Nucleotide sequences and three-dimensional modelling of the V_H and V_L domains of two human monoclonal antibodies specific for the D antigen of the human Rh-blood-group system

Nevin C. HUGHES-JONES,* Jacqueline M. BYE, Dennis BEALE and John COADWELL

M.R.C. Mechanisms in Immunopathology Unit and Department of Immunology, A.F.R.C. Institute of Animal Physiology and Genetics Research, Babraham, Cambridge CB2 4AT, U.K.

The nucleotide sequences were determined for the V_H and V_L domains of two human IgG1 antibodies, Pag-1 and Fog-B, specific for the D antigen of the Rh-blood-group system. The V_H-region genes of the two antibodies were derived from separate germ-line genes within the V_{H-IV} gene family, but both antibodies used the same J_{H6} gene. The D-region genes differed from each other, and no similarity was found to known D regions. The light chain of Fog-B belongs to the V_{L-1} subgroup and that of Pag-1 probably belongs to the V_{L-2} subgroup; both light chains used the J2/3 gene. Three-dimensional models of the variable domains were made, based on those of known crystallographic structure. The surface contours at the combining sites are clearly different, consistent with the evidence that the antibodies recognize different but overlapping epitopes. Some details of the molecular modelling of hypervariable regions have been deposited as Supplementary Publication SUP 50155 (6 pages) at the British Library Document Supply Centre, Boston Spa, Wetherby, West Yorkshire LS23 7BQ, U.K., from whom copies can be obtained on the terms indicated in Biochem. J. (1990) 265, 5.

INTRODUCTION

The V domains of the H and L chains of immunoglobulins are coded in the mature B-cell by single exons. These exons are derived during B-cell maturation by selection and rearrangement from many V_H, V_L, D, J_H and J_L segments (see review by Taussig, 1988). It is of considerable interest to determine whether or not certain gene segments are selectively expressed in the V domains of human antibodies specific for a single antigen.

The recent development of methods for establishing stable cell lines derived from human B-lymphocytes by heterohybridoma formation has led to the acquisition of a number of monoclonal antibodies specific for the D antigen of the human Rh-blood-group system (Thompson *et al.*, 1986; Melamed *et al.*, 1987). We present here the nucleotide sequences of the variable regions of two of these monoclonal antibodies, Pag-1 and Fog-B. The polypeptide carrying the D antigen is an erythrocyte membrane protein of molecular mass about 30 kDa (Moore *et al.*, 1982), and there is evidence that Pag-1 and Fog-B recognize different but overlapping epitopes on the molecule (Gorick *et al.*, 1988; Lomas *et al.*, 1989).

In addition to nucleotide sequencing, we have also carried out preliminary modelling of the antibody combining site by taking advantage of the similarities between the V-domain sequences of these two antibodies and the sequences of those structures whose three-dimensional conformations have already been elucidated.

METHODS

The human monoclonal antibodies Pag-1 and Fog-B were derived from different donors and have been previously described (Melamed *et al.*, 1987). They are IgG antibodies produced by human-mouse heterohybridomas; both are of the IgG1 subclass and both have lambda light chains.

Total cytoplasmic RNA was extracted from the cell line by using the Nonidet P40 lysis method (Favaloro *et al.*, 1980). Poly(A)-mRNA was isolated by oligo(dT)-cellulose column chromatography. First-strand cDNA was synthesized by priming with oligonucleotides complementary to the 5' end of the CH1 [5'-GGGAAGGTGTGCACGCC-3'] and the C λ [5'-AGCTCCTCAGAGGAGGG-3'] exons (kindly donated by Dr. G. Winter) and then cloned essentially as described by Gubler & Hoffman (1983). The G-tailed cDNA was annealed with C-tailed pUC18 plasmid vector. Recombinant plasmids were used to transform *Escherichia coli* DH5- λ by the high-efficiency method described by Hanahan (1983). The resultant colonies were screened by using the relevant 5'-end-labelled oligonucleotide primer as a probe. Positive clones were sequenced by using the dideoxy chain-termination method (Sanger *et al.*, 1977), as modified by Wilson *et al.* (1988) for super-coiled plasmids. Sequencing was carried out on the entire length of both strands of the cDNA.

Molecular modelling was carried out on an Evans and Sutherland PS390 graphics work station linked to a Vax 11/750 computer using the FRODO program devised by Dr A. J. Jones and modified by Dr P. Evans and Dr R. Hubbard. Atomic co-ordinates for the crystallographic antibodies Fab NEWM (Saul *et al.*, 1978), Fab KOL (Marquart *et al.*, 1980), Fab MOPC603 (Segal *et al.*, 1974), Fab J539 (Suh *et al.*, 1986), Fab HYHEL5 (Sheriff *et al.*, 1987), Bence-Jones protein REI (Epp *et al.*, 1975) and Bence-Jones protein RHE (Furey *et al.*, 1983) were obtained from the Brookhaven Data Base (Bernstein *et al.*, 1977). They are referred to in the text as crystallographic antibodies. Hypervariable regions were modelled as described by Chothia *et al.* (1986) and Padlan & Kabat (1988) with the use of the canonical forms described by Chothia & Lesk (1987). Conformational data for reverse turns were obtained from Chothia & Lesk (1987), Wilmot & Thornton (1988) and Sibanda *et al.* (1989). Side chains

Abbreviation used: CDR, complementary determining region.

* To whom correspondence should be addressed.

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers X52109 (Fog-B L chain), X52110 (Fog-B H chain), X52111 (Pag-1 H chain) and X52112 (Pag-1 L chain).

were positioned by using the criteria of Ponder & Richards (1987) and McGregor *et al.* (1987).

General sequence information and location of the complementary determining regions (CDRs) were obtained from Kabat *et al.* (1987), and their numbering has been used whenever possible. There are a few instances where residues that would be expected to fulfil the same conformational role in different domains have different numbers in Kabat *et al.* (1987). They have been given the same numbers as each other in the present text.

DNA sequences were obtained from the GenBank and EMBL Databases and amino acid sequences from the Leeds Composite Database Package. They were compared with the Pag-1 and Fog-B sequences by using the GAP program of the WISCONSIN package. Secondary-structure predictions were made by using the PEPLOT program.

RESULTS

Nucleotide and derived amino acid sequences

Heavy chains. The nucleotide and derived amino acid sequences

of the heavy-chain variable regions of both IgG Pag-1 and IgG Fog-B are given in Fig. 1.

V_H-region genes. Both heavy-chain variable regions (residues 1–94) of the two antibodies belong to the V_{H-IV} family (Lee *et al.*, 1987); the closest similarity in nucleotide sequence for Pag-1 is with the V_{71.2} germ-line gene (95% identity) and for Fog-B is with V_{2.1} germ-line gene (92% identity).

J_H-region genes. There is similarity between the antibodies in that they both use the J6 gene (Ravetch *et al.*, 1981). In the case of Fog-B, the 17 5' nucleotide residues of the germ-line J6 region appear to have been lost during D–J joining; on the other hand, it is probable that the entire J6 gene is utilized in Pag-1, if it is postulated that there are three deletions (GGT) at positions 16–18 from the 5'-end (Fig. 2). If this is so, then there are seven nucleotide changes for Pag-1 and three for Fog-B, resulting in one amino acid change in Fog-B (Val-102 to Leu-102) and three in Pag-1 (Tyr-100d to Gln-100d, Val-100f to Ser-100f and Gln-105 to Lys-105).

D-region genes. There is no close similarity between the nucleotide sequences of the D regions (residues 95h–100h) of Pag-1 or Fog-B and any of the 18 D regions that were examined

Fig. 1. Nucleotide and amino acid sequences of V_H domains of Fog-B and Pag-1

Fig. 2. Comparison between the J_H regions of Fog-B and Pag-1 and the germ-line gene J6 (Ravetch *et al.*, 1981)

Fig. 3. Nucleotide and amino acid sequences of V₁ domains of Fog-B and Pag-1

J₁-region genes. Both light chains use either the J₁2 or the J₁3 gene, which are identical (Udey & Blomberg, 1987; Dariavach *et al.*, 1987); that of Pag-1 shows complete identity with J₁2/3, and that of Fog-B shows two differences in substituting G for A at position 357 and A for G at position 388.

Fig. 4. Stereodrawings showing hypothetical models of the α -carbon skeletons of the V domains of IgG Fog-B and IgG Pag-1

These models are viewed looking along the pseudo-axes towards the binding sites. Heavy chains are on the right and light chains on the left. CDRs are shown as thick lines. The CDR1 loops of the V_L chain (starting at 24L) and the V_H chain (31H) of Fog-B protrude further into the solvent than the Pag-1 counterparts. The V_L CDR3 loop of Pag-1 (end of loop labelled 97L) protrudes further into the solvent than its counterpart in Fog-B. This drawing is for illustrative purposes only. The models need refining as more information on CDRs becomes available and when extensive energy minimization is carried out. We expect that modelled CDRs will require relatively little adjustment except for CDR3 of both V_H domains. These may require considerable revision.

Molecular modelling of hypervariable regions

The following is a brief account of the comparisons between the CDRs of Pag-1 and Fog-B and the five crystallographic antibodies upon which modelling was based. Further details are contained in the Supplementary Publication SUP 50155.

Heavy chains. CDR1. All the crystallographic V_H domains have similar main-chain conformations in their first CDR region, and V_H Pag-1 also has the same number of CDR1 residues. Modelling this region was therefore straightforward. V_H Fog-B has two extra CDR1 residues and these were placed at positions 32a and 32b.

CDR2. Both Pag-1 and Fog-B have the same number of CDR2 residues as in V_H NEWM, and modelling was therefore based on the latter.

CDR3. The number of residues in the third CDR regions of the crystallographic domains varies from seven to 17, and conformations are much more difficult to predict. Modelling of the CDR3 regions of Fog-B and Pag-1 was based on those of V_H MOPC603 and of V_H R19-9 (Lascombe *et al.*, 1989).

Framework regions. The framework region of V_H Pag-1 was based on that of V_H NEWM (74% identity), and that of Fog-B on V_H KOL (68% identity).

Light chains. CDR1. V_L Fog-B has the same number of residues as V_L KOL, and eight of the 13 residues are identical. Modelling of the CDR1 region of Pag-1 was also based on V_L

KOL, although it has two fewer residues (residues 26 and 26a in V_L KOL).

CDR2. Modelling of this region was based on that of MOPC603. Both antibodies had the same number of residues, although Fog-B shows less sequence similarity than Pag-1.

CDR3. The main-chain conformations of the three available crystallographic V_L chains are closely related. Modelling of this region for Fog-B was based on V_L KOL, which has the same number of residues. V_L Pag-1 has two additional residues (93b and 93c), which were used for extending the reverse turn in this region, also based on V_H KOL.

Framework regions. The greatest sequence similarity was to V_L KOL in both instances (Pag-1, 70% identity; Fog-B, 74% identity).

Stereo drawings of the three-dimensional model are shown in Fig. 4; the viewing position looks into the binding site along the pseudo-axis of the Fab arm. This is the same orientation as used by Kabat *et al.* (1987).

DISCUSSION

One of the aims of elucidating nucleotide sequences of antibodies specific for a single antigen is to determine the extent of diversity or restriction in the use of germ-line genes. Crews *et al.* (1981) studied 19 mouse immunoglobulins binding phosphocholine and concluded that virtually the entire immune response was derived from a single V_H -coding sequence. It is more usual,

however, to find diversity. Sanz & Capra (1987) reported that three families of antibodies were found in the response to *p*-azidobenzenearsonate by the A/J strain of mice, but that there was some restriction in that the light chains within one of the families were probably all derived from a single germ-line gene. Mudgett-Hunter *et al.* (1985) found broad diversity in the response to digoxin, and Clarke *et al.* (1985) found that the mouse response to the Sb epitope of influenza haemagglutinin also showed V_H diversity, even between antibodies derived from a single mouse. Berek *et al.* (1985) found restriction in the primary response to oxazolone but considerable diversity in the secondary response.

The two anti-D antibodies described in the present paper show similarity in the use of J-region genes, but not in the use of V_H , D or V_L genes. Both antibodies use the same J6 gene in their heavy chains and the same J2/3 gene in their light chains. The V_H genes both belong to the relatively uncommon V_{H-IV} family (Lee *et al.*, 1987), but are almost certainly derived from different germ-line genes, as is suggested by the finding that there is only 89% sequence identity between them; Fog-1 shows a closer sequence similarity to V_{2-1} (95% identity), and Pag-1 has 92% sequence identity with V_{71-2} . The D-region genes are completely different from each other, and the V_L genes belong to different subgroups. These findings indicate the use of a relatively diverse range of germ-line genes for human antibodies against the D antigen.

The D antigen of the Rh-blood-group system is a polypeptide of molecular mass about 32 kDa placed deeply in the erythrocyte surface and probably globular in structure. Seven different epitopes have so far been described (Lomas *et al.*, 1989). The two monoclonal antibodies described in the present paper recognize similar numbers of sites on each erythrocyte (approx. 23000), but inhibition experiments showed that, whereas Fog-B could completely inhibit the uptake of Pag-1, the latter would only partially inhibit the uptake of Fog-B (Gorick *et al.*, 1988). Although the exact mechanism underlying the differential inhibition results is not clear, it has been tentatively concluded that the two antibodies recognize different epitopes, which either overlap or are so close together that the occupancy of one epitope by antibody sterically hinders access to the other. This belief that the antibodies recognize different epitopes is given some support by the finding that the binding sites differ in their three-dimensional structure. Even at the level of main-chain conformation the CDR1 regions of Fog-B V_L domain (loop starting at 24L; Fig. 4) and of Fog-B V_H domain (31H) protrude further into the solvent than their counterparts in the V_L domain (24L) and the V_H domain (31H) of Pag-1. The main chain of the CDR3 of Pag-1 V_L domain (end of loop labelled 97L; Fig. 4) also protrudes further into solvent than that of the V_L domain (97L) of Fog-B. Moreover, there are several conformational differences in the CDR3 regions of the V_H domains (end of loop, 102H). Although side-chain conformations in the preliminary model have not been energy-minimized, numerous differences that affect surface topology and charge are to be expected between analogous CDRs in Pag-1 and Fog-B even where main-chain conformations have been predicted as being very similar. For example, at position 58 in V_H CDR2 Fog-B has a tyrosine residue but Pag-1 has an asparagine residue. This involves a difference in side-chain volume of 0.076 nm^3 ($76 \text{ \AA}^3$), a change from aromatic to aliphatic properties and the loss of a potential hydrogen-bonding atom. In V_L CDR2 at position 53 Fog-B has a lysine residue but Pag-1 has an aspartic acid residue. This involves a charge reversal and a gain of one potential hydrogen-bonding atom. There are thus substantial structural differences in the antibody combining sites of the two antibodies.

Contact residues between antibody and antigen have been

published for three antibodies against different epitopes on the lysozyme molecule (Amit *et al.*, 1987; Sheriff *et al.*, 1987; Padlan *et al.*, 1989). Each antibody uses all six CDRs in binding antigen. The main difference in structure between Pag-1 and Fog-B and the anti-lysozyme antibodies lies in the CDR3s of the V_H domains. In the two blood-group antibodies there are 16 and 17 residues in contrast with only five to eight residues in the anti-lysozyme antibodies. However, without knowledge of the D-antigen structure it is at present impossible to speculate as to how the anti-D CDRs might be used.

Anti-D antibodies bind more firmly to the antigen as the ionic strength of the medium is lowered, and there is evidence that at least three charged groups are involved (Hughes-Jones *et al.*, 1964). The CDRs of Fog-B and Pag-1 contain ten and eight charged groups respectively, which could be involved in the interaction.

We thank Dr. Greg Winter for his valuable advice on technical matters and for the supply of the oligonucleotide primers.

REFERENCES

- Amit, A. G., Mariuzza, R. A., Phillips, S. E. V. & Poljak, R. J. (1987) *Science* **233**, 747–753
- Berek, C., Griffiths, G. M. & Milstein, C. (1985) *Nature (London)* **316**, 412–418
- Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112**, 535–542
- Buluwela, L., Albertson, D. G., Sherrington, P., Rabbitts, P. H., Spurr, N. & Rabbitts, T. H. (1988) *EMBO J.* **7**, 2003–2010
- Chothia, C. & Lesk, A. M. (1987) *J. Mol. Biol.* **196**, 901–917
- Chothia, C., Lesk, A. M., Levitt, M., Amit, A. G., Mariuzza, R. A., Phillips, S. E. V. & Poljak, R. J. (1986) *Science* **233**, 755–758
- Clarke, S. H., Huppi, K., Ruzinsky, D., Staudt, L., Gerhard, W. & Weigert, M. (1985) *J. Exp. Med.* **161**, 687–702
- Crews, S., Griffin, J., Huang, H., Calame, K. & Hood, L. (1981) *Cell* **25**, 59–66
- Dariavach, P., Lefranc, G. & Lefranc, M.-P. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 9074–9078
- Epp, O., Lattman, E., Schiffer, M., Huber, R. & Palm, W. (1975) *Biochemistry* **14**, 4943–4952
- Eulitz, M. (1974) *Eur. J. Biochem.* **50**, 49–69
- Favaloro, J., Treiman, R. & Kamen, R. (1980) *Methods Enzymol.* **65**, 718–749
- Furey, W., Wang, B. C., Yoo, C. S. & Sax, M. (1983) *J. Mol. Biol.* **167**, 661–692
- Gorick, B. D., Thompson, K. M., Melamed, M. D. & Hughes-Jones, N. C. (1988) *Vox Sang.* **55**, 165–177
- Gubler, U. & Hoffman, B. J. (1983) *Gene* **25**, 263–269
- Hanahan, D. (1983) *J. Mol. Biol.* **166**, 557–580
- Hughes-Jones, N. C., Gardner, B. & Telford, R. (1964) *Immunology* **7**, 72–81
- Ichihara, Y., Abe, M., Yasin, H., Matsuoka, H. & Kurosama, Y. (1988) *Eur. J. Immunol.* **18**, 649–652
- Kabat, E. A., Wu, T. T., Reid-Miller, M., Perry, H. M. & Gottesman, K. S. (1987) *Sequences of Proteins of Immunological Interest*, 4th edn., pp. 41–235, U.S. Department of Health and Human Services, National Institutes of Health, Bethesda
- Kishimoto, T., Okajima, H., Okumoto, T. & Taniguchi, M. (1989) *Nucleic Acids Res.* **17**, 4385
- Lascombe, M.-B., Alzari, P. M., Boulot, G., Saludjian, P., Tougaard, P., Berek, C., Haba, S., Rosen, E. M., Nisonoff, A. & Poljak, R. J. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 607–611
- Lee, K. H., Matsuda, F., Kinashi, T., Kodaira, M. & Honjo, T. (1987) *J. Mol. Biol.* **195**, 761–768
- Lomas, C., Tippet, P., Thompson, K. M., Melamed, M. D. & Hughes-Jones, N. C. (1989) *Vox Sang.* **57**, 261–264
- Marquart, M., Deisenhofer, J., Huber, R. & Palm, W. (1980) *J. Mol. Biol.* **141**, 369–391
- McGregor, M. J., Suhail, A. M. & Sternberg, M. J. E. (1987) *J. Mol. Biol.* **198**, 295–310

- Melamed, M. D., Thompson, K. M., Gibson, T. & Hughes-Jones, N. C. (1987) *J. Immunol. Methods* **104**, 245–251
- Moore, S., Woodrow, C. F. & McClelland, D. B. L. (1982) *Nature (London)* **295**, 529–530
- Mudgett-Hunter, M., Anderson, W., Haber, E. & Margolies, M. N. (1985) *Mol. Immunol.* **22**, 477–488
- Padlan, E. A. & Kabat, E. A. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6885–6889
- Padlan, E. A., Silverton, E. N., Sheriff, S., Cohen, G. H. & Smith-Gill, S. J. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 5938–5942
- Ponder, J. W. & Richards, F. M. (1987) *J. Mol. Biol.* **193**, 775–791
- Ravetch, J. V., Siebelist, U., Korsmeyer, S., Waldmann, T. & Leder, P. (1981) *Cell* **27**, 583–591
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Sanz, I. & Capra, J. D. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 1085–1089
- Saul, F., Amzel, L. M. & Poljak, R. J. (1978) *J. Biol. Chem.* **253**, 585–597
- Segal, D., Padlan, E., Cohen, G., Rudikoff, S., Potter, M. & Davies, D. R. (1974) *Proc. Natl. Acad. Sci. U.S.A.* **71**, 4298–4302
- Sheriff, S., Silverton, E. W., Padlam, E. A., Cohen, G. H., Smith-Gill, S. J., Finzel, B. C. & Davies, D. R. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8075–8079
- Sibanda, B. L., Blundell, T. L. & Thornton, J. M. (1989) *J. Mol. Biol.* **206**, 759–777
- Siebenlist, U., Ravetch, J. V., Korsmeyer, S., Waldman, T. & Leder, P. (1981) *Nature (London)* **294**, 631–635
- Suh, S. W., Bhat, T. N., Navia, M. A., Cohen, G. H., Rao, D. N., Rudikoff, S. & Davies, D. R. (1986) *Proteins Struct. Funct. Genet.* **1**, 74–80
- Taussig, M. J. (1988) *Immunology Suppl.* **1**, 7–16
- Thompson, K. M., Hough, D. W., Maddison, P. J., Melamed, M. D. & Hughes-Jones, N. C. (1986) *J. Immunol. Methods* **94**, 7–12
- Udey, J. A. & Blomberg, B. (1987) *Immunogenetics* **25**, 63–70
- Wilmot, C. M. & Thornton, J. M. (1988) *J. Mol. Biol.* **203**, 221–232
- Wilson, P. W., Rogers, J., Harding, M., Pohl, V., Patlyn, G. & Lawson, D. E. M. (1988) *J. Mol. Biol.* **200**, 615–625
- Yamasaki, N., Komori, S. & Watanabe, T. (1987) *Mol. Immunol.* **24**, 981–985

Received 12 October 1989/20 November 1989; accepted 24 November 1989