Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

1.7 Å structure of the stabilized REI_v mutant T39K. Application of local NCS restraints

Isabel Usón,[a]* Ehmke Pohl,[a,b]† Thomas R. Schneider,[c]‡ Zbigniew Dauter,[c]§ Arno Schmidt,[b]¶ Hans-Joachim Fritz[b] and George M. Sheldrick[a]

[a]Institut für Anorganische Chemie der Universität Göttingen, Tammannstrasse 4, D37077 Göttingen, Germany, [b]Institut für Molekulare Genetik der Universität Göttingen, Griesebachstrasse 8, D37077 Göttingen, Germany, and [c]European Molecular Biology Laboratory, c/o DESY, Notkestrasse 85, D22603 Hamburg, Germany

† Present address: European Molecular Biology Laboratory, c/o DESY, Notkestrasse 85, D22603 Hamburg, Germany.
‡ Present address: Institut für Anorganische Chemie der Universität Göttingen, Tammannstrasse 4, D37077 Göttingen, Germany.
§ Present address: National Cancer Institute, Frederick and Brookhaven National Laboratory, Building 725A-X9, NY 11973, USA.
¶ Present address: BRAIN GmbH, Darmstädter Strasse 34, D64673 Zwingenberg, Germany.

Correspondence e-mail: uson@shelx.uni-ac.gwdg.de

Received 12 November 1998
Accepted 9 March 1999

PDB Reference: REL_v-T39K, 1bww.

The X-ray structure of the T39K mutant of the variable domain of a human immunoglobulin κ light chain has been determined at room temperature to 1.7 Å resolution with a conventional R factor of 0.182. T39K crystallizes in the triclinic space group $P1$ [a = 35.4 (1), b = 40.1 (1), c = 43.1 (1) Å, α = 66.9 (1), β = 85.4 (1), γ = 73.8 (1)°]. The unit-cell contains two monomers, related by a non-crystallographic twofold axis. The use of a novel type of local non-crystallographic symmetry restraints on related isotropic displacement parameters and 1–4 distances as incorporated in the refinement program *SHELXL* improves the model and quality of the maps, but local differences between both monomers in areas subject to different packing contacts can still be observed. 12 overall anisotropic scaling parameters were refined. These may have compensated for the difficulties in accurately scaling single rotation axis image-plate data from a triclinic crystal, because of the scarcity of common equivalent reflections. The final model has been used to perform a number of tests on anisotropic scaling, non-crystallographic symmetry, anisotropic refinement, determination of standard uncertainties and bulk solvent correction. It is remarkable that removal of the NCS restraints from the final model caused $R_{\rm free}$ to increase. These tests clarify the strategies for optimum use of *SHELXL* for refinement at medium as opposed to atomic resolution.

1. Introduction

Immunoglobulin domains consist of a sandwich of two antiparallel β-sheets, linked by a conserved disulfide bridge, which is buried in the hydrophobic interior of the domain (Amzel & Poljak, 1979). Whether antibodies actually require this disulfide bridge for folding stability had been a matter of debate for a long time (see, for example, Glockshuber *et al.*, 1992, or Goto & Hamaguchi, 1979). The contribution made by the central disulfide bridge to the overall folding stability of the immunoglobulin fragment REI_v, the variable domain of a human κ-light chain was investigated (Frisch *et al.*, 1996) by introducing stabilizing amino-acid replacements followed by removal of the disulfide bridge *via* chemical reduction or genetic substitution of the cysteine residue(s). The disulfide contribution was estimated to be, on average, 19 kJ mol^{-1}. The crystal structure of the disulfide-free mutant REI_v-C23V/Y32H showed within the experimental error the exact conservation of the fold, despite the absence of the disulfide bridge (Usón *et al.*, 1997). Removal of the disulfide bridge from the REI-wild type would not yield a stable protein, showing the necessity of at least one stabilizing mutation.

Table 1
Reflection data statistics for REI_v-T39K.

The percentage of unique reflections measured and with $I > 2\sigma(I)$ are both expressed in terms of the theoretical number of unique reflections.

Resolution range (Å)	Unique	% Complete	Redundancy	% ($I > 2\sigma$)	Mean $I/\sigma(I)$	$R_{sym}(I)$†	$R(I > 2\sigma$/all)
45–3.66	2313	99.4	3.6	98.6	29.4	0.037	0.164/0.164
3.66–2.91	2250	98.0	3.2	95.1	22.7	0.046	0.142/0.145
2.91–2.54	2222	97.1	1.8	90.0	12.6	0.053	0.147/0.157
2.54–2.31	2223	95.6	1.8	84.5	11.1	0.063	0.148/0.166
2.31–2.14	2247	96.6	1.8	83.5	9.8	0.075	0.153/0.175
2.14–2.02	2226	96.2	1.8	76.9	8.0	0.096	0.162/0.195
2.02–1.91	2230	96.4	1.8	69.0	6.1	0.129	0.167/0.215
1.91–1.83	2209	96.2	1.8	59.7	4.5	0.179	0.180/0.258
1.83–1.76	2242	96.6	1.8	48.5	3.4	0.228	0.183/0.299
1.76–1.70	2104	90.1	1.8	41.1	2.6	0.318	0.207/0.340
45–1.70	22266	96.2	1.8	74.4	11.1	0.048	0.158/0.182

† $R_{sym}(\mathrm{I}) = \sum |I - \langle I\rangle| / \sum(I)$, where $\langle I\rangle$ is the mean intensity of a set of equivalent reflections. For comparison purposes, a conventional $R = \sum ||F_o| - |F_c|| / \sum |F_o|$ is given.

The choice of residues to exchange was taken on the basis of multiple sequence alignment in a database of 116 immunoglobulin κ light chains most similar to REI_v, assuming as working hypothesis that amino-acid replacements in REI_v towards a canonical immunoglobulin sequence lead to stabilization (Steipe *et al.*, 1994). Here we describe the crystal structure of REI_v-T39K, containing one of the stabilizing mutations previously investigated; 104 of the 116 sequences considered contain lysine in position 39, and the folding stability of this mutant was found to be 5.3 kJ mol^{-1} higher than that of the wild type (Frisch *et al.*, 1996).

Refinement with *SHELXL* was performed on the 1.7 Å data set, and a series of tests were carried out to establish suitable refinement strategies with *SHELXL* at medium rather than atomic resolution; the majority of reported *SHELXL* refinements have involved data to 1.4 Å or higher resolution. In particular, the application of non-crystallographic symmetry as a local restraint, rather than as a global constraint or restraint, and the effect of anisotropic scaling have been investigated.

2. Methods

2.1. Production and crystallization

REI-T39K was expressed and purified as described previously (Kolmar *et al.*, 1992). Single crystals were grown by vapour diffusion at 277 K in sitting drops (10 µl) from a protein solution containing 10 mg ml^{-1} in 50 m*M* potassium phosphate buffer at pH 7, using 20% PEG 8000 in 100 m*M* HEPES buffer (pH 7) as a precipitant in a 1:1 ratio. The crystals grew over a period of two months. The crystals belong to the triclinic space group *P*1, with cell dimensions a = 35.4 (1), b = 40.1 (1), c = 43.1 (1) Å, α = 66.9 (1), β = 85.4 (1), γ = 73.8 (1)°. V = 53100 $Å^3$. The asymmetric unit contains a dimer, the Matthews coefficient (Matthews, 1968) being V_m = 2.26 $Å^3$ Da^{-1}, corresponding to 45% solvent.

2.2. X-ray data collection

Room temperature data to 1.7 Å resolution were collected at EMBL, c/o DESY, Hamburg, on the BW7B synchrotron beamline with an X-ray wavelength of 0.87 Å using a MAR image-plate scanner. Data processing and merging were performed using the *HKL* system (Otwinowski & Minor, 1997). A total of 22 266 unique data were measured at 293 K from a crystal of size 0.25 × 0.10 × 0.05 mm (R_{int} = 0.048, completeness = 96.3%) mounted in a capillary. Higher resolution data to 1.7 Å were collected at a crystal-to-detector distance of 150 mm with 180° φ range and a φ increment of 2.0°, low-resolution data at 300 mm (corresponding to 3.0 Å) using the same φ range and an increment of 3.0ú. Refinement of a per-frame *B* factor in *SCALEPACK* to account for crystal decay showed neither a systematic drift nor improved the consistency of equivalent reflections, so *B* factors were not refined in the final scaling of the data. One scale factor was refined for each frame; for the high-resolution pass they decreased by about 25% between first and last image, probably caused by imperfect centering of the small crystal in the X-ray beam. Reflection data statistics are shown in Table 1.

2.3. Structure solution

The structure of the T39K mutant was solved by the molecular-replacement method with the program package *AMoRe* (Navaza, 1994) using the REI wild-type dimer structure (Epp *et al.*, 1975) as a search model. The corresponding rotation-function search resulted in a well contrasted solution. After rigid-body refinement using data between 20 and 3.5 Å the *R* factor and correlation coefficient had values of 0.31 and 0.78, respectively.

2.4. Refinement progress

The model was refined against F^2 using the intensities generated by *SCALEPACK* (Otwinowski & Minor, 1997) with no resolution cut-off except when explicitly stated, with *SHELXL*97 (Sheldrick & Schneider, 1997), following the stages summarized in Table 2. Fig. 1 shows a graph of the refinement progress. First, the model was subjected to rigid-body refinement by full-matrix least-squares for data between 7 and 3 Å resolution, with each monomer treated as an independent rigid group. From this point on, 95% of the measured data to 1.7 Å were used for conjugate-gradient least-squares minimization, keeping a random set of 5% for cross validation with R_{free} (Brünger, 1992*a*). A constant weighting scheme for the reflection data was maintained throughout refinement:

Table 2
Refinement progress.

Action	Protein atoms	H atoms	Water	Parameters	Restraints	$R1[I > 2\sigma(I)$/all]	$R_{free}[I > 2\sigma(I)$/all]
Rigid group, 3–7 Å	1654	—	—	13	5061	0.319/0.323	0.301/0.305
Coordinates, NCSY	1654	—	—	2488	6441	0.358/0.387	0.373/0.407
U	1654	—	—	4142	8942	0.308/0.341	0.330/0.366
SWAT	1654	—	—	4144	8945	0.286/0.318	0.319/0.353
Changes in the model	1704	—	—	6821	9237	0.211/0.240	0.253/0.284
Anisotropic scaling	1704	—	—	6833	9243	0.196/0.221	0.234/0.262
Add solvent	1704	—	65	7093	9226	0.165/0.191	0.209/0.236
H atoms	1704	1666	65	7093	9279	0.163/0.187	0.204/0.229
Changes in model, solvent	1704	1670	92	7201	9271	0.156/0.181	0.195/0.221
Final	1704	1670	92	7201	9275	0.158/0.182	—/—

$$w = 1/\sigma(F_o^2) + (0.1P)^2,$$

with $P = [2F_c^2 + \mathrm{Max}(F_o^2, 0)]/3$, and, in contrast to normal practice for small-molecule refinement, not optimized to give a flat analysis of variance and a goodness of fit close to unity. This acknowledges the fact that the final model is neither perfect nor complete for a protein refinement at 1.7 Å.

Geometric restraints were applied on bond distances and 1–3 distances using the Engh & Huber (1991) values as targets. Planarity and chiral volume restraints, antibumping restraints, and similarity restraints on the isotropic displacement parameters were included. In addition, local non-crystallographic symmetry restraints were applied to corresponding atoms of the two monomers for 1–4 distances and for isotropic displacement parameters. By 'local restraints' we understand restraints that act over a short distance, and so are not affected by accumulated small conformational changes over larger distances. *SHELXL* requires no mask, matrix or NCS operator to define them; just a statement of the offset in the numeration between NCS-related residues, estimated standard deviations to weight both types of NCS restraints, and the first and last atom in the list to which the restraints are to be applied.

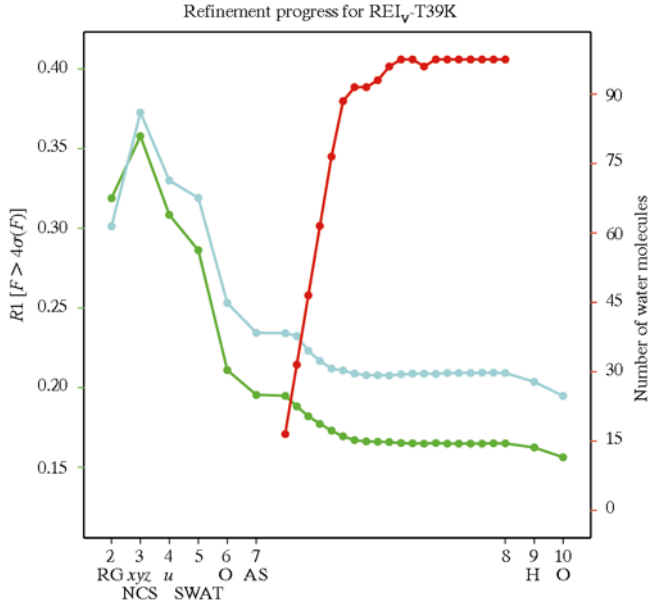

Figure 1
Graphical display of the behaviour of R_1 (green line) and R_{free} (blue line) throughout the different refinement stages and the number of water molecules (red line) incorporated during automatic water divining. The action taken at each step is described in the text.

The additional two residues at the N-termini (relative to the wild type) and the T39K mutations were included in the model after examining $3F_o - 2F_c$ and σ_A weighted maps (Read, 1986), displayed using *O* (Jones *et al.*, 1991).

Diffuse solvent was modelled using Babinet's principle (Moews & Kretsinger, 1975). 12 anisotropic scaling parameters (Parkin *et al.*, 1995) were refined simultaneously with the other least-squares parameters. Two residues in each chain, Cys25 and Arg63, were refined as discretely disordered. The alternative conformations were assigned different *SHELXL* 'part numbers' within the same residue. The program automatically avoids generating antibumping restraints between atoms assigned different part numbers and between atoms for which the sum of occupancies does not exceed 1.1. The occupancies of both alternative conformations were refined under the constraint that their sum has to add to 1.0000. The automatic water divining program *SHELXWAT* (Sheldrick & Schneider, 1997) was used to incorporate 92 water molecules into the model, selecting among the highest electron-density peaks (5σ level) those making reasonable contacts. The approximately spherical shape of the electron density corresponding to water molecules was confirmed in the σ_A map using interactive graphics. H atoms were included in geometrically calculated positions and refined using a riding model without introducing further parameters. Finally, some minor adjustments were performed manually using interactive graphics and the model refined to convergence. This model was used for the tests reported below, where it is referred to as the 'final model'. This structure was finally refined against all data (*i.e.* including the reflections flagged for R_{free}), without introducing any additional parameters or changing the restraint weights. The resulting geometrical parameters and statistics are summarized in Table 3. The atomic coordinates for REI$_v$-T39K have been deposited with the Protein Data Bank together with observed and calculated structure factors. Figures were drawn with *MOLSCRIPT* (Kraulis, 1991), *BOBSCRIPT* (Esnouf, 1997), *SHELXTL* (Sheldrick, 1990) and *SHELXPRO* (Sheldrick & Schneider, 1997).

Table 3
Refinement statistics.

No. of residues	218
No. of discretely disordered residues	4
No. of protein/solvent atoms	1670/92
R_{free} (%)† [$I > 2\sigma(I)$/all data]	19.5/22.1
$R1$ (%)‡ [$I > 2\sigma(I)$/all data]	15.6/18.1
R.m.s. deviations from idealized geometry	
Bond lengths (Å)	0.007
Bond angle distances (Å)	0.023
Zero chiral volumes (Å^3)	0.106
Non zero chiral volumes (Å^3)	0.098
Distances from restraint planes (Å)	0.027
Mean B factor (Å^2)	
Main-chain atoms	23.0
Side-chain atoms	30.7
Solvent atoms	40.5

† 5% data [1113 (no cutoff)/878 reflections with $I > 2\sigma$]. ‡ 95% data [21 153 (no cutoff)/16 351 reflections with $I > 2\sigma$].

3. Results and discussion

3.1. Description of the structure

As usually observed for immunoglobulin light chains, a dimer is formed in the crystal. The structure of a monomer of REI$_v$-T39K is shown in Fig. 2. The triclinic unit cell contains one dimer, in which the two monomers are related by a pseudo-twofold axis. A stereoview of the dimer seen along the non-crystallographic twofold axis is shown in Fig. 3. The packing consists of layers in the *XZ*-plane, with pseudo-twofold axes through and between dimers perpendicular to layers (Fig. 4*a*). However, adjacent layers are displaced relative to one another so that the pseudo-twofold axes are not valid for the next layer (Fig. 4*b*). Thus, the unit cell is indeed triclinic.

The crystal structure of REI-wild type was determined by Epp *et al.* (1974) and refined to a resolution of 2.0 Å (Epp *et al.*, 1975). The T39K variant contains two additional amino acids at the N-terminus that are not present in the wild type (Thr1 and Pro2). The mutant notation corresponds to the wild-type sequence, so the exchange of threonine for lysine is actually located at position 41 of the mutant. Fig. 5 shows the σ_A weighted electron-density map in the area of Lys41 of monomer I, prior to the introduction of the lysine side chain into the model. The final model is shown as well, and it can be seen how the ζ-amino group of Lys41 resides in an environment of negative electrostatic potential provided by Glu83 and Asp84 and makes an additional hydrogen bond with the carbonyl O atom of Glu83, which accounts for the stabilization by 5.3 kJ mol^{-1} of this mutant relative to the wild type (Frisch *et al.*, 1996). It establishes an additional link between two reverse turns connecting β-strands in both β-sheets, located at the end of the monomer opposite to the hypervariable regions and the buried disulfide bridge (Fig. 2). The S$^{\gamma}$ atom in Cys25, involved in the only disulfide bridge, is disordered between two discrete positions with relative occupancies of 0.6 and 0.4. This feature was already observed in the wild type. The side chain of Arg63 is also disordered. Both conformations form a salt bridge to Asp84. The major component (occupancy 0.6) contacts the carboxylic group of Glu83 as well.

The overall structures of the wild type and the T39K mutant are very close to each other, with an average root-mean-square deviation (r.m.s.d.) of the Cα atoms between the dimers of 0.45 Å, and r.m.s.d.s of 0.38 Å when the two monomers are fitted separately. In view of the r.m.s.d. of a Cα fit between the monomers in the wild type structure of 0.34 Å, the differences are not significant within the experimental error. The Ramachandran plot (Ramakrishnan & Ramachandran, 1965) shown in Fig. 6, contains only two outliers according to the values re-determined by Kleywegt & Jones (1996), both being situated in sharp turns and also present in the wild type. The 92 water molecules present in the model can be divided in two groups. 68 of them have an NCS-related counterpart (*i.e.* are conserved for each monomer) while the remaining 24 are unique.

3.2. Non-crystallographic symmetry

The presence in the asymmetric unit of several copies of a structure is usually exploited in the form of global non-crystallographic symmetry (NCS) constraints (see for example Hendrickson, 1985; Kleywegt, 1996), which lead to a reduction in the number of parameters and a simplification of the model being traced or refined, as the different NCS-related entities are regarded as being identical. While this approach is useful to improve the data-to-parameter ratio of structures determined at modest resolution, it may well prove too rigid for higher resolution structures, where averaging over the different equivalent copies can blur the details that would otherwise be significant at that resolution, thus leading to a less precise model. Therefore, the use of NCS constraints is an all-or-nothing situation. Better suited for higher resolution are NCS restraints as implemented in *X-PLOR* (Brünger, 1992*b*), *TNT* (Tronrud *et al.*, 1987) or *PROLSQ* (Hendrickson & Konnert, 1980) in the form of minimization of positional differences between the current atomic positions and the average positions of a least-squares fit with NCS-related

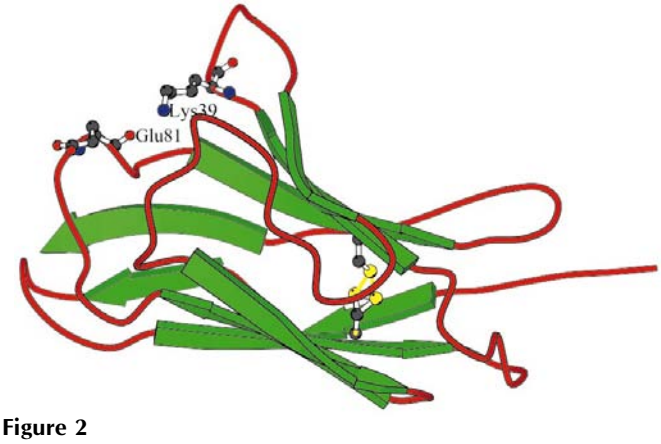

Figure 2
Ribbon plot of a REI$_v$-T39K monomer, showing its secondary structure and the location of the mutated residue 39, Glu81 whose main chain forms a hydrogen bond with N$^{\zeta}$ of Lys39, and the conserved disulfide bridge, with both disordered sites for the side chain of Cys23.

groups of atoms. Even though this procedure allows one to break up the NCS copies into groups of atoms and so still works well when parts of the NCS monomer can move relative to one another as rigid bodies, *e.g.* with a hinge motion (see for example, Boisvert *et al.*, 1996), it is still a global restraint in as much as the target position results from averaging over a large set of atoms and so is more sensitive to small accumulated differences. A more flexible alternative is the application of NCS as a local restraint, applied to 1–4 distances involving equivalent atoms in the different NCS-related monomers that are restrained to be equal without specifying a target value. The 1–4 distance restraint would be a way of restraining torsion angles to have similar values in the different copies, except that it allows ambiguity between plus and minus *gauche* for side-chain torsion angles, which is a reasonable possibility. It is not necessary to apply NCS restraints to 1,2 or 1,3 distances, because these are normally tightly restrained to target values anyway in macromolecular refinement.

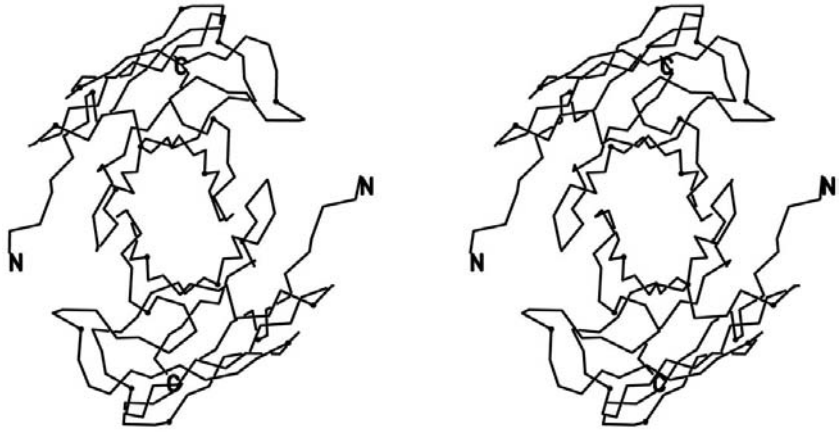

Figure 3
Stereoview of the REI_v-T39K dimer along the non-crystallographic twofold axis.

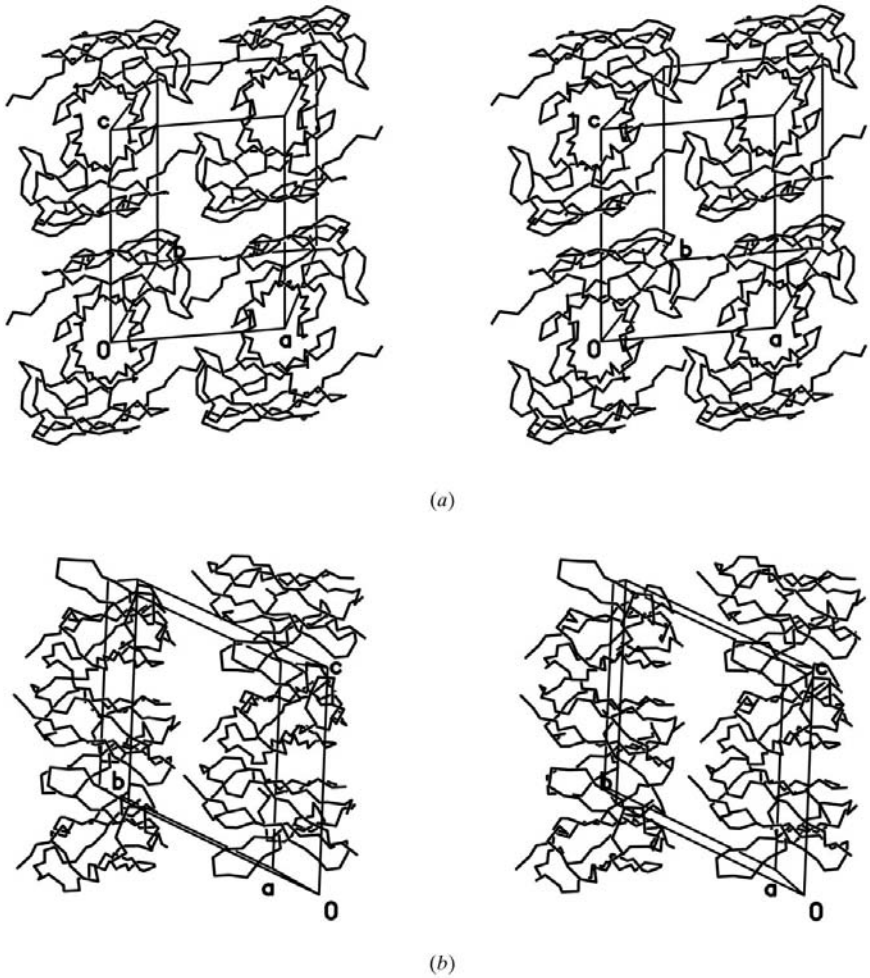

Figure 4
Packing diagrams showing (*a*) view of the *XZ* plane with pseudo-twofold axes through and between dimers; and (*b*) view of the *YZ* plane showing how the *XZ* layers are displaced relative to one another. The pseudo-twofold axes are local to one *XZ* layer.

Applying NCS locally in this way also allows more freedom to leave unrestrained any part of the structure where differences would be justified (*e.g.* side chains subject to different packing environments, hinges between rigid domains).

Fig. 7(*a*) shows a Kleywegt plot of the REI dimer structure (Epp *et al.*, 1975) that was our starting model in the molecular-replacement solution. In the Ramachandran plot, NCS-related residues are linked by lines, allowing one to see conformational differences at a glance. Fig. 7(*b*) shows the Kleywegt plot for the final REI_v-T39K model, which (unlike the wild type) was refined subject to local NCS restraints. Fig. 7(*c*) shows differences in the three main-chain torsion angles between the two monomers. Fig. 7(*d*) depicts differences in the isotropic displacement parameters of main- and side-chain atoms, which for NCS-related atoms were restrained to have similar values, along with their average values. Main-chain parameters are colour coded according to secondary structure, while plots for side-chain parameters use a colour scheme reflecting the residue type.

The differences in main-chain torsion angles φ and ψ are smaller than 3° with two exceptions: φ in the C-terminal residue Thr109 and ψ in

the disordered Cys25 differ by 6.6 and 5.6°, respectively. As the ω values are restricted by the planarity restraints applied to the peptide units, their low variability is not consequential. Also the NCS differences in side-chain torsion angles are in general below 7°, the only exception being the side chains of lysines and arginines, and that of Glu5, which are disordered into the solvent and lack well defined interactions or are involved in different specific packing contacts.

Isotropic displacement parameters of NCS-related atoms were restrained to be equal without specifying a target value. In the case of anisotropic refinement, no NCS restraints were applied on the displacement parameters. The differences in the isotropic displacement parameters of NCS-related atoms around residues 16 and 80 are associated with packing contacts (see Table 4). The B values in the chain II (labelled 2001–2109) are consistently higher than those in chain I (1001–1109) (average main-chain B: 23.9/22.0 Å, average side-chain B: 31.4/29.7 Å). The r.m.s.d. of a Cα superposition of the two monomers is 0.15 Å.

3.3. Anisotropic scaling

Overall anisotropy in the diffraction of macromolecular crystals is a common phenomenon (see for example Sheriff & Hendrickson, 1987). In the case of REI$_V$-T39K it could be due to both the laminal shape of the crystals and the difficulties in adequately scaling single rotation-axis image-plate data from an apparently decaying triclinic crystal, because of the scarcity of common equivalent reflections.

The anisotropic scaling correction implemented in the refinement program *SHELXL* makes use of the linear formula,

$$\begin{aligned} F^2_{\rm corr} = F^2_{\rm calc}[&h^2a^{*2}(a_1s + a_7) + k^2b^{*2}(a_2s + a_8) \\ &+ l^2c^{*2}(a_3s + a_9) + 2klb^*c^*(a_4s + a_{10}) \\ &+ 2hla^*c^*(a_5s + a_{11}) \\ &+ 2hka^*b^*(a_6s + a_{12})], \end{aligned}$$

where $s = \sin^{-2}\theta$.

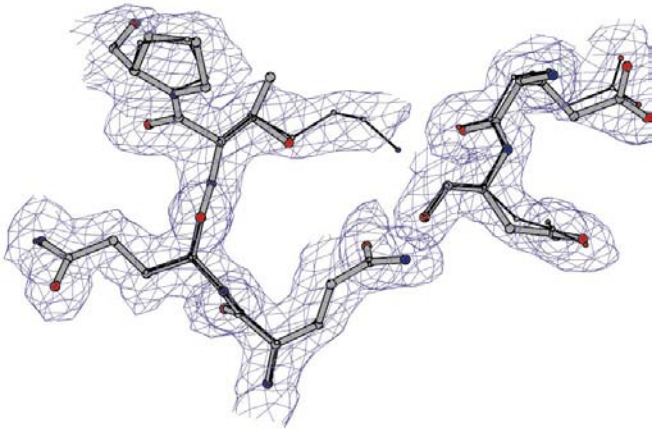

Figure 5
σ_A-weighted map (Read, 1986) showing the environment of the T39K mutation site, calculated whilst the model still contained a threonine side chain. The final refined model is shown as well.

This corresponds to the $n = 3$ and $n = 4$ terms in equation (11) as proposed by Parkin *et al.* (1995) (with some constant terms included in the coefficients $a_1 \ldots a_{12}$) and implemented by them for the empirical correction of absorption in the program *XABS*2. The 12 parameters $a_1 \ldots a_{12}$ are refined in each cycle. Their incorporation in the model caused both R_1 and $R_{\rm free}$ to drop by 2% (see Table 2). Tests including the additional terms in $\sin^2\theta$ suggested by Parkin *et al.* (1995) did not lead to significantly larger falls in $R_{\rm free}$ in this and other test cases.

In this case, there are no symmetry restrictions on the anisotropic scaling parameters, as the crystals are triclinic. It is debatable whether one should in general apply space-group symmetry constraints to the anisotropic scaling parameters, because not all sources of anisotropic diffraction would be expected to comply with space-group symmetry (*e.g.* physical restrictions and variation of the conditions during crystal growth, mechanical strains on mounting or freezing the crystal, *etc.*).

3.4. Refinement tests

After refinement was complete, we used the 'final model' (not yet refined against the $R_{\rm free}$ reflections) to perform a series of tests to assess different elements of the refinement strategy on the optimal model. The results are summarized in Table 5.

Removal of the non-crystallographic symmetry local restraints does not have a marked influence on R_1, which decreases slightly by 0.2%, once the refinement has converged to a minimum. It is remarkable that $R_{\rm free}$ increases by 0.5%. $R_{\rm free}$ reflections were chosen at random. There are good

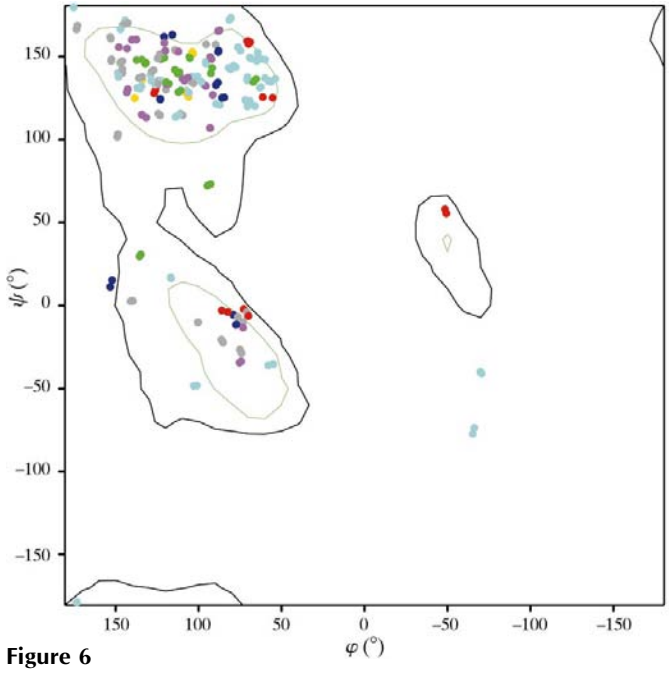

Figure 6
Ramachandran plot according to the expected distribution redetermined by Kleywegt & Jones (1996). Colour coding according to residue type (yellow: Cys, Met; green: Phe, Tyr, Trp, His; cyan: Gly, Ala, Leu, Ile, Val, Pro; red: Glu, Asp; blue: Arg, Lys; purple: Gln, Asn; grey: Ser, Thr).

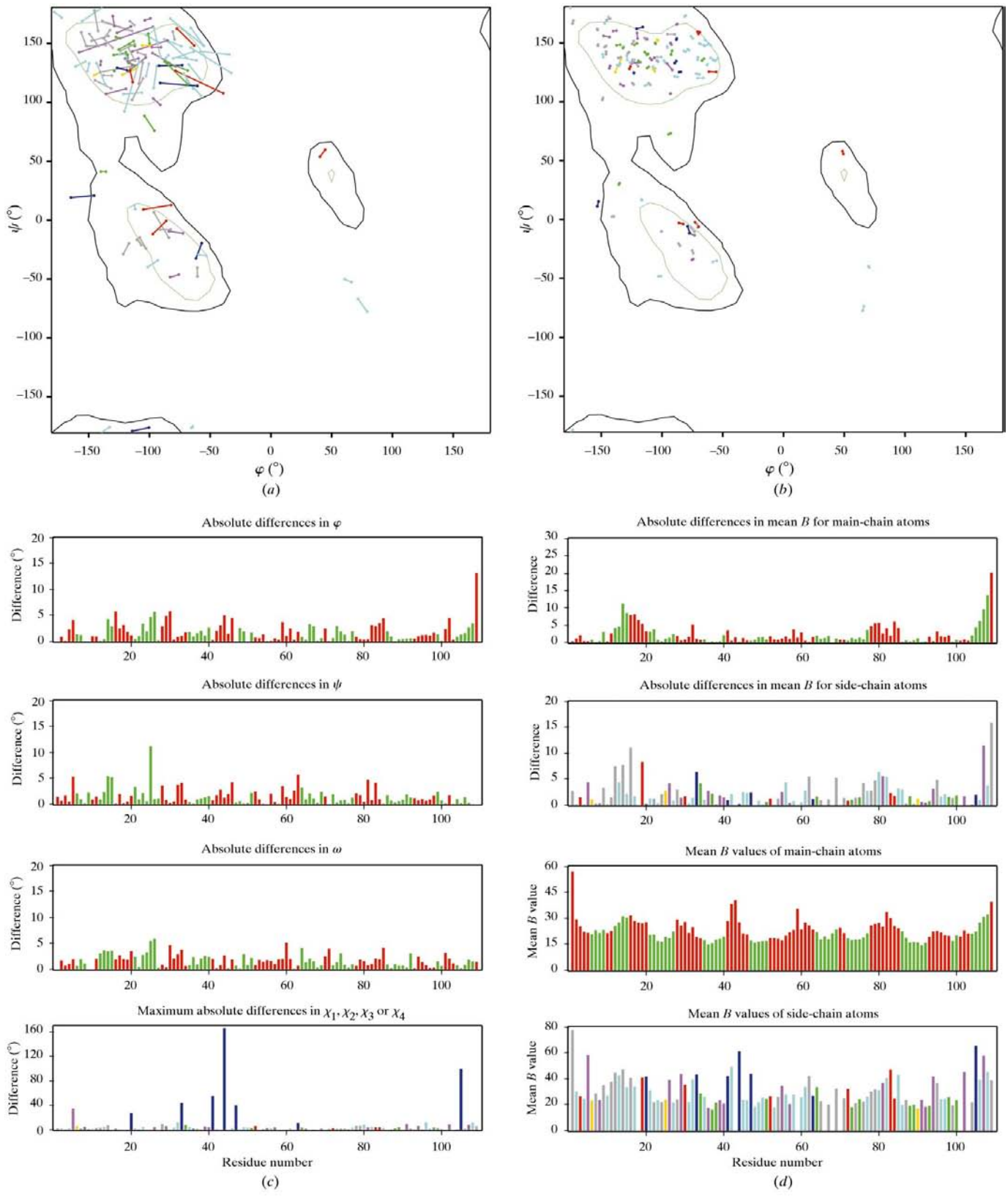

Figure 7
Plots illustrating the NCS analysis. (*a*) Kleywegt plot of the REI wild type; points representing equivalent residues in both monomers are linked by lines. Colour coding as in Fig. 6. (*b*) Kleywegt plot of REI$_v$-T39K. Colour coding as in Fig. 6. (*c*) Histogram showing differences in equivalent torsion angles between the two REI$_v$-T39K monomers. The residues are colour coded according to secondary structure (green for β-sheet, red for random coil) for main-chain parameters and according to residue type for side-chain parameters (as in Fig. 6). (*d*) Histogram showing differences between the two REI$_v$-T39K monomers in the average of equivalent *B* values over the main-chain (1) and side-chain (2) atoms of equivalent residues, and average values over all main-chain (3) and side-chain (4) *B* values in equivalent residues. Colour coding as in (*c*).

Table 4
Crystal contacts.

No.	Symmetry†	Contact area‡ (Å^2)	Polar interactions		Distance (Å)
1	(x, y, z)	744 (13.5 %)	Gln1040 NE2	Gln2040 OE1	3.1
	(dimer		Gln1040 OE1	Gln2040 NE2	3.1
	formation)		Gln1091 NE2	Tyr2038 OH	3.3
			Tyr1038 OH	Gln2091 NE2	3.3
2	$(x, y, z + 1)$		Arg1020 NH1	Gln2026 OE1	3.0
			Gln1026 OE1	Arg2020 NH2	2.7
			Thr1024 OG1	Thr2022 OG1	3.1
			Thr1022 OG1	Thr2024 OG1	3.0
3	$(x, y - 1, z)$		Tyr1034 OH	Ser1016 OG	2.9
4	$(x + 1, y, z)$		Gly1059 N	Asp1003 OD2	3.0
			Ser1062 N	Gln1005 OE1	2.9
			Ser1062 OG	Gln1005 NE2	2.9
			Ser1062 OG	Gln1005 OE1	3.0
			Asp2003 OD2	Gly2059 N	3.0
			Gln2005 OE1	Ser2062 N	3.0
			Gln2005 OE1	Ser2062 OG	2.3

† Crystallographic symmetry operator including translation vector. ‡ Water-accessible surface that is buried in the crystals, given as an absolute area per molecule pair and as a fraction of the total water-accessible surface of T39K (5748 Å^2).

Table 5
Refinement tests.

Action	Protein atoms	H atoms	Water	Parameters	Restraints	$R1[I > 2\sigma(I)/\text{all}]$	$R_{\text{free}}[I > 2\sigma(I)/\text{all}]$
Reference	1704	1670	92	7201	9271	0.156/0.181	0.195/0.221
No NCSY				7201	7067	0.154/0.179	0.201/0.227
NCSY only for 1–4 distances				7201	8419	0.156/0.180	0.196/0.222
NCSY only for B values				7201	7919	0.155/0.180	0.201/0.226
No anisotropic scaling				7189	9273	0.176/0.205	0.214/0.240
No anisotropic scaling + protein anisotropic				15709	21365	0.132/0.155	0.196/0.221
Protein anisotropic				15721	21375	0.133/0.156	0.198/0.222
No bulk solvent correction				7199	9278	0.174/0.199	0.201/0.229

arguments for choosing them in thin shells in the case of NCS, although this tends to have a deleterious effect on the quality of maps. Correlation between R_{free} reflections would be expected to cause a drop in R_{free}, so the fact that a slight increase is observed is a clear indication that removing the restraints has not improved the structure. The change in the geometry upon elimination of the NCS restraints can be quantified through the r.m.s.d. of a fit between the restrained and unrestrained models, which for all protein atoms averages 0.10 Å, and for all Cα atoms 0.03 Å.

Removing only the NCS restraints on equivalent 1–4 distances has almost the same effect on R_1 and R_{free} as removing all NCS restraints, and correspondingly, removal of the NCS restraints on equivalent B values has at this stage little or no effect on the residual values. Surprisingly, omitting either the 1–4 distance or the B-value NCS restraints (or both) leads to an increase in the average B values of main- and side-chain atoms in both monomers (by on average 0.6 and 0.8 Å^2, respectively, when both types of NCS restraints are omitted).

The positional NCS restraints are in this case the only restraints on torsion angles, although they do not tie the torsion angles to target values. Their presence may help to improve the geometry of the second monomer (labelled 2001–2109), which is less well defined in the electron-density maps than the first and shows consistently higher B values.

The absence of a diffuse solvent correction at this stage has a less marked effect than at the beginning of refinement, prior to modelling solvent molecules. The R_1 and R_{free} values increase by approximately half of the original drop if this correction is omitted at this stage.

Removal of the anisotropic scaling correction causes all residuals to increase by 2%. At the time it was introduced, it caused roughly the same drop.

Anisotropic refinement of the protein atoms, with or without anisotropic scaling, causes R_{free} to slightly increase, despite the drop in R_1 by more than 2%. The model is clearly getting worse because of the lower data-to-parameter ratio, the effect on R_1 being purely cosmetic as indicated by the increase in the unrestrained goodness-of-fit from 3.29 to 4.44 (g.o.f. = $\{\sum[w(F_o^2 - F_c^2)^2]/(N - P)\}^{1/2}$, where N is the number of reflections and P the number of parameters refined).

Finally, full-matrix least-squares refinements were performed after the structure had been refined to convergence against all data (including the R_{free} set) to calculate standard uncertainties (s.u.) both in the isotropic and the anisotropic cases. For this calculation all restraints were switched off, and the parameters were not allowed to shift. Fig. 8 plots the positional s.u. so obtained on the positional parameters against the B values of the corresponding atoms. The values calculated fit remarkably well with the formula proposed by Cruickshank (1999),

$$\sigma(x_i) = k(N_i/N - P)^{1/2} \times [g(B_i)/g(B_{\text{avg}})]C^{-1/3}\, R_1\, d_{\text{min}},$$

where k is a coefficient between 0.65 and 1.00, depending on the structure, N is the number of data, P is the number of parameters, $N_i = \sum Z_j^2/Z_i^2$, B_{avg} is the average B for fully occupied sites, $g(B)$ is an empirical quadratic function of B, and C is the fractional completeness of the data to d_{min}. Again, the consistently higher values of the standard uncertainties in

the anisotropic case indicate a less precise model than in the isotropic case.

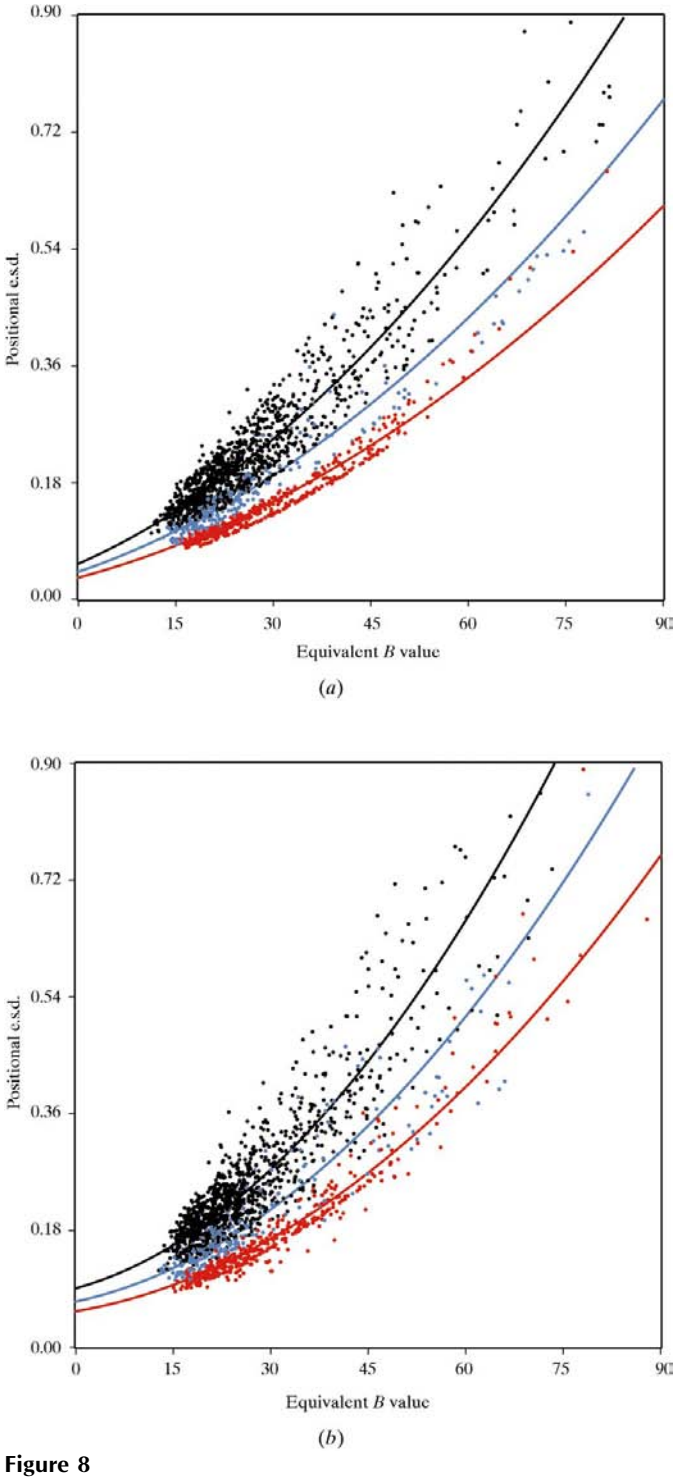

Figure 8
Standard uncertainties (s.u.) of the positional parameters plotted against B values. Black dots represent C, blue, N and red O atoms. The curves represent quadratic equations for each atom type as suggested by Cruickshank (1999). (*a*) Isotropic refinement of thermal displacement parameters. The average B values for main-chain, side-chain and solvent atoms are 23.0, 30.7 and 40.5, respectively. (*b*) Anisotropic refinement of thermal displacement parameters. The standard uncertainties are consistently higher than in the isotropic case, indicating a less precise model.

3.5. Conclusions

Medium-resolution protein refinement can be performed with *SHELXL*, and in favourable cases such as this one it is even possible to derive standard uncertainties in the parameters by inverting a full matrix (or a large full-matrix block, *e.g.* including all the *x, y, z* parameters). However, except when estimating the standard uncertainties, at this resolution it is important to exploit the many restraint options available; such as antibumping, distance, chiral volume, planarity, displacement parameter and non-crystallographic symetry (NCS) restraints. Local NCS restraints improve the quality of the model, while being flexible enough to account for differences relevant at medium resolution. Anisotropic refinement of B values may not be advantageous, given the increase in the number of parameters involved and the more limited number of available data than in the atomic resolution case. As refining anisotropically more than doubles the number of parameters, unless the decrease in $R_{\rm free}$ on going anisotropic exceeds at least 1% and parallels the drop in R_1, we suggest that the isotropic model should be retained; the goodness-of-fit also provides a useful guide as to whether anisotropic refinement is justified. Alternatively, overall anisotropic scaling involves only 12 extra parameters, and should be tested and judged on $R_{\rm free}$. One should bear in mind that $R_{\rm free}$ is subject to fluctuations because of the small number of reflections used to determine it. In other words, $R_{\rm free}$ itself has a standard deviation that is not much less than 1%, so small changes in $R_{\rm free}$ should not be overinterpreted. Although the diffuse solvent correction based on Babinet's principle is simple and robust, a more sophisticated model involving more than two variable parameters might be preferable; this correction is less important at very high resolution because a larger percentage of waters can be included in the model and the smaller percentage of reflections affected.

We are grateful to Dr Eila Cedergren-Zeppezauer for kindly giving us enough of her beamtime to complete our data collection. TRS thanks EMBL for a Predoctoral Fellowship. H-JF and GMS thank the Fonds der Chemischen Industrie for financial support.

References

Amzel, M. L. & Poljak, R. J. (1979). *Annu. Rev. Biochem.* **48**, 961–997.
Boisvert, D. C., Wang, J., Otwinowski, Z., Horwich, A. L. & Siegler, P. B. (1996). *Nature Struct. Biol.* **3**, 170–177.
Brünger, A. T. (1992*a*). *Nature (London)*, **355**, 472–475.
Brünger, A. T. (1992*b*). *X-PLOR*. A system for crystallography and NMR, Yale University, New Haven, Connecticut, USA.
Cruickshank, D. W. J. (1999). *Acta Cryst.* D**55**, 583–601.
Epp, O., Colman, P., Fehlhammer, H., Bode, W., Schiffer, M., Huber, R. & Palm, W. (1974). *Eur. J. Biochem.* **45**, 513.
Epp, O., Lattmann, E. E., Schiffer, M., Huber, R. & Palm, W. (1975). *Biochemistry*, **14**, 4943–4952.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 132–134.
Frisch, C., Kolmar, H., Schmidt, A., Kleemann, G., Reinhardt, A., Pohl, E., Usón, I., Schneider, T. R. & Fritz, H.-J. (1996). *Folding Des.* **1**, 431–440.

Glockshuber, R., Schmidt, T. & Plückthun, A. (1992). *Biochemistry*, **31**, 1270–1279.
Goto, Y. & Hamaguchi, K. (1979). *J. Biochem.* **86**, 1433–1441.
Hendrickson, W. A. (1985). *Methods Enzymol.* **115**, 252–270.
Hendrickson, W. A. & Konnert, J. H. (1980). *Computing in Crystallography*, edited by R. Diamond, pp. 13.01–13.25. Bangalore: Indian Academy of Sciences.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kleywegt, G. J. (1996). *Acta Cryst.* D**52**, 842–857.
Kleywegt, G. J. & Jones, T. A. (1996). *Structure*, **4**, 1395–1400.
Kolmar, H., Ferrando, E., Hennecke, F., Wippler, J. & Fritz, H.-J. (1992). *J. Mol. Biol.* **228**, 359–365.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Moews, P. C. & Kretsinger, R. H. (1975).*J. Mol. Biol.* **91**, 201.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Parkin, S., Moezzi, B. & Hope, H. (1995). *J. Appl. Cryst.* **28**, 53–56.
Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.
Read, R. (1986). *Acta Cryst.* A**42**, 140–149.
Sheldrick, G. M. (1990). *SHELXTL* software package for the determination of crystal structures, Release 5.10.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Sheriff, S. & Hendrickson, W. (1987). *Acta Cryst.* A**43**, 118–121.
Steipe, B., Schiller, B., Plückthun, A. & Steinbacher, S. (1994). *J. Mol. Biol.* **240**, 188–192.
Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). *Acta Cryst.* A**43**, 489–501.
Usón, I., Bes, M. T., Sheldrick, G. M., Schneider, T. R., Hartsch, T. & Fritz, H.-J. (1997). *Folding Des.* **2**, 357–361.