Acta Cryst. (1996). D**52**, 1058–1066

Pitfalls of Molecular Replacement: the Structure Determination of an Immunoglobulin Light-Chain Dimer

D.-B. HUANG,[a] C. AINSWORTH,[a] A. SOLOMON[b] AND M. SCHIFFER[a]*

[a]*Center for Mechanistic Biology and Biotechnology, Argonne National Laboratory, Argonne, Illinois 60439-4833, USA, and* [b]*Human Immunology and Cancer Program, Department of Medicine, University of Tennessee Medical Center/Graduate School of Medicine, Knoxville, Tennessee 37920, USA. E-mail: schiffer@anlcmb.bim.anl.gov*

(Received 6 February 1996; accepted 18 June 1996)

Abstract

The structure of protein Cle, a human light-chain dimer from the λIII subgroup, was determined using 2.6 Å data; the R value is 18.4%. The structure was solved, after a false start, by molecular replacement with the λII/V Mcg protein as a search structure. When the refinement did not proceed beyond an R value of 27%, it was discovered that while the constant domains were in their correct positions in the unit cell, the incorrect variable domains were used for defining the molecule. The correct solution required a rotation of 180° around the local twofold axis that relates the two constant domains of the dimer. The correct variable domain positions overlap about 70% of the same volume as the incorrect ones of a symmetry-related molecule. The refinement distorted the geometries of the domains. Though the constant domains were in their correct positions, the r.m.s. (root-mean-square) deviation of the Cα atom position was 1.2 Å when the two constant domains were compared. For the correct structure, this value is 0.5 Å. The φ and ψ angles, the r.m.s. chiral value and the free R value, even when calculated *a posteriori*, were good indicators of the correctness of the structure. The quaternary structure of the Cle molecule is similar to that in Mcg (crystallized from ammonium sulfate); the elbow bend is 115°. However, the arrangement of the variable domains differs from that observed in other variable domain dimers. The variable domains of Cle are 0.7 Å closer than in Mcg or variable dimer Rei. The hydrogen bonding at the interface of the two domains is novel. Residues Tyr36 from both monomers form a hydrogen bond that is part of a network with the Gln89 residues from both monomers. For the first time hydrogen bonds were observed between the main-chain peptide N and O atoms of the complementarity-determining region CDR2 and CDR3 segments of both monomers.

1. Introduction

Immunoglobulin light chains have an N-terminal variable (V) and C-terminal constant (C) domain of about 100 residues each. These two domains are connected by a flexible 'switch' peptide that allows one domain to move relative to the other. Three segments of the V domains are hypervariable and form the complementarity-determining regions (CDR's). Based on amino-acid sequences, κ and λ light chains have been identified. The λ light-chain dimers consist of two covalently linked light chains which are disulfide bonded at their penultimate residues. The angle between the local twofold axis of the V domain dimer and the local twofold axis of the C domain dimer defines the 'elbow' bend. The structures of λI light-chain dimer Loc (Chang *et al.*, 1985; Schiffer *et al.*, 1989; Huang, Ainsworth, Stevens & Schiffer, 1996) and λII/V light-chain dimer Mcg (Schiffer, Girling, Ely & Edmundson, 1973; Ely, Herron, Harker & Edmundson, 1989) have been previously determined; they were crystallized from several solvents. The Bence-Jones protein from patient Cle was characterized serologically and by its amino-acid sequence as a λIII light chain with Mcg$^-$Kern$^-$Oz$^-$ constant region (Eulitz, Murphy, Weiss & Solomon, 1991). It was excreted as the covalent dimer, and Cle protein formed casts in the kidney tubules of the patient (Myatt *et al.*, 1994). Light chains of the λIII subgroup represent about 43% of human λ chains (Solomon & Weiss, 1995). Protein Cle is the first λIII light-chain dimer to be studied by X-ray crystallographic techniques. The light chain in Fab fragment HIL (pdb8-fab.ent) also belongs to the λIII subgroup. The λIII proteins are of special interest because they differ from proteins of the other five λ-chain subgroups by having an unusually short first CDR region, as is found in the κI subgroup (Kabat, Wu, Perry, Gottesman & Foeller, 1991). Protein Cle also has a Gln at position 89, a residue well conserved in κ subgroups I, III and IV, but only predominant in λ subgroups III and VI (Kabat *et al.*, 1991). Residue 89 occurs at the interface of the V domains; in κI light chains such as Rei (Epp, Lattman, Schiffer, Huber & Palm, 1975) and Wat (Huang *et al.*, 1994) it participates in the dimer formation.

Molecular-replacement methods play an important role in the systematic study of the structures of light chains, antibodies and numerous other proteins. Molecular replacement for multidomain proteins has been

© 1996 International Union of Crystallography
Printed in Great Britain – all rights reserved

made substantially easier by the development in *X-PLOR* of Patterson correlation refinement and the 'elbow' function (Brünger, 1990, 1992). However, because the molecule consists of several identical domains and the inherent properties of the Patterson function, molecular replacement can give a reasonable but incorrect solution, as we found during this structure determination. We report here the structure determination of protein Cle by molecular replacement and the structure is described. The association of the V domains is different than previously observed in other light-chain dimers: they are held together by ten hydrogen bonds, and the domains are close together so a hydrogen bond can be formed between the main-chain segments of CDR2 and CDR3.

2. Materials and methods

2.1. *Crystallization and data collection*

Protein Cle was obtained from the urine of a patient who had multiple myeloma. As reported previously (Stevens, Westholm, Panagiotopoulos, Solomon & Schiffer, 1981), it was crystallized at pH 7.0 from 1.8 *M* ammonium sulfate in the orthorhombic space group $P2_12_12_1$ with $a = 113.17$, $b = 72.54$, $c = 49.25$ Å. The asymmetric unit consists of one dimer; the volume fraction of protein in the crystal is 0.57.

X-ray diffraction data were collected at room temperature on a multiwire area detector at Monsanto Company in St Louis. Two detectors were used simultaneously, one collecting the reflections which lie between infinity and 3.2 Å and the other an overlapping data set comprising reflections between 10.5 and 2.65 Å. The data were processed using a Fourier–Bessel scaling method (Weissman, 1982), the scaling *R* value was 4.7%. A data set consisting of 55 038 measurements was obtained, of which 11 373 reflections were unique. Between 10 and 2.65 Å 8684 reflections were above 2σ, representing 71% of the possible reflections.

2.2. *Naming conventions used*

The naming of light-chain domains followed the convention of Schiffer *et al.* (1973); V1 and C1, and V2 and C2 for the variable and constant domains of chain 1 and chain 2, respectively. Residues were numbered according to Kabat *et al.* (1991), 300 was added to the residue numbers of chain 2.

2.3. *Structure determination and refinement*

The structure was determined by molecular replacement using Patterson correlation (PC) refinement (Brünger, 1990) as implemented in *X-PLOR*, version 3.1 (Brünger, 1992) with λII/V light-chain dimer Mcg that was crystallized from ammonium sulfate, Mcg(AS), (1DCL; Xu & Schiffer, 1988) as the search model. The V domain of Cle differs from Mcg in 45 residues, and the CDR1 of Cle is three residues shorter than that of Mcg. PC refinement of the highest 6000 grid points of the rotation function was carried out using 15–4.0 Å resolution data. PC refinement revealed a solution with a correlation coefficient 1.23 times larger than the second solution. Translation searches using the method of Fujinaga & Read (1987) as implemented in *X-PLOR* were carried out with the correctly oriented and PC-refined Mcg dimer. The maximum peak of the translation function was 7.7σ above the mean and 3.0σ above the next-highest peak.

The orientation and position of the starting model were subsequently refined by rigid-body refinement, starting with the refinement of the whole molecule, followed by refinement of the two V and the two C domains, and finally by allowing each domain to move independently, resulting in an *R* value of 0.45 at 10.0–3.5 Å resolution. After the residues of Mcg were replaced by the residues of Cle, one round of simulated-annealing refinement using the slow-cooling method (Brünger, Krukowski & Erickson, 1990) reduced the *R* value to 0.30 for 10–2.65 Å resolution data. After this point, the structure did not improve. The *R* value stayed at 0.27 even after several manual adjustments based on $(2F_o - F_c)$ maps and repeated refinements with *X-PLOR* and *PROLSQ* (Hendrickson, 1985) were carried out. We noted that the geometries of the two variable domains deteriorated during the refinements and hydrogen bonds connecting β-sheets were broken. Some close contacts were observed between the V2 domain and the C2 domain in the molecular packing. At the beginning, we thought that the problems were probably from an incorrect 'elbow bend' in the starting model. In the light-chain dimers and Fab structures, the elbow bends can vary from 96 to 227° (Chang *et al.*, 1985; Wilson & Stanfield, 1994). Since in a light-chain dimer the two chains are equivalent, the elbow bend of the search model was modified in 10° steps from 90 to 180°. The modified models were refined by rigid-body refinements with *X-PLOR*, but the *R* value did not improve.

Finally, after examination of the structure, we came to the conclusion that the two constant domains were in the correct position but the two variable domains were in an incorrect position. The constant domains were related by the twofold screw axis along the 72.5 Å *b* axis of the unit cell; this packing arrangement of the constant domains is characteristic, observed in crystals of λ light-chain dimers (Schiffer, Chang & Stevens, 1985). We rotated the molecule 180° around the local twofold axis relating the two constant domains. This rotation moved the two variable domains into a new position, while the positions of the two constant domains did not change, as shown in Fig. 1. The rotated model was

easily refined to an R value of 0.20 with an overall temperature factor, using the same numbers of reflections and parameters used to develop the original erroneous model.

All manual adjustments were performed on an Evans & Sutherland ESV color graphics workstation supporting a version of program O (Jones, Zou, Cowan & Kjeldgaard, 1991) to display molecular fragments and electron density. Water molecules were identified from the $(2F_o - F_c)$ and $(F_o - F_c)$ maps. Electron density was accepted as a water molecule if it was within 4.0 Å of a likely hydrogen-bonding partner on the protein or another water molecule. A total of 123 positions were located as water molecule sites. The occupancies of the waters were refined with the same overall temperature factor applied for the protein atoms. The final occupancies of the waters ranged from 0.50 to 1.0. Of the 123 water molecule sites identified, 15 waters are in equivalent positions on the V domains, and 20 waters are on equivalent positions on the C domains. Ten waters act as bridge molecules that link the two monomers, of which five are located between the two variable domains and the other five between the two constant domains. 17 waters form hydrogen bonds with two symmetry-related molecules.

When the waters were included in the phase calculations, the electron-density maps were further improved. A few cycles of manual adjustment were made based on these density maps. Because of the resolution of the data and the relatively low number of reflections in the data set, only the overall temperature factor was refined. The number of waters included in the refinement was 123, less than one water per three amino-acid residues. At the final stage of the refinement, the R value was 0.184 for 8684 reflections with $F_o > 2\sigma(F_o)$ between 10.0 and 2.65 Å resolution. The mean positional error based on Luzzati plots (Luzzati, 1952) was 0.3 Å. The overall temperature factor for the protein and solvent molecules is 15.5 $Å^2$. The results of the refinement are shown in Table 1, and the R value as a function of resolution is shown in Table 2. The coordinates and structure factors have been deposited with the Protein Data Bank.*

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1LIL, R1LILSF). Free copies may be obtained through the Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: AM0041). At the request of the authors, the atomic coordinates and stucture factors will remain privileged until 13 May 1997.

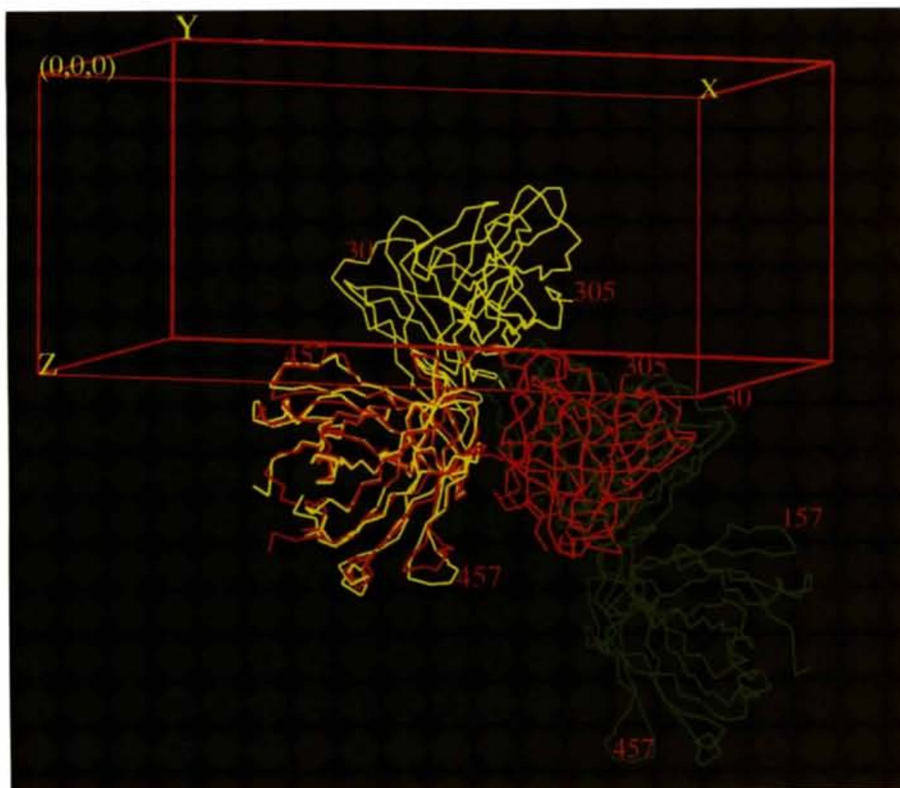


Fig. 1. Superposition of the correct dimer (in red) on two asymmetric units of the incorrect dimers (in yellow and green). The correct dimer was generated by a rotation around the local twofold axis that relates the two C domains. (300 has been added to the residue members of chain 2). The positions of the C domains in both models are the same except for the interchange of the C1 and C2 domains. The V dimer in the correct position superimposes on a V dimer of another asymmetric unit. The correct and incorrect V dimers overlap approximately 70% by volume.

Table 1. *Comparison of restrained least-squares refinement parameters in two models of Cle*

8684 reflections with $F_o > 2\sigma(F_o)$, between 10 and 2.65 Å resolution were used; the number of protein atoms is 3190.

	σ values		
	Incorrect model	Correct model*	Target
Diffraction data			$0.50\langle\|\|F_o\| - \|F_c\|\|\rangle$
Bonding distances (Å)			
Bond length	0.016	0.015	0.02
Angle-related distance	0.045	0.039	0.03
Intraplanar (1–4) distance	0.051	0.047	0.05
Deviation from plane (Å)	0.014	0.013	0.02
Chiral volume (Å^3)	0.24	0.169	0.15
Non-bonded contacts (Å)			
Single torsion contact	0.24	0.21	0.50
Multiple torsion contact	0.36	0.31	0.50
Possible hydrogen bond	0.31	0.28	0.50
Conformation torsion angle (°)			
Planar (ω)	2.6	2.3	3.0
Staggered	26.5	24.8	13.0
Orthonormal	40.4	27.0	20.0
R factor†	0.283	0.184	
Solvent waters		123	

* Refined final values. † $R = (\sum \|\|F_o\| - \|F_c\|\|)/\sum F_o$.

Table 2. *R value as a function of resolution*

With $F_o > 2\sigma(F_o)$.

Resolution (Å)	No. of observed	Percent of total	*R* (shell)
10.0–5.40	1127	85	0.188
5.40–4.40	1198	96	0.148
4.40–3.85	1242	93	0.164
3.85–3.50	1196	89	0.177
3.50–3.20	1397	84	0.199
3.20–2.95	1323	68	0.226
2.95–2.65	1201	35	0.234
10.0–2.65	8684	71	0.184

Free *R* value was calculated *a posteriori* (Brünger, 1996) and is shown in Table 3. A test set of 5% of the data selected randomly were removed from the refinement, and water molecules were not included in the coordinate set.

3. Results and discussion

3.1. *Structure determination: molecular replacement*

Comparisons of the course of refinement and the parameters from the restrained least-squares refinement for the correct and incorrect models are given in Tables 3 and 1, respectively. It can be seen from Table 2 that initially the refinement of the incorrect model appeared to proceed well. The *R* value of 27% and most of the refinement parameters for the incorrect model were acceptable. As shown in Table 1, the most notable difference of deviation in these two models is in their r.m.s. chiral volumes. It was 0.24 Å^3 in the incorrect model and 0.18 Å^3 in the correct one. An r.m.s. chiral volume of 0.24 Å^3 still can be considered to be in the acceptable range. The φ and ψ angles were the only values that suggested that the structure might be incorrect. Only 49% of the residues of the wrong structure had the φ and ψ angles in the most favoured region of the Ramachandran plots shown in Fig. 2, whereas 76% in the correct structure were in this region as given by *PROCHECK* (Laskowski, MacArthur, Moss & Thornton, 1993). The typical value for the percentage of residues in the most favoured regions is 74% for 2.65 Å structures. Free *R* value calculated *a posteriori* (Brünger, 1996) shown in Table 3; it clearly differentiated between the correct and incorrect structures.

Table 3. *The courses of refinement of the two models in X-PLOR*

Water molecules were not included in the refinements.

Refinement	Incorrect model	Correct model
R rigid body	0.45	0.38
R pre-stage	0.34	0.27
R slowcool	0.25	0.19
R free*	0.48	0.34
R.m.s. lengths (Å)	0.030	0.015
R.m.s. angles (°)	5.68	3.63

* Free $R = \sum_T \|F_o - F_c\| \sum_T \|F_o\|$, where T is a test set containing a randomly selected 5% of the observations omitted from the refinement.

The quarternary structures of the correct and incorrect models are comparable (Table 4). In both cases there is a local twofold axis between the two V domains and the two C domains, as expected. In the incorrect model, the local twofold relationship between the V domains is distorted: the rotational angle is 173° instead of 179°. The difference between the two models shows up in the differences between the structures of the equivalent domains within the molecule, as reflected in the r.m.s. deviation of Cα atom coordinates when these domains are superimposed. To compensate for the incorrect positioning of the V domains, the refinement distorted the structure within the constraints given. The distortion was worse, 2.40 Å compared with 0.77 Å, for the V domains that were in incorrect positions. The Cα distribution was also distorted in the C domains, 1.24 Å compared with 0.51 Å for the correct structure, though the C domains were in approximately correct positions. The r.m.s. deviation found for the V and C domains in the correct Cle structure is similar to that observed in three different crystal forms of the Loc protein (Huang *et al.*, 1996), determined with 2.3 Å resolution data. In another immunoglobulin structure, CD4, the r.m.s. deviation was 0.66 Å between the D1 domains determined in two crystal forms (Ryu, Truneh, Sweet & Hendrickson, 1994). The two V domains in protein Rei (Epp *et al.*, 1975) and in protein Wat (Huang *et al.*, 1994) and in protein M29b (Essen & Skerra, 1994) that were determined with higher, approximately 2 Å, resolution data are more similar to each other, the

r.m.s. deviation is 0.35, 0.36 and 0.35 Å, respectively. In general the cores of the domains with identical sequences became more similar with increased resolution of the diffraction data, and as the refinement progressed. The differences between equivalent domains can be observed in surface loops that have different crystal contacts. Based on our study, the structure of the Mcg protein, crystallized from distilled water (Ely *et al.*, 1989) probably should be re-examined; the r.m.s. deviation of the superimposed Cα-atom positions is 1.77 and 1.97 Å for the V and C domains, respectively, and the chiral volume is above 0.3 Å^3. Kleywegt (1996) tabulated the r.m.s. deviation of Cα-atom coordinates on comparing identical domains related by non-crystallographic symmetry; most of the values were below 1 Å and the values above 1.5 Å were clearly outliers. We did not use non-crystallographic symmetry restraints; their use might have alerted us to problems with our model earlier.

Table 4. *Comparison of the geometry of the individual domains for the correct and incorrect models of the Cle structure*

	Standard deviation (Å)	Rotation angle (°)	Rot-trans* angle (°)	C.m.† distance (Å)	Elbow bend (°)
Incorrect model					
V1—V2	2.40	173.2	90.9	22.8	112.7
C1—C2	1.24	179.2	89.7	15.7	
Correct model					
V1—V2	0.77	179.3	90.7	23.1	115.0
C1—C2	0.51	180.0	89.6	15.4	

* Angle between rotation axis and the vector that connects the center of mass of the two domains. † Center of mass.

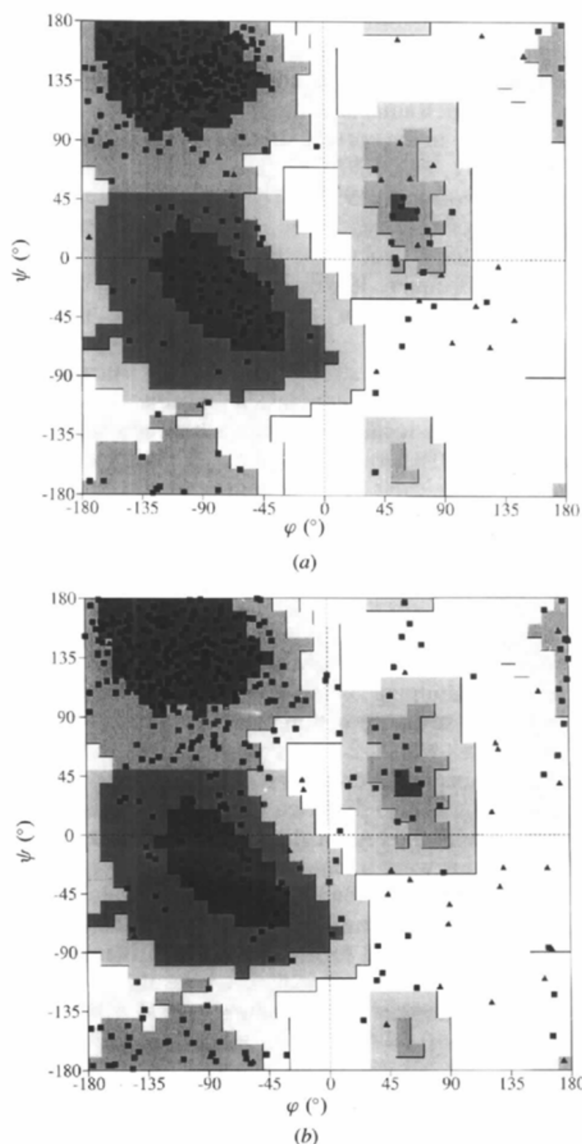


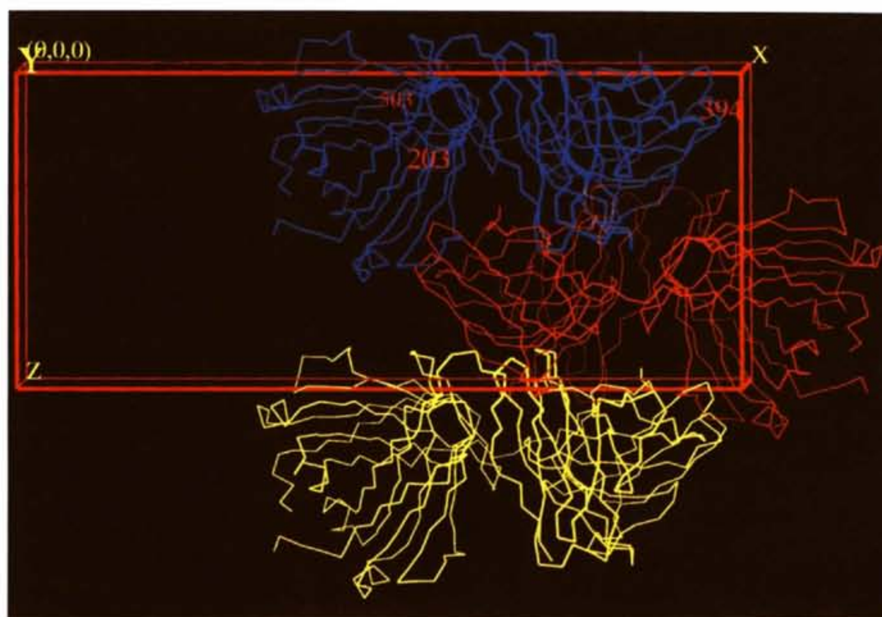
Fig. 2. Ramachandran plots of the (*a*) correct model and (*b*) the incorrect model of protein Cle calculated by *PROCHECK* (Laskowski *et al.*, 1993).

Crystal packing was reasonable for both models (Fig. 3), though in the incorrect model a few bad contacts exist between the V2 domain and a symmetry-related C2 domain. The *b* cell dimension of 72.54 Å of the Cle crystal is similar to that in other λ light-chain dimers. This results from similar packing of adjacent constant domains along the *y* axis direction, as previously described (Schiffer *et al.*, 1985), where the constant domains are related by crystallographic twofold screw axes. The partial refinement was possible because, in the wrong structure, the C domains were in their correct position and the V domains of another asymmetric unit partially (about 70%) overlapped with the correct solution.

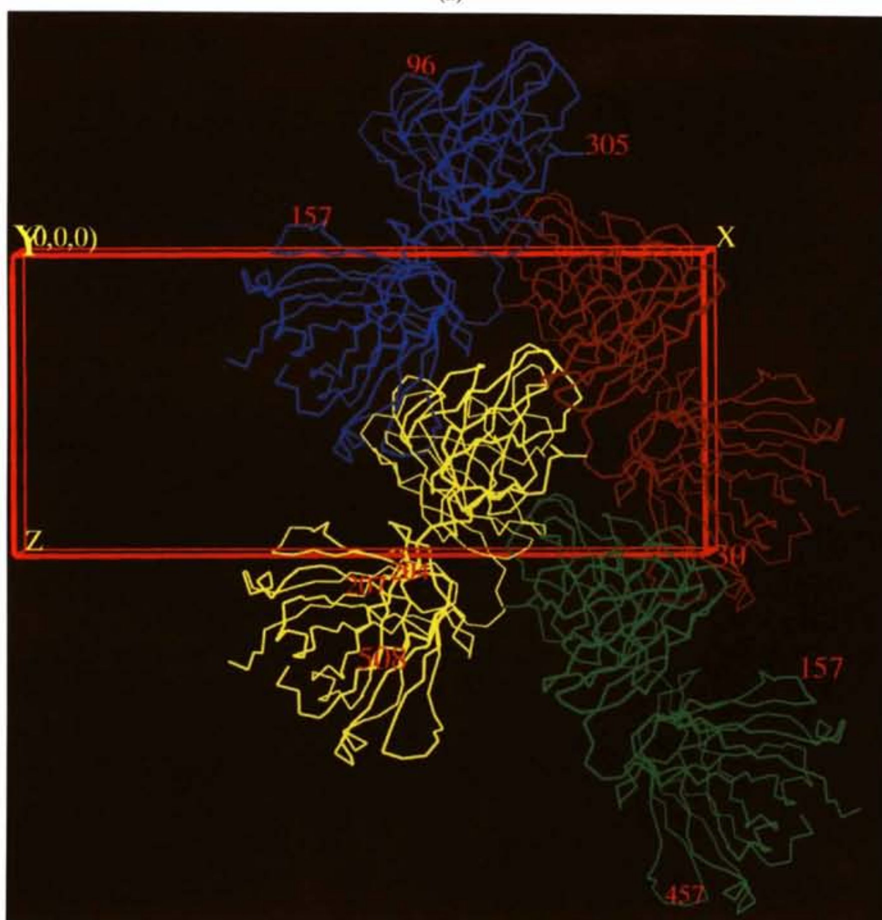
After finding the correct orientation for the model, we checked the rotation function, it did not have the correct rotation peak. Previously, rotation searches were also carried out with the program *MERLOT* (Fitzgerald, 1988); as search structures Mcg (AS), three different structures of protein Loc and the constant domain dimer were used, but no clear solution was obtained. Finally a direct rotation search (Brünger, 1992) with Mcg (AS) as the search structure, the structure used in the original search had the correct peak as its highest peak.

3.2. *Description of the molecule*

The amino-acid sequence for the V domain (Eulitz *et al.*, 1991) was corrected and additional residues that were missing in the CDR3 region were identified (M. Eulitz, personal communication). Residue 79 was found to be Gln instead of Val, and residue 91 was identified as Trp instead of Ala. The electron density for residue 91 found in both monomers was consistent with Trp, though originally they were refined as Ala residues. Residues 98 to 103 were identified by amino-acid



(a)



(b)

Fig. 3. Packing of Cle molecules in the unit cell, illustrating the packing interactions that are reasonable in both the correct (a) and incorrect (b) orientation of the molecules.

sequencing as Phe, Gly, Gly, Gly, Thr and Lys, homologous to other λIII-type proteins. The remaining six residues could not be sequenced; these residues were identified from the electron density in homology with other sequences (Kabat *et al.*, 1991). These are: 93 Ser (most frequently found in that position), 94 Asn or Asp (hydrogen-bonding pattern suggests that the residue is Asn and not Asp since it forms a hydrogen bond with the carbonyl O atom of residue 92), 95 Ala, 95A Ser, 96 Val, and 97 Val (Val is the most frequent residue observed at positions 96 and 97).

The tertiary structures of the individual variable and constant domains of protein Cle are very similar to those of other λ light chains. The arrangement of the domains in the Cle molecule is similar to that in Mcg (from ammonium sulfate). The elbow bend in Cle is 115°, compared with 113° in Mcg(AS). The relative positions of the C domains are like those observed in other λ light chains. Comparison of the relative positions of the V domains in the crystals of Cle, Mcg, and Rei is shown in Table 4. The V domains in all three proteins are related by local twofold axes. Though the Mcg is a complete light chain (Schiffer *et al.*, 1973) and the Rei protein is a V domain dimer (Epp *et al.*, 1975), the distance between the V domains is essentially the same (24.6 and 24.7 Å, respectively), but the V domains of the Cle protein are 0.7 Å closer to each other than they are in Mcg. The r.m.s deviations between equivalent Cα-atom positions within each molecule reflect the resolution of the data. The cross comparison between the V domains from the different proteins using the identical 80 Cα atoms shows that the domains have highly homologous structures (see Table 5). The buried

Table 5. *Comparison of the homologous domains of Cle, Mcg(AS) and Rei*

80 Cα atoms were included in the calculation. These residues are 5–7, 11–24, 33–38, 42–56, 58–66, 69–91, 96–105 (using the numbering convention of Kabat *et al.*, 1991).

V1—V2	R.m.s. deviation (Å)	Rotation angle (°)	Rot-trans* angle (°)	C.m.† distance (Å)	Buried areas (Å^2)
Cle	0.76	178.8	90.3	23.9	1523
Mcg(AS)	0.49	180.0	89.0	24.6	1444
Rei	0.31	179.8	89.0	24.7	1498

Cross comparisons of domains‡

	Rei		Cle	
	V1—V1	V2—V2	V1—V1	V2—V2
Mcg(AS)	0.71	0.86	0.88	0.67
Rei	—	—	0.87	0.87

* Angle between rotation axis and the vector that connects the center of mass of the two domains. † Center of mass. ‡ R.m.s. deviation is given.

surface area (Lee & Richards, 1971) between the V domains is similar, approximately 1500 Å^2. The effect of the solvent on the structures should be similar since all three proteins were crystallized from ammonium sulfate. The superposition of the V dimers of Cle, Mcg, and Rei is shown in Fig. 3. Though essentially the same face for each individual domain is involved in dimerization, the mode of dimerization is different in Cle relative to Mcg and Rei. The Mcg and Rei V dimers are similar: 3.5° rotation and 1.2 Å translation were required to superimpose domain 1's after domain 2's were superimposed. Domain 2 of Cle can be superimposed on Mcg by 11.4° rotation and 1.4 Å translation and on Rei by 14.0° rotation and 1.4 Å

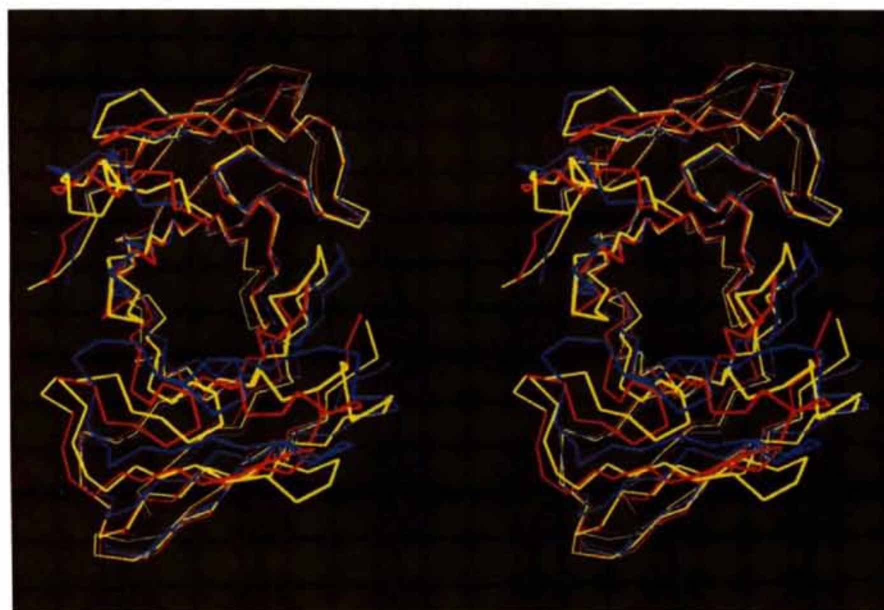


Fig. 4. Superposition of the Cle (cyan), Mcg (yellow), and Rei (red) (V) domain dimers by superposing domain 2 of each protein. The Mcg and Rei dimers are more similar to each other than they are to Cle.

Table 6. *Hydrogen bonds between the V domains in the crystal structures of Cle, Mcg(AS) and Rei (atom–atom distance < 3.5 Å)*

Cle				Mcg(AS)				Rei			
V1		V2	(Å)	V1		V2	(Å)	V1		V2	(Å)
Tyr36	OH···OH	336Tyr	3.09								
Tyr36	OH···OE1	389Gln	3.13					Tyr36	OH···OE1	389Gln	3.19
Gln89	OE1···OH	336Tyr	2.52								
Gln38	OE1···NE2	338Gln	3.47	Gln38	OE1···NE2	338Gln	3.38	Gln38	OE1···NE2	338Gln	3.06
Gln38	NE2···OE1	338Gln	3.02	Gln38	NE2···OE1	338Gln	2.58	Gln38	NE2···OE1	338Gln	2.89
				Tyr87	OH···O	342Lys	3.40	Lys42	O···OH	387Tyr	3.42
Ser56	N···O	395Ala	3.29	Pro55	O···ND2	396Asn	2.84				
Ala95	O···N	356Ser	3.02	Asp95	OD2···N	356Ser	3.22				
Ser56	OG···O	395Ala	3.28	Asp95	OD1···OG	356Ser	3.46				
Ser56	OG···O	394Asn	2.80	Try49	OH···OD2	395Asp	2.86				
Ser56	OG···N	396Ser	2.84	Ser94	O···OH	349Tyr	2.84				

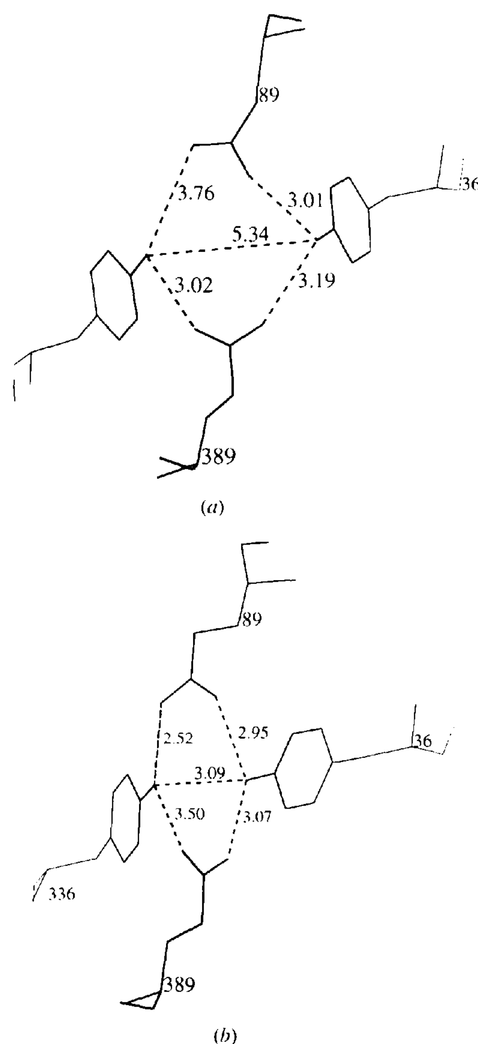


Fig. 5. The arrangements of identical residues Tyr36 and Gln89 from both monomers in the Rei protein (*a*) and in the Cle protein (*b*). In Rei, these residues form a hydrogen-bonded square, where the two Tyr residues are more than 5 Å apart, while in Cle the 'square' is distorted and the two Tyr residues are 3.1 Å apart. In the Cle protein the centers of mass of the two V domains are 0.8 Å closer than they are in Rei. In addition, a twist in the domains relative to Rei makes the above arrangement of hydrogen bonds possible.

translation. [For the above comparisons, the method and residues designated by Colman *et al.* (1987) were used.]

The hydrogen bonds between the variable domains are listed in Table 6. There are four hydrogen bonds with a distance less than 3.5 Å between the two Rei domains, eight between the Mcg domains, and ten between Cle domains. The hydrogen bonds formed between the Gln residues 38 from each domain are conserved in all three structures. This is indeed a conserved feature of light-chain dimers and Fab fragments. Mcg has an Ala residue at position 89, but both Rei and Cle have Gln at position 89. A square-like arrangement of hydrogen bonds with Tyr36, previously seen in the Rei structure (Epp *et al.*, 1975), was expected in Cle (see Fig. 4*a*). Instead (as shown in Fig. 4*b*) in Cle the two Tyr residues come into hydrogen-bonding distance, forming a strongly hydrogen-bonded network of residues. To achieve this hydrogen-bonding pattern, the two V domains in Cle had to move closer together and the relative orientation of the domains had to change. The relative positions of the domains in Cle also resulted in hydrogen bonds between the peptide N atoms of residue 56 (located in CDR2) and the peptide O atom of residue 95 (located in CDR3) of both monomers (see Fig. 5). This is the first time that a main-chain-to-main-chain hydrogen bond between CDR's has been observed in a light-chain dimer.

4. Conclusions

The identity of residues that determine the arrangement of the Cle V domains, relative to the V domain arrangement in both Mcg and Rei, is not clear. Cle has some residues in common with both the Mcg and the Rei protein. It is a λ-type protein like Mcg, but since it is in the λIII subgroup, it has a Gln residue at position 89. Clearly in domain–domain interaction, both packing forces and polar interactions have to be favorable; the observed structures must represent the lowest energy form of the system. Detailed comparative studies of the three structures will be required to understand the

Fig. 6. Stereo figure of the V domain dimer of Cle. Residues 36, 38, 89 which form hydrogen bonds between the V domains of protein Cle and also obey the local twofold symmetry between the domains are shown. The hydrogen bond between the CDR2 residue 56 carbonyl O atom and the other domain's CDR3 residue 95 amide N atom is observed for the first time in light-chain dimers. The two domains are in an orientation that allows the formation of these bonds.

function of the residues that determine the structures of the V dimers in each protein.

The structure determination of protein Cle shows that molecular replacement can give a reasonable but incorrect solution for this multisubunit protein. In a multidomain protein connected with long flexible hinges, especially when the domains are homologous and the fraction of the unit cell occupied by protein is high, molecular replacement can identify a structure with incorrect connectivity. The φ and ψ angles and (to lesser extent) the chiral volume are good indicators of the correctness of the structure as well as the free R, even calculated *a posteriori*. For immunoglobulins, where it has been shown that the basic core domain structure is maintained, comparison of the arrangement of the $C\alpha$ atoms of equivalent domains is also useful as an index of refinement.

We thank Chong-Hwan Chang, Huey-Sheng Shieh for collecting the diffraction data; Fred J. Stevens and Andrzej Joachimiak for critically reading the manuscript. This work is supported by the US Department of Energy, Office of Health and Environmental Research, under Contract No. W-31-109-ENG-38, and by the US Public Health Service Grants DK 43757and CA10056. This research was also supported in part by a grant from the Pittsburgh Supercomputing Center through the NIH National Center for Research Resources cooperative agreement. AS is an American Cancer Society Research Professor.

References

Brünger, A. T. (1990). *Acta Cryst.* A**46**, 46–57.

Brünger, A. T. (1992). *X-PLOR, Version 3.1. A System for X-ray Crystallography and NMR.* New Haven: Yale University Press.

Brünger, A. T. (1996). *Methods Enzymol.* In the press.

Brünger, A. T., Krukowski, A. & Erickson, F. W. (1990). *Acta Cryst.* A**46**, 585–593.

Chang, C.-H., Short, M. T., Westholm, F. A., Stevens, F. J., Wang, B.-C., Furey, W. Jr, Solomon, A. & Schiffer, M. (1985). *Biochemistry*, **24**, 4890–4897.

Colman, P. M., Laver, W. G., Varghese, J. H., Baker, A. T., Tulloch, P. A., Air, G. M. & Webster, R. G. (1987). *Nature (London)*, **326**, 358–363.

Ely, K. R., Herron, J. N., Harker, M. & Edmundson, A. B. (1989). *J. Mol. Biol.* **210**, 601–615.

Epp, O., Lattman, E. E., Schiffer, M., Huber, R. & Palm, W. (1975). *Biochemistry*, **14**, 4943–4952.

Essen, L.-O. & Skerra, A. (1994). *J. Mol. Biol.* **238**, 226–244.

Eulitz, M., Murphy, C., Weiss, D. T. & Solomon, A. (1991). *J. Immunol.* **146**, 3091–3096.

Fitzgerald, P. M. D. (1988). *J. Appl. Cryst.* **21**, 273–278.

Fujinaga, M. & Read, R. J. (1987). *J. Appl. Cryst.* **20**, 517–521.

Hendrickson, W. (1985). *Methods Enzymol.* **115**, 252–270.

Huang, D.-B., Ainsworth, C., Stevens, F. J. & Schiffer, M. (1996). *Proc. Natl Acad. Sci. USA*, **93**. 7017–7021.

Huang, D.-B., Chang, C.-H., Ainsworth, C., Brünger, A. T., Eulitz, M., Solomon, A., Stevens, F. J. & Schiffer, M. (1994). *Biochemistry*, **33**, 14848–14857.

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

Kabat, E. A., Wu, T T., Perry, H. M., Gottesman, K. S. & Foeller, C. (1991). Editors. *Sequences of Proteins of Immunological Interest*, 5th ed. Bethesda, MD: National Institutes of Health.

Kleywegt, G. J. (1996). *Acta Cryst.* D**52**, 842–857.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lee, B. & Richards, F. M. (1971). *J. Mol. Biol.* **55**, 379–400.

Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.

Myatt, E. A., Westholm, F. A., Weiss, D. T., Solomon, A., Schiffer, M. & Stevens, F. J. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 3034–3038.

Ryu, S.-E., Truneh, A., Sweet, R. W. & Hendrickson, W. A. (1994). *Structure*, **2**, 59–74.

Schiffer, M., Ainsworth, C., Xu, Z.-B., Carperos, W., Olsen, K., Solomon, A., Stevens, F. J. & Chang, C.-H. (1989). *Biochemistry*, **28**, 4066–4072.

Schiffer, M., Chang, C.-H. & Stevens, F. J. (1985). *J. Mol. Biol.* **186**, 475–478.

Schiffer, M., Girling, R. L., Ely, K. R. & Edmundson, A. B. (1973). *Biochemistry*, **12**, 4620–4631.

Solomon, A. & Weiss, D. T. (1995). *Clin. Diagn. Lab. Immunol.* **2**, 387–394.

Stevens, F. J., Westholm, F. A., Panagiotopoulos, N., Solomon, A. & Schiffer, M. (1981). *J. Mol. Biol.* **147**, 179–183.

Weissman, L. (1982). *Computing in Crystallography, International Summer School*, edited by D. Sayer, pp. 56–63. Oxford University Press.

Wilson, I. A. & Stanfield, R. L. (1994). *Curr. Opin. Struct. Biol.* **4**, 857–867.

Xu, Z.-B. & Schiffer, M. (1988). *Am. Crystallogr. Assoc. Meet. Abstracts*, p. 74.