The Complete Murine Immunoglobulin Class Switch Region of the \( \alpha \) Heavy Chain Gene-hierarchic Repetitive Structure and Recombination Breakpoints*

(Received for publication, July 31, 1992)

Hirosi Arakawa, Takuji Iwasato, Hidenori Hayashida, Akira Shimizu‡, Tasuku Honjo‡‡, and Hideo Yamagishi†

From the Department of Biophysics, Faculty of Science, the †Center for Molecular Biology and Genetics, and the ‡Department of Medical Chemistry, Faculty of Medicine, Kyoto University, Kyoto 606, Japan

A 7255-base pair (bp) sequence, including the previously sequenced murine segments of \( \text{l}_{\alpha} \), \( \text{S}_{\alpha} \), and \( \text{C}_{\alpha} \), has been completed. Homology matrix comparison revealed a switch repetitive region of 4.2 kilobases (kb) composed of 20–80-bp homology runs, including the previously assigned \( \text{S}_{\alpha} \) region. We distinguished several stretches of duplication, \( \text{i.e.} \) the central 0.8-kb repetitive region, with some 80-bp staggered consensus repeats containing 20–30-bp subsets, made up of the primordial pentamers CTG(A/G)G. All the breakpoints of the \( \text{S}_{\alpha} \) switch recombination, including those generated by the translocation of the \( \alpha \)-myc protooncogene and those catalyzed by bacterial extracts, are located within the consensus sequence subsets of the 4.2-kb repetitive region.

The constant region genes of the immunoglobulin heavy chain (IgH) in the mouse genome are arranged in linear order as follows: 5'-C\( \gamma_{5} \)-C\( \gamma_{6} \)-C\( \gamma_{7} \)-C\( \gamma_{1} \)-C\( \gamma_{26} \)-C\( \gamma_{52} \)-C-C\( \alpha_{2} \)-C\( \alpha_{3} \)-3'. The Ig class, which is defined by the C region of the IgH chain, changes during the course of differentiation of a single B cell. This phenomenon, called class switching, is accompanied by DNA rearrangement, which takes place between switch (S) regions located 5' of each \( \text{C}_{\alpha} \) gene, except for the \( \text{C}_{5} \) gene. The S regions were first defined functionally as regions responsible for switch recombination, and subsequent analyses of nucleotide sequences around the recombination breakpoints also revealed characteristic structural features (2–4).

The target of most switch recombination events is the S\( \alpha \) region, which comprises the simple tandem repetitions CTGAG and GTGGG (2). Other S regions have also been assigned by means of related structural motifs in addition to cross-hybridization with the \( \text{S}_{\mu} \) probe (1–3). By analyzing the circular DNAs produced as deletion products of switch recombination (5–8), recombination breakpoints were identified in switch-repetitive regions. This new approach seems to reflect the primary switch recombination, because chromosomal breakpoints in myeloma cells, virus-transformed cell lines, and hybridomas (4, 9–14) were often found outside such structurally distinct switch-repetitive regions, possibly due to frequent secondary deletions. Upon analyzing lipopolysaccharide-transforming growth factor \( \beta \)-induced switch circular DNA, we found that two \( \text{S}_{\alpha} \) breakpoints mapped to a new region that was comprised of a simple repetition of 5-base consensus sequences (15) outside the previously assigned \( \text{S}_{\alpha} \) region. Since only limited sequence data on \( \text{S}_{\alpha} \) regions are available, it has been difficult to relate the \( \text{S}_{\alpha} \) breakpoints to longer stretches of switch-repetitive sequences. We isolated several overlapping deletion fragments located between already sequenced segments and connected four data base sequences of a total of 7255 bp. We searched for repetitive sequences by a computerized sorting method and found a switch-repetitive sequence of 4.2 kb that displays an underlining hierarchic repetitive structure defined by the primordial pentamers CTG(A/G)G, which organize into larger repetitive units of 10, 20, 30, 80, and 800 bp.

EXPERIMENTAL PROCEDURES

The nucleotide sequencing strategy for repetitive sequences is described under “Results.” Several deletion fragments of pCS18\( \beta \) and pCS14\( \alpha \) derived from BALB/c mice were prepared as follows. Both fragments were recloned into the polylinker cloning sites of pHSG399 and digested by SpfI. Purified SpfI fragments were digested by XbaI. These SpfI-XbaI fragments were treated with exonuclease III and VII as described (16). Deletion fragments were self-ligated after deleting a stretch of nucleotides with T4 DNA polymerase using a DNA blotting kit (Takara Shuzo Ltd., Kyoto, Japan) and transferred into Escherichia coli DH5\( \alpha \) recA1 endA1. A computer program was used to generate diagonal lines indicating segments of 20 bases long that show homology above a threshold level (17).

RESULTS

Sequencing Strategy—The 10-kb EcoRI fragment (IgH703) isolated previously from BALB/c mice (18) contains the region between the \( \text{l}_{\alpha} \) segment and the \( \text{C}_{\alpha} \) region in which the entire \( \text{S}_{\alpha} \) region was supposed to be located. Four segments of this fragment have been sequenced and registered in the data bases M29011, X62548, J00474, and J00475 as shown in Fig. 1. The sequence data available under accession number X62548 were previously constructed from the overlapping \( \text{S}_{\alpha} \) circular DNA clones (15). To connect these four segments, we subcloned five gene fragments covering the three sequence gaps between the sequenced segments (a 4.2-kb XbaI fragment (pCS15), a 0.6-kb XbaI fragment (pCS51), a 2.5-kb SacI-XbaI fragment (pCS16\( \gamma \)), a 0.6-kb PstI-XbaI fragment (pCS16\( \alpha \)), and a 1.8-kb XbaI-EcoRI fragment (pCS14\( \alpha \)) (Fig. 1). These genomic subclones were stable during their propagation in the
Hierarchic IgH Sa Repetitive Sequences and Breakpoints

New 7255 bp database (accession no. D11468)

|   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|
| M29011 | X62548 | J00474 | J00475 |
| (1 - 597) | (1152 - 2554) | (2834 - 4265) | (5589 - 7255) |

EX X KH X X HS P X P X P E S S

Iα

pCS51

R M

4.2 kb

RC

pCS15

1 kb

pCS16α

TG TG

-35 -36

J

H

G

F

E

D

C

B

A

pCS16γ

M

R

pCS14α

Internal Homologies—To find internal homologies, we generated a homology matrix between identical pairs of the complete sequences. Segments of 20 bp having more than 80 and 90% homology to other segments are shown by thin and bold (double width) diagonal lines, respectively (Fig. 2a). Internal homologies are clustered in a 4.2-kb area. We found three long direct repeats, DR1, DR2, and DR3, in the 5' to 3' orientation. DR1 and DR3 repeats share homology with other repetitive sequences, whereas the DR2 repeat shows a unique sequence unit. We looked back directly at the original data to find the maximum unit length of these repeats. DR2 is composed of homopurine stretches of 130 bp at position 3140-3269 reading (AGGAG)₂AAGAG(AGGAG)₃, and DR1 consists of 65 nucleotides at position 3070-3134 composed of (CTAGG)₁₃ as found previously (21). DR3 comprises 132 nucleotides at position 4369-4500 reading ((TTAGT)(CTGG-G)(CTAGG)(CTAGG))₉(TTAGT)(CTAGG)CT.

Internal homologies of longer stretches are most enriched in the central part at position 2.3-3.1 kb followed by the downstream region. Internal homologies of only shorter stretches were found in the region upstream of the central part. We aligned the long internal homologies at position 2269-3871 excluding DR2 and tried to create maximal sequence similarity alignment by manually inserting gaps (Fig. 3a). A consensus sequence of an 80-bp repeating unit showing more than 60% homology to its individual repeats was identical to the prevalent sequence of Sa (18) previously processed from data base J00474 (Fig. 3b). The 80-bp consensus sequence contains two 30-bp repeating units, which are synonymous to the 5'-'Ca consensus sequence (9) and three 20-bp repeating units (Fig. 3b). Further analysis of the 80-bp con-
Hierarchic IgH Sα Repetitive Sequences and Breakpoints

(a) Nucleotide sequence around Sα (accession no. D11468)

(b) Homology matrix between identical 7255-bp sequences (accession no. D11468) including the complete Sα region. DR1, DR2, and DR3 represent long direct repeats. No significant internal homologies were seen in the sequence downstream of 5000 bp. b, stretches of the 80-bp consensus sequence in the Sα region shown by homology matrix. Recombination breakpoints are indicated by numbered vertical lines; T2-7 (1), pCS31 (2, 11), pCS43 (3), pCS44 (4), pCS20 (5), pCS19 (6), T2-11 (7), pCS41 (8), pCS18 (9), T2-2 (10), and pCS32 (12) for switch circular DNA (6, 15); Ea.Jα (13), 41.3/40.1 (14), Sμκα (15), ABPC45 (16), Sκκα (17), M167 (18), BF0.5/0.6 (19), J558 (20), and Sκκα (21) for switch recombination in myeloma (6, 11, 21, 23, 31); ABPC45 c-myc (22), HOPC1R (23), W267R (24), W267M (25), M315M (26), HOPC1M (27), J558M (28), M167c7 (29), and M603a30 (30) for c-myc translocation in myeloma (32-34); MOPC467 (31) for Jκ/Sκ joint in myeloma (35); Ig μ-α-15 (32), Ig μ-α-20 (33), Ig μ-α-11 (34), and Ig μ-α-1 (35) for bacterial extracts (36). Breakpoint positions are 922 (T2-7), 1332 (pCS31), 1490-1498 (Igκ-15), 1770 (pCS43), 1875-1877 (ABPC45 c-myc), 1906 (pCS44), 2020 (HOPC1R), 2054 (pCS20), 2236 (pCS19), 2332 (W267R), 2366 (T2-11), 2371 (pCS41), 2372-2377 (pCS18), 2378 (T2-2), 2436 (W267M), 2452 (M315M), 2855 (Ea.Jα), 2904 (pCS31), 2914-2915 (41.3/40.1), 2918-2919 (Sκκα), 3008 (ABPC45), 3077-3079 (pCS32), 3288 (Sκκα), 3474 (M167), 3475-3486 (BF0.5/0.6), 3539 (Igκ-20), 3788 (HOPC1M), 3936-3940 (J558), 3946 (J558M), 3955 (M167c7), 4075 (Sκκα), 4077 (M603a30), 4120-4131 (Igκ-11), 4373-4378 (Igκ-1), and 4469 (MOPC467).
sensus sequence revealed eight 10-bp repeats (NTGRGCTGGG) as well as the primordial pentamers (CTGRGG). These pentamers were nearly homologous to both of the pentamers units of the Sa region, CTGAG and GTGGG.

Recombination Breakpoints—To analyze the prevalence of the 80-bp consensus sequence motif enriched in the central part of the internal homology area, we compared it with the whole 7255-bp sequence to see how far the matches would continue and printed out those matches and lengths by generating a homology matrix where diagonal lines indicate segments of 20 bp, which show more than 75% homology (thick line) or more than 80% homology (double width bold line) (Fig. 2b).

These consensus sequence motifs are most strikingly distributed within the internal homology area of 4.2 kb between positions 627 and 4806 excluding DR2 (Fig. 2b). We assigned part of the internal homology area, we compared it with the 80-bp consensus sequence motif enriched in the central part of the internal homology region. Two long internal repeats, DR1 and DR2, are congeners in different unit lengths of hierarchic classes (4.2 and 0.8 kb for myeloma, 10 for myc translocation or deletion in myeloma, and 4 for the recombination catalyzed by bacterial extracts). Furthermore, one of the primordial pentamers (CTGGG) was actually acts as a functional Sa region.

We screened a complete 7255-bp sequence for a heptamer consensus motif (YAGGTTG), which had been found near the majority of switch recombination sites in plasma cell tumors and hybridoma lines (23, 24). Six of seven heptamers were found in a small region within So from positions 3281 to 3766, and another heptamer was outside Sa (5240–5246). This paucity of the heptamer consensus motif in the flanking region of Sa might correlate with the fact that all the switch-α breakpoints in myeloma were located within the So region, whereas myeloma switch-μ breakpoints were often located outside the Sμ (6, 15).

**DISCUSSION**

Three major classes of highly repetitive DNAs have been identified in primates: tandemly repeated sequences, such as the α-satellite, as well as both short (SINE) and long (LINE) interspersed repetitive nucleotide sequences (25). However, the So switch-repetitive region described here is different from these highly repetitive DNAs, since the So sequence repeats are well organized in different unit lengths of hierarchic classes (4.2 and 0.8 kb and 80, 30, 20, 10, and 5 bp). This suggests the duplication of an ancestral chromosomal region. Two long internal repeats, DR1 and DR2, are conserved in the central part of the internal homology region. DR1 is composed of tandem primordial pentamers constituting an 80-bp consensus sequence of recombination hot spots. The central 0.8-kb region containing DR1 is rich in this consensus sequence. Several duplications of the 0.8-kb region may have formed the 4.2-kb So switch-repetitive region, while DR2 missing recombination hot spots remained unduplicated. This duplication of the So switch region may have proceeded independently of the evolution of the Cγ gene locus presumably generated by Cy duplication in the ancestral chromosomal region containing Cα-Cγ-Cy-Cα (26, 27). The higher
order structure of switch-repetitive regions as found in the Sα region seems to be a common feature in the S region of longer stretches, Sγ2, Sγ3, and Sυb (28–30). Duplication events in switch regions may have elongated S regions. This can be correlated with the relative contents of immunoglobulin classes in mouse serum (3). During duplication events, the primordial pentamer sequences degenerated in all the Sγ regions but remained well conserved in the Sμ, Sε, and Sα regions (3). The Sα region is common to the Sμ and Sε regions with respect to the conservation of primordial pentamers and also to the Sγ regions with respect to the higher order duplication structure.

Switch recombination sites tend to share little homology as shown previously (3, 6–8, 15). Thus, the enzymes that mediate the cutting-religation reaction do not seem to recognize an obviously conserved sequence. However, the close correlation between recombination breakpoints and the Sα consensus sequence suggests that switches to Sα involve recognition of at least a 20-bp motif or a subset of the 80-bp consensus sequence including the primordial pentamer CTGGG. Longer stretches of the GC-rich sequence may be required for the sequence structure defined by internal homology agrees with the sequence including the primordial pentamer CTGGG. Longer stretches of the GC-rich sequence may be required for the sequence structure defined by internal homology agrees with the structure of switch-repetitive regions as found in the Sα region of longer stretches.

Acknowledgment—We thank Dr. Martina Nazarea for critical reading of the manuscript.

REFERENCES

1. Shimizu, A., Takahashi, N., Yaota, Y., and Honjo, T. (1982) Cell 28, 499–506.
2. Nikaido, T., Nakai, S., and Honjo, T. (1981) Nature 293, 845–848.
3. Nikaido, T., Yamawsaki-Kataoka, Y., and Honjo, T. (1982) J. Biol. Chem. 257, 7322–7329.
4. Kataoka, T., Kawakami, T., Takahashi, N., and Honjo, T. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 919–923.
5. von Schwedler, U., Jack, H.-M., and Wahi, M. (1990) Nature 345, 452–458.
6. Maruoaka, M., Yoshida, K., Maeda, T., Uenda, S., and Sakano, H. (1990) Cell 62, 135–142.
7. Iwasato, T., Shimizu, A., Honjo, T., and Yamagishi, H. (1990) Cell 62, 143–149.
8. Yoshida, K., Maruoaka, M., Usuda, S., Mori, A., Ishizaka, K., and Sakano, H. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 7859–7863.
9. Davis, M. M., Kim, S. K., and Hood, L. E. (1980) Science 209, 1590–1595.
10. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W., and Tonegawa, S. (1980) Nature 286, 670–673.
11. Kim, S., Davis, M., Sinn, P., Pattin, P., and Hood, L. (1981) Cell 27, 573–584.
12. Lang, R. B., Stanton, L. W., and Marcus, K. B. (1982) Nucleic Acids Res. 10, 611–630.
13. Akira, S., Sugiyama, H., Yoshida, N., Kikutani, H., Yamamura, Y., and Kishimoto, T. (1983) Cell 34, 545–556.
14. Winter, E., Krawinkel, U., and Radbruch, A. (1987) EMBO J. 6, 1663–1671.
15. Iwasato, T., Arakawa, H., Shimizu, A., Honjo, T., and Yamagishi, H. (1992) J. Exp. Med. 175, 1539–1546.
16. Guo, L.-H., and Wu, R. (1983) Methods Enzymol. 100, 60–96.
17. Miyata, T., Hayashida, H., Kikuno, R., Toh, H., and Kawade, Y. (1985) in Interferon Biology (Gresser, I., ed.) pp. 1–30, Academic Press Ltd., London.
18. Obata, M., Kataoka, T., Nakai, S., Yamagishi, H., Takahashi, N., Yamawaki-Kataoka, Y., Nikaido, T., Shimizu, A., and Honjo, T. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 2479–2484.
19. Marcus, K., Banerji, J., Pennacavage, N. A., Lang, R., and Arheim, N. (1980) Cell 22, 157–169.
20. Sanger, F. (1981) Science 214, 1205–1210.
21. Collier, D. A., Griffin, J. A., and Wells, R. (1986) J. Biol. Chem. 261, 7397–7405.
22. Fahlender, F. O., Sugimura, J., Yang, J.-Q., Wiener, F., Marcus, K. B., and Klein, G. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 3746–3750.
23. Marcus, K. B., Lang, R. B., Stanton, L. W., and Harris, L. J. (1982) Nature 298, 87–89.
24. Ott, D. E., and Marcus, K. B. (1989) J. Immunol. 1, 582–591.
25. Sanger, F., and Nicklen, S. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 5463–5467.
26. Flanagan, J. G., and Rabbits, T. H. (1982) Nature 300, 709–713.
27. Hayashida, H., Miyata, T., Yamawaki-Kataoka, Y., Honjo, T., Wels, J., and Blattner, F. (1981) EMBO J. 3, 2047–2053.
28. Wu, T. T., Reid-Miller, M., Ferry, H. M., and Kabat, E. A. (1984) EMBO J. 3, 2033–2040.
29. Sanger, F., Petreni, J., and Dunnick, W. (1985) J. Immunol. 135, 620–626.
30. Mowatt, M. R., and Dunnick, W. A. (1986) J. Immunol. 136, 2674–2683.
31. Dunnick, W., Wilson, M., and Stavnezer, J. (1988) Mol. Cell. Biol. 8, 1850–1856.
32. Calame, K., Kim, S. L., and Lee, R. H., Davis, M., and Hood, L. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 6994–6998.
33. Piccione, S. P., Catini, P. G., and Cole, M. D. (1984) Nature 310, 527–530.
34. Gerondakis, S., Cory, S., and Adams, J. M. (1984) Cell 36, 973–982.
35. Aguiller, R. J., Hope, T. J., and Sakano, H. (1982) EMBO J. 1, 2898–2903.
36. Kataoka, T., Takeda, S., and Honjo, T. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 2066–2070.