PROTEINS: Structure, Function, and Genetics 10:240–250 (1991)

Crystallographic Refinement of Ricin to 2.5 Å

Earl Rutenber, Betsy J. Katzin, Stephen Ernst, Edward J. Collins, Debra Mlsna, Michael P. Ready, and Jon D. Robertus

Clayton Foundation Biochemical Institute, Department of Chemistry and Biochemistry, University of Texas, Austin, Texas 78712

ABSTRACT **The plant cytotoxin ricin consists of two disulfide-linked chains, each of about 30,000 daltons. An initial model based on a 2.8 Å MIR electron density map has been refined against 2.5 Å data using rounds of hand rebuilding coupled with either a restrained least squares algorithm or molecular dynamics (XPLOR). The last model (9) has an R factor of 21.6% and RMS deviations from standard bond lengths and angles of 0.021 Å and 4.67°, respectively. Refinement required several peptide segments in the original model to be adjusted translationally along the electron density. A wide range of lesser changes were also made. The RMS deviation of backbone atoms between the original and model 9 was 1.89 Å. Molecular dynamics proved to be a very powerful refinement tool. However, tests showed that it could not replace human intervention in making adjustments such as local translations of the peptide chain. The R factor is not a completely satisfactory indicator of refinement progress; difference Fouriers, when observed carefully, may be a better monitor.**

Key words: ricin, refinement, molecular dynamics, molecular models

INTRODUCTION

An initial three-dimensional model of the heterodimeric plant toxin ricin, based on a 2.8 Å MIR electron density map, was reported earlier.[1] The protein is a heterodimer of Mr≈60,000. It is composed of a 267 residue A-chain (RTA), linked to a 262 residue B-chain (RTB) by a disulfide bond. The overall structure and action of the protein are described in references 2 and 3. In this study we report details of the crystallographic refinement of ricin to a resolution of 2.5 Å. The refinement progressed in two parts; the first employed conventional restrained least-squares methods, and the second employed simulated annealing with molecular dynamics. Simulated annealing has a larger radius of convergence than restrained least-squares methods and is reported to reduce the need for human intervention in the refinement process.[4] Complete descriptions of the refined structures of ricin A- and B-chains appear in this issue.[5,6] Key aspects of the determination of the initial model are reviewed below to give a complete story of its development.

High resolution data for native ricin were first collected in 1978 by oscillation photography and automated diffractometry. Data for five isomorphous heavy metal derivatives were also collected; heavy metal positions and occupancies were calculated and refined and phases determined. The resulting maps, although displaying certain proteinlike features such as alpha helices in the A-chain and some side chain density, were uninterpretable. Partial model techniques and phase recombination failed to yield a map in which the backbone of either chain could be traced unambiguously. Difficulties in bootstrap methods were ascribed to poor starting MIR phases, which resulted from the tendency of ricin crystals to change cell parameters.

Data were later collected for native ricin (2.5 Å) and two isomorphous derivatives on the Mark II multiwire area detector at the University of California San Diego. Metal positions and occupancies were refined and phases determined to 2.8 Å, including anomalous dispersion data to 4 Å. An electron density map was calculated using the MIR phases.[1]

Although the resulting map appeared clearer and more continuous than earlier ones, the backbone of ricin still could not be traced unambiguously. Phases were then improved using the solvent flattening method of Wang.[7] Although Wang suggests applying his method to SIR phases, it was found that the electron density resulting from the solvent flattened MIR map was clearer than that of either of the solvent flattened SIR maps. Indeed, the solvent-flattened MIR map was noticeably superior to all previous maps and allowed the chains to be traced readily using the computer graphics program FRODO[8] on an Evans and Sutherland Multi Picture System.

After the Cα backbone was built, amino acid residues were positioned individually. Guide angles were used to define regions of alpha helix or beta sheet and the residues were moved into density as a

Received July 20, 1990; revision accepted February 4, 1991.

Address reprint requests to Jon D. Robertus, Dept. of Chemistry, College of Natural Sciences, University of Texas, Austin, TX 78712-1167.

unit. In general, the Cα positions were preserved and side chains were oriented within their appropriate density when present. In the B-chain, nine tryptophan side groups and four disulfide bridges served as markers to ensure a proper chain tracing. The homology between the two B-chain domains was also exploited to verify the protein path. Only a few carbonyl bulges in the backbone density were clear enough to position carbonyl oxygens.

As building proceeded, zones of approximately 10 residues were subjected to the regularization routine MODL included in the FRODO package. MODL executes cycles of least-squares optimization of the peptide geometry and allows for the selection of atoms that are to remain fixed. One round of MODL with key atoms fixed followed by a second round with all but the terminal atoms free to move generally yielded a reasonable structure. The routine was used repeatedly with overlapping refinement zones until the entire molecule was optimized. The resulting model fit the MIR/solvent flattened density rather well, but some atoms were in close contact and some bond lengths and angles were distorted. The R factor was 50% for the hand-built model. We then began a systematic refinement of the model using several methods.

MATERIALS AND METHODS

Ricin Preparation

Ricin D was prepared by agarose affinity chromatography and gel filtration as described by Nicolson and Blaustein.[9] The Sephadex G-100 ricin fraction was further fractionated by ion exchange chromatography; 65 mg ricin D was dialyzed into 20 mM sodium phosphate pH 6.0—4 mM lactose and applied to a 30 ml column of Whatman P11 phosphocellulose equilibrated in the same buffer. The column was rinsed with approximately 100 ml buffer until two flow through peaks were eluted. The column was then developed with a 500 ml gradient of 0–500 mM NaCl in the starting buffer. A large peak eluting at < 100 mM NaCl was readily crystallized.

Recombinant ricin A-chain was isolated from *E. coli* JM105 carrying the ricin A-chain expression plasmid pUTA. This construction and the purification of recombinant ricin A are described elsewhere.[10]

Lectin Binding Blots

Native ricin D and resuspended crystals of phosphocellulose-purified ricin were electrophoresed by 10.5% SDS-PAGE and blotted onto a nitrocellulose membrane (Micron Separations). Unglycosylated recombinant ricin A-chain was also run as a control. The protein binding capacity of the blot was blocked by immersion in Tris-buffered saline (TBS, 10 mM Tris-Cl, pH 7.2—150 mM NaCl) containing 5% bovine serum albumin (BSA) for 30 minutes. The blot was then immersed for 30 minutes in TBS supplemented with 1 mM $MgCl_2$, and 10 μg/ml concanavalin A. The blot was washed twice in supplemented TBS, then incubated for 30 minutes in supplemented TBS containing 0.05% TWeen 80, 0.5% BSA, and rabbit antiConA (United States Biochemical). The blot was again washed twice with supplemented TBS—0.05% Tween 80—0.5% BSA, then incubated with horseradish peroxidase—conjugated goat anti-rabbit IgG in the same buffer. After again washing, the blot was developed by adding 5 ml of 3 mg/ml 4-chloro-1-naphthol in methanol and 25 ml of TBS containing 15 μl H_2O_2.

Automated Refinement

Crystallographic refinement of ricin made use of two program packages. From 1987 through 1988, a package of programs based on the Konnert and Hendrickson restrained least-squares refinement protocol was used.[1] The package centered on the program PROFFT, a fast Fourier transform version of the Konnert and Hendrickson program PROLSQ. From January 1989 to the present, the XPLOR package, created by Axel Brunger, was used.[12] We define a round of refinement as a rebuilding, by hand, of the model, followed by a variable number of cycles of automated refinement using one or the other program package. A total of four rounds of refinement using PROFFT (on a VAX 11/750) and five rounds of refinement using XPLOR (on a CRAY X-MP/24) were executed. Nine models designated model 1 through model 9 were involved, one during each round of refinement. In general, the model number was incremented at the beginning of hand rebuilding.

Restrained Least-Squares Package

Restrained least-squares minimization is perhaps the best-known and most widely employed automated crystallographic refinement technique for proteins.[13] The method minimizes an error function of the following form:

$$\Delta = \sum_h (|Fo| - |Fc|)^2 + \sum_j \sum_i d_{ij}^2$$

where h is summed over all crystallographic reflections, j is the type of restraint (typically bond lengths, bond angles, van der Waals distances and so on), and i is the number of penalty functions on the generalized distances, d, of type j.

The Konnert and Hendrickson least-squares refinement package used in this laboratory contained the program PROFFT, a version of the original program PROLSQ incorporating Agarwal's version of the fast Fourier transform.[14] Also included was the program PROTIN, which analyzes input model coordinates and lists unfavorable interactions. Such a list was used before and after running PROFFT to identify those regions of the model that were poorly built. PROFFT was initially configured as suggested

by the authors. In the last two rounds of refinement using PROFFT, however, the overall distance restraint weighting factor WDSKAL was varied to allow greater deviations from ideality.

Molecular Dynamics Package

XPLOR is a set of modular programs combining energy minimization with high temperature restrained molecular dynamics. XPLOR calculates an empirical energy term composed of a summation of the energies representing bonds, angles, dihedral angles, nonbonded interactions (van der Waals and electrical contacts), and improper angles (in this context an improper angle is used to define a set of planar atoms or a chiral center). A pseudo energy term representing the discrepancy between the observed and calculated structure factors is added as an additional restraint. Refinement attempts to minimize this total energy by simultaneously solving the classical equations of motion for each atom in the structure for a given time period and temperature. At high temperatures, atomic motions forbidden by least-squares energy minimizations are possible, and the method can overcome local energy barriers and move toward the global minimum. The radius of convergence is reported to be in excess of 5 Å for XPLOR as compared to 1–2 Å for other standard methods.[4] In addition, XPLOR is relatively insensitive to the choice of initial resolution range and the need to introduce shells of successively higher resolution terms is alleviated.[15]

XPLOR accepts input as individual Protein Data Bank (PDB) coordinate files each representing one macromolecular segment. Six of these files, representing two peptide chains, two covalently bound sugars, and two noncovalently bound sugars were assembled into one master PDB file by the XPLOR routine GENERATE. Coordinates for those hydrogen atoms that could be involved in hydrogen bonding were calculated by GENERATE and included in the master PDB file.

The XPLOR routine CHECK was used to assign weights to the x-ray structure factors that serve as restraints during refinement. The master PDB file was input to PREPSTAGE, which conducts multiple cycles of conjugate gradient energy minimization. This serves to relax the initial structure, that is, to separate atoms that may be clashing and to optimize bond lengths and angles, before proceeding with molecular dynamics. The first portion of energy minimization applied purely repulsive harmonic energy terms to the nonbonded atomic interactions. This allowed close contacts to be relieved without creating an excessively high energy term. The second portion of energy minimization invoked the standard Lennard-Jones potential energy function to describe the atomic interactions. As suggested by the authors, charges on the side chains of polar residues were set to zero. All temperature factors and occupancies were set to 20 and 1, respectively.

The relaxed, energy-minimized structure was then subjected to XPLOR's high temperature molecular dynamics routine. Of several options available, we chose to run SLOWCOOL, which raises the temperature of the structure to some target value by assigning a random velocity vector to each atom. The initial velocity vector is composed of a random direction and a magnitude derived from a Maxwellian distribution at the target temperature. The velocity for each atom is scaled according to the local energy gradient and applied for some small period of time. We configured SLOWCOOL to begin at a temperature of 4,000 K and execute 50 cycles of molecular dynamics for each of 148 25 K steps down to a final temperature of 300 K. Each of the 7,400 cycles lasts 0.5 fs for a total of 3.7 ps. The high initial temperature permits large random changes in the atomic positions and allows conformations to be explored that are unattainable without added kinetic energy. Local energy minima can be overcome and, as the temperature gradually decreases, the structure may settle into a new energy minimum. The number of cycles of molecular dynamics we performed at high temperatures is believed to be sufficient for equilibrium to be attained.[4]

Due to the random nature of the atomic motions during molecular dynamics, the final room temperature structure may not be at the true energy minimum and a final round of energy minimization, similar to PREPSTAGE, is necessary. FINALSTAGE executes a set of energy minimization cycles using standard Lennard-Jones type energy terms as in the second portion of PREPSTAGE. Throughout the above procedures, the overall B-factor was set to 20.0 Å.[2] As a final step in this phase of the refinement, the XPLOR routine BREFI was used to refine the individual isotropic B-factors.

The XPLOR routine ANALYZE was used to analyze the stereochemistry of the final structure and to compare that structure to others.

Maps

At the end of each round of refinement, that is, after a particular model had been hand built and refined using either PROFFT or XPLOR, a binary Fourier map (BFM) was computed using the measured structure factors and phases calculated from the newly refined model. Such a BFM, potentially biased by the calculated phases, was then used to construct a "no electron density" (NED) map. A NED map is similar to an omit map except that a slab portion of the starting density is omitted and the remaining map is back-transformed to obtain new phases. These phases are then combined with the measured structure factors to produce an unbiased slab of density, corresponding to that which was omitted. The process is repeated until the entire

NED map is constructed. An in-house program, NEDFFT, was used to produce a NED BFM with 2Fo-Fc character. Typically, maps were calculated from 10 Å to either 2.8 Å or 2.5 Å according to the resolution range used during refinement. A difference map, using structure factors of the form Fo-Fc and phases calculated from the current model, was also produced. The difference maps were typically contoured at 2–3 times the estimated standard deviation (sigma) of the electron density distribution from zero.

Rebuilding a Model

Following automated refinement, the model was inspected for errors and inconsistencies. The output from either PROTIN or ANALYZE was used to identify specific areas of the structure that had improbable geometries or close contacts between atoms. The program FRODO was used to display simultaneously the new model, the difference density, and the NED density on an Evans and Sutherland PS 390 color graphics system. The entire model was examined in short sections of the peptide chain and compared to the current density. When appropriate, atoms (frequently entire side chains) were shifted to better fit the electron density. As during the initial building of the model, the FRODO routine MODL was used to regularize a peptide zone after rebuilding.

Crystallographic Data

Ricin crystallizes in space group $P2_12_12_1$ and diffracts to approximately 2.5 Å. Crystallographic data were collected as previously described.[1] The rounds of least-squares refinement included all available data in the resolution range of five to 2.8 Å (13,553 reflections); the rounds of molecular dynamics refinement included all available data in the resolution range of five to 2.5 Å (18,572 reflections). Unless otherwise noted, the crystallographic R factor for a model was computed for all reflections in the resolution range used in the refinement of that model.

RESULTS

Least-Squares Refinement

Figure 1a summarizes the entire refinement as a plot of R factor versus refinement round number. The program PROFFT was used in the first four rounds of crystallographic refinement and produced models 1–4. Figure 1b is a plot of R factor versus cycle number for round 2 of this least-squares refinement.

Table I summarizes the character of each model and compares it to the final model (9). The original model (1) had a total RMS deviation from ideal bond lengths and angles of 0.080 Å and 7.48°, respectively. The RMS difference between the backbone atoms of model 1 and model 9 was 1.89 Å. After four

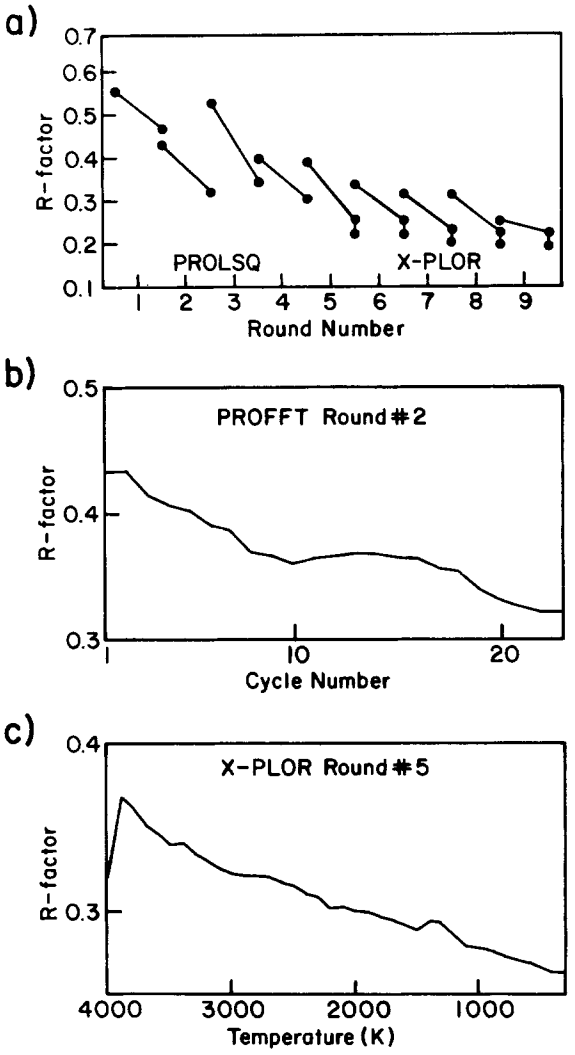

Fig. 1. Summary of refinement progress. **(a)** The R factor is shown for each round of refinement. The first point plotted for a round indicates the R factor after hand-rebuilding and prior to computer-assisted refinement. The second point represents the R factor after computer-assisted refinement. The third point, plotted only for rounds 5–9, represents the R factor after refinement of the isotropic B-factors. **(b)** The R factor is plotted as a function of cycle number for a typical round (2) of restrained least-squares refinement using the program PROFFT. **(c)** The R factor is plotted as a function of temperature for a typical round (5) of molecular dynamics refinement.

rounds of least-squares refinement and hand rebuilding, model 4 had significantly improved stereochemistry with a total RMS deviation from ideal bond lengths and angles of 0.037 Å and 6.15°, respectively, and a total RMS difference from model 9 of 1.52 Å.

Analysis of the gross structural differences between the various models is facilitated by Figure 2. Here the RMS difference between the backbone atoms of the final model (9) and several earlier models is shown as a function of residue number. The most dramatic differences between the final model and

TABLE I. Summary of Model Statistics

Model	R factor*	Δbond (Å)†	Δangle (degrees)†	RMS difference from model 9 (Å)§ Main Chain	RMS difference from model 9 (Å)§ Side Chain	<ΔΦ>** from model 9 (degrees)
1	0.467	0.080	7.48	1.89	3.49	57.5
2	0.322	0.023	5.211	1.76	3.33	44.5
3	0.345	0.060	7.18	1.56	2.79	56.7
4	0.304	0.037	6.15	1.52	2.76	53.5
5	0.227	0.021	4.80	1.34	2.57	38.8
6	0.223	0.022	5.08	1.20	2.40	40.5
7	0.209	0.021	4.60	0.92	1.94	32.4
8	0.202	0.020	4.57	0.67	1.39	20.9
9	0.198	0.021	4.61			
MD1	0.235	0.022	4.93	1.72	3.41	43.5

*The crystallographic R factor is computed for all data in the resolution range of 5-2.8 Å to facilitate comparisons between the various models.
†The RMS deviations from ideality of bond lengths and bond angles.
§The atomic RMS differences from model 9.
**The mean difference in phase angles from model 9 for all reflections between 10 and 2.8 Å.

the earliest model are in the A-chain. The peak at the amino terminus of the A-chain and those centered on residues 130, 190, and 230 represent fairly major adjustment during the refinement process. In particular, translations of several amino acid residues along the peptide path were made during the course of refinement. The segments in question were initially misbuilt due to weaknesses in the MIR map and the relative lack of landmark residues (e.g., tryptophans and bridging cysteines) in the A-chain sequence. These initial model errors, discussed in more detail below, were gradually resolved over the course of refinement by manually rebuilding portions of the peptide, usually in conjunction with difference density analysis.

Overall, the progress of the refinement during the first four rounds was steady. A minor but very favorable rebuilding after round 1 improved the R factor and allowed further refinement. A much improved map calculated after round 2 lead to a major rebuilding in which many side chain positions were altered and approximately 20 A-chain residues, centered about residue 190, were translated along the peptide path in the direction of the N-terminus. A dramatic increase in the R factor resulted and subsequent automated refinement during round 3 failed to reach the previous R factor minimum. Rebuilding with the map produced after round 3 was particularly productive and the refinement during round 4 converged rapidly. This result, along with the poor correlation between model 3 and its map, lead us to believe that the structure was trapped in a local energy minimum. During the fourth round of refinement, approximately midway, we loosened the stereochemical restraints applied to the model by PROFFT in hopes that the R factor minimization would drive the structure to a global minimum. The overall weighting of the van der Waals contacts was set to zero and the weight on bonds, angles, etc. (WDSKAL) was decreased from 1.0 to 0.7. Although this strategy had some effect in lowering the R factor, the benefit was not large and refinement became unstable with large fluctuations in the R factor and stereochemical parameters. Both weights were reset to their original value for the last few cycles of the round.

In our experience, the results of PROFFT were uneven. At the end of a round of refinement, the model was compared to its NED map constructed from the observed structure factors and calculated phases. In general, portions of the model fit the density quite well; however, many atoms, particularly those of side chains, were not positioned in the center of their density and occasionally an entire side chain was positioned outside of its density. In most cases, these misplaced atoms could be easily moved by hand into density while maintaining reasonable bond distances and geometries.

Molecular Dynamics Refinement

The XPLOR package of programs was used in the last four rounds of refinement and produced models 5–9. The results of the XPLOR refinement are shown graphically in Figure 1a as round numbers 5–9. A plot of R factor versus temperature for round five is shown in Figure 1c. The crystallographic data were extended from 2.8 Å to 2.5 Å for these last five rounds of refinement and, at the end of each round, the isotropic temperature factors for the individual, nonhydrogen atoms were refined.

XPLOR restrains the stereochemistry of the structure by computing energy terms for bond lengths, angles, and atomic interactions. The total energy of the structure prior to running XPLOR in round 5 was 228,000 kcal/mol and the R factor was 35.0%. The energy of the structure upon completion of FI-

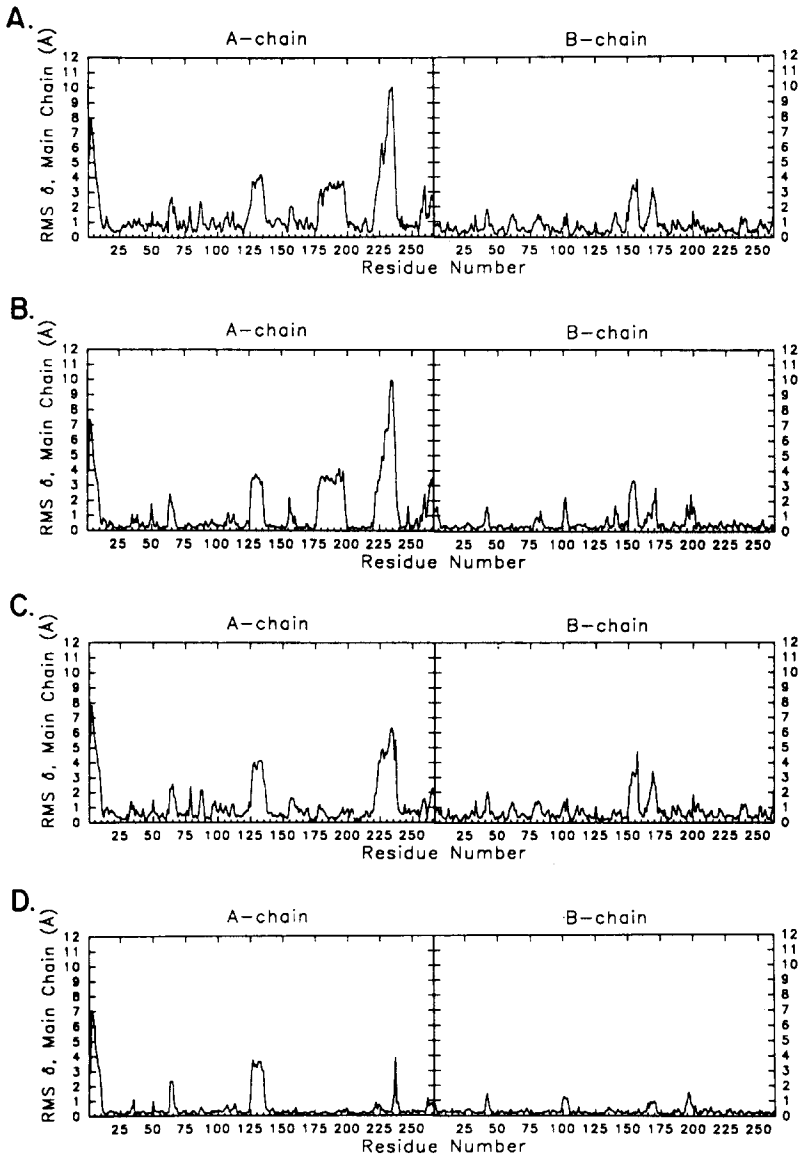

Fig. 2. Comparisons of various models with the final refined model. Each panel shows the total RMS deviations for the backbone atoms plotted as a function of residue number. The deviations were calculated between model 9 and **(A)** the initial hand-build model (1) after one round of least squares, **(B)** the initial hand-built model 1 after one round of molecular dynamics refinement (MD1), **(C)** model 4 after four rounds of least-squares refinement, and **(D)** model 7.

NALSTAGE in round 9 was 4,300 kcal/mol and at the end of BREFI the R factor, computed for all data in the resolution range of 5-2.5 Å, was 21.6%. Although the value of total energy is arbitrary, we were encouraged to notice that the energy for the final structure was less than 2% of that for the initial structure reflecting, in part, a great improvement in its stereochemistry. The total RMS deviation from ideality of bond lengths and angles was 0.021 Å and 4.67°, respectively, for our final model (9).

As seen in Figure 2, most of the major differences between model 1 and model 9 had been reconciled by the end of round 7. Model 7 differs structurally from model 9 in only a few areas. In the A-chain, the amino terminal nine residues and 10 residues centered on position 130 were shifted during rebuilding of model 8.

Overall refinement during rounds 5–9 proceeded as for rounds 1–4. That is, the program XPLOR was essentially substituted for PROFFT and the routine of analyzing maps, rebuilding the model by hand and applying computer-assisted refinement was retained. The most notable feature of the last five rounds of refinement is that the R factor dropped dramatically during round 5, as seen in Figure 1,

but the overall improvement in the R factor for subsequent rounds of refinement was marginal despite the gradual resolution of several significant discrepancies in the model. This phenomenon is almost certainly related to our primary observation concerning the XPLOR refined model. That is, after only one round of molecular dynamics, a comparison of the model and the resulting map revealed that nearly all of the atoms in the structure appeared to be positioned in the center of reasonable density. Occasionally, misbuilt residues were distorted to accommodate the local electron density although, in general, the stereochemistry of the XPLOR refined models was excellent. In our hand rebuilding, such distortions were regarded with suspicion and an analysis of the difference density usually lead to a more reasonable structure.

In addition to placing atoms in density, molecular dynamics also excelled at relieving stereochemical stress in the model. Upon completion of the first round of XPLOR refinement, we investigated a region of the structure that was known to be problematic. Throughout the rounds of PROFFT and hand rebuilding, a clash between three carbonyl oxygens in the B chain had persisted. In the XPLOR refined model, these carbonyl oxygens (O_{20}, O_{37}, and O_{38}) had been reoriented to produce a much more energetically favorable structure. The most dramatic change in the region involved a peptide bond that had been flipped by 180°. As discussed below, however, hand rebuilding was required to adjust other, more serious discrepancies in our model.

Molecular Dynamics Refinement of Model 1

XPLOR became available only after we were well advanced in the ricin refinement and was applied after round 4. Since molecular dynamics appeared to be very powerful, it was decided to test the method by using it to adjust our earliest hand-built, once-refined ricin model (1), preserving as nearly as possible the conditions used to refine the later models. Since molecular dynamics is not as sensitive to resolution as other methods, the refinement included data between 5 and 2.5 Å. Model 1 was energetically quite poor and additional cycles of energy minimization, both repulsive and standard Lennard-Jones, were required during PREPSTAGE. The initial energy of the structure was 316,000 kcal/mol and the R factor was 43.7%. At the end of FINALSTAGE the energy of the structure was 7,240 kcal/mol and at the end of BREFI the R factor was 23.5%. This test model was called MD1. The total RMS deviation from ideality of bond lengths and angles was 0.022 Å and 4.93° respectively. Note that R factor, total energy, and deviations from standard bonding parameters for model MD1 are only slightly higher than those for the final model (9).

Figure 2 shows the major structural differences between model MD1 and model 9. In general, the backbone atoms of model MD1 moved toward the positions eventually assigned in model 9. The major exceptions are those areas of the structure where a section of the peptide required translation by hand during model rebuilding. One such region, misbuilt in the initial model, is described below. It is an illustration of the course of our refinement and an indication of the power and limitations of the molecular dynamics method.

Treatment of a Major Model Error

The first major error in the initial model was noticed during the rebuilding of model 2 and involved the placement of a section of peptide around residue 180 in the A-chain. Figure 3 shows a portion of the A-chain near Arg 180 as represented by (a) model 1, (b) model MD1, (c) model 4, and (d) model 9. Also shown is the NED electron density as computed from the observed structure factors and phases calculated from the cognate model. The difference electron density is included as an indication of its usefulness during the course of the refinement.

In model 1 it is clear that the side group of Arg 180 has been placed in density, which is inadequate for its bulk. Also, the charged side chain of the residue is imbedded in a hydrophobic region contacting Leu 214 and Leu 248. The ring of Phe 181 is also in poor density, which extends toward solvent, an unlikely location for a nonpolar side group. The electron density around the side group of Ala 179 is weak and hard to interpret. The difference density, however, clearly indicates that a larger side chain is required at this position. A difference peak off the site of Arg180 also indicates that this group is misbuilt.

Figure 3b shows this same area for MD1, that is, after one round of molecular dynamics refinement with no hand rebuilding. The misplaced residues discussed above were not moved into their correct positions, although atoms are placed in density. The side chain of Arg 180, for example, fills the local electron density more completely than in model 1, but it is distorted from its energetically favorable, extended state. In fact, the Arg side chain has been distorted into a conformation that resembles the phenylalanine ring, which actually belongs in that position. The difference peak seen near the Arg 180 side group in model 1 is not present in the map for MD1. It appears that the distorted arrangement of the arginine side chain atoms fit the phenylalanine

Fig. 3. The progress of refinement in a problem area. A view of the region of the A-chain around Arg 180 in **(a)** the initial hand-built model (1) after one round of least squares, **(b)** the initial hand-built model 1 after one round of molecular dynamics refinement (MD1), **(c)** model 4 after four rounds of least-squares refinement, and **(d)** the final refined model (9). In each panel the NED electron density, calculated for the appropriate model, is shown as a basket of solid lines, and the positive difference density is shown as dashed contours. The NED density was contoured to display 12% of the map and the difference density was contoured at 2.5σ.

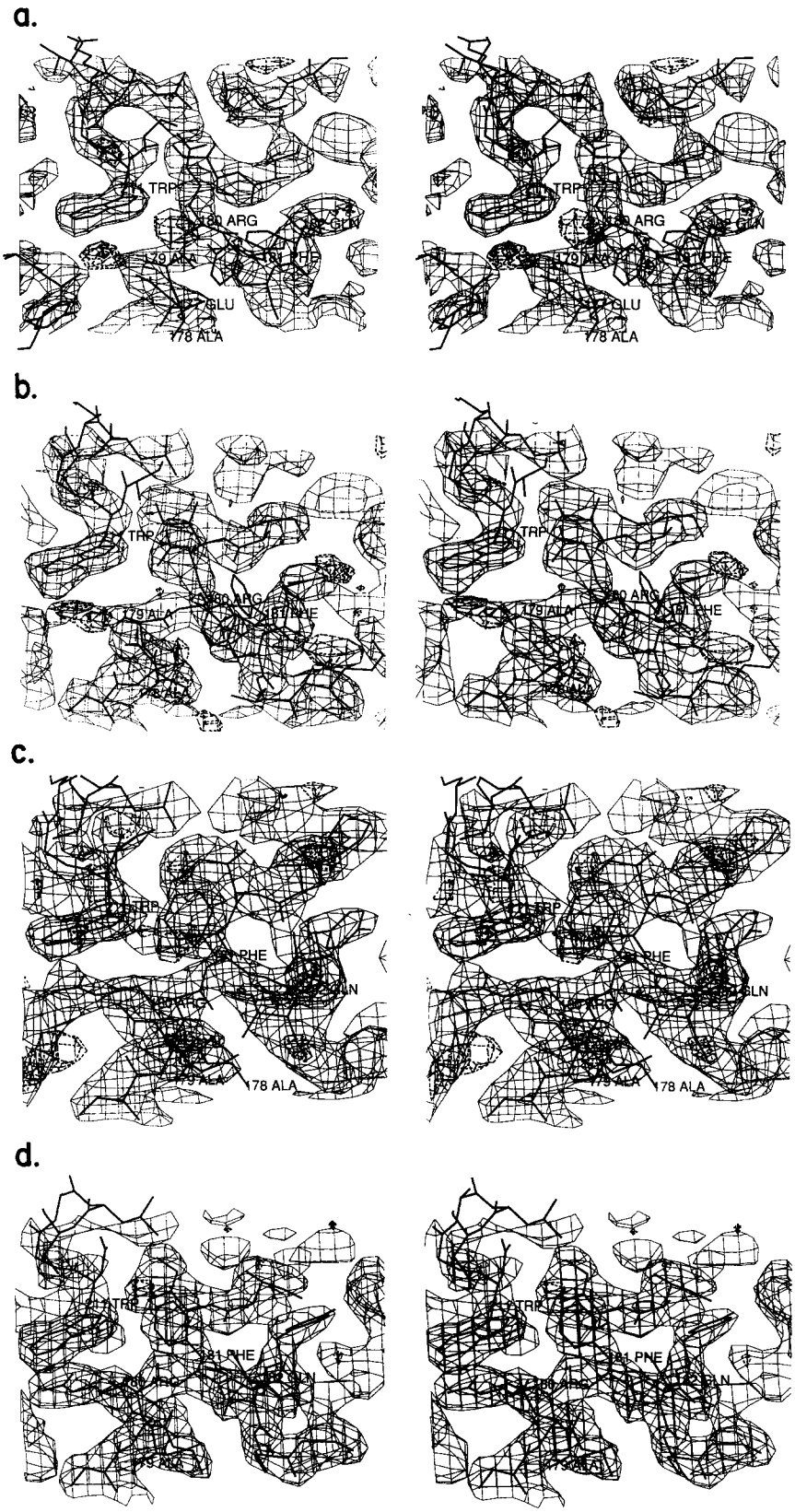

Fig. 3. Legend appears on page 246.

density well enough to evade detection by conventional difference Fourier methods at this resolution. In contrast, the difference peak near the beta carbon of Ala179 is still present, indicating that a larger group is required.

Figures 3c,d shows the same area from model 4 and model 9, respectively. Model 4 reflects the major changes made during the rebuilding of model 2. Ten residues were translated, to varying degrees, along the peptide path in the direction of the N-terminus. Arg180 was moved into the density initially occupied by Ala179, Phe181 was moved into the density initially occupied by Arg180 and so on. The side groups clearly fit the density in a more satisfying manner and the stereochemical sense of placing the side chain of Arg180 into solvent and Phe181 into nonpolar contacts, is evident. The NED map for model 4 is relatively clean, although the difference density does suggest that some improvements are possible. The stereochemistry of model 9 is clearly superior to that of the other models and the fit between the structure and the NED map density is very good. The higher resolution data affords a more detailed electron density distribution in which carbonyl bulges and other features are evident. The difference map calculated for model 9 is also relatively flat with virtually no peaks above 2.5σ, and indicating no need for structural changes.

Maps

Because the progress of our refinement ultimately depended upon the interpretation of electron density and hand rebuilding, the relative value of the various maps constructed during the course of the refinement warrants special emphasis. In general, the difference map calculated for the first model was rather noisy and difficult to interpret. However, there were significant differences between model 1 and the NED electron density, and this information allowed rebuilding to progress. Successive rounds of refinement gradually improved the phasing capacity of the model and the utility of the difference maps improved. In general, difference maps following a round of molecular dynamics refinement appeared flat, save in those local areas that contained structural errors. In contrast, because XPLOR effectively drives atoms into density, some misbuilt features were not revealed by the difference density. These errors could be found through attention to peculiar bond geometries and unorthodox arrangements of side groups.

Carbohydrate and Solvent Molecules

Ricin is a glycoprotein and has potential sites for the attachment of high mannose carbohydrate chains at asparagines 10 and 236 of the A-chain and asparagines 95 and 135 of the B-chain. However, some heterogeneity in the carbohydrate composition is observed at all four sites and the extent of glycosylation on the A-chain is quite variable.[16] In addition, the B-chain is a lectin and the crystal structure contains two lactose sugars held in reversible binding sites as described in the accompanying study.[4]

After the initial chain tracing, the positions of the high mannose sugars attached to the B-chain were quite evident from the strong electron density extending from the protein at the appropriate locations. The chemical structures of these biantennary complexes are known,[17] but the electron density was insufficient to accommodate all the sugars. In each case, a core containing the first five sugars, including the first mannose of each antenna, was attached to the asparagine side chain. The density for residues suspected to reside further out on each antenna vanished, probably due to high thermal motion of the termini. The complete density for each of the noncovalently bound lactose molecules was clear and these sugars were positioned in the model and included in the refinement process.

We were unable to locate any appropriate electron density in the vicinity of the glycosylation sites of the A-chain. As discussed above and demonstrated by Figure 2, refinement of the A-chain was troublesome and our inability to locate these sugars eroded our confidence in the correctness of the model. We used a lectin antibody assay, described in Materials and Methods, to examine the glycosylation levels of the ricin used for crystallization. This chemical test showed that relative to the B-chain, the A-chain of purified ricin was roughly 50% glycosylated. Furthermore, the level of glycosylation observed for the A-chain from dissolved ricin crystals was no more than 10% that of the B-chain. It appears that crystal growth selects for protein with unmodified A-chain.

Although well separated in the tertiary structure of the ricin molecule, it is interesting to note that the A-chain glycosylation sites are closely positioned in the three-dimensional crystal lattice. That is, Asn 236 is adjacent to the amino terminal residues, including Asn 10, of a symmetry related A-chain molecule. Perhaps due to the occasional presence of carbohydrate at one or both of the A-chain glycosylation sites, the electron density in this local region of the map was persistently weak and contributed to the initial misplacement of several residues. The A-chain residues around position 130 are a surface loop connecting two helices. This loop is also located in the general region of the glycosylation sites and, as seen from Figure 2, was also misbuilt in the first model and remained troublesome until round 8.

After the eighth round of refinement, the resulting NED map and the difference map were used in conjunction with the model to locate and position solvent molecules. Some tightly bound solvent molecules could be clearly seen in both maps; more often, the selection of a potential solvent site was

Fig. 4. Superposition of A-chain models 1 and 9. In this stereo pair the backbone of model 1 is shown as a light line. Model 9, the end point of this refinement, is shown in heavy lines. Four alpha carbon positions from model 9 are labeled. Each is located in a section of the protein that required rebuilding during refinement. These positions can be correlated with the differences illustrated in Figure 2.

based on the difference density. A potential water molecule would only be accepted if it had appropriate hydrogen bonding partners on the protein. In general, a solvent site was identified by a roughly spherical portion of density whose center was positioned approximately 2.8 Å from a nitrogen or oxygen atom. A total of 120 solvent molecules were added to the structure. The hand rebuilding during round 9 was minor and the solvent containing model was refined in PREPSTAGE without the use of molecular dynamics.

DISCUSSION

Ricin is a scientifically and commercially important protein. There is considerable interest in engineering the toxin both to understand its mechanism of action and to produce mutant proteins that may have improved properties for use in therapeutic agents. An accurate model of the protein should be a valuable guide to that engineering. In this light it is important to consider the results and overall value of crystallographic refinement.

Figure 4 shows a superposition of the alpha carbon backbones of the A chains from models 1 and 9. As indicated in Figure 2, the A-chain was altered more profoundly during refinement than B-chain and the superposition of the B-chains shows very few differences. It is clear from Figure 4 that the overall folding of the protein has not been changed by refinement. However, the details of the two structures show some important differences. A comparison with Figure 2 shows the nature of some major rebuilding. The loop containing reside 235 has been extended during refinement. It was pulled into electron density that had initially been thought to belong to the high mannose olgosaccharide expected to be bound to Asn236. The rebuilding was driven by the unfavorable stereochemistry resulting from the compression of the chain, however. It was made before we proved by chemical analysis that the A-chains of the crystal were unglycosylated. The interaction of A- and B-chains are described in an accompanying study.[6] Suffice it to say here that the interactions center on hydrophobic bonding between this loop and the C terminal region of the B-chain. The refinement radically altered our view of the residues involved in the intermolecular contact. For example, in the initial model Phe226, near the base of the loop, was thought to project outward and contact Leu132 of the B-chain. As a result of rebuilding during the refinement, Phe226 was translated three residues. It makes no contact with B-chain but is part of an internal hydrophobic cluster, surrounded by residues Ile218, Ile230, Val242, and Leu248.

Another major change brought about by refinement occurs in the active site. Notice that the position labeled 180 in model 9 superimposes with a Cα in model 1. As described above (Fig. 3), this position was initially occupied by Ala179. Since Arg180 is now known to be among the most important residues in the active site, it is ironic that this crucial change is not very apparent in the superposition plot. The effects of the translation of 10 residues are subtly apparent in the image as one inspects the chain in either direction from 180. It seems clear that a refined structure is required to assess the role of key residues or to rationalize the results of chemical or genetic changes to the wild-type protein.

As the science of protein crystallography has progressed from its infancy, a fueling force in its maturation has been the steady development of increasingly sophisticated and powerful computer software designed to facilitate evermore sophisticated computational tasks. Since power and complexity are often

related and their combination may lead to disaster, it is important that the crystallographer examine carefully the capabilities, limitations, and hazards of a new technique. Such insight is, of course, gained by trial and error and by borrowing from the experiences of others. We have employed just such a landmark method, molecular dynamics, in the crystallographic refinement of ricin and have reported some of our observations regarding its use and compared it to conventional restrained least squares refinement techniques.

The least-squares crystallographic refinement of ricin was interrupted by the introduction of the molecular dynamics method to our laboratory. We terminated conventional refinement with PROFFT and began refinement with XPLOR using molecular dynamics. Overall, the results of XPLOR were remarkable. The method resulted in some net atomic movements of 5 Å or more and returned models with excellent stereochemistry. Peptide bonds were flipped and entire side chains were repositioned to alleviate unfavorable interactions. Furthermore, molecular dynamics seemed to be less sensitive to local energy minima in which a structure may be trapped during the course of conventional refinement. Prior to running XPLOR, our structure was thought to be in such a minimum. Application of molecular dynamics to our structure reduced the need for manual rebuilding of the model and allowed refinement to continue.

The crystallographic R factor is the primary indicator of the overall "goodness" of a protein structure and the measure by which we have traditionally gauged the progress of a structure's refinement. In our experience, the method of molecular dynamics is capable of diminishing the R factor of an atomic model to a very desirable value while failing to correct major discrepancies known to exist in the structure. These errors required group translations of many angstroms to correct, distances well beyond the theoretical limits of the method. Less dramatic structural alterations in the model are frequently corrected by XPLOR and, in general, the stereochemistry of a model after refinement is extremely good. A comparison of such a model with its cognate map shows that atoms are positioned in density. The number of conformations attainable with molecular dynamics appears to be almost limitless and the method converges on those local conformations that place diffracting matter where it belongs. This results is a low R factor and a general flattening of the difference density. However, we found that the difference maps are quite sensitive even to minor errors and were very useful in identifying and correcting misbuilt regions. Indeed, drawing little security from the R factor, we executed rounds of refinement until the difference density was reduced to small, spurious peaks.

ACKNOWLEDGMENTS

We are grateful to Raquelle Smalley for her help in preparing the figures. This work was supported by grants GM 30048 and GM 35989 from the National Institutes of Health and by a grant from the Foundation for Research.

REFERENCES

1. Montfort, W., Villafranca, J.E., Monzingo, A.F., Ernst, S., Katzin, B., Rutenber, E., Xuong, N.H., Hamlin, R., Robertus, J.D. The three-dimensional structure of ricin at 2.8 Å. J. Biol. Chem. 262:5398–5403, 1987.
2. Olnes, S., Pihl A. The molecular action of toxins and viruses. In: "The Molecular Action of Toxins and Viruses," Cohen, P. and Van Heynigen, S. (eds.). New York: Elsevier, 1982:52–105.
3. Robertus, J.D. Toxin Structure. In: "Immunotoxins," Frankel, A.E. (ed.). Boston: Kluwer Academic, 1988.
4. Brunger, A.T. Crystallographic refinement by simulated annealing: Application to a 2.8 Å resolution structure of aspartate aminotransferase. J. Mol. Biol. 203:803–816, 1988.
5. Katzin, B.J., Collins, E.J., Robertus, J.D. The structure of ricin A-chain at 2.5 Å. 10:251–259, 1991.
6. Rutenber, E., Robertus, J. The structure of ricin B-chain at 2.5 Å resolution. Proteins 10:260–269, 1991.
7. Wang, B.C. Resolution of phase ambiguity in macromolecular crystallography. Methods in Enzymology 115:90–112, 1985.
8. Jones, T.A. In: "Computational Crystallography," Sayre, D. (ed.). Oxford: Oxford University Press, 1982:303.
9. Nicolson, G.L., Blaustein, J. The interaction of *ricinus communis* agglutinin with normal and tumor cell surfaces. Biocim. Biopys. Acta 266:543–547, 1972.
10. Ready, M.P., Kim, Y.S., Robertus, J.D. Directed alteration of active site residues in ricin A chain. Proteins (in press).
11. Konnert, J.H., Hendrickson, W.A. A restrained-parameter thermal factor refinement procedure. Acta. Cryst. A36: 344–350, 1980.
12. Brunger, A.T. Crystallographic refinement by simulated annealing. In: "Crystallographic Computing 4: Techniques and New Tecnologies," Issacs, N.W. and Taylor, M. R. (eds.). Oxford: Clarendon Press, 1988.
13. Stout, G.H., Jensen, L.H. X-ray structure determination, 2nd ed. New York: John Wiley & Sons, 1989.
14. Agarwal, R.C. New least squares refinement technique based on fast Fourier transform algorithm. Acta. Cryst. A34:791–809, 1978.
15. Bunger, A.T. Crystallographic refinement by simulated annealing: Application to crambin. Acta. Cryst. A45:50–61, 1989.
16. Foxwell, B.M.J., Donovan, T.A., Thorpe, P.E., Wilson, G. The removal of carbohydrates from ricin with endoglycosidases H, F, and D and α-mannosidase. Biocim. Biophys. Acta 840:193–203, 1985.
17. Kimura, Y., Hase, S., Kobayashi, Y., Kyogoku, Yd., Ikenaka, T., Funatsu, G. Structures of sugar chains of ricin kD. J. Biochem. 103:944–949, 1988.