PROTEINS: Structure, Function, and Genetics 35:25–33 (1999)

New Features and Enhancements in the X-PLOR Computer Program

John Badger,* R. Ajay Kumar, Ping Yip, and Sándor Szalma
Molecular Simulations Inc., San Diego, California

ABSTRACT **This article describes new methods for X-ray crystallographic refinement and nuclear magnetic resonance (NMR) structure determination that are available in the recent release of the X-PLOR software, X-PLOR 98.0. The major new features of the X-PLOR 98.0 software are: (i) the introduction of maximum likelihood methods (Pannu and Read, Acta Crystallogr 1996;A52:659–668) for X-ray crystallographic refinement with structure factor amplitude, intensity and phase probability targets, (ii) the addition of the Andersen thermal coupling method for temperature control during simulated annealing refinements, (iii) a new utility function for converting reflection data in to the X-PLOR format, (iv) validated scripts and performance enhancements for structure determination from NMR distance restraints using torsion angle dynamics, (v) fast code for direct nuclear Oberhauser effect (NOE) refinement using matrix doubling and gaussian quadratures, (vi) methodologies for using ambiguous restraint information to perform automated iterative peak assignment and structure determination (Nilges et al., J Mol Biol 1997;269: 408–422). Additional developments in methodology for refining crystal structures from poor initial models include the implementation of a fast adaptive bulk solvent scattering correction and an energy minimization routine that makes use of second derivative information. Trial crystallographic refinements with an energy minimization protocol that includes these enhancements indicate significantly improved convergence. The quality of the resulting models appears comparable to models obtained from refinement protocols that incorporate torsion angle dynamics. Test applications of the new energy minimizer to NMR structure refinement with using NOE calculations also show improved convergence, leading to more optimized final models. Proteins 1999;35:25–33.**

Key words: crystallographic refinement; NMR structure determination

INTRODUCTION

The impetus for the development of the first version of the X-PLOR program was the realization that molecular dynamics methods for searching conformational space could be usefully applied to the crystallographic refinement problem.[1] This original program (built partially from codes from the CHARMM[2] program for molecular simulations) rapidly grew into a flexible system for carrying out crystallographic refinement, molecular replacement calculations and nuclear magnetic resonance (NMR) structure determination embedded within the infrastructure of a versatile system for macromolecular simulation and analysis. This version, X-PLOR 3.1, which is still widely used in many laboratories, is definitively described in the X-PLOR 3.1 manual.[3]

The next major release of the X-PLOR software, X-PLOR 3.851, included an algorithm for torsion angle dynamics.[4] Tests of this methodology showed that it could be used to increase the radius of convergence of X-ray crystallographic refinement[4] and could also provide an efficient mechanism for structure determination from (NMR) distance restraints.[5] Other major new features in X-PLOR 3.851 included technology for probabilistic MAD phase determination,[6] a bulk solvent model for X-ray crystallographic refinement,[7] direct NMR refinement against three-bond HN–CaH coupling constants,[8] proton chemical shifts[9] and secondary carbon chemical shifts.[10]

The aim of this article is to document the validation of new methods and applications for crystallographic refinement and NMR structure determination that are available in a new version of the X-PLOR program, X-PLOR 98.0. In addition, we describe some recent advances in structure determination methodology which further improve the convergence of structures determined from crystallographic or NMR data and which will be available in the next release of the X-PLOR software.

METHODS AND APPLICATIONS

Maximum Likelihood Refinement of Macromolecular Models

Background

Over the last few years, several workers have pointed out theoretical deficiencies in X-ray crystallographic refinement methods that employ the least-squares residual as the target function and have shown that better target functions can be rederived from first principles using

Dr. Kumar's present address is Computational Biology Center, IBM Thomas J. Watson Research Center, PO Box 704, Yorktown Heights, NY 10598.

Dr. Yip's present address is Sequenom Inc., 11555 Sorrento Valley Road, San Diego, CA 92121.

*Correspondence to: John Badger, Computational Biology Center, Building 0537, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093. E-mail: badger@sdsc.edu

Received 10 August 1998; Accepted 30 November 1998

maximum likelihood as a statistical basis.[11–13] Readers interested in the basic theory of the maximum likelihood approach are referred back to these articles. Superficially, the target functions employed in maximum likelihood refinement resemble the traditional least-squares residual used by most previous refinement programs. For example, the maximum likelihood target based on structure factor amplitudes, E_{ml}, is

$$E_{\mathrm{ml}} = \Sigma(1/\sigma^2_{\mathrm{ml}})(|F_{\mathrm{obs}}| - \langle|F_{\mathrm{obs}}|\rangle)^2 \qquad (1)$$

In this equation the maximum likelihood weighting factors $1/\sigma^2{}_{\mathrm{ml}}$ are calculated as a function of resolution using a cross-validation set of data. The $\langle|F_{\mathrm{obs}}|\rangle$ are the expectation values of the observed structure factor amplitudes, replacing the direct calculations of $|F_{\mathrm{calc}}|$ that would be used in a least-squares refinement.

The computer code for the maximum likelihood approach to crystallographic refinement described in ref .11 has been included in X-PLOR 98.0. This article describes the efficient calculation of maximum likelihood targets based on both structure factor amplitudes and structure factor intensities, and also shows that gradient-based energy minimizations using the maximum likelihood targets give much more convergent refinements than the conventional least-squares[11] method. A refinement protocol involving torsion angle dynamics with a maximum likelihood target has also been tested and shown to give a wider radius of convergence than other refinement methods.[14] More recently, a maximum likelihood target that makes use of phase probability information has also been devised,15 allowing full use of all experimental information in the structure refinement, and this target is also available in X-PLOR 98.0.

Test system

Our illustrative test system for validating the implementation of the maximum likelihood refinement method in X-PLOR 98.0 is a 51-amino-acid protein model of the cubic insulin crystal structure[16] from which multiple conformations (discrete disorder in several side chains) and solvent atoms have been removed. Atomic temperature factors were uniformly reset to 15 Å^2 and positional errors were introduced into the model by a short period of molecular dynamics at a temperature of 600 K in the absence of crystallographic data. The resulting model had the relatively large root-mean-square backbone error of 1.6 Å from the solved structure. The model was refined using a standard slow-cooling torsion angle dynamics protocol (Table I) with minor adaptation for use with the maximum likelihood target. The Engh-Huber parameter set[17] was used for the covalent parameters governing the stereochemistry of the protein model. Nonbonded interactions were treated using a short-range quartic potential with a weight of 100 kcal/mol for the torsion angle dynamics portion of the protocol which was then reduced to 16 kcal/mol. Approximately 10% of the data was reserved from cross-validation purposes and a set of 10 refinements was run for each refinement trial, each begun with a different set of random velocities and using a different set of cross-validation data. The initial R factor was approximately 0.52 for the working data between 10.0 and 3.0 Å resolution.

TABLE I. Protocol for Crystallographic Refinement Using Torsion Angle Dynamics[†]

Stage	Method
1	Restrained chemical energy minimization, 200 steps
2	Slow cooling torsion angle dynamics, 5000–300 K
3	Constant temperature molecular dynamics, 300 K
4	Powell energy minimization, 120 steps

[†] For refinements that used the maximum likelihood targets the estimates for σ_A were calculated after stage 1 and updated after stage 3.

We have also attempted refinement of this starting model by conjugate gradient minimization, without torsion angle dynamics. For these trials the Engh-Huber parameter set was used and the nonbonded interactions were treated by a short-range quartic potential with weight set to 100 kcal/mol. The minimization protocol consisted of 30 stages of 30 cycles using the Powell energy minimization routine, with the maximum likelihood parameters updated on each stage. The computer time needed for one refinement trial with this protocol is similar to the time required for one refinement trial using the torsion angle dynamics protocol. To obtain statistically useful indications of convergence these refinements were run 10 times for each target, using a different set of cross-validation data in each trial.

For all of these refinements the scale factor for the X-ray energy gradient was obtained using the standard heuristic procedure[1] in which the X-ray energy gradient is scaled to one-third of the chemical energy gradient obtained after a short period of unrestrained molecular dynamics.

Results

Compared to most other published tests of maximum likelihood refinements,[11–14] the data used in this test are limited to lower resolution and the errors in the initial model are relatively large. In fact, conventional refinements in which the crystallographic data are limited to 3 Å resolution are frequently problematic, exhibiting poor convergence (e.g., ref. 4), so this example represents a fairly stringent test of the application of the maximum likelihood methodology.

Table II compares the results from these refinements in terms of the final R values and R_{free} values. The intensity based maximum likelihood target gave final models that were the most accurate (lowest R_{free}), with an average value of R_{free} that is 0.025 lower than for the least-squares target. The intensity based maximum likelihood target also gave the greatest number of well-converged refinements. In addition, the degree of overfitting in the maximum likelihood refinements remained within an acceptable limit whereas the refinements that used the least-squares residual as the target contained a level of bias that might be considered unacceptably large. After incorporat-

TABLE II. Comparison of Torsion Angle Refinements of the Cubic Insulin Structure Using the Maximum Likelihood Targets and the Least-Squares Target[†]

Target	R	R_{free}	#<0.35	#>0.35, <0.40	#>0.40	Time
ML (F^2)	0.2774	0.3584	5	5	0	55.8
ML (F)	0.2905	0.3771	3	5	2	55.6
LS	0.2614	0.3831	2	4	4	53.8

[†] The values given for R and R_{free} are averages of 10 trials. The three columns labeled with the # symbol contain the number of refined structures for which R_{free} was less than 0.35, between 0.35 and 0.40 and greater than 0.40. The column labeled time corresponds to time in minutes per trial for computations on a Silicon Graphics R4400 with a 150-MHz processor.

ing improvements in the speed of the computer code for the intensity based maximum likelihood calculations (Pannu and Read, private communication) the average time required for refinements using this target was only 4% greater than for the refinements with the least-squares target.

When the intensity based maximum likelihood target was used in the refinement trials that used only energy minimization the average final value for R_{free} was 0.392. When the least-squares target was used, the average final value for R_{free} was 0.443. These numbers should be compared to the average values for R_{free} obtained using the torsion angle dynamics protocols, which were 0.358 with the maximum likelihood target and 0.383 with the least-squares target (Table II). These calculations demonstrate that a significant degree of refinement of this structure is possible without torsion angle dynamics when the maximum likelihood target is employed but that refinement of this structure using least-squares minimization is not very effective. The protocol that used torsion angle dynamics combined with the maximum likelihood intensity target gave a significantly better average solution than was achieved solely by maximum likelihood energy minimization. Furthermore, the multisolution strategy possible with molecular dynamics resulted in one model with R_{free} of 0.315, which was distinctly better than any model obtained by the other methods.

Andersen Thermal Coupling for Protein Model Refinement

Background

In the past, temperature control during simulated annealing calculations with the X-PLOR program was achieved using the Berendsen thermal coupling method.[18] In this approach, a frictional force is added to each atom that is proportional to the individual atomic velocity. The friction constant, b, is given by

$$b = b_{\text{initial}}(T_0/T - 1). \tag{2}$$

In this equation b_{initial} is a preset scale constant, T_0 is the desired temperature and T is the actual temperature of the system. Thus, the frictional forces increase with the difference between T_0 and T and vanish when T_0 is equal to T.

In X-PLOR 98.0 we have introduced the option of temperature control by the Andersen thermal coupling method.[19] In this method a set of atoms is randomly selected from the system on each time step. Random velocities, selected from a Boltzmann velocity distribution at the desired temperature, are then assigned to this set of atoms. The physical process that is modeled by the Andersen thermal coupling algorithm is the collision of a particle in the protein molecule with a particle in a heat bath at the desired temperature.

In contrast to the Berendsen thermal coupling method, which modifies the trajectories of all atoms in the system by adding forces along their current paths, the Andersen thermal coupling algorithm alters the velocities and directions of only the selected atoms at each time step. The general effect of the Andersen thermal coupling is to break up correlated motions involving groups of atoms. This behavior might be expected to increase local sampling of the conformational space but could inhibit more concerted changes in the protein model.

Test system

To test the usefulness of Andersen thermal coupling for crystallographic refinements we have carried out test refinements on trypsin,[20] cubic insulin,[16] and colicin[21] starting from models that are essentially correct but that are still incompletely refined. These refinements were then compared against a set of similar refinements that used the Berendsen thermal coupling method. The intensity based maximum likelihood target was used in all of these calculations. For each example, a set of 10 refinements were performed using a protocol consisting of slow cooling from 1,000 to 300 K using cartesian molecular dynamics, followed by 160 steps of energy minimization. Cooling was achieved by reducing the temperature of the heat bath by 25 K after every 50 steps of molecular dynamics. In the calculations using the Andersen thermal coupling method the collision frequency was set so, on average, each protein atom would have its velocity changed 2.5 times over each cooling stage. In the calculations using the Berendsen thermal coupling method the friction constant was set to 100.0. These parameters achieved the necessary degree of temperature control over each cooling stage for the Andersen and Berendsen thermal coupling methods.

Results

Our results (Table III) show that the refinements that used the Andersen thermal coupling method gave models in which R_{free} was, on average, 0.0006 to 0.007 lower than the parallel refinement runs using the Berendsen thermal coupling method. In addition, the differences between R and R_{free} in the runs that employed Andersen thermal coupling were smaller, indicating less biased models. The calculations that employed the Andersen thermal coupling method were typically about 10% faster than the calculations that used the Berendsen thermal coupling method.

TABLE III. Comparison of Simulated Annealing Refinements Using Andersen and Berendsen Thermal Coupling Methods[†]

Test	R(And.)	R_{free}(And.)	R(Ber.)	R_{free}(Ber.)
Trypsin, 3 Å	0.2576	0.3469	0.2550	0.3539
Trypsin, 2 Å	0.2776	0.3123	0.2748	0.3171
Insulin, 2.5 Å	0.2671	0.3050	0.2608	0.3056
Colicin, 2.5 Å	0.2169	0.2753	0.2158	0.2815

[†] The structure and the maximum resolution of the data used in the refinements are given. The refined values for R and R_{free} are averaged over 10 trials for each structure.

Torsion Angle Dynamics for Structure Determination With Distance Restraints

Background

The application of the torsion angle dynamics algorithm in X-PLOR to structure determination using distance restraint data derived from NMR measurements was described in a paper by Stein et al.[5] Torsion angle dynamics has also been implemented, using a different method for integrating the equations of motion, in the DYANA program.[23] These approaches are conceptually simple and appealing, since, aside from a limited use of van der Waals interactions to avoid very severe steric overlaps, the simulated annealing in torsion angle space occurs on an energy surface which is formed only from energy terms derived from the NMR data.

Since the scripts for NMR structure determination using torsion angle dynamics were not included in previous releases of the X-PLOR software we have developed a script based on the protocol described by Stein et al.[5] In addition, we have developed a new protocol for NMR structure determination that completely eliminates the use of conventional simulated annealing from the structure determination process (Table IV).

An analysis of the time spent on various parts of the calculations involved in NMR structure determination with torsion angle dynamics led to the (unexpected) conclusion that the integration of the equations of motion accounts for a large fraction of the total time used. For this reason this part of the code has been optimized in X-PLOR 98.0 so that the integration of the equations of motion is now approximately 40% faster than in X-PLOR 3.851 (Table V), typically leading to savings of approximately 20% in the total run time for these protocols.

Test system

To validate these protocols and confirm the correctness of the modified code we have carried out test calculations on a 120-amino-acid protein ,villin 14T[24] as well as an extremely large structure for NMR methods, human transcription factor tfIIβ,[25] which is 206 amino acids in size. In both cases the starting point for the structure determination process was a linear polypeptide chain. For both of these structures we ran sufficient trials to generate 50 correct solutions. The criteria used to define a correctly converged structure were the same as those used previously for calculations with the X-PLOR program.[5] We required that there should be no violations of NOE restraints greater than 0.5 Å, no dihedral angle violations greater than 5 degrees, a RMS deviation in bond lengths of less than 0.02 Å and an RMS deviation in bond angles of less than 2.0 degrees. We note that these criteria appear stricter than those used in calculations using the DYANA program[23] so a direct comparison of calculation times for structure determinations with X-PLOR and DYANA is not possible.

Results

For the villin 14T structure the two protocols that use torsion angle dynamics gave very high success rates, marginally better than previously reported[5] (Table VI). Clearly, the annealing time used in these protocols is more than sufficient to solve structures up to the size of the villin 14T and could probably be reduced for smaller protein molecules. However, initial trials indicated that the length of this protocol was insufficient to obtain adequate success rate for the very large tfIIβ molecule. Some experiments in which we have tried to use shorter 'cooking' (stage 1) and 'cooling' (stage 2) periods indicated that the length of the cooking period is the most critical for a successful structure determination. This finding was interpreted visually by examination of the folding trajectory using the InsightII/DeCipher software,[26] revealing incomplete folding during the 'cooking' stage of the protocols. For this reason we increased the length of this initial stage by a factor of three for structure determination trials with the tfIIβ molecule and this led to a high success rate of approximately 75% for these protocols (Table VI).

These results indicate that, provided the annealing time of the protocol is sufficiently long, even the structures of large protein molecules can be determined relatively quickly and automatically by mean of torsion angle dynamics.

Direct NOE Refinement

The idea and basic theoretical approach for direct refinement against NOE intensities have been known for some time but the computer intensive method previously used in the X-PLOR program[3] limited applications of the approach. We have implemented in X-PLOR 98.0 a fast and efficient method of calculating NOE intensities and their gradients. This new method is based on the matrix doubling property of the NOE data and an integral representation of the gradients of the NOE data.[27] Mathematical details of the calculation method are given in Appendix 1.

The implementation of this method has been tested by carrying out a direct NOE refinement on the five distance-refined structures for squash trypsin inhibitor[28] deposited in the Brookhaven Protein Data Bank (entry code 2CTI). The direct NOE refinement of these structures employed a protocol consisting of three 1,000 step stages of constant temperature (T = 300 K) cartesian molecular dynamics with the force constant for the NOE energy set to 100, 200, 400 kcal/mol, respectively, followed by energy minimization.[29]

A comparison of the new algorithm with the method previously available within X-PLOR shows that the new

TABLE IV. Torsion Angle Dynamics Protocol for Protein Structure Determination Using Distance Restraints Adapted From Stein et al.[5] (NMR Protocol 1) and a Protocol Which Excludes Use of Cartesian Molecular Dynamics (NMR Protocol 2)

NMR protocol 1

	Stage 1	Stage 2	Stage 3	Stage 4
	High temperature TAD	Cooling TAD	Cooling MD	1000 step Minimization
Temperature (K)	50,000	50,000–1,000	1,000–300	–
Time step (ps)	0.015	0.015	0.003	–
Period (ps)	15	15	6	–
W_{NOE}(kcal/mol)	150	150	150	50
W_{Dihe}(kcal/mol)	100	100	100	300
W_{VDW}(kcal/mol)	0.1	0.1–1.0	1.0	1.0

NMR protocol 2

	Stage 1	Stage 2	Stage 3
	High temperature TAD	Cooling TAD	2,000 steps Mimimization
Temperature (K)	50,000	50,000–0	–
Time step (ps)	0.015	0.015	–
Period (ps)	15	15	–
W_{NOE}(kcal/mol)	150	150	75
W_{Dihe}(kcal/mol)	200	200	300
W_{VDW}(kcal/mol)	0.1	0.1–1.0	1.0

TABLE V. Computer Time (in seconds) Spent per Integration Step for Torsion Angle Dynamics[†]

Model size (amino acids)	X-PLOR 3.851	X-PLOR 98.0
52	0.23	0.14
102	0.57	0.36
206	1.98	1.05

[†] Timings were carried out on Silicon Graphics R4400 with a 150-MHz processor.

TABLE VI. Success Rate (%) for Structure Determination of Villin 14T and Human Transcription Factor tfIIβ Using NMR Protocols 1 and 2[†]

	Villin 14T	tfIIβ
NMR protocol 1 (ref. 5)	84.7	—
NMR protocol 1	90.1	76.9
NMR protocol 2	87.7	74.6

[†] The length of the first stage for these protocols was tripled in order to give a good success rate for the calculations using human transcription factor tfIIβ.

method is typically an order of magnitude faster (Table VII). The new methodology makes direct NOE refinement a computationally tractable approach for structure refinement with NMR data.

FUTURE PROSPECTS

In this section we describe some results with new methods implemented in a developmental version of the X-PLOR software.

X-Ray Crystallographic Refinement

Bulk solvent scattering correction

It is generally considered that an appropriate correction for bulk solvent scattering, allowing inclusion of extremely low-resolution data without distorting the scale factor between the observed and calculated structure factors, should have a beneficial effect on crystallographic refinement. Although effective for many applications, use of the bulk solvent scattering correction available in X-PLOR 3.851[7] may be problematic in refinements where the structural change in the model is large since the protein/solvent boundary may also undergo a significant change. It has been argued that simpler but more rapidly computable and adaptive methods are more suitable for these types of refinement problems.[30] A practical example of this difficulty appears to be provided by the trial refinement of amylase inhibitor described in Adams et al.,[14] in which the low-resolution data were omitted because the changing solvent correction impaired convergence.

For this reason, the Babinet solvent scattering correction method has been implemented in the X-PLOR program. This scattering correction is based on the premise that solvent occupation of the crystal is the complement of the protein density. The Babinet solvent scattering correction provides corrected calculated structure factors, F_{corr} from the original calculated structure factor, F_{calc}, according to

$$F_{corr} = (1 - K_{sol} \cdot \exp(-B_{sol}/4d^2)) \cdot F_{calc} \tag{3}$$

where K_{sol} is a solvent scale factor which has a theoretical value equal to the average solvent electron density divided by the average protein electron density (typically in the

TABLE VII. Comparison of Refinement Statistics and Computer Time for Direct NOE Refinement Using the Method in X-PLOR 3.851 and the New Matrix Doubling Method Available in X-PLOR 98.0[†]

	X-PLOR 3.851					X-PLOR 98.0				
Structure	E_{init}	R_{init}	E_{fin}	R_{fin}	CPU	E_{init}	R_{init}	E_{fin}	R_{fin}	CPU
1	1,599	0.59	195	0.46	1,064	1,639	0.63	177	0.47	154
2	854	0.60	171	0.43	1,064	912	0.64	193	0.45	153
3	909	0.58	172	0.45	1,054	959	0.62	170	0.45	176
4	926	0.61	179	0.44	1,106	996	0.66	181	0.44	167
5	16,700	0.69	148	0.41	1,398	16,763	0.72	150	0.42	215

[†] Initial (init) and final (fin) energies and R values are given for five different starting models of squash trypsin inhibitor. The R value quoted is the unweighted $R^{1/6}$ value. The computer time (CPU) is given in seconds for calculations on a Silicon Graphics R10000 Origin with a 185-MHz processor.

range 0.75–1.0) and B_{sol} is a solvent smoothing factor which is set to a high enough value (typically 100–400Å^2) to eliminate structural detail. Although an analytic solution to the calculation of scale factors K_{sol} and B_{sol} that optimize the fit to the diffraction data is possible, this solution has not been implemented in X-PLOR. In general, we believe that the solvent scale factor is reliably set according to the theoretical expectation since with some models and datasets the refinement leads to unphysical values. However, an X-PLOR macro is available for trial and error optimization of the solvent scale factor. This form of solvent correction does not require explicit information on the protein/solvent boundary, as this is implicit in the use of F_{calc} and is extremely fast to compute. For these reasons the Babinet bulk solvent correction is implemented in most of the other major refinement programs.[12,31,32]

Improved energy minimization routine

Many users of the X-PLOR program have observed that the Powell[33] energy minimization routine that the program employs for gradient based minimization sometimes terminates when the energy gradient is significantly nonzero. Furthermore, cogent arguments and examples have been presented to show that simplistic energy minimization methods, which only consult first derivative information, give poorer convergence in crystallographic refinements than energy minimization methods which make some use of second derivative (or curvature) information.[34] For these reasons we have implemented an adopted basis Newton-Raphson energy minimization routine[2] (ABNR), which makes use of numerical approximations to second-derivative information, into our developmental version of the X-PLOR program.

Examples of crystallographic refinement

To investigate the refinement capabilities of the X-PLOR program when the intensity-based maximum likelihood target is used in conjunction with the Babinet bulk solvent correction and the ABNR energy minimization routine we have carried out several test calculations. The cubic insulin model with initial coordinate error of 1.6 Å that was used in previous calculations was re-refined using all the available low-resolution data (d_{min} = 32 Å) with bulk solvent scattering correction parameters K_{sol} = 0.75 and B_{sol} = 100 Å^2. As in the previous test refinements, the high resolution limit for the data was limited to 3.0 Å and the Engh-Huber parameters were used with a quartic nonbonded potential. The model was refined both using the standard torsion angle dynamics protocol (Table I) and by a minimization protocol consisting of 30 stages of 30 cycles of minimization with the maximum likelihood parameters updated on each stage. For each type of refinement, a series of 10 runs was performed with a different set of cross-validation data used in each run.

In order to test the relative performance of these methods with higher resolution data, the calculations were also repeated with diffraction data extending to 2.2 Å resolution. To probe the limits of convergence for these protocols, additional calculations were performed with the 3.0 Å resolution data using starting models in which the RMS backbone errors were increased to 1.8 Å and 2.0 Å by prolonging the period of free molecular dynamics.

Results

In contrast to the earlier refinements results given in this paper and results reported elsewhere,[4,14] where the torsion angle dynamics protocol gave significantly better results than protocols based solely on energy minimization, the refinements at 3.0 Å resolution that employed the ABNR energy minimization routine (without dynamics) gave slightly better results than the refinements that employed the torsion angle dynamics protocol (Table VIII). With data extending to 2.2 Å resolution the reverse is the case and the torsion angle dynamics protocol gave slightly better results than the protocol based on energy minimization. The computational cost of refinement using the ABNR energy minimization routine was only approximately 5% greater than the refinements with the Powell energy minimization routine and the overall computational cost of refinement using torsion angle dynamics is similar to the cost of the refinement based solely on gradient-based minimization.

In the calculations described above, almost all trials led to reasonably converged models. When these protocols were tried with 3 Å resolution data and the model in which the RMS backbone error was 2.0 Å, none of the trials using either the torsion angle dynamics protocol or the energy minimization protocol converged. When the RMS backbone error in the starting model was 1.8 Å the more

TABLE VIII. Results of Maximum Likelihood Crystallographic Refinement for Cubic Insulin Using All Data to the Resolution Limit and Incorporating the Babinet Bulk Solvent Correction†

	R	R_{free}	#<0.35	#>0.35, <0.40	#>0.40
ABNR min, 3.0 Å	0.257	0.348	6	4	0
Powell min, 3.0 Å	0.270	0.357	5	4	1
TAD (ABNR), 3.0 Å	0.279	0.365	2	7	1
TAD (Powell), 3.0 Å	0.280	0.364	3	6	1
ABNR min, 2.2 Å	0.299	0.371	0	10	0
Powell min, 2.2 Å	0.309	0.379	0	9	1
TAD (ABNR), 2.2 Å	0.305	0.370	2	7	1
TAD (Powell), 2.2 Å	0.307	0.372	2	6	2

†For both the minimization (min) trials and torsion angle dynamics (TAD) trials the ABNR and the Powell energy minimization routines were tested. For each type of refinement the results were averaged over 10 trials using a different set of cross-validation data. The three columns labeled with the # symbol contain the number of refined structures for which R_{free} was less than 0.35, between 0.35 and 0.40 and greater than 0.40.

stochastic searching available with torsion angle dynamics produced two solutions with $R_{\text{free}} < 0.35$, whereas the best model obtained by the ABNR energy minimization had R_{free} of 0.37. Thus, there may be a very narrow range of refinement problems where this multisolution torsion angle dynamics protocol is able to find a better refined structure than the protocol based purely on energy minimization.

These results imply that an effective gradient minimization procedure may, when coupled with the maximum likelihood target and an appropriate solvent scattering correction (leading to improving the modeling of the data), often be competitive in terms of radius of convergence with methods based on torsion angle dynamics. In their discussion of maximum likelihood methods for crystallographic refinement, Bricogne and Irwin[13] state this point of view quite forcefully "The increase in radius of convergence may rapidly overturn the present reliance on simulated annealing as a means of getting out of local least-squares minima." At present, the most convergent optimization method may be dependent on the specific refinement problem. Clearly more tests are required to properly evaluate the full capabilities of gradient-based energy minimization algorithms and compare these with dynamics-based approaches. As refinement technologies continue to develop a stronger preference for the best type of optimization method may also evolve.

NMR Refinement

Direct NOE refinement using the ABNR energy minimization routine

We have compared the results of the direct NOE refinement for squash trypsin inhibitor using both the ABNR and the Powell energy minimization routines. The starting models for these refinements were the same as those described in the section on direct NOE refinement with X-PLOR 98.0. The refinement protocol consisted of three stages of 2,000 steps of constant temperature ($T = 300$ K) cartesian molecular dynamics with a changing weight of 100, 200, and 400 kcal/mol, followed by 500 steps of energy minimization using either the ABNR or Powell method.

TABLE IX. Comparison of Final Energies, E_{fin} (kcal/mol), and R Values for Direct NOE Refinement†

	X-PLOR 3.851			X-PLOR (developmental)	
Structure	Steps	E_{fin}	R_{fin}	E_{fin}	R_{fin}
1	47	136	0.45	117	0.41
2	225	110	0.41	102	0.38
3	198	115	0.40	110	0.40
4	119	121	0.41	113	0.40
5	207	110	0.39	105	0.38

†The direct NOE algorithm and Powell energy minimization routine in X-PLOR 3.851 are compared with the new direct NOE refinement algorithm and ABNR energy minimization routine in the developmental version of X-PLOR. The R value quoted is the unweighted $R^{1/6}$ value. For both sets of refinements 500 minimization steps were requested to complete the protocol but with X-PLOR 3.851 the line search was abandoned at the step number indicated.

Table IX compares the final energies and R values for refinements using the ABNR and the Powell energy minimization routines. The refinements that used the ABNR minimization method led to more optimized structures since calculations using the Powell minimization method terminate prematurely.

SOFTWARE AVAILABILITY

X-PLOR 98.0 was released for Silicon Graphics computers in May 1998. Multiprocessor support is now available for Silicon Graphics computers and single processor versions for IBM and Compaq Alpha computers have been created. These platforms will also be supported for the developmental version of X-PLOR described in the text, which we expect to be released in January 1999 under the name X-PLOR 98.1. The X-PLOR source code (FORTRAN90 in the case of X-PLOR 98.1) is not normally supplied with these releases but is available upon request. Interested users should contact Molecular Simulations Inc. to obtain the most recent versions of the X-PLOR software.

ACKNOWLEDGMENTS

We appreciate the cooperation of Randy Read and Raj Pannu in providing the maximum likelihood computer code for X-ray crystallographic refinement and Michael Nilges for providing the code and methodology for the ARIA approach to NMR structure determination.

APPENDIX 1. Methodology for Direct NOE Refinement

Ping F. Yip

The present implementation of the direct NOE refinement is a modified and improved version of an algorithm that was first implemented in the Discover molecular dynamics/mechanics program.[35] We will briefly sketch the main ideas here. First, we discuss the calculation of NOE values using the matrix doubling approach. Denoting the NOE intensity at a mix time τ by **A** and the relaxation rate matrix by **R** we have

$$\mathbf{A}(\mathrm{t}) = \mathbf{exp}\ (-\mathbf{R}\tau). \qquad \text{(A1)}$$

Since **A** is an exponential function of **R**, we easily obtain the following 'doubling' property

$$\mathbf{A}(2\tau) = \mathbf{A}(\tau + \tau) = \mathbf{A}(\tau)\mathbf{A}(\tau). \qquad \text{(A2)}$$

Thus, by carrying out one matrix multiplication one can easily calculate NOE intensities at double the mixing time. One can recursively use equation A2 to the point where the NOE matrix at a long mixing time is an even power of the NOE matrix at a very small mixing time where its valid to use a linear approximation

$$\mathbf{A}(\mathrm{s}) = \mathbf{exp}\ (-\mathbf{R}\mathrm{s}) = 1 - \mathbf{R}\mathrm{s}, \qquad \text{for } \mathbf{R}\mathrm{s} \ll 1. \quad \text{(A3)}$$

For example, if we were interested in computing the NOE matrix at a mixing time of 128 ms, we can first calculate the NOE intensities at 4 ms using equation A3. By carrying out a series of five matrix multiplications we will then obtain the NOE intensities at mixing times 8, 16, 32, 64, and 128 ms, respectively. Furthermore, because of the rapid falloff ($\approx 1/r^6$) behavior of the rate matrix elements of *R*, one can safely adopt a distance cutoff (typically 7–8 Å) beyond which one can assume NOE intensities will be zero. Consequently, sparse matrix techniques may be employed to further expedite the matrix multiplications involved.

To obtain NOE intensities at any given time, *t*, along a buildup curve, we employ a simple fitting procedure. Suppose one is interested in buildup curves up to a mixing time of τ_m. We first determine the power, *n*, of 2 such that $\tau_m * 2^{-n}$ will be sufficiently small for the linear approximation to be valid. Then we compute the NOE intensities by sparse matrix multiplication of the NOE intensities at all the mixing times (set **T**): 2^{-n}, 2^{-n+1}, 2^{-n+2};, 2^{-1}, τ_m. We further compute the time derivatives of the NOE intensities at those times. This is easily accomplished by multiplying the rate matrix with the appropriate NOE intensities. With the NOE intensities and their time derivatives at all the mixing times in *T*, one can use a piecewise spline to interpolate between all the times in *T*. Thus, one has obtained complete buildup curves between 0 and τ_m.

The next main task in an efficient NOE refinement methodology is the calculation of the gradient of the NOE matrix with respect to atomic coordinates. Here we use the integral formula and gaussian quadrature techniques of Yip.[27] Briefly, we cast the gradient as an integral over time of a product of NOE intensities. This integral is further approximated by a summation of the products at specified gaussian points. Finally, by using the build-curves already calculated using the method outlined above, the desired NOE intensities may be read off at the gaussian points and the gradient calculated. We found that using a five-point gaussian quadrature is sufficiently accurate for gradient evaluation. For a system with a 5 ns correlation time the mixing time at which the linear approximation is used is 5 ms.

REFERENCES

1. Brünger AT, Kuriyan J, Karplus M. Crystallographic R factor refinement by molecular dynamics. Science 1987;235:458–460.
2. Brooks B, Bruccoleri R, Olafson B, States D, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and molecular dynamics calculations. J Comp Chem 1983;4:187–217.
3. Brünger AT. X-PLOR, Version 3.1: A system for x-ray crystallography and NMR. New Haven: Yale University Press, 1992.
4. Rice LM, Brünger AT. Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. Proteins 1994;19:277–290.
5. Stein EG, Rice LM, Brünger AT. Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. J Magn Reson 1997;124:154–164.
6. Burling FT, Weis WI, Flaherty KM, Brünger AT. Direct observation of protein solvation and discrete disorder with experimental phases. Science 1995;271:72–77.
7. Jiang J-A, Brünger AT. Protein hydration observed by x-ray diffraction:solvation properties of penicillopepsin and neuraminidase crystal structures. J Mol Biol 1994;243:100–115.
8. Garret DS, Kuszewski J, Hancock TJ, Lodi PJ, Vuister GW, Gronenborn AM, Clore GM. The impact of direct refinement against three-bond HN-CaH coupling constants on protein structure determination by NMR. J Magn Reson 1994;B104:99–103.
9. Kuszewski J, Gronenborn AM, Clore GM. The impact of direct refinement against proton chemical shifts in protein structure determination by NMR. J Magn Reson 1995;B107:293–297.
10. Kuszewski J, Qin J, Gronenborn AM, Clore GM. The impact of direct refinement against 13Cα and 13Cβ chemical shifts on protein structure determination by NMR. J Magn Reson 1995; B106:92–96.
11. Pannu NS, Read RJ. Improved structure refinement through maximum likelihood. Acta Crystallogr 1996;A52:659–668.
12. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr 1997;D53:240–255.
13. Bricogne G, Irwin JJ. maximum-likelihood refinement of incomplete models with BUSTER+TNT. In: Bailey S, Dodson EJ, editors. Macromolecular refinement. Proceedings of the CCP4 Study Weekend at Chester College, January 5–6. 1996:85–92.
14. Adams PD, Pannu NS , Read RJ, Brünger AT. Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. Proc Natl Acad Sci USA 1997;94:5018–5023.
15. Pannu NS, Murshudov GN, Dodson EJ, Read R. Incorporation of prior phase probability information strengthens maximum likelihood structural refinement. Acta Crystallogr 1997;D54:1285–1294.
16. Badger J, Harris MR, Reynolds CD, Evans AC, Dodson EJ, Dodson GG, North ACT. Structure of the pig insulin dimer in the cubic crystal. Acta Crystallogr 1991;B47:127–136.
17. Engh RA, Huber R. Accurate bond and angle parameters for x-ray protein structure refinement. Acta Crystallogr 1991;A47:392–400.
18. Berendsen HJC, Postma JPM, van Gunsteren NF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.
19. Andersen HC. Molecular dynamics simulations at constant pressure and/or temperature. J Chem Phys 1980;72:2384–2393.
20. Marquart M, Walter J, Deisenhofer J, Bode W, Huber R. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. Acta Crystallogr 1983;B39:480–.
21. Elkins P, Bunker A, Cramer WA, Stauffacher CV. A mechanism for toxin insertion into membranes is suggested by the crystal structure of the channel-forming domain of colicin E1. Structure 1997;5:443–458.
22. Hendrickson WA, Lattman EE. Representation of phase probability distributions for simplified combination of independent phase information. Acta Crystallogr 1970;B26:136–143.

23. Guntert P, Mumenthaler C, Wuthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 1997;273:283–298.
24. Markus MA, Matsudaira P, Wagner G. Refined structure of villin 14T and a detailed comparison with other actin-severing domains. Protein Sci 1997;6:1197–1209.
25. Bagby S, Kim S, Maldonado E, Tong KI, Reinberg D, Ikura A. Solution structure of the C-terminal domain of human tfIIβ: similarity to cyclin A an interaction with a tata-binding protein. Cell 1995;82:857–867.
26. DeCipher User Guide, San Diego: Molecular Simulations Inc, 1997.
27. Yip PF. A computationally efficient method for evaluating the gradient of 2D NOESY intensities. J Biomol NMR 1993;3:361–365.
28. Holak TA, Gondol D, Otlewski J, Wilusz T. Determination of the complete three-dimensional structure of the trypsin inhibitor from squash seeds in aqueous solution by nuclear magnetic resonance and a combination of distance geometry and dynamical simulated annealing. J Mol Biol 1989;210:635–648.
29. Nilges M, Habazettl J, Brünger AT, Holak TA. Relaxation matrix refinement of the solution structure of squash trypsin inhibitor. J Mol Biol 1991;219:499–510.
30. Badger J. Modeling and refinement of water molecules and disordered solvent. Methods Enzymol 1997;277B:344–352.
31. Tronrud DE. TNT refinement package. Methods Enzymol 1997; 277B:306–318.
32. Sheldrick GM, Schneider TR. SHELXL: high resolution refinement. Methods Enzymol 1997;277B:319–343.
33. Powell MJD. Restart procedures for the conjugate gradient method. Math Program 1977;12:241–254.
34. Tronrud DE. Conjugate-direction minimization: an improved method for the refinement of macromolecule. Acta Crystallogr 1992;A48:912–916.
35. NMRchitect User Guide. San Diego, CA: Molecular Simulations Inc., 1997.