Acta Cryst. (1996). D**52**, 1107–1113

Structure of the Fab Fragment of a Monoclonal Antibody Specific for Carcinoembryonic Antigen

M. J. BANFIELD,[a] D. J. KING,[b] A. MOUNTAIN[b] AND R. L. BRADY[a]*

[a]*Department of Biochemistry and Molecular Recognition Centre, University of Bristol, University Walk, Bristol BS8 1TD, England, and* [b]*Celltech Therapeutics Ltd, 216 Bath Road, Slough SL1 4EN, England*

(*Received 13 March 1996; accepted 5 June 1996*)

Abstract

The three-dimensional structure of the Fab fragment of the murine monoclonal antibody A5B7, which is specific for carcinoembryonic antigen (CEA) a protein expressed on carcinoma cell surfaces, has been determined. The structure was solved by molecular replacement and has been refined to an *R* factor for 21.2% (all data 8–2.1 Å). The conformation of the hypervariable loops, which form the antigen binding site, are consistent with canonical loop predictions. Hypervariable loop H3 is unusual in surface exposing many hydrophobic groups at the expense of burying an aspartic acid in the protein core. Other regions of the structure that influence the conformation of the binding site are identified. This structure forms a basis for analysing the effects of amino-acid substitutions in both hypervariable and framework regions in engineering studies of the A5B7 antibody.

1. Introduction

Monoclonal antibodies form one of the best studied examples of molecular recognition, demonstrating the fine specificity that can be achieved through subtle variations of protein surface. Additionally, as antigen binding sites are generated by combining a limited repertoire of hypervariable loop conformations on a highly conserved structural framework, antibodies have been the subject of extensive protein modelling and engineering studies (Adair, 1992). Many of these studies have aimed to extend the therapeutic uses of antibodies with demonstrated medical applications arising from their specificity for marker proteins associated with a particular disease. When introduced into a foreign host, antibodies themselves elicit an immune response leading to subsequent clearance of the antibody from the circulation, reducing their efficacy as therapeutics. To address this problem one approach has been to transfer the antigen binding site of the rodent antibody onto the framework of a human antibody, a process often referred to as 'humanization'. As well as transfer of the antigen-binding hypervariable loops, these engineering studies frequently require alteration of surrounding framework residues that interact with and affect the conformation of amino acids in the binding site (Adair, Athwal & Emtage, 1991; Foote & Winter, 1992). These subtle interactions are difficult to predict, and although success is often achieved through trial and error, a detailed three-dimensional structure of the donor antibody is invaluable in guiding these alterations.

A5B7 is an IgG_1-kappa class monoclonal antibody that binds the tumour associated glycoprotein CEA (carcinoembryonic antigen), and has potential application in the treatment of colorectal cancer (Lane *et al.*, 1994). Although precise details of the epitope of A5B7 are unknown, as CEA is heavily glycosylated it is possible that the epitope comprises both peptide and carbohydrate moieties. A5B7 has been the subject of extensive CDR grafting experiments aimed at producing improved therapeutics (Adair, Bodmer, Mountain & Owens, 1992). To assess the likely impact of grafting on the antigen binding site, we have crystallized and determined the structure of the murine form of the Fab fragment of A5B7. We describe the elucidation and refinement of this crystal structure to 2.1 Å resolution, and discuss the conformation of the binding site and implications for humanization.

2. Experimental

2.1. *Preparation of Fab*

A5B7 was purified from hybridoma cell culture as described in King *et al.* (1992). Fab fragments of A5B7 were generated by digestion with immobilized papain (Pierce), and separated from Fc by ion-exchange fast protein liquid chromatography on a mono-Q column equilibrated with 20 m*M* Tris, pH 8.2, and eluted with a 0–250 m*M* NaCl gradient. This produced a number of peaks differing in isoelectric point but all were confirmed to be Fab by SDS–PAGE analysis. The largest single peak of Fab was used for crystallization.

2.2. *Crystallization and data collection*

Crystal grew reproducibly by the hanging-drop method of vapour diffraction from equal volumes of a

ISSN 0907-4449

Table 1. *Summary of X-ray diffraction data for A5B7 Fab*

Resolution (Å)	Completeness (%)	$I/\sigma(I)$	R_{sym} (%)
20.00–3.42	100	29.1	3.9
3.42–2.71	99.9	25.4	4.9
2.71–2.37	99.5	20.3	5.9
2.37–2.15	98.8	15.4	7.0
2.15–2.00	98.0	9.7	11.5
Overall	99.2	24.1	4.7

Total no. of measured reflections = 95102
Total no. of unique reflections = 27439
Overall redundancy = 3.5

10 mg ml^{-1} protein solution (buffered at pH 6 with 20 m*M* Bis–Tris) and a precipitant solution containing 17–21% PEG 4000 buffered at pH 4 with 50 m*M* sodium acetate. For data collection crystals were harvested into a solution containing buffer with 25% PEG 4000 and 20% glycerol as a cryo-protectant, then mounted in nylon loops and flash-frozen in a cold stream at 100 K. The crystals belong to space group $P2_1$ with unit-cell dimensions of $a = 38.39$, $b = 79.68$, $c = 68.52$ Å and $\beta = 104.91$. The volume of the unit cell is consistent with one Fab molecule in the asymmetric unit. A full data set (99.2% complete to 2.0 Å) was collected at the Daresbury synchrotron radiation source (station 7.2, wavelength = 1.488 Å, with an 18 cm MAR image plate) from a single cryo-frozen crystal (100 K). The data were processed using *DENZO* and scaled with *SCALEPACK* (Otwinowski, 1993) and are summarized in Table 1.

2.3. *Structure solution*

The crystal structure was solved by molecular replacement using the coordinates of the Fab Hy-Hel-5 (Sheriff *et al.*, 1987) as a search model. Table 2 shows a comparison of the sequence of the hypervariable loops of both Hy-Hel-5 and A5B7. Two search models were generated from the HyHel-5 coordinates: one corresponding to the variable domain of the Fab (V_L:V_H), the other to the constant domain (C_L:C_H1). The hypervariable loops where included in the variable domain search model, but were later omitted during model building to remove bias. Each model was used in separate rotation searches within *AMoRe* (Navaza, 1994) using data in the range 10–3 Å and a sphere radius of 18 Å. Each search produced a clear peak.

Variable domain; correlation coefficient = 20.9 (next highest peak = 12.3); constant domain; correlation coefficient = 22.7 (next highest peak = 12.6).

The models were rotated by the corresponding angles and used in a number of translation searches. Initially each model was used separately in the translation searches, then the variable domain was translated with the independent constant domain solution fixed at the optimal translation vector, the similarly the constant

Table 2. *Sequence comparison of the hypervariable loops of A5B7 wth Hy-Hel-5*

Hypervariable loop (Kabat numbering)	Hy-Hel-5 sequence	A5B7 sequence
L1		
24	Ser	Arg
25	Ala	Ala
26	Ser	Ser
27	Ser	Ser
29	Ser	Ser
30	Val	Val
31	Asn	Thr
32	Tyr	Tyr
33	Met	Ile
34	Tyr	His
L2		
50	Asp	Ala
51	Thr	Thr
52	Ser	Ser
53	Lys	Asn
54	Leu	Leu
55	Ala	Ala
56	Ser	Ser
L3		
89	Gln	Gln
90	Gln	His
91	Trp	Trp
92	Gly	Ser
93	Arg	Ser
94	Asn	Lys
95	Pro	Pro
96	—	Pro
97	Thr	Thr
H1		
31	Asp	Asp
32	Tyr	Tyr
33	Trp	Tyr
34	Ile	Met
35	Glu	Asn
H2		
50	Glu	Phe
51	Ile	Ile
52	Leu	Gly
52a	Pro	Asn
52b	—	Lys
52c	—	Ala
53	Gly	Asn
54	Ser	Gly
55	Gly	Tyr
56	Ser	Thr
57	Thr	Thr
58	Asn	Glu
59	Tyr	Tyr
60	His	Ser
61	Glu	Ala
62	Arg	Ser
63	Phe	Val
64	Lys	Lys
65	Gly	Gly
H3		
95	Gly	Asp
96	Asn	Arg
97	Tyr	Gly
98	Asp	Leu
99	—	Arg
100	Phe	Phe
100j	—	Tyr
100k	—	Phe
101	Asp	Asp
102	Gly	Tyr

Table 3. *Positions of constant and variable domains*

	Rotation Eulerian angles (°)			Translation Fraction of unit-cell length		
Constant	199.00	45.16	66.58	0.4286	0.0000	0.3854
Variable	154.77	57.05	96.56	0.7702	0.4681	0.1305

domain was translated with the independent Fv solution fixed. Because of the arbitrary axis on y, models that were fixed were assigned to the origin in this direction. Fixing the constant-domain model with subsequent search for the variable domain gave the best solution as judged by highest correlation coefficient (36.3) with lowest R factor (50.3%, 10–3 Å data), compared with the next best solution (22.4 and 54.6%, respectively). With this solution the variable domain was correctly positioned to allow continuity of the H and L chains into the constant domain, and on examining the solution with computer graphics the crystal packing appeared reasonable. Positions of the two domains were defined as in Table 3.

An initial model for A5B7 Fab was generated by mutation of the Hy-Hel-5 sequence where this differed from A5B7 using *O* (Jones, Zou, Cowan & Kjeldgaard, 1991).

2.4. *Crystallographic refinement*

Elements of the initial model were refined as rigid bodies in *X-PLOR* (Brünger, 1992) by separating the model into four protein chains, corresponding to the conventional V_L, V_H, C_L, C_H1 domains of the antibody. This lowered the crystallographic R factor (15–2.0 Å, $F > 2\sigma F$) from 42 to 31.5% and was followed by a round of simulated-annealing refinement using the standard protocol which further lowered the R factor to 31%. At this stage the model was manually adjusted using *O*. Many clearly ordered solvent molecules were observed in the initial maps and water molecules were introduced judiciously at sites of appropriate geometry corresponding to positive peaks ($> 4\sigma$) in a $F_o - F_c$ difference map. A second round of simulated annealing further reduced the R factor to 28.4%. Two subsequent rounds of positional refinement and simulated annealing in *X-PLOR*, accompanied by manual rebuilds, lowered the R factor to 25.9%. During manual rebuilding a combination of $F_o - F_c$, $2F_o - F_c$, $3F_o - 2F_c$ and OMIT maps were used, OMIT maps were found to be useful in rebuilding poor regions of electron density, and were also used for positioning the hypervariable loops.

At this stage of refinement further water molecules were built into $> 3\sigma$ peaks in $F_o - F_c$ difference maps, although at all stages any water molecules that appeared positionally unstable or refined to temperature factors greater than 60 Å^2 were removed.

Further iterative cycles of simulated-annealing refinement with *X-PLOR* and positional refinement with *PROLSQ* (Collaborative Computational Project, Number 4, 1994), lowered the R factor to a final value of 21.2% (for all data in the range 8–2.1 Å). Statistics of the final model are summarized in Table 4.

Table 4. *Refinement statistics for final model of A5B7 Fab*

(*a*) Variation of crystallographic R factor with resolution

Resolution range (Å)	R factor* (%)
8.00–4.04	17.3
4.04–3.28	15.9
3.28–2.63	20.0
2.63–2.45	22.4
2.45–2.31	25.0
2.31–2.19	28.9
2.19–2.10	30.9
Overall	21.2

(*b*) Stereochemical ideality and B-factor analysis of final model

R.m s. bond lengths (Å)	0.015
R.m.s. angles (°)	1.9
Average B factor (Å^2) main-chain atoms	22.15 R.m.s. deviation = 1.78
Average B factor (Å^2) side-chain atoms	23.34 R.m.s. deviation = 2.67
Average B factor (Å^2) solvent atoms	34.13
Total number of non-H atoms	3327
Identified solvent sites	400

* Crystallographic R factors were calculated using all data in the resolution shells listed.

3. Structure description and discussion

The current model comprises all 213 residues of the light (L) chain, and 222 residues of the heavy (H) chain. Four residues at the C terminus (H211–214) which form part of the hinge region are well ordered in the structure and are included in the final model. The model also includes 400 ordered water molecules, consistent with collection of X-ray data at 100 K. A Cα trace of the protein overlaid on the Cα trace of the Hy-Hel-5 Fab used as search model can be seen in Fig. 1, illustrating the change in elbow angle. The conventional elbow angle relating the dyads of the variable and constant domains as calculated from the final model is 143.4°.

In general the quality of the electron-density map is very good, and the six hypervariable loops can all be traced with high confidence. Fig. 2 shows electron density of hypervariable loop H3 as an example. There is only weak electron density for residues L155–159 and H130–H134. All these residues are on loops in the constant domain that pack adjacent to one another in the crystal, and may represent a localized region of disorder. A Ramachandran plot (Ramachandran & Sasisekharan, 1968) shows only two outliers: L50Thr

and H134Ser (Fig. 3). The electron density for L50 is however excellent, and the residue appears to be maintained in an unfavourable conformation by a network of hydrogen bonds. H134 is in one of the disordered regions referred to above.

Comparisons of the observed conformations of the hypervariable loops with those described by the canonical structure approach of Chothia *et al.* (1989), show the first hypervariable loop of the L chain (L1) adopts at type 1 conformation primarily determined by the absence of a residue at position 31, and the presence of a valine at position 29. The second hypervariable loop (L2) is similar to all other L2 structures observed and L3 forms a type 1 conformation. In the H chain, the first hypervariable loop (H1) forms a type 1 structure, and H2 is consistent with type 4, determined by the presence of residues 52a, 52b and 52c, and an arginine at 71. This arginine forms hydrogen bonds with main-chain carbonyl groups of residues in both the H1 (residues H29 and H32) and H2 (residue H52) loops. The H2 loop forms a prominent protrusion from the surface of the binding site and is very likely to be a major determinant in antigen binding.

The H3 hypervariable loop, for which no canonical structures have been described, is relatively long

Fig. 1. Crystal structure of A5B7. Stereo pair depicting the Cα trace of A5B7 (bold) overlaid on the Cα trace for the murine antibody Hy-Hel-5 (thin lines) (Sheriff *et al.*, 1987). Variable domain top, constant (C_H1) domain bottom. The characteristic immunoglobulin fold is clearly evident.

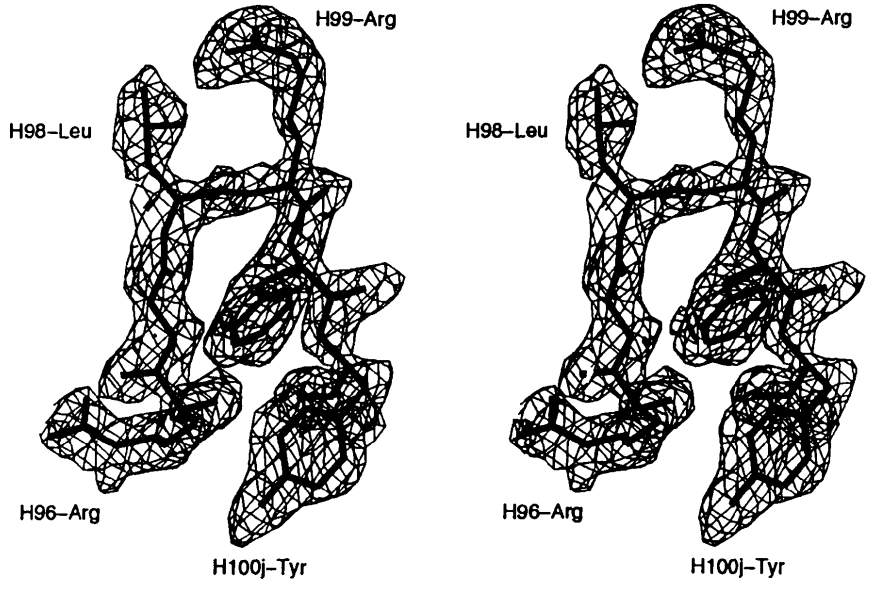

Fig. 2. Electron-density map. Stereo pair, showing a $2F_o - F_c$ electron-density map showing a portion of the H3 hypervariable loop, residues H96–H100j. The map is calculated with phases from the final model, and contoured at 1σ. Water molecules have been omitted for clarity.

compared with other antibodies. Unlike many antibodies that contain marked clefts within their antigen binding sites, the H3 loop of A5B7 largely fills the binding site leaving only a shallow depression (Figs. 4 and 5). This loop covers and contributes (H100kPhe, H103Trp) to the buried hydrophobic interface between the H and L chains. Interestingly, the acidic side chain of H95Asp is also buried within this region (Fig. 4) where it makes hydrogen-bond contacts with the main-chain H3 loop residues H100 and H100j and with the amine N atom of the indole ring of L90Trp (L3 loop). In this environment the side chain is presumably protonated, consistent with growth of the crystals at pH 4. It is interesting to speculate that these direct contacts of H3 loop residues with the L/H chain interface could provide a mechanism for altering the association of the V_L and V_H domains in concert with changes in the H3 loop structure on antigen binding. Significant changes in H3 loop conformations have previously been observed to accompany antigen binding (Schulze-Garmen, Rini & Wilson, 1993).

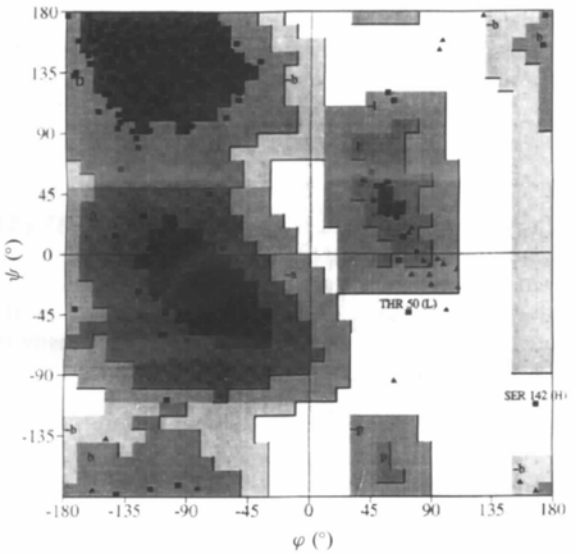

Fig. 3. Ramachandran plot (Ramachandran & Sasisekharan, 1968) of coordinates from the final model generated using the program *PROCHECK* as implemented in the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). The two outliers are labelled.

The positioning of the H3 loop in A5B7 could be further stabilized by a potential salt bridge between the side chain of H99Arg, at the apex of the H3 loop, and the spatially adjacent H58Glu (base of H2 loop). In the crystal structure the arginine participates in a triad of charged hydrogen bonds with the glutamic acid and the side chain of H214Asp from a symmetry-related molecule. This network is presumably encouraged by the pH at which the crystals are grown (pH 4) which is close to the expected pK_a values for both the glutamic and aspartic acid side chains. At neutral pH and in solution a direct ionic interaction between the glutamic acid and arginine side chains might be expected.

By anchoring the H3 loop across the 'floor' of the binding site H99Arg facilitates the surface exposure of a number of hydrophobic H3 loop residues including H98Leu, H100Phe, H100jTyr and H102Tyr, each of which could potentially contribute to the antigen–antibody interface. This unusual combination of

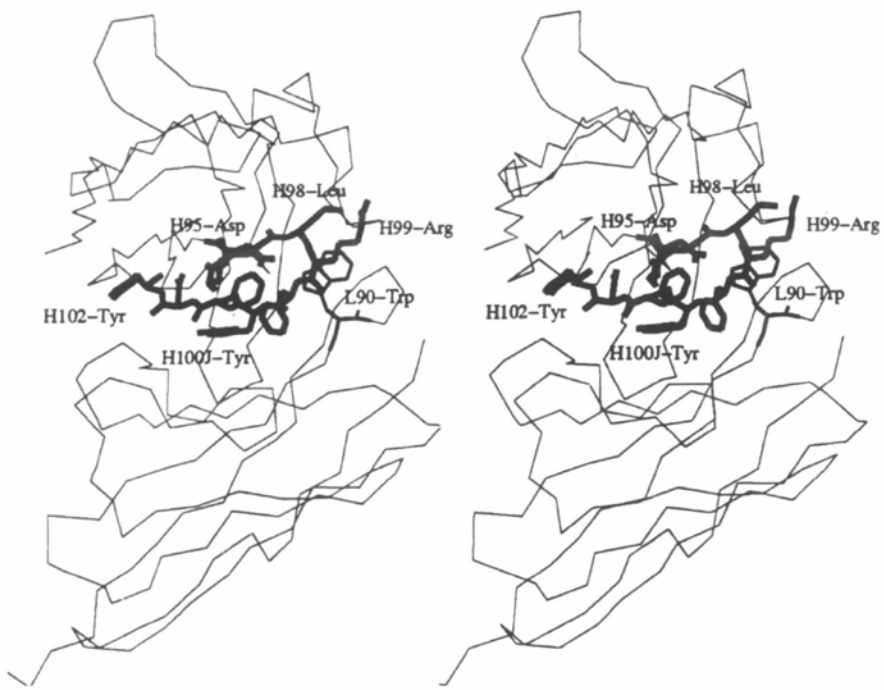

Fig. 4. Conformation of the H3 loop. Stereo pair looking down on the binding site showing the variable domain (H chain upper, L chain lower) as a Cα trace. All atoms of H3 loop residues are included and shown in bold, as is L90Trp from the L3 loop. Surface-exposed residues from the H3 loop include H98Leu, H100Phe, H100jTyr and H102Tyr. Note H95Asp is buried in the hydrophobic interface between the H and L chains, and forms a hydrogen bond to the amine N atom in the indole ring of L90Trp. The H3 loop lies across the floor of the binding site and contributes to the buried H/L-chain interface.

exposed hydrophobic groups adjacent to a buried charged side chain illustrates the complexity of predicting H3 loop conformations. Additionally, a number of charged residues in the H3 loop (H96Arg, H99Arg and H101Asp) are prominently exposed in the binding site. In total seven charged residues contribute to the surface of the binding site, namely L94Lys H31Asp, H58Glu, H64Lys, H96Arg and H101Asp, and are all potentially available for antigen binding. As the antigen may comprise carbohydrate in addition to protein, a high proportion of charged residues might be expected.

Previous studies have identified regions outside of the binding site that are important for antibody engineering (Adair *et al.*, 1991). Chothia *et al.* (1989) noted that in many antibodies the hydroxyl group of L71Tyr can form multiple hydrogen bonds with residues in the L1 loop. In A5B7 these contacts are also observed, the tyrosine contacting main-chain groups of L30, L31 and L68. The H2 loop sequence remains highly variable through to H65, and Eigenbrot *et al.* (1994) have illustrated that the COOH-terminal part of this sequence may contribute to antigen affinity, although their study was inconclusive as to whether its significance is due to direct antigen interaction or contributes a conformational effect. In the structure of A5B7 we observe a distinctive salt bridge immediately after this region (H66Arg to H86Asp), and if the conformation of this region is important it may need to be maintained in engineering studies.

Because of its close proximity to the H1 loop, it has often been found necessary to alter D/E loop region of the heavy chain in humanization experiments (Brady *et al.*, 1992). In the crystal structure of A5B7 much of this loop is involved in crystal contacts and is turned away from the H1 loop, making it difficult to assess the importance of this region for A5B7. However, as noted previously one residue in the loop, H71Arg, instigates a charged hydrogen-bond network with main-chain carbonyl groups of H29 and H32 (H1 loop) and H52 (H2 loop) and may be important in maintaining the conformation of the binding site (Fig. 5).

Finally, the NH_2 termini of both chains lie close to the binding site. Although partially disordered in the crystals, the amino terminus of the L-chain hydrogen bonds to the side chain of L92Ser and may influence the L3 loop conformation. In the H chain no side-chain contacts are observed to the adjacent H1 loop region, but the terminal glutamate is distorted by crystal contacts and its contribution to the binding site conformation cannot be assessed.

4. Conclusions

The crystal structure of the Fab fragment of A5B7 adds to the expanding repertoire of Fab structures contributing to our understanding of antigen recognition and binding. This structure further confirms many of the patterns noted previously, including the propensity of the hypervariable loops to form a limited number of canonical conformations and that these conformations can be influenced by neighbouring framework residues. The unusual H3 loop conformation, which favours solvent exposure of hydrophobic side chains at the cost of burying an aspartic acid, illustrates the complexity of reliably predicting structures for this region. Although interpretation of Fab crystal structures can be complicated by crystal contacts and the absence of bound antigen, this structure of A5B7 provides a detailed description of the binding site and identifies features outside of this region that may influence binding. The structure provides a useful guide for engineering and a basis from which to interpret the success of grafting experiments. We have recently obtained crystals of the Fab fragment of an engineered human construct of A5B7 and a structural analysis is under way. This should permit an interesting assessment of some of the

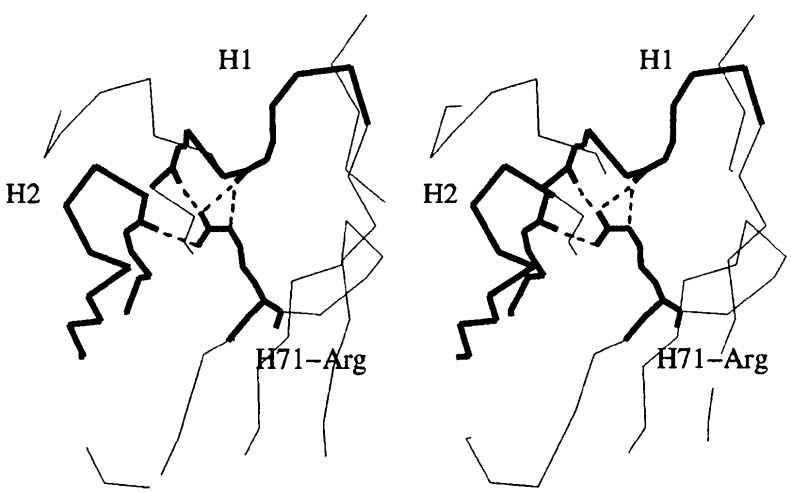

Fig. 5. Interaction of H71Arg with the H1 and H2 hypervariable loop regions. Stereo pair depicting the location of the H71Arg side chain and its interaction with residues in the antigen binding site. H71Arg forms a network of hydrogen bonds with main-chain atoms from H29 and H32 (H1 loop, bold lines) and H52 (H2 loop, bold lines).

features of the binding site noted above and their importance in maintaining binding activity.

Throughout this paper amino acids are numbered according to the Kabat numbering scheme (Kabat, Wu, Reid-Miller, Perry & Gottesman, 1987). Coordinates and structure factors have been deposited with the Brookhaven Protein Data Bank.* MJB is a recipient of a Medical Research Council (UK) studentship. Protein crystallography facilities at Bristol are supported by the BBSRC and the Wellcome Trust. We are grateful to the staff at Daresbury Laboratory for their assistance with synchrotron data collection.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1CLO, R1CLOSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: AD0018). At the request of the authors, the atomic coordinates and structure factors will remain privileged until 1 May 1997.

References

Adair, J. R. (1992). *Immunol. Rev.* **130**, 5–40.

Adair, J. R., Athwal, D. S. & Emtage, J. S. (1991). Patent WO91/09967.

Adair, J. R., Bodmer, M. W., Mountain, A. & Owens, R. J. (1992). Patent WO92/01059.

Brady, R. L., Edwards, D. J., Hubbard, R. E., Jiang, J.-S. Lange, G., Roberts, S. M., Todd, R. J., Adair, J. R., Emtage, J. S., King, D. J. & Low, D. C. (1992). *J. Mol. Biol.* **227**, 253–264.

Brünger, A. T. (1992). *X-PLOR Manual Version 3.1.* Yale University, Connecticut, USA.

Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G., Sheriff, S., Padlan, E. A., Davies, D., Tulip, W. R., Colman, P. M., Spinelli, S., Alzari, P. M. & Poljak, R. J. (1989). *Nature (London)*, **342**, 877–883.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Eigenbrot, C., Gonzalez, T., Mayeda, J., Carter, P., Werther, W., Hotaling, T., Fox, J. & Kessler, J. (1994). *Proteins Struct. Funct. Genet.* **18**, 49–62.

Foote, J. & Winter, G. (1992). *J. Mol. Biol.* **224**, 487–499.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kabat, E. A., Wu, T. T., Reid-Miller, M., Perry, H. M. & Gotteman, K. S. (1987). *Sequences of Proteins of Immunological Interest*, 4th ed. Bethesda, MD: National Institutes of Health.

King, D. J., Adair, J. R., Angel, S., Lowe, D. C., Proudfoot, K. A., Lloyd, J. C., Bodmer, M. W. & Yavranta, G. T. (1992). *Biochem. J.* **281**, 317–323.

Lane, D. M., Eagle, K. F., Begent, R. H. J., Hope-Stone, L. D., Green, A. J., Casey, J. L., Keep. P. A., Kelly, A. M. B., Lederman, J. A., Glaser, M. G. & Hilson, A. J. W. (1994). *Br. J. Cancer*, **70**, 521–525.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Otwinowski, Z. (1993). *DENZO* and *SCALEPACK*, *Data processing and scaling programs*, Yale University, New Haven, Connecticut, USA.

Ramachandran, N. G. & Sasisekharan, V. (1968). *Adv. Protein Chem.* **23**, 283–437.

Schulze-Garmen, U., Rini, J. M. & Wilson, I. A. (1993). *J. Mol. Biol.* **234**, 1098–1118.

Sheriff, S., Silveron, E. W., Padlan, E. A., Cohen, G. H., Smith-Gill, S. J., Finzel, B. C. & Davies, D. R. (1987). *Proc. Natl Acad. Sci USA*, **84**, 8075–8079.

Winter, G. & Milstein, C. (1991) *Nature (London)*, **349**, 293–299.