314
Current Topics in Microbiology and Immunology

Editors
R.W. Compans, Atlanta/Georgia
M.D. Cooper, Birmingham/Alabama
T. Honjo, Kyoto · H. Koprowski, Philadelphia/Pennsylvania
F. Melchers, Basel · M.B.A. Oldstone, La Jolla/California
S. Olsnes, Oslo · C. Svanborg, Lund
P.K. Vogt, La Jolla/California · H. Wagner, Munich
D. B. Moody (Ed.)

T Cell Activation by CD1 and Lipid Antigens

With 25 Figures and 5 Tables

Springer
D. Branch Moody, M.D.
Brigham and Women's Hospital
Harvard Medical School
Smith 514
1 Jimmy Fund Way
Boston, MA 02115
USA
e-mail: bmoody@rics.bwh.harvard.edu

Cover Illustration:
The cover illustration is a computer generated (Maxon CINEMA 4D and Photoshop 3D) image of
the CD1b structure with a lipid antigen in the pocket. The model is rendered as a solid abstract
glass sculpture held in place by a metallic base depicting the heavy chain (red) and the light chain
(purple). An abstract glass model of a hypothetical T cell receptor with alpha and beta subunits is
suspended over the CD1. Illustration by Chris Dascher (www.cdascher.com).

Library of Congress Catalog Number 72-152360
ISSN 0070-217X

This work is subject to copyright. All rights reserved, whether the whole or part of the material
is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation,
broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Dupli-
cation of this publication or parts thereof is permitted only under the provisions of the German
Copyright Law of September, 9, 1965, in its current version, and permission for use must al-
ways be obtained from Springer-Verlag. Violations are liable for prosecution under the German
Copyright Law.

Springer is a part of Springer Science+Business Media
springer.com
© Springer-Verlag Berlin Heidelberg 2007

The use of general descriptive names, registered names, trademarks, etc. in this publication does
not imply, even in the absence of a specific statement, that such names are exempt from the
relevant protective laws and regulations and therefore free for general use.
Product liability: The publisher cannot guarantee the accuracy of any information about dosage
and application contained in this book. In every individual case the user must check such
information by consulting the relevant literature.

Editor: Simon Rallison, Heidelberg
Desk editor: Anne Clauss, Heidelberg
Production editor: Nadja Kroke, Leipzig
Cover design: WMX Design, Leipzig
Typesetting: LE-TiX Jelonek, Schmidt & Vöckler GbR, Leipzig
Printed on acid-free paper SPIN 11614562 27/3150/YL 5 4 3 2 1 0
# List of Contents

## Section I. Molecular Biology

Evolutionary Biology of CD1 ........................................... 3  
   *C. C. Dascher*

Architecture of CD1 Proteins ........................................... 27  
   *D. M. Zajonc and I. A. Wilson*

Structure and Biology of Self Lipid Antigens ......................... 51  
   *G. De Libero and L. Mori*

Structures and Functions of Microbial Lipid Antigens Presented by CD1 73  
   *B. E. Willcox, C. R. Willcox, L. G. Dover, and G. Besra*

## Section II. Cellular Biology

CD1 Expression on Antigen-Presenting Cells .......................... 113  
   *S. K. Dougan, A. Kaser, and R. S. Blumberg*

Pathways of CD1 and Lipid Antigen Delivery, Trafficking, Processing, Loading, and Presentation ................................. 143  
   *M. Sugita, D. C. Barral, and M. B. Brenner*

TCR-Mediated Recognition of Glycolipid CD1 Complexes ............... 165  
   *B. A. Sullivan and M. Kronenberg*

Development and Selection of Vα14i NKT Cells ...................... 195  
   *H. R. MacDonald and M. P. Mycko*

## Section III. Disease

CD1-Restricted T Cells in Host Defense to Infectious Diseases .......... 215  
   *S. M. Behar and S. A. Porcelli*
NKT Cells and Autoimmune Diseases: Unraveling the Complexity 251
  S. Miyake and T. Yamamura

iNKT Cells in Allergic Disease 269
  E. H. Meyer, R. H. DeKruyff, and D. T. Umetsu

CD1-Restricted T Cells and Tumor Immunity 293
  J. B. Swann, J. M. C. Coquet, M. J. Smyth, and D. I. Godfrey

Harnessing NKT Cells for Therapeutic Applications 325
  V. Cerundolo and M. Salio

Subject Index 341
List of Contributors

(Addresses stated at the beginning of respective chapters)

Barral, D. C. 143
Behar, S. M. 215
Besra, G. 73
Blumberg, R. S. 113
Brenner, M. B. 143

Cerundolo, V. 325
Coquet, J. M. C. 293

Dascher, C. C. 3
DeKruyff, R. H. 269
Dougan, S. K. 113
Dover, L. G. 73

Godfrey, D. I. 293

Kaser, A. 113
Kronenberg, M. 165

Libero, G. De 51

MacDonald, H. R. 195

Meyer, E. H. 269
Miyake, S. 251
Mori, L. 51
Mycko, M. P. 195
Porcelli, S. A. 215
Salio, M. 325
Smyth, M. J. 293
Sugita, M. 143
Sullivan, B. A. 165
Swann, J. B. 293
Umetsu, D. T. 269

Willcox, B. E. 73
Willcox, C. R. 73
Wilson, I. A. 27
Yamamura, T. 251
Zajonc, D. M. 27
Section I
Molecular Biology
Evolutionary Biology of CD1

C. C. Dascher

Center for Immunobiology, Mount Sinai School of Medicine, 1 Gustave Levy Place, Box 1630, New York, NY 10029, USA
christopher.dascher@mssm.edu

1 Introduction ........................................... 4
2 The MHC Paralogy Group ............................... 5
3 Genome Duplication and the Origin of the Proto-MHC ............. 7
4 Origin of the MHC Antigen-Presenting Molecules ..................... 10
5 Origin and Age of CD1 .................................... 11
6 CD1 and the MHC Paralogy Group ............................ 15
7 Mammalian CD1 Isoform Diversification ............................ 16
8 Microevolution of Primordial CD1 ................................. 17
9 Conclusion ............................................. 19

References .................................................. 21

Abstract  The recognition more than a decade ago that lipids presented by CD1 could function as T cell antigens revealed a startling and previously unappreciated complexity to the adaptive immune system. The initial novelty of lipid antigen presentation by CD1 has since given way to a broader perspective of the immune system's capacity to sense and respond to a diverse array of macromolecules. Some immune recognition systems such as Toll-like receptors can trace their origins back into the deep history of sea urchins and arthropods. Others such as the major histocompatibility complex (MHC) appear relatively recently and interestingly, only in animals that also possess a jaw. The natural history of CD1 is thus part of the wider story of immune system evolution and should be considered in this context. Most evidence indicates that CD1 probably evolved from a classical MHC class I (MHC I) gene at some point during vertebrate evolution. This chapter reviews the evidence for this phylogenetic relationship and attempts to connect CD1 to existing models of MHC evolution. This endeavor is facilitated today by the recent availability of whole genome sequence data from a variety of species. Investigators have used these data to trace the ultimate origin of the MHC to a series of whole genome duplications that occurred roughly 500 million
years ago. Sequence data have also revealed homologs of the mammalian MHC I and MHC II gene families in virtually all jawed vertebrates including sharks, bony fishes, reptiles, and birds. In contrast, CD1 genes have thus far been found only in a subset of these animal groups. This pattern of CD1 occurrence in the genomes of living species suggests the emergence of CD1 in an early terrestrial vertebrate.

1 Introduction

The evolutionary history of CD1 is closely connected to the larger story of the evolution of the MHC. The MHC, in turn, has emerged as an important model for the more fundamental investigation of vertebrate genome evolution. Thus it is worth examining these broader topics to provide a conceptual framework for the evolution of CD1 [43]. Most of the theoretical models related to MHC evolution rely on known genetic mechanisms such as the expansion and contraction of gene families, duplication, deletion, neofunctionalization, chromosome translocation, and even whole genome duplication (polyploidization). These are processes that have played a fundamental role in the evolution of our own genetic history and that of all metazoan genomes. This history reaches hundreds of millions of years into the past and includes species that are long extinct but whose few descendants survive today, each carrying a genetic record that can illuminate both the present and the past. Piecing together this history can thus be likened to a form of paleo-molecular biology in which DNA sequences have replaced fossilized bone and rock-hammers have been traded in for computers. The recent flood of sequence data has ignited new interest in the study of genome evolution. Moreover, comparative immunology is a major driving force for many of the ongoing genome sequencing projects given that the evolution of the immune system is one of the major innovations that permitted the expansion of complex multicellular animals.

Given the structural similarities, it is often taken for granted that CD1 and the MHC share a common ancestry. This review examines the basis for this assumption and examines how CD1 fits into the established models of MHC evolution. To address this issue, we must piece together a phylogenetic history of CD1 using both protein sequences and the known ancestry of extant (living) animal species. By combining these data, together with relevant paleontological evidence, we can begin to formulate a model for the timing, origin, and evolution of CD1. Two essential theoretical principles underlie most evolutionary analysis. The first is the fundamental assumption of cladistics that two distinct species with a common characteristic likely shared a common
ancestor from which that characteristic was derived. In the context of molecular evolution, this principle is typically applied to comparing aligned protein or DNA sequences. The second underlying assumption is that of maximum parsimony. In the context of creating a phylogenetic relationship between two characters, minimizing the total number of evolutionary steps required to explain a given set of data is probably the correct answer. These principles guide the evaluation of competing hypotheses within this chapter.

Instead of reiterating recent reviews on CD1 molecular biology and genetics in mammals, this review attempts to synthesize a number of broader topics in comparative immunology and genomic evolution with which CD1 intersects. An overview of the evolution of MHC and CD1 in the context of vertebrate evolution is presented first. This is followed by several new insights into CD1 evolution that have emerged from the recent discovery of CD1 homologs in birds. Lastly, a potential model for CD1 emergence is presented.

2 The MHC Paralogy Group

Most immunologists are familiar with the MHC locus present in the genomes of all mammals, at least in terms of the major regions devoted to the critical task of antigen presentation. It is a daunting stretch of DNA spanning over 3,600 kb of chromosome 6 in humans with 128 functional genes and 96 pseudogenes [1, 6]. Many of these genes are directly involved in various aspects of the adaptive immune system, but there are others involved in innate immunity and some that have functions completely unrelated to the immune system [40]. Interestingly, comparison of the genes in the human MHC locus to the rest of the genome has revealed the presence of duplicated or paralogous genes on multiple chromosomes [6, 21, 32, 35]. Paralogous genes arise from gene duplication within a given species at some point in its evolution. In humans, for example, the CD1A and CD1B genes are paralogs. In contrast, human CD1D and mouse CD1D are orthologs, as they derive from the same gene that was present in a common human–mouse ancestor and thus arose as the result of speciation and not gene duplication. Importantly, it is the presence and pattern of paralogous genes in a succession of animal species that has revealed a potentially key mechanism for the evolution of the MHC and perhaps all metazoan genomes.

The critical observation is that the concept of paralogy extends not just to single genes but also to large blocks of genes and even whole chromosomes. Paralogous segments of chromosomal DNA containing multiple genes (referred to as paralogons) can be found throughout the human genome [64].
The MHC locus is itself part of linear array genes on chromosome 6 in humans that is at least partially duplicated on three other chromosomes in most jawed vertebrates [2, 15, 40]. These four paralogous regions located on chromosomes 1, 6, 9, and 19 in humans are collectively referred to as the MHC paralogy group, which is shown in Fig. 1. How did these paralogous segments arise and how is this related to the evolution of CD1? With these questions in mind, we can begin by exploring the evolutionary history of the MHC, and then later how this history relates to the emergence of CD1.
3 Genome Duplication and the Origin of the Proto-MHC

A certain debt is owed to the sea for the rapid progress in the CD1 field brought about by the discovery of alpha-galactosylceramide in a marine sponge [37, 54]. Similarly, to fully appreciate the evolutionary history of the MHC, we must once again return to the sea. Here, burrowed in the sandy bottom of many shallow tropical waters, it is possible to find the sea lancet or amphioxus (see photo in Fig. 2A). These modest invertebrates are cephalochordates that are thought to represent a critical transitional form that eventually led to the emergence of true vertebrates [25]. More importantly, for investigating the evolution of the immune system, amphioxus possess a locus of genes referred to as the proto-MHC. This nomenclature does not imply that amphioxus has genes encoding the peptide-binding MHC I and MHC II antigen-presenting molecules, but rather that these animals possess a linear array of anchor genes on a single chromosome that is conserved up to and including mammals [2, 12, 15]. Figure 1 illustrates some of the anchor genes (i.e., Notch, RXR, etc.) of the proto-MHC in amphioxus. It is the proto-MHC genes that are quadruplicated on four different chromosomes in jawed vertebrates, thus forming the core gene set of the MHC paralogy group (Fig. 1). The homologs of these proto-MHC genes are found within the present-day mammalian MHC locus, and it is this cluster of genes that is thought to be the primordial genomic scaffold on which MHC I, MHC II, and other members of the MHC-based adaptive immune system later evolved.

How did this proto-MHC go from a single copy in amphioxus to four in humans and other jawed vertebrates? A clue to the answer comes from the observation that it is not only the proto-MHC that exhibits this quadruplication phenomenon. Indeed, many other paralogous segments are found that exhibit a similar 1:4 ratio when comparing invertebrate to mammalian genomes [71]. Two non-mutually exclusive theories can account for the formation of these paralogous segments of DNA. One mechanism is simple en-bloc duplication in an ancestral species, possibly by copying a whole chromosome or a large segment of chromosome and incorporating this into a new germ line [2]. Although large- and small-scale chromosome duplication indeed occurs, in this stochastic model, a more random distribution from the observed 1:4 ratio might be expected [23]. A more sweeping model is thus required to explain the consistent ratio of paralogs.

A theory that accounts for the observed paralogy of genome segments is whole genome duplication or polyploidization; essentially a doubling of the number of chromosomes [64]. Genome duplication is a deceptively simple idea first proposed by Susumu Ohno in 1970 as a mechanism to provide the
Genome duplication and CD1 locus translocation. The proto-MHC is an ancient array of genes present in invertebrates. Two successive rounds of genome duplication resulted in up to four chromosomal copies with these anchor genes (A–C). The MHC paralogy group for human chromosomes 1, 6, 9, and 19 are indicated (E). The assignment of specific chicken chromosomes to the MHC paralogy group (F) is preliminary (CCD, unpublished data). The MHC I and MHC II genes evolved within one of paralogons in an ancient ancestor of jawed vertebrates after the split from jawless fishes (C). The CD1 genes have not been identified in fish despite some efforts to find them. The CD1 genes have thus far been found only in mammals and birds, implying that CD1 was present in a common ancestor of birds and mammals but after the separation from fish; possibly in an early tetrapod (D). Furthermore, CD1 is linked to the MHC in birds (F), which supports the removal of CD1 from the MHC paralogy group despite its location on chromosome 1 in humans (E). The CD1 gene(s) likely translocated from a primordial MHC locus in an early mammalian ancestor after the bird–mammal split 310 million years ago (D). The CD1 and MHC have remained linked in birds, which is more likely to represent the ancestral state.

The theory that two sequential rounds of genome duplication events occurred over the course of vertebrate evolution is commonly referred to as the 2R hypothesis [24, 57]. As described above, the prediction of this model is the 1:4 ratio rule of paralogous gene segments (X>2X>4X) when comparing early chordates prior to genome duplication and true jawed vertebrates after these duplication events (Fig. 1). Since not all genes follow this rule, there has been
some disagreement about the validity of these segments as paralogs and of the 2R hypothesis in general [28, 29, 44]. Controversy notwithstanding, analysis of whole genome sequence data increasingly supports the primary claim of the 2R hypothesis: that genome duplication has occurred at least once (and probably twice) during the early evolution of vertebrates from more primitive chordate ancestors [19, 21, 50, 64, 72]. Figure 2 illustrates the relevant steps in this model. Recent analysis of the amphioxus proto-MHC has provided strong evidence for an initial whole genome duplication [2, 63, 72]. These studies estimate that a duplication of the proto-MHC occurred approximately 600 million years ago (Mya) in an early jawless vertebrate ancestor after splitting off from the cephalochordate (amphioxus) lineage [2, 8, 79]. A second duplication is thought to have occurred in an early jawed vertebrate ancestor after splitting off from the jawless vertebrate lineage [19, 22]. This ancestral species subsequently gave rise to all jawed vertebrates, carrying with it the four paralogous copies of the proto-MHC found today in jawed vertebrates, one of which forms the core of the actual MHC locus found in all jawed vertebrates (Fig. 2). Whether two rounds of genome duplication took place, or a single round plus extensive localized segmental duplications, is still a matter of debate. However, the nature of these duplication events should become more apparent as additional whole genome sequences are analyzed.

4 Origin of the MHC Antigen-Presenting Molecules

Having set the stage for the emergence of the MHC I and MHC II antigen presentation genes found in all jawed vertebrates, it is now rather anticlimactic to reveal that the precise progenitor of these molecules remains a mystery. This is due to the absence of clear primordial MHC I or MHC II genes in any of the known extant jawed or jawless vertebrate species. Nevertheless, several theories have been put forward as to how these genes may have initially evolved [21]. One model suggests that the MHC II genes arose first by exon-shuffling that combined an immunoglobulin-like C domain with a peptide-binding region (PBR) whose origin is unclear [38]. The MHC I heavy chain was subsequently derived by the addition of another PBR exon to the MHC II β chain [38]. This relative order of MHC origin is supported by phylogenetic analysis of the relevant exons with the split between MHC I and MHC II estimated at approximately 500 Mya, just after the Cambrian explosion [27]. It is certainly plausible to speculate that the massive adaptive radiation of multicellular animals during that time period and the emergence of the MHC are somehow linked [27, 62]. Nevertheless, genes encoding a clear primordial
PBR that forms the critical antigen-binding domains of MHC I or MHC II genes have not been identified in any extant vertebrate or invertebrate species. Despite the uncertainty of its origins, what is clear is the unambiguous presence of the MHC antigen-presenting genes in all jawed vertebrates (Gnathostomes). Therefore, the MHC I and MHC II genes likely evolved in an ancestral jawed vertebrate species very early after its split from the jawless vertebrate lineage. This timing for the emergence of the MHC-based adaptive immune system is inferred from two critical observations. The first is that virtually all jawed vertebrates share a conserved set of genes that have clear evolutionary homologs with the extensively characterized mammalian MHC-based adaptive immune system [14]. At a minimum, these include the genes that encode the αβ and γδ T cell receptors, MHC I and MHC II, and immunoglobulin molecules. Homologs of all of these components can be found in Chondrichthyes (cartilaginous fishes such as sharks, skates, and rays), teleosts (bony fishes), amphibians, reptiles, birds, and mammals [21]. The second is that no clear MHC I or MHC II homologs have been found in the more primitive jawless vertebrate species [49, 75]. Only hagfish and lampreys survive as the two living examples of the ancient lineage of jawless vertebrates and, despite some effort, no evidence of a primordial MHC gene has been found in either group. Thus, the precise origin of the MHC antigen-presenting genes remains an enigma.

5 Origin and Age of CD1

A working hypothesis of this discussion is that CD1 and MHC I share a common ancestry. This hypothesis is strongly supported by sequence alignments and structural data. However, the precise age of CD1 relative to MHC I has remained unresolved. This is due primarily to the lack of sequence data for the immune genes of lower vertebrates. Therefore, the most straightforward way to trace the origin of the CD1 genes is to identify clear orthologous sequences or transitional forms of CD1 in more phylogenetically distant species. It may then be possible to infer the age of CD1 by correlating the emergence of these genes with established vertebrate phylogenetic histories based on the paleontological record. Cartilaginous fish are the most primitive vertebrates that possess genes encoding MHC I [5, 31, 60, 65]. Thus, early efforts to determine the age of CD1 began by searching for primitive homologs in the spiny dogfish (Squalus acanthias), an abundant species of shark found in many oceans. However, clear CD1 homologs were not detected by degenerate PCR [78] or by data-mining of the available S. acanthias EST database (unpublished data).
Initial attempts to establish the timing for the origin of CD1 were also frustrated by the apparent absence of CD1 homologs in teleosts (bony fishes). Extensive searches of the public databases have thus far failed to reveal clear CD1 homologs in any of the five completed teleost genome sequences (*Danio rerio*, *Fugu rubripes*, *Tetraodon nigroviridis*, *Gasterosteus aculeatus*, *Oryzias latipes*) or other teleost EST libraries [51]. Despite the apparent absence of CD1 in both teleosts and sharks, these groups do possess highly divergent nonclassical MHC I genes [39, 70, 78]. Bearing in mind that the collection of genomic sequence data from lower vertebrates is still ongoing, the data examined thus far suggests that CD1 may not be as old as MHC I, but is rather a more recent innovation.

It is difficult to build a strong case for the emergence of CD1 based solely on the absence of these genes in lower aquatic jawed vertebrates, but that all retain MCH I genes. Nevertheless, if these observations are supported by future genomic sequence data from additional species, then it strengthens the probability that CD1 emerged after MHC I. However, there are two important qualifications that may explain the apparent absence of CD1 in teleosts and sharks. One is that extant teleosts exhibit a highly fragmented MHC that effectively unlinks MHC I, MHC II, and distributes other MHC-locus genes to virtually every chromosome in zebrafish [67, 68]. Thus, it is possible that the chaotic structure of the teleost genome may have contributed to the loss of CD1 and other genes. This fragmentation occurred after the separation of teleosts from sharks and other cartilaginous fishes. The second caveat is simply that whole genome sequence data is currently unavailable for sharks (although the available shark EST database is negative for CD1) and other ancient cartilaginous fishes. Sharks, unlike bony fish, retain the linkage of the MHC I and MHC II in a single locus and also have an overall slower rate of evolution compared to mammals [46, 58]. Therefore, these animals may have retained more primitive features of the primordial MHC locus. Further whole genome sequence data from cartilaginous fishes, teleosts, and also from a lobe-finned fish such as the Coelacanth will be required to support the findings of these initial surveys.

The failure to find CD1 homologs in sharks or bony fishes prompted the search for CD1 genes in other nonmammalian vertebrates. The chicken genome was used for this purpose since birds represent an evolutionary intermediate between mammals and fish. Three independent groups have recently characterized two clear evolutionary homologs of CD1 in the chicken *Gallus gallus* [48, 51, 66]. As described above, phylogenetic analysis assumes that a characteristic shared between two organisms implies the existence of a common ancestor from which the characteristic was derived. Therefore, evidence of CD1 homologs in both birds and mammals implies that a primordial CD1
gene was present in the common ancestor of both groups (Fig. 2). What was this common ancestor and when did it live?

The separation of birds and mammals from a common reptilian ancestor into two distinct lineages is one of the major milestones in vertebrate evolution: the Synapsid-Diapsid (S-D) split. Mammals emerged from the Synapsid lineage while birds derive from the Diapsid [41]. For the curious, these terms refer to the presence of one or two large openings in the skulls of the Synapsid and Diapsid lineages, respectively, which permit muscle attachment to the jaw. It is generally accepted that modern Aves arose from within the Archosaurus group of Diapsid reptiles, more specifically the Theropod lineage of bipedal predatory dinosaurs [69]. Importantly, it is generally accepted from fossil evidence that the divergence of the Synapsid and Diapsid lineages occurred approximately 310 Mya [41, 42]. Thus, for the first time, the age of the CD1 gene family can be pushed back to before the existence of true mammals.

The timing of the S-D split from fossil evidence is well established in paleontology and therefore serves as a robust time calibration point for molecular clock analyses [41]. The molecular clock model uses the divergence time of two lineages derived from the geologic fossil record to calibrate an absolute time-scale for the branch lengths of a phylogenetic tree generated from sequence data. Based on the survey of the teleost genomes described above, our analysis assumes that CD1 genes evolved after the split of an ancestral tetrapod from a purely aquatic fish lineage. Figure 3 shows the resulting phylogenetic analysis with chicken CD1 representing the reptilian (Diapsid) branch and a broad array of mammal CD1 protein sequences representing the Synapsid branch. The node of these two branches represents the common ancestral CD1 protein at the time of the S-D split and is set to 310 Mya, thus allowing the remaining branch lengths to be interpreted as time. Using the zebrafish MHC I as the outgroup to root this tree, the split between teleost classical MHC I and CD1 is calculated to be 384 Mya (Fig. 3). Interestingly, this timing corresponds closely with the appearance of the first tetrapods in the Devonian fossil record roughly 365–385 Mya [3]. Briefly, this is the period in which primitive sarcopterygian fish made the initial transition from the aquatic to terrestrial environments; a point in evolution recently highlighted by the discovery of a Devonian fish with tetrapod-like limbs [3, 13]. In addition, the Xenopus amphibian MHC locus has recently been sequenced but lacks CD1 genes in the region examined [59]. These data suggest that the emergence of CD1 in the vertebrate genome occurred in the reptiliform lineage after the amphibian–reptile split. The early evolutionary history of frogs and other amphibians remains unclear, so a more precise timing of this split is not yet possible. However, the cumulative data thus far supports a model in which CD1 emerged in an early terrestrial vertebrate close to the
Molecular clock analysis of the CD1 protein family. A rooted tree of aligned CD1 α1–3 protein sequences from various mammal species and the two chicken CD1 homologs was generated. Zebra fish classical MHC I was also aligned and used as the outgroup based on the assumption that MHC I and CD1 share a common ancestor. The phylogenetic tree was generated using the Neighbor-Joining method and calibrated using the bird–mammal split at 310 Mya [41, 42]. The separation of MHC I and CD1 is calculated at 384 Mya, which correlates with the approximate time (365–385 Mya) when tetrapods first appear in the fossil record. The molecular clock also indicates the separation of most mammal isoforms occurred at approximately 154 Mya. This value is consistent with a period in the Mesozoic fossil record when there were only a few mammal species. Note the transorder conservation of CD1 isoforms among diverse mammal orders, which further supports the divergence of CD1 into a multigene family in an early mammal ancestor. Bootstrap values for the nodes leading to mammal CD1 isoform divergence are indicated. A list of the taxa used for this analysis is available upon request from the author.

transition from water to land (Figs. 2 and 3). More comprehensive genome sequence data from additional extant amphibian, reptile, and fish species will be required to support this model and to further define the origin and timing of CD1 evolution.