# Multiple-alphabet amino acid sequence comparisons of the immunoglobulin κ-chain constant domain

### (codon preferences/sequence conservation/β-sheet interactions)

SAMUEL KARLIN AND GHASSAN GHANDOUR

Department of Mathematics, Stanford University, Stanford, CA 94305

*Contributed by Samuel Karlin, August 22, 1985*

**ABSTRACT** We compare the amino acid sequences of the constant domains of the immunoglobulin κ chain of human, mouse, and rabbit by using four classification schemes ("alphabets") of the 20 amino acids based on their chemical, functional, charge, and structural properties. The comparison reveals three regions of pronounced similarity across the three species, independent of allotype. Two of these regions (residues 65–73 and 99–103) entail a high degree of identity at the DNA level and are distinguished from the rest of the constant domain in codon usage and in the dinucleotide sequence at abutting sites of adjacent codons. Residues 22–29 are highly conserved among the three species in the chemical and functional alphabets but do not show any three-sequence significant amino acid block identities. These results are discussed in terms of transcript processing, effector functions, and structural interactions within the constant domain and with the heavy chain.

This paper focuses on amino acid sequence comparisons of the human, mouse, and rabbit immunoglobulin κ-chain constant ($C_\kappa$) domain, using "natural" classification schemes ("alphabets") of amino acids based on chemical, functional, charge, and structural characteristics [see refs. 1 and 2 for DNA sequence analysis of the $J_\kappa$ (joining)-$C_\kappa$ region]. In each alphabet, we determine all identity blocks exceeding a prescribed length; "identity block" refers to a set of consecutive matches; the length is the number of matches.

In addition to the standard 20-letter array of amino acids, the alphabets employed are as follows. The *chemical alphabet* (following ref. 3, pp. 10–13) has 8 letters: Acidic (Asp, Glu); Aliphatic (Ala, Gly, Ile, Leu, Val); Amide (Asn, Gln); Aromatic (Phe, Trp, Tyr); Basic (Arg, His, Lys); Hydroxyl (Ser, Thr); Imino (Pro); Sulfur (Cys, Met). The *functional alphabet* (following ref. 4, pp. 67–71) has 4 letters: Acidic and Basic (same as in chemical alphabet); Hydrophobic nonpolar (Ala, Ile, Leu, Met, Phe, Pro, Trp, Val); Polar uncharged (Asn, Cys, Gln, Gly, Ser, Thr, Tyr). The *charge alphabet* has 3 letters: Acidic and Basic (as above); Neutral (all the other amino acids). The *structural alphabet* (after ref. 5, p. 11) has 3 letters: Ambivalent (Ala, Cys, Gly, Pro, Ser, Thr, Trp, Tyr); External (Arg, Asn, Asp, Gln, Glu, His, Lys); Internal (Ile, Leu, Met, Phe, Val). Groupings of amino acids into various alphabets different from those that we employ have been based on frequencies of evolutionary replacements among amino acids, chemical categorizations, and minimal base differences between codons (6–12). Other possible criteria in grouping amino acids could be based on codon degeneracy, physical properties (e.g., molecular weight, shape), kinetic properties (e.g., reaction velocity, Michaelis–Menten constant), or structure (e.g., α-helices, β-sheets, turns).

The use of different alphabets in comparative sequence analyses can provide insights on several levels. It may highlight "long" block identities common to all alphabets (see *Analysis*, section 1, for conditions characterizing statistically significant block identities). Significant block identities that appear in some but not in other alphabets may suggest contrasting functional or structural properties for different regions of the sequence and help elucidate features that simple amino acid comparisons do not.

## ANALYSIS

### Section 1. $C_\kappa$-Domain Block Identities

This section focuses on the $C_\kappa$-domain comparisons with special attention to the known three-dimensional structure. Some of the highly conserved amino acid block identities encompass stretches of strong similarity at the DNA level, whereas others emphasize chemical, charge, or structural amino acid characteristics.

The $C_\kappa$ domain embodies the following tertiary structure: both termini involve short α-helices but project in the central portion two β-pleated sheets, referred to as the X-face and Y-face, consisting of four and three segments, respectively (see Table 1 and ref. 13, p. 186). The $C_\kappa$-domain residues are numbered 1–106 in human and mouse (add 108 to obtain the coordinates in ref. 13). An assortment of weak electrostatic and nonpolar bonds operate between elements of the same face. The X-face and Y-face are bridged by a disulfide bond between the cysteines at positions 26 (in $X_2$) and 86 (in $Y_2$).

The human Km allotypes are polymorphic only in residues 45 and 83 (ref. 14, Chap. 11). No allelic variation has been discerned in mouse to date. There exists a high degree of $C_\kappa$ polymorphism in rabbit, with allotypes b4, b5, b9, and bas-N4 common in domesticated populations. These allotypes were distinguished originally by serologic methods and recently by detection of multiple amino acid substitutions (15, 16). The rabbit κ complex embodies two isotypes. The b4, b5, b9, and bas allotypes are part of the $K_1$ isotype, which is predominant over $K_2$. The latter is primarily expressed in wild rabbit populations. No $K_2$ allotypic variation has been detected.

The results obtained in our between-species comparisons (the identification of the significant block identities in the various alphabets) are robust to which $C_\kappa$ polymorphic sequence is used. This is demonstrated later in this section. The rabbit amino acid sequence used in the three-species comparison of Fig. 1 is that of allotype b4, which is the one predominantly expressed (16).

A block identity within and between the human, mouse, and rabbit $C_\kappa$ domains for a given alphabet is considered statistically significant if its length exceeds by two standard deviations the expected length of the longest block identity

Abbreviations: C, constant; J, joining; bp, base pair(s).

Table 1.  Tertiary structure of the human $C_\kappa$ domain

| X-face (residues) | Y-face (residues) | Turns (residues) |
|---|---|---|
| $X_1$  (8–12) | $Y_1$ (39–43) | $X_1$–$X_2$ (13–23) |
| $X_2$ (24–32) | $Y_2$ (85–91) | $X_2$–$Y_1$ (33–38) |
| $X_3$ (52–61) | $Y_3$ (94–100) | $Y_1$–$X_3$ (44–51) |
| $X_4$ (65–74) | | $X_3$–$X_4$ (62–64) |
| | | $X_4$–$Y_2$ (75–84) |
| | | $Y_2$–$Y_3$ (92–93) |

These coordinates are also indicated in Fig. 1.

for a corresponding "random" model that maintains the same alphabet-letter frequencies as in the observed sequences (see ref. 1 for elaborations and justifications). For the $C_\kappa$ domains, the lengths of the block identity required for statistical significance ($P \le 0.01$) are listed in Table 2. For example, in the chemical alphabet a block identity of length $\ge 8$ residues occurring in any two of the three sequences has probability less than 0.01 of occurring by chance.

On the DNA level (see ref. 1 for a detailed analysis), in all three species there occurs a statistically significant common sequence of 10 bp located at the same coordinates, 296-5′ in $C_\kappa$ [i.e., its start coincides with the 296th base pair (bp) of the $C_\kappa$ gene domain] in human and mouse, but at position 290-5′ in $C_\kappa$ in rabbit (overlapping amino acid residues 99–102). This block is embedded in a 15-bp match between the human and mouse DNA sequences. Between human and mouse there is a significant shared 15-bp sequence that is in exact alignment with two additional human–mouse block identities of 11 bp. These combined identities show only five mismatches over a stretch of 48 bp (coordinates 172–219-5′ in $C_\kappa$). There is one shared 9-bp sequence, near the sequence encoding the $NH_2$ terminus, at coordinates 23-5′ in $C_\kappa$ of human and mouse, and 26-5′ in $C_\kappa$ of rabbit.

The $C_\kappa$ domain of rabbit (b4 allotype) has two amino acid residues fewer than that of human or mouse. An amino acid alignment that yields the maximum number of matches and preserves the correspondence of all two- and three-sequence significant block identities entails one insertion and three deletions in the rabbit sequence, as follows. The third and fourth residues in rabbit (Pro-Val) replace the third residue (Ala) in both human and mouse. The three deletions in rabbit are placed at residues 34, 90, and 96 in human and mouse. This matching is utilized in Fig. 1, which displays all statistically significant amino acid block identities with respect to the five amino acid classifications.

In surveying the significant block identities over the $C_\kappa$ domain, four regions stand out: (*i*) residues 9–13 (covering mainly $X_1$), (*ii*) residues 22–33 (containing the segment $X_2$), (*iii*) residues 65–73 (essentially $X_4$), and (*iv*) the COOH-terminal residues 99–103. The most conserved region in all alphabets corresponds to residues 68–73 (contained in $X_4$) which subsumes the region of strongest DNA identity.

All the three-sequence significant block identities for the various alphabets are confined to the X-face, with the exception of an amino acid identity that overlaps the last residue of $Y_3$ (see Fig. 1). This amino acid identity may reflect exact structural requirements on the region connecting the light and heavy chains at the terminal cysteine residue. No parts of the $Y_1$ or $Y_2$ segments are significantly conserved relative to any of the alphabets, and that includes the neighborhood of the cysteine site (residue 86) of the $Y_2$ segment.

In all alphabets (except for the charge classification) there is a significant three-sequence block identity covering residues 67–73, which corresponds to the highly conserved DNA stretch 172–219 bp of the $C_\kappa$ domain.

Near the $NH_2$ terminus at the start of the $X_1$ segment, mouse and rabbit show a significant DNA block identity with

Table 2.  Significance levels of block identities for the $C_\kappa$ domain of human, mouse, and rabbit

| Alphabet | Length required for $P < 0.01$, no. of residues | |
|---|---|---|
| | Three-species identity blocks | Two-species identity blocks |
| DNA* | 10 | 14 |
| Amino acid | 3 | 5 |
| Chemical | 5 | 8 |
| Functional | 9 | 13 |
| Charge | 24 | 32 |
| Structural | 8 | 13 |

*Based on the comparison of immunoglobulin $\kappa$-chain genomic DNA sequences covering the $J$-$C$ region and flanks (see ref. 1 for further details), measured in bp.

modest concordance in the human sequence (see ref. 1 and Fig. 1). Residues 19–33 manifest significant three-sequence block identities in the chemical and functional alphabets (Fig. 1 *b* and *c*); this is not the case with the amino acid alphabet. Thus, the interaction of the X-face with the heavy chain on one side and the Y-face on the other side is adequately maintained through the preservation of the chemical and functional amino acid equivalents rather than the precise arrangement of amino acids in this segment. There are no significant block identities in the charge alphabet for the $C_\kappa$ domain (Fig. 1*d*).

As for pronounced nonsimilar regions, there are two. The first corresponds to residues 37–52, extending from the $X_2$-to-$Y_1$ turn and including the segment $Y_1$. The second (residues 77–93) covers the $X_4$-to-$Y_2$ turn and the segment $Y_2$. These regions protrude exterior to the tertiary structure of the combined light and heavy chains. The differences could relate to species-specific factors or may reflect a less functionally constrained region at the external parts of the protein.

**Conserved Interspecies Block Identities Are Invariant to Allotypic Sequence.** The human $C_\kappa$-domain Km allotypes are variable only in residues 45 (at the turn after segment $Y_1$) and 83 (at the turn preceding $Y_2$) (14). These positions are centered in the two principal nonconserved regions in all alphabets (Fig. 1).

The rabbit $C_\kappa$ allotypes (b4, b5, b9, bas) feature an extraordinary degree of amino acid replacement (15, 16). In fact, >40% of all nucleotide substitutions between these $C_\kappa$ sequences are nonsynonymous.

We examine briefly the nature of the similarities and differences of the rabbit $C_\kappa$ polymorphism in terms of the multiple alphabet comparisons. The longest amino acid block identity common to all four $C_\kappa$ allotypes is 8 residues long, starting at position 23 (in b4) and surrounding the cysteine residue of the $X_2$-face. The second longest four-way amino acid block identity consists of 5 amino acids from position 67 (embedded in a 10 amino acid block identity of b4, b5, and b9). This segment corresponds to the most extensive human–mouse conserved nucleotide stretch of the $C_\kappa$ domain. Codons 98–103 show no nucleotide substitutions among the $C_\kappa$ sequences except at codon 102 of b5, which replaces Asn with Ser.

The longest block identity for all rabbit $C_\kappa$ sequences in the chemical alphabet entails 11 residues from position 21, encompassing the three-species significant block identity of Fig. 1 at positions 22–29. This segment expands to a 15-residue block identity with respect to the b4, b5, and bas allotypes and to a match of 23 consecutive positions relative to b4 and b5. The rabbit $C_\kappa$ allotypes coincide in the chemical alphabet over positions 19–33 except for b9, which deviates at positions 20 and 32. The next longest four-way block
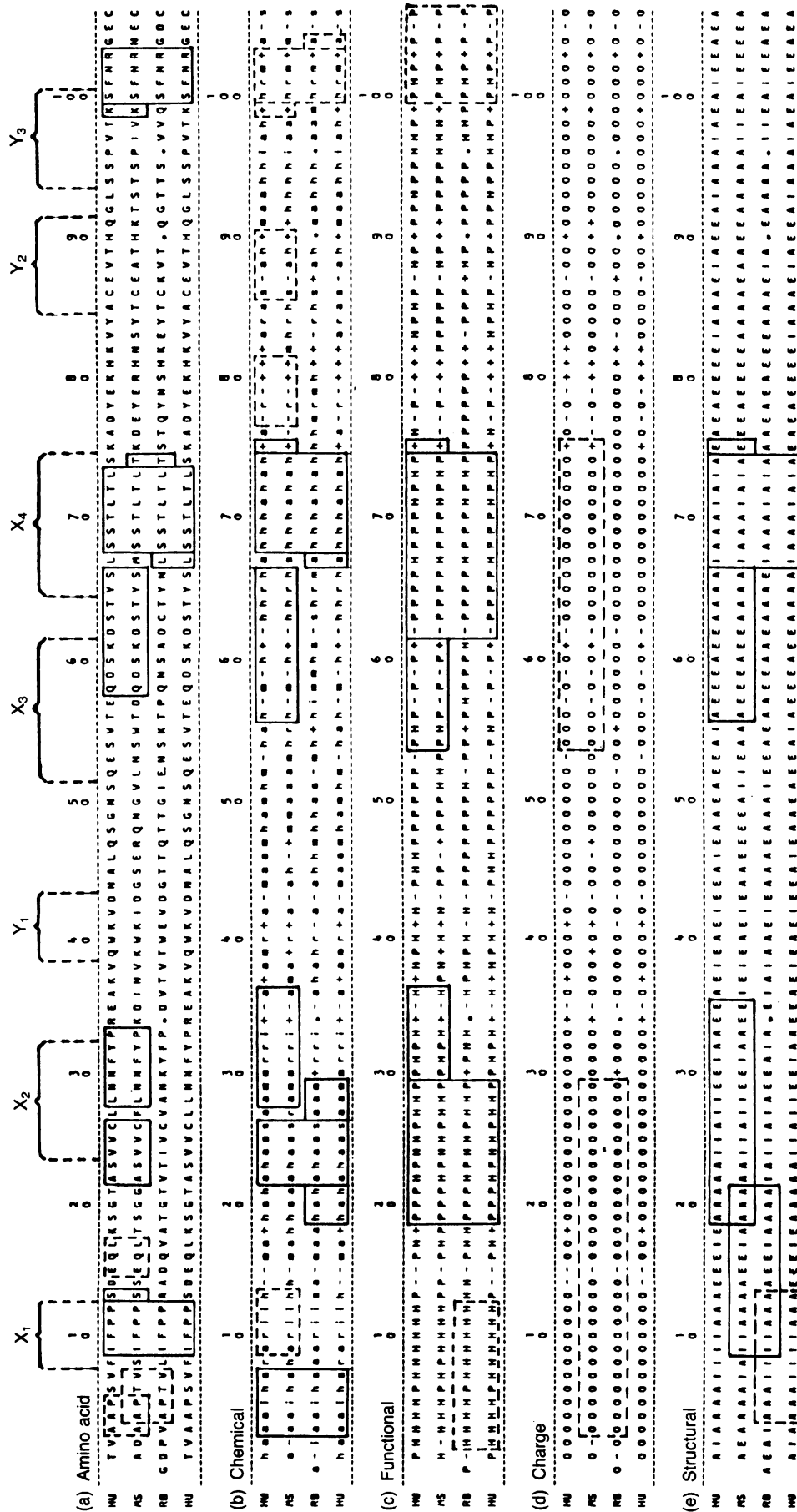
Genetics: Karlin and Ghandour

*Proc. Natl. Acad. Sci. USA* 82 (1985)     8599



FIG. 1. Statistically significant block identities involving at least two of the three [human (HU), mouse (MS), and rabbit (RB)] $C_\kappa$-domain sequences. (*a*) Amino acid alphabet: the standard one-letter abbreviations are used. (*b*) Chemical alphabet: Acidic (−), Aliphatic (a), Amide (m), Aromatic (r), Basic (+), Hydroxyl (h), Imino (i), Sulfur (s). (*c*) Functional alphabet: Acidic (−), Basic (+), Hydrophobic nonpolar (H), Polar uncharged (P). (*d*) Charge alphabet: Acidic (−), Basic (+), Neutral (0). (*e*) Structural alphabet: Ambivalent (A), External (E), Internal (I). The rabbit $C_\kappa$-domain is aligned as described in the text; the deletions at residues 34, 90, and 96 are each indicated by a dot. Solid boxes contain statistically significant block identities common to the enclosed sequences. Block identities longer than the theoretically expected length for the random model but not statistically significant that are in alignment with statistically significant blocks are indicated by broken-line boxes.

identities of the chemical alphabet are the 10-residue segment, 3–12, and the 7-residue segment, 67–73. Even the amino acid composition is invariant except at position 72, where b4 and b5 entail Thr, whereas b9 and bas use Ser. The bas and human sequences coincide in the chemical alphabet for the block of positions 60–74. It is telling that the three interspecies significant block identities in the chemical alphabet (Fig. 1b) correspond directly to the three longest rabbit $C_\kappa$ allotype matching blocks in this alphabet. Completely parallel results emerge in the functional and structural alphabet. Thus, the strongly conserved block identities of the human, mouse, and rabbit sequences persist in the various alphabets, independent of the rabbit allotype. The maintenance of the significant block identities of Fig. 1 independent of their allotypic representation strengthens the validity of these conserved regions in functional and evolutionary significance.

## Section 2. Nucleotide Versus Amino Acid Conservation

The existence of significant amino acid block identities for similar genes across species could be due to intrinsic conservation at the DNA level or to strict adherence to amino acid content that allows for moderate numbers of synonymous codon substitutions. What could account for strong nucleotide-sequence conservation, more than required for amino acid conservation, over some coding regions?

Before considering the details of the $C_\kappa$ domain, we discuss several recognized scenarios that may underly strong DNA conservation within coding regions: (i) conserved DNA segments may act vitally in transcriptional processing (e.g., activation, splicing); (ii) they may include features that modulate the stability of DNA duplex and/or mRNA transcripts; (iii) conserved DNA can assure more reliable codon–anticodon interactions; (iv) the conserved segments are simply evolutionary remnants without any specific function. It is useful to review briefly some of the arguments for these scenarios.

Significant DNA conservation is often a concomitant of the splice junction extending in both directions into the intron and exon. This is markedly the case for β-globin genes of animal species (17). The nature of the conserved coding DNA configuration may bear on requirements for avoidance of or conformance to certain DNA and RNA structures (secondary and tertiary) that enhance transcription, mRNA stability, and translational reliability. The nature (e.g., efficiency, binding energy) of codon–anticodon interactions is viewed as playing an important role in guiding selection among synonymous codons. The phenomenon of nonrandom codon usage for many genes has been documented (18, 19). It is observed that nonrandom codon usage is especially conspicuous in highly expressed genes, whereas codon biases are not as apparent for genes expressed at moderate to low levels (20, 21). Could one expect rare tRNA species to be more represented for gene expression required at low levels? The extent of this correlation has not been carefully evaluated.

Some proposals on relations of codon usage with nonrandom nucleotide-substitution rates are considered in some studies (for review, see ref. 22). The utilization of codons having abundant tRNA species for important coding segments can be expected to reduce translational errors. Nucleotide compositions manifesting certain codon preferences may be a way to avoid higher mutation rates and increase translational fidelity. Moreover, selection for moderation in stability of the mRNA secondary structure may induce nonrandom codon usage.

From these perspectives, we examine the $C_\kappa$-domain sections that are highly conserved at the DNA level compared to those sections conserved only in specific amino acid alphabets. Consider the 48-bp stretch from coordinates 172 to 219-5′ in $C_\kappa$ for which human and mouse agree except for 5 nucleotide mismatches. For the corresponding amino acids there is agreement from residues 58 to 73 except for residue 67. It is striking that the five Ser (codon degeneracy, 6) residues in the segment 58–73 are all encoded by the triplet AGC, whereas over the remaining $C_\kappa$ domain in each of the human and mouse sequences there are 11 Ser residues coded at roughly equal frequencies by the six synonymous codons. All 16 residues (except Lys in mouse) in the segment 58–73 have codons with strongly bonding bases (C or G) at the 3′ codon site, whereas over the whole $C_\kappa$ domain the total percent of strong bases at the third codon position is 69% in human and 59% in mouse. (The consensus cDNA of both α- and β-globin genes across mammalian species exhibits significantly more strongly bonding bases at the 3′ codon sites relative to its overall base frequencies, and this is especially pronounced in highly conserved segments.)

In assessing the dinucleotide composition of the third codon position (its 3′ site) with the subsequent 5′ codon site, we observe in the residue segment 58–73 of the human and mouse sequences a preponderance (9 of 15, 60%) of the doublet CA when compared to its occurrences over all the $C_\kappa$ domain (about 20%).

The other significant conserved DNA segment between the human and mouse $C_\kappa$ domains is a 15-bp stretch that corresponds to residues 99–103. Again, all five codons have a strongly bonding base at the 3′ site. Serine in this block is again encoded by AGC. Also the doublet CA appears twice in the third position and first position of abutting codons. In these conserved regions among the $C_\kappa$ rabbit sequences, the CA dinucleotide is represented more than twice as frequently as any other dinucleotide at the boundaries of consecutive codons. A marked preference for codon AGC for serine is also observed for all the $C_\kappa$ allotypes.

The foregoing features underscore the distinctive codon usage of these segments, possibly bearing on translational fidelity and codon–anticodon interactions. To compare with segments conserved only at the amino acid level, we examined the block identity covering residues 22–33 common to human and mouse and having a single amino acid mismatch at residue 27. We see seven nucleotide mismatches. Unlike the segment 58–73, 4 of the 12 codons of residues 22–33 in human end with a weakly bonding base, in conformity with the overall $C_\kappa$-domain nucleotide frequencies at the third position.

## DISCUSSION

Our amino acid alphabets (following refs. 3–5) emphasize properties of the amino acid side chains that relate to protein binding (functional and charge alphabets), tertiary conformation (structural alphabet), and chemical attributes (chemical alphabet). Finding significant block identities common to human, mouse, and rabbit $C_\kappa$ at the same position in many alphabets presumably supports the biological importance of this region. Significant block identities that occur in some alphabets and not in others may help highlight regions of particular functional characteristics.

The significant conserved segments (block identities) in the various amino acid alphabets with respect to the $C_\kappa$ domains of the human, mouse, and rabbit genes are independent of the allotypic sequence and display marked correspondence to tertiary structure (see *Analysis*, section 1). In contrast to the Y-face, the X-face (intermediate to and interacting with both the heavy chain and the Y-face) is well conserved in all but the charge alphabet. Statistically significant block identities in the amino acid alphabet emphasize the $X_1$ and $X_4$ segments. The $X_2$ segment shows significant three-species block identities in the chemical and functional alphabets but not at the simple amino acid level. This may indicate less stringent

constraints on the composition of this segment whereby only the chemical/functional equivalents of the residues are decisive and not the specific amino acid content.

There are known human Km and rabbit $C_\kappa$ sequence polymorphisms. The human allelic differences affect residue positions 45 and 83, which are well separated from any significant block identities for any of the alphabets. Even though the rabbit allotypes b4, b5, b9, and bas show an abundance of amino acid replacements, they all conserve two segments, at residues 22–30 and 67–73, in the chemical, functional, and structural alphabets. There is also complete identity, at the amino acid level, for all rabbit $C_\kappa$ sequences at residues 23–30; this segment contains the cysteine (residue 26) that serves to establish the intradomain disulfide bridge. Thus the triple significant block identities of Fig. 1 *b* and *c* do not depend on the $C_\kappa$ allotype sequence used.

In the charge alphabet, all five of the human $J_\kappa$ gene segments (each encoding 13 residues), the four functional mouse $J_\kappa$ segments, and the rabbit $J_\kappa 1$ segment encode the common charged-residue arrangement 0,0,0,0,0,0,0,+,0,−, 0,+,+. By contrast, the charge arrangement encoded by $J_\kappa 2$ of rabbit differs from the above consensus charge sequence at three positions. [This is curious because, to date, only the $J_\kappa 2$ products have been assayed, albeit under stringent selection, in response to a specific antigen (23)]. Thus, the charge alphabet shows remarkable similarity for the $J_\kappa$ gene segments of human and mouse but does not show any significant block identities relative to the $C_\kappa$ domain. This is in sharp contrast to the case of the human, mouse, and rabbit $\beta$-globin gene, where the charge and structural alphabets reveal an extraordinary degree of total matching (also for $\alpha$-globin; see ref. 17). Although the letter-frequency distributions of the charge alphabet are about the same for the $\beta$-globin gene and the $C_\kappa$ domain for these species (25% charged residues to 75% uncharged), the interspecies matching of charge configurations for the two gene products is markedly different. Between the human and mouse $C_\kappa$ domain, there are 11 charge-alphabet mismatches that are spread uniformly over the sequence. On the other hand, in the human versus mouse comparison of the $\beta$-globin gene, only three mismatches are found. Specifically, in the charge alphabet the human, mouse, and rabbit $\beta$-globin sequences match cumulatively in 143 out of 146 residues (the three mismatches all indicate replacements of a charged amino acid with an uncharged one); in contrast, over the $C_\kappa$ domain the common matches across all three species include only 77 of 106 residues. Apparently, the specific arrangement of charges is more important for $\beta$-globin function than it is for $C_\kappa$ performance.

The frequencies of charged residues among the different $C_\kappa$ rabbit allotypes [indicating only the numbers of positively (+) and negatively (−) charged residues] are as follows: b4 (9−, 6+); b5 (10−, 7+), b9 (12−, 7+), and bas (8−, 6+). The corresponding human and mouse charge occurrences are, respectively, (12−, 12+) and (13−, 12+). The discrepancy in this respect with rabbit is manifest. Among the four allotypes, each of the residue positions 41, 45, 82, and 87, contained in or proximal to the $Y_2$ and $Y_3$ faces, have representations from +, −, and 0 charged amino acids. Positions 45 and 83 are also the polymorphic sites of the human Km allotypes. Observe, in contrast, that segments 22–29 and 67–73, highly conserved among the three species, involve only uncharged amino acids.

At the DNA level, the segment containing amino acids 58–73, which overlaps the $X_2$ segment, is highly conserved, much more so than is required for amino acid conservation. Is it possible that the major active site of the $C_\kappa$ domain relates to this region? The appropriate amino acid composition could be germane for effector function, but judicious codon usage would provide more reliability in translation of this segment. Another cause for DNA conservation may relate to mRNA stability.

As suggested earlier, the residue segment 22–33 may be important for maintaining the essential structural interactions between the light and heavy chains and between the two $\beta$-sheets of the light chain in which only the chemical and functional equivalence of the amino acid sequence is vital. This might be investigated experimentally by introducing appropriate mutations in this segment that would alter the amino acid content but retain the chemical and/or functional equivalents.

To further assess the relevance of the comparisons using the multiple alphabets, we constructed a random four-letter alphabet obtained by randomly shuffling the amino acids and coalescing the first five to one letter, the next five to a second letter, etc. This random alphabet presumably entails no biological meaning. There was no significant block identity involving more than one mismatch position (i.e., a position showing distinct amino acids that happened to be identified in the random alphabet). This result enhances the significance ascribed to the block identities revealed by the chemical, functional, and structural alphabets but not seen with the 20-letter amino acid alphabet.

1. Karlin, S., Ghandour, G. & Foulser, D. E. (1985) *Mol. Biol. Evol.* **2,** 35–52.
2. Karlin, S. & Ghandour, G. (1985) *Mol. Biol. Evol.* **2,** 53–65.
3. Mahler, H. R. & Cordes, E. H. (1966) *Biological Chemistry* (Harper & Row, New York).
4. Lehninger, A. L. (1970) *Biochemistry* (Worth, New York).
5. Dickerson, R. E. & Geis, I. (1983) *Hemoglobin, Structure Function, Evolution, and Pathology* (Benjamin/Cummings, Menlo Park, CA).
6. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
7. Fitch, W. M. (1966) *J. Mol. Biol.* **16,** 9–16.
8. Haber, J. E. & Koshland, D. E. (1970) *J. Mol. Biol.* **50,** 617–639.
9. McLachlan, A. D. (1972) *J. Mol. Biol.* **64,** 417–437.
10. Grantham, R. (1974) *Science* **185,** 862–864.
11. Goodman, M. & Moore, G. W. (1977) *J. Mol. Evol.* **10,** 7–47.
12. Dayhoff, M. O. (1978) *Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, Suppl. 3.
13. Klein, J. (1982) *Immunology, The Science of Self-Nonself Discrimination* (Wiley, New York).
14. Weir, D. M. (1978) *Handbook of Experimental Immunology* (Blackwell, Oxford).
15. Emorine, L., Sogn, J. A., Trinh, D., Kindt, T. J. & Max, E. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1789–1793.
16. McCartney-Francis, N., Skurla, R. M., Jr., Mage, R. G. & Bernstein, K. E. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1794–1798.
17. Karlin, S. & Ghandour, G. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 6186–6190.
18. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. & Mercier, R. (1981) *Nucleic Acids Res.* **9,** R43–R75.
19. Ikemura, T. & Ozeki, H. (1983) *Cold Spring Harbor Symp. Quant. Biol.* **42,** 1087–1097.
20. Grosjean, H. & Fiers, W. (1982) *Gene* **18,** 199–209.
21. Ikemura, T. (1985) *Mol. Biol. Evol.* **2,** 13–35.
22. Li, W.-H., Luo, C.-C. & Wu, C.-I. (1985) in *Molecular Evolutionary Genetics*, ed. McIntyre, R. J. (Plenum, New York), 1–94.
23. Emorine, L., Dreher, K., Kindt, T. J. & Max, E. E. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 5709–5713.