and TLR Repertoire in Protostomes and Deuterostomes

Genomic data indicate that protostomes generally contain low to moderate numbers of Toll genes. Most arthropods (insects, crustaceans, myriapods, and chelicerates, belonging to the superphylum of Ecdysozoa) and some species of mollusks and annelids (superphylum Lophotrochozoa) have between 2 and 27 Toll genes.\(^1,29–33\) No Toll genes have been detected yet in the lophotrochozoan Platyhelminthes, which may indicate that these animals have secondarily lost their ancestral Toll genes.\(^1\) Among protostomes, extremes in the abundance of Toll genes exist. The nematode *Caenorhabditis elegans* has only one Toll gene, whereas the annelid worm *Capitella capitata* has 105 predicted Toll genes, an extensive repertoire likely created via many gene-duplication events.\(^33\)

Interestingly, vertebrates (deuterostomes) contain roughly equal numbers of TLR genes, ranging from 10 TLRs in humans to 21 TLRs in amphibians. However, the invertebrate deuterostome *S. purpuratus* (Echinodermata) has massively expanded its repertoire to 253 TLR genes. Most of the sequence diversity in these genes exists in the LRR motifs, and the TLRs more resemble each other than TLR genes from other animals. This suggests that the vast TLR repertoire in the sea urchin results from gene duplication, conversion, and/or
recombination events. Independent expansion of the TLR repertoire seems to have occurred in the invertebrate deuterostome Amphioxus (*Branchiostoma floridae*, subphylum Cephalochordata), resulting in 48 TLR genes (mixed protostome- and deuterostome-like TLRs). Perhaps these animals have compensated for their lack of an adaptive immune system by expanding their innate receptor diversity to cope with the diverse array of encountered microbial structures, a feature that may be of particular importance in the aquatic environment which these species inhabit. In contrast, the aquatic living ascidian *Ciona intestinalis*, which is an urochordate closely related to vertebrates, has only three TLR genes. Unlike vertebrate TLRs, two of the *Ciona intestinalis* TLRs recognize more than one type of ligand. This multiligand recognition of these TLRs, together with an expansion of complement factors, may have reduced the need for an expanded TLR repertoire in *C. intestinalis*.

### 3.4 Vertebrate TLR Phylogeny

Most knowledge on TLRs is based on studies of vertebrates (Fig. 13.3).

**FIGURE 13.3** Knowledge on the function of the numerous TLR genes is still marginal and dominated by studies on vertebrate TLRs. The pie charts indicate the number of TLRs per animal group, with a known function (white) or without any functional data (gray). The highest number of TLR genes currently identified per animal group are given under each chart. The estimated number of species per group is given under the group name, to put our knowledge on TLR function in perspective with species diversity.
To date, 16 TLR genes have been identified in the lamprey (jawless vertebrate), 13 have been identified in mammals, 10 in birds, 21 in amphibians, and 20 in teleost fish. Reptiles are predicted to have at least nine TLR genes. Based on sequence homology, most vertebrate TLRs can be grouped into six major families that, in general, have retained the ability to recognize distinct ligands. The large TLR1 family, consisting of TLR1, 2, 6, 10, 14, 15, 16, 18, and 25 recognizes lipoproteins (e.g., di- and tri-acylated lipopeptides). A typical member of this family is TLR15 that is activated by microbial proteolytic cleavage and TLR10 that functions as a negative regulator of TLR2. The TLR3, 4, and 5 families recognize double-stranded RNA, LPS, and bacterial flagellin, respectively. The family of TLR7, including TLR7, 8, and 9, recognize nucleic acid motifs. The sixth major family contains TLR11, 12, 13, 19, 20, 21, 22, 23, and 26. The receptors in this family that are functionally characterized, sense either protein (TLR11 and TLR12 respond to profilin of the protozoan *Toxoplasma gondii*) or nucleic acid motifs, like the TLR7 family. Especially from the large TLR1 and TLR11 families, some TLR genes appear to have been lost in various lineages, perhaps due to functional redundancy. Yet, almost all vertebrate species carry at least one gene from each of the major TLR families, underscoring the importance of innate recognition of a diverse array of microbial ligands.

An interesting exception to the conservation of the vertebrate TLR repertoire is the lack of TLR4 in some teleost fish, such as *Takifugu rubripes*. TLR4, combined with its coreceptors MD-2 and CD14, recognizes LPS, and this is of critical importance in the mammalian immune response to bacterial infections. Some fish, including zebrafish (*Danio rerio*) and common carp (*Cyprinus carpio*), do have multiple TLR4 copies but lack the TLR4 coreceptor genes. As a result, LPS sensing in fish is not mediated by TLR4. Instead, fish TLR4 appears to be a negative regulator of the proinflammatory NF-κB transcription factor. The factors driving this divergent evolution are unknown and may await analysis of TLR4 in intermediate amphibian and reptile species.

Another example of dynamic TLR evolution in vertebrates is TLR15 in the clade of Diapsida (reptiles and birds). TLR15 is only present in avian and reptilian genomes, and its TIR domain is related to members of the TLR1 family. However, unlike TLR1 family members that recognize lipopeptides, extensive sequence diversification of the TLR15 LRR motifs has led to the unique ability of this receptor to become activated by microbial proteases. Why only diapsid animals have developed this trait, and whether it provides these animals with a significant immunological benefit is unknown.

4 TLR FUNCTION

Current knowledge on the evolution of Tolls and TLRs mainly results from predictions based on genome analysis. As evolution is primarily function-driven, the challenge is to corroborate the predictions with functional evidence. At this time,
TLR-related proteins in Porifera have not been functionally characterized, and the only functional studies on cnidarian TLRs have been performed in Hydra. Hydra recognizes bacterial flagellin via an intermolecular interaction between the LRR motif and TIR domain containing proteins. In addition, Hydra deficient in the primary TLR adaptor protein MyD88 are more susceptible to infection with *Pseudomonas aeruginosa*. Although limited, these functional studies suggest that ever since its origin in the eumetazoan ancestor, the TLR system of Hydra functions in the host response to microorganisms. After the divergence of bilaterian animals, TLRs evolved independently in the protostomes and deuterostomes, which has led to functional diversification of the Toll and TLR systems.

### 4.1 Protostome Toll Function

Most knowledge of the function of protostomian Tolls comes from studies on the arthropod *D. melanogaster* (Hexapoda, Insecta), which has nine Toll genes. *D. melanogaster* Toll-1 (or simply, Toll) regulates the formation of the dorso-ventral axis in the fruit-fly embryo, but is also involved in other developmental processes, including the regulation of organogenesis, alignment, and migration of cardioblasts in the embryonic heart and neural-network development. In addition, Toll-1 initiates the synthesis of antimicrobial peptides from the fat body (equivalent of vertebrate liver) in adult flies after fungal infection. Although less well-characterized, other *D. melanogaster* Tolls also play a role in either development or immunity; Toll-2, -6, and -8 are involved in regulating the anterior–posterior axis formation in the fly embryo, whereas Toll-3, -8, and -9 are involved in the elimination of unfit cells from the developing embryo. Toll-5 is phylogenetically most similar to Toll-1 and shares the ability to activate the promoter of an antimicrobial peptide gene. Interestingly, Toll-9 is structurally more similar to deuterostomian TLRs, suggesting a function in fly immunity, but functional studies on this Toll produced inconsistent results.

Functional studies on Tolls have also been performed in other insects. In the mosquito *Aedes aegypti*, Toll-5A, a gene duplicate homologous to *D. melanogaster* Toll-5, is also involved in immunity, as deduced from increased susceptibility of the Toll-5A knockdown mutant to the fungus *Beauveria bassiana*. Other arthropods, including crustaceans, myriapods, and chelicerates, contain roughly equal numbers of Toll genes compared to insects, but their function has not been determined.

Studies on Toll in the protostomian phylum Nematoda have been limited to the model nematode *C. elegans*. *C. elegans* contains only one Toll gene (TOL-1) and lacks the MyD88 adaptor protein and NF-κB transcription factor. Loss of function of the TOL-1 gene causes severe developmental defects, as well as increased susceptibility to bacterial infection, suggesting a dual (MyD88 and NF-κB-independent) function of *C. elegans* Toll in development and in interaction with microbes, although the latter function appears less pronounced than in other protostomes like *D. melanogaster*. 
For species of the protostomian superphylum Lophotrochozoa, functional studies on Tolls and TLRs have not been reported, although genes encoding Tolls, TLRs, and signaling molecules, including MyD88 and NF-κB transcription factor, have been identified and are subject to regulation upon bacterial and fungal infection. As functional studies on arthropod and nematode Tolls have revealed diverse and specialized functions for a structurally highly related family of proteins, the study on Toll function in lophotrochozoans may provide key insights in the original function of Toll proteins in the protostomian ancestor, and the events that have driven functional diversification of protostomian Tolls.

4.2 Invertebrate Deuterostome TLR Function

Investigation of the function of TLRs and TLR-signaling molecules in diverse deuterostomian animals is still in its infancy. TLR gene-expression profiling in various tissues during different life stadia or after infection indicate that the identified genes are expressed. The presence of TLR transcript in immune cells but not sea urchin (Echinodermata) embryos suggests that TLRs play a role in the sea urchin immune system. Cephalochordates like amphioxus also have many putative TLR genes, most of which are not yet functionally explored. The TLR1 receptor of Chinese amphioxus (Branchiostoma belcheri tsingtauense) functionally interacts with an amphioxus MyD88 adaptor and activates NF-κB in a human-cell background.

Two of the three TLRs of the urochordate C. intestinalis recognize diverse microbial ligands (including flagellin and nucleic acids) and activate NF-κB in human cells. These ligands also induce expression of the proinflammatory cytokine Tumor Necrosis Factor alpha (an ancient and highly conserved inflammatory mediator) in the C. intestinalis gastrointestinal tract, suggesting that the TLRs have an immunomodulatory function. The role of the TLRs in C. intestinalis development has not been investigated.

4.3 Vertebrate TLR Function

In vertebrates, the different TLRs are variably expressed in virtually all cell types. The TLRs are considered to scan the environment for microbial ligands and to orchestrate an adequate immune defense. These events are relatively rapid and of low specificity, and form the basis of the innate immune response.
Activation of TLRs ultimately results in enhanced production of antimicrobial peptides and proinflammatory cytokines and chemokines that attract and activate professional immune cells like neutrophils and macrophages. T and B lymphocytes also express TLRs. TLR activation in diverse T-cell subsets promotes proliferation, migration, and proinflammatory cytokine production. In B-cells, activation of TLRs also induces the production of cytokines as well as the expression of costimulatory molecules and differentiation of B-cells into immunoglobulin-producing plasma cells. As indicated in several parts of this book, the immunoglobulin-based adaptive immune system is unique to vertebrates, and provides a much slower but more specific immune response than the innate response.

Besides microbial ligands, TLRs also respond to endogenous danger signals, the so-called damage associated molecular patterns (DAMPs). Endogenous activators include, among others, heat shock protein 60 (HSP60), extracellular matrix components, antimicrobial peptides, and self-nucleic acid motifs. DAMP activation of TLRs can promote a potent inflammatory response that causes excessive and detrimental immune stimulation. Indeed, the presence of high levels of endogenous TLR ligands are associated with several autoimmune diseases. The protostomian Toll receptors may also respond to endogenous danger signals. DAMPs released in *D. melanogaster* hemolymph result in the proteolytic cleavage of endogenous pro-Spätzle to form mature Spätzle, which binds to Toll-1 and initiates Toll signaling.

Overall, the apparent role of invertebrate and vertebrate deuterostomian TLRs in immunity, together with the immune function of the studied Hydra TLR-like molecules and some protostomian Tolls, strongly suggests that a role in immunity is one of the ancestral functions of these receptors. Whereas Toll has expanded its function to regulate development, TLRs in vertebrates seem to have expanded their function to initiate and regulate the vertebrate-specific adaptive immune response. However, evidence is growing that vertebrate TLRs may also have additional functions. In cardiomyocytes, TLR9 activation reduces ATP synthesis via a MyD88-independent pathway, increasing stress tolerance in these cells. This could be beneficial during myocardial ischemia. TLR9 may also be involved in correct neuron and muscle development, as mice lacking TLR9 show abnormalities in sensitivity, activity, and coordination. Other TLRs, including TLR2, 3, and 8, negatively regulate the proliferation of neuronal progenitor cells as well as the outgrowth of axons from neurons in the developing mouse brain. Stimulation of these TLRs in neurons does not activate the transcription factors that drive the immune responses, suggesting that TLRs in neurons function via an (unknown) different signaling network. These vertebrate TLR functions may resemble, to some extent, the functions of Tolls in *D. melanogaster* neuronal development. Besides its role in embryonic axis formation, Toll-1 of *D. melanogaster* is also involved in proper development of motor neurons and musculature, and Toll-6 and Toll-7 are receptors for neurotrophic factors.
5 MICROBE-DRIVEN TLR EVOLUTION

5.1 Microbial Methods to Avoid Detection by TLRs

Throughout evolution, microbes and their hosts are in a Red Queen’s race to prevent their own extinction. In order to survive and reproduce, microbes have to invent strategies to resist or evade host defenses, whereas hosts have to retaliate these strategies to prevent becoming overexploited. This also holds true for TLRs. One of the primary functions of TLRs is to detect microbes and to limit their numbers via activation of the immune system. Microbes, on the other hand, have evolved a great variety of tools to evade the TLR system.

One of the microbial evasion strategies is the degradation of TLR ligand. Some bacterial species, including the opportunistic pathogen *P. aeruginosa*, are motile by using a flagellum, which is composed of monomeric flagellin subunits. These flagellin subunits are potent activators of TLR5. However, to avoid activation of TLR5, *P. aeruginosa* secretes an alkaline protease (AprA), which degrades released monomeric flagellins (but not intact flagella), thus preventing activation of TLR5 and the development of an innate immune response. *P. aeruginosa* AprA cleaves flagellin in a domain that is conserved across bacterial species, and therefore also degrades flagellins of other bacterial species. Homologs of AprA have been identified in other flagellated bacteria, suggesting that flagellin degradation may be an evolutionary successful bacterial strategy for TLR5 evasion.

Microbes may also display virulence factors that physically block TLR recognition. Superantigen-like proteins (SSL3 and 4) of *Staphylococcus aureus* directly interact with the ECD of human and murine TLR2. This interaction presumably blocks the TLR2 ligand-binding pocket and hence prevents TLR2 from recognizing *S. aureus* cell-wall components. As a result, innate immune cells incubated with the SSLs and subsequently stimulated with TLR2 ligands show greatly impaired production of proinflammatory cytokines.

When microbes fail to escape recognition by TLRs, activation of the immune system may still be prevented by interfering with the TLR signaling cascade. Bacteria may engage host inhibitory receptors, which overrule activating receptors like TLRs. Alternatively, *S. aureus* secretes a TIR domain-containing protein (TirS), which interferes with TLR signaling and impairs NF-κB activation and cytokine production. Viruses also use molecular mimicry of signaling domains or degradation of molecules to interrupt TLR signaling.

The most obvious microbial strategy to avoid detection by TLRs is to alter the structure of main TLR ligands such as flagellins or LPS. Flagellin subunits of β- and γ-proteobacteria (eg, the genus *Salmonella*) are recognized by TLR5 and evoke an immune response. However, flagellin of α- and ε-proteobacteria (eg, the genus *Helicobacter*) is structurally different and lacks the TLR5 binding site, thus preventing detection by TLR5. Similarly, bacteria may alter their lipid A, which is the part of LPS that is recognized by TLR4, and thus impair detection by the TLR4/MD-2/CD14 receptor complex. At temperatures of its flea host (21 to 27°C), *Yersinia pestis* produces a lipid A structure that contains...
The Evolution of the Toll-Like Receptor System

Chapter | 13

The abilities of microbes to alter TLR ligand structures are however limited, as the ligands typically consist of conserved molecular patterns that are critical for microbial survival. This was nicely illustrated by substituting the TLR5 binding site in *Salmonella* flagellin with amino acids from *Helicobacter* flagellin. The *Salmonella* flagellin with *Helicobacter* amino acids could no longer be detected by TLR5, but at the cost of a loss of bacterial motility. 84

5.2 Purifying Selection on TLRs

The diversification of microbes and their TLR evasion strategies exert a selective pressure on the evolution of the TLR system. The “direction” of this selective pressure can be determined using phylogeny-based analyses of site-specific codon substitutions. By comparing TLR sequences among species, a site is identified to undergo positive selection when the ratio of nonsynonymous over synonymous codon substitutions is > 1. This indicates that a site has remained polymorphic and may provide a fitness advantage through adaptive evolution. When the ratio of nonsynonymous over synonymous codon substitutions is < 1, the codon shows no or little variation across species and has undergone purifying selection. This indicates that polymorphisms in such a site would generally be detrimental, and hence the site evolves under functional constraint. 87 Evolution under functional constraint is seen in TLR adaptor proteins, especially MyD88 and TRIF, due to their nonredundant role in signal transduction. 88, 89 The TLR adaptors interact with multiple proteins, and therefore polymorphisms would almost certainly impair their interaction with some of these proteins. 19 Maintaining function also dictates the evolution of the TIR domain 88, 90 because it shows a high level of identity across diverse species, and substituting even a single important site in the TIR domain can render the TLR inactive. 4, 91 In addition, the ECD of nucleic acid sensing TLRs (eg, TLR3, 7, 8, and 9) harbors polymorphisms, but these are rarely found in the ligand-binding region, indicating that ligand binding by these TLRs also evolves under functional constraint. 92–95 The reason for this constraint is likely the very similar structure of microbial and host nucleic acids, which poses the risk of inducing autoimmune responses. Through purifying selection, detrimental polymorphisms that may have increased the affinity to self-nucleic acids have likely been expelled from the population, thereby minimizing the risk of recognizing self-nucleic acids while maintaining adequate sensing of microbial nucleic acids.

5.3 Positive Selection on TLRs

In contrast to nucleic acid sensing intracellular TLRs, the ECD of surface-exposed TLRs (eg, TLR2, 4, and 5) displays a strong diversifying evolution
driven by positive selection of advantageous mutations. Genomic data from diverse species has allowed the identification of positively selected sites in TLR genes among primates,92 birds,93 rodents,94 cattle,96 pigs,97 and fish.98 Most of these sites are located either near to or directly in the ligand-binding region. The highly polymorphic makeup of TLR ligand-binding regions in different hosts may have been driven by antagonistic coevolution with host-specific pathogens and/or the necessity to discriminate between host-specific commensals and pathogens.99 Support for this diversifying evolution includes crystallography studies on both the human100 and mouse101 TLR4/MD2 complex that indicate that binding of LPS involves species-specific residues and functional studies that demonstrate that mouse TLR4 is activated by both hexa- and penta-acylated LPS,102,103 whereas human TLR4 only responds to hexa-acylated LPS.

Diversifying evolution may also explain the differential TLR5 response to flagellin in different species. Most of the zebrafish TLR5 residues physically interacting with Salmonella flagellin10 are poorly conserved across vertebrates. Furthermore, functional studies indicate that chicken and mouse TLR5 are more sensitive than human TLR5 to flagellin of S. enterica serovar Typhimurium.91 The recently discovered TLR5 of the reptile Anolis carolinensis (Carolina anole) proved to be more sensitive than human TLR5 to flagellins of P. aeruginosa (unpublished results).104 The species-specific recognition of TLR ligands, which has likely developed as a result of host-specific adaptations to coevolving microbial communities, may form a basis for the differential susceptibility or resistance to infection as seen among host species. Understanding the evolution of species-specific TLR functioning may thus aid in unveiling fundamental concepts behind zoonotic diseases.

6 CONCLUDING REMARKS

TLRs are among the most extensively studied innate immune receptors. Yet, there still remains much to be discovered about the evolution of this receptor family. One topic that merits more in-depth investigation is the greatly expanded number of TLR genes present in early invertebrate deuterostomes. Functional characterizations of (combinations of) TLR genes from these animals may provide evidence of independent evolution of a highly specific, sophisticated microbial recognition system based on innate immune receptors that is similar to the immunoglobulin-based adaptive immune system of vertebrates.

Another intriguing issue is the role and evolution of deuterostome TLRs beyond the immune system. Deciphering the roles for Tolls and TLRs in different organismal systems may question our view of the original function of the TLR system. Among the protostomes, the Lophotrochozoa seem to deserve more attention to elucidate the functions of the TLR genes in this superphylum.1 Potential roles for TLRs outside the immune system may also be studied in more species of Cnidaria and early deuterostomes, so that the positional origin of TLR functions can be more accurately estimated.
Finally, understanding of the evolution of TLRs would be greatly aided by much broader functional studies involving ligands from a large variety of microbes. Phylogeny-based analyses of molecular TLR evolution are able to predict residues with potential relevance for TLR function. Functional studies may provide experimental evidence and unveil the selective pressures that are at the basis of the purifying or diversifying selection of TLRs. Combined, these analyses could be instrumental in deciphering the molecular basis for antagonistic host-specific coevolution with microbes, and the resistance to disease that naturally follows. Such knowledge may be of particular interest in animal breeding to select for disease-resistant genotypes, as well as in predicting an individual’s susceptibility to disease.

REFERENCES

The Evolution of the Immune System


Chapter 14

The Evolution of Major Histocompatibility Complex in Teleosts

Masaru Nonaka, Mayumi I. Nonaka

Department of Biological Sciences, Graduate School of Science,
The University of Tokyo, Tokyo, Japan

1 GENOME ORGANIZATION

The human major histocompatibility complex (MHC) extends over four mega-base pairs, and contains more than 100 genes. The MHC’s proper function is believed to present peptides to T lymphocytes, which is performed by the alpha chains of MHC class I molecules encoded in the class I region, and the alpha and beta chains of MHC class II molecules encoded in the class II region. In addition, there are several genes with immunological functions (eg, complement Bf, C2, and C4) in the class III region of the mammalian MHC, although the physiological or evolutionary significance of this clustering of immunologically important genes has yet to be clarified. Phylogenetic studies of MHC gene organization, using the representative species of each vertebrate class, revealed that the avian chicken, amphibian *Xenopus*, and cartilaginous-fish shark have a centrally organized MHC, in which the class I, II, and III genes are linked similarly to the mammalian MHC (Fig. 14.1A). As cartilaginous fishes are considered to be the most ancient extant group to possess the MHC, such a central organization appears to be the original configuration of the MHC, which is conserved by most vertebrate groups. However, in all the teleost species analyzed thus far, class I alpha chain and class II alpha or beta chain genes are not linked (Fig. 14.1B). These results suggest that vigorous genomic rearrangements occurred in the MHC region of a common ancestor of teleost. Recent genomic analyses indicate that the linkage between class I and II genes is conserved in the basal ray-finned fish, such as the spotted gar, *Lepisosteus oculatus*, and the bichir, *Polypterus senegalus* (N.T. Fujito and M. Nonaka, unpublished), suggesting that the linkage between the MHC class I and II genes was conserved in the common ancestor of ray-finned fish, and was broken in the teleost lineage after the divergence of these basal ray-finned fish. It is tempting
FIGURE 14.1 Organization of the MHC class I region of the various animals. The brief genomic organization of the MHC region of human, chicken, Xenopus, and shark (A), and teleost medaka (B) are schematically shown. The genomic organization of human MHC region is generally conserved in other vertebrates, except for medaka. In the medaka, the MHC class II genes and the genes in the human class III region, such as complement C4, C2, and factor B genes, are not located in the MHC region but spread over several chromosomes. The general feature of the medaka MHC region is well-conserved among various teleost species.
to speculate that the dispersal of the MHC genes to several teleost chromosomes was triggered by the postulated teleost-specific tetraploidization and the following large-scale genome rearrangement, although the direct evidence to show their direct causal relationship is still missing.

Despite the absence of the linkage between the MHC class I and II genes, the teleost classical class I genes show a curious linkage with the genes directly involved in class I antigen processing/presentation, defining the teleost MHC class I region (Fig. 14.1B). These genes are proteasome subunit beta type-8 (PSMB8), PSMB9, and PSMB10 encoding subunits of the immunoproteasome responsible for production of the antigenic peptides, ATP-binding cassette, subfamily B (MDR/TAP), member 3 (ABCB3), and TAP binding protein (TAPBP). ABCB3 encodes a subunit of TAP (transporter associated with antigen presentation) responsible for peptide transportation into the endoplasmic reticulum lumen, and TAPBP encodes tapasin, believed to help peptide loading to the class I molecule by mediating interaction between class I and TAP molecules (Fig. 14.2). There is no structural similarity among the TAP, immunoproteasome, and tapasin molecules. Therefore, the linked genes in the teleost MHC class I region are structurally unrelated, but functionally they are intimately linked genes. Most of these genes are also present in the mammalian MHC region. However, some of them are not closely linked in the placental mammalian MHC, where the class I genes and the TAP/immunoproteasome subunit...
The Evolution of the Immune System

genes are separated by the class III region, which spans approximately 1 M bases (Fig. 14.1A). The close linkage between the class I and PSMB8 genes is conserved by the shark, ray-finned fish, amphibian Xenopus, reptilian anole, and basal mammals such as the platypus and wallaby. These results suggest that the close linkage between the class I gene and the genes directly involved in the class I antigen processing/presentation pathway is the original configuration of the MHC, and that the MHC of placental mammals who have lost this close linkage represents a derived situation.

Another example of the linkage between the structurally unrelated but functionally intimately linked genes found in the MHC is the linkage of the complement factor B (Bf), C2, and C4 genes, whose role has been described in detail in “The Evolution of Complement System Functions and Pathways in Vertebrates.” The Bf and C2 genes are gene-duplication products, and these two genes show a close structural similarity. On the other hand, the C4 gene has a similar structure to the C3 and C5 genes, which are totally different from the Bf and C2 structure. However, C2 and C4 are closely linked functionally because they assemble to form the classical pathway C3 convertase. The linkage between the C4 and Bf/C2 genes is conserved in shark, amphibian Xenopus, reptilian anoles, and mammals, whereas it is broken in all teleost fish analyzed thus far (Fig. 14.1B). Again, this linkage is conserved in bichir (N.T. Fujito and M. Nonaka, unpublished), suggesting that the linkages between the class I and II genes and between the C4 and Bf/C2 genes had broken simultaneously by the gross genome rearrangement that is believed to have occurred in the teleost common ancestor.

These results suggest that the original MHC configuration was the linkage among the class I genes, the class II genes, and the class III complement genes, as well as the close linkage among the class I antigen processing/presentation genes. The former linkage seems to be broken at the early stage of teleost evolution, whereas the latter linkage was retained by most teleost. On the other hand, the close linkage among the class I antigen processing/presentation genes was disrupted in the placental mammalian lineage.

2 CLASS I

The MHC class I molecules have been identified in all vertebrates except for the jawless fishes such as lampreys, suggesting that they had been established in the common ancestor of Gnathostomata after the divergence of Agnatha in concert with other molecules involved in the adaptive immune system. Agnathan immunity and its functioning in absence of MHC is presented in “Lymphocyte Populations in Jawless Vertebrates: Insights Into the Origin and Evolution of Adaptive Immunity.” The MHC class I molecule is composed of two polypeptide chains: (1) an α chain encoded by the IA gene, and (2) a β-2 microglobulin encoded by the B2M gene (Fig. 14.3A). The α chain has three extracellular domains, termed α1, α2, and α3, and a transmembrane and a cytoplasmic tail
region. The β-2 microglobulin polypeptide chain is instead composed of one IgC domain and binds to the α chain noncovalently. Most animals possess multiple copies of the IA gene within and outside the MHC region, whereas the B2M gene is located outside the MHC region in all species investigated so far, except for shark.

The IA genes are classified into two categories: classical (Ia) and nonclassical (Ib). The products of the Ia genes, such as human leukocyte antigen (HLA)-A, HLA-B, and HLA-C are expressed on most cell surfaces and functions in the presentation of intracellular peptide antigens, including the peptides derived from foreign pathogens to cytotoxic CD8+ T cells, which kill the infected cells directly. The α1 and α2 domains form a peptide-binding site and are highly polymorphic with many allelic products, which could bind various peptide antigens, whereas the α3 domain is not variable. The α3 domain interacts with the CD8 molecule, expressed on the cytotoxic T lymphocytes during recognition of the MHC class I molecules by those cells. The Ia genes are located within the MHC region in all animals, and are linked to the genes encoding the molecules directly involved in class I antigen presentation, such as immunoproteasome subunits, TAP transporter, and TAPBP. In contrast, the Ib genes are not polymorphic and are expressed in more restricted tissues. In addition, most of the amino-acid residues assigned as important for binding of the antigen peptides in the Ia molecule are often substituted by other amino acids in the Ib molecules, suggesting that Ib molecules are not involved in antigen presentation. The Ib genes contain two types of genes; one is represented by the evolutionarily ancient gene-duplication products such as the human MHC class I chain-related (MIC) genes and MHC class I-related (MIR1) genes, which have evolved independently from the Ia genes. The other type of Ib gene is represented by

![Figure 14.3](image_url) Domain structure of the MHC class I and II molecules. Both MHC class I and II molecules are heterodimer composed of the α and β chains. They have four extracellular domains, and peptide-binding groove is formed by the two membrane-distal domains, α1 and α2 in the case of class I, and α1 and β1 in the case of class II.)
relatively recent gene-duplication products such as the human \textit{HLA-E}, \textit{HLA-F}, and \textit{HLA-G}. Although the precise functions of the Ib molecules are not understood fully, the Ib genes are important for evolutionary study of the IA genes.\footnote{16}

In the teleost, the IA genes have been analyzed in various species and many IA genes are detected in each species. They are classified into several lineages by phylogenetic analysis. The most popular lineage is the U lineage, identified in all teleost species analyzed, and includes both Ia and Ib genes. The U lineage is also identified in the basal ray-finned fishes such as sturgeon\footnote{17} and spotted gar.\footnote{18} The other lineages, such as the Z lineage, include only Ib genes and have expanded mainly in the basal teleost, such as cyprinids and salmonids,\footnote{19–21} suggesting that these lineages had been established in the early days of the teleost evolution but diminished during diversification of later species, although the ZE gene in the Z lineage has been identified in a few higher teleosts, such as fugu, flounder, and medaka. Interestingly, the $\alpha_1$ and $\alpha_2$ domains of the ZE gene cluster with those of an IA gene of lungfish in the phylogenetic analysis.\footnote{22,23} However, no genes in the other lineages, including the U lineage, cluster with any IA genes of other vertebrates.

The most characteristic feature in the evolution of the teleost IA genes is the differential evolution of the three extracellular domains found in the U lineage.\footnote{22} In other vertebrates, the Ia and recently duplicated Ib genes evolved generally in a species-specific manner in all three domains, and orthologous relationships of the genes are found only among the closely related species. However, phylogenetic analysis of the teleost IA genes showed that the $\alpha_1$, $\alpha_2$, and $\alpha_3$ domains evolved in a quite different manner, and that the orthologous relationships among the distantly related species have been maintained in the $\alpha_1$ domain, whereas the $\alpha_3$ domain experienced sequence homogenization among the loci. Fig. 14.4 shows the phylogenetic trees constructed separately for the $\alpha_1$ and $\alpha_3$ domains, using the nucleotide sequences of various teleost Ia genes and some Ib genes. The $\alpha_1$ domains of the teleost IA genes, except for the zebrafish, formed four deeply diverged lineages, termed L1, L2, L3, and L4, supported with high bootstrap values (Fig. 14.4A). L1 included the genes from all acanthopterygians analyzed thus far and salmonids, suggesting that this group is the major lineage of the euteleost IA genes. L2 included the genes from the medaka and its close relatives such as Amazon molly and platyfish. L3 included the genes from all acanthopterygians, such as molly, platyfish, tilapia, cichlids, and pufferfish. L4 included the genes from the medaka, molly, tilapia, and salmonids. Three lineages except for L3 included the Ia genes. These results clearly showed the orthologous relationships among the various euteleost IA genes. Lineages L1 and L4 contain acanthopterygians and salmonids, indicating that the establishment of these lineages predated the divergence of the acanthopterygians and salmonids, which occurred approximately 260 million years ago.\footnote{24} Similarly, the establishment of L2 and L3 predated at least the divergence of medaka and molly or platyfish at 135 million years ago, and of medaka and pufferfish at 180 million years ago, respectively. The IA genes of zebrafish were
FIGURE 14.4 Phylogenetic analysis of the teleost IA genes. The phylogenetic trees were constructed by the NJ method, using the nucleotide sequences of various teleost IA genes in the U lineage separately for the α1 (A) and α3 (B) domains. The numbers on each branch represent the >750 bootstrap values calculated from 1000 replications. The genes used were the Ia and Ib genes, detected in the MHC class I region of medaka (22), fugu (11), platyfish (NCBI, genome database), guppy (NCBI, genome database), Amazon molly (NCBI, genome database), salmons (25,26) and zebrafish, and some unmapped IA genes from tilapia and cichlid. In the genes whose locus number is apparent, one allele from one locus was used in the phylogenetic trees, except for the salmonid Ia genes. Salmonids possess only one Ia gene, termed UBA, but it is exceptionally highly polymorphic in both Atlantic salmon and rainbow trout. Therefore, several alleles were added for the salmonid UBA. In addition, because Atlantic salmon and rainbow trout contain two MHC class I regions generated by tetraploidization, these regions are distinguished as _IA and _IB. The putative classical genes are shown in bold. The IA genes of sharks and some other vertebrates were added as outgroups. The accession numbers are as follows: medaka, Orla-UAA*0101, Orla-UAA*0101, and Orla-UAA*0101 (AB183488), Orla-UAA*0201 (BA000027); guppy, Pore-CAA90792 (ZS4086); Amazon molly, Pofo-UBAlike (XM_007564296), Pofo-UBAlike (XM_007564294), Pofo-UAAlike (XM_007564324); platyfish, Xima-UAAlike (XM_00580181), Xima-UAAlike (XM_005802885), Xima-UAAlike (XM_005801815); cichlid, Aua-517 (AF308550), Aua-705 (AF308551); Nile tilapia, Orni-UAA1 and Orni-UAA1 (AB270897), Orni-100701791 (XM_003459617); pufferfish, Furu-I03 (AJ271723), Teni-CR733998 (CR733998), Teni-CAG005166 (CAAE01014601); Atlantic salmon, Sasa-UAA*0601_IA (AF504013), Sasa-UAA*0101 (AF504019), Sasa-UAA*0301 (AF504022), Sasa-UAA*0101 (EF441211), Sasa-UCAp_IB (EF427379); rainbow trout, Onmy-UAA*0501_IA (AB162342), Onmy-UAA*0501_IA (AF287488), Onmy-UBA*0101, Onmy-UBA*0101, and Onmy-UBA*0101 (AB162343); zebrafish, Dare-UAA1 (Z46777), Dare-UAA1 (BC097061), Dare-UAA1 (AF220063), Dare-UBA1 (AF220063), and human, Hosa-A*290901 (AM944568).
The Evolution of the Immune System

not included in any of the four lineages, and some of them clustered with some salmonid genes with a rather lower bootstrap value, suggesting that the mode of the diversification of teleost IA genes was different in the early days of fish evolution. The results are summarized in Fig. 14.5.

In contrast, the \( \alpha_3 \) domain clustered by species or related fish groups with no exceptions, regardless of Ia or Ib types, suggesting that homogenization occurred among multiple loci within each species after speciation (Fig. 14.4B). The topology of the phylogenetic tree is the same as that believed to be true for the species evolution (also see Fig. 14.5). Because the \( \alpha_3 \) domain is known as a constant domain interacting with CD8 on cytotoxic T cells, homogenization might be necessary for retaining the potential for CD8 binding, as well as binding of the \( \beta_2M \). On the other hand, the \( \alpha_2 \) domains showed no lineages with high bootstrap values, in contrast to the \( \alpha_1 \) domains, suggesting that partial gene conversion or unequal crossing-over occurred among multiple loci during the divergence of teleosts (data not shown).\(^{22}\)

The conservation of orthologous relationships of the teleost Ia genes for over 260 million years ago is exceptionally long. The orthologs of the human Ia genes, \textit{HLA-A} and \textit{HLA-B}, are found in macaques, and the orthologs of the human Ib genes, \textit{HLA-E} and \textit{HLA-F}, are found in new world monkeys. The \textit{HLA-F} locus was estimated to have occurred at about 46–66 million years ago.\(^{32}\) However, it is hard to detect orthology between more distantly related species of the mammals. In marsupials, some Ib genes show orthology between the opossum
and wallaby that diverged about 80 million years ago, but it is difficult to detect one-to-one orthology between the Ia genes of these species.

It is probable that differential evolution of three α domains contributed to the long-term conservation of the α1 domain, the diversification of the α2 domain, and the homogenization of the α3 domain. It is speculated that genetic events such as recombination and/or gene conversion have occurred frequently in the teleost IA genes, in addition to gene duplication and deletion. Such genetic events also contributed to the allelic diversification, as seen in the salmonid Ia gene. Salmonids possess only one Ia gene termed UBA. However, the α1 domains of the UBA alleles were detected in at least two of four ancient lineages, and one was out of the lineages (Fig. 14.4A), suggesting that a sequence of an α1 domain of other loci was introduced into salmonid UBA as an allele. Interallelic recombination between the exons encoding the α1 and α2 domains has been observed in the medaka and salmonids.

The teleost MHC class I region differs from other vertebrates in that it contains no class II genes and no class III region, and that the IA genes are linked more closely with the antigen-processing genes essential for antigen presentation by the class I molecules. The tight linkage of the Ia genes and the antigen-processing genes within approximately 150 kb might facilitate coevolution of these genes, possibly providing the genetic basis for the dynamic evolution of the teleost Ia genes.

3 CLASS II

The MHC class II molecules are cell-surface glycoproteins, which bind peptides of extracellular origin, and display them to CD4+ T cells. The MHC class II molecules consist of two noncovalently-associated polypeptide chains, the α and β chains, which comprise two extracellular domains, α1 and α2, or β1 and β2, respectively, a connecting peptide, a transmembrane region, and a cytoplasmic tail (Fig. 14.3B). The mammalian MHC class II genes are classified into two categories, classical and nonclassical. The classical MHC class II genes are polymorphic, polygenic, and are constitutively expressed in professional antigen-presenting cells, such as B cells, macrophages, and dendritic cells. Interferon-γ, secreted upon infection by Th1 T cells and NK cells, induces the expression of the classical MHC class II molecules in various other cell types. The α1 and β1 domains of the classical MHC class II molecules are highly variable, and are encoded by class IIA (DXA) and class IIB (DXB) genes, respectively, where D stands for class II, X for an unspecified class-II family, and A or B for the α or β subclass. The classical MHC class II genes are present in all jawed vertebrates analyzed to date, with the exception of Atlantic cod. In addition to the classical MHC class II genes, most jawed vertebrates possess the nonclassical MHC class II genes. Although the basic structure of the nonclassical molecules are similar to their classical counterparts, they are not polymorphic, and some of them play roles other than
presenting peptides to T cells. The functions of most of them remain poorly understood.

A newly synthesized classical MHC class II molecule, HLA-DP, -DQ, or -DR in humans, is transported from the endoplasmic reticulum to the MHC class II compartment (MIIC) of endosomal compartments by binding to a protein called the invariant chain (Ii), which blocks the peptide-binding groove of a classical MHC class II molecule so that other endogeneous peptides cannot bind to the groove. In the MIIC, the Ii is cleaved by a cysteine protease, cathepsin, leaving a short fragment, called the class II-associated invariant-chain peptide (CLIP). Subsequently, the nonclassical class II molecule, HLA-DM, binds to the class II–CLIP complex and helps the class II molecule to load antigen peptides. After appropriate binding of a peptide to a class II molecule, the class II–peptide complex is transported to the cell surface, where it displays the peptide to CD4+ T cells. The DM genes have been identified not only in mammals but also in chickens and frogs, indicating phylogenetic conservation of its role to help peptide loading throughout tetrapods.

Compared to the class I genes, information on the class II genes of ray-finned fish is limited, although class II genes have been cloned from spotted gar, Ostariophysi (catfish, zebrafish, and carp), Proacanthopterygii (salmon and trout), and Acanthopterygii (fugu, stickleback, cichlid, sea bass, etc.). These class II genes of ray-finned fish, both class IIA and class IIB genes, are classified into several lineages; one lineage called the A lineage or the DA group contains all the classical class II genes identified from ray-finned fish thus far. From the primitive ray-finned-fish spotted gar, several class IIA and IIB genes have been identified. However, none of them belonged to the A lineage, and it is still an open question as to whether spotted gar has the classical MHC class II genes or not. Interestingly, these identified class II genes of spotted gar are linked to each other, and linked with the classical MHC class I gene, suggesting that the non-linkage between the MHC class I and class II genes was established only within the teleost lineage. Zebrafish belonging to Ostariophysi has both A lineage and non-A lineage MHC class II genes. However, the exact number of these genes is still not clear since the genome assembly (Zebrafish Genome Project (http://www.sanger.ac.uk/resources/zebrafish/genomeproject.html; Wellcome Trust Sanger Institute, UK) is still imperfect, and it is difficult to discriminate between alleles and loci. In Atlantic salmon, there are one pair of A lineage class II genes, IIA and IIB, and several non-A lineage class II genes. Some of them, located on chromosomes 2 and 5, are believed to have been generated by the salmonid-specific whole genome duplication. In Xiphophorus fishes, two divergent, unlinked class IIB genes, DAB and DXB, were identified. In Nile tilapia, expression of several classical class IIA and IIB genes, and at least one nonclassical class IIA gene, are reported, and the number of expressed classical genes is different, depending on the haplotype. Sato et al. reported that Nile tilapia have two divergent families of class II genes, the IIA and IIB families, which differ in their chromosomal locations, and might have split before the
separation of the teleost orders and experienced a different mode of evolution. Recently, Atlantic cod was shown to have no CD4, class II, or Ii genes. The most comprehensive analysis of MHC class II genes was performed in medaka, using the in-depth and draft genome information for the highly diverged inbred strains HdrR and HNI, respectively (http://www.shigen.nig.ac.jp/medaka/genome/top.jsp; Hyodo-Taniguchi and Sakaizumi, 1993). Medaka has five pairs of expressed class II genes, comprising one IIA and one IIB gene, located on chromosomes 3, 5, 9, 16, and 18. None of them are linked with the class I genes. Three pairs of the class II genes on chromosomes 3, 9, and 18 belong to the A lineage, and one of them on chromosome 18 shows a high degree of polymorphism, indicating that only this pair are classical class II genes.

The most characteristic feature of the teleost MHC class II genes is the absence of the DM gene. The DM molecules encoded by the DM genes, which reside in the MHC region of all tetrapod species analyzed thus far plays a pivotal role in the class II antigen presentation pathway in peptide editing and stabilization of the classical class II molecules. Thus the absence of the DM gene in the ray-finned fish suggests that the ray-finned fish possess quite different MHC class II antigen presentation system. Molecular mechanism of this system is still mostly unknown, and it is interesting to clarify whether some nonclassical class II genes of ray-finned fish are involved in class II antigen presentation or not.

4 PSMB8

PSMB8 is a catalytic subunit of immunoproteasome, which processes intracellular proteins into potentially antigenic peptides that bind with MHC class I molecules in the endoplasmic reticulum lumen. The MHC class I bound with peptides in the endoplasmic reticulum is transported to the cell surface to present antigens to the cytotoxic T lymphocytes. The 20S proteasome, a catalytic core of the larger 26S proteasome, is a large complex composed of four stacks of two outer α-rings and two inner β-rings containing seven α and seven β subunits, respectively. Among them, only three beta subunits have catalytic activity. In the constitutive proteasome, PSMB5, PSMB6, and PSMB7 are responsible for chymotrypsin-like, caspase-like, and trypsin-like activities, respectively. Upon immunological reaction, secreted IFN-γ induces PSMB8, PSMB9, and PSMB10 that replace with PSMB5, PSMB6, and PSMB7, respectively. Immunoproteasomes formed by replacement of these three subunits show a higher activity to cleave the C-terminal side of hydrophobic amino acids or basic amino acids. PSMB8-deficient mice show reduced cell-surface expression of MHC class I molecules and increased susceptibility to pathogens. In contrast, two missense mutations of the human PSMB8 gene, one affecting only the chymotrypsin-like activity and the other affecting all three peptidase activities, are reported to cause autoinflammation and lipodystrophy in homozygotic carriers, without increasing their susceptibility to pathogens. These reports suggest that PSMB8 plays some roles outside of the class I antigen
processing-pathway, and that the full physiological function of PSMB8 is still to be understood.

Dichotomous types of PSMB8 have been reported from sharks, ray-finned fish, and tetrapods. These two types possess either bulky aromatic amino-acid residues such as Phe and Tyr (termed F type), or smaller amino-acid residues such as Ala and Val (termed A type) at the 31st position of the mature protein. The 31st residue of PSMB8 is located at the entrance of the S1 pocket and is believed to affect cleaving specificity. Only the A-type allele of the PSMB8 gene has been reported from placental mammals, and human PSMB8 possessing Val at the 31st position has been reported to show chymotrypsin-like specificity, which cleaves the C-terminal side of large hydrophobic amino-acid residues. Although the cleaving specificity of F-type PSMB8 has not been directly analyzed, 3D modeling predicted elastase-like specificity cleaving the C-terminal side of smaller neutral amino-acid residues.

These two types of the PSMB8 gene show curious evolution in the ray-finned fish lineage. As shown in Fig. 14.6, the A- and F-type PSMB8 of basal ray-finned fish such as bichir (Polypteriformes), butterfly fish (Osteoglossiformes), zebrafish (Cypriniformes), salmon and trout (Salmoniformes), and pond smelt (Osmeriformes) formed type-specific clusters, termed A- and F-lineage, respectively. The allelic status of the A- and F-type PSMB8 was demonstrated in bichir and zebrafish by segregation analysis and genome analysis, demonstrating that PSMB8 is a single-copy gene in these species (N.T. Fujito and M. Nonaka, unpublished). These results indicate that the A- and F-type PSMB8 represent an extremely long-term trans-species polymorphism (TSP) that has been maintained for more than 400 million years since the divergence of Polypteriformes from the teleost lineage. TSP originally described with the mammalian class I and II genes is believed to be maintained by balancing selection, such as overdominance selection, where heterozygous individuals show a higher fitness than homozygous individuals of either alleles. All TSPs reported thus far continued for less than 100 million years, indicating that the dimorphism of the PSMB8 gene of basal ray-finned fish maintained more than 400 million years should be under unprecedentedly strong balancing selection. In contrast, all PSMB8 sequences of higher ray-finned fish such as mirror dory (Zeiformes), soldierfish (Beryciformes), medaka (Beloniformes), and puffer fish (Tetraodontiformes) are included in the A lineage. Although the possibility that the presence of the F lineage alleles in these groups were overlooked due to an insufficient number of analyzed samples cannot be totally excluded, these results suggest that the F lineage was lost in the higher teleost lineage after the divergence of Osmeriformes, but before the divergence of Zeiformes. Interestingly, however, both the A- and F-type PSMB8 are present in some Beloniformes and Tetraodontiformes fish, and both the A- and F-type alleles of these species belong to the A lineage. Phylogenetic analysis indicates that the A- and F-type alleles of each species form a species-specific cluster, suggesting that the F-type alleles were revived multiple times independently within the A lineage, although the loss of the F lineage occurred only once.
The Evolution of Major Histocompatibility Complex in Teleosts  Chapter | 14

in the common ancestor. The opposite situation is observed in Anguilliformes and Aulopiformes. Out of three Anguilliformes species analyzed, *Gymnothorax kikado* and *Anguilla japonica* possess both the A- and F-type alleles, and the *Conger myriaster* possesses only the A-type allele. All these alleles belong to the F lineage irrespective of their types, suggesting that the A lineage had been lost in the common ancestor of Anguilliformes, and A-type allele revived at least three times independently within the F lineage.

Two Aulopiformes species, *Aulopus japonicus* and *Chlorophthalmus albatrossis*, show a similar evolutionary pattern of the PSMB8 gene (Fig.14.6). Both the long-term trans-species dimorphism and recurrent recovery of dimorphism indicate that the PSMB8 dimorphism is under
exceptionally strong balancing selection.\textsuperscript{57,59,60} However, if the balancing selection is so strong, it is strange that the original dimorphism based on the ancient A and F lineages has been lost at least four times during vertebrate evolution: the loss of the F lineage in higher teleosts and tetrapods, and the loss of the A lineage in Anguilliformes and Aulopiformes. An alternative explanation of the phylogenetic tree compatible with the continuous maintenance of the PSMB8 dimorphism is interallelic sequence homogenization excluding the 31st amino acid-encoding region,\textsuperscript{56,59} as schematically shown in Fig. 14.7. Evidence for recent interallelic sequence homogenizations have been reported from Anguilliformes\textsuperscript{59} and tetrapods.\textsuperscript{56} Thus, interallelic sequence homogenization does not seem to be an exceptionally rare event. It is tempting to speculate that the evolutionary event underlying the apparent loss of the A lineage in the common ancestor of Anguilliformes and Aulopiformes, and apparent loss of the F lineage in the common
ancestor of higher teleost and in the common ancestor of tetrapods, were also interallelic sequence homogenization except for the close vicinity of the 31st residue. Once such interallelic sequence homogenization occurred between the A and F lineage sequences, the A and F alleles should show a high degree of sequence identity that could facilitate further interallelic sequence homogenization. However, if sequence homogenization has not occurred for a while just by chance, the nucleotide sequences of the A and F alleles could diverge so much that sequence homogenization cannot occur easily anymore. Sequence homogenization in each species after speciation could generate the phylogenetic pattern, apparently suggesting recurrent recoveries of dimorphism, as shown in Fig. 14.7. Whether the evolution of the PSMB8 dimorphism actually followed this sequence homogenization scenario, or the recurrent recoveries scenario, has yet to be clarified. However, the totally unprecedented pattern of evolution of the PSMB8 dimorphism suggests the presence of an exceptionally strong balancing selection that has maintained the PSMB8 dimorphism of basal ray-finned fish for more than 400 million years. Although the molecular basis for this balancing selection has yet to be clarified, there is a curious correlation between the PSMB8 dimorphism and the close linkage between the PSMB8 and MHC class I genes. The close linkage between these two genes is conserved by most jawed vertebrate groups except for birds, who lost the PSMB8 gene itself, and placental mammals. Since all PSMB8 genes of placental mammals analyzed thus far do not show dimorphism, the close linkage between the MHC class I and PSMB8 genes seems to have meaning only in the presence of the PSMB8 dimorphism. One possible explanation is that the close linkage between the MHC class I and PSMB8 genes facilitates coevolution of these two genes.

5 CONCLUSIONS

The teleost MHC is characterized by the tight linkage of the classical class I gene with the genes directly involved in class I antigen processing/presentation. The linkage between the class I and PSMB8 genes is of special interest, since this linkage seems to have enabled exceptionally dynamic evolution of these genes in teleosts. The most important evolutionary meaning of the vertebrate MHC could be to provide a genome region for the class I antigen processing/presentation genes to perform coevolution.

REFERENCES


Page left intentionally blank
Subject Index

A
Activation pathways, 152
Adaptive immune responses, 1
Adaptive immune system (AIS), 51, 151
Adenosine 5'-triphosphate (ATP), 88
Adhesion molecules, 56
Adrenalin, 231
Aedes aegypti, 319
Aeromonas hydrophilia, 89
Aeromonas salmonicida, 91
Agglutinins, 51, 52
Agnathan lymphocytes, 56
Agnathan T-like cells, antigen recognition by, 57
AIS. See Adaptive immune system (AIS)
Allantoicase, 218
Allocentrotus fragilis, 298
Amphibians, 74, 78, 282
Amphioxus, 316
AMPs. See Antimicrobial peptides (AMPs)
Anaphylatoxin receptors, 162
Angiopoietin, 279
Anguilla anguilla, 112
Anolis carolinensis, 75, 127, 324
Anopheles gambiae, 256, 257
Anopheles gambiae s.s., 241
Antibacterial activity, 39
Antigen receptors, 51
Antimicrobial peptides (AMPs), 2
Apoptosis, 51, 111
Apoptosis-inducing ligand (APRIL), 122
Appendicularians, 29
Appendicularias, 29

APRIL. See Apoptosis-inducing ligand (APRIL)
Arabidopsis thaliana, 296
Argopecten irradians, 17
Arthropoda, 6, 288
Aryl hydrocarbon receptor (AHR), 56
Ascaridia galli, 111
Ascidia, 29
Ascidian circulation, 30
Ascidian hemocytes, 31
and oxidative stress response, 40
Ascidian pigmentation, 33
Ascidians, 29
Astakines, 5
ATP-synthase, 5
Australian purple sea urchin, 298
AUUUA destabilization sequence, 92
Avian type I IFNs, 184

B
Bacterial mimicry, 290
Bacteriovorax marinus, 277
BAPs. See Bioactive peptides (BAPs)
BCAP. See B-cell adaptor protein (BCAP)
B-cell adaptor protein (BCAP), 57
B-cell CLL/lymphoma 6 (BCL6), 57
B-cell receptor-associated protein (BCAP), 52
B-cell receptors (BCRs), 52, 71, 295
B cells, 59, 70, 72
Bf/C2 genes, 334
Bf/C2-like molecules, 161
β-1,3-glucan, 4, 298
Bichir, 342
Bilaterians, 288, 315
Bilateria, 288, 315

351
Bioactive peptides (BAPs), 228
*Biomphalaria glabrata*, 7, 18, 231
*Biomphalaria straminea*, 18
Bird IL-1F3 and IL-1F5, 94–95
B-lymphocyte-induced maturation protein 1 (BLIMP-1), 57
*Bombyx mori*, 4, 229
Bony fish, 73, 77
activities of T cells against viruses, 77
allograft rejection, 77
antibody-secreting cells, 73
antigen-induced antibody production, 73
B cells and antibody production in response to pathogens, 73
dendritic cells (DC) role, 78
distribution of B cells, 73
ELISPOT technology, 73
functional activities of CD8- and CD4-bearing T cells, 78
genes coding for T-cell activities, 78
IgM-bearing lymphocytes, 73
IgM-secreting B cells, 73
immunological memory and B-T-cell cooperation, 77
innate-type B lymphocytes, 74
mucosa-associated IgT-B cells, 73
presence of cytotoxic T cells (CTL), 77
role for IgD-producing B cells, 73
T-cell activities, 77
T-cell markers, 77
teleost-related molecule CCR7, 73
in vitro cellular activities of T cells, 77
in vivo observation of functional T-cell activity, 77
*Botryllus schlosseri*, 228
*Branchiostoma belcheri tsingtauense*, 320
*Branchiostoma floridae*, 286, 316
*Branchiostoma japonicum*, 118
Brownian motion, 33
Butterfly fish, 342

**C**

*Caenorhabditis elegans*, 316
California purple sea urchin, 298
*Calliphora vomitoria*, 10, 228
*Callorhinichus mili*, 97
*Calotes versicolor*, 75
Calpains, 88
CAMs. See Cell adhesion molecules (CAMs)
*Capitella capitata*, 316
Carbohydrate recognition domains (CRDs), 205
*Carcinus maenas*, 17
Cardiotrophin-like cytokine (CLC), 100
Cartilaginous fish, 72, 76
anatomical localization of T cells, 76
B-cell subpopulations, 72
genes coding for IgM, IgW, and IgNAR, 72
IgM-mediated cytotoxicity in, 72
RAG1 and RAG2 genes, 76
Th2-related master gene regulators, 77
thymocytes expressing TRβ6, 76
TR gene systems, 76
Caspase-1, 88
Cathepsin G, 88
Caudofoveata, 6
CC. See Coiled-coil (CC) domain
CC-NBS-LRR (CNL), 296
c3-convertases, 152
CDA. See Cytidine deaminase (CDA)
CD8+ T cells, 335
Cell adhesion molecules (CAMs), 241, 260
Cell-shape changes, 230
Cell surface molecules, 56
Cellular components, 1
Central nervous system (CNS), 29, 225
Cephalothorax, 1
*Cerastoderma edulis*, 17
*Cerastoderma glaucum*, 17
*Cg-BMP7* gene, 6
*C2* genes, 334
*C3* genes, 334
*C4* genes, 334
*C5* genes, 334
*Cg-runx1* gene, 6
Chemokine receptor 9/7(CCR9/7), 56
Chemokines, 320
Chemotaxis, 152, 230
Chinese amphioxus, 320
Chitin-binding protein, 282
*Chlamys farreri*, 6, 7
Choanoflagellate proteins, 277
Chordata, 29
Chordate lineage, 275
Chordates, 281
evolution of functional diversity, 281
Chromosomal segmental duplication, 305
Chromosome 8 ORF 70 (C8orf70) genes, 98
Chymase, 88
Cl-inhibitor (C1INH), 152, 163
*Ciona intestinalis*, 32, 37, 118, 228, 282, 286, 316
Ciona savignyi, 118
Circulating hemocytes, 8
  in the hemolymph of larva of the dipteran, 13
Circulating phagocytes, 31
  of colonial ascidian Botryllus schlosseri, 32
Circulatory fluids, 1
Class II-associated invariant-chain peptide (CLIP), 340
CLC. See Cardiotrophin-like cytokine (CLC)
CLC genes, 104
CLIP. See Class II-associated invariant-chain peptide (CLIP)
Clustering, 301
Cnidaria, 315
CNL. See CC-NBS-LRR (CNL)
CNS. See Central nervous system (CNS)
CNTF genes, 103
Coagulation factor, 158
Coiled-coil (CC) domain, 296
Collagenase, 88
Common carp, 318
Complement, 38–39
Conchifera, 6
Connective tissues, 1
Corticotropin-releasing hormone (CRH), 227
C1q-receptor, 162
Crassostrea gigas, 6, 116, 217
Crassostrea rhizophorae, 17
Crassostrea virginica, 17
CRDs. See Carbohydrate recognition domains (CRDs)
CRH. See Corticotropin-releasing hormone (CRH)
Crustacean hematopoiesis, 5
Cryptosporidium baileyi, 115
C-terminal FIMAC domains, 161
Cupriavidus metallidurans, 277
CXC chemokine, expression of, 91
CXC chemokine receptor (CXCR) 4, 52
Cynoglossus semilaevis, 107
Cyprinus carpio, 318
Cytidine deaminase (CDA), 51, 52, 72
Cytofluorimetric analysis, 228
Cytokines, 37–38, 56, 87, 225
  receptors, 56
Cytolysis, 160
Cytoplasmic factors, 260
Cytosolic sensors, 173
Cytotoxic cells, 32
Cytotoxicity, 35

D
Damage associated molecular patterns (DAMPs), 321
DAMPs. See Damage associated molecular patterns (DAMPs)
Danio rerio, 71, 318
Daphnia magna, 257
Daphnia pulex, 264
Daphnia pulex, 259
Decay-accelerating factor (DAF), 152
Degranulation, 152
Degree of granulation, 15
D-E loop, 123, 124
Dendritic cells, 174
Dendritic epidermal T cells (DETC), 57
DETC. See Dendritic epidermal T cells (DETC)
Deuterostomes, 1, 286, 316
time-line extention, 286
Diapsida, 318
Distinct binding spectra of His-Type and Non-His-Type C3 Isotypes of Carp complement, 165
Diverging animals, 287
Doliolida, 29
DOPA-oxidase activity, 37
Downstream modulators of interferon response, 188
Down syndrome cell adhesion molecule (Dscam), 241, 295
alternative splicing, 243
Dscam-hv, immune-related roles of, 244
arthropod immunity, 257
bind to antigens, 259–260
isofoms, amplification of, 257
membrane-bound acts as receptor for phagocytosis, 259
pancrustaceans, immunity of, 244–257
“shedding” process, 258–259
Dscam-hv, in arthropods, 260
duplicated exons, clusters of, 265
evolutionary history in pancrustaceans, 264–265
molecular evolution compatibility, 266
origin of, 260
radiation in arthropods, 262–264
experimental methodology
  experimental host and parasite/antigen, 267
  future perspectives, 268
hypervariability, 244
  Ig domains, 242
Subject Index

immunoglobulin, 241
protein structure, 242
isoform diversity, 242
protein isoforms, 242
serpentine shape (S-shape), 242
structural aspects of, 242
variable exons, somatic combinatorial arrangements of, 266
Drosophila melanogaster, 1, 2, 4, 10, 116, 241, 257, 259, 311
Dscam. See Down syndrome cell adhesion molecule (Dscam)
Dscam genes of myriapods, 262
Dscam-hv, 263. See also Down syndrome cell adhesion molecule (Dscam)
diversity, 265
in D. melanogaster, 243
Ig domains of, 259
Ig7 orthologous domains, 265
mRNA expression, 256
overview of studies examining, 245
proteins, 259, 267
RNAi, 259
RNAi/mutagenesis, 256
role, in immunity, 268

F
Fenneropenaeus chinensis, 17
FIBCD1. See Fibrinogen C domain connecting 1 (FIBCD1)
Fibrin blood clots, 275
Fibrinogen C domain connecting 1 (FIBCD1), 280
protein, 279
Fibrinogen-related domains (FReDs), 275
and neighbor domains, 278
coled-coil domains, 278
collagen, 278
fibronectin type III, 278
immunoglobulin, 278
paralogs, 278
phylogenetic occurrence, 276
bioinformatics survey, 276
choanoflagellate Monosiga brevicollis genome, 277
choanoflagellate proteins, 277
evolutionary appearance, in sponges, 277
in protochordates, 286
cephalochordate amphioxus, 286
sea squirt, 286
structural features, 275
amino-terminal subdomain, 275
carboxy-terminal P domain, 276
NEC domain, 276
subdomains, 285
Fibrinogen-related proteins (FREPs), 295
Fibroblast growth factor receptor (FGFR), 4
Fibroleukin, 279
Fibronectin type III (Fn3), 278
Ficolin (FCN), 151, 279
Fishes, 71
acquisition of master genes, 72
antigen-specific antibodies, 72
IFNs, 185–186
lymphocytes of, 71
proteins, 285
Fn3. See Fibronectin type III (Fn3)
Fowlpox virus, 92
Francisella noatunensis, 91

E
EBI3 gene, 107
ECD. See Extracellular domain (ECD)
Echinodermata, 316
Echinoderms, 286
ECM. See Extracellular matrix (ECM)
EGF. See Epidermal growth factor (EGF)
Eimeria-infected chickens, 111
Eimeria tenella, 115
Eisenia fetida, 228
Elastase, 88
Electron microscopy, 51
Element patterns, 298
Elephant shark genome, 98
Embryogenesis, 6
Encapsulation, 34
Endogenous ligand, 311
Endoplasmic reticulum lumen, 341
Ensis siliqua, 17
Enterogona, 29, 33
Epidermal growth factor (EGF), 229
Epinephelus coioides, 106
Escherichia coli, 257, 259
ETotic cells, 15
Eukaryotes, 295
Euprepocnemis shirakii, 4
European eel agglutinin, 213
Exon repertoires, 256
Extracellular domain (ECD), 311, 312
Extracellular immunoglobulin domain, 315
Extracellular matrix (ECM), 280
FReD-containing proteins (FRePs), 275
  human, 279
  putative immune-related, 275
FReDs. See Fibrinogen-related domains (FReDs)
FREPs. See Fibrinogen-related proteins (FREPs)
FRePs. See FReD-containing proteins (FRePs)
FTLD. See F-type lectin domain (FTLD)
F-type lectin domain (FTLD), 213
F-type lectins, 213
  evolutionary aspects, 213
  functional aspects, 217
  molecular aspects, 213
  oyster binding, 217
  structural aspects, 213
fugu genome, 104

G
Galactose oxidase, 218
Galectins, 207
  evolutionary aspects, 208
  functional aspects, 210
  molecular aspects, 208
  structural aspects, 208
Gastropods, 1
Gene diversification, 300
  allelic polymorphism, 300
Gene duplication, 278, 334
  185/333 gene family, echinoids, 298
Gene organization, 125
Genetic diversification, 242
Genetic recombination, 205
Genomic modifications, 304
Germline gene, 295
GH. See Growth hormone (GH)
Glucocorticoid-induced TNF receptor ligand (GITRL), 125
Glycosphingolipids, 211
Glycosylation site, 282
Glycosylphosphatidylinositol-anchored protein, 54
Gnathostomata, 334
Gnathostomes, 69
G protein-coupled receptor, 39
Granulocytes, 14, 19
Growth hormone (GH), 227
Gut microbes, 295

H
Haemopoietic cell lines, 92
Haliotis discus discus, 116
Halocynthia roretzi, 37, 286
Heliocidaris erythrogramma, 298
Hematopoiesis, 1, 5
  crustacean hematopoiesis, 5
  insect hematopoiesis, 2–4
  molluscan hematopoiesis, 6
  in pancrustacean models, 2
Hematopoietic organs, 1, 4
Hematopoietic producing organ (HPO), 4
Hematopoietic tissues, 31
  in the gastropod, P. canaliculata, 8
Hemocoeol, 1, 9
Hemocytes, 1, 260
  circulating hemocytes in hemolymph of larva of the dipteran, 13
described in Pancrustacea on the basis of their functions, 14
  humoral factors, produced by, 37
  identified among insect orders, 11
  lineages in invertebrates, 1
  morphology, 31
  in pancrustacean and molluscan models, 9
  of pelagic tunicates, 40
  and their lineages in crustaceans, 15–17
Hemopoiesis, 1
  in pancrustacean and molluscan models, 9
Hen egg-white lysozyme (HEL), 55
Hepassocin, 279
Herpesvirus saimiri (HVS 13), 115
Heterochromatin, 51
Heterodimerization, 312
Heterophilic binding, 260
HK cells, 94
HK leukocytes, 125
HLA. See Human leukocyte antigen (HLA)
Homeostasis, 98
Homophilic binding, 242, 260
Horseshoe-crab lectin, 276
Host-microbe coevolution, 311
Host-pathogen interactions, 217
HPO. See Hematopoietic producing organ (HPO)
HR. See Hypersensitivity response (HR)
HUGO Gene Nomenclature Committee (HGNC), 116
Human genome, 279
Human leukocyte antigen (HLA), 335
Human polymorphonuclear cells, 231
Humoral components, 1
Humoral factors, produced by hemocytes, 37
Hyaline cells, 15
Hyas arenaceus, 17
**Hydra magnipapillata**, 315
Hydrolytic enzymes, 31
Hypersensitivity response (HR), 296

**IA genes**, 335
  classical, 335
  nonclassical, 335
IFN-γ inducing factor (IGIF), 92
IFN receptors, 187–188
IFN stimulated genes (ISGs), 173
IGFs. See Insulin-like growth factors (IGFs)
IgM genes, 70
Igs. See Immunoglobulin (Ig)
IL-17A/F genes, in vertebrates, 114
IL-1β converting enzyme (ICE), 88
IL-1 cytokine family, 87
IL-2 cytokine family, 96
  IL-7, 98
  IL-9, 98
  IL-21, 99
  IL-4 and IL-4/13, 97
  IL-2 and IL-2L, 96
  IL-15 and IL-15L, 99
IL-16 expression, 57
IL-10 family, 109
  IL-10, 110
  IL-22, 109
  IL-26, 110
  vIL-10, 112
IL-12 family
  in 2R Mammals, 106
  in 4R Salmonids, 106
  in 3R Teleosts, 106
IL-17 family, 112–115
IL-1F genes, 96
IL-4 gene, 97
IL-6/IL-12 superfamily, 100
  IL-6 family, 101
    CLC, CT1, and CT2, 104
    CNTF and CNTF-Like, 103
    IL-6, 101
    IL-11, 101
    LIF, OSM, and M17, 103
IL-12 family, 104
  fish IL-12 isoforms, 106
    p35, p35α, and p35β, 104
    p40, p40α, p40β, and p40c, 105
  IL-23, 106
  IL-27, 107
  IL-35, 108
IL-2-Like gene (IL-2L), 96
IL-8 receptor, 57
IL-17 receptor, 57
IL-1 receptor antagonist (IL-1RN), 95
IL-36 receptor antagonist (IL-36RN), 95
IL-1 receptor related protein (IL-1Rrp2), 95
IL-1RI receptor protein, 96
Immune cells, 1
Immune defense, adequate, 320
Immune-mobile brain, 233
Immune-neuroendocrine integration, 233
Immune-responses, 1, 229
  modifiers, 89
Immune systems, 6, 295
  adaptive, 295
  innate, 295
Immunocytes, 227, 233
  evolution of, 69
  as immune-neuroendocrine cell, 227
Immunoglobulins (Igs), 153, 205, 278
  based molecular motifs, 69
  IgM production, 69
  superfamily, 205
    Dscam, 205
    FREP, 205
    hemolin, 205
Immunological memory, 51
Immuno-procoagulant, 280
Immunoproteasome, 341
Immunostimulants, 89
Immunotoxins, 89
Incilaria fruhstorferi, 7
Infectious bursal disease virus (IBDV), 95
Infectious salmon anemia virus (ISAV)
  protein, 191
Inflammation, 35–36
Innate immune responses, 1
Insects, 289
  hematopoiesis, 2–4
Insulin-like growth factors (IGFs), 229
Integrins, 211, 282
Integrins aL (ITGAL), 56
Interferon regulatory factor (IRF), 313
Interferon (IFN) system, 173
Interleukin-1 receptor, 313
Intraepithelial γδ T cells, 57
Intra-epithelial lymphocytes (IEL), 70
Intron-exon organization, 90
Invertebrates, 205
  immune systems, 1
IRF. See Interferon regulatory factor (IRF)
isg15 gene, 191
ISG15 transcript, 191
Isoform diversity
for immunity, 241
*Ixodes scapularis*, 262
Dscam molecules, 264

J
JAK/STAT-dependent process, 109
Japanese oyster, 217
Jawless vertebrates, 51, 52
Jellyfish, 287

L
Lamprey lymphocytes
populations, features, 55
Larveaceans, 29
LAT (linker for activation of T cells), 56
*Lates calcarifer*, 189
Lectins, 37, 206
C-type, 206
evolutionary history, 206
F-type, 206
structural analysis, 206
taxonomic distribution, 206
*Lepisosteus oculatus*, 331
*Lethenteron japonicum*, 112
Leucine-rich repeat (LRR), 295, 312
based receptors, 70
family of proteins, 52
motifs, 69, 312
Leukocytes, 152
LIF gene, 103
LIF/OSM-related gene, 103
Ligand-recognition spectrum, 205
Lipopolysaccharide (LPS), 116, 151, 228,
298, 311
mediated upregulation, 95
Lipoproteins, 312
*Litopenaeus vannamei*, 228, 256, 258
*Locusta migratoria*, 4
*Lottia gigantea*, 6
LPS. See Lipopolysaccharide (LPS)
LRR. See Leucine-rich repeat (LRR)
LRRNT loop, 56
*Lymnaea stagnalis*, 18, 228
Lymph gland, 1
Lymphocyte development, in jawless
vertebrates, 59
cell fate determination of VLRA+ cells and
VLRC+ cells, 61
development of VLRA+ cells and VLRC+ cells, 59
development of VLRB+ cells, 59
evidence for selection in “thymoids,”, 61
evolution of lymphocytes in vertebrates,
61–63
lymphocyte development in jawless
vertebrates, 60
Lymphocytes, 174, 226
clones, 51
of lamprey, features of populations, 55
as neuroendocrine cell, 226
Lymphoid lineage, 1
Lymphoid organ, 9, 116
*Lytechinus pictus*, 298
Lytic pathway, 152

M
Macrophages, 88, 174
derived serine proteases, 88
progenitor cells, 103
Major histocompatibility complex (MHC), 52,
122, 153, 331
class III molecules, 334
class II molecules, 331
class I molecules, 331
Mammalian, 1
complement system, 152
CT1 and CT2, 104
EBI3, 107
genomes, 163
IFNs, 181
IL-2, 96
IL-15, 99
IL-21, 99, 100
IL-4 and IL-13, 97
IL-12 cytokine family, 100
IL-10 family in, 109
MAMP. See Microbe associated molecular
patterns (MAMP)
*Manduca sexta*, 4, 229
Mannose-binding lectin (MBL), 151, 206
*Marsupenaeus japonicus*, 116
MASP. See MBL-associated serine proteases
(MASPs)
MASP/C1r/C1s, Bf/C2, and C3/C4/
C5 families, hypothetical
diversifications of, 166
MASP2 gene, 158
MASP-like proteins, 38
Matrix metalloproteinases (MMP), 88, 92
MBL. See Mannose-binding lectin (MBL)
MBL-associated serine proteases (MASPs),
151, 206
MBL-like lectin, 153
MCP. See Membrane-cofactor protein (MCP)

Medaka, 339

Meiotic recombination, 303

α-Melanocyte-stimulating hormone (α-MSH), 227

Membrane-attack complex (MAC), 152

Membrane-bound proteins, 151

Membrane-cofactor protein (MCP), 152

Metazoan evolutionary, 314

MFAP. See Microfibril-associated protein (MFAP)

MHC. See Major histocompatibility complex (MHC)

MHC class I chain-related (MIC) genes, 335

MHC class II compartment (MIIC) of endosomal compartment, 340

MHC class II molecules, 339

classical, 339

nonclassical, 339

MHC class I molecules, 334

domain structure, 335

polypeptide chains, 334

β-2 microglobulin, B2M gene encoded, 334

α chain, IA gene encoded, 334

MHC class I-related (MR1) genes, 335

MHC-instructed T cells, 70

MIC genes. See MHC class I chain-related (MIC) genes

Microbe associated molecular patterns (MAMP), 311

Microbial neuraminidases, 213

Microfibrill-associated protein (MFAP), 279

MIIC. See MHC class II compartment (MIIC)

Miniature inverted repeat transposable elements (MITEs), 306

MITEs. See Miniature inverted repeat transposable elements (MITEs)

Mitosis, 5

Molluscan, 1, 228, 288

hematopoesis, 6

bivalves, 6

echolapods, 8

gastropods, 7

Molugula occidentalis, 228

MR1 genes. See MHC class I-related (MR1) genes

MsFBP32, 213

α-MSH. See α-Melanocyte-stimulating hormone (α-MSH)

Multiple alignment of selected vertebrate IL-18, 93

Multiple IL-1F2 genes in fish, 90

Mx proteins, 189

Mycobacterium marinum, 107

MyD88 transcription factor, 320

Myriapod S. maritima, 262

Mytilus galloprovincialis, 6

Myxovirus-resistance (Mx) proteins, 188

N

National Center for Biotechnology Information (NCBI), 277

NBS-LRR proteins, 296

Nematostella vectensis, 315

Nephrocytes, 33

Nereis diversicolor, 228

Neural-network development, 319

Neuroendocrine hormones, 229

role, 229

stress response, 229

Neuroendocrine systems, 225

Neuropeptides, 225

NF-κB transcription factor, 318

Nitric oxide (NO) production, 120

NK cell stimulatory factor (NKSF), 104

NOD-like receptor family, 179

Nonchordates, 285

Nondeuterostome animals, 286

Nonglycosylated proteins, 207

Nonlectin proteins, 218

Nonmammalian IL-1F2, 89

processing of, 91–92

Novel fish IL-1F, 96

Nuclear export sequence (NES), 95

Nuclear membrane, 51

Nucleotide-binding oligomerisation domain (NOD)-like receptors (NLRs), 173

Nucleotide binding site leucine rich repeat (NBS-LRR) genes, 295

gene family, higher plants, 296

O

Octopus briareus, 8

Octopus vulgaris, 8

Olfactoria, 29

2,5-Oligoadenylate synthase (OAS), 188

Opioid peptides, 229

Oplegnathus fasciatus, 106

Opsonic-only complement system of invertebrate chordates as a prototype, 154

Organogenesis, 319

OSA. See 2,5-Oligoadenylate synthase (OAS)

OSM knockout mice, 103
Pacifastacus leniusculus, 5
Pagocytosis, 231
Paired box protein 5 (PAX5), 57
PAMPs. See Pathogen-associated molecular patterns (PAMPs)
Pancrustacea-crustaceans, 289
Pancrustaceans, models, 2
Paralogs, 278
Pathogen-associated molecular patterns (PAMPs), 90, 151, 173, 295
recognition, 173
sensor, 154
triggered immunity (PTI), 296
Pattern recognition receptors (PRRs), 173, 206, 295, 311
PDGF. See Platelet-derived growth factor (PDGF)
p35/EBI3 heterodimer, 108
Penaeus monodon, 5, 258
PENK. See Proenkephalin (PENK)
PENK B. See Prodynorphin (PENK B)
Peptidoglycans (PGN), 116, 298
Peripheral genes, 301
Perna viridis, 17
Petromyzon marinus, 112
PGN. See Peptidoglycans (PGN)
Phagocytes, 40, 152
Phagocytosis, 33–34
Phallusia mammillata, 32
Pharynx, 29
Phenoloxidase (PO), cascade, 15
PROSITE database, 87
Protein kinase R (PKR), 188, 190
Proteolytic cascade, 151
Protochordates, 205, 282, 295
Protocollins, 283
Protostomes, 1, 316
PRRs. See Pattern recognition receptors (PRRs)
Pseudomonas aeruginosa, 318
Pseudopodoces humilis, 126
PSMB8, dimorphism evolution, 342
Psyllid, 15
PTI. See PAMPs, triggered immunity (PTI)
Putative magical binding sites, 290
Pyrosomida, 29
Radioimmunoassay (RIA), 227
RAG. See Recombination-activating genes (RAG)
Reactive oxygen species (ROS), 5, 110
Receptors, 152
Recombinant flounder FasL protein, 125
Recombination-activating genes (RAG), 70
containing immunocytes, 70
enzymes, 52
Regulators, of mammalian complement system, 152
Reptiles, 75, 78
B cells in tissues of a lizard, 75
cell-surface-associated IgM, 75
### Subject Index

#### Reptiles (cont.)
- gene sequences of Ig classes, 75
- genome cloning, 75
- IL-1F3 and IL-1F5, 94–95
- immunology, 75
- mitogenic responses to seasonal variations, 78
- mixed leucocyte reaction (MLR), 78
- phagocytic activity of B cells, 76
- presence of cytotoxic and regulatory T cells, 78
- type I IFNs, 184

### RGC2
- Resistance Gene Candidate 2 (RGC2), 302

### Resistance genes
- 295

### Reticulo-hematopoietic organ
- 4

### RIG-I (RLRs)
- Retinoic acid-inducible gene-I (RIG-I)-like receptors (RLRs), 173

### RIA
- See Radioimmunoassay (RIA)

### RNA interference (RNAi)
- 180

### RNA sequence
- of *Anopheles gambiae*, 256

### Rough endoplasmic reticulum
- 51

### Runt gene
- 6

### Salmonids
- 339, 342

### Salpida
- 29

### Samia cynthia
- 229

### Scaphopods
- 6

### Schistosoma mansoni
- 7, 164, 231

### Scophthalmus maximus
- 190

### Scyliorhinus caniculus
- 89

### Sea urchins
- 298

### Somatic diversity
- 266

### Somatic recombination
- 72

### Sperm-gamete recognition
- 217

### Sp185/333 genes, diversification mechanisms
- 303

### Spleen tyrosine kinase (Syk)
- 57

### Splice variants (SVs)
- 95

### Sponges
- 287

### Radioimmunoassay (RIA)
- 287

### Interfacial region alignment
- 288

### 20S proteasome
- 341

### 26S proteasome
- 341

### ssRNA
- See Single stranded RNA (ssRNA)

### Staphylococcus aureus
- 259

### Stat1, stat2, and irf9 genes
- 188

### Strongylocentrotus purpuratus
- 298, 316

### Structure-based ligand specificity
- 312

### Styela clava
- 228

### Styela plicata
- 228

### Suberites domuncula
- 315

### Synteny analysis
- of IL-17A/F loci from 2R tetrapods, 2R and 3R teleost species, 113

### Takifugu rubripes
- 318

### TAP binding protein (TAPBP)
- 333

### TAPBP
- See TAP binding protein (TAPBP)

### Tapes philippinarum
- 6

### TATA box
- 92

### T-cell factor 1 (TCF1)
- 56

### T-cell-like VLRA+ cells
- 57

### T-cell receptors (TCRs)
- 52, 71, 153, 295

### like gene
- 62

### T cells
- 57, 59, 76, 331, 341

### amphibians
- 78

### bony fish
- 77

### cartilaginous fish
- 76

### reptiles
- 78

### TCR/BCR genes
- 59, 62

### TCRs
- See T-cell receptors (TCRs)

### Teleost complement components
- 164

### of multiple isoforms, evolutionary significance
- 164–166

### Teleost p40c molecules
- 105

### Teleost species
- 336

### evolutionary relationships
- 338

### orthologous relationships
- 338

### phylogenetic analysis
- 337

### Teleost-specific tetraploidization
- 331

### Tenascins
- 279
T follicular helper (Thh) cells, 99
TGF. See Transforming growth factor (TGF)
Thaliaceans, 29
*Thalia* democrita, 40
Th cell markers CD4, T-bet, and GATA3, 100
Th1-like response, 92
Thymoids, 61
β-Thymosins, 5
Thyrotropin (TSH), 227
Tipartite motif-containing (TRIM) protein, 191
TIR. See Toll/interleukin-1 receptor (TIR)
TIR domain-containing protein (TirS), 322
TIR-NBS-LRR (TNL), 296
TirS. See TIR domain-containing protein (TirS)
TIR–TIR domain interactions, 313
TLRs. See Toll-like receptors (TLRs)
T lymphocytes. See T cells
TNF family, 116
amphibians, 126
basal vertebrates, 118
birds, 126
cephalochordates, 118
cyclostomata, 118
fish, 119
remaining TNFSF members, 124–126
TNFSF1-3, 119–122
TNFSF13 and 13B, 122–124
primitive chordates, 118
protostomian invertebrates, 116
reptiles, 126
urochordates, 118
TNF-like weak inducer of apoptosis (TWEAK), 125
TNF receptors, 38
TNFSF2 genes, 120, 127
TNFSF members within vertebrates, 117
TNL. See TIR-NBS-LRR (TNL)
Toll/interleukin-1 receptor (TIR), 296
Toll-like receptor 3 (TLR3), 56
Toll-like receptors (TLRs), 57, 69, 173, 311
distribution, 314
evolution, 314
family, 175–177
function, 318
invertebrate deuterostome TLR, 320
protostome toll, 319
vertebrate TLR, 320
microbe-driven evolution, 322
detection avoidance methods, 322
positive selection, 323
purifying selection, 323
origin, 314
protostomian and deuterostomian, structural difference, 316
structure, 312
extracellular domain, 312
intracellular domain, 313
transmembrane domain, 313
TLR-related proteins, 315
vertebrate phylogeny, 317
Toll proteins, 316
Toll signaling, 321
*Toxoplasma gondii*, 318
*Trachemys scripta*, 76
Trachidermus fasciatus, 283
Transcription factors, 6, 57
Transcriptome analysis, 39, 52
Transforming growth factor (TGF), 229
Transglutaminases (TGases), 5
Transmembrane (TM) domain, 207, 263, 313
Transmission electron microscopy, 2
Trans-species polymorphism (TSP), 342
*Tridacna crocea*, 18
Trophocytes, 33
Trot, 342
Trout, 342
Trout rIL-15, 99
*Trypanosoma borreli*, 90, 110
*Trypanosoma carassii*, 115
TSH. See Thyrotropin (TSH)
TSP. See Trans-species polymorphism (TSP)
Tumor necrosis factor alpha, 320
Tumor necrosis factor-alpha converting enzyme (TACE), 119
Tumor necrosis factor superfamily (TNFSF), 116
Tunicates, 29
hemocytes, immune role of, 33
phylogenetic tree of, 30

**U**

Urochordates, 29

**V**

Vanadocytes, 33
Variable domain-containing chitin binding proteins (VCBPs), 295
Variable lymphocyte receptor B (VLRB), 160
Variable lymphocyte receptors (VLR), 52
V-binding proteins, 33
VCBPs. See Variable domain-containing chitin binding proteins (VCBPs)
Vertebrates
antigen-receptors, 244
IFNs and IFN receptors, 182
Vertebrates (cont.)

- IL-1F2, 88
- IL-1F4, 92–94
  - jawless, 244
  - melanocortin system, 228
- *Vibrio anguillarum*, 283
- Viperin, 190
  - *viperin* gene, 190
- Viral-sensing pathways
  - by PRRs in vertebrates, schematic representation of, 175
- Virus-sensing pattern recognition receptors, 174
- *Viviparus contectus*, 228
- VLRA and VLRB receptors, 55
- VLRA and *VLRC* genes, 56
- VLRA⁺ cells, 59
  - and VLRC⁺ cells resemble gnathostome T cells, 56
- VLRA gene, 61
- VLRA/VLRC receptors, 57, 61
- VLR-based lymphocytes, 72
- VLRB⁺ cells resemble gnathostome B cells, 56
- VLRB gene, 56, 59, 61
- VLRB-HEL complex, 55
- *VLRC* gene, 61
- VLRC receptors, 59
- *VLR* gene, 52

- hagfish, 54
- lampreys, 54
- VLR proteins, 52
  - antigen recognition, 55
  - crystal structure, 55
  - domain organization, 53
  - gene assembly, 53
  - tertiary structure, 53

**W**

- Wallaby, 333
- WGD. See Whole-genome duplication (WGD)
- White gland, 1
- Whole-genome duplication (WGD), 97, 153
- Wine grape, 296

**X**

- *Xenopus laevis*, 74, 127

**Y**

- *Yersinia pestis*, 322
- *Yersinia ruckeri*, 109

**Z**

- Zebrafish, 71, 284, 318, 342
- Zoonotic diseases, 324
THE EVOLUTION OF THE IMMUNE SYSTEM

CONSERVATION AND DIVERSIFICATION

EDITED BY DAVIDE MALAGOLI

Davide Malagoli Department of Life Sciences, University of Modena and Reggio Emilia, Modena, Italy.

The Evolution of the Immune System: Conservation and Diversification is the first book of its kind that prompts a new perspective when describing and considering the evolution of the immune system. Its unique approach summarizes, updates, and provides new insights on the different immune components, highlighting conserved traits and remarking on the important features that diversified along evolutionary lineages.

The aim of the book is to help scientists from different educational paths and at different stages of their career to appreciate the conserved traits and the original innovations of the metazoan immune system.

Key Features

• Provides a complete overview of the most studied and hot topics in comparative and evolutionary immunology
• Reflects the conserved organization of the immune system in metazoans
• Helps the reader to gain a modern idea of the evolution of the immune system in multicellular organisms
• Brings concepts and new perspectives on the evolution of the immune system to a wide readership