Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structures of thymus and activation-regulated chemokine (TARC)

Oluwatoyin A. Asojo,[a] Cyril Boulègue,[b] David M. Hoover,[a] Wuyuan Lu[b] and Jacek Lubkowski[a]*

[a]Macromolecular Crystallography Laboratory, National Cancer Institute at Frederick, Frederick, MD 21702, USA, and [b]Institute of Human Virology, University of Maryland Biotechnology Institute, 725 West Lombard Street, Baltimore, MD 21201, USA

Correspondence e-mail: jacek@ncifcrf.gov

Received 21 January 2003
Accepted 29 April 2003

PDB References: TARC, triclinic, *P*1 space group, 1nr4, r1nr4sf; tetragonal, $P4_1$ space group, 1nr2, r1nr2sf.

Thymus and activation-regulated chemokine (TARC) is a CC chemokine that is mainly expressed in the thymus. TARC interacts primarily with the CCR4 receptor and to a lesser extent with the CCR8 receptor. The structures of TARC have been solved by molecular replacement in two space groups, triclinic (*P*1) and tetragonal ($P4_1$), and refined to resolutions of 1.72 and 2.1 Å, respectively, with *R* factors of 19.8% ($R_{\rm free}$ = 24.1%) and 19.8% ($R_{\rm free}$ = 27.7%), respectively. The search model originated from the crystal structure of another chemokine, RANTES, and proved to be only modestly similar to the refined structure of TARC. Whereas the tetragonal structure was easily solved using the program *AMoRe*, solution of the triclinic structure proved to be quite challenging and was obtained by combining the results from four different molecular-replacement programs (*AMoRe*, *CNS*, *BEAST* and *EPMR*), with subsequent extension of the gathered information. The tertiary structure of TARC is similar to that of other CC chemokines, with a three-stranded antiparallel β-sheet flanked by a C-terminal helix. Both quaternary structures consist of dimers, which in the triclinic crystals pack further into tetramers. The TARC dimers are similar to those observed previously in the crystal structures of both MCP-1 and RANTES.

1. Introduction

Thymus- and activation-regulated chemokine (TARC), also known by the systematic name CCL17, is a typical CC chemokine consisting of 94 amino acids and four conserved cysteine residues. The first 23 residues form a signal peptide. TARC was initially expressed using an Epstein–Barr virus vector signal sequence trap method and was the first CC chemokine shown to be chemotactic for lymphocytes but not for monocytes (Imai *et al.*, 1996, 1997).

TARC interacts specifically with the seven-helical transmembrane G-protein-coupled CCR4 receptor. This receptor is constitutively expressed by certain subsets of T cells in the thymus, where it may play a role in T-cell differentiation. The expression of CCR4 on TH2 cells (Bonecchi *et al.*, 1998; D'Ambrosio *et al.*, 1998; Sallusto *et al.*, 1998), platelets (Power, Mayer *et al.*, 1995), monocytes and basophils (Power, Clemetson *et al.*, 1995) may be important in the recruitment of these cells to certain inflammation sites. TARC also binds to CCR8, giving it an additional role in activation, migration and proliferation of lymphoid cells (Bernardini *et al.*, 1998). Apart from the thymus, lesser expression of TARC is observed in other tissues, such as the lung, colon and small intestine. The mature TARC protein has a 28% amino-acid sequence identity to MDC (CCL22), another CC chemokine that, like TARC, binds to the CCR4 receptor (Godiska *et al.*, 1997).

TARC shares a lower sequence identity with other CC chemokines, including RANTES, MIP-1β and I-309, the latter two of which also bind to CCR8.

TARC is involved in allergy-related diseases including atopic dermatitis (Vestergaard *et al.*, 2001), allergic airway inflammation (Kawasaki *et al.*, 2001) and psoriasis vulgaris (Rottman *et al.*, 2001), as well as in classic Hodgkin's lymphoma (Peh *et al.*, 2001). It also binds more tightly to glycosaminoglycans (GAGs) than do other CC chemokines such as RANTES or MCP-1. The high affinity of TARC for GAGs allows solid-phase gradients of TARC within the extracellular matrix of endothelial cells to attract and immobilize mast-cell granules (Patel *et al.*, 2001). This property causes TARC to attenuate inflammatory responses and may play a role in liver injury after systemic lipopolysaccharide administration (Yoneyama *et al.*, 1998). As a continuation of our efforts to determine the structural basis of the biological properties of chemokines, we have recently crystallized synthetic TARC in three forms belonging to the triclinic, hexagonal and tetragonal space groups (Asojo *et al.*, 2003). In this report, we describe the structure of TARC in triclinic and tetragonal space groups.

2. Materials and methods

2.1. Crystallization and data collection

The synthesis of TARC, the initial crystallization experiments and crystallization in the triclinic space group were carried out as described previously (Asojo *et al.*, 2003). Although the structural solution for the tetragonal form crystals proved to be relatively straightforward, it is necessary to mention that suitable X-ray data for the TARC crystals in the triclinic space group were collected prior to obtaining well diffracting tetragonal crystals.

Tetragonal crystals were obtained by vapor diffusion in hanging drops at 285 K. Drops were prepared by mixing equal volumes of protein solution (25 mg ml^{-1}) and reservoir solution, which consisted of 0.17 *M* ammonium acetate, 0.085 *M* trisodium citrate pH 5.6, 25.5%(*w*/*v*) polyethylene glycol 4000 and 15%(*v*/*v*) glycerol. The needle-like crystals grew to a maximum size (0.05 × 0.1 × 1.0 mm) within two weeks. For diffraction experiments, crystals were taken directly from the droplets with fiber loops and were flash-frozen in a stream of N_2. The X-ray data sets for both crystal forms were processed using the programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997).

The tetragonal data set was collected at the X9B beamline of the National Synchrotron Light Source at Brookhaven National Laboratory. These crystals diffracted to 2.1 Å and belong to space group $P4_1$, with unit-cell parameters $a = b = 47.8$, $c = 58.2$ Å. The X-ray data from triclinic crystals (unit-cell parameters $a = 44.35$, $b = 56.53$, $c = 76.66$ Å, $\alpha = 69.97$, $\beta = 85.56$, $\gamma = 72.74^\circ$) extended to 1.72 Å using a conventional radiation source (Asojo *et al.*, 2003). In addition, pseudo-*C*2 symmetry was observed when the data was indexed at a lower resolution (~4 Å). Although the X-ray data intensities could be integrated in this *C*-centered monoclinic cell, individual images could only be scaled within a limited oscillation range (<50°) and with the resolution limit not exceeding 3.5 Å. Furthermore, upon indexing at very low resolution (>5.5 Å), a pattern of reflections consistent with *F*222 symmetry was observed. In the latter case, however, the data-reduction process failed even for subsets of a few adjacent images. It is important to mention that three independent and complete X-ray data sets were collected from different crystals and the observations described above were consistent in all cases. With the unit-cell volume of the triclinic form crystals being 172 375 Å^3 and solvent occupying between 40 and 80% of the unit-cell volume (a Matthews number of between 2.1 and 5.4 Å^3 Da^{-1}), the unit cell of those crystals was expected to contain between four and ten molecules of TARC.

3. Results and discussion

3.1. Structure determination for the tetragonal crystal form

The tetragonal structure was solved using the program *AMoRe* (Navaza, 1994) with a monomer of the CC chemokine RANTES (Wilken *et al.*, 1999) as the search model. Similar solutions, corresponding to dimers of TARC in the asymmetric unit, were obtained from searches with high-resolution limits of 2.5, 3.0 and 3.5 Å, all with a low-resolution limit of 15.0 Å. No solutions were obtained in space group $P4_3$. The initial *R* factor of the unrefined solution from *AMoRe* was 40% ($R_{\rm free}$ = 44%).

3.2. Structure determination for the triclinic crystal form

3.2.1. The self-rotation searches. Self-rotation searches of the three independently collected X-ray data sets were conducted using both *AMoRe* and *CNS* (Brünger *et al.*, 1998) with various resolution ranges. The high-resolution limit varied between 2.5 and 5.5 Å and the low-resolution limit between 8.0 and 25 Å. The results obtained from all the self-rotation searches displayed a striking consistency in terms of detecting the non-crystallographic twofold rotation axes. The values of the Euler angles and direction cosines obtained from both *AMoRe* and *CNS* searches were equivalent and were characterized by a high intensity of the self-rotation signal as expressed by the correlation coefficient or *R* values. Additionally, the three non-crystallographic twofold axes are mutually perpendicular. Their orientations with respect to the orthogonal axes of the reference system, together with the smaller triclinic unit cell, are shown in Fig. 1(*a*). The mutual orthogonality of the non-crystallographic axes is consistent with the earlier observation of a pseudo-*F*222 pattern of reflections at very low resolution.

Detection of three non-crystallographic twofold axes, however, suggests that the unit cell consists of four monomers, all of which have to satisfy the same non-crystallographic symmetry. This conclusion was supported by the similar level of the self-rotation signal for all three twofold axes, suggesting that these axes relate a comparable number of atoms. The case of only four molecules in the unit cell, however, would result

in a very high solvent content (~80%) and has never been observed for chemokine crystals at high resolution and with good-quality X-ray diffraction. Therefore, a scenario in which eight molecules of TARC related by three twofold non-crystallographic axes are present in the small unit cell of the triclinic crystals seemed most plausible. Three mutually perpendicular twofold axes frequently indicate the presence of a tetrahedral arrangement of molecules and imply the possibility of two tetramers of TARC occupying the unit cell. Such a possibility was not confirmed, however, by the lack of translation-derived peaks in 'native' Patterson maps.

3.2.2. Molecular models used in MR searches. In all the MR searches the models were based on the monomeric X-ray structure of RANTES. These models comprised residues 7–67; all the side chains of residues that differed from alanine, serine and cysteine were converted to serine, with the positions of the C^{β} and O^{γ} atoms being the same as those of the 'equivalent' atoms in the original residues. The cysteine residues were retained unchanged, as the disulfide bonds formed by these residues are structurally conserved in all CC chemokines.

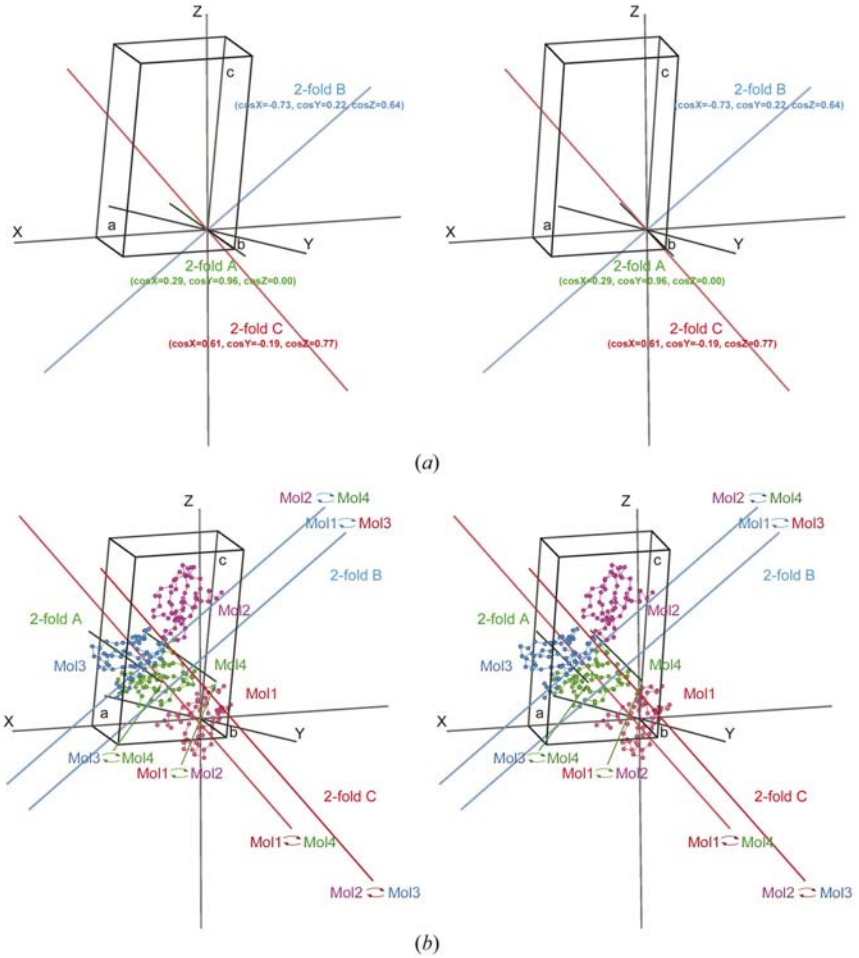

Figure 1
Three non-crystallographic twofold axes. (*a*) The three straight lines accompanied by their direction cosines are parallel to the non-crystallographic twofold axes identified during the self-rotation searches with *AMoRe* and *CNS*. Note that axis *A* lies parallel to the *ab* face of the unit cell. (*b*) The four monomers of TARC are placed in the unit cell according to the MR solutions obtained from *BEAST* and *EPMR*. The non-crystallographic twofold axes relating the molecules within the pairs are also shown.

3.2.3. Molecular replacement using *BEAST*.

(i) *Cross-rotation searches using BEAST*. As previously described for tetragonal crystals, a number of cross-rotation searches (Read, 2001) were conducted varying several parameters, among which the most significant were the resolution range, rejection criteria for reflections and the assumed degree of similarity between the search-model structure and the anticipated solution. In all searches, the assumed content of the asymmetric unit (a.u.) was eight monomers and the rotation step was 3.6°. Analysis of the angular relations between the cross-rotation peaks reveals that a group of peaks (1, 2, 4 and 7) and the group consisting of peaks 3, 5, 6 and 8 are related by three 180° rotations. Significantly, these angular relationships are identical to the non-crystallographic twofold rotations determined during the self-rotation searches.

(ii) *Translation searches using BEAST*. Following the encouraging results of the cross-rotation searches with *BEAST*, translation searches were continued with the same program. Applying a similar strategy, *i.e.* varying the resolution ranges and the reflection rejection criteria, the first four monomers were easily (and consistently) placed, as confirmed subsequently, in correct positions. Searches were performed in two different ways. In one approach, the first monomer, oriented according to the highest cross-rotation peak, was fixed at the center of the orthogonal reference system. In the three separate translation searches, the search models were monomers pre-oriented according to the cross-rotations 2, 4 and 7, respectively. The same results were obtained in the second approach, in which the individual monomers with already determined translations contributed to the partial structure were fixed in the subsequent searches. After the four monomers, oriented according to the cross-rotation peaks 1, 2, 4 and 7, were combined, the likelihood gain was 176.97, with root-mean-square (r.m.s.) 7.66, and the value of the signal-to-noise was 10.14. The arrangement of the four monomers in the unit cell is shown in Fig. 1(*b*). The monomers were packed very loosely in the unit cell and consequently no quaternary arrangement that would possibly ease subsequent MR trials could be elucidated. Using the same protocols, translation searches were conducted for the additional four monomers rotated according to the cross-rotation solutions 3, 5, 6 and 8. However, neither of these strategies resulted in any significant improvement of the likelihood. Also, every additional solution from translation searches introduced multiple overlaps with the four molecules of TARC already present in the unit cell. For confirmation of the correctness of the MR solution for the

initial four monomers, the monomers were subjected to structural refinement using the program *CNS*, with maximum likelihood as a target. A divergence of the R_{free} values was not observed in the course of refinement. However, the decrease in both R factor and R_{free} was minimal; the R_{free} value remained above 0.5. The electron-density maps also did not reveal any features that would unambiguously support the correctness of the solution. Despite numerous trials, we could not extend the solution beyond the four monomers using MR searches with *BEAST*.

3.2.4. The MR searches for TARC using *AMoRe* and *EPMR*.

(i) *The MR searches for TARC using EPMR: first four monomers.* Lacking a decisive confirmation of the correctness of the MR solution of the first four monomers of TARC, different MR programs were used to tackle this problem. Using the results obtained from searches with *BEAST* both as a starting point (the results of the cross-rotation searches) and as a reference (the results from translation searches), we conducted extensive calculations initially with *AMoRe* and then with *EPMR* (Kissinger *et al.*, 1999). The results of calculations using *AMoRe* did not provide any significant progress. The program *EPMR* was considered particularly useful, as it employs a substantially different approach to solving the MR problem. Using the results of cross-rotation searches from *BEAST*, the translation searches only were conducted with *EPMR*. The MR solutions for the first four monomers of TARC in the triclinic crystals found using *BEAST* and *EPMR* were identical. Although a population size of 300 and 50 generations (default values in *EPMR*) led to successful results, we found that increasing both parameters led to more efficient searches.

(ii) *The searches for the remaining four monomers using EPMR.* Initially, translation searches were attempted with the additional four monomers of TARC rotated according to the cross-rotation solutions 3, 5, 6 and 8 obtained using *BEAST*. Similar to the results of trials with the other programs, the translation solutions could not be identified for any additional molecules in the unit cell. Such a result, as well as those previously obtained, strongly indicated that the cross-rotation solutions 3, 5, 6 and 8 may be incorrect. Therefore, subsequent searches were conducted with *EPMR* according to the general protocol, *i.e.* simultaneous rotation and translation using the same model of monomer as a search probe. No restriction on intermolecular distances was imposed in any of the searches. The most significant observation from *EPMR* was the revelation that the eight crystallographically independent monomers of TARC were arranged into four structurally conserved dimers. Fig. 2(*a*) represents the alignment of the C^{α} traces of these dimers. Such a result was not obtained with any of the other MR packages. The structural similarity of those dimers can be additionally shown in terms of the values of the r.m.s deviation of the positions of their C^{α} atoms calculated after their alignment. The r.m.s. deviation values varied between 0.17 and 0.47 Å for alignments of different dimers. Dimeric assemblies of TARC resemble those found earlier for other CC chemokines. This similarity is illustrated in Fig. 2(*b*), where in addition to the TARC dimer, those determined for MCP-1 and RANTES are also shown. It is interesting to point out that in our hands this dimer was the only search model capable of successfully solving the MR problem in both *CNS* and *AMoRE*.

In addition, *EPMR* was the only program independently capable of solving the triclinic structure of TARC. However, the rather small differences between the statistical measures obtained for correct and wrong solutions would be likely to make the analysis of the results, selection and the following cross-validation of correct solutions difficult. The last statement particularly applies to the values of R, which are quite high even for the correctly located model.

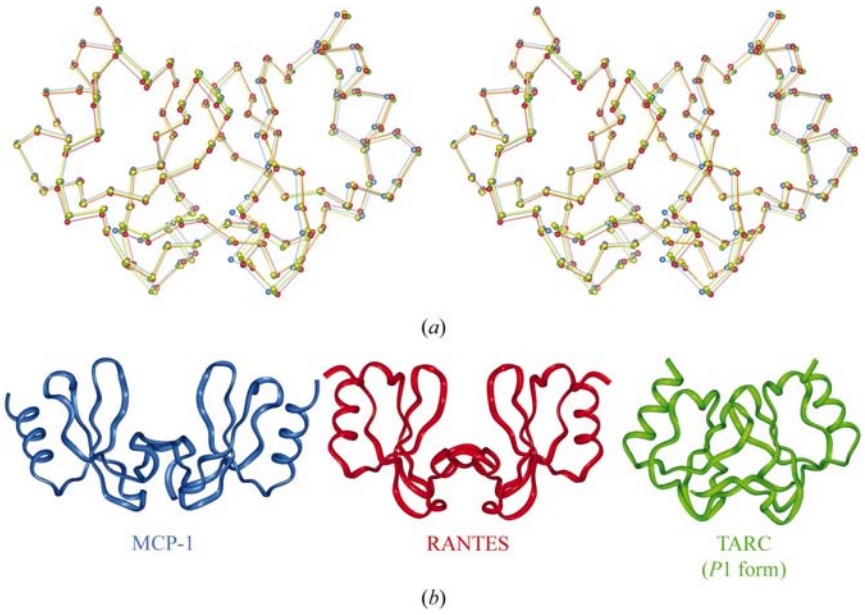

Figure 2
(*a*) Stereoscopic projection of the TARC dimers identified in the triclinic structure. The C^{α} traces of four dimers have been aligned using the program *ALIGN*. Despite the use of monomers as the search models, the four dimers show striking similarity and their topology is reminiscent of that found previously for other CC chemokines, as shown in (*b*). In the majority of CC chemokines, dimerization is promoted by interactions between the N-termini of both monomers, while the C-terminal helices are the most distant regions of the molecules. Although both of these features are present in the case of TARC, the dimers found are more compact compared with other CC chemokines.

3.3. Validation of the MR solutions: electron-density maps

After carrying out limited structural refinement of the MR solutions, consisting of rigid-body and least-square minimizations using the program *CNS* with maximum-likelihood as the target function, $2F_o - F_c$ and $F_o - F_c$ electron-density maps

were calculated for the resolution ranges 15.0–2.7 Å. Visual inspection of these maps indicated satisfactory agreement with the model. In addition, the status of several sites provided convincing validation of the correctness of the MR solutions. The amino-acid sequence of TARC has only one tryptophan residue, at position 27 (LKT**W**YQTS), while the corresponding residue in RANTES is tyrosine (IKI**Y**FYTS), which was replaced by serine in the MR model. Fig. 3(*a*) shows the $2F_o - F_c$ and $F_o - F_c$ electron-density maps around the site of Trp27, together with the MR model (ball-and-stick representation). The peaks present in both maps clearly indicate the location of the side chain of the missing tryptophan residue. In contrast, no electron-density peaks reminiscent of a tryptophan residue are found at the site corresponding to amino acid 57 (Fig. 3*b*), which is tryptophan in RANTES and arginine in TARC. Similar observations could be made for all ten independent monomers, eight in the triclinic and two in the tetragonal crystal form.

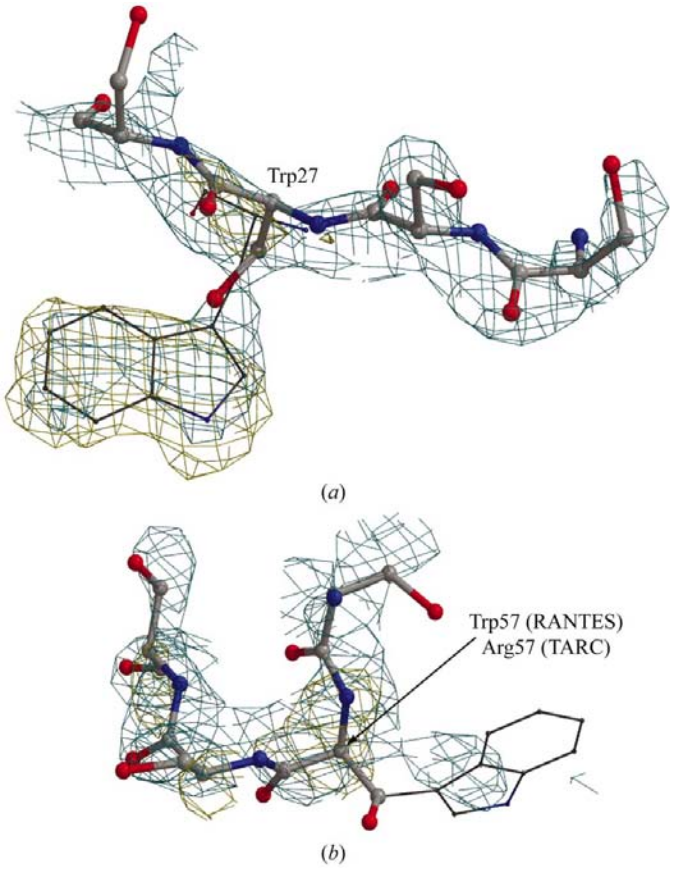

Figure 3
Examples of the electron-density peaks that provide added validation of the MR solution. Since all the residues that were neither Gly, Ala, Ser or Cys in TARC were substituted by Ser in the MR model, the presence of specifically shaped peaks in the $F_o - F_c$ electron density (shown above in khaki) provided additional confirmation of the correctness of the MR solutions. Sites of particular interest owing to the size of the side chain are shown. (*a*) The Trp residue present in position 27 (an unusual position for chemokines) is very well reflected in both the $F_o - F_c$ and $2F_o - F_c$ maps; the electron-density peaks, with the Trp side chain modeled into these peaks, are shown as thin sticks. (*b*) The Trp residue that is usually present in the C-terminal region of chemokines (position 57 in TARC) is instead Arg in TARC. The absence of the tryptophan is in agreement with the lack of electron density. The poorly defined $2F_o - F_c$ electron density possibly reflects flexibility of the Arg side chain. The small differences between the geometries of TARC and the MR model are the likely cause of the minor $F_o - F_c$ peaks. The contours of the $F_o - F_c$ and $2F_o - F_c$ electron-density maps are drawn at levels of 3.0 and 1.0σ, respectively.

Table 1
Data and refinement statistics.

Values in parentheses are for the highest resolution shell.

Data		
Space group	*P*1	$P4_1$
Resolution (Å)	20–1.72 (1.84–1.72)	25–2.18 (2.26–2.18)
R_{merge}† (%)	7.5 (27.9)	6.2 (19.4)
Completeness (%)	92.79 (92.9)	93.5 (92.2)
Redundancy	7.0 (3.0)	5.5 (2.5)
$I/\sigma(I)$	12.0 (5.1)	19.0 (2.1)
Refinement statistics		
Resolution (Å)	20–1.72 (1.84–1.72)	10–2.18 (2.26–2.18)
R factor‡ (%)	19.8 (27.5)	19.8 (26.0)
R_{free}§ (%)	24.1 (35.4)	27.7 (39.1)
Geometry		
Average *B* (Å^2)	32.7	29.1
R.m.s. deviation		
Bond length (Å)	0.029	0.022
Bond angles (°)	2.5	3.4
Backbone deviation (Å)	1.4	1.1
Coordinate error¶ (Å)	0.120	0.208
Ramanchandran plot		
Core (%)	92.1	83.9
Allowed (%)	6.8	15.2
Generously allowed (%)	0.4	0.9
Disallowed (%)	0	0
Model information		
No. of monomers	8	2
No. of amino-acid residues	516	122
No. of waters	645	225
No. of sulfates	4	0

† $R_{merge} = \sum|I - \langle I\rangle|/\sum I$, where I is the observed intensity and $\langle I\rangle$ is the average intensity obtained from multiple observations of symmetry-related reflections after rejections. ‡ *R* factor $= \sum||F_o| - |F_c||/\sum|F_o|$, where F_o are the observed and F_c the calculated structure factors. § The R_{free} set uses 5% of randomly chosen reflections (Brünger *et al.*, 1998). ¶ The coordinate error is based on *R* values.

3.4. Structural refinement

Iterative cycles of model building into electron density with the program *O* (Jones *et al.*, 1991) and structure refinement were carried out. The models were initially refined to 2.3 Å using rigid-body refinement followed by positional refinement and simulated annealing with a maximum-likelihood refinement procedure using Engh and Huber stereochemical parameters (Brünger *et al.*, 1998). Refinement to the full resolution limits of 1.72 Å (triclinic) and 2.18 Å (tetragonal) were carried out using restrained refinement and the maximum-likelihood refinement procedure with isotropic *B*-factor refinement using the program *REFMAC*5 (Pannu *et al.*, 1998). The final models of TARC were refined to *R* factors of 19.8% (R_{free} = 24.1%) and 19.8% (R_{free} = 27.7%) for the triclinic and tetragonal models, respectively. The data and model statistics are summarized in Table 1.

3.5. Structural features

The monomers have the characteristic fold of CC chemokines consisting of a three-stranded antiparallel β-sheet

flanked by a C-terminal α-helix (Fig. 4). There are two identical monomers in the asymmetric unit of the tetragonal crystal form and only 122 of 142 amino-acid residues are visible in the model, with the first seven N-terminal residues and last three C-terminal residues disordered in both monomers. The dimer packing observed for the tetragonal space group appears to be an artifact of crystallization, as there is no significant buried surface area between the crystallographically independent monomers; this dimer topology is different from any observed in previously solved structures of CC chemokines and no dimers that are similar to those in other CC chemokines can be generated from all symmetry mates. The maximum amount of buried surface between symmetry-related monomers is 1074 Å^2, which represents about 11% of the total solvent-accessible surface area.

There are four dimers in the asymmetric unit of the triclinic crystal form. The primary differences between monomers were found within a loop region involving residues Gln29–Ala38 as well as in the orientation of the flexible N-terminal loops. In the triclinic structure, 516 of 568 amino-acid residues could be built and the C-termini are ordered in three of the eight monomers. The first five residues in the N-terminus, which form part of the flexible N-terminal loops, still exhibited disorder and could not be built in four monomers. In monomer *A* residues 2–68 were ordered, while residues 8–71, 4–69, 7–69, 3–71, 6–71, 3–68 and 8–70 were ordered in monomers *B*, *C*, *D*, *E*, *F*, *G* and *H*, respectively. The four equivalent dimers involve monomers *AB*, *CD*, *EF* and *GH*. The interactions within dimers are mediated *via* an intermolecular antiparallel β-sheet between the N-termini (residues 7–13) and are similar to those observed in the crystal structure of other CC chemokines such as RANTES. The total solvent-accessible surface area buried between monomers ranges from 1552 to 1718 Å^2 (16–17% of the total area). In comparison, MCP-1 buries 18% of the accessible surface area (Lubkowski *et al.*, 1997), MIP-1β buries 18% (Lodi *et al.*, 1994) and RANTES buries 20% (Wilken *et al.*, 1999).

Dynamic light-scattering experiments performed using a DynaPro DLS Instrument (Protein Solutions, Inc.) suggest the formation of dimeric species in solutions of TARC at concentrations between 10 and 20 mg ml^{-1} buffered with 0.1 *M* sodium acetate pH 4.6. In addition, the formation of large-molecular-weight aggregates was not observed under the measurement conditions. It is likely that the triclinic dimer represents the solution dimer, as the buffer and pH are identical to those in light-scattering measurements, and that this dimer forms contacts similar to those observed in other CC chemokines. In addition, when the pH of the crystallization for the tetragonal crystals was changed to 4.6, triclinic crystals were predominantly formed. The same behavior was observed when at least 25 m*M* ammonium sulfate is added to the original tetragonal crystallization conditions. These observations further buttress the assertion that the dimer form observed in the triclinic crystals is the solution dimer observed in the dynamic scattering experiments.

An analysis of the dimers in the triclinic crystal form showed that Glu9 is buried within the dimeric interface. Further, the carboxylic acid groups from both monomers within the dimer are quite close to each other and in one dimer

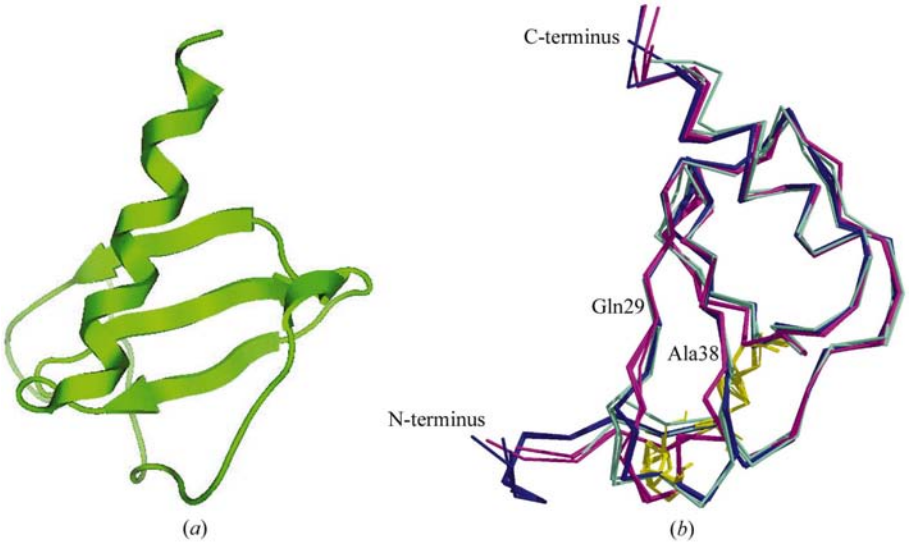

Figure 4
(*a*) Each monomer of TARC has the characteristic fold of CC chemokines consisting of a three-stranded antiparallel β-sheet flanked by a C-terminal α-helix. (*b*) Overlay of the ten TARC monomers with the identical tetragonal monomers in magenta; the triclinic monomers are shown in magenta and blue. The characteristic disulfide bonds are shown in yellow. The regions of highest variability are the N-terminus as well as the loop involving residues Gln29–Ala38.

```
                    *         20        *         40        *         60
RANTES     : -----SPYSSDTTPC---CFAYIARPLPRAHIKEY--FYTSGK-CSNPAVVFVT--RKNRQV :  49
mutRANTES  : -----SPYSSDTTPC---CFAYIARPLPRAHIKEY--FYTSGK-CSNPAVVFVT--AANRQV :  49
MIP-1β     : ----APMGSDPPTAC---CFSYTARKLPRNFVVDY--YETSSL-CSQPAVVFQT--KRSKQV :  50
MIP-1α     : ----ASLAADTPTAC---CFSYTSRQIPQNFIADY--FETSSQ-CSKPGVIFLT--KRSRQV :  50
MCP-1      : ----QPDAINAPVTC---CYNFTNRKISVQRLASY--RRITSSKCPKEAVIFKT--IVAKEI :  51
SLC        : GIPRTQGSDGGAQDC---CLKYSQRKIPAKVVRSYR-KQEPSLGCSIPAILLLPRKRSQAEL :  58
FKN        : ------MQHHGVTKCNITCSKMTSK-IPVALLIHYQQNQAS---CGKRAIILET--RQHRLF :  50
TARC       : -----ARGTNVGREC---CLEYFKGAIPLRKLKTW--YQTSED-CSRDAIVFVT--VQGRAI :  49

                    *         80        *        100        *        120
RANTES     : CANPE-KKWVREYINSLEMS---------------------------------------- :  68
mutRANTES  : CANPE-AAWVREYINSLEMS---------------------------------------- :  68
MIP-1β     : CADPS-ESKVQEYVYDLELN---------------------------------------- :  69
MIP-1α     : CADP-SEEWVQKYVSDLELSA-------------------------------------- :  70
MCP-1      : CADPK-QKWVQDSMDHLDKQTQTPKT--------------------------------- :  76
SLC        : CADPK-ELWVQQLMQHLDKTPSPQKPAQGCRKDRGASKTGKKGKGSKGCKRTERSQTPKGP : 118
FKN        : CADPK-EQWVKDAMQHLDRQAAALTRNG------------------------------- :  77
TARC       : CSDPN-NKRVKNAVKYLQSLERS------------------------------------ :  71
```

Figure 5
Primary structures of SLC, TARC, RANTES, MIP-1β, MCP-1, MIP-1α, mutRANTES and FKN show well conserved sequences.

(CD) the glutamate side chains are hydrogen bonded. It is likely that raising the pH above 5 causes these side chains to deprotonate and forces the dimer apart owing to electrostatic repulsion. Countering this effect, the presence of sulfate ions allows dimerization at pH 5.6. There are five clearly identified sulfate ions in the triclinic structure, two of which form electrostatic bridges across the interfaces of dimers *AB* and *CD*. All the interactions are between main-chain atoms, with three from one chain (Glu9 O, Glu9 N, Ser31 N) and two from the other (Ala48 N and Ala48 O). These sulfate ions are solvent-accessible and could mimic the positions of the sulfate groups of GAGs. Thus, while inhibited under physiological pH, the presence of polysulfonated GAGs may allow dimerization on cell surfaces.

3.6. Interactions with glycosaminoglycans

It was previously observed that TARC binds more tightly to GAGs than do other CC chemokines such as RANTES or MCP-1 (Patel *et al.*, 2001). This affinity for GAGs allows solid-phase gradients of TARC within the extracellular matrix of endothelial cells to attract and immobilize mast-cell granules. The same authors have also shown that the avidity of GAG, and specifically heparin, binding is in the following order: SLC > TARC > RANTES > MIP-1β > MCP-1 > MIP-1α > mutRANTES > FKN. All the CC chemokines compared in the study referred to above are similar in size except for SLC, which has a >40 amino-acid residue extension rich in basic residues that may provide additional sites for heparin binding (Fig. 5).

Through site-directed mutagenesis studies of RANTES, certain basic residues, Arg44, Lys45, Lys47, Lys55 and Lys56 (shown in Fig. 6*a*), were hypothesized to constitute the GAG-binding site, as mutating these residues to Ala resulted in lowered heparin binding for mutRANTES (Patel *et al.*, 2001). However, mutRANTES still binds heparin more tightly than FKN and by comparing the differences between structures of the CC chemokines, a further understanding as to which residues may contribute to tighter heparin binding may be gained. The structures of RANTES, MIP-1β, MIP-1α, MCP-1 and FKN (PDB codes 1eqt, 1hun, 1b50, 1dok and 1f2l, respectively) overlay quite well with TARC as shown in Fig. 6(*a*), with the locations of the basic residues implicated in heparin binding for RANTES indicated by the side chains.

In TARC, the position of residue 44 is occupied by a non-polar valine, which suggests that position 44 may not contribute to TARC's increased binding avidity for heparin. This suggestion is validated by the fact that FKN, with the lowest heparin-binding affinity, has an arginine in this position; mutRANTES and MCP-1, which bind heparin more tightly than FKN, have a non-polar alanine and isoleucine, respectively, in that position. Position 45 in TARC is occupied by a glutamine, which has uncharged polar side chains, so this position also may not contribute significantly to the increased avidity of heparin binding in TARC.

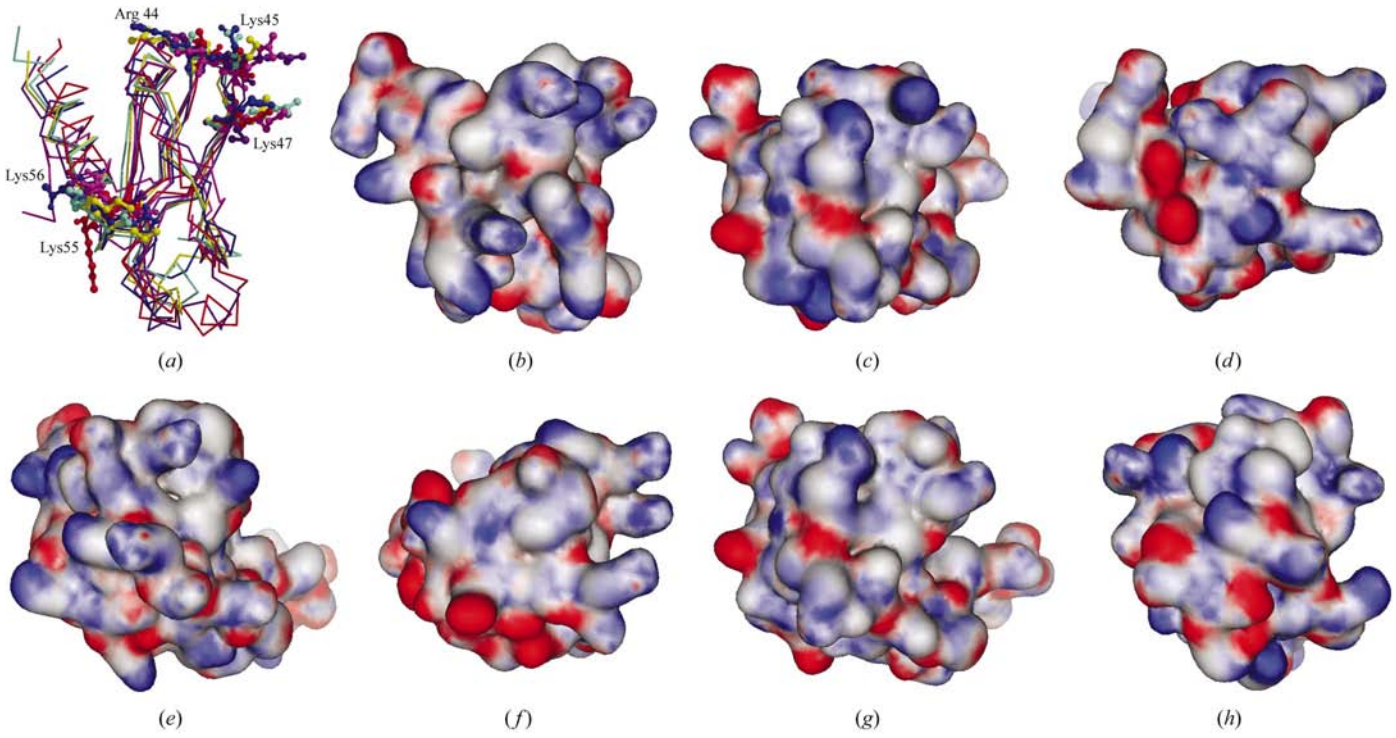

Figure 6
(*a*) Overlay of the structures of TARC (aquamarine), MIP-1α (magenta), MCP-1 (red), FKN (blue), MIP-1β (blue–violet) and RANTES (yellow) show how well the CC chemokines align; the putative heparin-binding residues are shown in ball-and-stick representation. (*b*)–(*g*) Charge distribution on surfaces of the CC chemokines, (*b*) TARC, (*c*) RANTES, (*d*) MIP-1β, (*e*) MCP-1, (*f*) MIP-1α, (*g*) mutRANTES and (*h*) FKN, shown in the same orientation as Fig. 6(*a*), correlates with the avidity of heparin binding. Note that regions around residues Lys47, Lys55 and Lys56 are more basic (blue) in TARC than in the others.

Position 47 (arginine in TARC) is highly conserved in all the CC chemokines compared and is occupied by either arginine or lysine. This residue is likely to play a role in heparin binding, but its exact function is unclear. Likewise, all the CC chemokines compared contain a basic residue at either position 55, position 56 or both. Although TARC does not have a charged residue in either position, position 57, which in the other CC chemokines is a conserved tryptophan, is instead occupied by an arginine, which presents a possible additional GAG-binding site.

While it is still unclear which residues in TARC may contribute to the high heparin-binding affinity, an electrostatic display of TARC shows a greater distribution of basic charge on the surface than other chemokines in the regions proposed to be important for heparin binding. Assuming there to be a correlation between the distribution of surface basic charge and the avidity of heparin binding, it is obvious that TARC has a greater total basic surface area in these critical regions than other chemokines, while FKN has the least (Fig. 6).

3.7. Comparison with other chemokine structures

The binding of TARC to chemokine receptors CCR4 and CCR8 is not a monogamous relationship. TARC shares the chemokine receptor CCR4 with one other chemokine, MDC/CCL22; TARC also shares the chemokine receptor CCR8 with two other chemokines, I-309/CCL1 and MIP-1β/CCL4. Based on sequence alignments, there are several regions of conservation within these proteins. However, most of the residues that are conserved between these proteins are somewhat universal, in that the residues are conserved throughout most of the CC chemokine family. Thus, the overlaps in functionality of CC chemokines are not surprising.

There are a few common elements which arise upon structurally aligning TARC with MIP-1β and I-309, CC chemokines of known structure (Lodi *et al.*, 1994; Keizer *et al.*, 2000; PDB codes 1hum and 1el0, respectively). The sequence Pro20-Leu21-Arg22 in TARC is matched in MDC and I-309. This region contains a highly conserved 3_{10}-helix between the N-terminal loop and the first β-strand and is known to contribute to receptor binding and discrimination. Nearby is a conserved arginine (Arg47 in TARC, Arg46 in MIP-1β and I-309) that can interact with receptors. Also, the guanidinium group of Arg57 of TARC, which substitutes for the highly conserved tryptophan in the C-terminal helix, is located in the same space as that of Arg61 of I-309 and that of the ε-amino group of Lys19 of MIP-1β. While these residues are not conserved within CC chemokines, the functionality based on charge placement may play a role in receptor-binding overlaps.

In a larger comparison to chemokine structures, the effect of mutating tryptophan from its highly conserved position to an arginine in TARC (Arg57) was analyzed. The role of the tryptophan is presumably to fasten the C-terminal helix to the β-sheet by burying the large hydrophobic side chain in the core of the monomer. The loss of tryptophan in TARC is accompanied by motion of other hydrophobic side chains. Ile19 of TARC rotates into the core of the molecule and is contacted by Ile49 and Phe15 to replace the functionality of the tryptophan. Interestingly, in the recently determined X-ray structure of MIP-3α, the same rearrangements were seen when the side chain of the tryptophan was rotated out of the core (Hoover *et al.*, 2002). Thus, although the tryptophan in this position is highly conserved, it may not be critical to the structures of chemokines.

4. Concluding remarks

The structure of TARC provides insights into the charge distribution and other issues relevant to GAG binding and seems to suggest a correlation between surface basic charge and heparin binding. In addition, new questions are raised about the occurrence of a dimer that is similar to that observed in other CC chemokines and the relevance of this dimeric form to the activity of TARC.

The results presented here also clearly show that the MR technique has become a very powerful tool for solving the structures of homologous macromolecules. Not long ago, it would have been nearly impossible to tackle successfully the solution of a multimolecule structure using MR. Such progress is a combination of three major components: technological improvements (faster computers, X-ray data-collection equipment delivering more accurate data), software development (computationally more efficient algorithms, novel protocols) and increase of structural knowledge (the rapidly growing number of entries in the PDB). Currently available computer programs can conduct MR calculations very rapidly (*i.e. AMoRe*), are less sensitive to the model and X-ray data quality (*i.e. BEAST*) and complete computationally very demanding tasks in a modestly short time (*i.e. EPMR*).

For simpler MR problems, such as solving the structure of crystals containing less than four molecules in the a.u. (with the availability of an adequate model), each of the MR programs currently used is likely to be able to provide successful results. In such cases, *AMoRe*, owing to its efficiency, may possibly be considered the program of choice, as was the case in the determination of the structure of TARC in the tetragonal space group. When the contents of the a.u. increase, however, it was observed that the strategy utilized in *AMoRe* may not lead to the solution. The same holds true when the search model does not approximate the target structure sufficiently well, which is frequently the case for models based on NMR structures. In the last few years, programs performing six-dimensional MR have become available (Chang & Lewis, 1997; Kissinger *et al.*, 1999; Glykos & Kokkinidis, 2001). Their relatively low popularity results primarily from their computational demands. Owing to the rapid increase of computational performance, however, six-dimensional searches may become more common, especially since the strategy employed by these programs is usually quite different from conventional MR. The use of multiple strategies for conducting MR always provides additional insight into the structural properties of the crystal organization and validation of the putative solutions. In some cases, however, when

the complexity of the MR problem increases, a continuation of MR searches conducted with different computational tools may be necessary.

The authors would like to thank Dr Randy J. Read for providing us with the program *BEAST* prior to its official release, and Drs Alexander Wlodawer, J. K. Mohana Rao and Zbigniew Dauter for helpful discussions during this project and suggestions during the manuscript preparation. This research was sponsored in part by the Intramural AIDS Targeted Antiviral Program of the Office of the Director, National Institutes of Health. All figures were generated using *Raster3D* (Merritt & Bacon, 1997), *MOLSCRIPT* (Kraulis, 1991) and *BOBSCRIPT* (Esnouf, 1997), except for surfaces, which were generated using *MOE* (Chemical Computing Group).

References

Asojo, O. A., Cater, S., Hoover, D. M., Boulegue, C., Lu, W. & Lubkowski, J. (2003). *Acta Cryst.* D**59**, 163–165.

Bernardini, G., Hedrick, J., Sozzani, S., Luini, W., Spinetti, G., Weiss, M., Menon, S., Zlotnik, A., Mantovani, A., Santoni, A. & Napolitano, M. (1998). *Eur. J. Immunol.* **28**, 582–588.

Bonecchi, R., Sozzani, S., Stine, J. T., Luini, W., D'Amico, G., Allavena, P., Chantry, D. & Mantovani, A. (1998). *Blood*, **92**, 2668–2671.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Chang, G. & Lewis, M. (1997). *Acta Cryst.* D**53**, 279–289.

D'Ambrosio, D., Iellem, A., Bonecchi, R., Mazzeo, D., Sozzani, S., Mantovani, A. & Sinigaglia, F. (1998). *J. Immunol.* **161**, 5111–5115.

Esnouf, R. M. (1997). *J. Mol. Graph. Model.* **15**, 132–133.

Glykos, N. M. & Kokkinidis, M. (2001). *Acta Cryst.* D**57**, 1462–1473.

Godiska, R., Chantry, D., Raport, C. J., Sozzani, S., Allavena, P., Leviten, D., Mantovani, A. & Gray, P. W. (1997). *J. Exp. Med.* **185**, 1595–1604.

Hoover, D. M., Boulegue, C., Yang, D., Oppenheim, J. J., Tucker, K., Lu, W. & Lubkowski, J. (2002). *J. Biol. Chem.* **277**, 37647–37654.

Imai, T., Baba, M., Nishimura, M., Kakizaki, M., Takagi, S. & Yoshie, O. (1997). *J. Biol. Chem.* **272**, 15036–15042.

Imai, T., Yoshida, T., Baba, M., Nishimura, M., Kakizaki, M. & Yoshie, O. (1996). *J. Biol. Chem.* **271**, 21514–21521.

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

Kawasaki, S., Takizawa, H., Yoneyama, H., Nakayama, T., Fujisawa, R., Izumizaki, M., Imai, T., Yoshie, O., Homma, I., Yamamoto, K. & Matsushima, K. (2001). *J. Immunol.* **166**, 2055–2062.

Keizer, D. W., Crump, M. P., Lee, T. W., Slupsky, C. M., Clark-Lewis, I. & Sykes, B. D. (2000). *Biochemistry*, **39**, 6053–6059.

Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Lodi, P. J., Garrett, D. S., Kuszewski, J., Tsang, M. L., Weatherbee, J. A., Leonard, W. J., Gronenborn, A. M. & Clore, G. M. (1994). *Science*, **263**, 1762–1767.

Lubkowski, J., Bujacz, G., Boqué, L., Domaille, P. J., Handel, T. M. & Wlodawer, A. (1997). *Nature Struct. Biol.* **4**, 64–69.

Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.

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

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Pannu, N. S., Murshudov, G. N., Dodson, E. J. & Read, R. J. (1998). *Acta Cryst.* D**54**, 1285–1294.

Patel, D. D., Koopmann, W., Imai, T., Whichard, L. P., Yoshie, O. & Krangel, M. S. (2001). *Clin. Immunol.* **99**, 43–52.

Peh, S. C., Kim, L. H. & Poppema, S. (2001). *Am. J. Surg. Pathol.* **25**, 925–929.

Power, C. A., Clemetson, J. M., Clemetson, K. J. & Wells, T. N. (1995). *Cytokine*, **7**, 479–482.

Power, C. A., Meyer, A., Nemeth, K., Bacon, K. B., Hoogewerf, A. J., Proudfoot, A. E. & Wells, T. N. (1995). *J. Biol. Chem.* **270**, 19495–19500.

Read, R. J. (2001). *Acta Cryst.* D**57**, 1373–1382.

Rottman, J. B., Smith, T. L., Ganley, K. G., Kikuchi, T. & Krueger, J. G. (2001). *Lab. Invest.* **81**, 335–347.

Sallusto, F., Lenig, D., Mackay, C. R. & Lanzavecchia, A. (1998). *J. Exp. Med.* **187**, 875–883.

Vestergaard, C., Kirstejn, N., Gesser, B., Mortensen, J. T., Matsushima, K. & Larsen, C. G. (2001). *J. Dermatol. Sci.* **26**, 46–54.

Wilken, J., Hoover, D., Thompson, D. A., Barlow, P. N., McSparron, H., Picard, L., Wlodawer, A., Lubkowski, J. & Kent, S. B. (1999). *Chem. Biol.* **6**, 43–51.

Yoneyama, H., Harada, A., Imai, T., Baba, M., Yoshie, O., Zhang, Y., Higashi, H., Murai, M., Asakura, H. & Matsushima, K. (1998). *J. Clin. Invest.* **102**, 1933–1941.