The 48-base-long primordial building block of immunoglobulin light-chain variable regions is complementary to the primordial building block of heavy-chain variable regions

(primordial repeats/complementarity/antigen binding sites)

**SUSUMU OHNO AND TAKESHI MATSUNAGA**

City of Hope Research Institute, Duarte, California 91010

**Contributed by Susumu Ohno, December 23, 1981**

---

**ABSTRACT** The ancestral gene for immunoglobulin light-chain variable regions (Ig V_L) of the κ as well as the λ class apparently arose from about 12 tandem repeats of the 48-base-long primordial building block sequence TCT-TGC-GCA-GTA-AGT-CCA-CCT-CAG-GTC-ATA-TCC-AGT-CAG-GCT-GCT-GAA. Even today, amino acid residues 67 to 82 of each Ig V_L are still specified by a direct descendant in toto of the above-noted primordial building block, whereas amino acid residues 14 to 25 are invariably specified by its truncated copy. The Ig V_H primordial building block presently identified is 100% complementary to the Ig V_H (heavy-chain variable region) primordial building block previously identified. In the recognition of specific antigenic determinants by antibodies, Ig V_L and Ig V_H of light-chain-heavy-chain dimers have to complement each other. It is perhaps fitting that the primordial building blocks of the two are represented by the complementary strands of the same 48-base-pair-long DNA sequence.

In our previous papers, we have shown that about 12 tandem repeats of the 48-base-long prototype building block sequence TCT-TGC-GCA-GTA-AGT-CCA-CCT-CAG-GTC-ATA-TCC-AGT-CAG-GCT-GCT-GAA gave rise to the ancestral gene for immunoglobulin heavy-chain variable regions (Ig V_H). Although Ig V_H and Ig V_L are very similar in their size and general configuration, their amino acid sequence homology is not very impressive. Not surprisingly, when the published Ig V_L gene base sequences of the κ and λ classes of mice and humans were examined (3-7), we found in them only a few recognizable copies, representing mostly small fragments of the above-noted 48-base-long Ig V_H prototype building block sequence. This suggested the utilization of a different primordial building block for the construction of the ancestral Ig V_L gene. We now show that the primordial building block of Ig V_L genes was the complementary strand of the Ig V_H prototype building block sequence.

In Fig. 1, pertinent features of our previous findings on Ig V_H genes (1, 2) are recapitulated, using the mouse anti-phosphocholine T-15 (V2) gene as an example (8). Only the recognizable copies of the 48-base-long primordial building block sequence, shown in Fig. 2, that occupy the invariant positions within the V_H coding sequence proper are marked. It should be noted that the last 48 coding bases, specifying amino acid residues to 100 (residues 83 to 98 in most other V_Hs), still demonstrate 54% base sequence homology to the primordial building block sequence. In addition, amino acid residues 4 to 14, 17 to 23, 29 to 37, and 38 to 48 are specified by four truncated copies of the primordial building block (2).

In Fig. 2, the previously identified Ig V_L primordial building block sequence (2) is presented in the inverse, right-to-left order. Placed above it is the consensus amino acid sequence of Ig V_H, positions 83 to 98. This consensus is derived from the 1979 sequence compilation of various Ig V_H subclasses of man, mouse, rabbit, guinea pig, and dog (9). It should be noted that the codons for two of the four absolutely invariant amino acid residues (Cys-96 and Ala-97) are contained within the primordial building block, and so are the codons for four of the five nearly invariant residues; Leu-86, Asp-90, Tyr-95, and Arg-98. This region of all the published and in press coding sequences of mouse Ig V_H invariably demonstrates greater than 50% base sequence homology to the primordial building block (2). Using the mouse Ig V_H anti-NP (186-2) gene (which codes for the V_H chain of an antibody to nitrophenyl) as an example (10), this point is illustrated at the very top of Fig. 2.

Placed directly below the Ig V_H primordial building block base sequence is its complementary strand aligned in the proper left-to-right order. This complementary strand then is the primordial building block sequence of Ig V_L, as well as Ig V_L genes. The consensus amino acid sequence of positions 67 to 82 representative of Ig V_L of both κ and λ classes is shown immediately below it. This consensus is again derived from the 1979 sequence compilation of Ig V_L and Ig V_L of man, mouse, rat, rabbit, dog, and guinea pig (9). It may be noted that three of the four absolutely invariant residues of this region are specified by the codons contained in the primordial sequence: Ser-67, Leu-73, and Gln-79. As shown in Fig. 3, the homology between the Ig V_L primordial building block and the 48-base-long sequence specifying amino acids 67 to 82 of the mouse Ig MOPC41 V_L gene (3) is maintained at 50%. The corresponding region of the human Ig V_L 101 gene (6) maintains a 48% base sequence homology and that of the mouse Ig V_L 303A gene (7), a 42% homology, as shown at the bottom of Fig. 2. There remains little doubt that this region of all mammalian Ig V_L genes is specified by direct descendants in toto of the 48-base-long Ig V_L primordial building block sequence.

In Fig. 3, the germ-line base sequence of the mouse Ig MOPC41 V_L gene is shown (3). In it, 15 recognizable descendents derived from one or two segments of the 48-base-long Ig V_L primordial building block are identified. All these descendents were greater than 20 bases in length and demonstrated a 50% or greater base sequence homology to the corresponding segments of the Ig V_L primordial building block.

Abbreviations: H, heavy chain; L, light chain; V, variable region.
Not a single deletion of an insertion was allowed to maximize the homology with a given segment. As with all Ig VH genes (1, 2), segmental descendants of the primordial building block were found in adjacent noncoding sequences (5' and 3' spacer terminals and an intervening sequence) as well as in the hydrophobic leader coding sequence, thus confirming the previous finding on Ig VH (1, 2) that not only the VH coding sequence proper, but also its adjacent noncoding sequences and its hydrophobic leader coding sequence arose from tandem repeats of the one prototype building block which, in the case of Ig V L, is TCT-TGC-GCA-GTA-AGT-CCA-CTC-CAG-AGT-GCT-GCT-GAA. About the same number of recognizable descendants were found in two other mouse Ig V L genes (4, 5) as well as in one human V L gene (6) and one mouse Ig V AL gene (7). Within the VH coding sequence proper, however, only two fixed positions common to Ig V L and Ig V AL were invariably specified by recognizable descendants of the Ig VH primordial building block sequence; aside from amino acid residues 67 to 82 specified by a direct descendant in toto, amino acid residues 14 to 25 were specified by a 36-base-long truncated descendant of the primordial building block (Fig. 3). This paucity of descendants occupying the fixed positions within the VH coding sequence proper is not surprising, because the separation of Ig V L from Ig V AL apparently occurred 100 million or more years ago, before extensive adaptive radiation of mammalian species. In the case of Ig VH coding sequences, the most conserved triplet, Tyr-Tyr-Cys, constituting the residues 94 to 96, is specified by a part of the direct descendant of the Ig VH primordial building block (Fig. 2). By contrast, the origin of codons for the same amino acid triplet almost invariably occupying the positions 86 to 88 of Ig V L, as well as Ig V AL, remains rather obscure, in spite of the fact that, by a single-base substitution, the first two codons of the Ig VH primordial building block readily specify Tyr-Cys. It is a curious fact that positions 75 to 78 of certain Ig V AL s, but none of the Ig V Ls, of man and mouse are often represented by Ile-Ser-Ser-Leu (9). Because these quadruplet residues are quite homologous with the first four amino acid residues specified by the Ig VH primordial building block (Fig. 2), the 48-base-long region specifying amino acid residues 75 to 90 that included Tyr-Tyr-Cys at positions 86 to 88 of certain V AL s often exhibit striking base sequence homology with the Ig VH, instead of the Ig V L, primordial building block. A case in point is the Ig MOPC41 VH gene (3) shown in Fig. 3. While the region specifying amino acid residues 75 to 92 of this gene is 50% homologous with the Ig VH primordial building block, the partially overlapping 48-base-long region specifying residues 75 to 90 reveals a 46% base sequence homology with a complementary strand of the Ig V L primordial building block that is the Ig VH primordial building block (Fig. 2). Because the above is not a universal feature of all Ig VH s, its meaning is not very clear.

This work was supported in part by Grant AI 15620 from the National Institutes of Health and research grants from the Wakunaga Pharmaceutical Company of America and the Bixby Foundation.
Mouse $V_H$ anti-NP$^b$ (186-2)

Consensus $V_H$ sequence (five species)

$V_H$ building block

$V_L$ building block

Consensus $V_L$ sequence (six species)

Human $V_L$, 101

Mouse $V_L$, 303A

**Fig. 2.** The previously identified 48-base-long primordial building block ($V_H$ building block) of Ig $V_{H8}$ (2) is here presented in the inverted right-to-left direction, while its bases are numbered in the 5'-to-3' direction. Immediately below it is its complementary sequence in the proper left-to-right direction, which is the primordial building block of Ig $V_L$ (of both $\kappa$ and $\lambda$ classes, $V_L$ building block). Bases of these primordial building blocks are presented in large capital letters and so are the amino acid residues specified by them.

Because an intact copy in toto of each primordial sequence specifies only one fixed position within Ig $V_H$ or Ig $V_L$, the consensus amino acid sequence of this fixed position is shown immediately above or below the primordial building block. The consensus sequences were derived from the 1979 sequence compilation by Kabat et al. (9). In the case of positions 83 to 98 of Ig $V_{H8}$, five species were included in the compilation, and Ig $V_{H8}$ of rabbit, guinea pig, and dog were not divided into subgroups. Thus, the most common amino acid at a given position of each of the three species is given the value of 25, whereas the commonest amino acid of each of the four subclasses of mouse Ig $V_H$ is given the value of 6.25, and that of each of the three subgroups of man the value of 8.3. Accordingly, the value of 100 means the evolutionary invariant residues only with regard to the above criterion. An individual Ig $V_H$ may present a different amino acid in the invariant position. Of the consensus amino acid sequence so obtained, amino acid residues directly derived from the primordial building block are shown in large capital letters and others in small capital letters.

In constructing the consensus amino acid sequence for positions 67 to 82 of Ig $V_{L8}$, the same principle was employed. Two classes ($\kappa$ and $\lambda$) and six species (man, mouse, rat, rabbit, dog, and guinea pig), however, were included in the compilation, and only Ig $V_{L8}$ of man and mouse, and Ig $V_{L8}$ of man were divided into subgroups (9). With regard to residues 83 to 98 of Ig $V_{L8}$, the base and amino acid sequences of this portion derived from the mouse Ig $V_L$ anti-NP$^b$ (186-2) gene (10) are provided as an individual example. Amino acid sequence homology of 88% with the consensus sequence and base sequence homology of 52% with the primordial building block are evident. Amino acid residues not corresponding to the consensus are shown in italics. Other examples are provided by Fig. 1 and our previous publications (1, 2). Two individual examples for positions 67 to 82 of Ig $V_L$ are provided by the human Ig $V_L$, 101 gene (6) and the mouse Ig $V_L$, 303A gene (7).

**Fig. 3.** The 646-base-long mouse germ-line DNA segment that contained the coding sequence for Ig MOPC41 \( V_L \) (−4 to 95) and its attendant hydrophobic leader coding sequence (−22 to −4) is reproduced (3). Fifteen recognizable copies, each being longer than 20 bases and demonstrating greater than 50% base sequence homology with one or two segments of the 48-base-long \( V_L \) primordial building block (Fig. 2) are identified by solid bars. The two landmark cysteine residues for the ever-present intrachain disulfide bridge are shown in large capital letters, while those amino acid residues specified by original codons of the primordial building block are shown in small capital letters. Of the remainder, those homologous with human Ig \( V_L \) 101 (6) are shown in ordinary letters. Others are shown in italics. The 68% amino acid sequence homology reveals the extent of interspecific sequence conservation by Ig \( V_L \).