Dr M. M. Harding and Mr S. Maginn of Liverpool University and Dr P. Murray-Rust of Glaxo are thanked for discussions. The SERC is thanked for the provision of a research grant to JRH. The SERC, Daresbury Laboratory, is thanked for the provision of synchrotron radiation and facilities.

References

CAMPBELL, J. W., HABASH, J., HELLIWELL, J. R. & MOFFAT, K. (1986). *Inf. Q. Protein Crystallogr.* **18**, 23–32.

CLUCAS, J. A., HARDING, M. M. & MAGINN, S. J. (1988). *J. Chem. Soc. Chem. Commun.* pp. 185–187.

CRUICKSHANK, D. W. J., HELLIWELL, J. R. & MOFFAT, K. (1987). *Acta Cryst.* A**43**, 656–674.

DEBAERDEMAEKER, A., TATE, C. & WOOLFSON, M. M. (1985). *Acta Cryst.* A**41**, 286–290.

FARBER, G. K., MACHIN, P., ALMO, S. C., PETSKO, G. A. & HAJDU, J. (1988). *Proc. Natl Acad. Sci. USA*, **85**, 112–115.

HAJDU, J., MACHIN, P. A., CAMPBELL, J. W., GREENHOUGH, T. J., CLIFTON, I. J., ZUREK, S., GOVER, S., JOHNSON, L. N. & ELDER, M. (1987). *Nature (London)*, **329**, 178–181.

HARDING, M. M., MAGINN, S. J., CAMPBELL, J. W., CLIFTON, I. & MACHIN, P. (1988). *Acta Cryst.* B**44**, 142–146.

HELLIWELL, J. R., HABASH, J., CRUICKSHANK, D. W. J., HARDING, M. M., GREENHOUGH, T. J., CAMPBELL, J. W., CLIFTON, I. J., ELDER, M., MACHIN, P. A., PAPIZ, M. Z. & ZUREK, S. (1989). *J. Appl. Cryst.* 483–497.

HELLIWELL, J. R., HARROP, S., HABASH, J., MAGORRIAN, B. G., ALLINSON, N. M., GOMEZ DE ANDEREZ, D., HELLIWELL, M., DEREWENDA, Z. & CRUICKSHANK, D. W. J. (1989). *Rev. Sci. Instrum.* In the press.

MOLINES, H. & WAKSELMAN, C. (1984). *Synthesis*, **84**, 838–839.

NORTH, A. C. T., PHILLIPS, D. C. & MATHEWS, F. A. (1968). *Acta Cryst.* A**24**, 351–359.

SHELDRICK, G. M. (1976). *SHELX76*. Program for crystal structure determination. Univ. of Cambridge, England.

STILL, W. C., KAHN, M. & MITRA, A. (1978). *J. Org. Chem.* **43**, 2923–2925.

WOOD, I. G., THOMPSON, P. & MATTHEWMAN, J. C. (1983). *Acta Cryst.* B**39**, 543–547.

Acta Cryst. (1989). B**45**, 488–499

Crystallographic Refinement by Incorporation of Molecular Dynamics: Thermostable Serine Protease Thermitase Complexed with Eglin c

BY PIET GROS, MASAO FUJINAGA, BAUKE W. DIJKSTRA, KOR H. KALK AND WIM G. J. HOL

Biomolecular Study Centre (BIOS), Department of Chemistry, University of Groningen, Nijenborgh 16, 9747 *AG Groningen, The Netherlands*

(*Received* 30 *January* 1989; *accepted* 16 *May* 1989)

Abstract

In order to investigate the principles of protein thermostability, the crystal structure of thermitase from *Thermoactinomyces vulgaris*, a thermostable member of the subtilisin family of serine proteases, has been determined in a complex with eglin c. Eglin c is a serine protease inhibitor from the leech *Hirudo medicinalis*. After data collection with a television area-detector diffractometer and initial structure solution by molecular-replacement methods, crystallographic refinement proceeded with incorporation of molecular-dynamics techniques. It appeared that this refinement procedure has a large convergence radius with movements of more than 5 Å for many atoms. Two procedures for the crystallographic molecular-dynamics refinement have been tested. They differed mainly in time span and weight on the X-ray 'energy'. The best results were obtained with a procedure which allowed the molecular-dynamics technique to search a large area in conformational space by having less weight on the X-ray restraints and allowing more time. The use of molecular-dynamics refinement considerably simplified the laborious and difficult task of fitting the model in its electron density during the refinement process. The final crystallographic R factor is 17·9% at 2·2 Å resolution.

Introduction

Thermitase is a heat-stable serine protease from *Thermoactinomyces vulgaris* containing 279 amino-acid residues (Meloun, Baudys, Kostka, Hausdorf, Frömmel & Höhne, 1985; Frömmel & Höhne, 1981; Teplyakov, Strokopytov, Kuranova, Popov, Harutyunyan, Vainshtein, Frömel & Khene, 1986; Dauter, Betzel, Höhne, Ingelman & Wilson, 1988). It is homologous to the subtilisins Carlsberg, BPN′ and proteinase K, for which high-resolution X-ray structures have been reported (Bode, Papamokos & Musil, 1987; McPhalen, Schnebli & James, 1985; Katz & Kossiakoff, 1986; McPhalen, Svendsen, Jonassen & James, 1985; Betzel, Pal & Saenger, 1988). The amino-acid sequences of these enzymes have been compared by Meloun *et al.* (1985). The observed sequence identity between subtilisins Carlsberg and BPN′ is 69%, whereas the sequence identity of either of the subtilisins with thermitase is only 47%. A difference of 25 K in optimum tempera-

0108-7681/89/050488-12$03.00

ture for activity between thermitase and subtilisin Carlsberg has been reported by Kleine (1982) and Gusek & Kinsella (1987). Knowing the structure of thermitase in detail allows structural comparison of proteins with distinctly different stability properties. Comparison of thermolabile and thermostable protein structures is an important approach in determining origins of protein stability (see *e.g.* Alber, Dao-Piu, Wilson, Wozniak, Cook & Matthews, 1987; Argos, Rossmann, Grau, Zuber, Frank & Tratschin, 1979; Bryan, Rollence, Pontoliano, Wood, Finzel, Gilliland, Howard & Poulos, 1986; Matthews, 1987; Pauptit, Karlsson, Picot, Jenkins, Niklaus-Reimer & Jansonius, 1988; Perutz, 1978).

Structure solution

High-quality crystals of thermitase complexed with eglin c, a serine protease inhibitor from the leech *Hirudo medicinalis* (Seemüller, Fritz & Eulitz, 1981), were used in the study. The inhibitor consists of 70 amino-acid residues of which only 63 residues were observed in crystal structures reported by Bode *et al.* (1987) and McPhalen, Schnebli & James (1985). The crystals of thermitase:eglin c of space group $P2_12_12_1$, $a = 63{\cdot}25$, $b = 72{\cdot}10$ and $c = 89{\cdot}25$ Å, were grown by using the hanging-drop technique. The complex was crystallized at pH 6·0 in the presence of 50 m*M* bis-(2-hydroxyethyl)iminotris(hydroxylmethyl)methane and 15% polyethylene glycol (MW = 6000). The diffraction data were collected up to 2·2 Å resolution from one crystal on a FAST television area-detector diffractometer (Arndt, 1982; Renetseder, Dijkstra, Kalk, Verpoorte & Drenth, 1986) with a 1·5 kW sealed tube with a nickel filter operated at 46 kV and 32 mA. The crystal-to-detector distance was 65·4 mm. The data were processed using the *MADNES* program package (Pflugrath & Messerschmidt, 1986). A total of 63206 measurements were made yielding 16 310 unique reflections with an $R_{\rm merge}$ ($= \sum|F - \langle F\rangle|/\sum|\langle F\rangle|$) of 5·0%.

The structure was solved by applying the molecular-replacement method using the subtilisin Carlsberg:eglin c complex (Bode *et al.*, 1987) as a starting model. Its coordinates were kindly made available to us by Bode *et al.* (1987). The orientational and positional parameters for this model were easily obtained using the fast-rotation (Crowther, 1972) and translation functions (Crowther & Blow, 1967). The fast-rotation function was applied using 10 to 3·5 Å resolution data and a radius of integration of 21 Å. It resulted in a peak that was 1·4 times higher than the second highest peak. Data between 10 and 4 Å resolution were used for the translation functions. The correct translation was found from peaks in the sections $x = \frac{1}{2}$, $y = \frac{1}{2}$ and $z = \frac{1}{2}$, which were all twice as high as any other peak in those sections. Ambiguities in the resultant initial electron density distribution of thermitase:eglin made model building a difficult task. This is a problem often encountered, especially in the early stages of refinement when electron density maps are of low quality.

Molecular-dynamics refinement

The difficulties in interpreting an initial low-quality electron density map are alleviated to a considerable extent by the use of molecular dynamics in the crystallographic refinement procedure (MD refinement) (Brünger, Kuriyan & Karplus, 1987; Fujinaga, Gros & van Gunsteren, 1989). The basic principle of the method is that the crystallographic discrepancy term $E_x = (1/\sigma_x^2)\sum_{hkl}[|F_{\rm obs}(hkl)| - k|F_{\rm calc}(hkl)|]^2$ (where $F_{\rm obs}$ corresponds to observed structure factors, $F_{\rm calc}$ to calculated structure factors based on the current model, k is the scaling factor between $F_{\rm obs}$ and $F_{\rm calc}$, and σ_x is a weighting term) is added as a pseudo-energy term to the potential energy, $E_{\rm MD}$, of the system. Then, the total potential energy, $E_{\rm MDX} = E_{\rm MD} + E_{\rm X}$, is minimized during the refinement. This is achieved by solving the Newtonian equations of motion for all atoms using the forces derived from the energy terms, including the X-ray 'energy'. The temperature is kept constant by coupling to a heat bath (Berendsen, Postma, van Gunsteren, Dinola & Haak, 1984). Owing to the presence of kinetic energy in the simulated system, energy barriers can be overcome in this process. This means that, for example, peptide planes can be flipped and loops can be rearranged by exploring large areas in conformational space. Thus, in principle, it is possible to reach a lower energy minimum than the nearest minimum. This is in contrast to least-squares methods which will only find a local minimum.

The method of crystallographic refinement using molecular dynamics, as first described by Brünger *et al.* (1987), has been incorporated with some modifications in the molecular-simulation package *GROMOS* (van Gunsteren & Berendsen, 1987) by Fujinaga *et al.* (1989). Unlike Brünger *et al.* (1987), who perform their refinement at 3000 K, Fujinaga *et al.* (1989) have found that the method works well at 300 K. Here we report on the application of the latter procedure in the case of the thermitase:eglin c where the starting model is derived from a molecular-replacement solution.

Refinement of thermitase:eglin c

The starting model was constructed by changing the amino-acid sequence of the correctly oriented and translated subtilisin Carlsberg:eglin c complex into

Table 1. *Deletions and insertions of the thermitase sequence in comparison to subtilisin Carlsberg*

Addition	7 N-Terminal residues	Tyr1–Tyr7
Deletion	1 Residue	Trp24, *V*, Asp25
Insertion	2 Residues	Ala49, Gly50
Insertion	1 Residue	Thr83
Deletion	4 Residues	Asn170, *V*, *V*, *V*, *V*, Tyr171
Deletion	2 Residues	Ser244, *V*, *V*, Ala245
Insertion	1 Residue	Ser260
Addition	1 C-terminal residue	Tyr279

'*V*' indicates a deleted residue from the subtilisin Carlsberg sequence.

that of thermitase:eglin c, using the *MUTATE* option (Read, 1988) of the molecular-modelling and analysis program *WHATIF* (Vriend, 1989). The resulting model was neither inspected nor corrected manually in any way. A seven-residue N-terminal extension and a single-residue C-terminal addition of thermitase were not included in the model at this stage. Nor were any insertions added to the polypeptide chain. Bond breaks were created at the three insertion sites giving the refinement the possibility to move the chains apart. At three other sites (Table 1), residues had to be deleted from the starting model. Here extremely long bonds were made to close the gap. The resulting model contained five polypeptide chains: one eglin c (unchanged in amino-acid sequence) and four chains making up the structure of thermitase.

The model was subjected to a combination of energy minimization with X-ray restraints (EMX), 'pure' energy minimization (EM), molecular-dynamics refinement with X-ray restraints (MDXREF) and individual X-ray temperature-factor refinement (see Fig. 1). Initially EMX with data between 6 and 3 Å was performed to reduce the highest energies in the system. After 300 of these steps, at point *A* (see Fig. 1), a small manual correction was applied to the model involving only residues 180 and 266. This was followed by an additional 100 EMX steps. To speed up MDXREF the *SHAKE* algorithm (Ryckaert, Ciccotti & Berendsen, 1977) was used to restrain bond distances. However, the bond lengths have to be within acceptable limits before *SHAKE* can be used. This had not been obtained after a total of 400 steps of EMX. Therefore, 100 steps of EM, *i.e.* without X-ray information, were applied to meet the *SHAKE* criteria. From this point on, two procedures were tested.

Procedure I involves the 'gentle' approach. First, MDXREF was carried out at 600 K, initially, using only the lower-resolution X-ray data from 10 to 5 Å resolution. Subsequently, the resolution was increased gradually (Fig. 1*b*). Besides the slow increase of resolution in procedure I, the weight applied to the X-ray energy was initially significantly lower than the weight applied in procedure II described below (σ_x used in procedure I are given in Fig. 1*b*). The X-ray weight was kept constant during the MDXREF runs at 10–5 and 6–4 Å resolution. Typically MDXREF was performed in series of 0·2 ps, consisting of 100 molecular-dynamics steps with $\Delta t = 2$ fs. Then the balance between improvement of the crystallographic *R* factor and the energies was evaluated. At 6–3, 6–2·5 and 6–2·2 Å resolution the X-ray weights were increased slowly, see Fig. 1(*b*). After addition of 140 water molecules to the model, at point *I* in Fig. 1, the temperature in MDXREF was decreased from 600 to 300 K. In procedure I, molecular dynamics was given a significant period of time and amount of freedom for searching through conformational space.

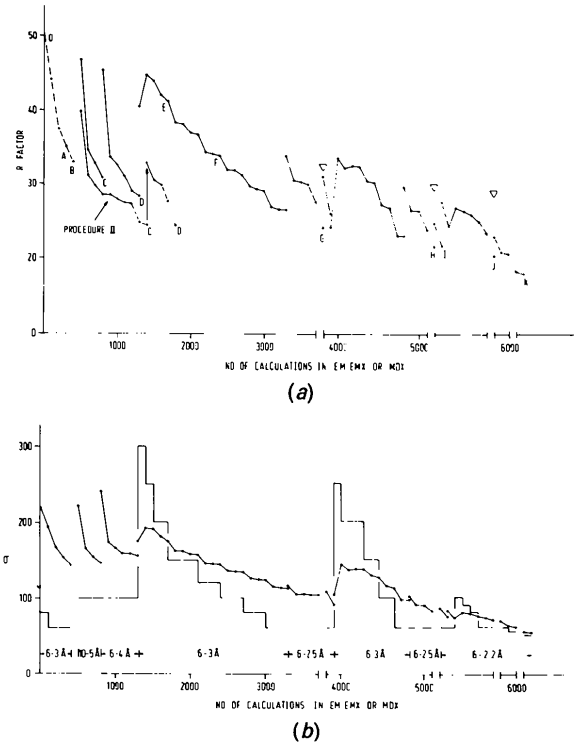

Fig. 1. Course of the refinement as indicated by the crystallographic *R* factor. (*a*) The crystallographic *R* factor ($R = \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$) as a function of the number of steps in the refinement. The energy minimization with X-ray restraints (EMX) is indicated by ---; the molecular-dynamics refinement (MDXREF) is indicated by —, and the individual temperature-factor refinement by ···. No *R* factor was calculated during energy minimization without X-ray restraints (EM); this is indicated by the gap between calculations 400 and 500. Labels at points along the refinement refer to models that are discussed in the text. Model-building sessions are indicated by ▽. The *R* factor during the refinement according to procedure II is also included. (*b*) σ_F $\{\sigma_F = [\langle(|F_{obs}| - F_{calc})^2\rangle]^{1/2}\}$ as observed during the refinement and σ_x as applied, in a stepwise manner, in the weighting of the X-ray energy. The resolution limits used in the refinement are shown. The CPU time on a CONVEX C1-XP required for 100 calculations of MDXREF using 6·0–3·0 Å resolution data was 1·6 and 2·5 h for MDXREF with diffraction data between 6·0 and 2·2 Å resolution.

In procedure II, MDXREF was run for only 1·4 ps (700 steps) at 600 K using 6 to 3 Å resolution X-ray data. Here a relatively strong weight was put on the X-ray energy term ($\sigma_x = 60$). Subsequently, 200 steps of EMX were performed using the same data and the same weights. The procedure was continued at 2·5 Å resolution for 0·4 ps using MDXREF with $\sigma_x = 60$ at 600 K followed by 200 steps of EMX with $\sigma_x = 55$. It was stopped at this point, because procedure I gave much better results.

After MDXREF, the model was energy minimized by EMX before any temperature-factor refinement or map calculation and model refitting were performed. Individual temperature-factor refinement was accomplished by using Agarwal's fast Fourier refinement program *FFTREF* (Agarwal, 1978). Three sessions of model fitting in $2m|F_o| - D|F_c|,\alpha_c$ (Read, 1986) maps were carried out using the molecular-graphics program *FRODO* (Jones, 1978) on an Evans & Sutherland PS390 colour-graphics system. Also, omit maps were calculated in which troublesome regions of the model were omitted from the structure-factor calculation, in order to minimize bias from incorrectly placed atoms. The resolution limits used for temperature-factor refinement and electron density map calculation were always the same as in the latest previous EMX run. The stereochemistry of the model was analyzed by looking at the distribution of all energy terms throughout the model (program kindly provided by J. Postma). Large deviations from the mean energy indicated local regions of bad stereochemistry. These parts of the model were inspected on the graphics system and refitted according to the electron density. The insertions and the C-terminal addition were introduced at point *G* (see Fig. 1) in the refinement. At point *H*, the protein model was completed by adding the seven N-terminal residues. The calculations for the refinement were performed on a CONVEX C1-XP of the BIOSON Research Institute.*

Results

The application of molecular-dynamics crystallographic refinement in the case of the thermitase:eglin c complex shows that large errors in the starting model can be properly corrected. The starting model for refinement was derived from the molecular-replacement solution, for which the structure of the subtilisin Carlsberg:eglin c complex was used. The sequence identity between thermitase and subtilisin Carlsberg is only 47%. Besides the amino-acid substitutions there are also a number of insertions, deletions and additions. These are listed in Table 1. To test the power of MDXREF, the starting model was constructed without manual intervention. In this starting model a large number of bad van der Waals contacts existed owing to the automatic amino-acid substitution procedure used. Furthermore, extremely long bonds were present at deletion sites. There were bond lengths of 3·3 Å between C of Trp24 and N of Asp25, 12·9 Å between C of Asn170 and N of Tyr171, and 7·15 Å between C of Ser 244 and N of Ala245. The crystallographic *R* factor of the starting model was 51% for data between 6 and 3 Å resolution.

The progress of the refinement, monitored by the *R* factor, is shown in Fig. 1. In procedure I, which is discussed below if not specifically stated otherwise, the refinement was continued with MDXREF using data between 10 and 5 Å resolution. The resolution was gradually increased. The starting model, which was obtained without any manual model building, could be refined totally automatically (with one minor exception which will be discussed below) from an *R* factor above 50% to an *R* factor of 24·1% at 2·5 Å resolution (point *G* in Fig. 1). At this stage most of the electron density was readily interpretable. After insertion of the missing residues, Ala49, Gly50, Thr83, Ser260 and Tyr279 at point *G* in Fig. 1, the refinement was continued and reached an *R* factor of 22·8% for data between 6 and 2·5 Å resolution. Subsequent refinement at 2·2 Å resolution brought the *R* factor down to 21·4% at point *H* in Fig. 1. At this point, the seven-residue N-terminal extension could be built into the electron density map in a straightforward manner. No water molecules were included up to point *I*. Here 140 water molecules were located in the electron density map. Another 71 waters were positioned at point *J* with *R* = 20·1% using data from 6 to 2·2 Å resolution. Two high-density positions in the electron density map were identified which appeared to be analogous to the known strong and weak calcium-binding sites in subtilisins (Bode *et al.*, 1987; Pantoliano, Whitlow, Wood, Rollence, Finzel, Gilliland, Poulos & Bryan, 1988). In the last few steps involving EMX and individual temperature-factor refinement, a final *R* factor of 17·9% for data from 6 to 2·2 Å resolution was achieved.

As mentioned above, only one minor model correction was applied manually going from point *O* to point *G* in Fig. 1. The small correction was needed because side-chain mutations from subtilisin Carls-

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1TEC, R1TECSF), and are available in machine-readable form from the Protein Data Bank at Brookhaven or one of the affiliated centres at Melbourne or Osaka. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 37028 (as microfiche). Free copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England. At the request of the authors, the list of structure factors will remain privileged until 21 June 1990.

Table 2. *Energies* (kJ mol^{-1}) *of three models of thermitase:eglin c*

Energy term	Energy At point D'*	At point G	At point K	R.m.s. deviation from ideality‡ at point K
Bond	0.1875×10^4 (0.75)†	0.1422×10^4 (0.57)	0.8894×10^3 (0.34)	Bond length 0.020 Å
Angle	0.5611×10^4 (1.65)	0.4501×10^4 (1.32)	0.2643×10^4 (0.75)	Bond angle 3.16°
Dihedral	0.3165×10^4 (2.12)	0.2901×10^4 (1.94)	0.1787×10^4 (1.15)	Torsion angle 26.10°
Improper dihedral	0.1445×10^4 (1.30)	0.1311×10^4 (1.18)	0.8133×10^3 (0.70)	Trigonal non-planarity 0.024 Å
Lennard–Jones	-0.5954×10^4	-0.7665×10^4	-0.9890×10^4	Planar groups 0.058 Å
Electrostatic	-0.1092×10^5	-0.1116×10^5	-0.1225×10^5	Bad contacts 0.156 Å
No. of non-H atoms	2434	2434	2737	
R factor	24.4% (6–2.5 Å)	24.1% (6–2.5 Å)	17.9% (6–2.2 Å)	

* Points correspond to the labels in Fig. 1. K is the final model.
† The average energy per bond, angle, dihedral and improper dihedral (in kJ mol^{-1}) is given in parentheses.
‡ Calculations were performed using the program *TNT* (Tronrud *et al.*, 1987).

berg to thermitase were performed without checking for any physical impossibility. The lack of convergence after 300 steps of EMX was indicative of a serious model error. Inspecting the energy per residue at this point revealed immediately that one steric clash occurred in the model which could not be corrected by EMX. At point O a large number of short non-bonded interactions were present in the model and thus a large number of high Lennard–Jones energy terms were found. After 300 steps of EMX only two residues remained with very high energies, Lys180 and Trp266. The side chain of Lys180, which was positioned inside the six-membered ring of Trp266, was rotated to a physically reasonable position. Then the refinement was continued. This was the only manual correction needed to obtain a stereochemically reasonable model at 2.5 Å resolution with $R = 24.1\%$, starting from a model with only 47% sequence homology and an R factor above 50%.

During the entire course of the refinement, depicted in Fig. 1, the electron density of the model was inspected three times and the model corrected on a computer-graphics system. At these three points, G, H and J in Fig. 1, the model geometry was analyzed by inspecting the distribution of energies throughout the molecule. Only regions with large energies (*i.e.* more than three to four times the standard deviation above the mean energy) were corrected manually by fitting the model to the electron density. In the first model-building step, at point G in Fig. 1, corrections involved regions that had not reached their final position or that seemed to be 'locked' the wrong way in the electron density. One of the regions involved Trp266, the side chain of which occupied the space for the insertion Ser260. However, at this stage (point G) the map was unambiguous in this region and the insertion Ser260 and the loop 261–266 could be (re-)built easily. Another example is the loop 60 to 68. Here the electron density was less interpretable. Although the electron density in this area could not be interpreted convincingly at this stage, residues 60 to 65 were fitted as well as possible. In the two other model-building sessions, at points H and J in Fig. 1, the large deviations from the mean energies involved an increasing number of valines, threonines, isoleucines and tryptophans. These residues were placed correctly by MDXREF/EMX with respect to their main-chain atoms, but the side chains needed to be rotated 180° about their C^α—C^β bonds. In the case of Trp this involved a repositioning of the ring system (see also Fig. 6*b*). Also peptide planes, which were not (yet) flipped by MDXREF, were found by their local high energies. A Ramachandran plot was calculated at point K and is given in Fig. 2. The Ramachandran

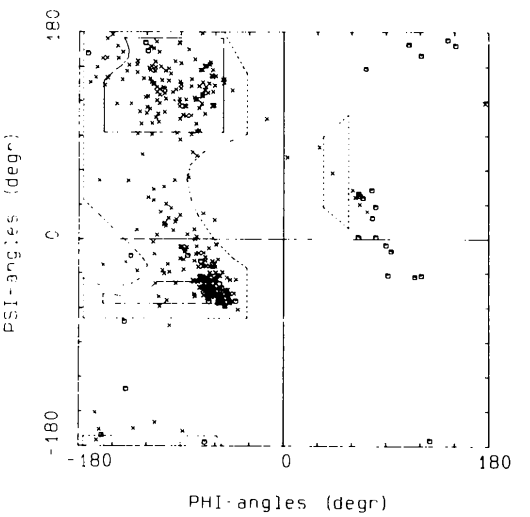

Fig. 2. A Ramachandran plot of the final model, at point K in Fig. 1, for both thermitase and eglin c. The φ–ψ angles of the non-glycine residues are indicated by ×, whereas glycine φ–ψ angles are denoted by □. The areas within the continuous lines define the fully allowed conformations for $\tau(N,C^\alpha,C) = 110°$. The broken lines show the regions obtained by relaxing the van der Waals contact constraints as well as by allowing τ to increase to 115° (Ramachandran & Sasisekharan, 1968). Asp188 ($\varphi = 3.6°$, $\psi = 71.3°$) and Ala49 ($\varphi = -15.0°$, $\psi = 104.5°$) are not within the allowed regions. Here the electron density map was not clear enough to allow better modelling. Also the active-site residue, Asp38 ($\varphi = -165.2°$, $\psi = -151.0°$), is not within the allowed regions; this phenomenon is also observed in the other subtilisins (Bode *et al.*, 1987; McPhalen, Schnebli & James, 1985; McPhalen, Svendsen, Jonassen & James, 1985; Betzel *et al.*, 1988).

plot shows the correctness of the main-chain geometry of the final model. Table 2 shows that the mean energies per residue were quite low and that geometric deviations as indicated by *TNT* (Tronrud, Ten Eyck & Matthews, 1987) were small. Hence, a well-refined model has been obtained with only three manual interventions. Model building could be kept to a minimum (i) by making many large shifts automatically as shown below; (ii) since the density improved considerably as a result of the large shifts which made the model phases, and hence the electron density, more accurate; and (iii) by inspecting mainly residues of higher energy.

Atomic shifts

The *R* factor gives only a very general impression of the course of the refinement and it is therefore interesting to see what kind of coordinate shifts have occurred. After the first series of EMX, EM and MDXREF the number of atoms deviating more than 1 Å from the final model has decreased from 1180 to 536 (the total number of non-H atoms in the model up to point *E* being 2428). The error in C^{α} positions along the polypeptide chain is depicted in Fig. 3 at three points in the refinement: in Fig. 3(*a*) at point *O*, the starting model; in Fig. 3(*b*) at point *B*, after EMX; and in Fig. 3(*c*) at point *G*, after MDXREF and before the first model-building session. Looking at Figs. 3(*a*)–(*c*), it becomes clear that some regions were not (fully) corrected, whereas there are also segments that were corrected within the limits of accuracy. These involve residues 10 to 35, 90 to 130, 210 and 240 and virtually the entire eglin molecule (residues '280' to '342'). It is noteworthy that the errors in the model became more localized; the peaks became sharper, going from Fig. 3(*a*) to 3(*b*) to 3(*c*). This is clear around residues 63, 243 and 270.

The advantage of a more global search in configurational space by molecular dynamics in comparison to finding the nearest local minimum by a least-squares approach is shown in Fig. 4. In those regions where the structure has to be rearranged locally (*e.g.* residues 60 to 68 and 240 to 279) and the model errors frequently exceed 1 Å, the shifts after MDXREF are significantly larger than after EMX only. Whereas the shifts in C^{α} positions in EMX do not generally exceed 1 Å, large shifts are not exceptional at all in MDXREF (Figs. 4*a* and 4*b*). Also the maximum atomic shift per residue in EMX steps (Fig. 4*c*) is a factor two, or more, smaller than the maximum atomic shift per residue due to molecular-dynamics refinement (Fig. 4*d*).

Having discussed general features, it is worthwhile looking at a few specific places, where the differences between molecular-dynamics refinement and conventional refinement became most evident. In the final

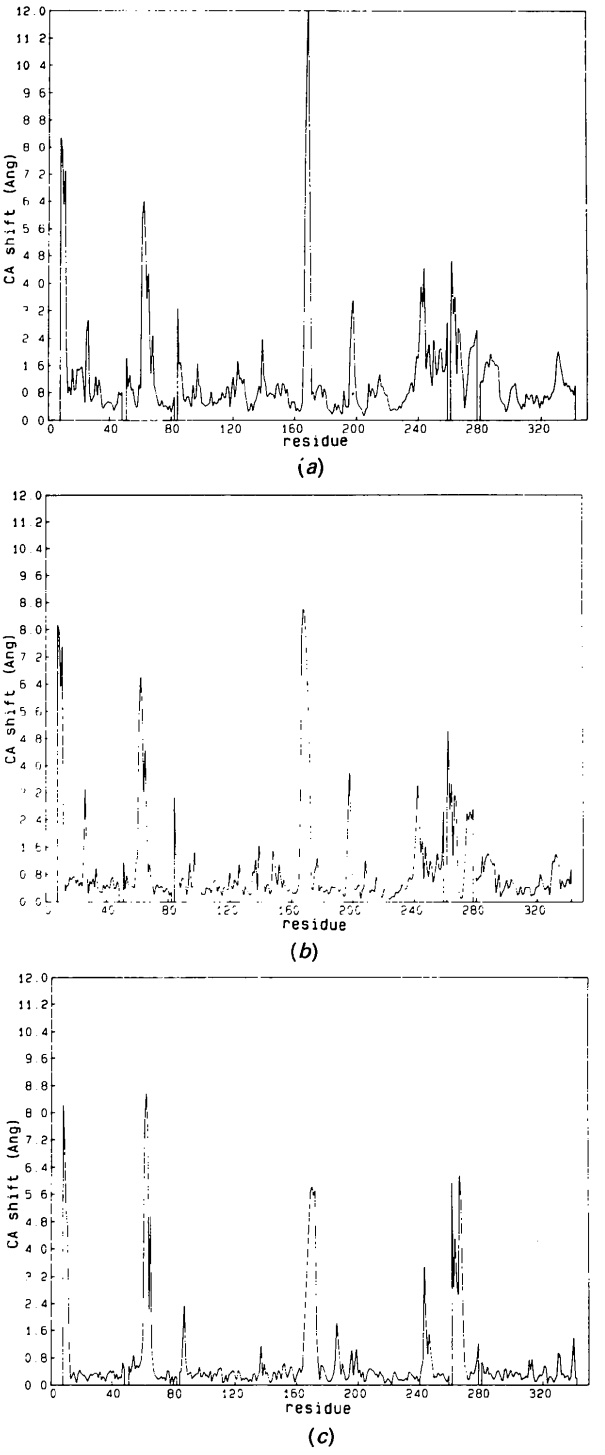

Fig. 3. Deviations in C^{α} positions during refinement. The deviation of C^{α} positions for each residue from the position in the final refined model (at point *K* in Fig. 1) are given for: (*a*) the starting model, point *O* in Fig. 1; (*b*) after 400 steps of EMX, point *B* in Fig. 1; (*c*) before the first model-building session, point *G* in Fig. 1. The thermitase residue numbers run from 1 to 279 and eglin c runs from 280 to 342. The gaps in (*a*), (*b*) and (*c*) indicate missing residues in the models at points *O*, *B* and *G*.

structure of thermitase, the side chain of Tyr274 sticks into the solution having a well-defined density. In the starting model, however, Tyr274 was pointing inwards. Movement of the tyrosine side chain from the interior to the exterior of the protein was blocked by the side chain of Gln22. Fig. 5 shows five stages in the refinement of Tyr274 and Gln22. At point *O*, the start, there is no way out for the tyrosine. At point *E* the tyrosine side chain is highly bent and the C^{α} position is more than 2·4 Å away from its correct position. Furthermore, Gln22 is pushed upwards to give more space to the Tyr274. This situation of Tyr274 trying to squeeze out by pushing Gln22 upwards was present for about 2 ps. Once Tyr274 had sufficient space to swing its hydroxyl group out (point *F*), it took much less time to get Tyr274 and Gln22 in almost their final positions (as can be seen in Fig. 5).

From the electron densities shown in Figs. 5(*c*) and 5(*d*), it can be seen that the position of the side chain of Tyr274 was already visible in the electron density at point *O*. It is also clear that the density for the main chain at the N-terminal side of Tyr274 was not present in the initial map. Fig. 5(*d*) shows that even though Tyr274 is at the surface of the protein the final electron density is well defined. This final position of Tyr274 has been obtained, without any manual intervention, by passing through local high-energy states in the molecular-dynamics refinement. The total shift of the hydroxyl O atom of Tyr274 during EMX and MDXREF steps was 7·2 Å and was carried out fully automatically.

Another example of the power of the MD refinement method is illustrated by the refinement of loop 24 to 27. A deletion of one amino acid, preceding Asp25, occurs in this region in comparison to

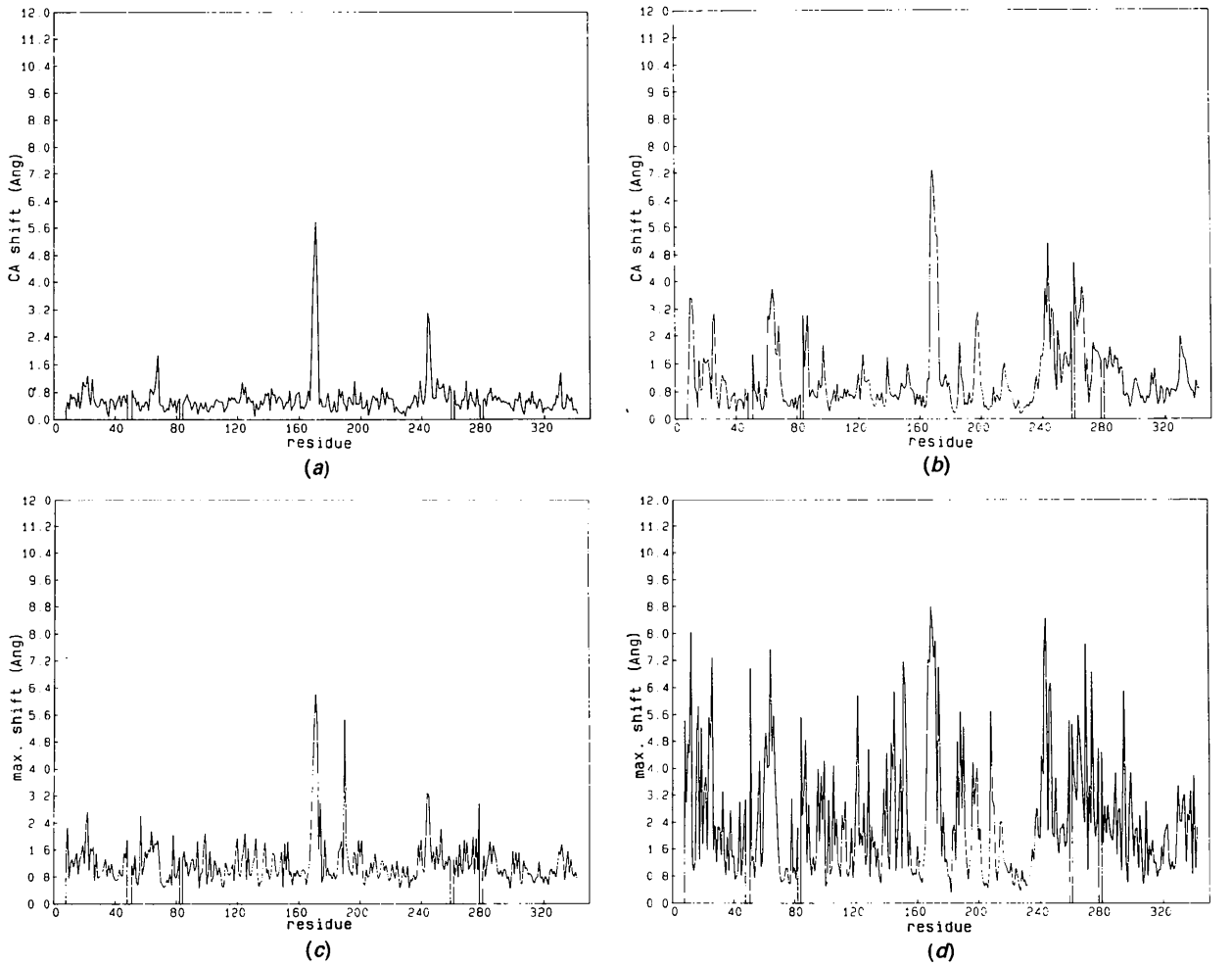

Fig. 4. Applied shifts during the refinement. The applied shifts to C^{α}'s and the maximal atomic shift per residue for thermitase (residue numbers 1–279) and eglin c (residues 280–342) are shown for two stages. (*a*) Shifts in C^{α} positions during EMX, going from point *O* to *B* in Fig. 1. (*b*) Shifts in C^{α} positions after EMX and MDXREF, going from point *O* to *G* in Fig. 1. (*c*) Maximal shift per residue during EMX, going from point *O* to *B* in Fig. 1. (*d*) Maximal shift per residue after EMX and MDXREF, going from point *O* to *G* in Fig. 1.

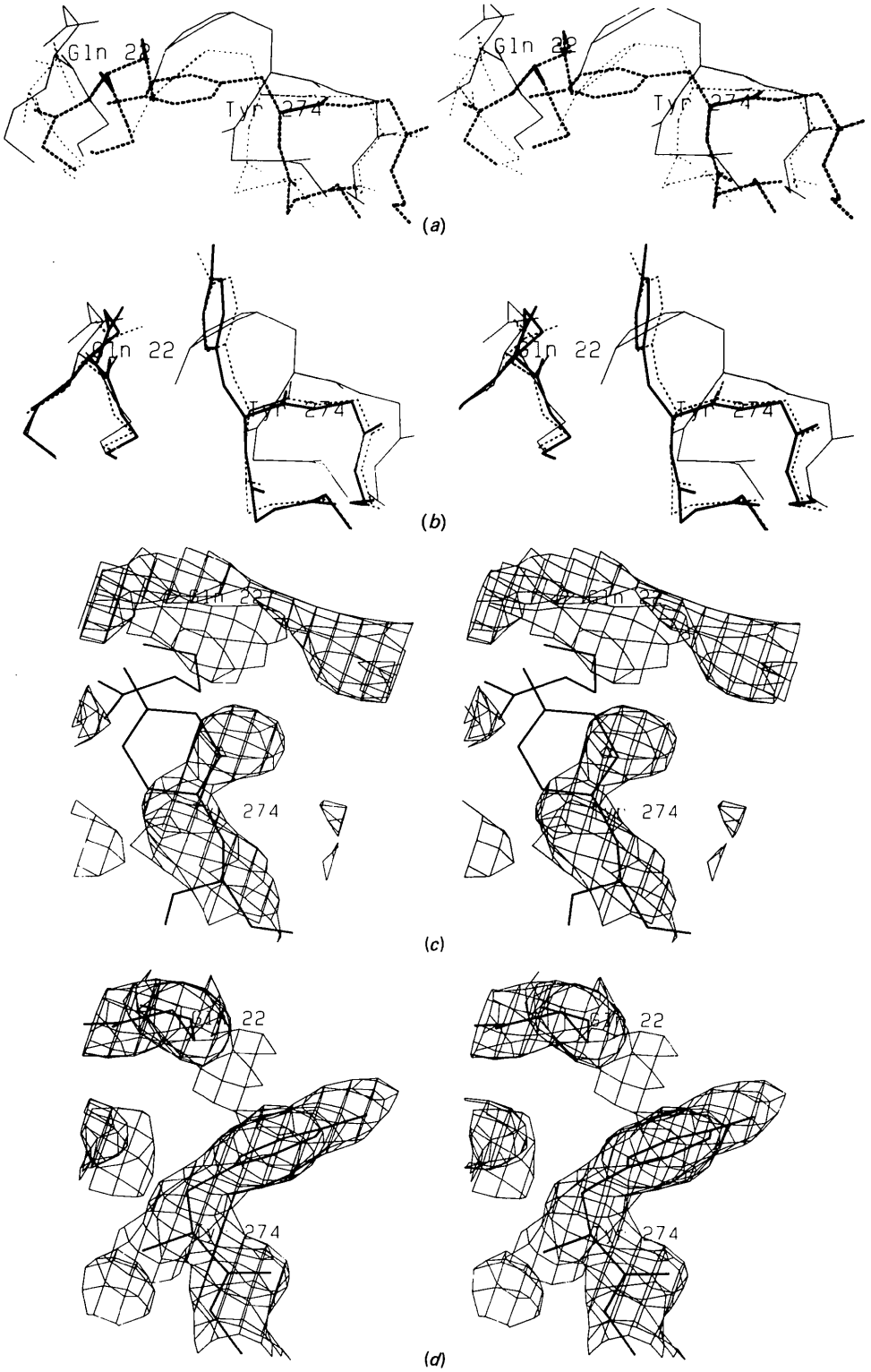

Fig. 5. Conformations of Tyr274 and Gln22 during the refinement. (*a*) Subsequent stages in the refinement of Tyr274 and Gln22. The model at point *O* (see Fig. 1) is indicated by ---, at point *C* by ---, at point *E* by —. (*b*) The model at point *E* (see Fig. 1) is indicated by —, at point *F* by ---, at point *K* by —. (*c*) The model at point *O* with the $2m|F_o| - D|F_c|,\alpha_c$ electron density map at point *O*, contoured at $1{\cdot}1\sigma$. (*d*) The final electron density map contoured at the same level, and the final model, *i.e.* at point *K* in Fig. 1. The figures show that at point *F* Gln22 is near its final position and the main chain around residue 274 is basically in the right conformation. At point *G* Gln22 and Tyr274 were, within the limits of accuracy, at their final positions.

subtilisin Carlsberg. In thermitase this loop adopts a different conformation. Figs. 6(*a*) and 6(*b*) show this loop at various stages in the refinement. In Figs. 6(*c*)–(*e*), they are shown with the corresponding electron densities at points *O*, *G* and *K*, respectively, of the refinement. Figs. 6(*a*) and 6(*c*) give an impression of the magnitude of the model errors that can be corrected by the molecular-dynamics method. From Fig. 6(*a*) it can be seen that the major correction of the positioning of Asp25 is accomplished during the refinement at low resolution. Whereas the density is discontinuous at point *O*, after EMX/MDXREF at point *G*, the main-chain electron density is clearly visible. At point *G* the loop Trp24-Asp25-Ile 26-Ala27 is essentially in the right place. The shifts in C^{α} positions of Asp25 and Ile26 from point *O* to *G* in Fig. 1 were 2·8 and 3·1 Å, respectively, and the maximal shifts found for these residues were 4·9 Å for $O^{\delta 1}$ of Asp25 and 7·3 Å for $C^{\delta 1}$ of Ile26.

The starting position of the side chain of Asp25 is very near to the correct position and electron density of Ile26, as seen by inspecting the models at points *O* and *K*, and their electron densities (Figs. 6*c* and 6*e*). Quite surprisingly, this caused no erroneous positioning of Asp25 and Ile26 by MDXREF. In procedure I, Asp25 is apparently given sufficient time to move 'upwards' to its correct density after which there is space enough for Ile26 to be correctly positioned. That Asp25 is corrected before Ile26 starts to move can be seen from the intermediate models displayed in Figs. 6(*a*) and 6(*b*).

Turning to procedure II, it may be recalled that this protocol takes less time to arrive at a comparable *R* factor (compare *D'* and *G* in Fig. 1). However, the result of procedure II for correcting loop 24–27 is very poor. Fig. 7(*a*) shows the different models obtained at points *O*, *D'* and *K*. In the case of procedure II there has been a much tighter restraint on the X-ray information. This caused Asp25 to be 'locked' in the wrong electron density, occupying the space for Ile26. Also, from Table 2, which gives the mean energies over all atoms of the

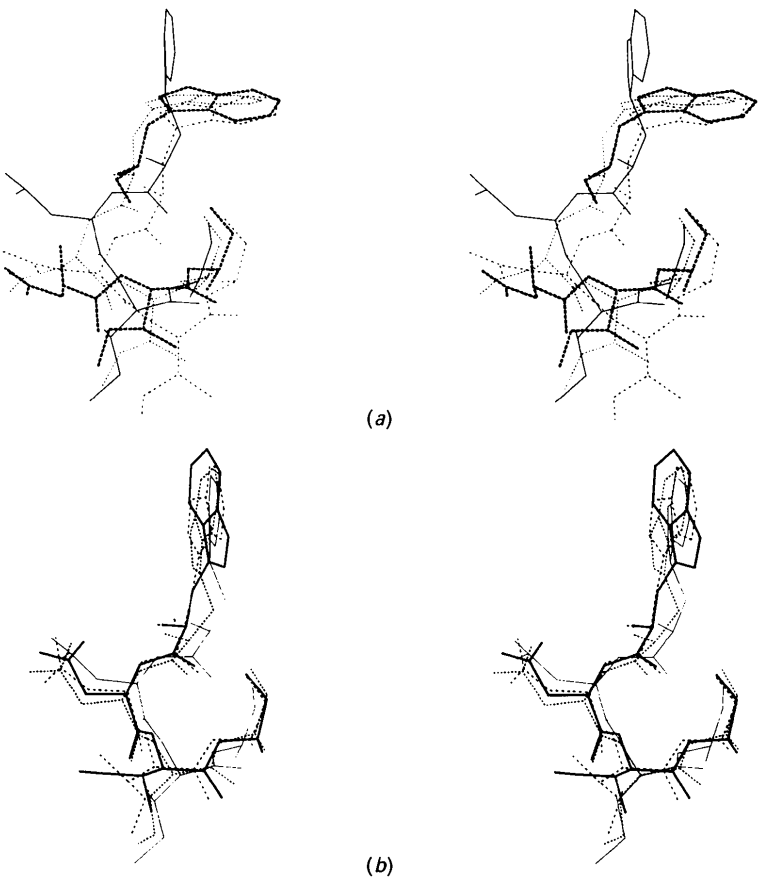

Fig. 6. Conformations of the loop 24–27 during the refinement. (*a*) and (*b*) show different models of the loop Trp24, Asp25, Ile26 and Ala27 along the refinement. (*a*) The models at point *O* (see Fig. 1), thick dashed lines (– – –); at point *B*, thin dashed lines (– – –); at point *C*, thin sparsely dashed lines (– –); and at point *D*, thin solid lines (—). The gap in the main chain of the model at point *O* indicates the deletion site between Trp24 and Asp25. (*b*) The models at point *D*, again thin solid lines (—); at point *F*, dashed lines (– – –); at point *G*, sparsely dashed lines (– –); and at point *K*, thick solid lines (—).

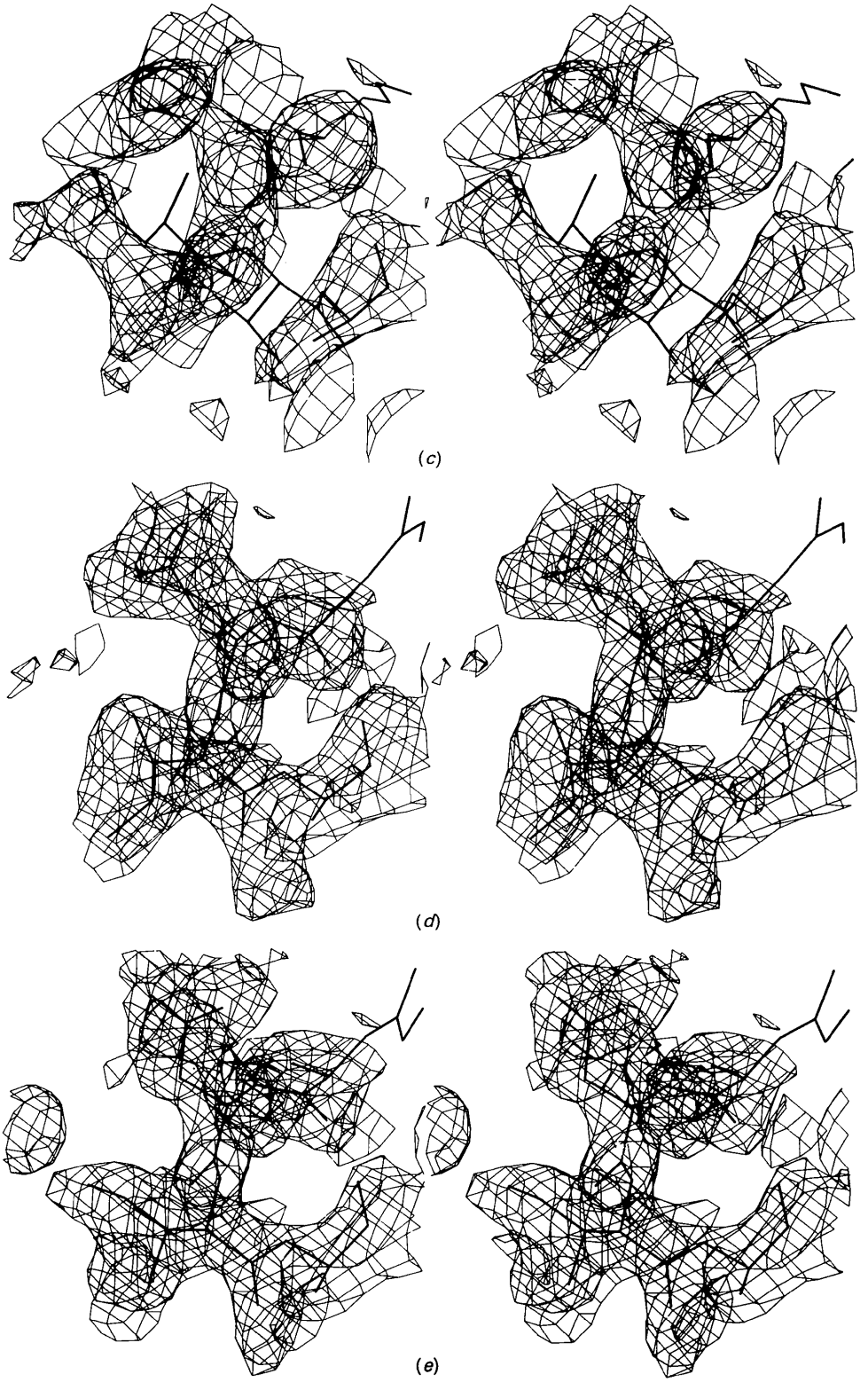

Fig. 6 (*cont.*) (*c*) The model at point *O* with its corresponding $2m|F_o| - D|F_c|,\alpha_c$ electron density, contoured at $1{\cdot}1\sigma$. (*d*) The model at point *G* with its corresponding electron density. (*e*) The final model, *i.e.* at point *K*, in its final electron density at 2·2 Å resolution. After refinement at 2·2 Å resolution the density indicated a staggered conformation, which is also stereochemically preferred, of the Ile26 side chain. The model was corrected manually. The second manual correction after refinement at 2·2 Å resolution was the flip of the ring system of Trp24.

models at D', G and K, it can be seen that procedure I produces lower energies than procedure II. Clearly, the gentle approach in procedure I leads to better results than procedure II, where the search through conformational space is restricted by the larger weight on the X-ray term.

Discussion

We have presented the application of the crystallographic molecular-dynamics refinement to a model obtained by the molecular-replacement procedure. The starting model could be refined at 2·5 Å resolution, without any prior model building. Large corrections to the model were applied by EMX and MDXREF. However, using procedure I, a small number of loops, *e.g.* loops around residues 10, 63, 170, 243 and 265 (Fig. 3*c*), were not positioned correctly. Manual corrections are thus still necessary. All corrections applied by MDXREF and EMX result in a better model and a better electron density map, which is more easily interpreted. The big advantage of the procedure described above is that from a model with only 47% amino-acid sequence identity, a refined model of the new protein is obtained almost fully automatically. The interpretation of the electron density map is first performed after refinement at 2·5 Å resolution when an R factor of 24·1% was reached. As discussed, most of the model errors had been corrected at that stage or, where the model was still incorrect, the map had improved significantly and most of the remaining errors could be corrected by manual model fitting. It is also important to have an easy survey of the quality of the model. By studying the distribution of energies throughout the molecule, local 'hot' regions with high-energy terms can be located. These local 'hot' regions contain model errors and need specific

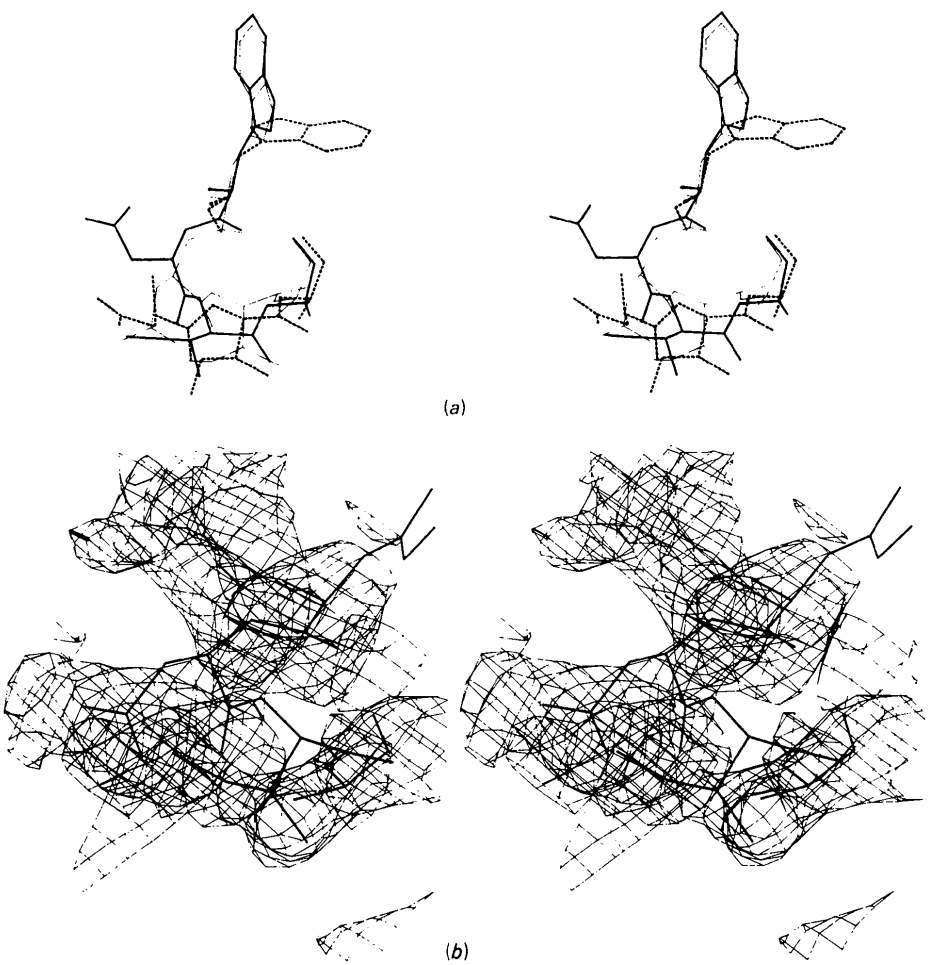

Fig. 7. Comparison between the results of procedures I and II for the loop 24–27. (*a*) Thick dashed lines (---), the starting model at point O (see Fig. 1); thick lines (—), the model at point K, obtained by procedure I; thin lines (—), the model at point D', obtained by procedure II. (*b*) The $2m|F_o| - D|F_c|,\alpha_c$ electron density map, contoured at $1{\cdot}1\sigma$ and the model at point D' obtained by procedure II. Comparison with Figs. 6(*d*) and 6(*e*) shows that a better result is obtained by applying procedure I.

attention during the model-building sessions. After refinement at higher resolution, *i.e.* 2·5 or 2·2 Å resolution, small coordinate errors, such as incorrectly placed side chains of Thr or Val, can easily be traced by inspecting the distribution of the various energies. The combination of molecular-dynamics crystallographic refinement with a good indicator of the quality of the model reduces the human time required for model inspection and refitting significantly.

In conclusion, we have shown that molecular-dynamics crystallographic refinement (Brünger *et al.*, 1987; Fujinaga *et al.*, 1989) is indeed a powerful method. The analysis of the energy distribution throughout the molecule is most helpful when inspecting the model for manual model correction. Rearrangements of loops have occurred during the refinement of the thermostable serine protease, thermitase, complexed with the serine protease inhibitor, eglin c. However, the probability of the occurrence of rearrangements depends on the strategy of the refinement, as shown by the comparison of procedures I and II. Our studies indicate that running MDXREF initially for a long time at low resolution, with relatively low weights on the X-ray restraints, before the structure is refined at higher resolution, yields better results.

We are grateful to Professors J. Drenth, H. J. C. Berendsen and W. F. van Gunsteren for stimulating discussions, H. Kelders for crystallization experiments, Dr Schnebli, Ciba Geigy, for a gift of eglin c and Dr J. P. M. Postma for providing the energy-analysis program. MF was supported by the Alberta Heritage Foundation for Medical Research. This research was carried out with support of the Dutch Foundation for Chemical Research (SON) and the BIOSON Research Institute with financial aid from the Netherlands Organisation for the Advancement of Pure Research (NWO).

References

AGARWAL, R. C. (1978). *Acta Cryst.* A**34**, 791–809.

ALBER, T., DAO-PIU, S., WILSON, K., WOZNIAK, J. A., COOK, S. P. & MATTHEWS, B. W. (1987). *Nature (London)*, **330**, 41–46.

ARGOS, P., ROSSMANN, M. G., GRAU, U. M., ZUBER, H., FRANK, G. & TRATSCHIN, J. D. (1979). *Biochemistry*, **18**, 5698–5703.

ARNDT, U. W. (1982). *Nucl. Instrum. Methods*, **201**, 21–25.

BERENDSEN, H. J. C., POSTMA, J. P. M., VAN GUNSTEREN, W. F., DINOLA, A. & HAAK, J. R. (1984). *J. Chem. Phys.* **81**, 3684–3690.

BETZEL, C., PAL, G. P. & SAENGER, W. (1988). *Acta Cryst.* B**44**, 163–172.

BODE, W., PAPAMOKOS, E. & MUSIL, D. (1987). *Eur. J. Biochem.* **166**, 673–692.

BRÜNGER, A. T., KURIYAN, J. & KARPLUS, M. (1987). *Science*, **235**, 458–460.

BRYAN, P. N., ROLLENCE, M. L., PONTOLIANO, M. W., WOOD, J., FINZEL, B. C., GILLILAND, G. L., HOWARD, A. J. & POULOS, T. L. (1986). *Proteins*, **1**, 326–334.

CROWTHER, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. ROSSMANN, pp. 173–178. New York: Gordon and Breach.

CROWTHER, R. A. & BLOW, D. M. (1967). *Acta Cryst.* **23**, 544–548.

DAUTER, Z., BETZEL, C., HÖHNE, W. E., INGELMAN, M. & WILSON, K. S. (1988). *FEBS Lett.* **236**, 171–178.

FRÖMMEL, C. & HÖHNE, W. E. (1981). *Biochim. Biophys. Acta*, **670**, 25–31.

FUJINAGA, M., GROS, P. & VAN GUNSTEREN, W. F. (1989). *J. Appl. Cryst.* **22**, 1–8.

GUNSTEREN, W. F. VAN & BERENDSEN, H. J. C. (1987). *BIOMOS. Biomolecular Software*. Laboratory of Physical Chemistry, Univ. of Groningen, The Netherlands.

GUSEK, T. W. & KINSELLA, J. E. (1987). *Biochem. J.* **246**, 511–517.

JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

KATZ, B. A. & KOSSIAKOFF, A. (1986). *J. Biol. Chem.* **261**, 15480–15485.

KLEINE, R. (1982). *Acta Biol. Med. Germ.* **41**, 89–102.

MCPHALEN, C. A., SCHNEBLI, H. P. & JAMES, M. N. G. (1985). *FEBS Lett.* **188**, 55–58.

MCPHALEN, C. A., SVENDSEN, I., JONASSEN, I. & JAMES, M. N. G. (1985). *Proc. Natl Acad. Sci. USA*, **82**, 7242–7246.

MATTHEWS, B. W. (1987). *Biochemistry*, **26**, 6885–6888.

MELOUN, B., BAUDYS, M., KOSTKA, K., HAUSDORF, G., FRÖMMEL, C. & HÖHNE, W. E. (1985). *FEBS Lett.* **183**, 195–199.

PANTOLIANO, M. W., WHITLOW, M., WOOD, J. F., ROLLENCE, M. L., FINZEL, B. C., GILLILAND, G. L., POULOS, T. L. & BRYAN, P. N. (1988). *Biochemistry*, **27**, 8311–8317.

PAUPTIT, R. A., KARLSSON, R., PICOT, D., JENKINS, J. A., NIKLAUS-REIMER, A.-S. & JANSONIUS, J. N. (1988). *J. Mol. Biol.* **199**, 525–537.

PERUTZ, M. F. (1978). *Science*, **201**, 1187–1191.

PFLUGRATH, J. W. & MESSERSCHMIDT, A. (1986). *MADNES User's Guide*. Max-Planck-Institut für Biochemie, Martinsried, Federal Republic of Germany.

RAMACHANDRAN, G. N. & SASISEKHARAN, V. (1968). *Adv. Protein Chem.* **23**, 283–438.

READ, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

READ, R. J. (1988). Personal communication.

RENETSEDER, R., DIJKSTRA, B. W., KALK, K. H., VERPOORTE, J. & DRENTH, J. (1986). *Acta Cryst.* B**42**, 602–605.

RYCKAERT, J.-P., CICCOTTI, G. & BERENDSEN, H. J. C. (1977). *J. Comput. Phys.* **23**, 327–341.

SEEMÜLLER, U., FRITZ, H. & EULITZ, M. (1981). *Methods Enzymol.* **80**, 804–816.

TEPLYAKOV, A. V., STROKOPYTOV, B. V., KURANOVA, T. P., POPOV, A. N., HARUTYUNYAN, E. G., VAINSHTEIN, B. K., FRÖMMEL, C. & KHENE, V. (1986). *Sov. Phys. Crystallogr.* **31**, 553–556.

TRONRUD, D. E., TEN EYCK, L. & MATTHEWS, B. W. (1987). *Acta Cryst.* A**43**, 489–501.

VRIEND, G. (1989). In preparation.