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 isotypes) 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.

cladistic molecular markers | immunoglobulin κ , λ , and σ chain genes | genomic organization | recombination signal sequence

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 unresolvable 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 denominations 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) 1,329 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



Fig. 1. Neighbor-joining tree of functional IGLV sequences from human, chicken, lizard, and frog. The tree is condensed at the 50% bootstrap value level. Filled circles indicate that the interior branches are supported by \geq 90% bootstrap value. The tree was constructed by using the pair-wise deletion option and the Poisson correction distance. The known human IGL κ and λ sequences and the frog (*Xenopus laevis*) σ sequence are indicated by IGVK, IGVL, and IGVS, respectively.

Cladistic Markers of IGLV Sequences. Because of these difficulties, we resorted to an alternative approach to distinguish the IGL isotypes in tetrapods based on the use of cladistic molecular markers characterizing IGL (IGLV, IGLJ, and IGLC) protein sequences. For the IGLV, we aligned the framework regions (FR1-FR3) of functional sequences identified from the 12 tetrapod species. We used only FRs because CDR sequences are highly variable and difficult to align. Our analysis revealed the presence of several conserved amino acid residues or alignment gaps that were useful for distinguishing IGLV sequences and assigning them into three groups (groups 1-3 in Fig. 2). The distribution and the percentage of occurrence of each of these markers in the functional IGL sequences are given in Table 2 and Dataset S1, Dataset S2, and Dataset S3. Groups 1 and 2 include human sequences of proteins originally identified as bearing the κ and λ antigenic determinants, respectively. Group 3 includes sequences previously assigned to a separate σ isotype found only in frogs (11). Group 3 sequences differ from the group 1 and 2 sequences by the presence of three additional amino acid residues in FR3 (Fig. 2). Group 1 and 2 sequences could be discriminated by the length of the FR1 region, which is 23 aa long in group 1 sequences and contains Ser or Thr at position 7, whereas the length is 22 aa in group 2 sequences. In addition, the two groups could be distinguished by the amino acid position 53, which contained a bulky aromatic side chain (Phe or Tyr) in group 1 and a residue with a smaller side chain (mainly Ala or Gly) in group 2 sequences. The former characteristic was previously reported for V κ and the latter for V λ sequences using a small number of sequences (5). Furthermore, there is a fairly conserved (≈91% conservation) 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 IGLJ 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, 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 $C\kappa$ domains in all of the species tested except frog have amino acid residues at positions 94 and 95, whereas the C λ domains lack residues at these positions (Fig. 3). In the frog, both $C\kappa$ and $C\lambda$ domains lack the residues at these two positions. In the second case, at position 79 C κ and C λ domains have Tyr and Trp, respectively, in all of the species except frog, in which $C\kappa$ has Trp at this position. The two alternative explanations of these exceptions 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 RSSs. Comparison of RSSs revealed that group 1 (κ) and group 3 (σ) *IGLV* genes have an RSS with a 12-bp spacer

Table 1. Numbers of Ig light chain genes in 12 tetrapod species

	К							λ						σ											
	IGVK		10	IGVK		IGVK		JΚ	IG	СК			IGVL	10	ĴΓ	IG	CL		IG	VS	10	JS	IG	CS	
Species	F	Р	F	Р	F	Р	Loc	F	Р	F	Р	F	Р	Loc	F	Р	F	Р	F	Р	Loc				
Human	34	38	5	0	1	0	Ch 2	32	42	4 [1]	2	4	3	Ch 22	0	0	0	0	0	0	_				
Mouse	80	78	4	1	1	0	Ch 6	3	0	4	0	4	0	Ch 16	0	0	0	0	0	0	_				
Rat	76	98	5	1	1	0	Ch 4	8	2	2	1	2 [1]	0	Ch 11	0	0	0	0	0	0	—				
Dog	16	9	5	0	1	0	Ch17	51	65	9	0	9	0	Ch 26	0	0	0	0	0	0	_				
Microbat	0	0	0	0	0	0	_	93	39 (7)	7	1 (5)	10 [1]	0 (2)	_	0	0	0	0	0	0	_				
Cow	9	13	4	1	1	0	Ch11	28	29 (1)	2 [1]	2	2 [1]	2	Ch 17	0	0	0	0	0	0	—				
Horse	22	25	4	1	1	0	Ch15	25	20	4 [1]	2	4 [1]	2	Ch 8	0	0	0	0	0	0	_				
Opossum	76	48	4	1	1	0	Ch 1	51	10	8	0	8	0	Ch 3	0	0	0	0	0	0	_				
Platypus	9	9 (1)	6	1	1	0	_	14	7 (2)	3	0 (1)	3	0 (1)	_	0	0	0	0	0	0	—				
Chicken	0	0	0	0	0	0	_	1	24	1	0	1	0	Ch 15	0	0	0	0	0	0	_				
Lizard	13	3	4	0	1	0	_	29	9	2	1	2 [1]	0	_	0	0	0	0	0	0	_				
Frog	45	10	5	4	1	0	—	8	4	5	0	3	0	—	9	4	3	1	1	0	—				

F, functional gene; P, pseudogene; Loc, location. The number in the parentheses is that of partial genes. In the λ locus, the *IGJL-IGCL* block can be nonfunctional if one of them is a pseudogene. The number in brackets indicates the *IGJL* or *IGCL* genes that are nonfunctional because the partner gene is defective.

at their 3' end, whereas group 2 (λ) *IGLV* genes have an RSS with a 23-bp spacer. The κ and σ *IGLJ* genes have an RSS with a 23-bp spacer, whereas the $J\lambda$ genes are flanked by an RSS with a 12-bp spacer at the 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 IGJK 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 *IGJL* 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 RPIA and EIF2AK3, whereas the 5' end is flanked by LSP1 and DRD5P1 (see Fig. 4 and ref. 10 for these abbreviated genes). The 3' end of the κ -encoding locus is flanked by the RPIA and EIF2AK3 genes in other tetrapods as well, except in frogs, in which only the RPIA 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 RTDR1 genes, and the 5' end contains the TOP3B 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 locus are not conserved. In the lizard the chromosomal position of GNAZ and RTDR1 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 *IGJK* 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 cladistic molecular markers 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 character states. Underlying this redefinition is the congruency 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 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/or λ . An extension of the cladistic analysis to the fishes would provide a test of these predictions.

	4		FR1	CDR1	FR2	CDR2/0 I	IGLV	FR3	60 7	RSS	RSS	IGLJ	12
	1		10	20 V	30	✓ 40 a	a b	30	80	0	1		12
Human	ΕI	VMTQS	PATLSLSPGE	RATLSCWYQQI	KPGQAPRLL	IYGIP-	ARFSGSG	GSGTDFTLTI	SSLQPEDFAVYYC	12bp	23bp W T F	GQGTKV	EIK
Mouse	ΕI	VLTQS	PALMAASPGE	K V T I T C W Y Q Q I	KSETSPKPW	IYGVP -	VRFSGSO	GSGTSYSLTI	S S ME A E D A A T Y Y C	12bp	23bp W T F	GGGTKL	EIK
Rat	DI	VMTQS	PTSMSISVGD	O R V T M N C W Y Q Q I	KTGQSPKLL	IYGVP-	D R F T G S G	GSGTDFTFTI	S N M Q A E D L A V Y Y C	12bp	23bp Y T F	GAGTKL	ELK
Dog	ΕI	VMTQS	PASLSLSQEE	K V T I T C W Y Q Q I	KPGQAPKLL	IYGVP -	S R F S G S G	GSGTDFSFTI	SSLEPEDVAVYYC	12bp	23bp W T F	GAGTKV	EIK
Cow	DI	VLTQT	PLSLSVFPGE	TASISCWHLQI	KPSQSPQLL	IYGVP -	D R F T G S G	GSGTDFTFKI	S R V E A E D A G V Y Y C	12bp	23bp F T F	GPRTRV	EIK Gp1
Horse	DI	VMTQS	PDSLAVSLGQ	RVEMKCWYQQI	KPVKAPKLL	IYGVP-	D R F S G S G	GSGTDFTLTI	SSLQAEDVAVYYC	12bp	23bp W T F	GAGTKL	ЕΙΚ (к)
Opossum	ΕI	VLTHS	PASVSVSLGE	RVTIKCWFQQI	KQDRPIKAL	IYGIP-	P R F S G S G	GSGTDFTLTI	S S L E A E D A A I Y Y C	12bp	23bp W T F	GGGTKV	EIK
Platypus	QI	VLTQT	PESLAKAPGE	S A T I K C W Y Q Q I	KPGQAPKLI	IKGIP-	ARFSG SG	GSGTDFTLSI	N S V E A G D A A D Y Y C	12bp	23bp W T F	GKGTKL	EIK
Lizard	DI	VVTQT	° P S S L Q K N P G D	RVDVQCFYQLI	LPGQKPKLL	IYGMP-	DRFSGRO	GSGTDFTFTI	NGVRPEDEGEYYC	12bp	23bp F T F	GVGTKL	EIK
Frog (Xt)	QI	VLTQS	PDYVSVSPGE	TATLTCWYQQI	KSGQAPKLL	IYGTP-	ERISG SO	GSGTDFTLTI	S R M E A E D A A D Y Y C	12bp	23bp R T F	GGGTRV	EIK
Human	S Y	ELTQ-	PPSVSVSPGQ	TASITCWYQQI	KPGQSPVLV	IYGIP-	ERFSG SN	N S G N T A T L T I	SGTQAMDEADYYC	23bp	12bp Y V F	GTGTKV	TVL
Mouse	Q A	VVTQ-	ESALTTSPGE	TVTLTCWVQEI	K P D H L F T G L	IGGVP-	ARFSG SL	LIGDKAALTI	T G A Q T E D E A I Y F C	23bp	12bp W V F	GGGTKL	TVL
Rat	S Y	ELIQ-	PPSASVTLGN	T V S I T C WY Q Q I	K P D K S I V R V	IYGIS-	DRFSGSS	SSGTTATLTI	RDTQAEDEADYYC	23bp	12bp W V F	GKGTRL	TVL
Dog	QS	VLTQ-	PASVSGSLGQ	RVTISCWYQQI	LPGTGPRTV	IYGVP -	DRFSG SK	SGSTATLTI	SGLQAEDEADYYC	23bp	12bp A V F	GGGTHL	TVL
Microbat	Q S	VLTQ-	PSSVSGSLGQ	RVTISCWYQQI	LPGKIPKLL	IYGIP-	DRFSG SR	RSGSSASLTI	TGLQAEDEADYYC	23bp	12bp W V F	GGGTQL	TVL
Cow	Q A	VLTQ-	PPSMSGSLGQ	RVSITCWYOMI	I P G S A P R T L	IYGVP-	DRFSVSR	RSGNTATLTI	SSHQAEDEADYYC	23bp	12bp A V F	GSGTTL	TVL Gp2
Horse	S S	ELTO-	PPSVSVTLIC	T V S T I C WYHOI	KPGOVSLLN	IYGIS-	DRFTG SK	SENTATLTI	GGTOTEDEADYYC	23bp	12bp Y I F	GSGTHL	TVL (λ)
Opossum	QD	оvvтò-	PPSVSAALGS	SARLSCWHQQS	SPGKPPRYL	LYGIP-	ERFSGSO	GSGTNRFLSI	SNVQPEDEADYFC	23bp	12bp H I F	GKGTKL	ТVТ
Platypus	S S	VPTO-	PPSASVALGO	TASITC WYOOI	KPGRAPVLV	IYGIP-	DRFSSSK	SGSAATLTI	S R A Q A E D E A D Y Y C	23bp	12bp Y I F	GSGTHL	TVL
Chicken	0.4	ALTO-	PSSVSANPGE	тукі те муооі	K P G S A P V T L	LYNLP -	SRESGSK	SGSTATLTL	TGVOADDEAVYYC	23bp	12bp G I F	GAGTTL	TVL
Lizard	0 F	VLTO-	PPSLAADTGT	TVOLSCWYOOI	KPGNAPKLV	LYGLP-	DRETGSK	SANSATLSI	SNLOAEDEADYYC	23bp	12bp F V F	GGGTHN	TYL
Frog (Xt)	0.1	VLTO-	PASVSVSVGG	TVKLTCWYOOI	MLPSAPRLI	LYGLP-	ERESG KN	N S G N T A S L E L	S G A O A G D E A D Y Y C	23bn	12bp WI F	GGGTOL	TVL
Frog (Xt)		V - SLT	PSNNAVNLGE	RATENCLLKOI	PGNAPOLI	LYGIST	DRYTATINS	SAATEYOFLL	K K A E T A D T A H Y Y C	12hn	23hn Y V F	SOSSKL	L V T Gp3
	~ ~		D C NNAVNI CE	PVTINCIOKOI	PGNVPOLL	LAGIST	DRVTATVNA	AATEVOELL	KKAETAHTAHVVC	P	V V F	SOSSKI	

Fig. 2. Alignments of IGLV and IGLJ sequences. One IGLV and one IGLJ sequence (randomly chosen) from each isotype of each species was taken from the large-scale alignment. The cladistic molecular markers that distinguish among the three isotypes (Gp 1–3) are highlighted. The lengths of the RSS spacer sequences are given for IGLV and IGLJ sequences. The numbering of amino acid positions is based on human IGVK and IGVL sequences (Ensembl annotation), and the gaps relative to the frog IGVS sequences are indicated by "a" and "b." The aligned sequences of all functional *IGLV* and *IGLJ* genes are given in Datasets S1 and S2, respectively. Xt and XI stand for frog species X. *tropicalis* and X. *laevis*, respectively. The RSS information for XI sequence (accession no. NP_001087883) is not available.

The long-term persistence of the molecular markers is contrasted by the regional instability of the IGL loci. The genomic maps of these loci reveal multiple gene rearrangements in the evolutionary process. The rearrangements have resulted in variation in the total length of the loci and the number of component

Table 2. Summary of the molec	ular markers in IGL protein
sequences	

Gene	Position	Markers	к	λ	σ
IGLV	<i>IGLV</i> 7 Gap		- (100)	+ (100)	- (100)
		S/T	+ (97)	- (97)	+ (100)
	41a	Gap	+ (100)	+ (100)	- (100)
	46 a–b	Gap	+ (100)	+ (100)	- (100)
	53	F/Y	+ (100)	- (100)	+ (100)
	64–67	DEAD	- (100)	+ (91)	- (100)
IGLJ	2	Т	+ (100)	- (100)	- (100)
	4	G	+ (100)	+ (98)	- (100)
	7	Т	+ (100)	+ (100)	- (100)
	10–12	EI(L)K TVL(A/T)	+ (98)	- (100)	- (100)
			- (100)	+ (94)	- (100)
IGLC	14	S	+ (100)	+ (100)	- (100)
	17	QE	+ (100)	- (100)	- (100)
			- (100)	+ (100)	- (100)
	20a	Gap	+ (100)	+ (100)	- (100)
	32	F	+ (100)	+ (100)	- (100)
	34	Р	+ (100)	+ (98)	- (100)
	56	Gap	- (100)	+ (100)	+ (100)
	60	D/E K/R	+ (100)	- (100)	- (100)
			- (100)	+ (100)	+ (100)
	65	Т	+ (100)	- (100)	+ (100)
	91	Н	+ (100)	+ (100)	- (100)
	102	F	+ (100)	- (100)	- (100)

+, -, 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 IGL isotypes except frog, are not listed here (see Fig. 3).

genes and their orientation. Remarkably, these local rearrangements have not affected the overall organization of the loci, which can be used as one of the cladistic markers differentiating the three isotypes. Therefore, it seems that the evolution of the IGL loci (isotypes) is driven by two opposing forces, one diversifying and the other stabilizing these genomic regions. The diversifying force is easy to comprehend. Any multigenic region containing closely related genomic segments is prone to unequal crossing-over, gene duplication, deletion, inversion, and other diversifying forces (17–19). By contrast, the nature and mechanism of the stabilizing force is much less obvious. Probably the cause is not the same for the different markers of the IGL loci and proteins. For example, the structure of RSSs is dictated by the nature of the recombination mechanism, which requires an interaction between the RSS elements flanking the IGLV and IGLJ genes: one must have the heptamer-12-bp spacer-nonamer and the other the nonamer-23-bp spacer-heptamer structure and the orientation to be functional.

The conservation of amino acid markers may be determined by constraints on the 3D structure and function of the IGL isotypes. To get a glimpse into what these constraints might be, we have modeled the 3D structure of some of the protein sequences using the known human IGL structures as reference (Fig. 5). Of the 21 conserved amino acid markers, 8 are located on the surface of the IGL molecule (Fig. 5); 5 are partially buried, and the remaining markers are buried inside the molecule. The surface amino acids markers might provide functional differentiation of the isotypes through interaction with different ligands (Fig. S3a; ref. 20). Some of these surface amino acid residues could function during IGH-IGL chain pairing because they are located at the interface between the heavy and light chains (Fig. S3b; ref. 21). The internal markers might constrain the structure of the molecule somewhat differently in each of the three isotypes (Fig. S3c; ref. 6). Experimental support for the function of most of the identified markers is missing, however.

The conservation of these markers for more than 350 million years suggests the presence of different functional constraints on the three IGL isotypes. Although κ - and λ -bearing antibodies seem to be functionally equivalent, the two isotypes of IGL

	1 10	20 a	30	40	50	60	70	80	90	100 107
Human	- T V A A P S V F I F P P S D I	EQLKS - GTAS	V V C L L N N F Y P R	REAKVQWKVDN	ALQSGNSQE	S V T E Q D S K D S T	YSLSSTLTLSKA	DYEKHKVYACE	VTHQGLSSPV	TKSFNRGEC
Mouse	- ADAAPTVSIFPPSSI	EQLTS - GGAS	VVCFLNNFYPK	K D I N V K W K I D G	SERQNGVLN	S WTDQDSKDST	YSMSSTLTLTKD	EYERHNSYTCE	ATHKTSTSPI	VKSF NRNEC
Rat	- RAAAPTVSIFPPSTI	EQLAT - GGAS	V V C L M N N F Y P R	R D I S V K W K I D G	TERRDGVLD	S V T D Q D S K D S T	YSMSSTLSLTKA	DYESHNLYTCE	VVHKTSSSPV	VKSFNRNEC
Dog	- N D A Q P A V Y L F Q P S P I	DQLHT - GSAS	VVCLLNSFYPK	K D I N V K W K V D G	VIQDTGIQE	SVTEQD-KDST	YSLSSTLTMSST	EYLSHELYSCE	ITHKSLPSTL	IKSFQRSEC
Cow	- S D A E P S V F L F K P S D I	EQLKT - GTVS	V V C L V N D F Y P K	K D I N V K W K V D G	VTQSSNFQN	SFTDQDSKKST	YSLSSILTLPSS	EYQSHDAYTCE	VSHKSLTTTL	VKSFSKNEC Gp1
Horse	- D D A K P S A F I F P P S S I	EQLSS - GSAS	V V C L V Y G F Y P S	G A T I N W K V D G	LAKTSSFHS	SLTEQDSKDNT	YSLSSTLTLPKA	DYEAHNVYACE	VSHKTLSSPL	VKSFKRQDC (K)
Opossum	I - S V A T P S A F I F Q P S P I	EQLQT - GSAS	LVCLVNNFYPK	K E A S V V W K V D N	AVKNTGIST	SVTEQDSKDNT	YSLSSTLALSAN	DYNNHETYSCE	IIHQSLPSPL	VKSFQRSQC
Platypus	- ENARPSVFIFHPSEI	EQLTQ - GSAS	I V C L V N N F Y P Q	Q N A K V Q W K L D N	KPTESGVLT	S T T A Q D S K D S T	YSLSSTLSLSKA	EYEKHNQYSCE	ITHETLTSPL	VKSFQRNEC
Lizard	- D D A A P S A S I F P P S Q I	EQLKT - NSAT	MVCLVTGFYPS	5 D L N V A WK A D G	KAVTGDVQT	SSAASE - ADKT	YKLSSTLTLSRD	EYNSHDTYDCE	VTHKTLTKPL	VKSFRRSEC
Frog (Xt)	- A D A I P V V F I F K P S E I	EQVKE - GNPT	AVCLINNFYPR	R S V T I K W K V D G	QVIASDIKN	SDYMQE-SDNT	YSQSSMLTLTKD	KWDKSEKFECL	V Q H K T Q Q L	AQSFIKSQC
Human	Q P K A N P T V T L F P P S S I	EELQA-NKAT	LVCLISDFYPC	G A V T V A WK A D G	SPVKAGVET	T - K P S K Q S N N K	YAASSYLSLTPE	QWKSHRSYSCQ	VTHEGSTV	EKTVAPTEC
Mouse	Q P K S S P S V T L F P P S S I	EELET - NKAT	LVCTITDFYPC	G V V T V D W K V D G	TPVTQGMET	T - Q P S K Q S N N K	YMASSYLTLTAR	AWERHSSYSCQ	VTHEGHTV	EKSLSRADC
Rat	Q P K S T P T L T V F P P S T I	E E L Q G • N K A T	LVCLISDFYPS	S D V E V A WK A N G	APISQGVDT	A • N P T K Q • G N K	YIASSFLRLTAE	QWRSRNSFTCQ	VTHEGNTV	EKSLSPAEC
Dog	Q P K A S P S V T L F P P S S I	EELGA•NKAT	LVCLISDFYPS	S G V T V A W K A D G	SPVTQGVET	T - K P S K Q S N N K	YAASSYLSLTPD	K W K S H S S F S C L	V T H E G S T V	EKKVAPAEC
Microbat	Q P K S A P S V S L F P P S A I	EELAN - NKAT	LVCLMSDFYPO	G S V T V A WK A N G	TPVTQGVET	T - K P S K Q S N N K	YAASSYLSVSSQ	DWKSASAYSCQ	VTHDGKTV	EKTVAPSEC
Cow	Q P K S P P S V T L F P P S T I	EELNG - NKAT	LVCLISDFYPO	G S V T V V W K A D G	STITRNVET	T - R A S K Q S N S K	YAASSYLSLTSS	DWKSKGSYSCE	V T H E G S T V	TKTVKPSECGp2
Horse	GPTSAPSVSLFPPSSI	EELSA - NKAT	VVCLISDFSPS	S D L T V S WK V N G	A A I S Q G V Q T	T - K P S K Q S N G K	YAASSYLTLTPA	Q W K S S S S S S C Q	V T H Q G K T V	EKKLSPSEC (A)
Opossum	Q P K A S P M V M A Y G P S Q I	DELRT - PTAT	LTCLISGFYPC	GSLEVTWTKNG	SPVSSGVLT	S - Q P V R Q T D N K	YSASSYLSVNPN	QWGENNQYMCK	V T H E G Q V I	EKELSASQC
Platypus	Q P K A S P T V N M F A P S S .	AELDS - KKAT	LVCLMGGFYPC	G S V T V T WK A D G	ATVSEGVET	T - K P S K Q T D N K	YMASSYLTLTPA	QWKAKNSYTCQ	VTHDGKVF	EKSVSPSEC
Chicken	Q P K V A P T I T L F P P S K I	E E L E A • T K A T	L V C L I N D F Y P S	S P V T V D W V I D G	S T R S G E T	T - A P Q R Q S N S Q	YMASSYLSLSAS	DWSSHETYTCR	V T H N G T S I	TKTLKRSEC
Lizard	Q P A V P P V V S L F P P S Q I	EELSK - NKAT	LACLMDGFHPC	GVVQVEWLGDG	RTISSGVET	T - K P I K Q - G D K	YIASSYLTLSRS	EWEANESYTCK	. V T H E S K T I	EKVVSRLAC
Frog (Xt)	G D V K S P S V S I F P P S E I	ЕЕІАТ - ККАТ	LVCSLSDFTPR	RGATVKWQVDG	KEKTDGVVS	S - G L S K Q S D S L	YMESSYLSLTSD	EWLKHDLYSCK	V T H Q G K E I	IQTLKRSDC
Frog (Xt)	DKFPEPALLVFPPYTI	EDNESKDSST	LTCHISKLAVS	S L V N V N W L K D G	TTLQDGVST	S - N P V R E S D N T	FSMSSYLTLSSK	DINKDSLYSCM	II Q Q E G S S A F V	SKGVKLSQCGp3
Frog (XI)	DKFPEPALLVFPPYTI	D D N E S K D S S T	LTCHISKLEVS	S L V N V N W L I D G	TTVQDGVTT	S - N P V R E S D N T	FSMSSYLTLSSK	DINKDRLYSCM	II KQEGSSAFI	s k g V k l s q c (σ)

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 κ and λ 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 XI stand for frog species *X. tropicalis* and *X. laevis*, respectively. The aligned sequences of all functional *IGLC* genes are given in Dataset S3.

sequences have been related to different pathophysiologic conditions. The λ -bearing antibodies have been found in more than 60% of light-chain amyloidosis, whereas the κ 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 λ chains has been shown to be preferentially used (23). Yet another well-established functional difference between κ and λ antibodies is the interaction of κ chains with the



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; *DRD5P1*, dopamine receptor D5 pseudogene 1; *GNA2*, guanine nucleotide binding protein; *RTDR1*, rhabdoid tumor deletion region gene 1; *TOP3B*, topoisomerase III β ; and *PPM1F*, protein phosphatase 1F. Ch and Sc correspond to chromosome and scaffold, respectively.

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 κ and λ antibodies was indicated by Wardemann *et al.* (2004) (26). These investigators found that λ genes are more effective than κ genes as silencers of autoantibody production. Taken together, all these reports suggest that κ and λ isotypes are not functionally equivalent.



Fig. 5. Molecular markers on the surface on the structure of 1GC1 (κ) light chain. The conserved amino acids shown in Figs. 2 and 3 are mapped on the 3D structure of the κ 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 IGL), and 14, 32, 60, 79, 91, and 102 of IGLO) are located on the surface, and 5 markers (positions 64–67 of IGLV, 11 of IGL), 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 IGCL from human, and 1 IGCS from frog) were used as queries. The accession numbers of IGLV queries are L33854, M38267, AB064080, CAA26318, X53936, X57811, M99606, D86993, S78544, 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 IGLV 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 con-

1. Klein J, Hoøejší V (1997) Immunology (Blackwell Science, Oxford).

- Akira S, Okazaki K, Sakano H (1987) Two pairs of recombination signals are sufficient to cause immunoglobulin V-(D)-J joining. Science 238:1134–1138.
- Korngold L, Lipari R (1956) Multiple-myeloma proteins. III. The antigenic relationship of Bence Jones proteins to normal gammaglobulin and multiple-myeloma serum proteins. Cancer 9:262–272.
- Hill RL, Delaney R, Fellows RE, Lebovitz HE (1966) The evolutionary origins of the immunoglobulins. Proc Natl Acad Sci USA 56:1762–1769.
- Chothia C, et al. (1989) Conformations of immunoglobulin hypervariable regions. Nature 342:877–883.
- Kabat EA, Wu TT, Perry HM, Gottesman KS, Foeller C (1991) Sequences of Proteins of Immunological Interest (US Department of Health and Human Services, Washington, DC), NIH publication no. 91–3242.
- Ota T, Sitnikova T, Nei M (2000) Evolution of vertebrate immunoglobulin variable gene segments. Curr Top Microbiol Immunol 248:221–245.
- Pilstrom L (2002) The mysterious immunoglobulin light chain. Dev Comp Immunol 26:207–215.
- Criscitiello MF, Flajnik MF (2007) Four primordial immunoglobulin light chain isotypes, including lambda and kappa, identified in the most primitive living jawed vertebrates. *Eur J Immunol* 37:2683–2694.
- Qin T, et al. (2008) Genomic organization of the immunoglobulin light chain gene loci in Xenopus tropicalis: evolutionary implications. Dev Comp Immunol 32:156–165.
- Klein SL, et al. (2002) Genetic and genomic tools for Xenopus research: The NIH Xenopus initiative. Dev Dyn 225:384–391.
- Zezza DJ, Stewart SE, Steiner LA (1992) Genes encoding Xenopus laevis Ig L chains. Implications for the evolution of kappa and lambda chains. J Immunol 149:3968–3977.
- Haire RN, et al. (1996) A third Ig light chain gene isotype in *Xenopus laevis* consists of six distinct VL families and is related to mammalian lambda genes. *J Immunol* 157:1544–1550.
 Saluk PH, Krauss J, Clem LW (1970) The presence of two antigenetically distinct light
- Saluk FF, Madus J, Clam LW (1970) The presence of two antigenetically distinct right chains (k and λ?) in alligator immunoglobulin. *Proc Soc Exp Biol Med* 133:365–369.
 Solomon A (1976) Bence-Jones proteins and light chains of immunoglobulins (second
- of two parts). N Engl J Med 294:91–98.
- Schwager J, Burckert N, Schwager M, Wilson M (1991) Evolution of immunoglobulin light chain genes: Analysis of *Xenopus* IgL isotypes and their contribution to antibody diversity. *EMBO J* 10:505–511.
- 17. Ohno S (1970) Balanced polymorphism and evolution by the birth-and-death process in the MHC loci. *Evolution by Gene Duplication* (Springer, Berlin).
- Nei M, Hughes AL (1992) in 11th Histocompatibility Workshop and Conference, eds. Tsuji K, Aizawa M, Sasazuki T (Oxford Univ Press, Oxford), pp 27–38.

served 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 IGLV 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 gene trees 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, Zhenguo 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.).

- 19. Das S, Nozawa M, Klein J, Nei M (2008) Evolutionary dynamics of the immunoglobulin heavy chain variable region genes in vertebrates. *Immunogenetics* 60:47–55.
- Housden NG, et al. (2004) Observation and characterization of the interaction between a single immunoglobulin binding domain of protein L and two equivalents of human kappa light chains. J Biol Chem 279:9370–9378.
- Kabat EA, Padlan EA, Davies DR (1975) Evolutionary and structural influences on light chain constant (CL) region of human and mouse immunoglobulins. Proc Natl Acad Sci USA 72:2785–2788.
- 22. James LC, et al. (2007) Beta-edge interactions in a pentadecameric human antibody V kappa domain. J Mol Biol 367:603–608.
- Forconi F, et al. (2008) Selective influences in the expressed immunoglobulin heavy and light chain gene repertoire in hairy cell leukemia. Haematologica 93:697–705.
- Nilson BH, Solomon A, Bjorck L, Akerstrom B (1992) Protein L from Peptostreptococcus magnus binds to the kappa light chain variable domain. J Biol Chem 267:2234–2239.
- Montano RF, Morrison SL (2002) Influence of the isotype of the light chain on the properties of IgG. J Immunol 168:224–231.
- Wardemann H, Hammersen J, Nussenzweig MC (2004) Human autoantibody silencing by immunoglobulin light chains. J Exp Med 200:191–199.
- Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389–3402.
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positionspecific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.
- Nei M, Kumar S (2000) Molecular Evolution and Phylogenetics (Oxford Univ Press, Oxford).
- Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8:275–282.
- Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425.
- Eck RV, Dayhoff MO (1966) Atlas of Protein Sequence and Structure (National Biomed Res Foundation, Silver Springs, Maryland).
- Rzhetsky A, Nei M (1992) Statistical properties of the ordinary least-squares, generalized least-squares, and minimum-evolution methods of phylogenetic inference. J Mol Evol 35:367–375.
- 34. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599.

Supporting Information

Das et al. 10.1073/pnas.0808800105

<



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 κ and λ sequences and the frog (X. laevis) σ 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 κ, λ, and σ 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 σ 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 Gln-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 Vernier zone have a structural role in supporting the CDR loop conformations and/or their relative dispositions. The position 53 in σ chains is substituted to Tyr (see Fig. 2).

Other Supporting Information Files

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