PROTEINS: Structure, Function, and Genetics 10:340–358 (1991)

Exploration of Disorder in Protein Structures by X-Ray Restrained Molecular Dynamics

John Kuriyan,[1,2] Klara Ösapay,[2] Stephen K. Burley,[3,4] Axel T. Brünger,[5] Wayne A. Hendrickson,[6] and Martin Karplus[7]

[1]Howard Hughes Medical Institute and [2]Laboratory of Molecular Biophysics, The Rockefeller University, 1230 York Avenue, New York, New York 10021; [3]Harvard Medical School, Health Sciences and Technology Division, 25 Shattuck Street, Boston, Massachusetts 02115; [4]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; [5]Howard Hughes Medical Institute, Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511; [6]Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biophysics, Columbia University, 630 W. 168th Street, New York, New York 10032; and [7]Department of Chemistry, Harvard University, 12 Oxford St., Cambridge, Massachusetts 02138

ABSTRACT **Conformational disorder in crystal structures of ribonuclease-A and crambin is studied by including two independent structures in least-squares optimizations against X-ray data. The optimizations are carried out by X-ray restrained molecular dynamics (simulated annealing refinement) and by conventional least-squares optimization. Starting from two identical structures, the optimizations against X-ray data lead to significant deviations between the two, with rms backbone displacements of 0.45 Å for refinement of ribonuclease at 1.53 Å resolution, and 0.31 Å for crambin at 0.945 Å. More than 15 independent X-ray restrained molecular dynamics runs have been carried out for ribonuclease, and the displacements between the resulting structures are highly reproducible for most atoms. These include residues with two or more conformations with significant dihedral angle differences and alternative hydrogen bonding, as well as groups of residues that undergo displacements that are suggestive of rigid-body librations. The crystallographic *R*-values obtained are ≈ 13%, as compared to 15.3% for a comparable refinement with a single structure. Least-squares optimization without an intervening restrained molecular dynamics stage is sufficient to reproduce most of the observed displacements. Similar results are obtained for crambin, where the higher resolution of the X-ray data allows for refinement of unconstrained individual anisotropic temperature factors. These are shown to be correlated with the displacements in the two-structure refinements.**

Key words: ribonuclease A, crambin, conformational disorder, protein crystallography, simulated annealing, X-ray refinement

INTRODUCTION

Spectroscopic experiments and molecular dynamics simulations demonstrate that protein molecules sample a large number of distinct conformations at room temperature, with anisotropic and anharmonic probability distributions for each atom.[1,2] Recent analysis of X-ray diffuse scattering from crystals of insulin indicate that liquid-like motions, with short-range correlations, account for most of the atomic mean-square displacements.[3] In contrast, the results of X-ray diffraction analyses are usually presented as unique structures having only isotropic and harmonic motional parameters for the atoms.[4,5] This simple description, which may seem to imply a relatively uncomplicated and well defined structure, is actually an economy in parameterization that is forced by the limited extent of diffraction from protein crystals. Although it is clear that lattice contacts can affect the displacements,[6] the mobility of protein molecules in crystals (which contain 30 to 80% solvent) is quite high, with rms displacements for ordered regions typically ranging from 0.5 to 1.5 Å.[4] The high mobility leads to the intensities of the Bragg reflections being very weak for most protein crystals beyond reflections corresponding to a resolution limit of about 2 to 1.5 Å, and results in a substantial decrease in the number of independent X-ray reflections that can be measured per atom when compared to results for small molecules. Refinement using sophisticated models that more ac-

Received September 12, 1990; revision accepted February 11, 1991.

Address reprint requests to Dr. John Kuriyan, Howard Hughes Medical Institute, Laboratory of Molecular Biophysics, The Rockefeller University, 1230 York Avenue, New York, NY 10021.

Present address of Stephen K. Burley: Howard Hughes Medical Institute, Laboratory of Molecular Biophysics, Rockefeller University, 1230 York Avenue, New York, NY 10021.

Present address of Klara Ösapay: Research Institute of Scripps Clinic, 10666 North Torrey Pines Road, La Jolla, CA 92037; On leave from EGIS Pharmaceuticals, Budapest, Hungary.

curately reflect the underlying disorder* is necessarily difficult unless constraints or restraints that reduce the number of free parameters are introduced, and the formulation of such restraints is one of the major remaining problems in protein crystallography.

The need for improved models can be seen by considering the final R-values obtained on refining protein structures. The conventional crystallographic R-value, R, is a measure of the agreement between the model and the X-ray data, and is given by

$$R = \sum ||F_o| - |F_c|| / \sum |F_o|$$

where $|F_o|$ and $|F_c|$ are the observed and calculated structure factors, respectively, and the summation is over all the measured reflections. The R-value for a random acentric structure is 59%, and for initial models obtained from map fitting the R-values are usually between 35 and 45%.[7] Protein refinements are often considered complete when R-values below 20% are obtained, though for the best refined structures the R-values are between 11 and 15%.[7] For small molecules it is not unusual to have R-values as low as 3–5%.[7] However, for well-refined protein structures with good data, the final R-values are usually a factor of 3 greater than the estimated errors in the data. In the case of crambin, after refinement of 0.945 Å resolution, with a model that incorporated disorder, hydrogen atoms, and limited anisotropy, the conventional R-value converged to 11% whereas the "error R-value," obtained with experimentally verified standard deviations in the numerator, was 4% for the same data (Hendrickson and Teeter,[8] unpublished results). The careful refinement by Wlodawer et al. of ribonuclease-A at a nominal resolution of 1.3 Å shows similar results.[9] R_{sym}† (on $|F|^2$) is only 5%; nevertheless, the conventional R-value (on $|F|$) for their model including 188 water molecules and alternative conformations for 13 side chains is 15%.[9] Along with the poorly modeled solvent scattering[10,52] and systematic errors due to absorption and radiation damage, the lack of a proper treatment of disorder in the protein structure is likely to be an important contribution to the high R-value.

The large residual R-values in the X-ray experiments may be compared with the results of test refinements against *simulated* data. A molecular dynamics simulation of myoglobin was used to generate time-averaged structure factors, and refinements of isotropic harmonic models for the protein against these simulated data resulted in R-values that were above 12%.[11] In this case the data have no error (they are calculated directly from the simulation by Fourier transformation and averaging), and the high R-values can be ascribed to the failure of the model used to describe the atomic probability distributions generated by the dynamics trajectory, most notably the lack of inclusion of anisotropy and anharmonicity, the latter often a result of multiple conformations.

Improvements in the characterization of disorder have become possible recently, with the availability of X-ray data to high resolution, improved refinement methods, and powerful graphics systems for interpreting electron density maps. Thorough investigations into disorder have been reported for crambin, erabutoxin, lamprey hemoglobin, and myohemerythrin, at resolutions ranging from 2 to 0.945 Å[12] and for ribonuclease-A at 1.3 Å resolution.[13] About 10% of the residues in these proteins exist in more than one localized conformation, and similar results have been obtained for CO-myoglobin.[14]

In all of these refinements, the crystallographers relied on visual inspection of electron density maps to locate and model the disorder. At 1.5 Å resolution, atomic positions separated by less than ≈1.2 Å are unlikely to result in distinct peaks in the electron density.[15] For side chains with widely separated conformations (for example, alternative conformers involving rotation about the side chain dihedral angle χ_1 = N–C_α–C_β–C_γ) this does not pose serious problems. For more subtle disorder all conformers may be within a continuous envelope of density, and the modeling requires considerable effort at interpretation.[12] Thus alternative conformations are usually included in a crystallographic model only when the electron density can be interpreted unambiguously, even though the presence of disorder may be evident in features of the electron density maps.

Analyses of molecular dynamics simulations of several proteins have revealed internal motions with important consequences for the interpretation of structures obtained by refinement against X-ray data.[11,16–19] These studies have not involved direct comparison with experiment, but rather have used the simulations as model systems in which to study the effects of various approximations. The atomic probability distributions are seen to be strongly anharmonic and anisotropic.[16] Studies on lysozyme[18,19] and myoglobin[11] indicated that the most important anharmonic effect is the existence of multiple peaks in atomic distribution functions, i.e., multiple potential wells or substates, rather than third- or fourth-order corrections to the harmonic distribution.

This paper is concerned with the extent to which information about alternative conformations can be

*Because of the difficulty of separating time-independent and time-dependent contributions experimentally, we use the term "disorder" to include all effects in this paper.

†The symmetry R-factor is given by $R_{sym} = \sum_j (I_j - \langle I \rangle) / \langle I \rangle$ where I_j is the jth measurement of the intensity of a particular reflection or a symmetry related reflection, and $\langle I \rangle$ is the average value of the intensity.

extracted from X-ray data at high resolution (minimum Bragg spacings of 1.5 Å or better). The restrained molecular dynamics method for refinement against X-ray data (simulated annealing)[20] is used for the analysis because it provides a powerful way of searching the energetically accessible conformational space of a protein for alternative structures that are consistent with the X-ray data. The results of restrained molecular dynamics are compared to those obtained using conventional "least-squares" optimization. In contrast to previous applications of restrained molecular dynamics, which use a single protein structure, the present study uses two noninteracting structures that contribute equally to F_c. The purpose of this extended model is to take account of the fact that when electron density peaks associated with a disordered atom are closely spaced, refinement using single-site models may not reveal information about conformational heterogeneity since these converge to final structures close to the mean position.[18] This model decreases the ratio of observations to parameters by a factor of two, and we limit ourselves to only two structures because the inclusion of any more is likely to make the system underdetermined and unstable. Even with only two structures, the increase in the number of parameters could lead to the introduction of meaningless shifts. We therefore consider only features that are consistently reproduced between independent annealing runs that are initiated with different sets of random velocities.

It must be emphasized that this procedure is not being suggested as an improved refinement method, but only as a search procedure to characterize the extent of disorder. We refer to this search procedure as "twin refinement." It should also be noted that the restrained molecular dynamics runs described here cannot be used to calculate physically meaningful correlation functions or other dynamic quantities. Specifically, a direct comparison with the results of X-ray diffuse scattering experiments[3,21,22] is not possible. While molecular dynamics simulations have been used to provide theoretical estimates of the time-dependent properties of proteins,[2] the inclusion of X-ray restraints and the use of unrealistic temperatures necessary for the simulated annealing calculations described here mean that the resulting structures are the only meaningful quantities.

Most of the work discussed here is focused on bovine ribonuclease-A, which was chosen as typical of proteins that diffract to high resolution. Ribonuclease-A has been previously studied extensively by crystallography[9,23,24] and computer simulation.[25] X-Ray data to 1.53 Å are used in the refinements (S.K. Burley and G.A. Petsko, to be published). We also consider results for crambin for which X-ray data to Bragg spacings of 0.945 Å are used.[12,26]

DATA COLLECTION AND LEAST-SQUARES REFINEMENT

Ribonuclease

Monoclinic crystals of bovine pancreatic ribonuclease (Sigma Chemical Co., Type III-A) were obtained in space group $P2_1$ with one molecule per asymmetric unit using the batch crystallization method with methanol as the precipitating reagent (40% methanol, pH 5.5 at 4°C). The protein:solvent volume ratio of this crystal form of ribonuclease-A is 53:47, which is typical of crystalline globular proteins. After crystallization was complete, the mother liquor was slowly exchanged with methanol to give an estimated final concentration of about 70% methanol/water by volume, with a freezing point of −80°C.[27] A large flawless crystal ($0.5\times\ 0.5\times 0.7$ mm^3) was mounted with its b axis parallel to the cylinder axis in a 0.7-mm quartz capillary tube, sealed with wax and 5-min epoxy, and then mounted on a conventional goniometer head.

Intensity data were collected using a Nicolet P3 automated 4-circle diffractometer with Ni-filtered CuK_α radiation from a sealed tube X-ray source operated at 40 kV and 25 mA. The diffractometer was equipped with a modified LT-1 low-temperature device, and the crystal was slowly cooled to −30°C over 24 hr. All intensity data were recorded at −30°C to reduce radiation damage. Fourteen reflections ($38.3 < 2\theta < 44.2°$) were used to determine the unit cell parameters [a = 30.15(3) Å, b = 38.22(3) Å, c = 53.02(4) Å, β = 106.45(5)°]. Reflection profiles recorded by ω-scan documented uniform peak shape throughout reciprocal space with a full width at half maximum of 0.25°, and Wyckoff scans were used for peak height measurements.[28] The crystal was moved through an ω-scan of 11 steps of 0.03° and the peak heights were calculated from the maximal 7 steps in the scan. Scan rates were chosen to permit collection of an entire unique data set to Bragg spacings of 1.53 Å (17,130 reflections) in 4 days from a single crystal. Diffracted intensity decay was monitored by measuring 5 standard reflections ($30.9 <2\theta<42.7°$) every 300 intensity measurements. The decay was quasilinear with crystal exposure time and isotropic over the 2θ range. The total loss of intensity during the 4-day data collection was 16%. We elected the use of low temperature and Wyckoff scans to allow collection of the unique data from single crystal rather than attempt to scale data from different crystals together, because we have found variables such as the degree of crystal disorder and differing absorption corrections difficult to control during scaling. The extent of radiation damage in this case is not significantly greater than the value of 10%, which is usually taken as the threshold for rejecting protein crystals during data collection.[29-31] An absorption curve was measured in steps of φ = 15° from a strong reflection at 2θ = 31°

TABLE I. Difference in Side Chain Dihedral Angles in Ribonuclease Between the Two Twin Structures Averaged for 15 Simulated Annealing Runs and the Same Dihedral Angle Differences Obtained by Pseudo-energy Minimization*

Residues	Average difference in dihedral angle and standard deviation Structures B1-1 to B1-15		Minimization alone Structure C	
1 Lys	$\chi_2 = 33 \pm 7$	$\chi_3 = 49 \pm 25$	$\chi_2 = 45$	$\chi_3 = 21$
7 Lys	$\chi_3 = 21 \pm 5$	$\chi_4 = 18 \pm 7$	$\chi_3 = 19$	$\chi_4 = 33$
11 Gln	$\chi_3 = 44 \pm 11$		$\chi_3 = 61$	
15 Ser†	$\chi_1 = 116 \pm 5$		$\chi_1 = 128$	
18 Ser†	$\chi_1 = 89 \pm 6$		$\chi_1 = 78$	
28 Gln	$\chi_3 = 164 \pm 4$		$\chi_3 = 2$	
31 Lys	$\chi_2 = 33 \pm 8$		$\chi_2 = 12$	
33 Arg	$\chi_2 = 25 \pm 8$		$\chi_2 = 23$	
34 Asn	$\chi_1 = 64 \pm 28$	$\chi_2 = 100 \pm 41$	$\chi_1 = 19$	$\chi_2 = 39$
35 Leu	$\chi_2 = 44 \pm 27$		$\chi_2 = 58$	
39 Arg	$\chi_4 = 27 \pm 5$		$\chi_4 = 18$	
41 Lys	$\chi_1 = 26 \pm 11$	$\chi_2 = 31 \pm 10$	$\chi_1 = 44$	$\chi_2 = 37$
		$\chi_3 = 31 \pm 10$		$\chi_3 = 52$
43 Val†	$\chi_1 = 120 \pm 5$		$\chi_1 = 136$	
44 Asn	$\chi_2 = 25 \pm 7$		$\chi_2 = 14$	
48 His	$\chi_2 = 21 \pm 6$		$\chi_2 = 15$	
50 Ser†	$\chi_1 = 116 \pm 4$		$\chi_1 = 116$	
51 Leu	$\chi_1 = 73 \pm 16$	$\chi_2 = 41 \pm 9$	$\chi_1 = 23$	$\chi_2 = 7$
59 Ser†	$\chi_1 = 133 \pm 4$		$\chi_1 = 131$	
61 Lys	$\chi_2 = 28 \pm 19$	$\chi_3 = 82 \pm 25$	$\chi_2 = 26$	$\chi_3 = 98$
		$\chi_4 = 97 \pm 70$		$\chi_4 = 130$
66 Lys	$\chi_1 = 24 \pm 8$	$\chi_2 = 14 \pm 5$	$\chi_1 = 28$	$\chi_2 = 2$
		$\chi_3 = 23 \pm 10$		$\chi_3 = 0$
70 Thr	$\chi_1 = 59 \pm 6$		$\chi_1 = 65$	
74 Gln	$\chi_3 = 38 \pm 11$		$\chi_3 = 7$	
76 Tyr	$\chi_2 = 23 \pm 3$		$\chi_2 = 16$	
83 Asp†	$\chi_1 = 92 \pm 15$		$\chi_1 = 86$	
85 Arg	$\chi_3 = 122 \pm 53$	$\chi_4 = 100 \pm 47$	$\chi_3 = 125$	$\chi_4 = 141$
86 Glu	$\chi_1 = 64 \pm 3$	$\chi_3 = 87 \pm 4$	$\chi_1 = 58$	$\chi_3 = 75$
87 Thr	$\chi_1 = 60 \pm 5$		$\chi_1 = 54$	
91 Lys	$\chi_2 = 26 \pm 10$	$\chi_3 = 38 \pm 23$	$\chi_2 = 10$	$\chi_3 = 17$
98 Lys	$\chi_1 = 16 \pm 13$	$\chi_2 = 73 \pm 51$	$\chi_1 = 67$	$\chi_2 = 25$
103 Asn	$\chi_1 = 72 \pm 5$	$\chi_2 = 117 \pm 6$	$\chi_1 = 3$	$\chi_2 = 41$
104 Lys	$\chi_2 = 75 \pm 17$	$\chi_3 = 101 \pm 23$	$\chi_2 = 27$	$\chi_3 = 1$
105 His	$\chi_2 = 26 \pm 7$		$\chi_2 = 23$	
111 Glu	$\chi_3 = 40 \pm 10$		$\chi_3 = 45$	
113 Asn	$\chi_2 = 77 \pm 70$		$\chi_2 = 2$	
114 Pro	$\chi_1 = 44 \pm 21$		$\chi_1 = 41$	
118 Val	$\chi_1 = 71 \pm 6$		$\chi_1 = 3$	
119 His†	$\chi_1 = 148 \pm 3$	$\chi_2 = 148 \pm 2$	$\chi_1 = 146$	$\chi_2 = 35$
123 Ser	$\chi_1 = 31 \pm 7$		$\chi_1 = 6$	

*The residues listed have at least one dihedral angle that satisfies the condition that the average dihedral angle difference be greater than 20° and the standard deviation be less than one-third this average. Additionally, sidechains that have separations of greater than 1.45 Å in the average twin structures are also listed. Pro 114 is listed although the separation is only 1.35 Å at the Cγ position as the alternatively conformers are clearly resolved.

†Disordered residues in Burley and Petsko structure.

with the crystal set at $\chi = 90°$, and was well fit by the function $A \sin^2(\phi-\delta)+B$.[32] Background data were collected at the point of maximum transmission between 2 layer lines in steps of $2\theta = 1°$, and the 2θ dependence was typical.[28] The ϕ dependence of the background recapitulated the form of the absorption curve.

Intensity data were corrected for Lorentz and polarization effects, and then corrected for crystal absorption,[32] followed by corrections for background scatter and linear intensity decay to give $|F_{obs}|$. Measurement of 17,130 reflections in the range $2<2\theta<60°$ resulted in 94.5% (16,172) of intensities greater than $2\sigma(I)$, which were retained for further analysis. The ribonuclease structure was refined against these X-ray data using the restrained least-squares program of Konnert and Hendrickson.[5] The final structure has an R-value of 15.6% at 1.53 Å resolution and includes 165 water molecules and 2 sulfate ions in addition to the protein (S.K. Burley and G.A. Petsko, to be published). In addition, 7 side chains were modeled with two conformations that

differed beyond the C_γ atom (see Table I). This model for ribonuclease will be referred to as the Burley–Petsko structure.

Crambin

The diffraction data used in these refinements of crambin derive from the measurements described in the structural report on crambin at 1.5 Å resolution.[26] However, processing of the limited-range step-scan data from beyond 1.5 Å spacings required software modifications that were not implemented until later. To account for the splittings of the α1 and α2 X-ray emission lines at high scattering angles, the program of Hanson et al.[3] was modified to fit appropriately fixed gaussian doublets to the peak profiles. In addition to this change, a decay correction based on a standard reflection was applied; otherwise the data reduction procedures were as described previously. Only the positive members of Bijvoet mates were recorded for reflections beyond 1.5 Å spacings. The data set for refinement at high resolution was composed from these reflections together with the positive set from the complete 1.5 Å data used in phase determination. The resulting data set comprises 22,145 reflections and is 100% complete to spacings of 0.945 Å. Overall 94% of the reflections have $|F| > 2\sigma_F$, and in the outermost shell between 1 and 0.945 Å, 89% meet this criterion (Hendrickson and Teeter, unpublished results).

The model that had already been refined at 1.5 Å resolution[26] was then refined against the extended data to 0.945 Å spacings. The refinement successively included the effects of anomalous scattering, the placement of hydrogen atoms, incorporation of heterogeneity and disorder, revisions to the solvent structure, and the use of anisotropic temperature factors with fixed ellipsoid orientations. Refinement of this model against the 18,889 reflections between spacings of 10 and 0.945 Å which have $|F| > 4\sigma_F$ gave $R = 11.1\%$ (Hendrickson and Teeter, unpublished results). The model comprises 729 atomic sites including 86 waters, 2 ethanols, and 7 discretely disordered sidechains. The stereochemical ideality of this model is typified by 0.015 Å rms deviation from ideal bond lengths between heavy atoms and 0.005 Å rms deviation for bonds to hydrogen atoms and, for thermal factors, by rms thermal fluctuations[34] of 0.054 and 0.025 Å for main chain bonds between heavy atoms and those involving hydrogens, respectively. This model, which is the same as that for which the conformational heterogeneity was described earlier,[12] was the starting point for the refinements described here, except that anisotropic temperature factors were not used.

TWIN REFINEMENT METHOD

Structure optimizations were done using molecular dynamics and energy minimizations using a pseudo-energy given by

$$E = E_{CHARMM} + S \sum_{hkl} w(hkl)\,(|F_o| - |F_c|)^2.$$

E_{CHARMM} is an empirical energy function given by the PARAM19 version of the CHARMM forcefield,[35] as modified by Brünger.[36] Electrostatic effects, which in the CHARMM potential include hydrogen bonding as well as ionic interactions, were not included in any of the calculations described here. This is consistent with most refinements at high resolution reported previously, where no restraint terms based on hydrogen bonding were introduced.[5] The program X-PLOR[20,36,37] was used for all the calculations described here, with unit weights assigned to all reflections. Starting models for twin refinements were generated by duplicating the original model, i.e., two identical protein molecules were created. The two molecules do not interact with each other, but both contribute equally to the X-ray diffraction term in the pseudo-energy function. In practice this was done by translating one of the structures by 3 or more unit cell lengths along one of the crystallographic axes. The nonbonded cutoffs used in the calculations are sufficiently small (8 Å) that the two structures are invisible to each other in terms of the empirical energy; however, the translational symmetry of the crystal guarantees their equivalence in terms of the X-ray structure factor calculation. The second molecule was translated back to the primary unit cell for analysis. X-Ray restrained molecular dynamics runs were carried out as described previously,[38,39] except that no intermolecular packing forces were computed (the use of two molecules in equivalent positions related to translations along cell edges prevents this).

X-Ray restrained molecular dynamics runs were initiated with these structures. The isotropic temperature factors from the previous least-squares optimizations were retained for all atoms. The weight on the X-ray term (S) was chosen so that reasonable geometry was maintained and the R-value did not rise above ≈30% during the dynamics. The dynamics stages were followed by pseudo-energy minimizations in which the atomic positions and temperature factors were optimized by a conjugate gradient procedure. The scale factor S was increased at this stage, so as to result in final structures with rms deviations of bond lengths from ideality of ≈0.025 Å in all cases. The positional and isotropic temperature factor parameters were optimized in alternating cycles, with no restraints being placed on the temperature factor variations.

DISCUSSION OF REFINEMENT STRATEGY

The twin refinement method doubles the number of variable parameters in the refinements and consequently the results must be interpreted with caution. Unfortunately, the incorporation of stere-

ochemical restraints, in the form of the empirical energy term, makes it extremely difficult to estimate the true number of free parameters in the system, and precludes the use of available statistical methods for evaluating the significance of the resulting R-values.[40] Our strategy is to use simulated annealing protocols that are similar to those used for the optimization of approximate models in early stages of refinement[39] and that have been shown to result in large structural changes in regions where the model is in error.[38] We expect that in twin refinements following these protocols, and initiated with different sets of random velocities, the displacements that are *reproducible* will be indicative of significant structural disorder.

The molecular dynamics runs, while important for sampling conformational space and for establishing the reproducibility of the results, are extremely time consuming. One twin refinement run requires approximately 25–100 hr of CPU time on a STARDENT GS1000 computer. We have therefore examined the extent to which pseudo-energy minimizations alone can delineate the disorder in ribonuclease. This procedure corresponds to carrying out just the last stage of the complete simulated annealing, and requires approximately 5 hr of CPU time. In this case, the reproducibility of the results can be estimated by introducing different random displacements into the initial structure, and repeating the optimizations (see below).

The presence of multiple conformations (i.e., conformations that are separated by a free energy barrier, or anisotropic distributions, with no barrier between the underlying conformations) will both result in displacements between the two molecules in the twin refinement. Test refinements were carried out for crambin, using as "data" structure factor amplitudes calculated from a model that had a single position for every atom, but with anisotropic temperature factors. Refinement of a twin model by pseudo-energy minimization resulted in relative displacements along the direction of the longest principal axis of the thermal ellipsoids that had been used in the "data" generation. This effect, which arises in the absence of alternative conformations so long as there is significant anisotropy, has to be considered along with the possibility of true multiple conformations when evaluating the results. The ability to distinguish between the two effects depends on the magnitude of the displacement and the resolution of the X-ray analysis (see below).

Ribonuclease Refinements

The Burley–Petsko structure includes seven residues in two alternative conformations each (Table I). Calculations were performed with these alternative conformations included from the beginning (the initial twin structures then differed at these seven residues) or not included at all (the major conformer was present in both twin structures initially). The calculations are grouped into the following four classes, based on the initial model and on whether they included a dynamics stage.

(A) The first refinement followed the protocol recommended by Weis et al.[39] for refinement of single structure models obtained by multiple isomorphous replacement or molecular replacement. However, in this case the starting model is the highly refined Burley–Petsko structure, with a single conformation for all side chains (the conformation of highest occupancy being retained in cases where the side chain was modeled with more than one in the BP structure). Velocities were assigned at 4,000 K and the system was coupled to a heat bath.[41] The temperature of the heat bath was lowered to 300 K over 1.85 psec, followed by pseudo-energy minimization and unrestrained temperature factor optimization (A1). The scale factor, S, was 90,000 during the dynamics and 550,000 during the minimization. A constant temperature (500 K) restrained molecular dynamics run of 1 psec duration, followed by minimization with the pseudo-energy functions and temperature factor optimization, gave very similar results (A2). The largest features in F_o-F_c difference maps calculated using these structures were near residue His-119. This residue is modelled by two alternative conformations in the Burley–Petsko model, and neither run A1 nor A2 resulted in the second conformation being located (see Results).

(B) For all further calculations, the initial structures had alternative conformations for seven side chains as in the Burley–Petsko model. Simulations were run for 1, 2, or 3 psecs at constant temperature (500 K) followed by cooling and pseudo-energy minimization (runs B1,B2,B3). The results of runs with varying total times were similar. Fifteen different twin refinement runs of 1 psec duration were carried out, starting with identical initial structures (the structures in each set differing only in alternate coordinates for the 7 disordered side chains in the Burley–Petsko model), but with different random velocities assigned at the beginning (runs B1-1 to B1-15) The results of these 15 runs are taken together for much of the following discussion.

(C) Twin refinement by pseudo-energy minimizations were initiated starting from the initial structures with no perturbations introduced. The starting twin structure was the same as in the B runs.

(D) Random displacements (rms value of 0.35 Å) were introduced in the initial model, followed by pseudo-energy minimization with no dynamics stage.

In addition, refinements (E) were also initiated with only single structures in the model. The starting structures in these cases were taken from the final structures obtained from the twin refinement protocol, i.e., the twin structures obtained by X-ray

restrained molecular dynamics were separated and refined individually as single structures.

Analysis of the results of the various twin refinements led to the identification of several residues as potentially existing in discrete conformations that had not been modeled as such originally (Table I). A composite model was created at this final stage that had a single structure for the bulk of the molecule, but allowed these residues to adopt more than one conformation for the backbone and side chain. Nonbonded interactions between atoms in alternative conformations of the same residue were not included. This treatment is similar to that described by Smith et al.[12] and, following these authors, the occupancy factors for disordered groups were not refined. Instead, they were set to values that were consistent with the corresponding electron density features and manually adjusted during the refinement so as to yield similar temperature factors for the alternative conformers.

A major unresolved problem in these refinements is the treatment of the solvent molecules for which localized electron density is present. The Burley and Petsko model includes 165 water molecules and 2 sulfate ions. The electron density shows features that are inconsistent with simultaneous occupation of many solvent sites. For this reason, and also because of the lack of covalent attachment, many water molecules were observed to move out of electron density during high temperature stages in trial refinement runs. To focus solely on effects due to disorder in the protein, the Burley and Petsko solvent model is kept fixed throughout the refinements. The two twin structures in each refinement thus include the same (fixed) solvent model. In the course of refinement of the composite single structure/alternative conformer model, the solvent model was reevaluated on the basis of difference Fourier maps and hydrogen bonding possibilities. Twenty-nine additional water molecules were included at this stage and the protein and solvent parameters were simultaneously optimized, the only optimization of solvent structure performed in this work.

Crambin Refinements

The resolution in the ribonuclease analysis (1.53 Å) makes it difficult to distinguish in many cases between the effects of anisotropy and discrete conformational disorder. We therefore turned to crambin, which diffracts to beyond 0.8 Å spacings[42] and for which data to 0.945 Å are available[26] and for which the discrete disorder has been described.[12] The twin annealing procedure was repeated for this small protein. The only difference from ribonuclease was that the dynamics runs were performed at 1.5 Å resolution and a scale factor of 60,000, followed by pseudo-energy minimizations at 0.945 Å resolution, and a scale factor of 480,000. All hydrogens were included in the structure, as in the original model, but were not included in the X-ray structure fact or calculation. All structure factor calculations for ribonuclease and crambin at 1.5 Å resolution used the subgrid fast Fourier transform algorithm.[43] For crambin at 0.945 Å resolution, the finite sampling of the electron density led to significant error in the high resolution shells and a direct summation structure factor calculation was used instead.

The very high resolution of the crambin analysis makes possible the refinement of reliable unconstrained anisotropic temperature factor parameters for each atom. These were refined for a single structure, using a conjugate-gradient procedure implemented in the program X-PLOR (J. Kuriyan, unpublished). The positions of atoms were kept fixed at their original values during the refinement, and anisotropic temperature factors were determined for nonhydrogen protein atoms. For disordered residues, only the major conformer was included in this treatment. The reliability of the results was estimated by initiating refinements from different initial conditions such as uniform or random tensor elements or tensor elements equivalent to the isotropic temperature factors. At 0.945 Å resolution, the tensors resulting from all refinements were very similar, with 86% of the atoms having principal axes of greatest displacement that do not deviate by more than 10° in the different refinements and with 95% of the atoms having deviations less than 20°. When the minimal Bragg spacing of the data used in the refinement is lowered to 1.2 Å, only 50% of the atoms have the axes within 10° of the axes determined at 0.945 Å, and only 75% have them within 20°. On further reducing the resolution to 1.5 Å, the numbers drop to 18 and 50%, respectively. Refinement of unconstrained anisotropic parameters for ribonuclease at 1.5 Å was not well-behaved and led to many tensors having nonphysical negative determinants. Only tensors for crambin, refined at 0.945 Å, are used in the analysis. Anistropic temperature factors were not used for the twin refinements.

RESULTS

The refinements on ribonuclease are presented first. We show that the results of refinements using single structures do not deviate significantly from the initial X-ray model. However, when two molecules are simultaneously included in the model, relatively large displacements between the two structures are observed. These are described next, with particular emphasis being placed on the degree to which the displacements are reproducible in the 15 independent B1 runs. The displacements are analyzed in detail in terms of dihedral angle differences, alternative hydrogen bonding possibilities, and segmented motions of parts of the protein. Finally the R-values obtained on refining the various models are discussed, along with the results of refining a composite model for the protein that has selected residues in alternative conformations. Refinements

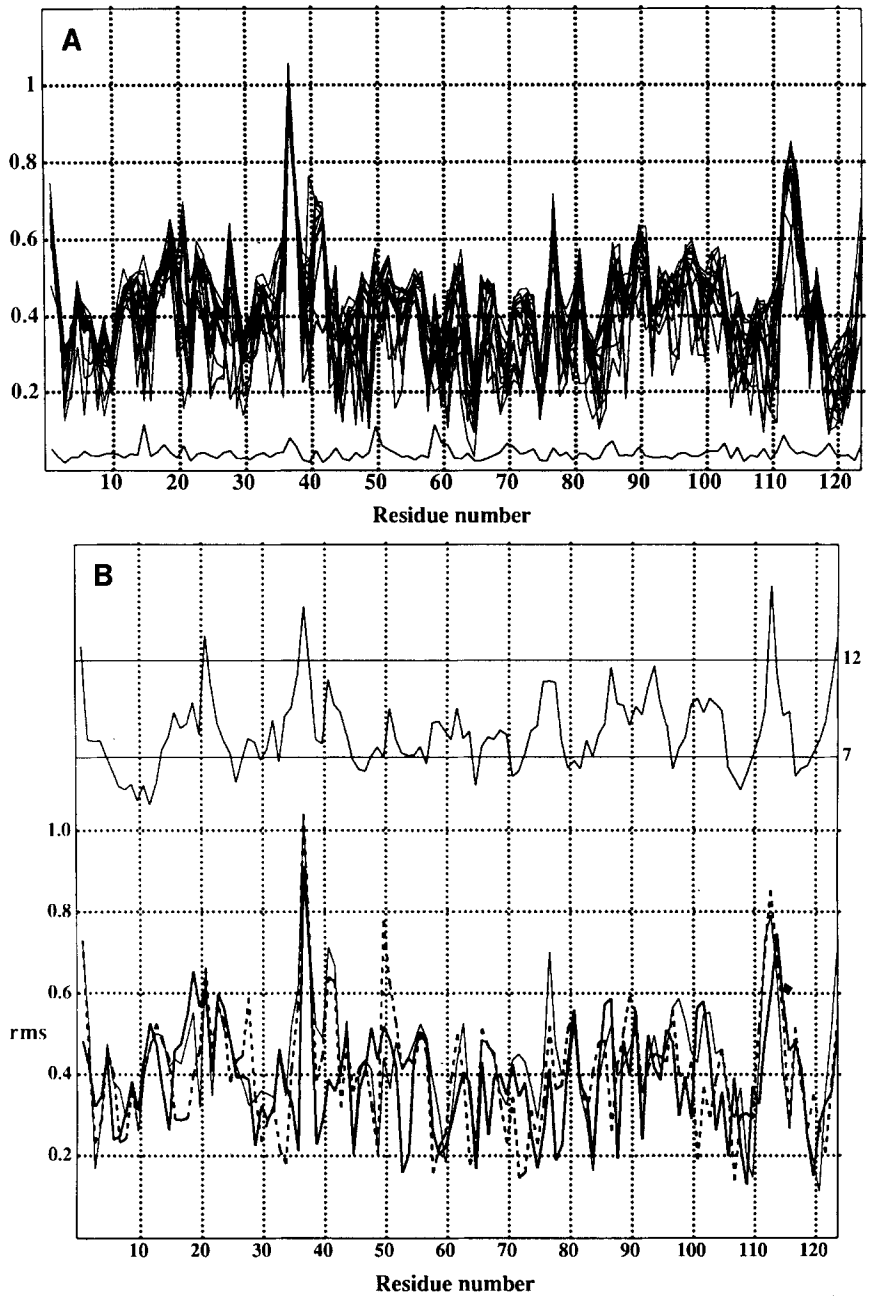

Fig. 1. Backbone rms deviations (in Å) for ribonuclease, as a function of residue number. (**A**) rms deviations between the twin structures obtained by runs B1-1 to B1-15 (upper lines) and deviations in backbone positions for one of the single structure refinement relative to the Burley and Petsko model (lower line). (**B**) Rms deviations between the two molecules in runs A1 (dashed line), C (thin line), and D (thick line). The upper part of the figure shows the backbone residue-averaged temperature factors ($Å^2$) for the Burley–Petsko structure.

of crambin are presented in the last part of this section, and the results of refinement of unconstrained anisotropic temperature factors for each atom are described.

Ribonuclease

Deviations between refined ribonuclease structures

Figure 1 shows the deviations in backbone (N,C_α,C) positions between the two structures for each of the 15 independent twin refinement runs (B1-1 to B1-15). The rms deviation in backbone position between the two protein molecules in each of the 15 twin structures is between 0.43 and 0.45 Å. For each of the twin structures there is considerable variation in the backbone deviations, which range from ≈0.05 Å for residue 65 to >1 Å for residue 37. A striking feature is that there is close agreement in the pattern of backbone displacements for all 15 independent sets of twin structures. Figure 1 also

shows the deviations in backbone positions relative to the initial X-ray structure for one of the single structure refinements. This structure (E) is obtained by deleting one of the two molecules in a twin set after one of the B1 runs, and optimizing the single molecule against X-ray data. The displacements shown are typical of single structure refinements, and are rarely greater than 0.1 Å (rms value = 0.055 Å).

The rms displacement of 0.055 Å between structures (E) that are obtained by refinement of single structures is related to the reproducibility of the backbone positions and corresponds to an estimated standard deviation of approximately 0.04 Å. This is a measure of the extent to which alternative local minima and the limited number of steps taken during least-squares optimization prevent convergence to the original structure. This standard deviation, which takes no account of systematic errors in the data, is smaller than the estimates of ≈0.15 Å for the rms error in backbone positions in ribonuclease obtained by comparing the results of refinements against independently measured data sets.[44] In contrast, the deviations between the twin structures are about an order of magnitude higher. For example, the deviation between the backbone atoms of residue 37 in two structures refined simultaneously is > 1 Å in all cases, whereas the deviation between structures refined individually, after completion of the twin refinement phase, is less than 0.04 Å for the same residue. The structures (E) were obtained by selecting individual structures from the twin B1 structures, and the relatively low rms displacement between the singly refined structures indicates that the separation observed in the twin structures is not due to the restrained molecular dynamics having moved the atoms beyond the radius of convergence of the pseudo-energy minimizations.

The large deviations observed on refining 2 structures simultaneously are correlated with the temperature factors of the backbone (Fig. 1B). A question that arises is whether these displacements are merely manifestations of increased uncertainties associated with the positional coordinates due to the doubling of the variable parameters in the optimization. For the most part this appears not to be the case, because the displacements are highly reproducible in all 15 of the independent runs as well as in runs A1 and A2. Figure 2 shows portions of the ribonuclease molecule corresponding to some of the regions with largest displacement in the twin structures. Each of the panels shows 30 different molecules (2 each from the 15 independent twin runs), yet in many places the structures are so closely overlapped that only 2 or three structures appear to be present. For any one pair, the structures are invariably separated into two.

The extent to which the structures obtained in the 15 B1 twin refinements are clustered has been quantified. For each atom, a mean-square displacement tensor is calculated based on all 30 positions of the atom. The principal axis of greatest displacement is then determined. All positions corresponding to positive displacement along this axis are placed in one "cluster," and those with negative displacement in the other. Distances d_1 and d_2 are then calculated, where d_1 is the average distance between atomic positions within the same cluster, while d_2 is the average distance between atomic positions that are not in the same cluster. The ratio d_2/d_1 is a measure of the extent to which the 30 positions for each atom fall into well separated clusters (Fig. 3A). This ratio is greater than 3 for 81% of the atoms. For comparison, a spherically symmetric random distribution of 3,000 points was partitioned in a similar way using 30 randomly picked points at a time. The mean value of d_2/d_1 was 1.2 ± 0.1, indicating the nonrandom character of the twin simulation results (Fig. 3A). Figure 3B shows that for the 30 structures from the B1 runs, the atoms within a cluster are very tightly grouped together (modal deviation from the mean of ≈0.05 Å).

Two twin refinement runs were also carried out without any molecular dynamics stage. In one (C), the starting structure was the same as in the 15 runs with 1 psec dynamics, and the twin structure was simply subjected to least-squares optimization of the positional and temperature factor parameters, as in the final stages of the other runs. Figure 1B shows backbone deviations between twin structures obtained by slow cooling (A1), minimization alone (C), and by minimization starting with random displacements (D). Most of the shifts observed in the molecular dynamics runs are reproduced in the calculations that do not include a molecular dynamics stage (Fig. 1B, and see below).

Dihedral angle and hydrogen bonding differences

Table I lists the average difference in dihedral angles between the two structures in each of the 15 B1 twin sets, and the standard deviations in these differences. Some residues, particularly charged side chains, show dihedral angle deviations that are large, but with standard deviations that are also comparable in magnitude. These are residues that are not well localized in the 15 refinements, such as Lys-98 (Fig. 2B). The electron density for these residues is very weak, and usually below the noise level of the maps. The second class includes residues that show large shifts in dihedral angle, with standard deviations that are much smaller. These correspond to residues with 2 or 3 well-defined conformations, such as Asn-34 (Fig. 2A). The disorder in this residue was not accounted for in the original model, but is clearly visible in the electron density (Fig. 4).

It is interesting to note that some of the conformations found by the twin refinements of the Burley

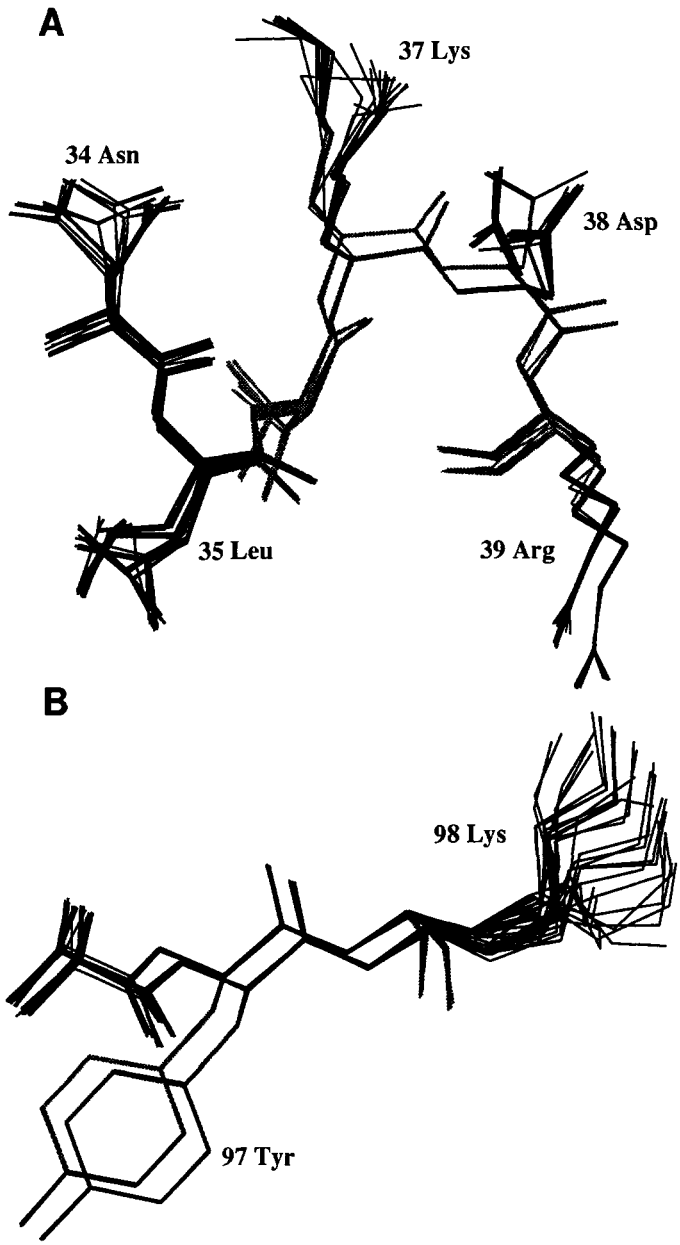

Fig. 2. (**A–E**) Portions of ribonuclease-A obtained by runs B1-1 to B1-15 after the twin structures had been translated back into the same unit cell. Each figure shows 30 structures (two from each of the twin structures). The residue numbers are indicated. Note that in many places the structures are so closely superimposed that a smaller number of structures appear to be present. Figures 2C and 2D appear on page 350 and Figure 2E on page 351.

and Petsko data correspond to those reported by Wlodawer et al.[9] For example, the two conformations for Asn-34 shown in Figure 4 have torsion angles of (χ_1 = −56 and −159) and (χ_2 = 114 and 38), respectively. Wlodawer et al. report (χ_1 = −55 and −148) and (χ_2 = 132 and 43), for the same residue, close to the conformations found by simulated annealing. The large standard deviations in the torsion angle differences for this residue (Table I) are due to the presence of a third conformation ($\chi_1 \approx -100$, $\chi_2 \approx 90$) found in 23% of the structures (Fig. 2A); this had not been reported previously.

Some of the side chains with two or three localized conformations are observed to have potential hydrogen-bonding interactions that are different for the alternative conformers. No electrostatic or hydrogen-bonding interactions are included in the empirical energy term used, which implies that the potential hydrogen-bonding interactions in the refined structures reflect features in the electron density. An interesting example is Lys-41, at the active site of the enzyme (Fig. 2E). This residue, which shows a small systematic difference between the torsion angles in all 15 structures, is seen to have potential

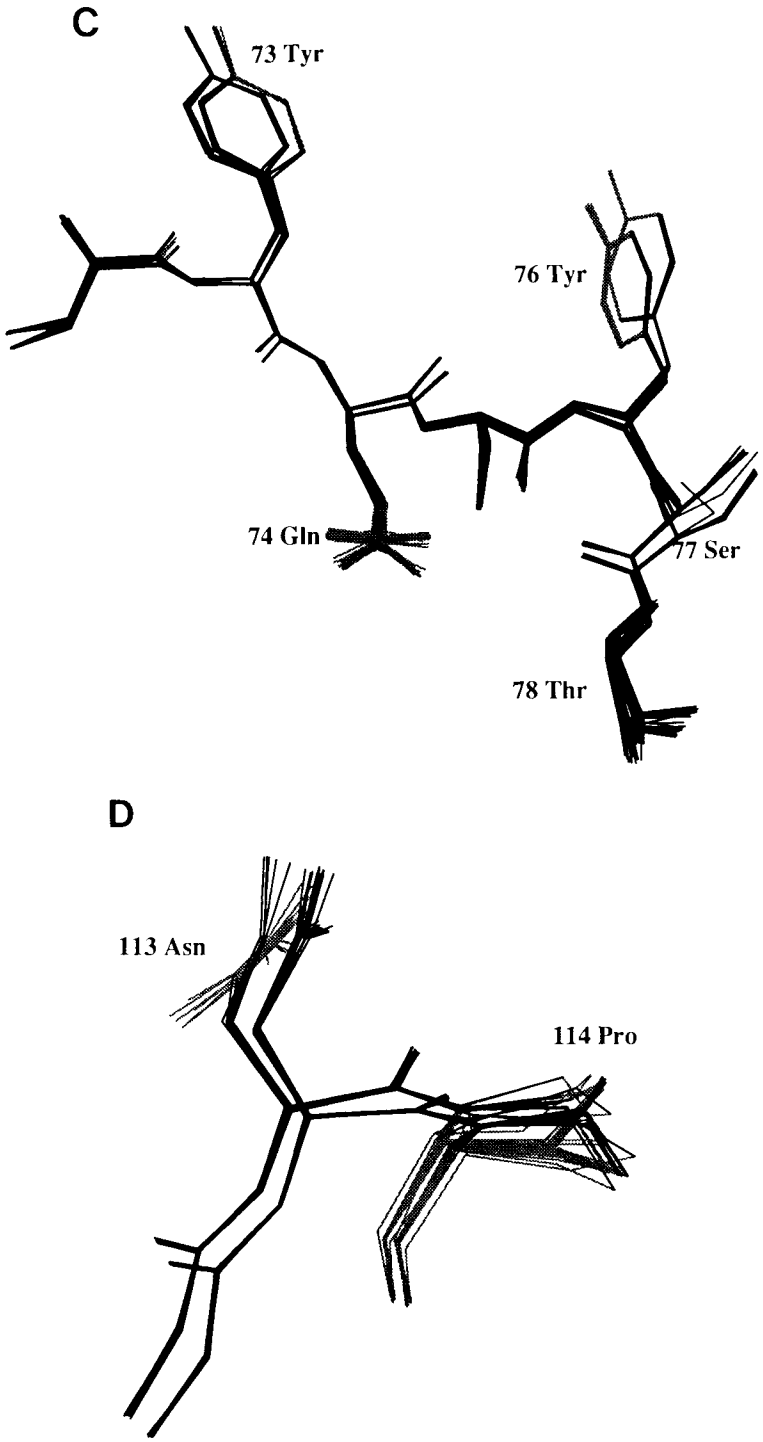

Fig. 2C,D. Legend appears on page 349.

hydrogen-bonding interactions with Gln-11 in one conformation, and with Asn-44 in the other.

Five other residues have alternative hydrogen-bonding interactions involving protein atoms. These are Asp-83, Ser-15, Lys-37, Ser-50, and Lys-61. Lys-61 has an interaction to one of the sulfates in one of its conformations and forms a salt bridge to Glu-9 in a symmetry related molecule, in accordance with other reports.[9,13] Several other residues, for example, Gln-28, Lys-31, Ser-59, Tyr-76, Thr-87, and Asn-113 have alternative H-bonding interactions to water molecules.

In the 15 B1 annealing runs, the initial structures used had 7 residues in the alternative conformations

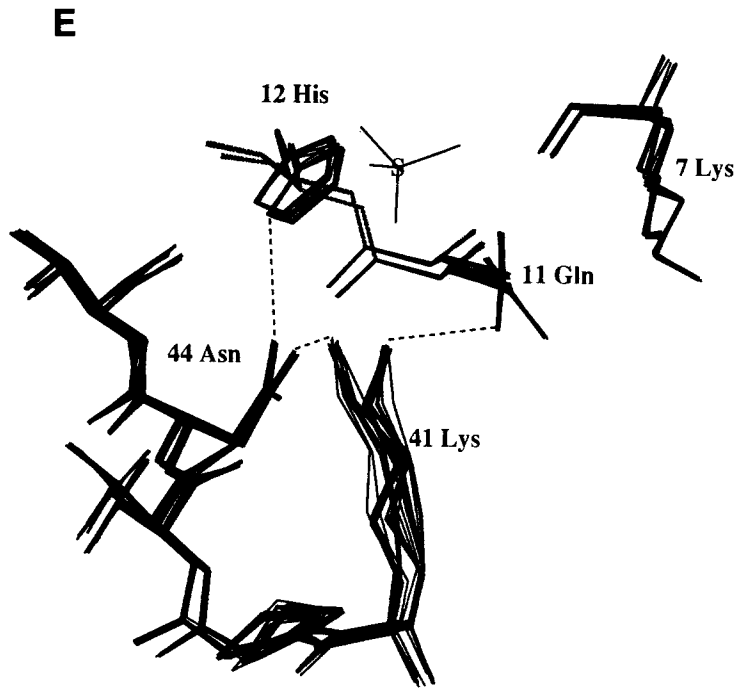

Fig. 2E. Legend appears on page 349.

determined by Burley and Petsko (Table I). Several simulated annealing runs (A1,2) were initiated with identical structures used for these residues (i.e., no disorder present initially). None of the refinements, even one initiated at 4,000 K (A2), was able to locate the second conformation for the active site histidine 119. Increasing the lengths of the runs from 1 to 2 or 3 psec had no effect on the results. The inability of these refinements to locate the alternative conformation, which is separated from the other by a 148° rotation about χ_1, is probably due to the solvent structure being kept fixed in these calculations. The His-119 side chain is close to one of the two sulfate ions, and rotation about the χ_1 dihedral brings it into collision with the sulfate (interatomic distances less than 2 Å). Rotation in the opposite direction brings the side chain into equally close contact with the backbone atoms of the adjoining β-strand. For the other 6 disordered residues, all of the disordered conformations were located by the annealing procedure.

Table I lists the residues for which the average difference in any dihedral angle is more than 20° and more than 4 times the standard deviation. There are 31 such residues in addition to the 7 already modelled as disordered by Burley and Petsko. Twin minimization without dynamics (Run C) is able to reproduce most of the dihedral angle changes, including some that are as large as 60° (residues 61, 85, 86; see Table I). There are very few residues with well localized alternative conformations in the 15 runs (B1-1 to B1-15) that are not reproduced in the direct minimization (e.g., 28, 103, 104, 113, 118). The optimizations starting from an initial structure with random perturbations introduced into the positions of every atom (rms shift of 0.35 Å) yield overall results that are very similar (Fig. 1B).

Segmented displacements

Another class of residues shows fairly large deviations (0.5–1 Å), with highly clustered separation of the two structures, but with very small dihedral angle changes. The extremely tight clustering of the two sets of structures indicates that the observed shifts reflect an underlying feature of the electron density. However, since the displacements involved are about 1 Å or less, any alternative conformations that may be present cannot be clearly resolved in the 1.53 Å resolution electron density map. Examples are Tyr-73 and Tyr-97 (Fig. 2) where the dihedral angle differences between the two structures are less than 10°. For Tyr-97 (Fig. 2B), the displacement can be described qualitatively as a rigid body shift that extends over neighboring residues. These displacements are correlated over several residue segments, as is schematically shown for the whole protein backbone in Figure 5. The displacements between the C_α atoms of 2 structures obtained by averaging over the 15 B1 pairs are shown to be strongly correlated over the helices, sheets, and loops in the protein (Fig. 5).

We have explored the extent to which rigid-body motions can account for these displacements in one

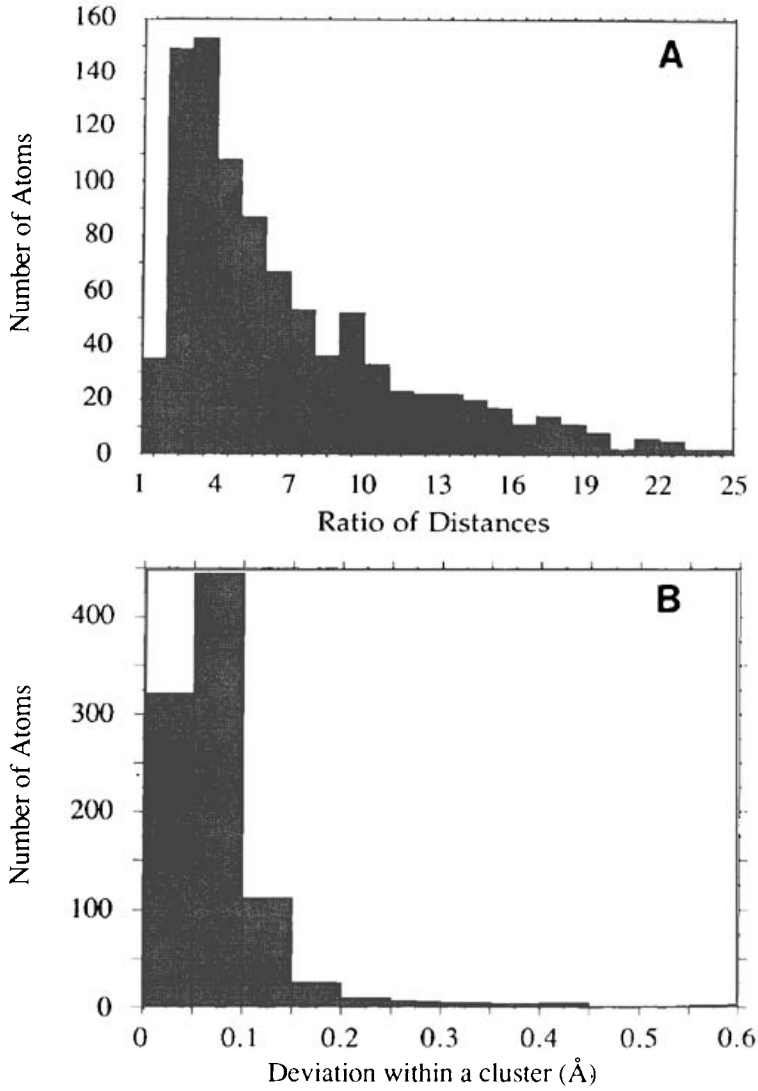

Fig. 3. Analysis of the clustering of atomic positions in the 15 31 ribonuclease structures. Every atom in the protein has 30 positions that are placed in one of two clusters (see text). For each atom, distances d_1 and d_2 are calculated where d_1 is the average distance between positions that are in the same cluster and d_2 is the average distance between positions that are in different clusters. Distances d_3 are also calculated for each cluster, where d_3 is the distance between any particular position and the mean position within the cluster. (**A**) Histogram of the values of d_2/d_1, for all atoms in the protein. (**B**) Histogram of the values of d_3 for all 30 positions of each of the atoms.

region of the protein, residues 94 to 97 (Fig. 6A). Figure 6B shows the thermal ellipsoids[45] corresponding to the mean-square displacement tensors calculated from one of the B1 twin structure pairs. These ellipsoids have their longest axes along the direction of displacement between the two structures, and have two smaller and equal axes, the lengths of which are given by the average temperature factor of the two atoms. With these mean-square displacement tensors as target values, anisotropic temperature factors were refined for atoms in residues 94 to 97 using the rigid-body librational model of Schomaker and Trueblood (Fig. 6C).[46] This method considers these residues to constitute an internally rigid body, and anisotropic temperature factors result from translations and screw-librations. The deviation between the principal axes of the rigid-body anisotropic tensors and the direction of displacement in the twin structures is 27 ± 9°. While the anisotropic tensors were obtained using only one pair of structures, Figure 6A shows all 30 structures obtained from the B runs. It is clear that the simple rigid-body model provides a qualitative description of the overall displacements in this region.

Moss and co-workers have shown that refinement of rigid-body parameters for aromatic rings and other small groups in ribonuclease leads to lower standard deviations in difference electron density maps.[24] Our results suggest that larger segments of

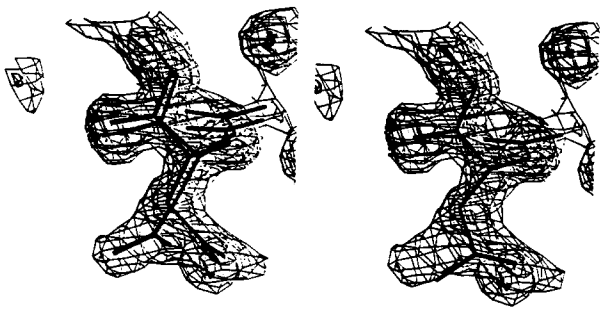

Fig. 4. Electron density maps for Asn-34 in ribonuclease. Left, two molecules from the B1 twin refinement. Right, one molecule from the random displacement refinement. The electron density map is the same in both figures, and was calculated using phases from the B1 twin refinement structure and amplitudes given by $(2|F_o| - |F_c|)$. Note that the random displacement/refinement results in a structure that is intermediate between the two conformations in the twin refinement.

the protein (3 to 4 residues or more) might also be treated in this manner. Residues 94 to 97 contain 31 atoms, and 20 parameters suffice to describe all the anisotropic temperature factors in the rigid body approximation, as opposed to the 186 parameters that would be required for the full treatment of anisotropy. The refinement of reliable unconstrained anisotropic thermal parameters is not possible for ribonuclease at this resolution, and the correlations evident in the C_α displacements (Fig. 5) point out the importance of developing simplified models for anisotropic motion.

R-values and refinement of a composite model for ribonuclease

The R-values for the various structures are given in Table II. The twin structures (B1), including the 7 disordered side chains of the Burley–Petsko model, have R-values of ≈13%, as compared to 15.3% for a model with one conformation only for all residues (E). Runs that were initiated with identical conformations for the disordered residues in the original crystallographic model (A1,2) converged to a slightly higher R-value (13.3%) than the others. The longer runs (B2, B3) did not lead to improved R-values, nor did the run initiated at 4,000 K (A1). The result of minimization alone (without dynamics, C) is an R-value of 13%, essentially the same as that of the B runs. The introduction of random displacements before minimization leads to a slightly higher R-value (13.2%). The statistical significance of the decreased R-value in the twin refinements is difficult to establish. The refinements have been done without proper weighting of the structure factors, and the incorporation of stereochemical restraints via the empirical energy makes it difficult to determine the number of independent parameters.

In order to reduce the number of parameters in the model while still accounting for the alternative conformations indicated by the twin refinements, a composite model was generated in which selected residues were allowed to adopt more than one conformation, similar to the procedure described by Smith et al.[12] Thirty-eight residues were treated as disordered; these are listed in Table I. Only one occupancy factor was used for all atoms in one conformation of a residue, and the sum of the occupancies of all conformations of a residue was always unity. The occupancies were adjusted periodically so that the resulting temperature factors of atoms in alternative positions were similar. None of the conformations has temperature factors above 50 Å^2 or occupancies below 30%. F_o-F_c difference maps were periodically examined during the refinement. Several positive peaks were found at 5 to 10 standard deviations above the average value of the maps. Most of these were in positions that are reasonable for water molecules and were in hydrogen bonding distance of protein or water atoms. Of these peaks 29 were finally interpreted as water molecules. At this stage, the water and sulfate ions were also allowed to move. The occupancy factors for all waters were set to 1.0 and nonbonded interactions between water molecules that approached each other closely were not included, because the electron density clearly indicated several sites where water molecules are present on overlapping positions. The total number of water molecules now included is 194, as compared to 165 in the original Burley–Petsko model and 188 in that of Wlodawer et al.[9] The R-value for this model is 12.8%.

Crambin

Results obtained on refinement of twin structures at very high resolution (0.945 Å) for crambin allow comparison with unrestrained anisotropic temperature factors. The overall results for the X-ray restrained molecular dynamics runs are very similar to those for ribonuclease. For example, Tyr-29 is found to occupy two sites, analogous to the displacements found in ribonuclease (Fig. 7A). This residue has been modeled previously with two conformations by Hendrickson and co-workers[12] and even though the displacements are only on the order of 1 Å, at this resolution the electron density clearly shows that a two site model is better than a single site with anisotropy (Fig. 7). Another example is Pro-36, for which two puckerings of the ring have been reported,[12] with χ_1 dihedrals of −18° and 28°. The twin refinements produce two puckerings as well, with this dihedral at −22° and 28°.

The rms deviation between backbone atoms of the two structures included in the twin refinements is 0.31 Å, and a number of residues in crambin show significant deviations and yet are within a continuous envelope of density with no evidence of discrete disorder. These displacements are highly correlated with the principal axis of greatest displacement calculated from the anisotropic temperature factors obtained by unconstrained refinement. For example,

Fig. 5. Displacements between the C_α atoms of 2 ribonuclease structures obtained by averaging over the 15 B1 pairs. The average structure is displayed, with arrows showing the direction and magnitude of the displacements in the averaged twin structure. For clarity, the length of the arrows is multiplied by 4. The residue numbers are indicated.

thermal ellipsoids from the anisotropic temperature factor refinement are shown for Arg-10 in Figure 7C and comparison with the twin structures (Fig. 7B) shows that the terminal atoms of the side chain have these ellipsoids oriented along the direction of displacement between the twin structures. For all atoms in the protein, the mean angle between the longest principal axes of the thermal ellipsoids and the direction of displacement in the twin refinement is 19±13°, and the normalized distribution of this angle is shown in Figure 8. The distribution is peaked at 0°, and falls off to zero at high angle, in contrast to the sin(θ) distribution expected for randomly oriented vectors (Fig. 8).

The R-values at 1.5 and 0.945 Å for the various crambin refinements are given in Table II. The twin simulation including the alternative conformations of Smith et al. yields an R-value of 10.5% at 0.945 Å, i.e., increasing the number of parameters per atom by 4 leads to a drop in R-value of only 2.1%. Refinement of anisotropic temperature factors for the original model (6 parameters per atom) gives an R-value of 9.5%. It is difficult to compare these R-values, however, since the 6 additional parameters introduced for each atom in the anisotropic case are completely unconstrained, whereas those introduced by the twin refinement are subject to stereochemical restraints. As in ribonuclease, minimization alone yields results that are very similar to those obtained with molecular dynamics, and high temperature dynamics does not introduce any improvement.

DISCUSSION

For disordered materials such as glasses good agreement with X-ray experiments cannot be obtained unless the structure factors are averaged over a large number of Monte Carlo configurations.[47] The complex structure of proteins likewise suggests that a very large number of alternative configurations may have to be explicitly accounted for. Gros et al. have recently proposed a novel refinement method aimed at addressing this prob-

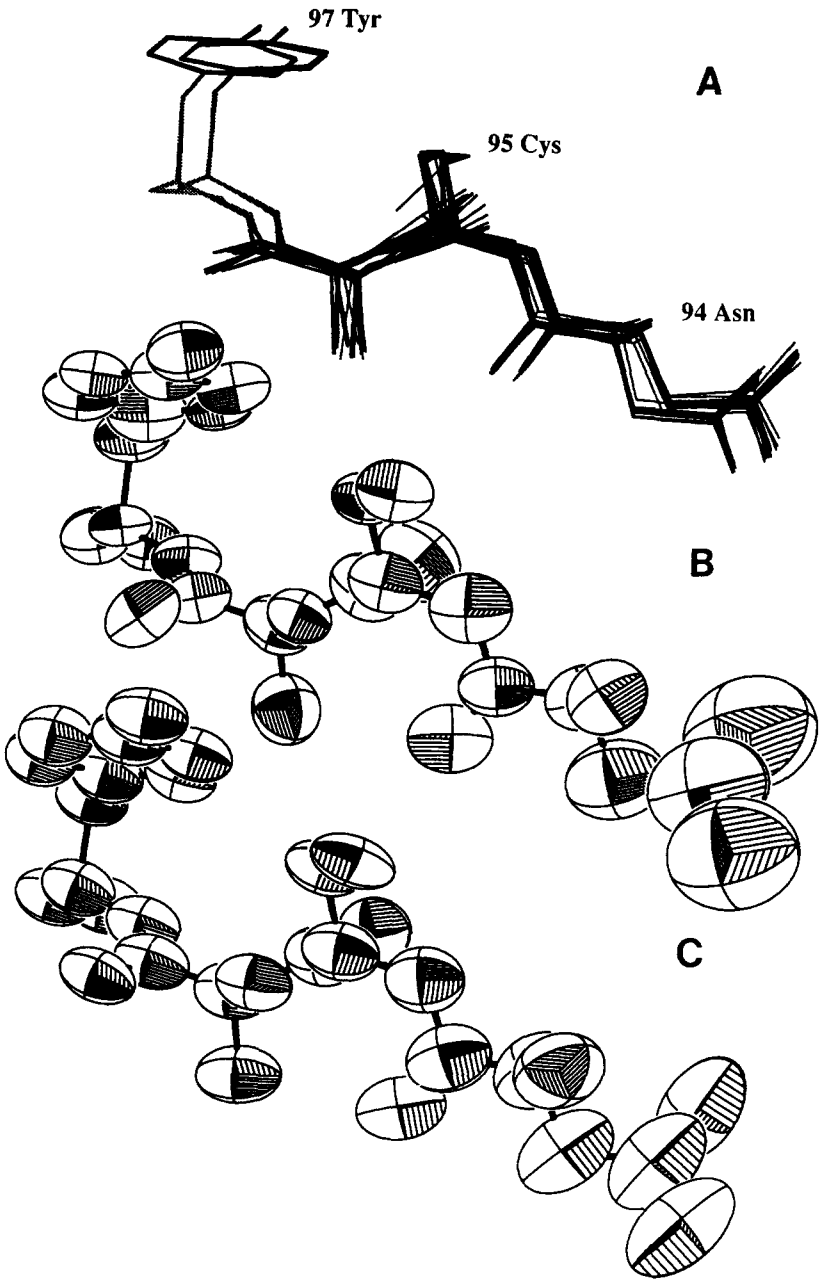

Fig. 6. Residues 94 to 97 in ribonuclease. (**A**) Thirty structures obtained by runs B1. (**B**) Thermal ellipsoids with mean-square displacements that are equivalent to those for a pair of B1 twin structures. The anisotropy arises from the displacements between the two structures. (**C**) The thermal ellipsoids obtained by rigid-body (TLS) refinement for these four residues against the thermal ellipsoids in (**B**).

lem.[48] They carry out X-ray restrained molecular dynamics simulations where the restraint term includes a time-average over a number of configurations of the protein, in contrast to the procedure described here where the restraint term depends only on the instantaneous structure. An ensemble of structures is generated by their procedure, and it results in a low time-averaged R-value ($< 10\%$), and accounts for the anisotropy and anharmonicity of the structural fluctuations. Caspar and Badger have recently obtained equally low R-values by a solvent modeling procedure.[52] However, these calculations involve the introduction of a large and uncertain number of additional free parameters (several tens of structures are involved in the averaging procedure of Gros et al., albeit linked by the potential function), and the significance of the results needs to be studied further.

We have determined whether significant improvements in the fit to X-ray data could be obtained by

TABLE II. *R*-Factors of the Various Structures for Ribonuclease-A and for Crambin*

	R-factor at 1.5 Å	Δbonds† (Å)	*R*-factor at 0.945 Å
RNase			
Original model	15.6	0.017	
Slow cooling A1	13.3	0.025	
1 psec dynamics 500K A2	13.3	0.025	
1 psec dynamics 500K B1-1 to B1-15	13.08 ± 0.12	0.025 ± 0.001	
2 psec dynamics B2	13.1	0.024	
3 psec dynamics B3	13.0	0.024	
Random displacements D	13.2	0.023	
Minimization C	13.0	0.025	
Single structure minimization E	15.3	0.024	
Composite model	14.0	0.027	
Composite model, refined solvent	12.8	0.027	
Crambin			
Original model	10.9	0.022	12.7
Original model + anisotropic B		0.022	9.5
Slow cooling A1	9.1	0.025	11.0
1 psec dynamics 500K A2	8.9	0.023	10.9
1 psec dynamics 500K B	8.7	0.022	10.5
Minimization C	8.8	0.022	10.6

*The A structures are obtained by refinement runs started with the same initial conformation for the twins, runs B, C, and D had alternate conformations for residues reported as disordered in the original structures.

†Δbonds: rms deviation of bond lengths from ideality. For the original structures of ribonuclease, the ideal values are those used in the PROLSQ program.[5] For all other structures, the reference values are those given in the parameter set used in X-PLOR.[35]

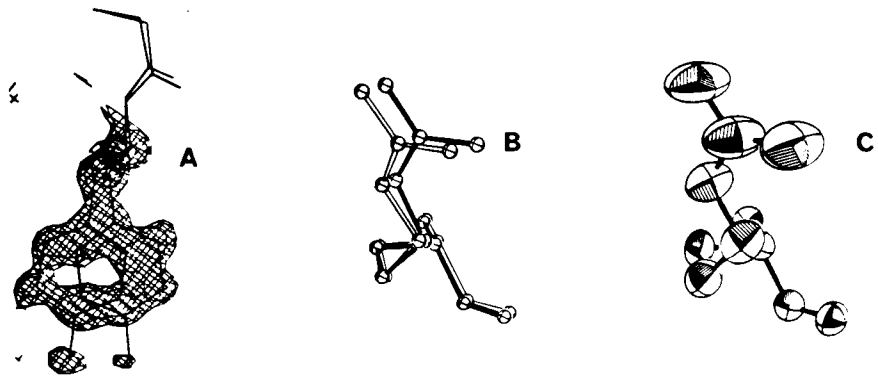

Fig. 7. (**A**) $|F_o| - |F_c|$ map for Tyr-29 in crambin, with the esidue omitted from the phase calculation. The two structures shown are from the twin refinement (B). Note the well separated peaks for the terminal OH group. (**B**) Two structures for Arg-10, from the twin refinement. (**C**) Thermal ellipsoids for Arg-10, obtained by refinement of unconstrained parameters, against the X-ray data.

averaging calculated structure factors over the many configurations generated by our restrained molecular dynamics runs. The method as described here does not lead to improved *R*-values on such averaging. In the case of ribonuclease, an *R*-value of 12.8% is obtained by averaging F_c over the 30 structures corresponding to the 15 B1 runs, which is not significantly better than values obtained by using any of the twin structures alone. Similarly, if structure factors are averaged over a 5 psec simulation using a single structure and a constant temperature of 300 K, with similar weights on the X-ray term as for the B1 runs, the *R*-value is 13%. Higher *R*-values are obtained on increasing the temperature or on reducing the weight on the X-ray term. The rms fluctuation in atomic positions is very low in these restrained MD simulations (<0.1 Å for all atoms), showing that the X-ray term tightly restrains the structures to the initial value.

Diamond has demonstrated recently that the use of rigid-body motions and a very small number of low frequency normal modes in the refinement of isotropic temperature factors reduces drastically the number of parameters required to obtain satisfactory fit to the X-ray data.[49] Similarly, the very simple model that considers only the oscillations of rigid protein molecules[46,50] is successful in reproducing the isotropic temperature factor profiles of a wide range of proteins ranging in size from lysozyme to influenza virus hemagglutinin.[51] The segmented displacements that are systematically present in our twin refinements of ribonuclease and crambin sug-

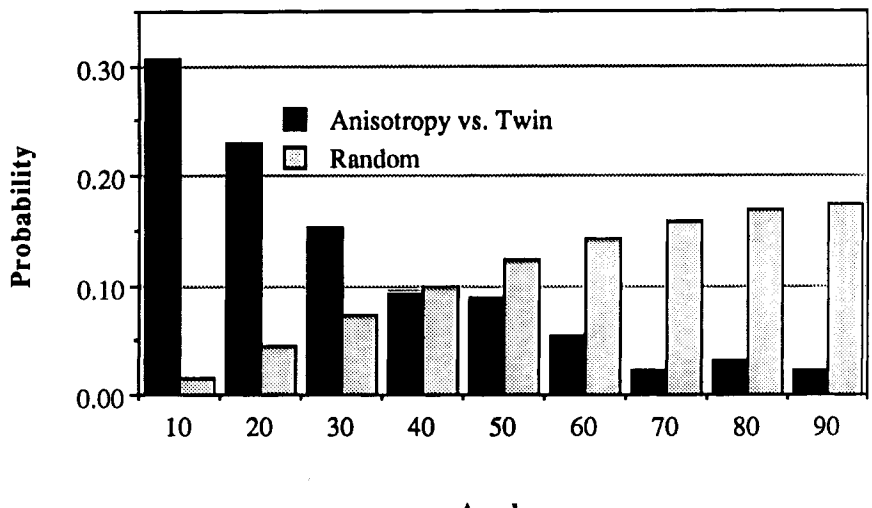

Fig. 8. Comparison between the twin refinement displacements in crambin with the principal axis of greatest displacement as obtained from the refinement of unconstrained anisotropic thermal parameters for each atom. The figure shows the normalized distribution of the angle between the direction of displacement in the twin structure and the principal axis of greatest displacement in the refined anisotropic temperature factors (dark bars). The sin(θ) distribution expected for the angles between randomly oriented vectors are also shown (light bars). Only the atoms not in disordered residues and having a ratio less than 0.75 between the smallest and the greatest axes of their anisotropic ellipsoid are used in the calculation.

gest that these simple librational models may be extended to treat the anisotropic displacements of small segments that extend over several residues.

The twin structure refinements also show that for many residues explicit account must be taken of distinct alternative conformations. Search procedures similar to those described here should prove useful in searching automatically for disordered regions in proteins. It appears that even though the molecular dynamics stage of the refinement has a larger radius of convergence than the conjugate-gradient minimizations, the latter suffices for characterizing most of the disorder observed for ribonuclease and crambin. This allows for rapid evaluation of the extent of disorder, and the specific regions identified by this process can then be modeled with alternative conformations in the conventional manner, as shown here for ribonuclease.

ACKNOWLEDGMENTS

Supported in part by a grant from the National Institutes of Health to J.K. (GM-43094) and from the National Science Foundation to W.A.H. (DMB 89-17570). J.K. is a Pew Scholar in the Biomedical Sciences. S.K.B. was supported by a Natural Sciences and Engineering Research Council of Canada Postdoctoral Fellowship. We thank Professor G.A. Petsko for helpful suggestions and encouragement.

REFERENCES

1. Ansari, A., Berendzen, J., Braunstein, D., Cowen, B.R., Frauenfelder, H., Hong, M.K., Iben, I.E.T., Johnson, J.B., Ormos, P., Sauke, T.B., Scholl, R., Schulte, A., Steinbach, P.J., Vittitow, J., Young, R.D. Rebinding and relaxation in the myoglobin pocket. Biophys. Chem. 26:337–355, 1987.
2. Brooks, C.L., Karplus, M., and Pettitt, B.M. "Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics." Adv. Chem. Phys. New York: Wiley, 1988.
3. Caspar, D.L.D., Clarage, J., Salunke, D.M., and Clarage, M. Liquid-like movements in crystalline insulin. Nature (London) 332:659–662, 1988.
4. Petsko, G.A., and Ringe, D. Fluctuations in protein structure from X-ray diffraction. Annu. Rev. Biophys. Bioeng. 13:331–371, 1984.
5. Hendrickson, W. A. Stereochemically restrained refinement of macromolecular structures. Methods Enzymol 115:252–270, 1985.
6. Sheriff, S., Hendrickson, W.A., Stenkamp, R.E., Sieker, L.C., Jensen, J.C. Influence of solvent accessibility and intermolecular contacts on atomic mobilities in hemerythrins. Proc. Natl. Acad. Sci. U.S.A. 82:1104–1107, 1985.
7. Jensen, L.H. Overview of refinement in maromolecular structure analysis. Methods Enzymol. 115:227–234, 1985.
8. Hendrickson, W.A. Diffraction studies of atomic mobility in proteins. Chem. Scripta 29A:109–112, 1989.
9. Wlodawer, A., Svensson, L.A., Sjölin, L., Gilliland, G.L. Structure of phosphate-free ribonuclease A refined at 1.26 Å. Biochemistry 27:2705–2717, 1988.
10. Blake, C.C.F., Pulford, W.C.A., Artymiuk, P.J. X-ray studies of water in crystals of lysozyme. J. Mol. Biol. 167:693–723, 1983.
11. Kuriyan, J., Petsko, G.A., Levy, R.M., Karplus, M. Effect of anisotropy and anharmonicity on protein crystallographic refinement. An evaluation by molecular dynamics. J. Mol. Biol. 190:227–254, 1986.
12. Smith, J.L., Hendrickson, W.A., Honzatko, R.B., Sheriff, S. Structural heterogeneity in protein crystals. Biochemistry 25:5018–5027, 1986.
13. Svensson, L.A., Sjolin, L., Gilliland, G.L., Finzel, B.C., Wlodawer, A. Multiple conformations of amino acid residues in ribonuclease A. Proteins: Struct. Funct. Genet. 1: 370–375, 1986.
14. Kuriyan, J., Wilz, S., Karplus, M., Petsko, G.A. X-ray structure and refinement of carbon-monoxy (Fe II)-myoglobin at 1.5Å resolution. J. Mol. Biol. 192:133–154, 1986.
15. Swanson, S.M. Effective resolution of macromolecular diffraction data. Acta Crystallogr., Sect. A 44:437–442, 1988.
16. Mao, B., Pear, M.R., McCammon, J.A. Anharmonic atomic displacements in cytochrome-c. Biopolymers 21:1979–1989, 1982.
17. Yu, H., Karplus, M., Hendrickson, W.A. Restraints in temperature factor refinement for macromolecules: An evalu-

ation by molecular dynamics. Acta Crystallogr., Sect. A 38:563–568, 1985.
18. Ichiye, T., Karplus, M. Anisotropy and anharmonicity of atomic fluctuations in proteins: Analysis of a molecular dynamics simulation. Proteins 2:236–259, 1987.
19. Ichiye, T., Karplus, M. Anisotropy and anharmonicity of atomic fluctuations in proteins: Implications for X-ray analysis. Biochemistry 27:3487–3497, 1988.
20. Brünger, A.T., Kuriyan, J., and Karplus, M. Crystallographic R-factor refinement by molecular dynamics. Science 235:458–460, 1987.
21. Doucet, J., Benoit, J.P. Molecular dynamics studied by analysis of the X-ray diffuse scattering from lysozyme crystals. Nature (London) 325:643–646, 1987.
22. Neinhaus, G.U., Heinzl, J., Huenges, E., Parak, F. Protein crystal dynamics studied by time-resolved analysis of X-ray diffuse scattering. Nature 338:665–666, 1989.
23. Borkakoti, N., Moss, D.S., Palmer, R.A. Ribonuclease-A: Least-squares refinement of the structure at 1.45Å resolution. Acta Crystallogr., Sect. B 38:2210–2217, 1982.
24. Howlin, B., Moss, D.S., Harris, G.W. Segmented anisotropic refinement of bovine ribonuclease A by the application of the rigid-body TLS model. Acta Crystallogr., Sect. A A45:851–861, 1989.
25. Haydock, K., Lim, C., Brünger, A.T., Karplus, M. Simulation analysis of structures on the reaction pathway of RNAse A. J. Am. Chem. Soc. 112:3826–3831, 1990.
26. Hendrickson, W.A., Teeter, M.M. Structure of the hydrophic protein crambin determined directly from the anomalous scattering of sulphur. Nature (London) 290: 107–113, 1981.
27. Douzou, P. "Cryobiochemistry: An Introduction." New York: Academic Press, 1977.
28. Wyckoff, H.W., Doscher, M., Tsernoglou, D., Inagami, T., Johnson, L.N., Hardman, K.D., Allewell, N.M., Kelly, D.M., and Richards, F.M. J. Mol. Biol. 27:563–578, 1967.
29. Phillips, S.E.V. Structure and refinement of oxymyoglobin at 1.6Å resolution. J. Mol. Biol. 141:441–484, 1980.
30. Honzatko, R.B., Hendrickson, W.A., Love, W.E. Refinement of a molecular model for lamprey hemoglobin from *petromyzon marinus*. J. Mol. Biol. 184:147–164, 1985.
31. Steigemann, W., Weber, E. Structure of erythrocruorin in different ligand states refined at 1.4 Å resolution. J. Mol. Biol. 127:309–338, 1979.
32. North, A.C.T., Phillips, D.C., Mathews, F.S. Acta Crystallogr., Sect. A 24:351–359.
33. Hanson, J.C., Watenpaugh, K.D., Sieker, L., and Jensen, L.H. A limited-range step-scan method for collecting X-ray diffraction data. Acta Crystallogr., Sect. A 35:616–621, 1979.
34. Konnert, J.H., and Hendrickson, W.A. A restrained parameter thermal factor refinement procedure. Acta Crystallogr. Sect. A 36(344–350), 1980.
35. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., Swaminathan, S., and Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. 4:187–212, 1983.
36. Brünger, A.T. Crystallographic refinement by simulated annealing: Application to a 2.8Å resolution structure of aspartate aminotransferase. J. Mol. Biol. 203:803–816, 1988.
37. Brünger, A.T. "X-PLOR (Version 1.5) Manual." 1988. The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University 260 Whitney Avenue, New Haven, CT 06511.
38. Kuriyan, J., Brünger, A.T., Karplus, M., Hendrickson, W.A. X-ray refinement of protein structures by simulated annealing: Test of the method on myohemerythrin. Acta Crystallogr. Sect. A 45:396–409, 1989.
39. Weis, W.I., Brünger, A.T., Skehel, J.J., Wiley, D.C. Refinement of the influenza virus hemagglutinin by simulated annealing. J. Mol. Biol. 212(4):737–761, 1990.
40. Hamilton, W.C. Significance tests on the crystallographic R factor. Acta Crystallogr. 18:502–510, 1965.
41. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., Haak, J.R. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81(3684), 1984.
42. Teeter, M.M., Hendrickson, W.A. Highly ordered crystals of the plant seed protein crambin. J. Mol. Biol. 127:219–233, 1979.
43. Brünger, A.T. A memory-efficient fast fourier transformation algorithm for crystallographic refinement on supercomputers. Acta Crystallogr., Sect. A 45:42–50, 1989.
44. Wlodawer, A., Borkakoti, N., Moss, D.S., Howlin, B. Comparison of two independently refined models of ribonuclease-A. Acta Crystalogr., Sect. B 42:379–387, 1986.
45. Willis, B.T.M., Pryor, A.W. "Thermal Vibrations in Crystallography." Cambridge, England: Cambridge University Press, 1975.
46. Schomaker, V., Trueblood, K.N. On the rigid-body motion of molecules in crystals. Acta Crystallogr., Sect. B. 24:63–76, 1968.
47. Keen, D.A., McGreevy, R.L. Structural modelling of glasses using reverse Monte Carlo simulation. Nature (London) 344:423–425, 1990.
48. Gros, P., van Gunsteren, W.G., Hol, W. Inclusion of thermal motion in crystallographic structures by restrained molecular dynamics. Science 249:1149–1152, 1990.
49. Diamond, R. On the use of normal modes in thermal parameter refinement: Theory and application to the bovine trypsin inhibitor. Acta Crystallogr. Sect. A 46:425–435, 1990.
50. Sternberg, M.J.E., Grace, D.E.P., Phillips, D.C. Dynamic information from crystallography. J. Mol. Biol. 130:231–253, 1979.
51. Kuriyan, J., Weis, W.I. Rigid protein motion as a model for crystallographic temperature factors. Proc. Natl. Acad. Sci. U.S.A., 88:2773–2777, 1991.
52. Badger, J., Caspar, D.L.D. Water structure in cubic insulin crystals. Proc. Natl. Acad. Sci. U.S.A. 88:622–626, 1991.