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 $a = 54·63(2)$ Å, $b = 52·22(3)$ Å and $c = 42·62(3)$ 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

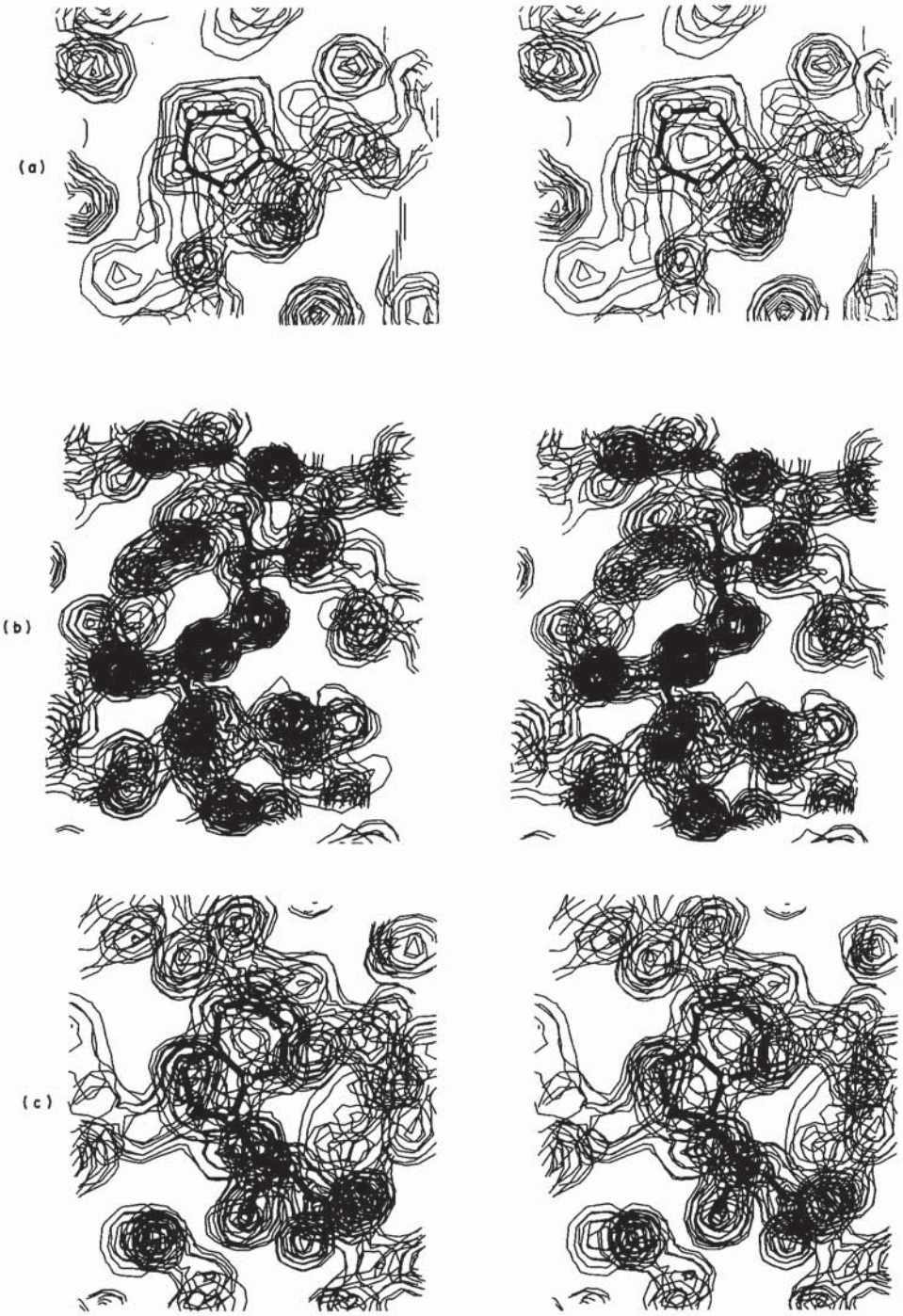

FIG. 1.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.