Journal of MOLECULAR BIOLOGY

EDITOR-IN-CHIEF: J. C. KENDREW

Volume 167, Number 3, 5 July 1983

ACADEMIC PRESS
London New York Toronto Sydney San Francisco
A Subsidiary of Harcourt Brace Jovanovich, Publishers
JMOBAK 167 (3) 523–756
ISSN 0022–2836

JOURNAL OF MOLECULAR BIOLOGY

Published three times a month on the 5th, 15th and 25th at 24–28 Oval Road, London NW1 7DX, England by Academic Press Inc. (London) Limited.

1983: Volumes 163–171, 36 Issues. Inland, £420.00 including postage and packing; abroad, $1100.00 including postage and packing. Index and Cumulative Contents of Volumes 1 to 20, 21 to 40, 41 to 60: prices on application.

Subscription orders should be sent to Academic Press Inc. (London) Limited, 24–28 Oval Road, London NW1 7DX, with the exception of those originating in the U.S.A., Canada, Central America and South America; these should be sent to Academic Press Inc., 111 Fifth Avenue, New York, New York 10003. Second class postage paid at Jamaica, N.Y., U.S.A. Air freight and mailing in the U.S.A. by Publications Expediting Inc., 200 Meacham Avenue, Elmont, N.Y. 11003, U.S.A. Send notices of change of address to the office of the Publishers at least 6–8 weeks in advance. Please include both old and new addresses. Postmaster, send changes of address to *Journal of Molecular Biology*, 111 Fifth Avenue, New York, New York 10003.

J. Mol. Biol. (1983) **167**, 661–692

Structure of a Novel Bence–Jones Protein (Rhe) Fragment at 1·6 Å Resolution

W. Furey Jr, B. C. Wang, C. S. Yoo and M. Sax

Biocrystallography Laboratory
Box 12055, V.A. Medical Center, Pittsburgh, Pa 15240, U.S.A.
and Department of Crystallography
University of Pittsburgh, Pittsburgh, Pa 15260, U.S.A.

(Received 13 October 1982, and in revised form 3 February 1983)

The crystal structure of Rhe, a λ-type Bence–Jones protein fragment, has been solved and refined to a resolution of 1·6 Å. A model fragment consisting of the complete variable domain and the first three residues of the constant domain yields a crystallographic residual R_F value of 0·149. The protein exists as a dimer both in solution and in the crystals. Although the "immunoglobulin fold" is generally preserved in the structure, there are significant differences in both the monomer conformation and in the mode of association of monomers into dimers, when compared to other known Bence–Jones proteins or Fab fragments. The variations in conformation within monomers are particularly significant as they involve *non*-hypervariable residues, which previously were believed to be part of a "structurally invariant" framework common to all immunoglobulin variable domains. The novel mode of dimerization is equally important, as it can result in combining site shapes and sizes unobtainable with the conventional mode of dimerization. A comparison of the structure with other variable domain dimers reveals further that the variations within monomers and between domains in the dimer are coupled. Some possible functional implications revealed by this coupling are greater variability, induced fitting of the combining site to better accommodate antigenic determinants, and a mechanism for relaying binding information from one end of the variable domain dimer to the other.

In addition to providing the most accurate atomic parameters for an immunoglobulin domain yet obtained, the high resolution and extensive refinement resulted in identification of several tightly bound water molecules in key structural positions. These water molecules may be regarded as integral components of the protein. Other water molecules appear to be required to stabilize the novel conformation.

1. Introduction

Rhe is a λ-type Bence–Jones protein fragment consisting of a complete variable domain and the first three residues of a constant domain. The protein exists solely as a non-covalent dimer both in solution and in the crystalline state. The crystals are orthorhombic, with space group $P2_12_12$ and cell constants $a = 54{\cdot}63(2)$ Å, $b = 52{\cdot}22(3)$ Å and $c = 42{\cdot}62(3)$ Å. The unit cell contains four monomers (114

0022–2836/83/190661–32 $03.00/0

residues each) and approximately 51% solvent by volume. The crystallization conditions have been reported (Wang & Sax, 1974), as has a preliminary analysis (α-carbon level) of the structure at 3·0 Å resolution (Wang *et al.*, 1979). In the 3 Å analysis, several aspects of the structure were described, the most significant being the discovery of a mode of association of variable domains into dimers different from that observed previously in either Bence–Jones proteins or Fab fragments. This novel mode of dimerization is correlated with a change in conformation *within* monomers, and does not represent merely a rotation of one domain relative to the other. Furthermore, the segment within each monomer that changed conformation consists solely of *non*-hypervariable residues. Previously, these residues (sequence numbers 36 to 50) were believed to be part of a "structurally invariant" framework common to all immunoglobulins.

In order to determine factors responsible for this novel conformation, it is necessary to extend the resolution beyond the 3 Å level and to complete the chemical sequencing of the protein. Details of the phase extension and refinement to 1·9 Å resolution were reported (Furey *et al.*, 1979), but no attempt was made to interpret the structure, since this was merely an interim step toward a detailed analysis at 1·6 Å resolution. The current high-resolution analysis confirms the 3 Å structure interpretation and indicates that, although the conformational change involves non-hypervariable residues, it is very likely caused by interactions with residues in the first and second hypervariable regions. The analysis also suggests a role for solvent in stabilizing the novel protein conformation. An alternative interpretation of the data suggests a plausible trigger mechanism for relaying information from the combining site end of variable domain dimers to the switch regions connecting them to constant domains.

2. Experimental Procedures

The X-ray data collection and reduction process has been described in detail (Furey *et al.*, 1979), so only aspects reflecting the quality of the data need be stated here. Rhe crystals diffract X-rays extremely well, so it was possible to collect all data to 1·6 Å resolution from 1 crystal with only a 15% decrease in standard reflection intensities. In addition, the moderate size of the unit cell enabled collection of diffractometer data by the θ-2θ scan method commonly used in small molecule crystallography. The overall quality of the data set is reflected in the fact that over 75% of all reflections to 1·6 Å resolution are considered "observed", even by the rather stringent $I/\sigma(I) > 3$ criterion. The percentage of observed reflections as a function of resolution is given in Table 1. Note that the data were collected at room temperature with a relatively low power (0·5 kW) X-ray source, hence it should be possible to collect even higher resolution data if low temperature techniques are applied and a more powerful (1·5 kW) X-ray tube is used.

Extending the procedure described in the 1·9 Å paper (Furey *et al.*, 1979), refinement was resumed after including the additional X-ray data to 1·6 Å resolution. The 1·9 Å refined co-ordinates ($R_F = 0{\cdot}28$, $R_F = (\Sigma||F_o|-|F_c||)/\Sigma|F_o|$) served as initial parameters. Since chemical sequence information was not available, electron density maps (either $2F_o - F_c$ or $F_o - F_c$) were examined frequently. Residues to be examined were always deleted from the phasing process before map calculation. Occasionally, errors in amino acid sequence or atomic positions were indicated by the maps. The errors were corrected by non-interactive computer graphics (Furey *et al.*, 1979), optical comparator techniques (Richards, 1968) and, in the last stages, interactive computer graphics. In all cases after sequence changes were implemented, restrained reciprocal space refinement (Hendrickson & Konnert, 1978) was

TABLE 1

Observed reflections as a function of resolution

d_{min} (Å)	Total	No. observed	% (in range)	% (cumulative)
3·0	2667	2582	96·8	96·8
2·4	5091	4781	90·7	93·9
1·9	10,073	8975	84·1	89·1
1·6	16,665	12,842	58·7	77·1

Observed if $I > 3\ \sigma(I)$

resumed. The computation was performed with the space group general array processor version (Furey *et al.*, 1982) of the Hendrickson–Konnert program.

The next step in the refinement process was to introduce individual isotropic thermal factors for each of the non-hydrogen atoms, and to include solvent molecules in the model. Despite the obvious presence of water molecules in some of the early maps, solvent atoms were not incorporated into the model until late in the refinement process ($R_F = 0{\cdot}23$) to avoid mistaking erroneously sequenced side-chain atoms for water molecules. Water oxygen positions were obtained from difference electron density maps by scanning through the largest peaks and determining the shortest distance between each peak and all atoms in the current model. If the shortest distance was between 2·3 and 3·4 Å and the model atom involved was capable of forming hydrogen bonds, a water oxygen was added at the peak position. The new model was then subjected to several cycles of least-squares refinement. The procedure was iterated several times and, after discarding water molecules that moved far from their original locations, resulted in the inclusion of 186 water oxygen atoms (102 partially occupied). The mean electron density at the water sites is 1·26 e/Å^3. The estimated error in the electron density function is ~0·20 e/Å^3 (Cruickshank, 1949), hence the water molecules should be reasonably well-determined. Thermal and occupancy factors for the water molecules ranged from 5 to 68 Å^2 and 0·30 to 1·00, respectively. Hydrogen atom contributions for hydrogen atoms in the protein were included in all structure factor calculations, but their positions were recomputed every few cycles rather than refined.

At this time, partial sequence information became available (W. Brown, personal communication) for comparison with the X-ray model. There were 11 discrepancies, all of which were resolved in the model by adding, deleting or changing the atomic type of, at most, 2 atoms/residue. The chemical sequence is now known except for distinctions within the pairs, Gln, Glu and Asn, Asp. Since these residues are isosteric and each can form hydrogen bonds, they cannot always be distinguished at 1·6 Å resolution by X-ray data. Therefore acid/amide distinctions in the model were based mostly on the sequences of other λ-type variable domains (Kabat *et al.*, 1977). In favorable cases, however, obvious hydrogen bonding partners and/or the chemical composition in short peptides enabled a choice to be made. The amino acid sequence in the final model is given in Table 2. When compared with the sequence deduced from the 3 Å map, significant discrepancies were revealed, although the original interpretation was reasonable (60% correct).

The final model contains 1019 non-hydrogen atoms, 833 from protein (all fully occupied), and the 186 water oxygen atoms. The model fits into the electron density extremely well, as indicated in Fig. 1. The mean isotropic thermal factor is 16·5 Å^2 when all atoms are included, and 12·5 Å^2 for protein atoms only. For solvent atoms, the mean B† is 34·2 Å^2. A histogram indicating the distribution of thermal factors for the protein is given in Fig. 2. The final crystallographic residual R_F value is 0·149 for the 12,763 observed reflections with d spacings ranging from 10 to 1·6 Å. The R factor is plotted as a function of resolution in

† B (isotropic thermal factor) $= 8\pi^2\bar{u}^2$, where $\bar{u}$ is the root-mean-square amplitude of atomic vibration.

TABLE 2

Amino acid sequence of Bence–Jones protein Rhe (V_L)

Residue No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Code	E	S	V	L	T	Q	P	P	S	A	S	G	T	P	G	Q	R	V	T	I
No.	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Code	S	C	T	G	S	A	T	D	I	G	S	N	S	V	I	W	Y	Q	Q	V
No.	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Code	P	G	K	A	P	K	L	L	I	Y	Y	N	D	L	L	P	S	G	V	S
No.	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
Code	D	R	F	S	A	S	K	S	G	T	S	A	S	L	A	I	S	G	L	E
No.	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
Code	S	E	D	E	A	D	Y	Y	C	A	A	W	N	D	S	L	D	E	P	G
No.	101	102	103	104	105	106	107	108	109	110	111	112	113	114						
Code	F	G	G	G	T	K	L	T	V	L	G	Q	P	K						

One-letter amino acid code: A, alanine; C, cysteine; D, aspartate; E, glutamate; F, phenylalanine; G, glycine; I, isoleucine; K, lysine; L, leucine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine; O, water.

Fig. 3. Throughout the refinement, the protein model was restrained with respect to stereochemistry, resulting in a final model with root-mean-square deviations from "ideal" bond distances of 0·024 Å. The root-mean-square deviation from planarity for the appropriate atomic groups (including peptide links) is 0·020 Å. The last refinement cycle resulted in a root-mean-square shift in atomic positions of 0·006 Å with a corresponding estimated standard deviation of 0·011 Å. The latter is most certainly an underestimate of the true error in atomic positions, which we estimate to be ~0·04 to 0·08 Å by the method of Luzzati (1952). Mean electron densities for the main-chain atoms, N, Cα, C and O are 2·89, 2·24, 2·44 and 3·17 e/Å^3, respectively.

3. The Molecular Structure

(a) *The monomeric subunit*

The Rhe monomer consists of nine strands, as is the case for all immunoglobulin variable domains of known structure. A stereoscopic drawing of the α-carbon structure is given in Figure 4. If we neglect the N-terminal residue, which is highly disordered and cannot be located accurately, the structure begins with a β-turn involving residues 2 to 5. The nine strands contain residues 6 to 12, 17 to 24, 33 to 39, 44 to 50, 54 to 57, 60 to 67, 72 to 79, 84 to 92 and 97 to 114. All strands are in an extended conformation; however, there is one short helix of roughly 1·5 turns (residues 25 to 32) connecting strands 2 and 3.

As seen in Figure 4, the nine strands are connected by ten turns, resulting in a pair of antiparallel β-sheets. The first, designated sheet A, contains strands 2, 6 and 7, while the second (sheet B) contains strands 3, 4, 8 and 9. The two sheets are also connected by a disulphide bridge between Cys22 (strand 2) and Cys89 (strand 8), and by a salt-bridge between Arg62 and Asp83. Only strands 1 and 5 are isolated from the sheets, although segments of strands 3 and 4 eventually break away from sheet B. This latter feature has important consequences with regard to the mode of domain–domain association upon dimerization. In previous

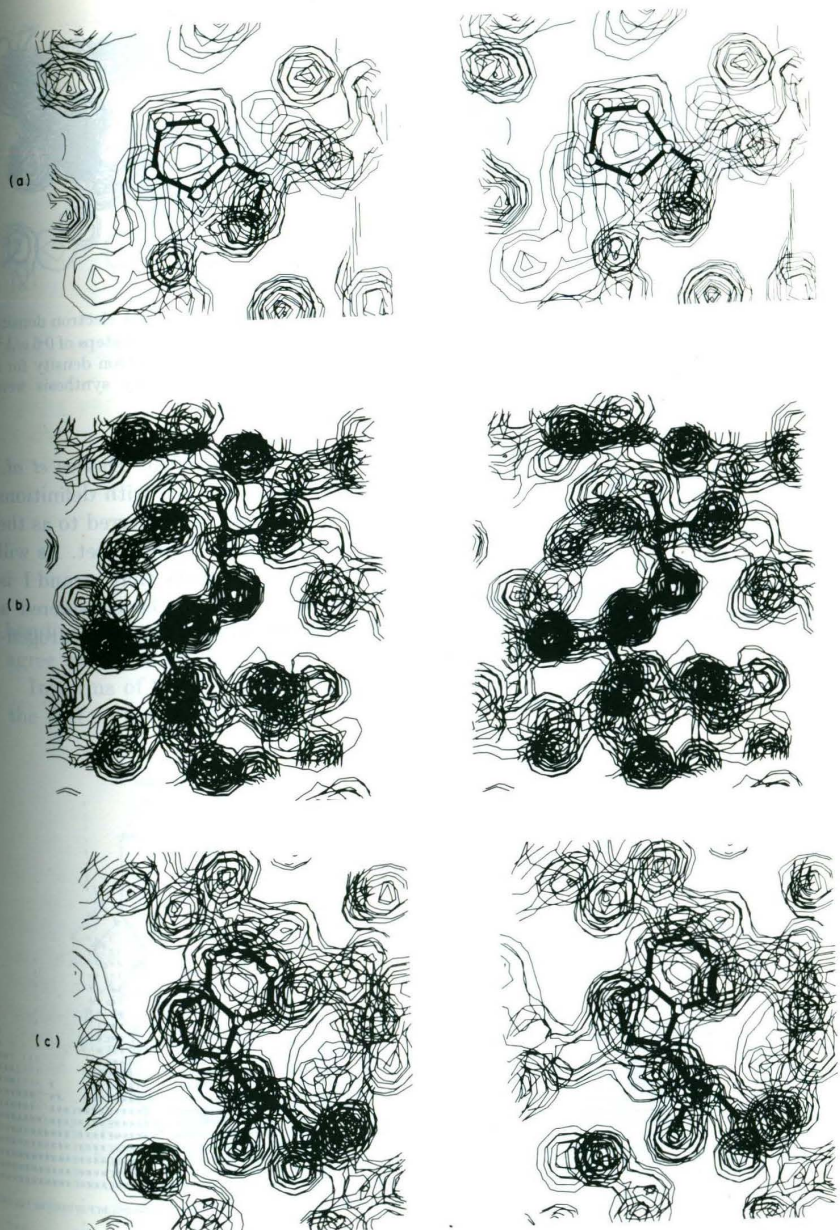


FIG. 1.

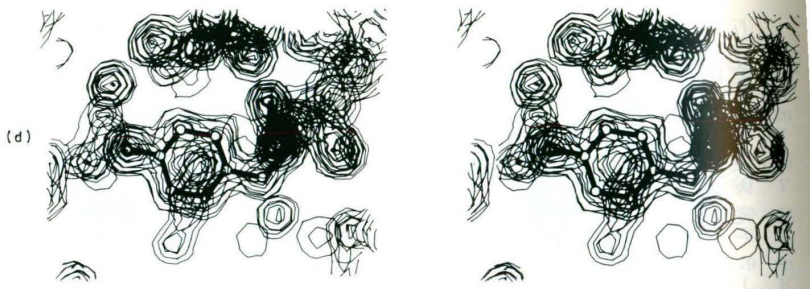


FIG. 1. Stereoscopic projections of some typical residues superimposed on the 1·6 Å electron density map illustrating the fit. The electron density contours begin at 0·6 e/Å^3 and increase in steps of 0·6 e/Å^3. Note that all rings, even prolines, have "holes" in their centers. Also note the electron density for a water molecule near the OH group in (d). Coefficients for the electron density synthesis were $F_o \exp(i\phi_c)$ with the illustrated atoms omitted from the phasing.

descriptions of the immunoglobulin fold (Amzel & Poljak, 1979; Davies *et al.*, 1975), both three and four-strand structures are mentioned, but with definitions opposite to those given above. Usually, strands 1, 2, 6 and 7 are referred to as the four-strand sheet with strands 3, 8 and 9 forming the three-strand sheet. As will be shown later, the hydrogen-bonding pattern in Rhe suggests that strand 1 is much more closely related to sheet B, and that it should not be considered a member of the "four-strand" sheet (A). Surprisingly, the corresponding hydrogen-

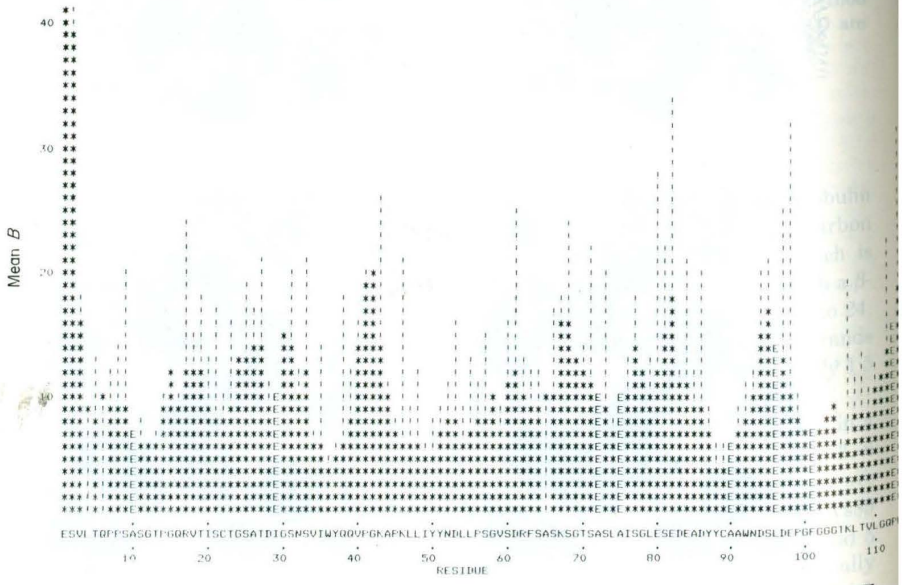


FIG. 2. Histogram of mean thermal factors (B) for each residue. (∗) Main chain; (!) side chain; E, main and side-chains equal.

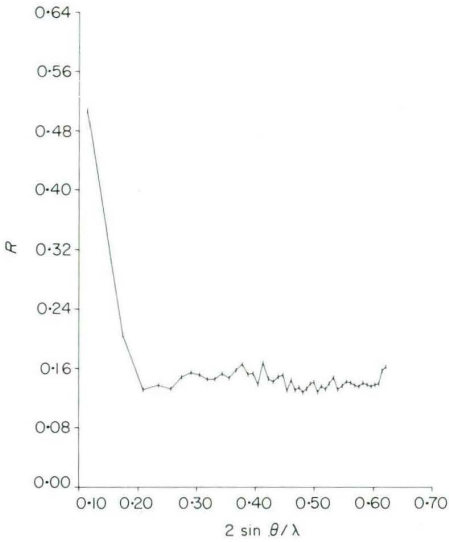


FIG. 3. Plot of R factor *versus* resolution.

bonding diagrams for other immunoglobulin domains (Amzel & Poljak, 1979) agree with that for Rhe, yet the sheet designations are different.

In terms of the variability of amino acid sequence, residues 23 to 35 represent the first hypervariable region, while the second and third hypervariable regions

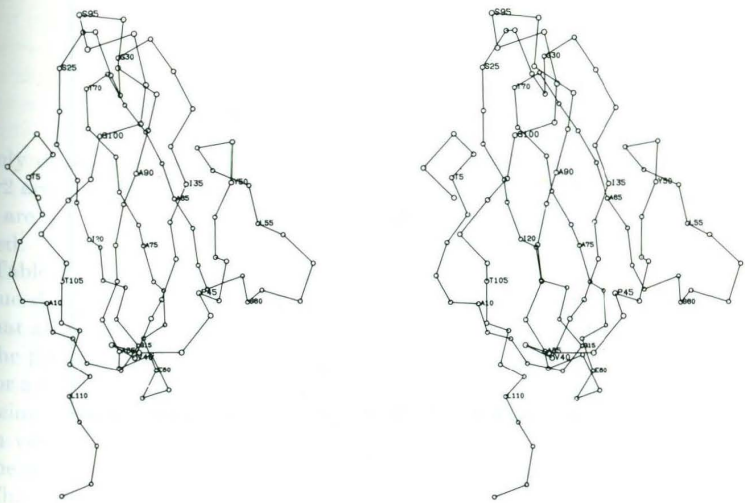


FIG. 4. Rhe monomer, α-carbon skeleton.

correspond to residues 51 to 57 and 90 to 100, respectively. All other residue positions usually show much less variation in sequence for λ-type V_L† domains.

(b) *Main chain torsion angles and turns*

In Figure 5 the distribution of ϕ, ψ angles is shown in the usual Ramachandran plot, with "allowed" regions marked according to Ramakrishnan & Ramachandran (1965). The most striking feature of the plot is the small number of non-glycine residues outside allowed regions, indicating a very stable conformation. There are only three violations, with extenuating circumstances explaining the occurrence of each. For example, the abnormal ψ value for Asp28 is readily explained when neighboring residues are also considered. From Figure 6(a), it is clear that any change in ψ would weaken hydrogen-bonding interactions with the main chain NH groups of Ser31 and Asn32. A change in ψ here would also affect the direction of the NH bond of Ile29. Since this group is already ideally oriented for hydrogen bond formation to the side-chain carboxyl group of Asp28, any change could only destabilize the system. There are even more compelling reasons for the unusual angles in Asn52. From Figure 6(b), it is clear that any change in ϕ would weaken the hydrogen-bond interactions between NH of Asn52 and O of Val34, and between main chain atoms O of Tyr51 and NH of Asp53. Likewise, any change in ψ would weaken the hydrogen bond between main chain O of Asn52 and water oxygen O118. A change in ψ would also weaken the NH—O hydrogen bond already mentioned between Asp53 and Tyr51.

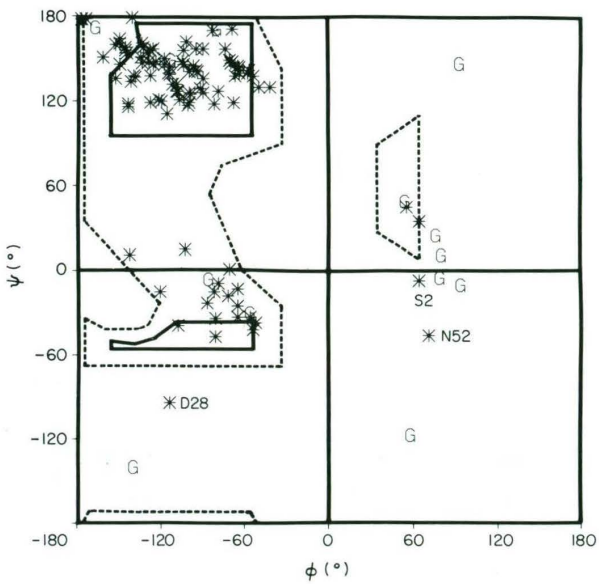


FIG. 5. The Ramachandran plot for Rhe. G indicates glycine residues.

† Abbreviation used: V_L, variable domain of light chain.

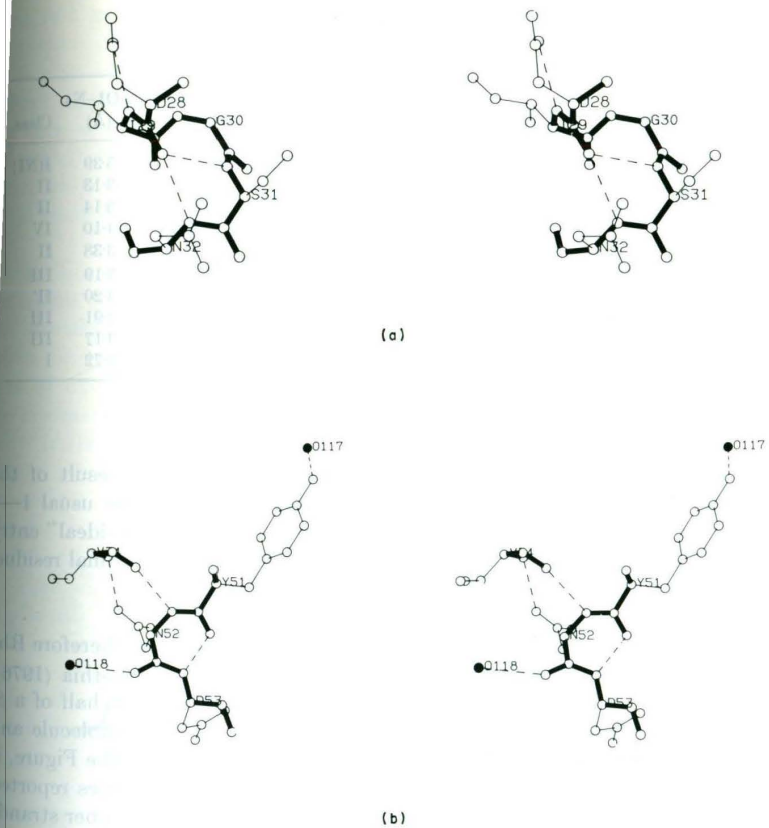


FIG. 6. Environments around (a) Asp28 and (b) Asn52.

The only other abnormal conformation occurs for Ser2. Since the torsion angles for Ser2 are likely to be adversely affected by the disorder in residue 1, and in any event are just slightly outside of an allowed region, no further comment is required.

In Table 3 the ten sharp turns in Rhe are classified according to the scheme introduced by Venkatachalam (1968), and extended by Lewis *et al.* (1973). The fact that all three turns of type II have Gly in position 3 is in excellent agreement with the findings of Venkatachalam (1968) and Crawford *et al.* (1973). They claim that for a 1—4 hydrogen bond to remain intact in a type II bend, position 3 must be glycine. In a more recent study on the frequency of occurrence of each amino acid in various positions of β-turns in 29 proteins, Chou & Fasman (1977) found that the most frequently occurring amino acid residues are Pro, Gly, Asn, Asp and Ser. The structure of Rhe is consistent with their findings in that of the 40 residues involved in the ten turns, 23 (57%) conform to those five amino acid

TABLE 3
β-Turn parameters

Residue no.	Sequence	$\phi 2$ (°)	$\psi 2$ (°)	$\phi 3$ (°)	$\psi 3$ (°)	*d* Cα1–Cα4 (Å)	*d* O1–N4 (Å)	Class
2–5	SVLT	−76	−47	−104	146	6·27	5·39	I(NI)
13–16	TPGQ	−50	130	95	−11	5·93	3·13	II
40–43	VPGK	−55	138	79	11	5·71	3·14	II
51–54	YNDL	70	−46	−143	12	5·86	4·10	IV
56–59	PSGV	−57	142	79	−6	6·04	3·38	II
60–63	SDRF	−72	3	−72	−18	6·10	3·19	III
68–71	SGTS	58	−117	−104	15	5·60	3·20	II′
80–83	ESED	−53	−37	−64	−14	5·49	2·91	III
93–96	NDSL	−65	−25	−81	−36	5·34	3·17	III
94–97	DSLD	−81	−36	−86	−21	5·00	3·72	I

NI, non-ideal, i.e. one angle off by more than 50°.

types. There is only one abnormal turn (type IV), which is the result of the unusual torsion angles for Asn52 discussed above. In this case, the usual 1—4 hydrogen bond is replaced by a 1—3 hydrogen bond. The only "non-ideal" entry (residues 2 to 5) is probably still an artifact of the disordered N-terminal residue.

(c) *Secondary structure*

The secondary structure of Rhe is almost exclusively β-sheet, and therefore Rhe is classified as a β-protein according to the scheme of Levitt & Chothia (1976). Sheet A (strands 2, 6 and 7), along with strands 1 and 5, forms one half of a β-barrel. These five strands make up approximately one-half of the molecule and the corresponding main chain atoms are shown in Figure 7(a). From the Figure, it is obvious that the sheet is quite regular, with no β-bulges of the types reported by Richardson *et al.* (1978). There are few hydrogen bonds from the inner strands to strands 1 or 5. Indeed, most of the carbonyl and NH groups in strands 1 and 5 are nearly perpendicular to the plane of sheet A. This is particularly the case for strand 1, with Thr5 and Thr13 being exceptions. The hydrogen bond to Thr13 is a consequence of the β-turn involving residues 13 to 16. In strand 5, even if the NH and carbonyl groups were ideally oriented, hydrogen bonds would not exist as this strand is considerably below the plane of sheet A. Note that strand 5 consists almost exclusively of members of the second hypervariable region, hence interactions between these residues and strand 6 are likely to vary considerably within V_L domains from different sources.

Analysis of sheet B (strands 3, 4, 8 and 9) is considerably more complicated. With the exception of the short helical region and a few turns linking to sheet A, these four strands complete the monomeric subunit of Rhe. The main chain atoms for this half of the molecule are shown in Figure 7(b). The most obvious features are the regular β-type hydrogen bonds between the two central strands, and the "wide" β-bulge (Richardson *et al.*, 1978) to the left involving residues Gly103, Gly104 and Tyr88.

The rightmost strand (strand 4) includes a "classic" β-bulge involving residues

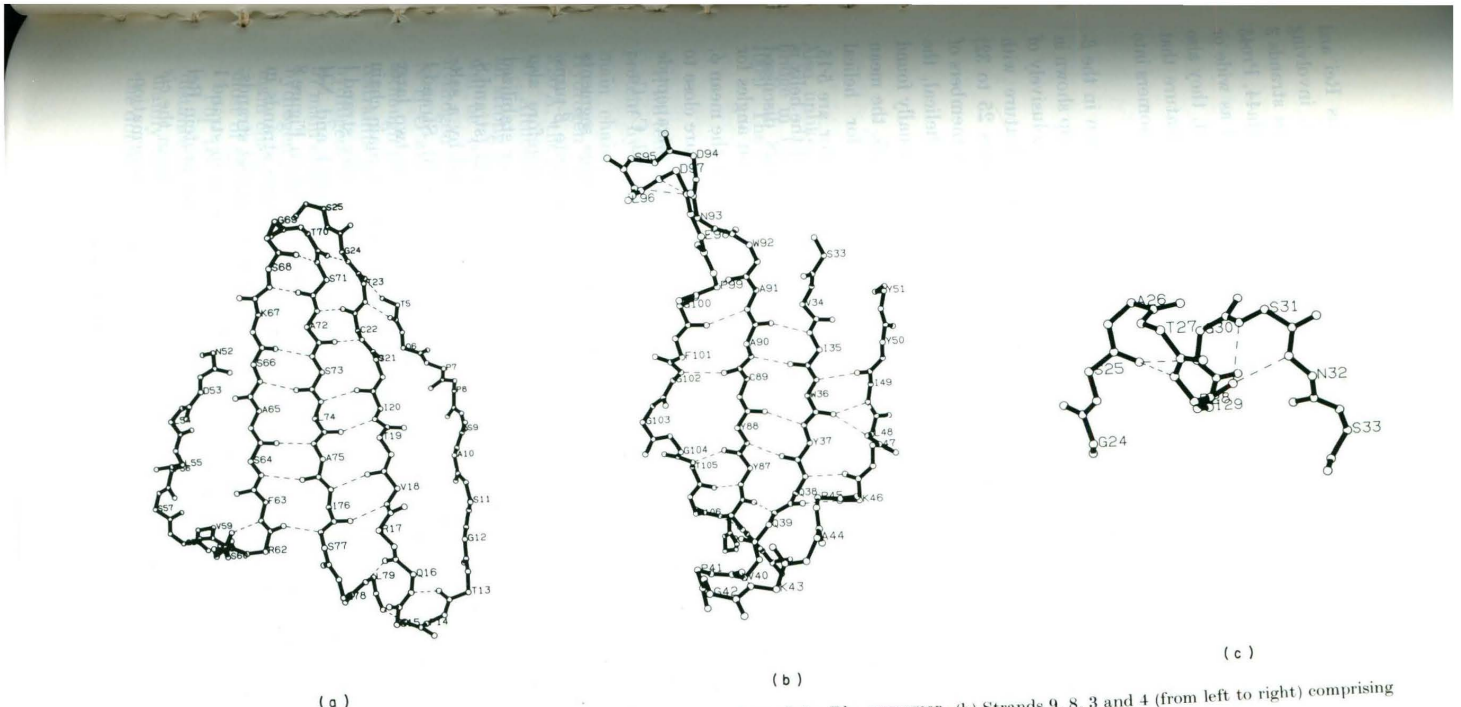


FIG. 7. (a) Strands 5, 6, 7, 2 and 1 (from left to right) comprising one-half of the Rhe monomer. (b) Strands 9, 8, 3 and 4 (from left to right) comprising the other half. (c) Helical region.

Leu48, Ile49 and Trp36. In similar regions of Bence–Jones proteins Rei and McPC603, a classic bulge was detected by Richardson *et al.* (1978), involving residues Leu47, Ile48 and Trp35. There is another bulge further down strands 3 and 4 after they turn away from sheet B. The residues involved are Ala44, Pro45 and Gln39, but it is not clear whether this bulge should be classified as wide or classic. Note in Figure 7(b) that as strands 3 and 4 proceed downward, they also move out, ending up nearly perpendicular to sheet B. It is this feature that ultimately determines the mode of association of variable domain monomers into a dimer.

In terms of the variability of amino acids, nearly all residues shown in the β-sheet portion of Figure 7(b) are generally considered conserved. The loop shown in the upper left containing residues 90 to 100, however, is made up exclusively of residues from the third hypervariable region. The only other feature with considerable secondary structure is the short helical region (residues 25 to 32) shown in Figure 7(c). Note that the region is made up exclusively of members of the first hypervariable region. Although this segment is obviously helical, the distances between α-carbons i and $i+3$ are considerably larger than usually found in helices. For the five consecutive tetrapeptides beginning with Ser25, the mean $C\alpha_i$–$C\alpha_{i+3}$ distance is 5·80 Å. The mean $C\alpha_i$–$C\alpha_{i+3}$ distances for helical tetrapeptides in myoglobin, lysozyme and pancreatic trypsin inhibitor are 5·15, 5·33 and 5·40 Å, respectively (Chou & Fasman, 1977). The extension of the helix is apparently required to accommodate the unusual angles for Asp28 discussed earlier. This is evident from an examination of the main chain torsion angles for residues 26 to 31, the inner residues of the helix. If Asp28 is excluded, the mean ϕ, ψ values for the five residues are $-73°$ and $-28°$, respectively, which are close to the normal helical values of $\phi \sim -60°$ and $\psi \sim -30°$. Therefore, if a tetrapeptide did not include Asp28, one would expect to observe normal helical $C\alpha_i$–$C\alpha_{i+3}$ distances.

The complete Rhe monomer is generated by placing the two large segments shown in Figure 7 back to back connecting them with the appropriate β-turns. The short helical segment connects strand 2 to strand 3, and therefore also connects sheet A to sheet B. The interaction between sheets is further stabilized by a disulphide bridge between Cys22 (strand 2, sheet A) and Cys89 (strand 8, sheet B), by hydrophobic interactions in the core of the protein, and by a salt-bridge between Arg62 (strand 6, sheet A) and Asp83 (strand 8, sheet B). Strand 1 also interacts with strand 9 (sheet B), forming a seam between the two large regions of β-structure. This is evident in Figure 8, where the complete main chain hydrogen bonding pattern is shown. Apparently, this is the reason why strand 1 interacts only weakly with sheet A, and also explains the carbonyl and NH orientations in strand 1 mentioned above. The skewed hydrogen bonds in Figure 8 result from the fact that these two strands are parallel. Of the nine strands in Rhe, this is the only occurrence of β-type interactions between parallel strands; all other interacting strands run in opposite directions. Note the kink in strand 1 caused by the Pro-Pro sequence for residues 7 and 8. In Bence–Jones protein Rei (Huber & Steigemann, 1974), residue 8 is also Pro, but was found to be in the *cis* conformation. All eight proline residues in Rhe are in the *trans* conformation.

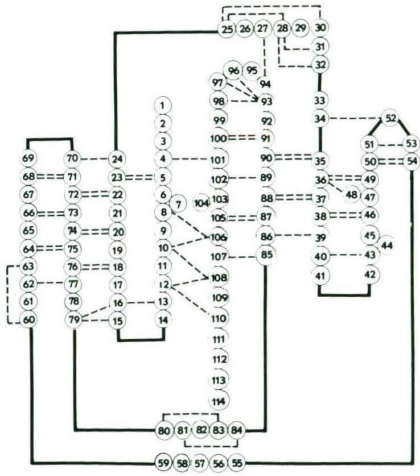


FIG. 8. Schematic diagram showing main chain hydrogen bonds.

Also, unlike in the Rei structure, there is no evidence of disorder in the disulphide bridge in Rhe. The S–S distance is 2·018 Å with C_β–S_γ–S_γ angles of 103·4° and 100·9°. The C_β–S–S–C_β torsion angle is 87·4°.

(d) *Hydrogen bonding*

If we temporarily restrict our attention to only one monomer, then the protein–protein hydrogen-bonding interactions can be separated into three categories: main chain–main chain, main chain–side chain, and side chain–side chain. Hydrogen bonds were included in Figure 8 if the N–O distance was less than 3·30 Å, the N–H–O angle greater than 120°, and the C–O–H angle greater than 90°. Although these criteria are somewhat arbitrary, they are reasonable in that all strong hydrogen bonds will certainly be detected, all totally unrealistic ones will be rejected, yet a considerable margin for error is provided for. Most hydrogen bonds are in fact quite normal, as indicated by mean N–O distances and N–H–O angles of 2·990 Å and 160°, respectively. It should be pointed out that although protein stereochemistry was restrained during the refinement process, no restraints were included for any of the hydrogen bond interactions. Therefore, the hydrogen bond parameters represent unbiased structural observations. There are 57 main chain–main chain, 21 main chain–side chain, and 11 side chain–side chain hydrogen bonds stabilizing the Rhe monomer. The dominance of main chain–main chain interactions is not surprising, considering that two β-sheets constitute most of the molecule.

For the main chain–side chain hydrogen bonds, the mean donor–acceptor distance decreases to 2·908 Å and the standard deviation rises to 0·213 Å. This is mainly the result of including three hydrogen bonds with oxygen atoms as both the donor and acceptor. If these three terms are neglected, the mean distance rises

TABLE 4
Side chain–side chain hydrogen bonds

Donor	Acceptor	*d* D–A (Å)	D–H–A (°)	C–O–H (°)
Q6NE2	T105OG1	3·005	174	114
R17NH2	S77OG	3·166	152	171
T27OG1	S25OG	3·209	155	140
T27OG1	N93OD1	3·148	164	99
N32ND2	D94OD2	3·115	178	102
R62NH1	D83OD1	2·830	174	123
R62NH2	D83OD2	2·994	169	121
K67NZ	N52OD1	2·912	172	103
N93ND2	D28OD1	2·858	169	119
S95OG	N93OD1	2·682	172	123
Q112NE2	E84OE2	3·254	156	96
Mean		3·016	167	119
Standard deviation		0·181	9	22

to 2·961 Å and the standard deviation drops to 0·174 Å. The adjusted statistics then are in reasonable agreement with the results for main chain–main chain hydrogen bonds. The mean donor–H–acceptor angle for these interactions is also 160°.

Of the 11 side chain–side chain hydrogen bonds (listed in Table 4), five correspond to interactions exclusively between hypervariable residues, and therefore are likely to be unique to the Rhe structure. Of the remaining six hydrogen bonds, the most important seem to be the two involved in the salt-bridge between Arg62 and Asp83 shown in Figure 9. In addition to connecting the two β-sheets, the residues in the Figure serve to insulate the non-polar core residues from solvent. Since these residues are not from hypervariable regions, this feature may be common to all V_L domains.

Although the mean donor–acceptor distances and donor–H–acceptor angles are similar in all three categories, the "acceptor angles" C–O–H show a marked tendency to decrease as the number of main chain atoms in the hydrogen bond decreases (146°, 132° and 119° for main chain–main chain, main chain–side chain and side chain–side chain interactions, respectively). This is probably the result of increased flexibility in the side-chains, enabling the more favorable acceptor

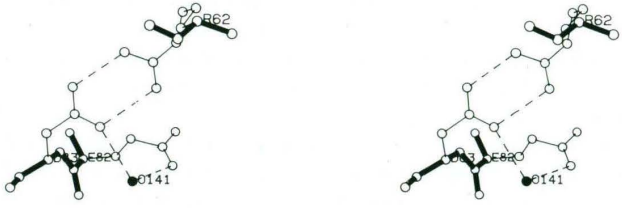


FIG. 9. Salt-bridge between Arg62 and Asp83.

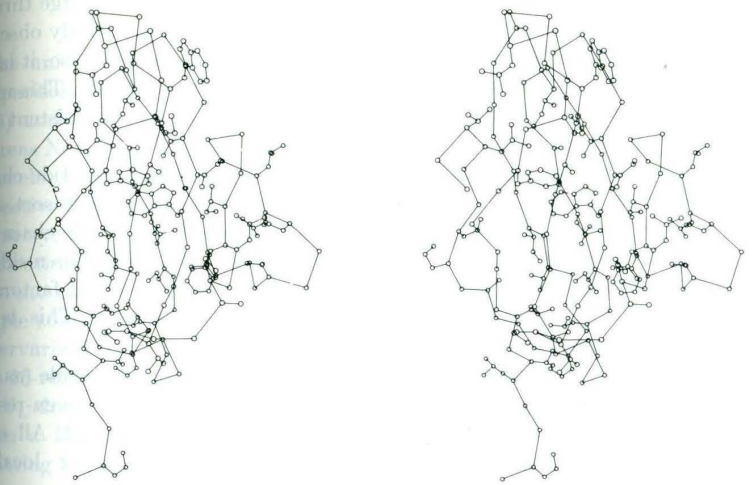


FIG. 10. All non-polar residues shown relative to the α-carbon frame.

angles to occur. For an ideal linear hydrogen bond, the donor–H–acceptor angle is 180°, and for a carbonyl oxygen as acceptor, the ideal C–O–H angle is 120°.

(e) *Non-polar residues*

The Rhe monomer contains 42 non-polar residues, of which 23 (55%) can be thought of as "core" because of their internal positions in the molecule. The locations of all non-polar residues relative to the α-carbon frame are shown in Figure 10. The core residues are mainly Leu, Val and Ile, but they also include the aromatic side-chains of Trp36 and Phe63. Trp36 is deeply buried and is almost at the center of gravity of the molecule, whereas Phe63 is near the bottom of the domain and is partially exposed to solvent. Of the 19 non-polar residues not designated as protein core, most participate in hydrophobic interactions across the 2-fold axis of dimerization.

(f) *Charged residues*

Considering the conditions of crystallization, 19 of the 114 amino acid residues in a Rhe monomer are expected to be charged. The seven Asp, five Glu, five Lys and two Arg residues produce a net charge of -5, in accord with the observed isoelectric point of 4·4. Surprisingly, there are no strong acid–base interactions other than the one salt-bridge in which Arg62 neutralizes Asp83. The positive charge on the side-chain of Arg62 might also be partially neutralizing Glu82 by transfer of charge through a solvent molecule. This was shown in Figure 9, where it is seen that water O141 links the side-chains of Glu82 and Asp83. Since the positive charge is strongly linked to the carboxyl group of Asp83, the charge may

also be propagated through water O141 to Glu82. The diffusion of charge through solvent molecules has been proposed by Finney (1977), and was recently observed in the crystal structure of actinidin by Baker (1980). In the present case, it is also possible that O141 is not a water oxygen at all, but an ammonium ion. This would certainly be more effective in neutralizing charge and is equally consistent with X-ray data at 1·6 Å resolution.

Although there are no other strong acid–base interactions between side-chains, nearly all charged residues are involved in polar interactions of some sort, with hydrogen bonds to water being the norm. The only charged side-chain not identified in any specific contacts is the C-terminal residue Lys114. This residue is unambiguously positioned in the electron density map, but its thermal factors are high (36 $Å^2$ for N_ζ) and considerable disorder may be present. This is not surprising, considering its exposed location (see Fig. 4).

In terms of the overall charge distribution, the bottom of the molecule (as seen in Fig. 4) is predominantly negative. This is caused mainly by the seven-residue sequence 80 to 86, which contains five acids expected to be ionized. All other regions in the monomer are relatively neutral, with no other localized concentration of charge.

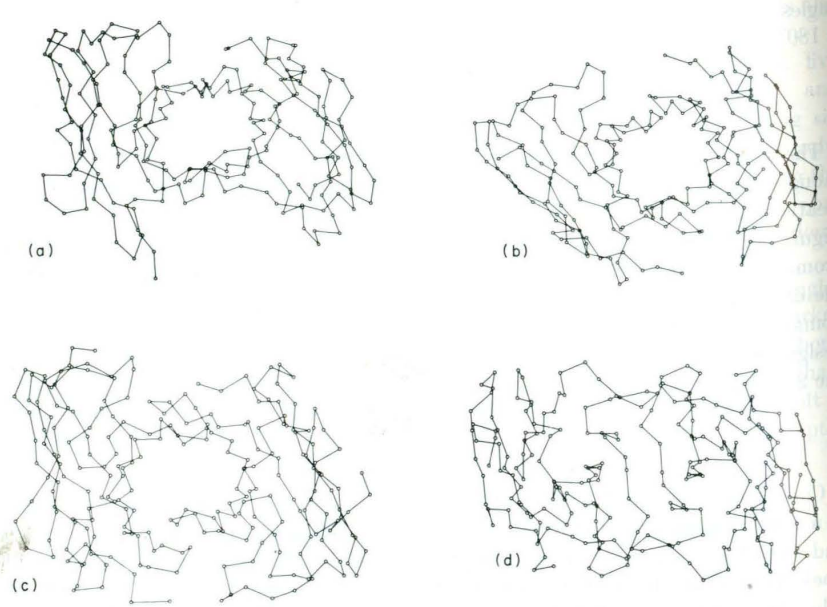


FIG. 11. Comparison of variable domain dimers viewed down their 2-fold or pseudo-2 fold axes. (a) Fab New; (b) Rei; (c) Mcg; (d) Rhe. Co-ordinates for Rhe are taken from the refined model. Co-ordinates for the others are taken from the protein data bank, Brookhaven National Laboratory Associated Universities Inc., Upton, Long Island, N.Y. In (a) and (c) the constant domains were omitted.

(g) *The Rhe dimer*

The unique mode of domain association exhibited by Rhe is best realized by comparison with the related structures New (Poljak *et al.*, 1973), Rei (Epp *et al.*, 1975) and Mcg (Schiffer *et al.*, 1973). Rei and Mcg are Bence–Jones proteins, whereas New is the antigen binding fragment (Fab) of an immunoglobulin G (IgG) molecule. Figure 11 presents the four variable domain dimers viewed down their respective 2-fold or pseudo-2-fold axes and shows the difference in combining site geometry. In Rhe, the combining site cavity is lined at the bottom instead of penetrating infinitely between domains. One might be tempted to associate the different cavity architectures with sequence changes resulting from the varied sample sources; however, the major structural change in Rhe occurs in the non-hypervariable residues 36 to 50, a region highly conserved in amino acid sequence. This 15-residue segment lies between the first and second hypervariable regions of the V_L domain and forms most of the interdomain contacts within the cavity. When a Rhe monomer is fit by a least-squares procedure to a Rei monomer (omitting hypervariable residues and the 15-residue loops from the minimization), there are deviations of up to 8 Å between corresponding atoms in the two loops. This is in spite of the fact that the amino acid sequences are nearly identical in these regions, with only one very conservative Thr/Val substitution distinguishing them.

In Figure 12, the complete Rhe dimer is shown both at the α-carbon level and with all atoms included. Since Rhe forms a non-covalent dimer, interactions between domains must either be due to hydrogen bonds, van der Waal's contacts or electrostatic forces. If we apply the same criteria for accepting bonds used within a monomer, then there are only six hydrogen bonds spanning the 2-fold axis. The 12 residues participating in hydrogen bonding across the 2-fold axis are Gln39, Ala44, Pro45, Lys46, Leu47, Ser57 and their symmetry-related counterparts. Note that only one of the hydrogen bonds (in fact, the weakest) involves a hypervariable residue (Ser57), and that all other hydrogen-bonded residues are members of the 15-residue loop.

Despite the presence of four lysine residues in the interface, there are no indications of any strong acid–base interactions between domains. The major forces maintaining the dimer are therefore hydrophobic interactions spanning the 2-fold axis. To identify strong hydrophobic interactions, all interatomic distances across domains were computed, and those less than 4 Å were deemed significant. The contacts involve 19 unique residues of which only seven (Ile35, Tyr51, Leu54, Pro56, Ser57, Trp92 and Pro99) are from hypervariable regions. Pro56 and Ser57 are the only hypervariable residues to form contacts with non-hypervariable residues, and 11 such contacts are formed. The strongest interactions that occur exclusively between hypervariable residues are between Tyr51 and Trp92 (7 contacts). The 19 residues involved in interdomain contacts are shown in Figure 13. In Figure 13(b), the contact region close to the 2-fold axis and at the bottom of the cavity is shown, highlighting the hydrogen bonding across the dyad. Although there are few direct contacts between hypervariable residues across the 2-fold axis, there are several cases where water molecules link residues

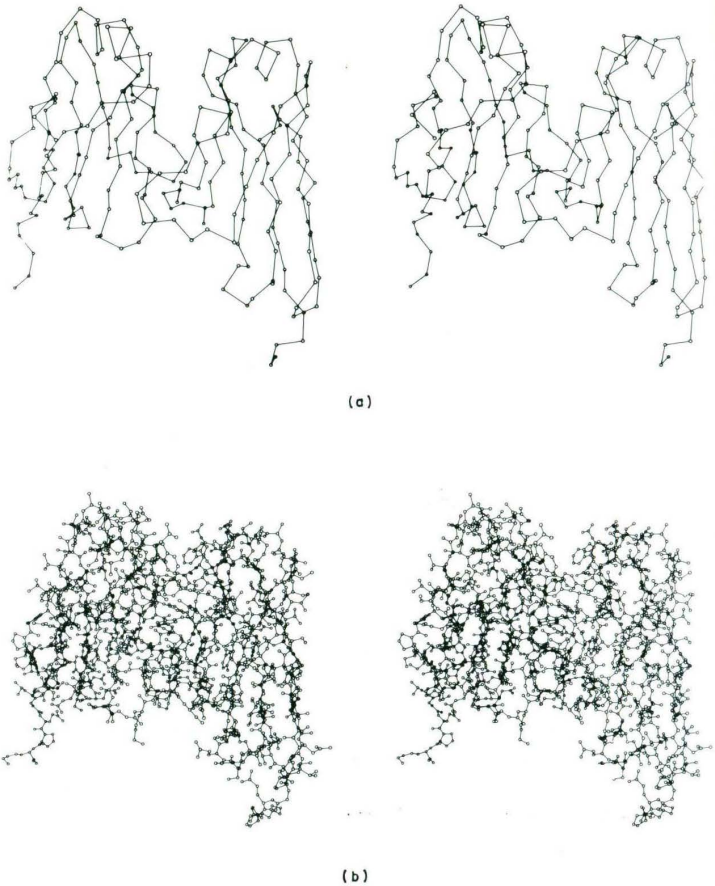
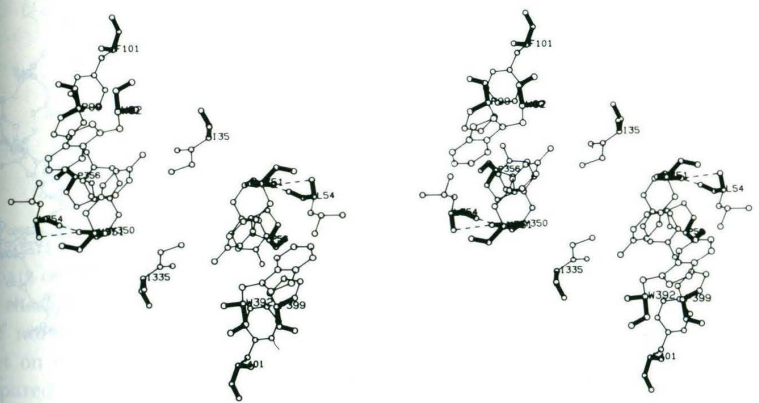


FIG. 12. (a) The Rhe dimer, α-carbon skeleton; (b) all atoms.

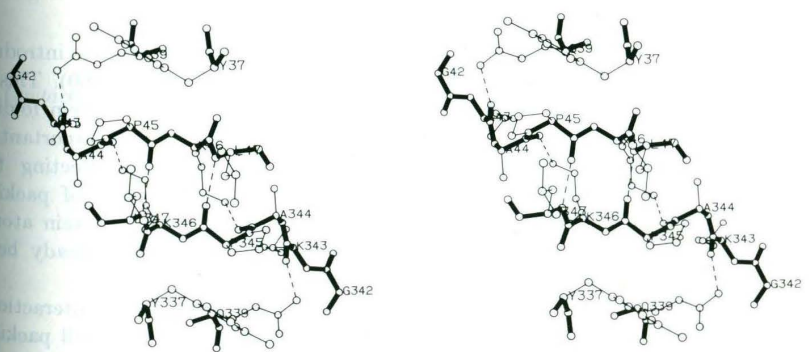
in dyad-related domains. These cases will be discussed later. The combining site cavity between domains is shown in Figure 13(c) in a view down the 2-fold axis. Only the hypervariable residues are included. From the Figure, obvious contenders for hapten binding are Ser33, Ile35 and Tyr51. Since most of the important contacts between domains are between non-hypervariable residues, we see no reason why other variable domain dimers cannot adopt this mode of domain–domain association.

(h) *Packing interactions*

In any crystal there will always be contacts to neighboring molecules that could conceivably influence molecular features within the unique part of the structure.

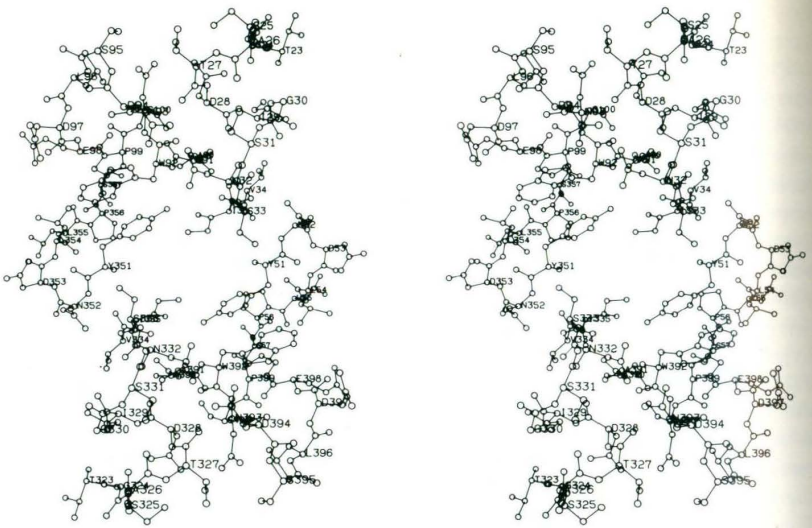


(a)



(b)

FIG. 13.



(c)

FIG. 13. Contacts across the 2-fold axis of dimerization. (a) Upper region near the top of the cavity; (b) lower region at the base of the cavity; (c) all hypervariable residues.

In the case of protein crystals, however, intermolecular contacts tend to introduce only minor effects on the overall molecular structure (Rupley, 1969). This is probably because the interactions believed to be responsible for protein folding are established long before crystallization occurs. Nevertheless, it is important to determine whether or not packing interactions are seriously affecting the conformation of Rhe within the crystal. To determine the extent of packing interactions, all interatomic distances between symmetry-related protein atoms were computed (neglecting the 2-fold related domain, which has already been discussed).

The unique interatomic contacts less than 4 Å are confined to interactions between 29 residue pairs involving seven different domains. The overall packing pattern is shown in Figure 14, which is a view down the *c* axis of the crystal. In addition to showing the packing interactions between molecules, the Figure also clearly reveals the presence of solvent channels parallel to the *c* axis and extending infinitely through the crystal. The 29 interacting residue pairs are listed in Table 5 and are grouped according to the symmetry operator used to generate the molecule in contact. Of the 29 pairs, only the six flagged with a footnote sign are hydrogen-bonded, hence the major packing interactions are hydrophobic in nature. The contacts are spread fairly evenly over the symmetry operations, so we do not expect packing forces between any particular pair of domains to dominate the structure.

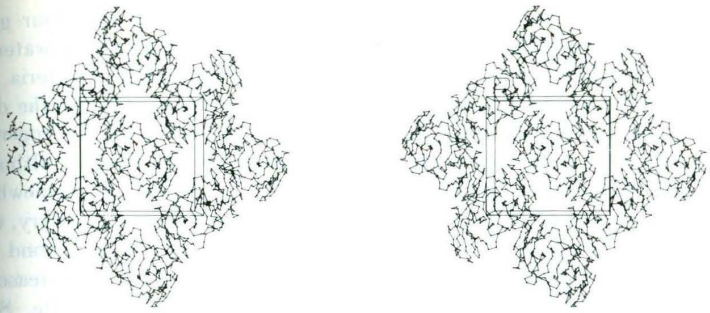


FIG. 14. Packing diagram viewed down the *c* axis.

Of particular interest are the contacts involving Gln39 and Pro41, since they are part of the 15-residue segment. Upon close examination, however, it is evident that these contacts are incidental and in no way force the 15-residue loop to orient itself normal to the β-sheets. We conclude that packing forces have no major effect on the folding pattern of a Rhe monomer, and in fact are probably minor compared to the forces holding the dimer together.

(i) *Water structure*

Since Rhe crystals are known to contain approximately 51% solvent by volume (Wang *et al.*, 1979), one expects many protein–solvent interactions within the

TABLE 5

Contacts to symmetry-related molecules: symmetry operation applied to second residue in pair

Symmetry operator $X-1/2, -Y-1/2, 1-Z$		
S9–D61†	S11–G58†	S11–S57
S11–G58	G12–S57	T13–S57
Q16–S57		
Symmetry operator $-X-1/2, Y-1/2, 1-Z$		
T13–F101	Q16–Y88†	Q16–Q39†
R17–P41		
Symmetry operator $1/2+X, -Y-1/2, 1-Z$		
L55–L110	P56–L110	S57–L110
Symmetry operator $-X-1/2, 1/2+Y, 1-Z$		
E1–P14	L96–P113	L96–K114†
L96–G111	D97–K114	E98–Q112
E98–K114	P99–G111	F101–G111
Symmetry operator $X, Y, Z-1$		
T70–K114†	G69–K114	G69–P113
S68–K114	S68–P113	
Symmetry operator $-X-1/2, Y-1/2, -Z$		
S68–S95		

† Contact includes hydrogen bond.

crystal. The interactions involving solvent can be separated into four general categories: main chain NH–water; main chain O–water; side chain–water; and water–water. Unlike the previously described hydrogen-bonding criteria, when water is involved we do not consider any angular information; only the donor–acceptor distance is used to detect hydrogen bonds. This is because hydrogen atom positions cannot be computed accurately for most of the interactions, although it is possible for those involving nitrogen as the donor. To compensate somewhat for the possible inclusion of interactions that may have unfavorable geometry, we use a more selective cutoff distance of 3·2 Å for acceptance of a hydrogen bond. Upon examination of the final structure, nearly all water molecules were reasonably positioned, thus indicating that the selection criteria was adequate. Similar criteria were applied to analyze the water structure in the iron–sulfur protein rubredoxin (Watenpaugh *et al.*, 1978).

Of the 114 main chain nitrogen atoms in Rhe, the presence of eight proline residues reduces the number available for hydrogen bonding to 106. Of the 106, 21 are found to be hydrogen-bonded to water, while 13 are found not to be hydrogen-bonded at all. The hydrogen-bonding potential of main chain NH groups is therefore realized to 88% of its maximum value. Of the 13 not designated as hydrogen-bonded, most are involved in contacts of less than 4 Å to suitable acceptors. The majority of NH groups bound to water are in the loops connecting strands, which is not surprising since most other NH groups are tied up in the β-sheets. The mean N–O distance is 2·976 Å with an estimated standard deviation of 0·177 Å, in good agreement with values found for main chain–main chain hydrogen bonds.

For main chain oxygen–water interactions, all 114 carbonyl oxygen atoms are theoretically available, although many are also tied up in β-sheets. Of the 114, 49 form hydrogen bonds to water, and 13 form no hydrogen bonds at all. Therefore, the hydrogen-bonding potential is also realized to the extent of 88% if we assume only one hydrogen bond per oxygen atom. As before, most of the atoms not designated hydrogen-bonded actually are involved in contacts just outside the cutoff. The mean O–O distance is 2·850 Å with an estimated standard deviation of 0·183 Å.

In a Rhe monomer, there are 59 residues with side-chain atoms capable of forming hydrogen bonds, and 53 (90%) actually do so. In terms of the side chain–water interactions, there are 62 hydrogen bonds involving 46 different protein atoms: 14 of the residues have charged side-chains, hence solvent is clearly aiding in charge delocalization. It is also possible that some of the "waters" may really be ammonium ions, although there are no clearcut cases. The mean donor–acceptor distance is 2·896 Å with an estimated standard deviation of 0·203 Å. Note that roughly 90% of the hydrogen bonding potential is realized in all categories.

For the 246 water–water hydrogen bonds in Rhe, the mean O–O distance is 2·763 Å with an estimated standard deviation of 0·266 Å. The estimated standard deviation is significantly higher for water–water interactions than for any other category. This is partially because large thermal factors prevent accurate determination of atomic positions, but it may also reflect the transition from ordered water to bulk liquid. A detailed analysis of all water–water interactions is

TABLE 6
Water oxygen hydrogen-bonding statistics

Number of hydrogen bonds to protein	Total number of hydrogen bonds						
	0	1	2	3	4	5	Total
0	13	31	30	14	2	0	90
1		14	24	16	4	2	60
2			7	12	5	1	25
3				5	3	0	8
4					2	1	3
5						0	0
Total	13	45	61	47	16	4	186

beyond the scope of this text, and it suffices to state that most water–water interactions are confined to the large solvent channels.

In the preceding paragraphs, interactions involving water were described from the point of view of protein atoms. An alternative and equally informative approach is to consider each water molecule and the hydrogen bonds it forms separately. A convenient way to summarize water interactions is to construct a matrix whose elements are the number of water molecules that form 0, 1, 2, etc. hydrogen bonds. This approach was introduced by Watenpaugh *et al.* (1978) and gives a satisfactory description of the hydrogen-bonding distribution. The matrix for Rhe is given in Table 6. The bottom row gives the total number of water molecules involved in 0, 1, 2, etc. hydrogen bonds, depending on the column, while intermediate row entries determine how many of the hydrogen bonds were made to protein. For example, there are 61 water molecules that form two hydrogen bonds, of which 30 form none to protein, i.e. they make hydrogen bonds only to other water molecules. Twenty-four must bridge protein to solvent, because only one of their hydrogen bonds is to protein, and seven bond exclusively to protein. This type of data organization draws attention to water molecules that are tightly bound, and suggests an order for examining solvent interactions. Presumably, water molecules that form four or more hydrogen bonds, with three

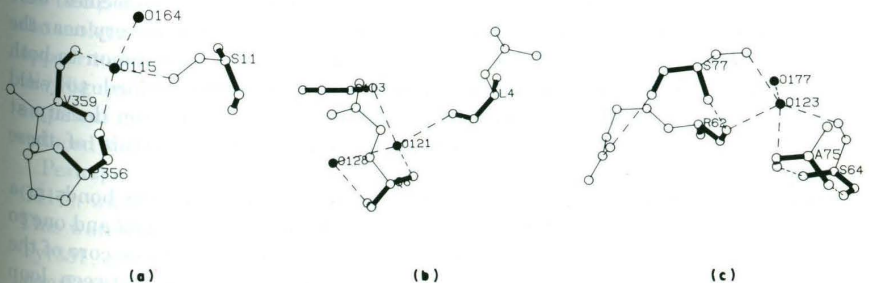


FIG. 15. Several multiply bound water molecules. (a) Water O115; (b) water O121; (c) water O123.

or more to protein are structurally significant, and may even be considered an integral part of the molecular structure. Several examples of these "structural" water molecules are shown in Figure 15.

In this and subsequent Figures, if a residue belongs to a symmetry-related domain, its residue number is incremented by 300. Thus, the first 300 residues (1 to 114 = protein, 115 to 300 = water) define the contents of the crystallographic asymmetric unit. Residue numbers for the water molecules indicate the order in which they were incorporated in the model, and therefore are a measure of their peak heights in the electron density maps. Not surprisingly then, solvent residue numbers are also well-correlated with their B values, and serve as indicators of their reliability. The first 84 water molecules, numbered 115 to 198, are fully occupied, while those numbered 199 to 300 were given variable occupancy factors during refinement.

In general, most of the tightly bound water molecules have tetrahedral geometry, although it is often distorted. In Figure 15(a), O115 is seen to hydrogen bond to the side-chain of Ser11, O of Pro56, O of Val59 and to water O164 with tetrahedral geometry. Note that O115 was the first water oxygen located, consistent with its tightly bound state. In Figure 15(b), a similar situation is shown for O121. In this case, hydrogen bonds are formed to main chain N of Glu6, N of Gly103, O of Leu4 and water O128. For O123, shown in Figure 15(c), there are five hydrogen bond interactions. The donors are the OH groups in Ser64, Ser77 and water O177, while the acceptors are the oxygen atoms, of Ala75 and Arg62. Note that in Figure 15(a) symmetry-related domains are involved, whereas in (b) and (c) all residues involved are in the same molecule. In fact, in the latter two cases all residues are from non-hypervariable regions. Thus, these features may be present in other V_L domains.

The previous section is not meant to imply that water molecules involved in less than four hydrogen bonds are not important. It merely illustrates the general tendency of water to fill in gaps in the protein surface and, wherever possible, to link side-chains in an energetically favorable manner. In cases where three hydrogen bonds are formed, the tetrahedral symmetry is usually maintained but with one of the lone electron pairs on the water not participating in bonding.

The identification of several strongly bonded water molecules in key structural segments of the protein indicates that solvent may help stabilize the novel Rhe conformation. In particular, two of these "structural" water molecules were identified, one in what is usually thought of as a hydrophobic region very near the 15-residue loop, and the other deep in the combining site cavity common to both V_L domains constituting the dimer. These water molecules refined to yield isotropic thermal factors of 9·7 and 8·9 Å^2, respectively. The electron density at both of these sites is 3·11 e/Å^3, reflecting the highly ordered nature of these waters.

Water O118 (Fig. 16(a)) is found to form three distinct hydrogen bonds in a roughly trigonal arrangement, one to NH1 of Trp36, one to O of Asn52 and one to O of Ser66. This water is interesting because of its location deep in the core of the protein, and also because it is curiously close to the junction between loop residues and hypervariable regions (note the position of Trp36 in the loop). In a

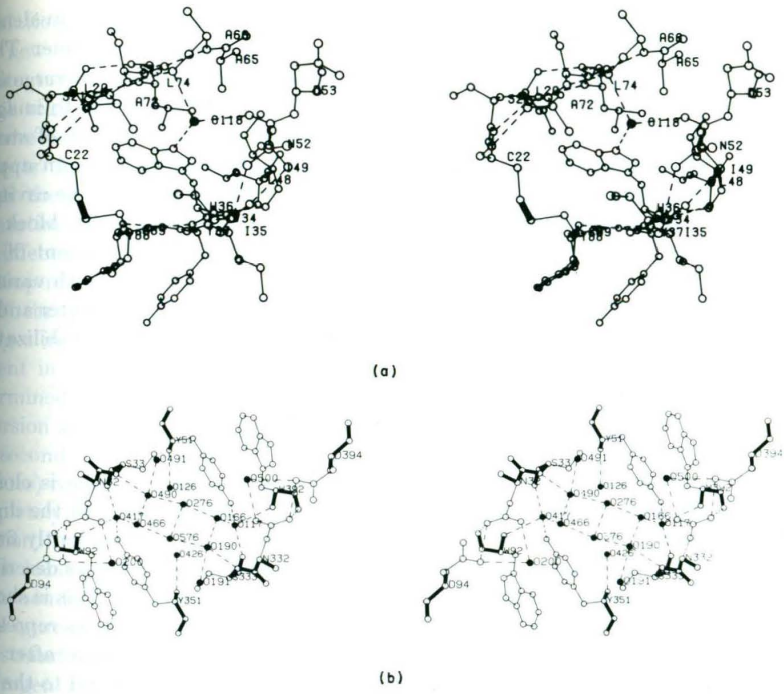


FIG. 16. (a) Water O118 deep in the protein core amongst hydrophobic residues. Note that it is Trp36, an upper loop residue, to which this water is hydrogen-bonded. (b) Arrangement of water molecules in the cavity between variable domains in the dimer. Note the location of water O417 (equivalent to O117), which forms hydrogen bonds to all 3 hypervariable region residues.

survey of immunoglobulin structures, Padlan (1977) classified Trp36 as being completely buried, and therefore it would not be expected to be accessible to solvent. This is clearly not the case, although detection of water molecules may not have been possible with the lower resolution data available for the immunoglobulins used in the study. Water O118 is actually sandwiched between the two β-sheets, which are viewed "end on" in the Figure. This water is very near the molecular center of the protein in a hydrophobic region, although there does appear to be an accessible route to the protein surface in the upper right corner of the Figure. In a study of the conformational energy of a Rhe monomer by semi-empirical methods (Furey, unpublished results) the presence of O118 in this region reduced the total energy of the system by $\sim$9·5 kcal/mol.

Perhaps the most intriguing solvent molecule is water O417 (equivalent to O117), which lies deep in the combining site cavity and is shown in Figure 16(b). This water forms four distinct hydrogen bonds, one to O of Ser33, one to OH of Tyr351, one to N of Trp92 and one to another water molecule. This water may be significant if one realizes that the three amino acid residues involved are from each of the three respective hypervariable regions. Ser33 and Trp92 represent the

first and third hypervariable regions in one monomer, while Tyr351 (equivalent to Tyr51) represents the second hypervariable region from the other monomer. Thus, representative residues from each of the three hypervariable regions converge on a single water molecule deep in the combining site cavity. It is difficult to imagine this water being displaced upon binding; first, because of the number of strong hydrogen bonds that must be broken, and second, because there does not appear to be any exit route for this water molecule (recall that the bottom of the cavity is blocked in Rhe, and any incoming antigen or hapten could conceivably block the top). It seems then, that this water may indeed be a structural element of the protein helping to stabilize the Rhe conformation by tying together both variable domains as well as the three hypervariable regions. We estimate this water and its dyad-related counterpart to contribute a total of ~22 kcal/mol stabilization energy to the dimer.

(j) *Loop conformation and domain–domain association*

It can be shown that the orientation of the 15-residue loop in Rhe is closely related to the mode of domain–domain association by comparison with the dimer of Bence–Jones protein Rei. If each domain of a Rhe dimer is independently fit by a least-squares procedure to the corresponding domain of a Rei dimer as described earlier, atomic collisions result near the local dyad-relating domains.This is shown in Figure 17 in a view down the pseudo-2-fold axis. The continuous lines represent the Rei structure, while the broken lines represent the Rhe structure after the least-squares fit. The broken lines in the center of the dimer correspond to the 15-residue loop, with the shortest contact being between Pro41 and its dyad-related counterpart. The distance between α-carbons in this contact is only 2·8 Å. If all atoms were included there would be many more bad contacts, with some atoms even crossing the 2-fold axis. Therefore, if the 15-residue loop is oriented normal to the β-sheets as it is in Rhe, there *must* be an associated rotation of one domain relative to the other to avoid collisions. Alternatively, if single domains from each of the Rhe and Rei dimers are aligned by a least-squares procedure, then one of the remaining domains requires a rotation of ~61° and a centroid

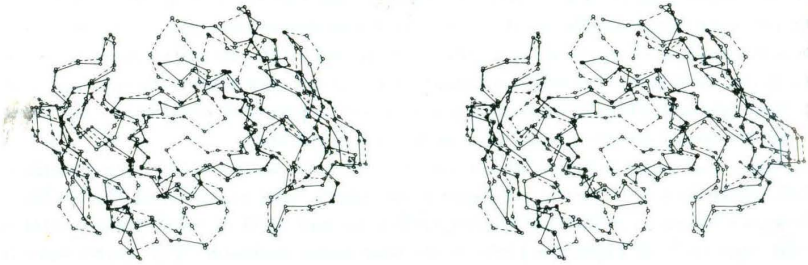


FIG. 17. Rhe structure (broken lines) after a least-squares fit to the Rei structure (continuous lines). Note the short contacts near the pseudo-2-fold axis.

translation of ~19 Å to become aligned with its counterpart. The major effect of this new mode of dimerization, in addition to providing a different cavity architecture, is to increase the distance between corresponding residues in the two switch regions by 6 to 9 Å.

There are no indications from the packing analysis of any intermolecular contacts that might favor the Rhe loop conformation other than those contacts across the axis of dimerization itself. Accordingly, we attribute the novel loop orientation to be an intrinsic feature of the molecule and not an artifact of crystal field effects. A comparison of main chain torsion angles between Rhe and Rei within the loop residues indicates no abrupt change, although there are deviations of up to 30° throughout with a root-mean-square deviation of 18°. Despite these subtle changes in torsion angle, the same hydrogen-bonding scheme between strands is present in both structures. It appears likely that the overall loop orientation is determined by the cumulative effect of sequence variations and/or subtle changes in torsion angle originating at or beyond the top of the two strands, i.e. in the first and second hypervariable regions. The largest atomic shift (for non-hypervariable residues) relative to the Rei structure occurs in Tyr37, in which the side-chain is completely reoriented, corresponding to an 8 Å shift in the terminal oxygen position.

The local conformation of the 15-residue loop is shown in Figure 18 with main chain bonds thickened and hydrogen bonds as broken lines. It is apparent from this Figure that the loop is not in a random conformation but behaves more like a two-strand sheet, implying a certain degree of structural rigidity in this region. Hydrogen bonding and van der Waals' contacts to neighboring hypervariable residues are apparent from Figure 8. In particular, interactions with residues 51 to 54 may play an important role in determining the loop orientation. The fact that the loop links to hypervariable regions 1 and 2 (residues 23 to 35 and 51 to 57,

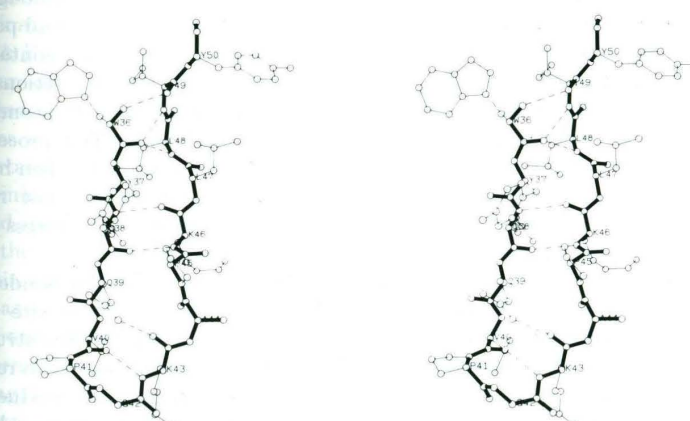


FIG. 18. (a) Local conformation of the 15-residue loop consisting of the non-hypervariable residues 36 to 50.

respectively) is significant, because it can relate the conformational change in the loop to amino acid changes associated with antigen specificity, although the sequence within the loop itself is invariant. First and second hypervariable region residues can apply the appropriate force to orient the loop *via* hydrogen bonding, van der Waals' contacts and subtle changes in torsion angle. The result of these forces, in addition to changing the local loop orientation, is a change in the mode of association of variable domains into dimers. The novel association of domains is a necessary consequence of the new loop orientation, since unfavorable atomic collisions across the 2-fold axis would otherwise occur. This reorienting of domains can ultimately generate combining site shapes and sizes that could not be obtained simply by plugging different amino acids into hypervariable segments of a *structurally invariant* framework.

4. Discussion

The existence of the thermodynamically stable Rhe structure is in contradiction to the hypothesis of a structurally invariant framework common to all variable domains. If this concept is to be retained, clearly it must be modified to exclude residues 36 to 50 from the framework. It should be obvious that researchers studying immunoglobulin crystal structures by the molecular replacement method (Rossmann, 1972) should use extreme caution in selection of search models, since their unknown structures may differ in the mode of dimerization as well as in the conformation of the isolated monomer. It might be best to consider only monomers as search models, and omit residues 36 to 50.

The accuracy of the model obtained from crystallographic refinement with high-resolution data enabled identification of structural features relating the domain–domain mode of association, i.e. combining site geometry, to the conformation of the non-hypervariable loop residues 36 to 50 of each monomer. This 15-residue loop appears to be somewhat variable in curvature, although it is highly conserved in amino acid sequence. Apparently, its orientation and position can be altered through hydrogen bonding and/or van der Waals' contacts to residues in the first and second hypervariable regions. These interactions with hypervariable residues may ultimately determine gross features of the combining site cavity by altering the mode of variable domain association. This process can result in a larger range of combining site geometries than is possible when holding all non-hypervariable residues structurally invariant, yet it does not require additional alterations in variable domain DNA sequences. Such a process would therefore extend the range of binding specificity very efficiently.

The detection of solvent molecules that are strongly hydrogen-bonded and intimately related to both the 15-residue loop and the combining site cavity suggests that solvent may play an important role in stabilizing the Rhe structure. Note that the two tightly bound water molecules form hydrogen bonds to residues 36, 51 and 52, all of which themselves form contacts to upper loop residues near the hypervariable region junctions. Since there is no other immunoglobulin structure known to this resolution, nor are there any for which the solvent positions have been determined so extensively, it is difficult to assess the effect of solvent in

stabilizing protein conformation. This problem will be solved only when more immunoglobulin structures have been determined at high resolution. However, insight into the problem might be gained by applying conformational energy minimization techniques to the Rhe crystal structure in the presence and absence of solvent. Work along this line is underway in our laboratory.

Most of the conclusions discussed in the preceding paragraphs are drawn directly from the observed correlation between gross structural features of the combining site, domain–domain association, and the conformation of loop residues 36 to 50 within each monomer. In general, we have considered variations in domain–domain association as a means of expressing antigen specificity, since origins of the variations can be traced to hypervariable residues. It is also possible that changes in the hypervariable regions that are less dramatic than amino acid substitutions might also reorient the loop, and consequently alter the mode of dimerization. Adjustments in domain–domain association in response to changes in the first and second hypervariable regions may also permit a certain degree of induced fitting of the combining site to better accommodate antigenic determinants. When viewed in this light, a plausible trigger mechanism for relaying information from the combining site to switch regions connecting variable and constant domains is unveiled. Before proceeding further, it must be stated that the Rhe V_L dimer is an antibody analogue (fragment) and not an intact antibody. However, it is well-known that Bence–Jones dimers and Fab fragments share both conformational features and many chemical properties. Indeed, it has even been suggested that Bence–Jones proteins may be considered as primitive antibodies of unknown specificity (Edmundson *et al.*, 1974). Hence the extrapolation to intact antibodies is not unwarranted.

The details for transmission of binding information from the antigen combining site to the complement fixing site on antibody molecules is perhaps the most intriguing unanswered question in structural immunology today. Whether the transmission is by an allosteric, associative or distortive mechanism, or by some combination of these mechanisms is still an open question (Metzger, 1974,1978; Reid & Porter, 1975; Beale & Feinstein, 1976; Givol, 1976). Although an allosteric mechanism has been proposed (Huber *et al.*, 1976), it does not consider actual changes in the combining site, instead it concentrates on the hinge and switch regions connecting domains. Although the hinge and switch regions are indeed important, all binding information must necessarily be initiated within the combining site, and be transmitted to the hinge and switch regions only as a second step. Since the combining site architecture is shown to be highly correlated with the 15-residue loop conformation, and ultimately with the mode of domain–domain association, it is reasonable to expect that changes in the combining site such as those which could occur upon binding might also reorient the 15-residue loop, and subsequently change the mode of association of domains. The large conformational change resulting from domain–domain reorienting is necessarily propagated to the switch regions, and could subsequently be passed to domains further down the chain. It is noteworthy that spectroscopic evidence (Lancet *et al.*, 1977) suggesting large symmetry-conserving conformational changes upon binding has been reported.

A simple primary and secondary structure analysis adds credence to this mechanism, although it does not necessarily prove the allosteric model. Since the complement fixing reaction is common to most antibodies yet must be initiated from within the variable domains, it is very likely dependent on the amino acids that are highly conserved within the variable domains. Since the majority of non-hypervariable residues in variable domains are tied up in β-sheets, there are few candidates left that might be associated with any trigger mechanism (unless one assumes that the sheets are destroyed upon binding, but this is highly unlikely). Residues 36 to 50 are the only realistic choices. The fact that they connect directly to hypervariable residues and are involved in most of the interdomain contacts serves to strengthen this hypothesis. It is also interesting to note that although the "immunoglobulin fold" is present in all immunoglobulin domains, the constant domains not involved in binding of antigen are seven-strand structures, while the variable domains are nine-strand structures. One of the extra strands associated with the antigen-binding domains lies within the 15-residue loop, implying a binding-related function for this loop (and its counterpart in the heavy chain variable domain).

If this hypothesis is indeed correct, there is an important implication with regard to crystallographic binding studies. It implies that in order to observe the conformation of a bound complex crystallographically, one must cocrystallize the immunoglobulin with antigen or hapten rather than simply soak the latter into previously formed immunoglobulin crystals. This is because the large conformational changes expected would either be prevented from occurring because of the rigid crystal lattice, or would occur but probably destroy the crystal in the process.

Also if the postulated trigger mechanism is correct, it remains to be explained why Rhe exists in this new conformation in the first place since there is no hapten or antigen bound in the crystal. It may be that variable domain dimers can exist in two forms, which are separated by a low energy pathway, with the structures of Fab New, Rei and Mcg representing one form and Rhe the other. If the activation energy for a transition from one form to the other is sufficiently low, from statistical considerations one would expect to observe both conformations to some extent even in the absence of antigen, although the unbound conformation should be dominant. Evidence for such an effect in antibodies has been reported (Chiang & Koshland, 1979), where a small amount of complement was found to be fixed by hapten-specific immunoglobulin M (IgM) antibodies even in the absence of hapten. The assumption of a low energy barrier between bound and free dimers is also consistent with the remarkable efficiency of the immune response. Crystal packing forces might even reduce the required activation energy for a transition, thereby increasing the frequency of occurrence of the lower probability state. A conclusive test of this hypothesis will be obtained only when an immunoglobulin is crystallized bound to its antigen.

The authors are grateful to the University of Pittsburgh for use of their computing facilities. Atomic co-ordinates have been deposited in the Protein Data Bank, Brookhaven

National Laboratory Associated Universities Inc., Upton, Long Island, N.Y. This work was supported by the Medical Research Service of the V.A. Medical Center and by National Institutes of Health grant Am-Ca 18827.

REFERENCES

Amzel, M. L. & Poljak, R. J. (1979). *Annu. Rev. Biochem.* **48**, 961–997.
Baker, E. N. (1980). *J. Mol. Biol.* **141**, 441–484.
Beale, D. & Feinstein, A. (1976). *Quart. Rev. Biophys.* **9**, 135–180.
Chiang, H. C. & Koshland, M. E. (1979). *J. Biol. Chem.* **24**, 2736–2741.
Chou, P. Y. & Fasman, G. D. (1977). *J. Mol. Biol.* **115**, 135–175.
Crawford, J. L., Lipscomb, W. N. & Schellman, C. G. (1973). *Proc. Nat. Acad. Sci., U.S.A.* **70**, 538–542.
Cruickshank, D. W. J. (1949). *Acta Crystallogr.* **2**, 65–82.
Davies, D. R., Padlan, E. A. & Segal, D. M. (1975). *Annu. Rev. Biochem.* **44**, 639–667.
Edmundson, A. B., Ely, K. R., Girling, R. L., Abola, E. E., Schiffer, M., Westholm, F. A., Favsch, M. D. & Deutsch, H. F. (1974). *Biochemistry*, **13**, 3816–3827.
Epp, O., Lattman, E. E., Schiffer, M., Huber, R. & Palm, W. (1975). *Biochemistry*, **14**, 4943–4952.
Finney, J. L. (1977). *Phil. Trans. Roy. Soc. ser. B*, **278**, 3–32.
Furey, W., Wang, B. C., Yoo, C. S. & Sax, M. (1979). *Acta Crystallogr. sect. A*, **35**, 810–817.
Furey, W., Wang, B. C. & Sax, M. (1982). *J. Appl. Crystallogr.* **15**, 160–166.
Givol, D. (1976). In *Receptors and Recognition, Series A* (Cuatrecasas, P. & Greaves, M. F., eds), vol. 2, pp. 3–38, Chapman and Hall, London.
Hendrickson, W. A. & Konnert, J. H. (1978). In *Biomolecular Structure, Function, Conformation and Evolution* (Srinivasan, R., ed.), vol. 1, pp. 43–57, Pergamon, Oxford.
Huber, R. & Steigemann, W. (1974). *FEBS Letters*, **48**, 235–237.
Huber, R., Deisenhofer, J., Colman, P. M. & Matsushima, M. (1976). *Nature (London)*, **264**, 415–420.
Kabat, E. A., Wu, T. T. & Bilofsky, H. (1977). In *Variable Regions of Immunoglobulin Chains*, Bolt Beranek and Newman, Inc., Cambridge, Mass.
Lancet, D., Licht, A., Schechter, I. & Pecht, I. (1977). *Nature (London)*, **269**, 827–829.
Levitt, M. & Chothia, C. (1976). *Nature (London)*, **261**, 552–557.
Lewis, P. N., Momany, F. A. & Scheraga, H. A. (1973). *Biochim. Biophys. Acta*, **303**, 211–229.
Luzzati, V. (1952). *Acta Crystallogr.* **5**, 802–810.
Metzger, H. (1974). *Advan. Immunol.* **18**, 169–207.
Metzger, H. (1978). In *Contemporary Topics in Molecular Immunology* (Inman, F. P. & Reisfeld, R. A., eds), vol. 7, pp. 119–153, Plenum, New York.
Padlan, E. A. (1977). *Quart. Rev. Biophys.* **10**, 35–65.
Poljak, R. J., Amzel, L. M., Avey, H. P., Chen, B. L., Phizackerley, R. P. & Saul, F. (1973). *Proc. Nat. Acad. Sci., U.S.A.* **70**, 3305–3310.
Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.
Reid, K. B. M. & Porter, R. R. (1975). In *Contemporary Topics in Molecular Immunology* (Inman, F. P. & Mandy, W. J., eds), vol. 4, pp. 1–22, Plenum, New York.
Richards, F. M. (1968). *J. Mol. Biol.* **37**, 225–230.
Richardson, J. S., Getzoff, E. D. & Richardson, D. C. (1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**, 2574–2578.
Rossmann, M. G. (1972). *The Molecular Replacement Method*, Gordon and Breach, Science Publishers Inc., New York.
Rupley, J. A. (1969). In *The Structure and Stability of Biological Macromolecules* (Timasheff, S. N. & Fasman, G. D., eds), pp. 291–352, Marcel Dekker, New York.
Schiffer, M., Girling, R. L., Ely, K. R. & Edmundson, A. B. (1973). *Biochemistry*, **12**, 4620–4631.

Venkatachalam, C. M. (1968). *Biopolymers*, **6**, 1425–1436.
Wang, B. C. & Sax, M. (1974). *J. Mol. Biol.* **87**, 505–508.
Wang, B. C., Yoo, C. S. & Sax, M. (1979). *J. Mol. Biol.* **129**, 657–674.
Watenpaugh, K. D., Margulis, T. N., Sieker, L. C. & Jensen, L. H. (1978). *J. Mol. Biol.* **122**, 175–190.

Edited by R. Huber