J. Mol. Biol. (1986) **190**, 593–604

Phosphocholine Binding Immunoglobulin Fab McPC603

An X-ray Diffraction Study at 2·7 Å

Yoshinori Satow†, Gerson H. Cohen, Eduardo A. Padlan and David R. Davies

Laboratory of Molecular Biology
National Institute of Arthritis, Diabetes and Digestive and Kidney Diseases
National Institutes of Health, Bethesda, MD 20892, U.S.A.

(Received 27 December 1985)

The crystal structure of the Fab of McPC603, a phosphocholine-binding mouse myeloma protein, has been refined at 2·7 Å resolution by a combination of restrained least-squares refinement and molecular modeling. The overall structure remains as previously reported, with an elbow bend angle between the variable and constant modules of 133°. Some adjustments have been made in the structure of the loops as a result of the refinement. The hypervariable loops are all visible in the electron density map with the exception of three residues in the first hypervariable loop of the light chain. A sulfate ion occupies the site of binding of the phosphate moiety of phosphocholine.

1. Introduction

Direct information about the three-dimensional structure of the antibody combining site is based on the crystal structures of three Fab species: Kol, a human myeloma protein with unknown specificity (Marquart *et al.*, 1980); New, a human myeloma protein that binds to a vitamin K1 derivative (Amzel *et al.*, 1974; Saul *et al.*, 1978); and McPC603, a mouse plasmacytoma protein specific for phosphocholine (Segal *et al.*, 1974). Although the structures of several monoclonal antibodies with known antigen binding specificities are being investigated (Mariuzza *et al.*, 1983; Silverton *et al.*, 1984; Colman *et al.*, 1981; Gibson *et al.*, 1985; Amit *et al.*, 1985), refined, high-resolution structures from them are not available.

The three-dimensional structure for the Fab of McPC603 (IgA,κ) protein that was previously reported (Segal *et al.*, 1974) was an unrefined structure based on 3·1 Å diffractometer data and with only partial sequence information. This work has now been extended through the collection of a complete set of 2·7 Å diffraction data using oscillation photography, augmented by the knowledge of the complete amino acid sequence and with the help of interactive molecular graphics procedures (Lipscomb *et al.*, 1981; Diamond, 1981). The crystal structure has been refined using restrained least-squares procedures (Hendrickson & Konnert, 1981). In this paper, we report the results of this investigation.

2. Materials and Methods

(a) *Crystal structure analysis and preliminary refinement at 3·1 Å resolution*

Crystals were prepared from concentrated solutions of ammonium sulfate as described (Rudikoff *et al.*, 1972). After preparation, the crystals were transferred into stabilizing solutions consisting of 50% saturated ammonium sulfate (pH 7·0), 0·2 M-imidazole or 0·1 M-sodium cacodylate. Isomorphous heavy-atom derivatives were prepared using $TmCl_3$ (American Potash & Chemical Corp.) $K_2Pt(CNS)_6$ (K&K Laboratories, Inc.), and KI (Fisher Scientific Co.). The thulium and platinum derivatives were prepared by soaking crystals in solutions containing cacodylate and 30 mM-$TmCl_3$ or 0·4 mM-$K_2Pt(CNS)_6$, respectively, for 3 to 4 weeks. An iodine derivative was prepared by soaking crystals in 50 mM-KI, 2·5 mM-Chloramine T (Eastman Kodak Co.) for 2 weeks, after which the iodinated crystals were washed with stabilizing solution to remove excess iodine. Double and triple derivatives were prepared using these heavy-atom compounds. The native and iodinated crystals were prepared using imidazole buffer; all other derivatives were prepared in cacodylate buffer.

Most of the 3·1 Å data were collected using a Picker FACS-I diffractometer (Segal *et al.*, 1974); the $TmCl_3$ derivative data set was collected from precession photographs. Co-ordinates for the heavy-atom sites in the thulium and platinum derivatives were obtained from difference Patterson syntheses at 4·5 Å resolution (Padlan *et al.*, 1973). Those for the iodine derivative were obtained from a difference Fourier synthesis with phases computed

† Present address: Photon Factory, National Laboratory for High Energy Physics, Oho-machi,

atom refinement and phase calculation (Dickerson *et al.*, 1961) were then computed using local versions of the programs of Busing *et al.* (1962) and Matthews (1966). Table 1 shows the refined heavy-atom parameters. A "best" Fourier synthesis (Blow & Crick, 1959) was computed and a Kendrew skeletal model was fitted to the electron density using an optical comparator (Richards, 1968). The model was improved using the computer graphics system GRIP at the Department of Computer Science, University of North Carolina (Tsernoglou *et al.*, 1977) and was then subjected to restrained least-squares refinement (Konnert, 1976; Hendrickson & Konnert, 1981) on the TI-ASC computer at the U.S. Naval Research Laboratory. The initial value of the conventional crystallographic R-factor was 0·41. This was reduced to 0·30 after 5 cycles with a single overall temperature factor and tight structural restraints. Further refinement using individual temperature factors for the atoms reduced the residual to 0·27. The r.m.s.† total shift from the original position was 0·76 Å. At this point, the fit of the model to a $2F_o-F_c$ map was examined using BILDER (Diamond, 1981), as implemented on a PDP 11/70 under the RSX-11M operating system (G. H. Cohen, unpublished results). In the second stage of refinement, the protein geometry was initially less restrained and subsequently tightened, yielding a final residual of 0·24. The r.m.s. total shift from the atomic positions at the start of the second stage of refinement was 0·42 Å. A sulfate ion, that had been located in the hapten binding cavity (Padlan *et al.*, 1973; Segal *et al.*, 1974), was included throughout the refinement.

The regions corresponding to residues 101 to 108 in the heavy chain and residues 31 to 35 in the light chain were not clearly defined in the original electron density function based on the heavy-atom phases. These regions remained poorly defined after the preliminary refinement.

(b) *Crystal structure analysis and refinement at 2·7 Å resolution*

A new set of crystals was prepared for the higher resolution, 2·7 Å, phase of this study. The imidazole buffer employed during crystallization was replaced by cacodylate when the crystals were transferred to stabilizing solutions. Intensity data to 2·7 Å were collected by rotation photography (Arndt & Wonacott, 1977) with Kodak No-Screen Medical X-ray films (3 in a pack) using Ni-filtered CuKα radiation from an Elliott GX-6 rotating anode X-ray generator operated at 40 kV and 40 mA. A Franks double bent mirror system (Harrison, 1968) purchased from Brandeis University was used to focus the X-ray beam. Data were recorded on an Enraf-Nonius Arndt–Wonacott rotation camera with a nominal 87 mm crystal-to-film distance. The data consisting of 88 film packs were collected from 26 crystals by oscillation about the 2 crystal axes, b and c, with an oscillation range of 1·25° and an overlap of 0·25°. The time for each exposure was 14 to 21 h.

Table 1
Heavy-atom parameters

Heavy-atom compound	Co-ordinates x	y	z	Occupancy†	Thermal factor (Å^2)
$K_2Pt(CNS)_6$	0·3807	0·3936	0·5422	0·0218	19·6
$TmCl_3$	0·3427	0·6617	0·2523	0·0245	34·8
Iodine-1	0·2336	0·7083	0·5569	0·0119	28·3
Iodine-2	0·2409	0·6794	0·5695	0·0184	7·5
Iodine-3	0·2328	0·6328	0·6404	0·0125	6·7
Iodine-4	0·1241	0·7017	0·4838	0·0044	6·8

† The site occupancy is on an arbitrary scale in which the average structure amplitude of the native protein is 14·3.

The films were scanned at 100 μm steps on an Optronics P-1000 film scanner. The initial film processing, which included preliminary refinement of the crystal orientations and successive evaluation of the integrated intensities, was made on a PDP 11/70 computer using the rotation program package written by G. Cornick & M. A. Navia (unpublished results). Intensities from each pack were further processed through intra-film-pack scaling, post refinement, scaling and averaging using programs specially written for these purposes (Y. Satow, unpublished results). Intensities in a single pack were scaled by refining non-linear response correction factors (Matthews *et al.*, 1972) and absorption factors for the film base and emulsion, then were corrected for Lorentz and polarization factors. The measured intensities from the films were merged, reduced to a unique set and processed by a scaling and averaging program, which refines relative scale and exponential fall-off factors for individual films. This program follows closely the formalism of the established algorithms of Hamilton *et al.* (1965) and Rossmann *et al.* (1979). Post refinement of the crystal orientations, lattice constants and rocking curve parameters were done as proposed by Winkler *et al.* (1979) and Rossmann *et al.* (1979). The total number of 147,706 intensities, including partially recorded reflections, were finally scaled and symmetry averaged, yielding 24,235 unique reflections for the resolution range of 10 Å to 2·7 Å. The agreement factor:

$$R = \sum_h \sum_j^{N_h} |I_{hj} - \bar{I}_h| / \sum_h N_h \bar{I}_h,$$

where the intensity I_{hj} for reflection h was measured N_h times, was 0·077 or 0·054 for the structure factor amplitudes F_{hj}. Within a 3 Å sphere, 95% of the data were retained; within the complete 2·7 Å sphere, 93% were retained. The earlier diffractometer data were not included in this final set of 2·7 Å film data.

The initial model for the 2·7 Å work was taken from the 3·1 Å results. It was subjected to a number of cycles of least-squares refinement (Hendrickson & Konnert, 1981) with periodic examination and rebuilding of the model using BILDER on a VAX 11/780 (R. Ladner, unpublished results). The sulfate ion located in the 3·1 Å analysis was not included in the early stages of least-squares refinement. With an overall temperature factor, the initial value of the R-factor was 0·41. After 5 cycles of positional parameter refinement with an overall temperature factor, the value was reduced to 0·33. Individual temperature factors for the atoms were used in the succeeding cycles. The program restrains the values of

† Abbreviations used: r.m.s., root-mean-square; F_o, observed structure factor amplitude; F_c, calculated structure factor amplitude; w, weight; CDR, complementarity determining region (Kabat *et al.*, 1983); VL, light chain variable domain; VH, heavy chain variable domain; CL, light chain constant domain; CH1, first constant domain of heavy chain; m.i.r., multiple isomorphous replacement; L1, L2 and L3, 1st, 2nd and 3rd CDR of the light chain; H1, H2 and H3, 1st, 2nd and 3rd CDR of the heavy chain.

the *B*-factors so that each is influenced by the *B*-factors of the atoms to which it is bound as well as the atoms 1 removed along a chain. In the final stages of least-squares refinement, particular care was taken to ensure that the stereochemistry was kept reasonable and that $w(|F_o|-|F_c|)^2$ was approximately constant over the range of the data used (8·0 Å through 2·7 Å).

The main-chain stereochemistry was constantly monitored by the program GEOM (G. H. Cohen, unpublished results) and points of significant departure from expected stereochemistry were examined and corrected *via* interactive computer graphics. In addition to a sulfate ion, 138 water molecules of variable occupancy were identified from examination of ΔF and $2F_o-F_c$ maps and refined with the protein molecule.

As sequence data became available, the "working" sequence was updated appropriately. The VL and VH sequences are listed (sequences 12, p. 45 and 1, p. 128, respectively) by Kabat *et al.* (1983), quoted from Rudikoff *et al.* (1981) and Rudikoff & Potter (1974), with the addition of the tetrapeptide Leu-Glu-Ile-Lys, which occurs at the end of VL (Rudikoff, unpublished results). The CH1 sequence given by Auffray *et al.* (1981), obtained by translation of the nucleotide sequence of cDNA complementary to alpha-chain mRNA from J558 tumor cells, (sequence 60, p. 175, Kabat *et al.*, 1983) was used. This sequence differs in 4 places from that obtained by Tucker *et al.* (1981) from BALB/c genomic DNA, while they both differ in 2 of these positions from the sequence reported by Robinson & Appella (1980) for MOPC511, as quoted by Kabat *et al.* (1983). Examination of the final electron density map at these positions did not permit a clear distinction to be made between these sequences. The sequence of MOPC21 (sequence 23, p. 167, Kabat *et al.*, 1983 from Svasti & Milstein, 1972) was used for the CL domain of McPC603.

Throughout this paper, the amino acid numbering is serial, starting from number 1 with the 1st residue of each chain of the molecule. The correspondence between our numbering scheme and that of Kabat *et al.* (1983) is presented in Table 2.

As noted by Segal *et al.* (1974), the molecule possesses 2 approximate local dyad symmetry axes, which relate the pair of variable domains and the pair of constant domains to each other. We examined the relationship and the similarity of these pairs of domains using the program ALIGN (G. H. Cohen, unpublished results), which iteratively rotates one set of atoms to another set to optimize their fit while preserving the order of the linear sequences of the 2 sets. The program uses the algorithm of Needleman & Wunsch (1970) to identify the structural homology while accounting for insertions and deletions.

Interdomain and intermolecular contacts were calculated with the aid of the program CONTAX (E. A. Padlan, unpublished results). Two atoms are defined to be in contact if their co-ordinates lie within the sum of their van der Waals radii plus 1·0 Å. Intramolecular main-chain hydrogen bonds were calculated by the program EREF (M. Levitt, personal communication).

Table 2

Correspondence between the numbering scheme used here and that of Kabat et al. *(1983)*

Light chain		Heavy chain	
This work	Kabat *et al.*	This work	Kabat *et al.*
1–27	1–27	1–52	1–52
28–33	27a–27f	53–55	52a–52c
34–220	28–214	56–85	53–82
		86–88	82a–82c
		89–106	83–100
		107–109	100a–100c
		110–139	101–130
		140–142	133–135
		143–163	137–157
		164–171	162–169
		172–184	171–183
		185–200	185–200
		201–205	202–206
		206–216	208–218
		217–222	220–225

3. Results and Discussion

Table 3 shows an estimate of the quality of the stereochemical parameters for the final model. It is expressed in terms of the r.m.s. deviations of the various classes of parameters from accepted values (Sielecki *et al.*, 1979). The ϕ,ψ plot for the main chain is shown in Figure 1. There are a few residues that have "forbidden" ϕ,ψ values. The quality of the map in these regions together with the low

Table 3

Summary of stereochemical criteria (Hendrickson & Konnert, 1981; Sielecki et al., *1979; Cohen* et al., *1981)*

	Final model	Target σ
$R=\frac{\sum\|\|F_o\|-\|F_c\|\|}{\sum\|F_o\|}$	0·225	
Average ΔF	136	†
Interatomic distances (Å)		
1–2	0·020	0·015
1–3	0·040	0·020
1–4	0·037	0·025
Planarity (Å)	0·027	0·020
Chiral volumes (Å^3)	0·257	0·150
Non-bonded contacts (Å)		
1–4	0·24	0·50
Other	0·34	0·50
Angles (deg.)		
ω, χ_5 (Arg)	27·0	15·0
$\chi_1, \cdots, \chi_4$	5·0	5·0
Temperature factors (A^2)		
Main chain 1–2	0·5	0·5
Main chain 1–3	1·0	0·7
Side chain 1–2	0·4	0·5
Side chain 1–3	0·7	0·7

The standard groups dictionary used is specified in Table 2 of Sielecki *et al.* (1979). In this Table, F_o refers to the observed structure factor, F_c is the calculated structure factor and ΔF is the quantity $\|F_o|-|F_c\|$. The target σ represents the inverse square-root of the weights used for the parameters listed. The values given are the r.m.s. deviations from the respective ideal values.

† The weight chosen for the structure factor refinement, the "target σ" of ΔF, was modeled by the function $w=(1/d)^2$ with:

$$d=40-500(\sin(\theta)/\lambda-1/6).$$

By this means, $w(\Delta F)$ was approximately constant over all of the data (w representing the weight for a given reflection). Other details of this Table are explained fully by Cohen *et al.* (1981).

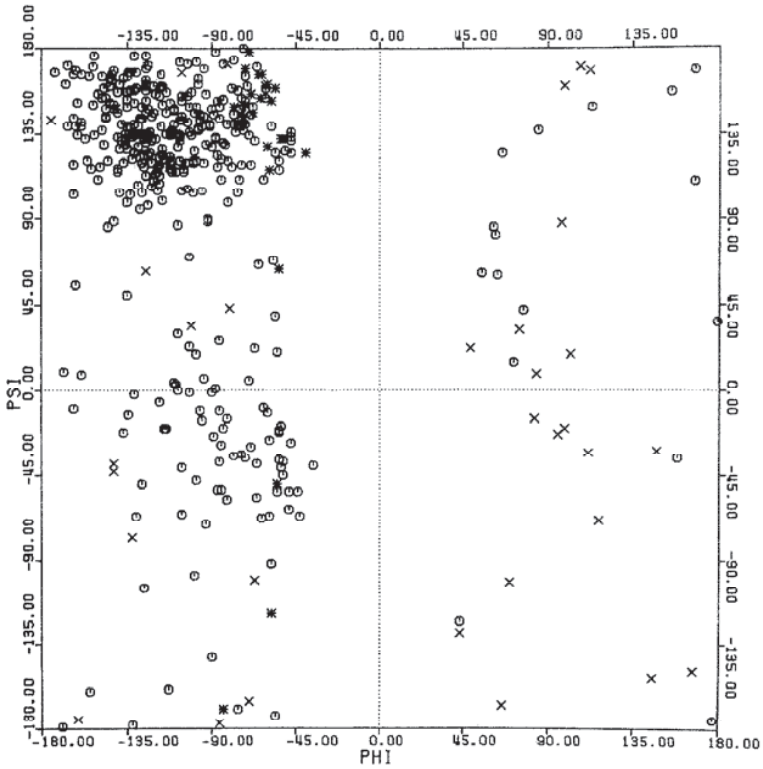

Figure 1. A plot of the dihedral angles at each alpha-carbon atom for the refined co-ordinates. Asterisk, Pro; cross, Gly; circle, any other residue.

resolution thwarted the assignment of more satisfactory geometry. The estimate of error in this model is 0·3 Å as determined by the method of Luzzati (1952). The final crystallographic *R*-factor is 0·225 for 23,737 reflections in the range of 8·0 Å to 2·7 Å. The diffraction data, co-ordinate data, temperature factors and solvent occupancies for the Fab of McPC603 have been deposited in the Protein Data Bank at Brookhaven National Laboratory (Bernstein *et al.*, 1977).

(a) *General structure of the Fab*

Several changes were observed in the molecule as a consequence of this refinement, although the overall structure remained similar to that described earlier (Segal *et al.*, 1974). Most of these changes are to be found in the details of the loops, in particular L1 and H3. The region of L1 is poorly defined in the electron density maps, even with the higher-resolution data. We have now rebuilt the neighboring H3, whose density has become less ambiguous and, concurrently, have altered the conformation of L1. In Figure 2, L1 may be noted to protrude from the general surface of the molecule. The direction of this protrusion is reasonably correct but the electron density is too weak to define the positions of the three outermost amino acid residues.

Three other residues have tentative placements due to problems in maintaining proper stereochemistry while fitting the electron density: Gln162 and Asn163 (residues 156 and 157 in the numbering used by Kabat *et al.*, 1983) of CL and Glu202 (residue 203, Kabat *et al.*, 1983) of CH1. In these cases, we permitted the stereochemical constraints to dictate the final placement. Also, several disordered side-chains were observed within the entire molecule. In all these cases, the configuration corresponding to the stronger density was chosen for the model.

The 442 residues of McPC603 Fab include 25 proline residues. Of these, five are *cis*-proline: residues 8 and 101 in VL, 147 in CL, and 143 and 155 in CH1. The assignment of the configuration of all five *cis*-proline residues was unambiguous. The structurally homologous Pro8 and Pro95 of Rei (Huber & Steigemann, 1974), and Pro147 (CL) and Pro155 (CH1) of Kol (Marquart *et al.*, 1980), also have been found in the *cis* conformation. Pro143 (CH1) of McPC603 has no counterpart in previously reported antibody structures.

(b) *Crystal packing*

The McPC603 Fab crystallizes in the space group $P6_3$ (Rudikoff *et al.*, 1972), with $a = 162{\cdot}53$ Å, $c = 60{\cdot}72$ Å. The molecules are situated in clusters

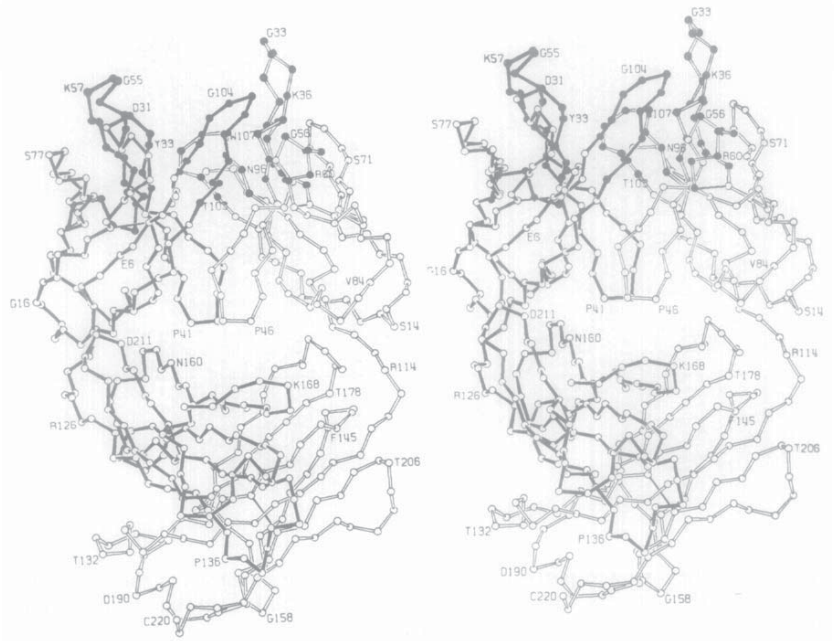

Figure 2. A stereo drawing of the alpha-carbon skeleton of McPC603. Continuous lines denote the heavy chain. The filled circles show the complementarity-determining residues (Kabat *et al.*, 1983).

of three at each of the crystallographic 3-fold axes of the unit cell (Fig. 3). Each cluster is related to neighboring clusters *via* the 2-fold screw situated midway between the 3-fold axes (Figs 3 and 4). The clusters of three are maintained principally by a set of hydrogen-bond and van der Waals contacts between neighboring molecules. Residues 14 to 20 and 71 of VL interact with residues 1, 26, 27, 100, 102, 104, 105 108 and 110 of the VH of the neighboring molecule while residues 18, 66, 67, 69,

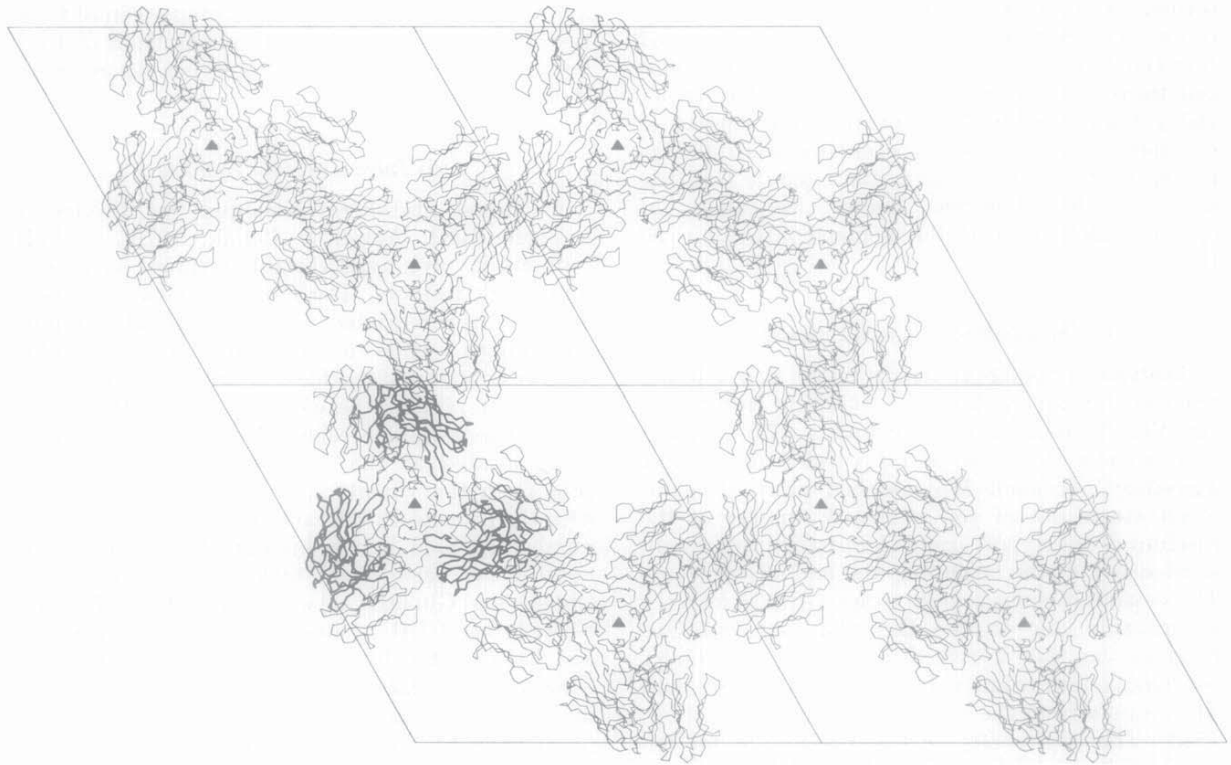

Figure 3. A projection of the alpha-carbon skeletons of 4 unit cells and the *ab* plane, illustrating the large channels that run parallel to the *c* axis through the crystal. The heavy chains have been drawn bolder in the 3 molecules that are clustered about a crystallographic 3-fold axis of the lower left-hand cell. The 3-fold axes are indicated by the symbol ▲: 2-fold screw axes are located midway between each adjacent pair of 3-fold axes; a 6_3 axis is located at each corner of

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.