neighborhood. We suggest that the second neighborhood of $T_{\mathbf{h},\mathbf{k}}$ should consist of the set $\{T_{\mathbf{h},\mathbf{k}}, T_{\mathbf{h}}, T_{\mathbf{k}}, T_{\mathbf{h}+\mathbf{k}}, T_{2\mathbf{h},2\mathbf{k}}, T_{\mathbf{h}-\mathbf{k}}, T_{\mathbf{h},\mathbf{h}}, T_{\mathbf{k},\mathbf{k}}, T_{\mathbf{h}+\mathbf{k},\mathbf{h}+\mathbf{k}}, T_{2\mathbf{h}}, T_{2\mathbf{k}}, T_{2(\mathbf{h}+\mathbf{k})}\}$. Besides the usual constraints $|T_{\mathbf{h},\mathbf{k}}| = R_{\mathbf{h}}R_{\mathbf{k}}R_{\mathbf{h}+\mathbf{k}}$, $T_{\mathbf{h}} = N^{1/2}R_{\mathbf{h}}^2$ *etc.*, one must also impose the following constraint: let $T_{\mathbf{h},\mathbf{k}} = T_1 e^{i\varphi_1}$, $T_{2\mathbf{h},2\mathbf{k}} = T_2 e^{i\varphi_2}$, $T_{\mathbf{h},\mathbf{h}} = T_3 e^{i\varphi_3}$, $T_{\mathbf{k},\mathbf{k}} = T_4 e^{i\varphi_4}$, $T_{\mathbf{h}+\mathbf{k},\mathbf{h}+\mathbf{k}} = T_5 e^{i\varphi_5}$; then there exist variables $\varphi_{\mathbf{h}}$, $\varphi_{\mathbf{k}}$, $\varphi_{\mathbf{h}+\mathbf{k}}$, $\varphi_{2\mathbf{h}}$, $\varphi_{2\mathbf{k}}$, $\varphi_{2(\mathbf{h}+\mathbf{k})}$ so that

$$\begin{aligned}\varphi_1 &= \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} - \varphi_{\mathbf{h}+\mathbf{k}}\\ \varphi_2 &= \varphi_{2\mathbf{h}} + \varphi_{2\mathbf{k}} - \varphi_{2(\mathbf{h}+\mathbf{k})}\\ \varphi_3 &= 2\varphi_{\mathbf{h}} - \varphi_{2\mathbf{h}}\\ \varphi_4 &= 2\varphi_{\mathbf{k}} - \varphi_{2\mathbf{k}}\\ \varphi_5 &= 2\varphi_{\mathbf{h}+\mathbf{k}} - \varphi_{2(\mathbf{h}+\mathbf{k})}\end{aligned} \tag{16}$$

thereby imposing the constraint

$$\varphi_3 + \varphi_4 + \varphi_5 = 2\varphi_1 - \varphi_2. \tag{17}$$

A similar model can be realized for the quartet. It is now clear that the basic quantities would then be

$$\begin{aligned}m(\mathbf{h},\mathbf{k},\mathbf{l}) = K_4^{-1} \int & f(\mathbf{u}_\alpha) f(\mathbf{v}_\alpha) f(\mathbf{w}_\alpha) f(\mathbf{u}_\alpha - \mathbf{v}_\alpha)\\ & \times f(\mathbf{u}_\alpha - \mathbf{w}_\alpha) f(\mathbf{v}_\alpha - \mathbf{w}_\alpha)\\ & \times \exp\{2\pi i[\mathbf{h}\,.\,\mathbf{u}_\alpha + \mathbf{k}\,.\,\mathbf{v}_\alpha + \mathbf{l}\,.\,\mathbf{w}_\alpha]\}\\ & \times d\mathbf{u}_\alpha\, d\mathbf{v}_\alpha\, d\mathbf{w}_\alpha\end{aligned}$$

$$\begin{aligned}&= [\langle A_{\mathbf{q}_1} A_{\mathbf{q}_2} A_{\mathbf{q}_3} A_{\mathbf{q}_1+\mathbf{q}_2+\mathbf{h}+\mathbf{k}+\mathbf{l}} A_{\mathbf{q}_1+\mathbf{q}_2+\mathbf{q}_3+\mathbf{k}+\mathbf{l}}\\ &\quad \times A_{\mathbf{q}_1+\mathbf{q}_3+\mathbf{l}} \rangle_{\mathbf{q}_1,\mathbf{q}_2,\mathbf{q}_3}]\\ &\quad \times [\langle A_{\mathbf{q}_1} A_{\mathbf{q}_2} A_{\mathbf{q}_3} A_{\mathbf{q}_1+\mathbf{q}_2} A_{\mathbf{q}_1+\mathbf{q}_3} A_{\mathbf{q}_1+\mathbf{q}_2+\mathbf{q}_3} \rangle_{\mathbf{q}_1,\mathbf{q}_2,\mathbf{q}_3}]^{-1}\end{aligned}$$

where $A_{\mathbf{q}} = (R_{\mathbf{q}}^2 - 1)$.

Let us finally notice that our model can easily be combined with additional chemical information if the latter is available.

References

BROSIUS, J. (1978). Doctoral thesis. Kathol. Univ. Leuven, Belgium.

BROSIUS, J. (1985). *Acta Cryst.* A**41**, 613-617.

BROSIUS, J. (1989). In preparation.

FISHER, J., HANCOCK, H. & HAUPTMAN, H. (1970). Rep. No. 7132. Naval Research Laboratory, Washington, DC, USA.

GIACOVAZZO, C. (1977). *Acta Cryst.* A**33**, 527-531.

GILMORE C. J. & HAUPTMAN, H. (1985). *Acta Cryst*, A**41**, 457-462.

HAUPTMAN, H. (1964). *Acta Cryst.* **17**, 1421-1433.

HAUPTMAN, H. (1985). *Acta Cryst.* A**41**, 454-457.

HAUPTMAN, H., FISHER, J., HANCOCK, H. & NORTON, D. (1969). *Acta Cryst.* B**25**, 811-814.

HAUPTMAN, H. & KARLE J. (1958). *Acta Cryst.* **11**, 149-157.

KARLE, J. (1970). *Acta Cryst.* B**26**, 1614-1617.

KARLE, J. & HAUPTMAN, H. (1957). *Acta Cryst.* **10**, 515-524.

VAUGHAN, P. A. (1958). *Acta Cryst.* **11**, 111-115.

VAUGHAN, P. A. (1959). *Acta Cryst.* **12**, 981-987.

Acta Cryst. (1989). A**45**, 471-477

Application of Molecular Dynamics in the Crystallographic Refinement of Colicin A

BY J. P. M. POSTMA, M. W. PARKER AND D. TSERNOGLOU

European Molecular Biology Laboratory, Meyerhofstrasse 1, *Postfach* 10.2209, 6900 *Heidelberg, Federal Republic of Germany*

(*Received* 19 *December* 1988; *accepted* 21 *February* 1989)

Abstract

Crystallographic refinement based on molecular dynamics (MD) has been applied to a 2·5 Å resolution X-ray structure of the pore-forming fragment of colicin A. The crystallographic *R* factor was reduced from 48 to 23% with a concomitant improvement in stereochemical parameters. The method considerably speeded up the refinement process but was associated with some pitfalls. In particular, some badly fitted segments of the structure required manual rebuilding, even after MD refinement and some problems with weighting schemes were encountered. Analysis of the effects of the refinement and ideas for improvements are presented.

Introduction

Crystallographic refinement of macromolecules is a laborious and often demanding task requiring extensive use of human and computing resources. Typically, conventional refinement methods require many cycles of c.p.u.-intensive computations interspersed by sessions of model rebuilding into difference electron density maps using interactive computer graphics. A major source of difficulty is the non-linear relationship between X-ray diffraction data and atomic parameters which can cause refinement to become trapped in false minima. The radius of convergence in conventional least-squares methods is about a quarter of the minimum Bragg spacing in the

0108-7673/89/070471-07$03.00

data set (Jack & Levitt, 1978). Thus atoms displaced by more than about 1 Å require rebuilding.

The recent application of restrained molecular dynamics appears to offer a great leap forward in refinement procedures (Brünger, Kuriyan & Karplus, 1987; Fujinaga, Gros & van Gunsteren, 1989; van Gunsteren, 1988). The molecular dynamics technique, in contrast to energy minimization, allows a molecular system to explore a greater range of conformational space due to the available kinetic energy. A recent example showed that atoms can move more than 10 Å towards their correct position without manual intervention (Gros, Fujinaga, Dijkstra, Kalk & Hol, 1988).

We have recently completed the refinement of an atomic model of the pore-forming fragment of colicin A by the methodology of molecular dynamics and energy minimization using the molecular dynamics package *GROMOS* (van Gunsteren & Berendsen, 1983, 1987; see Fujinaga *et al.*, 1989). The crystallographic refinement was dramatically speeded up by the procedure but nonetheless was accompanied by a number of problems. In this paper we present the results of the refinement, identify problems encountered using the new procedure and suggest some improvements in applying molecular dynamics to crystallographic refinement.

Theory and method

Molecular dynamics (MD) describes the time evolution of a system of interacting particles. The method is based on the numerical integration of Newton's equations of motion. In the case of a molecular system, the constituent particles build up a molecular arrangement through bonds based on a specific stereochemistry for all atom types.

A critical parameter in the MD method is the nature of the potential function $V(\mathbf{r})$ which describes bonded and non-bonded interactions. In our work we made use of the function

$$\begin{aligned} V(\mathbf{r}) = & \sum_{\text{bonds}} \tfrac{1}{2}K_b(b-b_0)^2 + \sum_{\text{angles}} \tfrac{1}{2}K_\theta(\theta-\theta_0)^2 \\ & + \sum_{\text{torsions}} \tfrac{1}{2}K_\zeta(\zeta-\zeta_0)^2 \\ & + \sum_{\text{dihedrals}} K_\varphi[1+\cos(n\varphi-\delta)] \\ & + \sum_{\substack{\text{non-bonded pairs} \\ (i,j)}} [C_{12}(i,j)/r_{i,j}^{12} - C_6(i,j)/r_{i,j}^6 \\ & + q_iq_j/4\pi\varepsilon_0\varepsilon_r r_{i,j}]. \end{aligned} \tag{1}$$

This potential consists of bonded interactions (bond, angle, torsion and dihedral components) and non-bonded interactions (van der Waals and electrostatic). This function, encoded in the *GROMOS* MD suite of programs, has been applied with encouraging results to a wide variety of chemical and biochemical problems (van Gunsteren, 1988).

The power of the MD technique lies in its ability to search large areas of conformational space in order to locate global minimum conformations or to pass through multiple minima. It is the available kinetic energy related to the macroscopic temperature T of the system which allows energy barriers of the order of kT to be surmounted. This ability can be contrasted with energy minimization (EM) methods which operate at $T=0$ and, as a result, can generally only find minimum energy conformations which closely resemble the initial one. The amount of conformational space to be explored can be reduced significantly by incorporation of experimental data into the potential function $V(\mathbf{r})$ (restrained MD). In this technique the experimental data are 'translated' into a suitable potential function with a weight factor (force constant) that can be varied to strengthen or weaken the effect of the experimental constraints. A classical example is the expression of individual proton-proton distances derived from 2D-NMR experiments into a distance-dependent potential (van Gunsteren, Kaptein & Zuiderweg, 1984); van Gunsteren, Boelens, Kaptein, Scheek & Zuiderweg, 1985; Brünger, Clore, Gronenborn & Karplus, 1986).

Application of the MD method to the crystallographic refinement of protein structures has been very recent (Brünger *et al.*, 1987; Fujinaga *et al.*, 1989). In this application, the residual minimized in conventional least-squares refinement (Agarwal, 1978), is translated into a potential function P,

$$P = \sum_{\mathbf{h}} [|F_o(\mathbf{h})|^2 - |F_c(\mathbf{h})|^2] \tag{2}$$

where F_o and F_c are observed and calculated structure factors and $\mathbf{h}$ is a reciprocal-lattice point. The global potential V_{total} then becomes

$$V_{\text{total}}(\mathbf{r}) = V(\mathbf{r}) + V_{\text{X-ray}}(\mathbf{r}) \tag{3}$$

where

$$V_{\text{X-ray}} = K_{\text{X-ray}}P. \tag{4}$$

The force constant $K_{\text{X-ray}}$ governs the relative strength of the X-ray information with respect to the molecular potential and can be written as (Fujinaga *et al.*, 1989)

$$K_{\text{X-ray}} = K/\sigma^2 \tag{5}$$

where σ^2 is equal to

$$\sum_{\mathbf{h}} [F_c(\mathbf{h}) - F_o(\mathbf{h})]^2/\text{nref}$$

(nref is the number of reflections) and K is a multiplication factor also carrying the correct units. Great care should be taken in the choice of the force constant used; too high a value may produce good agreement with experimental observations to the detriment

of restricting the structure to an improbable region of conformational space.

A flow chart of the methodology used in the *GROMOS* suite of programs is given in Fig. 1. Observed structure factors are expanded in space group *P*1 using the utility programs *PROISF* and *PROESF*. The utility program, *PRE-GROMOS*, written by one of us (JPMP), takes care of all necessary molecular data conversions, generates a *GROMOS* molecular topology file, appends polar hydrogens to the initial structure and prepares all necessary job and input files. The initial model is energy minimized in order to release energetic strain. Assignment of velocities to individual atoms was done from a Maxwellian distribution corresponding to the desired temperature. Because the recalculation of the X-ray potential *P* and its derivatives is time consuming, the harmonic approximation of Jack & Levitt (1978) is used in which the potential is only updated if any atom moves by more than a few tenths of an ångström. The MD calculation is normally continued until the conventional *R* factor converges. The resulting structure is then subjected to energy minimization in order to release the system's kinetic energy.

The case of the pore-forming fragment of colicin A

The thermolytic fragment of colicin A carries the pore-forming activity of the bacterial toxin. The fragment crystallized in space group $P4_32_12$ with cell parameters $a = b = 73{\cdot}0$, $c = 171{\cdot}3$ Å and contained two monomers per asymmetric unit (Tucker, Pattus & Tsernoglou, 1986). Each monomer consists of 204 amino-acid residues. The crystals diffract moderately to 3 Å resolution but diffraction spots can be observed out to at least 2·5 Å resolution. An electron density map was calculated to 3 Å resolution on the basis of isomorphous/anomalous phasing and solvent flattening techniques (Parker, Pattus, Tucker & Tsernoglou, 1989). The map was of sufficient quality to allow an unambiguous chain tracing which served as a starting model for MD refinement.

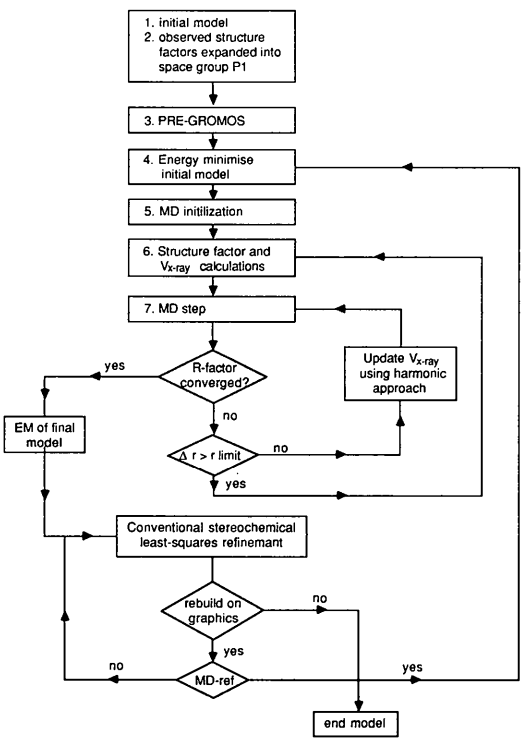

Fig. 1. Flow diagram of methodology used in the MD X-ray refinement process.

Results and discussion

The initial model of colicin was subjected to 50 steps of steepest descent and subsequently energy minimized using the conjugate gradient method. No account was taken of the experimental data at this stage. The MD refinement was initialized at a temperature of 300 K and an isothermal MD algorithm (Berendsen, Postma, van Gunsteren, DiNola & Haak, 1984) was used to keep the overall temperature constant, using a temperature relaxation time $\tau_t = 0{\cdot}1$ ps. The MD time step was chosen to be 0·002 ps. The *SHAKE* algorithm was used to restrain bond lengths (Ryckaert, Cicotti & Berendsen, 1977; Ryckaert, 1985). A cut-off radius of 8 Å was used in calculating non-bonded interactions. The present *GROMOS* MD X-ray refinement package does not yet include *B*-factor refinement. Therefore each atom was assigned a *B* factor of 15 Å^2. Structure-factor gradients were updated when an atom moved by more than 0·2 Å. Experimental data were restricted to the resolution range 6·0–2·5 Å in order to avoid the significant contributions of solvent structure at lower resolution. σ^2 in (5) was calculated to be (approximately) 300.

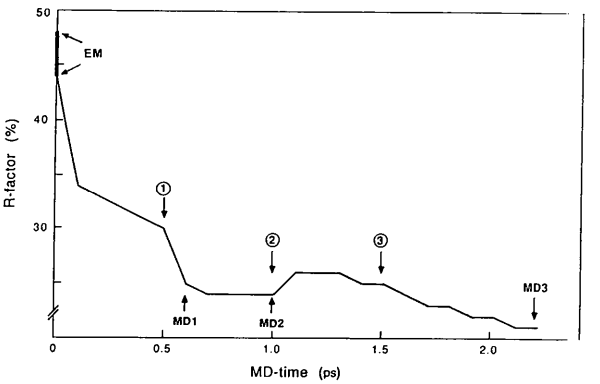

Fig. 2. Course of *R* factor during the MD refinement process. The meaning of the numbers is as follows: (1) X-ray force constant increased; (2) temperature increase from 300 to 1000 K; (3) temperature decrease from 1000 to 100 K. MD1, MD2 and MD3 denote conformations which have been analysed.

Table 1. *Refinement parameters and deviations from ideal geometry*

	Target (σ)	MIR initial	MIR final	MD1 initial	MD1 final	MD2 initial	MD2 final	MD3 initial	MD3 final	Current model
$R(=\sum\|\|F_o\|-\|F_c\|\|/\sum\|F_o\|)$	—	0·48	0·38	0·25	0·23	0·24	0·23	0·22	0·22	0·25
Number of cycles	—	—	9	—	3	—	2	—	4	34
Resolution range (Å)	—	10·0–3·0	5·0–2·5	5·0–2·5	6·0–2·5	6·0–2·5	6·0–2·5	6·0–2·5	6·0–2·5	6·0–2·5
Number of reflections ($F_o > 3\sigma F_o$)	—	8917	9180	9180	10 024	10 024	10 024	10 024	10 024	10 024
Distances (Å)										
Bond distance	0·03	0·057	0·020	0·027	0·025	0·029	0·027	0·031	0·025	0·024
Angle distance	0·04	0·091	0·047	0·089	0·057	0·091	0·057	0·073	0·057	0·048
Planar 1, 4 distance	0·045	0·177	0·127	0·141	0·073	0·141	0·076	0·132	0·049	0·045
Non-bonded distances (Å)										
Single torsion	0·50	0·307	0·271	0·195	0·237	0·205	0·240	0·180	0·215	0·244
Multiple torsion	0·50	0·611	0·421	0·304	0·343	0·306	0·310	0·304	0·328	0·374
Possible H bond	0·50	0·597	0·416	0·146	0·231	0·135	0·203	0·135	0·172	0·279
Torsion angles (°)										
Planar	4·0	22·5	14·3	22·6	5·2	24·0	4·3	26·0	9·5	2·1
Staggered	20·0	30·3	29·6	26·6	29·1	26·7	28·1	23·7	25·4	29·4
Orthonormal	20·0	53·9	50·8	42·7	37·8	46·1	40·7	40·5	36·8	37·3
Miscellaneous										
Plane groups (Å)	0·02	0·017	0·015	0·088	0·015	0·087	0·016	0·078	0·026	0·010
Chiral centres (Å^3)	0·15	0·476	0·155	0·348	0·238	0·344	0·248	0·269	0·224	0·174
Bond length distances $>4\sigma$	—	364	53	409	76	437	70	234	72	17

Table 2. *Root-mean-square deviations* (Å) *on superimposition of* α*-carbon atoms*

	MIR	MD1	MD2	MD3	Final
MIR	×				
MD1	0·98	×			
MD2	0·98	0·31	×		
MD3	0·96	0·41	0·42	×	
Final	1·02	0·58	0·55	0·44	×

Table 3. *Residues with maximum root-mean-square deviations on superimposition of* α*-carbon atoms*

Coordinate sets 1	Coordinate sets 2	Deviation range (Å)	Residues
MIR	MD3	1·5–3·7	Ala A5, Lys A6, Val A10, Ala A108, Asp A109, Trp A130, Gly A131, Pro A132, Gly B50, Asn B83, Ala B84, Met B93, Ala B94, Lys B103, Gly B124, Gly B131, Gly B144, Ala B160, Tyr B161, Ala B203
MD1	MD2	1·0–1·1	
MD2	MD3	1·0–1·3	Asn A48, Gly A50, Gly B50, Asn B75,
MD3	Final	1·0–3·0	Gly A50, Lys A51, Lys A103, Lys A106, Val A107, Ala A108, Gly A131, Gly B50, Ile B196, Ile B197,
MD1	Final	1·0–2·8	Gly A50, Lys A51, Lys A76, Gln A92, Asp A93, Lys A103, Lys A106, Val A107, Ala A108, Asp A109, Gly A131, Ala A202, Gln B49, Gly B50, Lys B51, Ile B74, Asn B75, Lys B76, Gly B99

A force constant K of 1·0 was used to weight the X-ray terms initially.

The trend of the R factor *versus* simulation time is shown in Fig. 2. The initial EM (without X-ray potential) lowered the R factor from an initial value of 48 to 44%. After 250 MD steps the R factor had dropped from 44 to 30%. As the R factor tended to level off we decided to increase the force constant by setting the value of K equal to 2·0. Subsequently the R factor dropped from 30 to 24% between 0·5 (MD1) and 0·7 ps (100 steps) but did not change significantly between 0·7 and 1·0 ps (MD2). At this stage we decided to raise the temperature of the system in order to let it search in a more extended conformational space and surmount possible energy barriers. The temperature was raised to 1000 K ($\tau_t = 0{\cdot}01$ ps) and after 0·5 ps the system was gradually cooled down to 100 K ($\tau_t = 0{\cdot}1$ ps). This process led to a final R factor of 22% (MD3).

The results of the MD refinements have been tabulated in Tables 1 to 3. The conventional crystallographic R factor for the initial model was 48%. Application of conventional least-squares refinement (*PROLSQ*; Hendrickson & Konnert, 1980) brought the R factor down to 38% with concomitant improvement in the geometry of the model. In contrast, the MD1 refinement reduced the R factor to about 23%. However, some of the geometrical parameters exhibited undesirable values due to the way stereochemical restraints are balanced in *GROMOS* compared with *PROLSQ*. In particular, deviations from planarity of peptide planes were high. However, it was a straightforward process to improve these deviations by application of a few cycles of *PROLSQ* without degrading the R-factor value.

The current model (henceforth referred to as the 'final' model) is based on the MD3 model (see Table 1). The MD3 model was rebuilt into a $2|F_o|-|F_c|$ density map and then subjected to 16 cycles of *PROLSQ* refinement followed by a final rebuild and

a further round of *PROLSQ* refinement. The details of the final refinement run are given in Table 1. The rebuilds were quick owing to the high quality of both the MD3 model and the difference maps. Significant differences between the MD3 and final models that required rebuilding are discussed below.

There was little to choose from between the MD refinement runs MD1, MD2 and MD3 on the basis of statistics (Tables 1 to 3). The final R factors and geometrical parameters after a few rounds of *PROLSQ* for all three runs were very similar. The rather large deviations in planarity parameters for MD3 were due to a glycine residue which took on arbitrary values of the planarity angle between *PROLSQ* cycles. Examination of the electron density map indicated that this residue had been displaced from density by MD3 (see below). Table 2 contains a list of root-mean-square (r.m.s.) deviations of α-carbon atoms when models derived from the MD runs were superimposed by the method of Rossmann & Argos (1975). R.m.s. deviations between the initial model (MIR) and models derived from the MD runs were of the order of 1 Å. It should be noted that the final model was based on the MD3 model and this is probably why the r.m.s. deviation between MD3 and the final model is slightly lower than that between the other MD runs and the final model. The changes brought about by MD refinement were of a global nature rather than attributable to a small number of large changes (Fig. 3).

A list of residues exhibiting the largest r.m.s. deviations is given in Table 3. A striking feature of the table is the repeated occurrence of glycines and

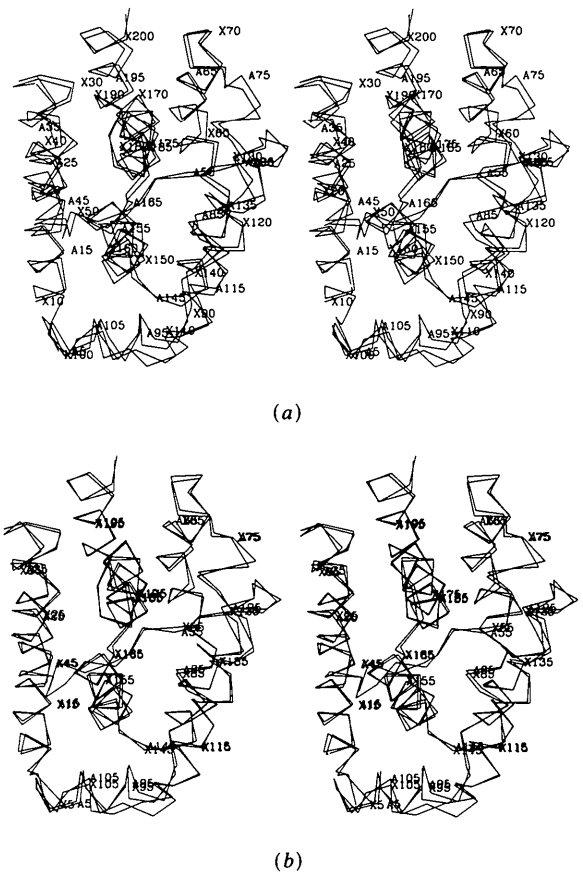

Fig. 3. Superimposition of α-carbon backbones between (*a*) the A monomer of the MIR model and the corresponding monomer of the MD1 model (labelled X); (*b*) the A monomer of the final model and the corresponding monomer of the MD1 model (labelled X).

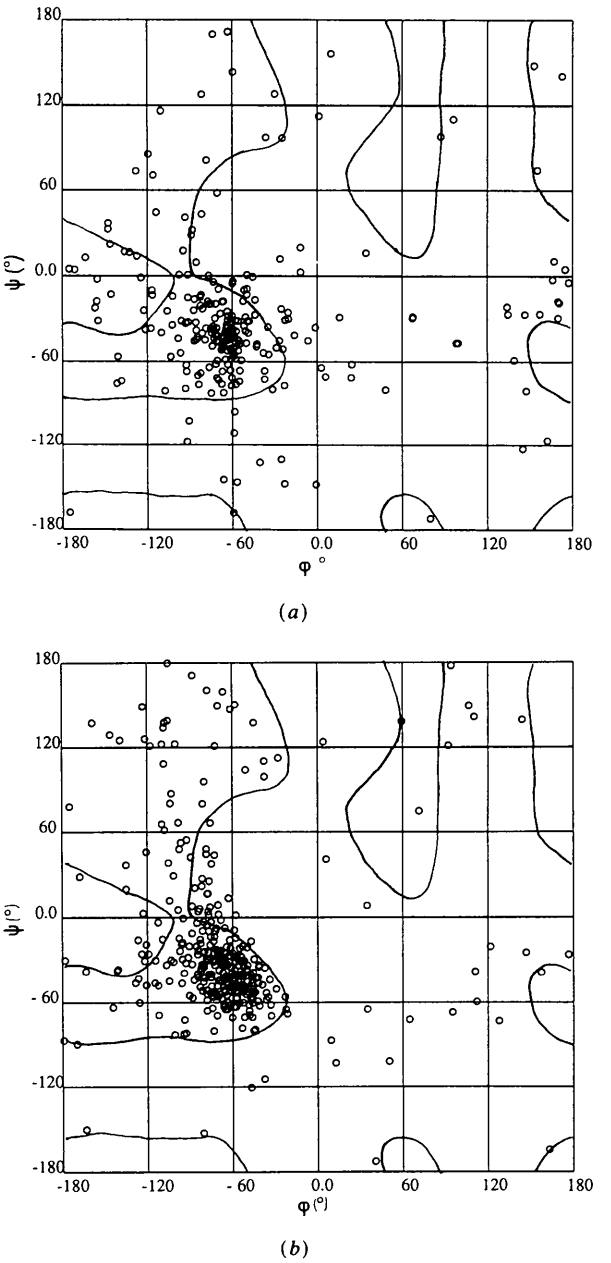

Fig. 4. Ramachandran plots superimposed on a conformational energy map. (*a*) MIR model; (*b*) MD1 model.

alanines which can be attributed to their greater flexibility. Again it is clear that there was little advantage to be gained by continuing the MD refinement past MD1. Some of the largest deviations occurred in a flexible loop region (residues 90 to 110) of the molecule. The density in this region was rather poor, with an appreciable lack of side-chain density leading to poor X-ray restraints. This region required further rebuilding into difference maps calculated after MD refinement (Fig. 3*b*).

The main-chain torsion angles φ, ψ and ω are good indicators of the quality of the model. Ramachandran plots for the initial and MD1 models are shown in Fig. 4. The majority of non-glycyl residues after MD1 have main-chain dihedral angles that lie within or very close to the acceptable regions of a φ-ψ plot (Ramakrishnan & Ramachandran, 1965). There are more than 50 outliers in the initial model compared with less than 20 after the MD1 run. All MD1 outliers occur in loops connecting secondary structure elements in the final model and occur in ambiguous portions of the electron density map.

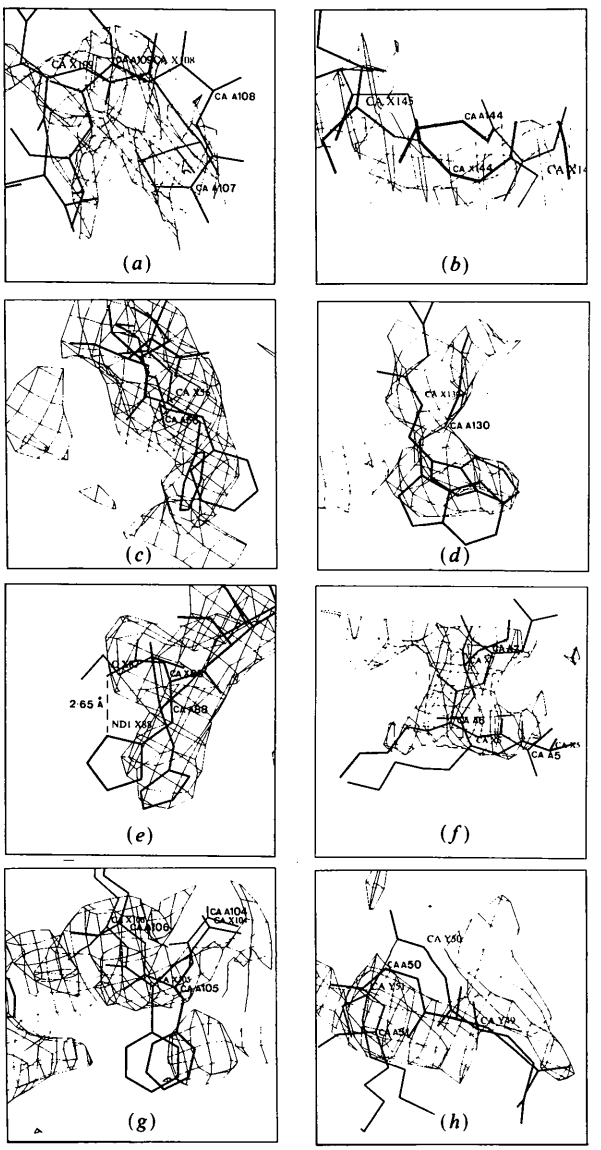

Fig. 5. Fit of residues into MIR electron density map before refinement (labelled *A*) and after MD1 (labelled *X*) except for (*h*) where fit is after MD3 (labelled *Y*). (*a*) Ala 108, (*b*) Gly 144, (*c*) Phe 56, (*d*) Trp 130, (*e*) His 88, (*f*) N-terminal, (*g*) Phe 105, (*h*) Gly 50.

Some of the changes on application of MD refinement lead to an improvement in the fit of main chain (Figs. 5*a* and *b*) and side chains (Figs. 5*c* and *d*) into the density. Some changes produced a worse fit which almost invariably occurred at the surface of the molecule indicating that the electrostatic contribution to the global potential was dominating over the X-ray contribution. Examples are the fit of His 88 (Fig. 5*e*) and the N-terminal residues (Fig. 5*f*). In some cases the MD method appears incapable of improving the fit of poorly built residue segments (Fig. 5*g*). In one case the main-chain fit worsened in a region of poor density (Fig. 5*h*). Flipping of peptide planes was observed for a number of residues.

Concluding remarks

Our experiences of MD refinement with colicin A leads to mild optimism. MD refinement produced a model with an excellent *R* factor and good geometry over the space of about 2 d c.p.u. time on a VAX 8650. In order to yield similar results, many months of conventional least-squares refinement interspersed with manual rebuilding had been envisaged. However, associated with the method are a number of problems. The problem of poor planarity was cured by application of a small number of conventional stereochemical least-squares refinement cycles. The problem of poorly fitted surface side chains could be tackled by a number of approaches. These include changing the balance of the non-bonded and X-ray contributions to the global potential, incorporation of a solvent model or a potential of mean force into the simulation and a possibly better definition of the bonded and non-bonded potential-function parameters, in particular electrostatics and polarizability. As an example, MD simulation studies of bovine pancreatic trypsin inhibitor (BPTI) including solvent compare better with structures derived from X-ray experiments than corresponding ones *in vacuo* (van Gunsteren, Berendsen, Hermans, Hol & Postma, 1983; McCammon & Harvey, 1987; Levitt & Sharon, 1988).

A number of features included in various conventional least-squares refinement programs would improve the current MD refinement package. Desirable features include incorporation of non-crystallographic symmetry, least-squares optimization of *B* factors and crystal-packing restraints. Brünger (1988)

has reported a number of improvements to the *X-PLOR* MD package including phase restraints. Also, fundamental research in the field of improving and extending the conformational search capacities of the MD algorithm could be of benefit in MD X-ray refinement techniques. The observation that MD refinement cannot always correct erroneously fitted peptide segments or produce satisfactory fits to poorly defined portions of electron density emphasizes the advantages of using structural databases (Jones & Thirup, 1986; Ponder & Richards, 1987) in the initial model building and careful inspection of difference electron density maps after refinement.

Brünger (1988) has shown that MD can generate an ensemble of structures, each of which agrees with the experimental observations. Furthermore, regions of large variations of the ensemble were shown to be due to badly fitted or disordered segments of the structure. We propose that a statistical analysis of these variations may provide an alternative to *B*-factor values as an indicator of poorly determined regions of macromolecular structures. The observation that the MD runs MD2 and MD3 led to no significant improvements of the refined model after the MD1 run suggests that the current MD methodology will not prove generally useful for further refinement of already well refined structures.

A major part of the MD X-ray refinement calculation was carried out using the *GROMOS* package (van Gunsteren & Berendsen, 1983; 1987). We thank Masao Fujinaga and Wilfred van Gunsteren for giving us a pre-release of their MD refinement program. We also wish to acknowledge the important contributions of Daniel Baty, Claude Lazdunski, Franc Pattus and Alec Tucker to the structure determination of colicin A.

References

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

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

BRÜNGER, A. T. (1988). *J. Mol. Biol.* **203**, 803-816.

BRÜNGER, A. T., CLORE, G. M., GRONENBORN, A. M. & KARPLUS, M. (1986). *Proc. Natl Acad. Sci. USA*, **83**, 3801-3805.

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

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

GROS, P., FUJINAGA, M., DIJKSTRA, B., KALK, K. H. & HOL, W. G. J. (1988). Personal communication.

GUNSTEREN, W. F. VAN (1988). *Protein Eng.* **2**, 5-13.

GUNSTEREN, W. F. VAN & BERENDSEN, H. J. C. (1983). *BIOMOS Biomolecular Software.* Laboratory of Physical Chemistry, Univ. of Groningen, The Netherlands.

GUNSTEREN, W. F. VAN & BERENDSEN, H. J. C. (1987). *Groningen Molecular Simulation (GROMOS) Library Manual.* BIOMOS b.v., Nijenborgh 16, Groningen, The Netherlands.

GUNSTEREN, W. F. VAN, BERENDSEN, H. J. C., HERMANS, J., HOL, W. G. J. & POSTMA, J. P. M. (1983). *Proc. Natl Acad. Sci. USA*, **80**, 4315-4319.

GUNSTEREN, W. F. VAN, BOELENS, R., KAPTEIN, R., SCHEEK, R. M. & ZUIDERWEG, E. R. P. (1985). *Molecular Dynamics and Protein Structure*, edited by J. HERMANS, pp. 92-99. Polycrystal Book Service, PO Box 27, Western Springs, IL 60558, USA.

GUNSTEREN, W. F. VAN, KAPTEIN, R. & ZUIDERWEG, E. R. P. (1984). Proceedings; NATO/CECAM Workshop on Nucleic Acid Conformation and Dynamics, Orsay, France. Edited by W. K. OLSON, pp. 79-92.

HENDRICKSON, W. A. & KONNERT, J. H. (1980). *Computing in Crystallography*, edited by R. DIAMOND, S. RAMASESHAN & K. VENKATESAN, pp. 13.01-13.26. Bangalore: Indian Academy of Science.

JACK, A. & LEVITT, M. (1978). *Acta Cryst.* A**34**, 931-935.

JONES, T. A. & THIRUP, S. (1986). *EMBO J.* **5**, 819-822.

LEVITT, M. & SHARON, R. (1988). *Proc. Natl Acad. Sci. USA*, **85**, 7557-7561.

MCCAMMON, J. A. & HARVEY, S. C. (1987). *Dynamics of Proteins and Nucleic Acids.* Cambridge Univ. Press.

PARKER, M. W., PATTUS, F., TUCKER, A. D. & TSERNOGLOU, D. (1989). *Nature (London)*, **337**, 93-96.

PONDER, J. W. & RICHARDS, F. M. (1987). *J. Mol. Biol.* **193**, 775-791.

RAMAKRISHNAN, C. & RAMACHANDRAN, G. W. (1965). *Biophys. J.* **5**, 903-933.

ROSSMANN, M. G. & ARGOS, P. (1975). *J. Biol. Chem.* **250**, 7525-7532.

RYCKAERT, J. P. (1985). *Mol. Phys.* **55**, 549-556.

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

TUCKER, A. D., PATTUS, F. & TSERNOGLOU, D. (1986). *J. Mol. Biol.* **190**, 133-134.