Evolutionary redefinition of immunoglobulin light chain isotypes in tetrapods using molecular markers

Sabyasachi Das1, Nikolas Nikolaidis2, Jan Klein, and Masatoshi Nei1

Institute of Molecular Evolutionary Genetics and Department of Biology, Pennsylvania State University, University Park, PA 16802

Contributed by Masatoshi Nei, September 4, 2008 (sent for review July 24, 2008)

The phylogenetic relationships of Ig light chain (IGL) genes are difficult to resolve, because these genes are short and evolve relatively fast. Here, we classify the IGL sequences from 12 tetrapod species into three distinct groups (κ, λ, and σ) using conserved amino acid residues, recombination signal sequences, and genomic organization of IGL genes as cladistic markers. From the distribution of the markers we conclude that the earliest extant tetrapods, the amphibians, possess three IGL isotypes: κ, λ, and σ. Of these, two (κ and λ) are also found in reptiles and some mammals. The λ isotype is found in all tetrapods tested to date, whereas the κ isotype seems to have been lost at least in some birds and in the microbat. Conservation of the cladistic molecular markers suggests that they are associated with functional specialization of the three IGL isotypes. The genomic maps of IGL loci reveal multiple gene rearrangements that occurred in the evolution of tetrapod species. These rearrangements have resulted in interspecific variation of the genomic lengths of the IGL loci and the number and order of IGL constituent genes, but the overall organization of the IGL loci has not changed.

A common method of reconstructing phylogeny of a protein is to compare statistically changes at all amino acid positions in its available sequences. Often, however, the branching pattern of phylogenetic trees based on such analyses turns out to be unresolved because no single pattern is significantly supported (Fig. 1). Nevertheless, in some cases meaningful information about the protein’s evolution can be gained by focusing on highly conserved positions and using alternative replacements at these positions as cladistic markers in a phylogenetic reconstruction. Here, we use this alternative approach to shed light on the controversial issue of Ig light chain evolution.

The Ig (IG) molecule consists of two identical heavy chains (IGH) and two identical light chains (IGL) (1). Each light chain has two domains: constant (C) and variable (V). The C domain of the light chain is encoded by the IGLC (constant) gene, whereas the V domain is encoded by two kinds of genes, IGLV (variable) and IGLJ (joining), each occurring in multiple copies. For the formation of a V domain, one copy of each of these two kinds of gene comes together by a process of gene recombination. The recombination is mediated by a recombination signal sequence (RSS) composed of conserved heptamer and nonamer sequences, separated by either 12 ± 1 bp or 23 ± 1 bp spacer sequence (2). The IGLV gene consists of the complementarity-determining regions (CDRs) and framework regions (FRs). In mammals two isotypes of light chains, κ and λ chains encoded by different loci, have been identified.

The classification of IGL chains into the κ and λ isotypes was originally based on serology using rabbit antisera against human myeloma (Bence-Jones) proteins (3). Later, sequence comparisons of the human κ and λ proteins revealed them to differ by conserved amino acids replacements at multiple positions (4–6). Sequences of IGLV proteins from other mammalian species had similarly been shown to fall into two groups, one corresponding to the human κ and the other to the human λ isotypes (6). The κ and λ denotations had been extended from human to other species, including the rabbit, which cannot be assumed to possess the distinguishing antigenic determinants because otherwise it would not recognize them as foreign. Initially, when only a small number of human IGLV proteins were examined, the groupings established by serologic typing and by sequencing were congruent. Later, however, when a large number of IGLV sequences had been accumulated, the sequence-based phylogenetic trees often failed to distinguish clearly the expected two groups. This finding signaled that difficulties could be expected in an extension of this kind of phylogenetic analysis to other species. Indeed, as Fig. 1 illustrates, such difficulties have been encountered in our studies as well as those of others (7, 8). These problems have led some researchers to the use of cladistic molecular markers, such as the genomic organization of IGL loci, order of RSSs, and conserved amino acid residues for resolving the incongruencies (9, 10). The aim of the present study was to classify the tetrapod IGL genes (proteins) without resorting to common methods of phylogenetic reconstruction; to expand the repertoire of suitable molecular markers; to test the reliability of these markers in a large-scale analysis; and to redefine the isotypes of IGL in tetrapods.

Results

Phylogenetic Analysis of Light Chain Genes. We identified (see Methods) 1329 IGLV, 129 IGLJ, and 78 IGLC genes in the genomes of 12 tetrapod species (Table 1). To understand the long-term evolution of IGL isotypes we reconstructed the phylogeny of the amino acid sequences encoded by the IGLV and IGLC genes. Fig. 1 shows that the neighbor-joining tree of IGLV sequences from the four classes of tetrapods fails to resolve the phylogenetic relationships of IGL isotypes because the bootstrap values for some clades are too low for the tree to be reliable. The use of other phylogenetic methods [see supporting information (SI) Fig. S1] and different models of substitution (not shown) results in similarly unresolved topologies. The phylogenetic tree based on the C domain does not resolve the evolutionary relationships of the different IGL-encoded isotypes either (Fig. S2). These results can be explained by the short length of IGL genes and their high evolutionary rates (7). These features affect the tree topology, which changes depending on the number of sequences used, choice of species, and tree-building methods. It seems, therefore, that the conventional phylogenetic analysis is not adequate to distinguish among different IGL-encoded isotypes and infer their long-term evolutionary history.

Author contributions: S.D., J.K., and M.N. designed research; S.D. and N.N. performed research; S.D. and N.N. analyzed data; and S.D., N.N., J.K., and M.N. wrote the paper.

The authors declare no conflict of interest.

1To whom correspondence may be addressed. E-mail: sud13@psu.edu or nxm2@psu.edu.

2Present address: Department of Biological Science, California State University, Fullerton, CA 92834.

This article contains supporting information online at www.pnas.org/cgi/content/full/0808800105/DCSupplemental.

© 2008 by The National Academy of Sciences of the USA

www.pnas.org/cgi/doi/10.1073/pnas.0808800105

PNAS | October 28, 2008 | vol. 105 | no. 43 | 16647–16652
sequences previously assigned to a separate IGL, and IGVS, respectively.

human sequences of proteins originally identified as bearing the Dataset S1, Dataset S2, and Dataset S3. Groups 1 and 2 include markers in the functional IGL sequences are given in Table 2 and assigning them into three groups (groups 1–3 in Fig. 2). The gaps that were useful for distinguishing IGLV sequences and presence of several conserved amino acid residues or alignment highly variable and difficult to align. Our analysis revealed the tetrapod species. We used only FRs because CDR sequences are (FR1–FR3) of functional sequences identified from the 12 sequences. For the IGLV, we aligned the framework regions markers characterizing IGL (IGLV, IGLJ, and IGLC) protein isotypes in tetrapods based on the use of cladistic molecular we resorted to an alternative approach to distinguish the IGL option and the Poisson correction distance. The known human IGL bootstrap value. The tree was constructed by using the pair-wise deletion level. Filled circles indicate that the interior branches are supported by 90% consensus (DEAD (Asp-Glu-Ala-Asp) motif in group 2 sequences, but this motif is degenerate in group 1 sequences. On the basis of these results, the human IGVK (κ variable) and IGVL (λ variable) sequences (Ensembl annotation) and the frog IGVS (σ variable) sequence (accession no. S78544 and NP.001087883) belong to group 1, 2, and 3 sequences, respectively.

Cladistic Markers of IGL Sequences. The three groups defined by the markers of IGLV sequences are also supported by a cladistic analysis of the IGLJ sequences (Fig. 2). In particular, the Gly at position 4 and Thr at position 7 discriminate the group 3 sequences from the groups 1 and 2 sequences. The Thr at position 2 discriminates the group 1 sequences from the other two groups. Group 1 sequences contain the EIK (or occasionally ELK) motif at positions 10–12, whereas group 2 and group 3 sequences contain the TVL (or TVT) and IVT motifs, respectively. These results again show that group 1, 2, and 3 sequences correspond to the human IGJK, IGJL, and frog IGJS sequences, respectively (Fig. 2).

Cladistic Markers of IGLC Sequences. Analysis of IGLC sequences confirmed the existence of the three isotypic groups (Fig. 3). Specifically, by the presence of distinct amino acid residues at positions 14, 32, 34, 79, and 91, group 3 sequences could be discriminated from the sequences belonging to groups 1 and 2. Residues at positions 17, 56, 60, 65, and 102 discriminated group 1 sequences from group 2 sequences. These markers discriminate IGL isotypes in all of the species examined, including the frog. There are also two other markers, which can also distinguish IGL isotypes in all of the species except the frog. In one case, the Cc domains lack the residues at these two positions. In the second case, at position 79 Cc and CA domains lack residues at these positions (Fig. 3). In the frog, both Cc and CA domains lack the residues at these two positions. In the second case, carbon at position 79 Cc and CA domains have Tyr and Trp, respectively, in all of the species except frog, in which Cc has Trp at this position. The two alternative explanations of these exceptional cases could be independent mutations at these sites in the frog lineage or retention of an intermediate stage in the evolutionary split of IGCK and IGCL genes.

Taken together, the cladistic markers in the IGLV, IGLJ, and IGLC proteins indicate that the sequences of groups 1, 2, and 3 correspond to the κ, λ, and σ isotypes, respectively. Moreover, these markers enable us to redefine several unclassified or misclassified IGL sequences from the frog (see Dataset S4). In additions, the cladistic markers strongly suggest that the rho and type III sequences (12, 13) previously identified in frogs are actually the κ and λ isotypes, respectively. In addition, the markers enable us to identify two IGL isotypes in lizards as the κ and λ isotypes. To our knowledge, this is the first evidence that reptiles possess these two isotypes. Previously Saluk et al. (1970) had shown the presence of two antigenically distinct light chain types in alligator (14), but the specific isotypes of these chains were not determined.

Analysis of RSS. Comparison of RSSs revealed that group 1 (κ) and group 3 (σ) IGLV genes have an RSS with a 12-bp spacer.
at their 3’ end, whereas group 2 (λ) IGLV genes have an RSS with a 23-bp spacer. The κ and σ IGLL genes have an RSS with a 23-bp spacer, whereas the JA genes are flanked by an RSS with a 12-bp spacer at their 5’ end (Fig. 2).

Genomic Organization of the Ig Light Chain Loci. Comparative analysis of the genomic organization of IGL loci provides yet another isotype-discriminating marker. As expected, we found that the genes encoding the three IGL isotypes (κ, λ, and σ) are located in different genomic regions (Table 1, Fig. 4). In addition, the genomic organizations of these three regions are different. In the κ-encoding locus of all tetrapods studied, multiple IGLK genes are present in a cluster, followed by a single IGCK gene, whereas in the λ-encoding locus IGJL and IGCL genes occur as IGJL–IGCL blocks, usually present in multiple copies (except in chicken, in which there is only one IGJL–IGCL block). In the λ locus, the IGIL gene is located within a 3-kb 5’ region from the IGCL gene, whereas in the κ locus the IGJK cluster is located within a 6-kb 5’ region from the IGCK gene. The σ locus of the frog has a genomic organization similar to that found in the κ locus of tetrapods (Fig. 4).

Analysis of Genes Flanking the IGL Loci. The 3’ end of the human κ-encoding locus is flanked by the non-IG genes RPLA and EIF2AK3, whereas the 5’ end is flanked by L5P1 and DRD5P1 (see Fig. 4 and ref. 10 for these abbreviated genes). The 3’ end of the κ-encoding locus is flanked by the RPLA and EIF2AK3 genes in other tetrapods as well, except in frogs, in which only the RPLA gene is present. The 5’ flanking genes of the κ-encoding locus either are present at different genomic locations, or their homologues could not be identified in other species. The 3’ end of the human λ-encoding locus contains the GNAZ and the RTRDRI genes, and the 5’ end contains the TOPSB and PPM1F genes. In other species studied these genes are not in conserved synteny with the λ-encoding locus, although they are generally located on the same chromosome (or scaffold) as that of the λ-encoding locus, and their distances from the λ-encoding loci are not conserved. In the lizard the chromosomal position of GNAZ and RTRDRI could not be identified with confidence owing to the preliminary nature of the genome assembly (Fig. 4).

Number of IGL Genes in Extant Species. The numbers of putatively functional genes and pseudogenes in different species of tetrapods are highly variable (Table 1). The numbers of IGVK and IGVL genes also vary among different evolutionary lineages. In rodents, IGVK genes are more abundant than IGVL genes. By contrast, no IGVK genes have been found in the microbat, in which IGVL genes are well represented. In the human, the IGVK and IGVL genes are present in similar numbers. The two nonplacental mammals (opossum and platypus) differ considerably in the numbers of IGVK and IGVL genes. There is only one functional IGVL gene but 24 IGVL pseudogenes in the chicken genome. In the lizard there are 19 IGVK and 38 IGVL genes. The copy numbers of IGCL, IGJL, and IGIK genes also vary somewhat from species to species. The IGCK gene, on the other hand, is present in a single copy in all species examined (Table 1).

Discussion

We have identified 21 clades in the class mammal in IGL polypeptides that differentiate IGL genes of tetrapods into three isotypic groups. All three of these groups are present only in frog, whereas most of other tetrapod species examined have two groups (the κ and λ isotypes), but some (birds and microbats) have only one group (the λ isotype). We retain the traditional designations of the isotypic groups (κ, λ, and σ) but redefine their characterization. Instead of the original characterization of the κ and λ isotypes based on the presence or absence of serologically detectable antigenic determinants (3, 15), or of all three isotypes on the basis of overall sequence similarity (11–13, 16), we now define them by the possession of the described 21 characteristic states in the three groups. Underlying this redefinition is the congruence in the distribution of the character states in the three groups. The congruence is so high that the isotype of an unknown IGL sequence (including pseudogenes and partial sequences) can be identified by determining the character state of only one or a few markers. The distribution of isotypes in the tetrapods indicates that (i) in tetrapods two of the three IGL isotypes (κ and λ) can be traced back from mammals to amphibians, (ii) the amphibian class possesses a third isotype (σ), and (iii) the κ isotype seems to have been lost in at least some birds and in the microbat. The large number of character state differences (including genomic organization) conserved from amphibians to mammals suggests an ancient divergence of the κ and λ lineages, probably preceding the splitting of the extant tetrapod lineages. Similarly, the sharing of character states in nearly equal numbers between the σ and κ as well as between σ and λ isotypes suggests ancient divergence of σ from the ancestor of the κ and/λ. An extension of the cladistic analysis to the fishes would provide a test of these predictions.
Table 2. Summary of the molecular markers in IGL protein sequences

<table>
<thead>
<tr>
<th>Gene</th>
<th>Position</th>
<th>Markers</th>
<th>α</th>
<th>λ</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IGLV</strong></td>
<td>7</td>
<td>Gap</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S/T</td>
<td>+ (97)</td>
<td>- (97)</td>
<td>+ (100)</td>
</tr>
<tr>
<td></td>
<td>41a</td>
<td>Gap</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
</tr>
<tr>
<td></td>
<td>46 a-b</td>
<td>Gap</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
</tr>
<tr>
<td></td>
<td>53</td>
<td>α</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
</tr>
<tr>
<td></td>
<td>64–67</td>
<td>DEAD</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
</tr>
<tr>
<td><strong>IGLJ</strong></td>
<td>2</td>
<td>T</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>G</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>T</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
</tr>
<tr>
<td></td>
<td>10–12</td>
<td>E(I)K TVL(AVT)</td>
<td>(98)</td>
<td>(100)</td>
<td>(100)</td>
</tr>
<tr>
<td><strong>IGLC</strong></td>
<td>14</td>
<td>S</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>Q E</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
</tr>
<tr>
<td></td>
<td>20a</td>
<td>Gap</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>F</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>P</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
</tr>
<tr>
<td></td>
<td>56</td>
<td>Gap</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>D/E K/R</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
</tr>
<tr>
<td></td>
<td>65</td>
<td>T</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
</tr>
<tr>
<td></td>
<td>91</td>
<td>H</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
</tr>
<tr>
<td></td>
<td>102</td>
<td>F</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
</tr>
</tbody>
</table>

*+, −: presence and absence of amino acid residue(s) or gap in the IGL sequences. The number in parentheses is the percentage of occurrence of each marker in the functional IGL sequences. The two markers in IGLC sequences at positions 79 and 94–95, which distinguish IGLC sequences except frog, are not listed here (see Fig. 3).
sequences have been related to different pathophysiologic conditions. The $\lambda$-bearing antibodies have been found in more than 60% of light-chain amyloidosis, whereas the $\kappa$ antibodies have been shown to mediate more than 85% of cases of the light-chain deposition disease (22). In addition, in the B cell receptor of hairy cell leukemia $\lambda$ chains has been shown to be preferentially used (23). Yet another well-established functional difference between $\kappa$ and $\lambda$ antibodies is the interaction of $\kappa$ chains with the bacterial surface protein L (20, 24). It has also been demonstrated that the isotype of the IGL chain can influence the kinetics of the intracellular assembly of antibodies and the susceptibility of the interchain disulfide bonds to reducing agents (25). A different kind of functional differentiation between $\kappa$ and $\lambda$ antibodies was indicated by Wardemann et al. (2004) (26). These investigators found that $\lambda$ genes are more effective than $\kappa$ genes as silencers of autoantibody production. Taken together, all these reports suggest that $\kappa$ and $\lambda$ isotypes are not functionally equivalent.

Fig. 3. Alignment of representative IGLC sequences. The conserved markers that can be used for identifying the three isotypes corresponding to group 1–3 genes are highlighted. The two markers at position 79 and 94–95 (not highlighted) can also distinguish the $\kappa$ and $\lambda$ isotypes except frog. The numbering of amino acid positions is based on the human IGCK and IGCL sequences (Ensembl annotation), and the gap relative to the frog IGCS sequence is indicated by “a.” Xt and Xl stand for frog species X. tropicalis and X. laevis, respectively. The aligned sequences of all functional IGLC genes are given in Dataset S3.

Fig. 4. Schematic diagrams of the genomic organizations of Ig light chain loci in tetrapods (not to scale). The red, blue, and black vertical rods represent IGLV, IGLJ, and IGLC genes, respectively. Long rods show functional genes, and short rods indicate pseudogenes. Rods above and below the lines indicate genes located on opposite strands. Here the gene symbols used are as follows: RPIA, ribose-5-phosphate isomerase; EIF2AK3, eukaryotic translation initiation factor 2-$\alpha$ kinase 3; LSP1, lymphocyte-specific protein 1; DRDSP1, dopamine receptor D5 pseudogene 1; GNAZ, guanine nucleotide binding protein; RTDRI, rhabdoid tumor deletion region gene 1; TOP3B, topoisomerase III $\beta$; and PPM1F, protein phosphatase 1F. Ch and Sc correspond to chromosome and scaffold, respectively.

Fig. 5. Molecular markers on the surface on the structure of 1GC1 ($\kappa$) light chain. The conserved amino acids shown in Figs. 2 and 3 are mapped on the 3D structure of the $\kappa$ IGL chain. The numbers correspond to the positions of the amino acids in the multiple sequence alignment of Figs. 2 and 3. Of the 13 aa markers of IGL sequences, 8 (positions 7 of IGLV, 10 of IGLJ, and 14, 32, 60, 79, 91, and 102 of IGLC) are located on the surface, and 5 markers (positions 64–67 of IGLV, 11 of IGLJ, and 17, 34, and 65 of IGLC) are partially buried. The positions in the IGLV region are shown in blue, the positions in the J region in red, and the positions in the IGLC region in green. (Left) Ribbons. (Right) Surfaces. The images in Lower are turned 180° over the horizontal axis compared those in with Upper. The structural information was downloaded from the Protein Data bank (www.rcsb.org), and the figures were created using Pymol (Delano Scientific).
Methods

Identification of IGL Genes. An exhaustive gene search was conducted to identify all of the IGLV, IGLJ, and IGLC genes in the draft genome sequences of 12 species (from the Ensembl genome browser) listed in Table 1. The information about genome assembly is given in Table S1. For all species, we performed two rounds of tBLASTn search (27) for IGLV and IGLC sequences separately with the cutoff E-value of 10^-15 against the genome sequences. In the first round, the amino acid sequences of 10 functional IGLV (4 known IGVK and 4 known IGVL sequences from human, and 2 known IGVS sequences from frog) and 4 functional IGLC sequences (1 IGCK and 2 IGLC from human, and 1 IGCS from frog) were used as queries. The accession numbers of IGLV queries are L33854, M39267, A0864080, CAA26318, XS5936, X57811, M99606, D86993, 578544, and NP.001087883. The accession numbers of IGLC queries are BC110394, BC007782, D87023, and NP.001087883. Because these queries are similar to one another, they hit the same genomic regions. We extracted only nonoverlapping sequences given by the best hit (with the lowest E-value). Any retrieved IGLV sequences that aligned with the query sequence without any frameshift mutations and/or premature stop codons in the leader sequence and the V-exon, possessed the two conserved Cys residues in FR1 and FR3 regions, respectively, and had a proper RSS, was regarded as a potentially functional IGLV gene. Other sequences (including truncated ones) were regarded as IGLC pseudogenes. The first-round Blast best-hit sequences of a specific organism were used as queries in the second round of tBLASTn search to find additional IGLV sequences. In a similar way, nonredundant sequences were retrieved. Any retrieved IGLC sequence that contained frameshift mutation and/or internal stop codons and/or did not have Cys residues at conserved positions was regarded as a pseudogene. For the detection of IGLJ genes, we screened the 7-kb region upstream sequences of the IGLC gene, considering the location of the RSS sequence upstream of the IGLJ gene. Because the IGLJ gene was very short, it could not be detected by Blast search. The list of IGL genes is given in Table S2.

Sequence Alignment and Phylogenetic Analysis. The amino acid sequences translated from functional IGLC genes and the FR regions translated from functional IGV genes were aligned separately using the CLUSTALW program (28). We used the default parameters of CLUSTALW and inspected the alignments manually to maximize similarity. For IGLV genes, only the FRs sequences were considered in the analysis, whereas the CDRs were excluded because they were highly variable and contained many insertions/deletions. The evolutionary distances were computed by using the Poisson correction (29), p-distance (29), and JTTmatrix-based (30) methods. The phylogenetic trees were constructed by the neighbor joining (31), maximum parsimony (32), and minimum evolution (33) methods based on the pairwise deletion option using the MEGA4.0 program (34). The IGLV and IGLC genes were rooted by using two IGHV (accession nos. S24660 and S24664) and two IGHC (accession nos. X07781 and X07783) sequences of elasmobranch species Heterodontus francisci, respectively. The reliability of the tree was assessed by bootstrap resampling with a minimum of 1,000 replications.

ACKNOWLEDGMENTS. We thank Masafumi Nozawa, Dimitra Chalkia, Zhen- guo Lin, and Hiroki Goto for their valuable comments and suggestions. This work was supported by National Institutes of Health grant GM020293–35 (to M.N.).

Fig. S1. Phylogenetic tree condensed at the 50% bootstrap value level for functional IGLV sequences from human, chicken, lizard, and frog. The trees are constructed by (a) minimum evolution and (b) maximum parsimony methods. The known human IGL $\kappa$ and $\lambda$ sequences and the frog (X. laevis) $\sigma$ sequence are indicated by IGVK, IGVL, and IGVS, respectively.
Fig. S2. Phylogenetic tree condensed at the 50% bootstrap value level for functional IGLC sequences from 12 tetrapod species. The trees are constructed by (a) neighbor-joining, (b) minimum evolution, and (c) maximum parsimony methods. The open circles, filled circles, and open squares represent $\kappa$, $\lambda$, and $\nu$ constant domains, respectively.
Fig. S3. The conserved differences presented in Figs. 2 and 3 are mapped on the 3D structure (Left Panels: ribbons; Right Panels: surface) of the 1GC1 (κ) light chain molecule. The positions of the amino acids known to (a) interact with protein L, (b) interact with the C domain of the heavy chains, and (c) comprise the Vernier zone are shown in blue, the markers in red, and the positions that overlap in green. The interaction of κ chains with the bacterial surface protein L is mediated by the FWR regions of the V domains and does not interfere with antigen binding (see ref. 20). Two of the residues that mediate this binding correspond to the conserved markers Ser-7 at the IGLV and Lys-12 at the IGLJ. In addition, the insertion in the α sequences at position 46a,b at the IGLV may interfere with the binding of protein L. It has been suggested that the interface between the constant domain of the light and the heavy chains (IGLC–IGHC interface) may play important roles in the perfect antigen–antibody complementarity. It may function to support the variable domains and to conserve the favorable interface between antigen and antibody (see ref. 21). Two of the residues at the IGLC (Ser-14 and Gin-17) that make contact with the IGHC and the gap at position 56 at the IGLC have been identified in this study as conserved markers that discriminate IGL isotypes. We speculate that these residues might have unique functions in the different light chains. The marker at position 53 of the V domain belongs to the amino acid residues constituting the Vernier zone (see ref. 6). The residues in the Vernier zone have a structural role in supporting the CDR loop conformations and/or their relative dispositions. The position 53 in α chains is substituted to Tyr (see Fig. 2).

Other Supporting Information Files

Table S1
Table S2
Dataset S1
Dataset S2
Dataset S3
Dataset S4