Protein Science (1996), *5*:1067–1080. Cambridge University Press. Printed in the USA.

Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases

JOHN KUSZEWSKI, ANGELA M. GRONENBORN, AND G. MARIUS CLORE

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520

(RECEIVED January 26, 1996; ACCEPTED March 20, 1996)

Abstract

A new conformational database potential involving dihedral angle relationships in databases of high-resolution highly refined protein crystal structures is presented as a method for improving the quality of structures generated from NMR data. The rationale for this procedure is based on the observation that uncertainties in the description of the nonbonded contacts present a key limiting factor in the attainable accuracy of protein NMR structures and that the nonbonded interaction terms presently used have poor discriminatory power between high- and low-probability local conformations. The idea behind the conformational database potential is to restrict sampling during simulated annealing refinement to conformations that are likely to be energetically possible by effectively limiting the choices of dihedral angles to those that are known to be physically realizable. In this manner, the variability in the structures produced by this method is primarily a function of the experimental restraints, rather than an artifact of a poor nonbonded interaction model. We tested this approach with the experimental NMR data (comprising an average of about 30 restraints per residue and consisting of interproton distances, torsion angles, $^3J_{HN\alpha}$ coupling constants, and ^{13}C chemical shifts) used previously to calculate the solution structure of reduced human thioredoxin (Qin J, Clore GM, Gronenborn AM, 1994, *Structure 2*:503–522). Incorporation of the conformational database potential into the target function used for refinement (which also includes terms for the experimental restraints, covalent geometry, and nonbonded interactions in the form of either a repulsive, repulsive-attractive, or 6-12 Lennard-Jones potential) results in a significant improvement in various quantitative measures of quality (Ramachandran plot, side-chain torsion angles, overall packing). This is achieved without compromising the agreement with the experimental restraints and the deviations from idealized covalent geometry that remain within experimental error, and the agreement between calculated and observed ^1H chemical shifts that provides an independent NMR parameter of accuracy. The method is equally applicable to crystallographic refinement, and should be particular useful during the early stages of either an NMR or crystallographic structure determination and in cases where relatively few experimental restraints can be derived from the measured data (due, for example, to broad lines in the NMR spectra or to poorly diffracting crystals).

Keywords: conformational databases; NMR; protein structure determination; refinement; X-ray

The determination and refinement of a protein structure by NMR or crystallography depends on the minimization of a target function (E_{tot}) comprising the experimental restraints (E_{exp}), and two groups of a priori restraints consisting of covalent geometry (E_{cov}) and nonbonded interactions (E_{nb}) (Jack & Levitt, 1978; Konnert & Hendrickson, 1980; Clore et al., 1985; Hendrickson, 1985; Braun, 1987; Brünger et al., 1987; Clore & Gronenborn, 1989; Havel, 1991; Brünger & Nilges, 1993):

$$E_{tot} = E_{exp} + E_{cov} + E_{nb}. \quad (1)$$

In the case of NMR, the experimental restraints principally comprise NOE-derived interproton distances (or NOE intensities) that may be supplemented by torsion angles, coupling constants and chemical shifts. For crystallography, the experimental terms are the observed structure factor amplitudes. For any structure to be acceptable, it should exhibit good agreement with the experimental restraints, while at the same time displaying very small

Reprint requests to: G.M. Clore and A.M. Gronenborn, Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520; e-mail: clore@vger.niddk.nih.gov and gronenborn@vger.niddk.nih.gov.

deviations from idealized covalent geometry and good nonbonded contacts. The absolute requirement for restraints relating to the covalent geometry and nonbonded interactions is due to the fact that the ratio of the number of experimental observables to degrees of freedom is less than 1 (i.e., the problem is ill-determined). This is in distinct contrast to the situation in small molecule crystallography where it is possible to measure a sufficient number of observables with which to determine the structure without the help of a priori restraints.

Although the various terms describing the covalent geometry (i.e., bond lengths, bond angles, planes, and chirality) are known to very high accuracy (Engh & Huber, 1991), there is a much larger degree of uncertainty in the description and parameterization of the nonbonded interactions. Empirical nonbonded potentials used in energy minimization and molecular dynamics simulations generally comprise terms for van der Waals, electrostatic, hydrogen bonding, and solvation interactions (Momany et al., 1974; Brooks et al., 1983; Weiner et al., 1986; Jorgensen & Tirado-Rives, 1988; Halgren, 1992; van Gunsteren et al., 1994; Cornell et al., 1995). In both NMR and X-ray structure determinations, it is also possible to represent the nonbonding interactions by a minimalist repulsive van der Waals potential designed to exclude approximately the same regions of space as those excluded by the Lennard-Jones van der Waals potential (Konnert & Hendrickson, 1980; Hendrickson, 1985; Nilges et al., 1988a). The limitations of the full empirical nonbonded interaction potentials with regard to proteins can be ascertained from the observation that, even in the most carefully performed molecular dynamics simulations, the coordinates for the backbone atoms drift by ~1.5 Å from the starting X-ray coordinates (Loncharich & Brooks, 1990; Chandrasekhar et al., 1992; Brunne et al., 1993; Eriksson et al., 1995). Moreover, both energy minimization and molecular dynamics (including solvent) starting from X-ray coordinates reduce the agreement between calculated and experimentally measured values of $^3J_{HN\alpha}$ coupling constants obtained by solution NMR (Kay et al., 1989).

Although the description of the nonbonded contacts may not be that critical for the accuracy of high-resolution (2 Å or better) X-ray structures, it does play a significant role in determining the limits of accuracy attainable in an NMR structure determination (Clore et al., 1993; Gronenborn & Clore, 1995). Thus, for example, the average pairwise backbone RMS difference between the three independently solved crystal structures of interleukin-1β at 2 Å resolution (R-factors less than 20%) using different refinement programs with different representations for the nonbonded contacts is only 0.3 Å (Clore & Gronenborn, 1991). In contrast, a very small change in the van der Waals radii used to calculate the NMR structure of the oligomerization domain of p53 with identical experimental restraints resulted in an atomic RMS shift of 0.4 Å on a background of a precision of 0.3 Å (Clore et al., 1995). There are several factors responsible for these observations. First, the number of restraints (experimental plus a priori) per degree of freedom is generally smaller for NMR relative to X-ray structure determinations. Second, the nature of the two techniques is very different. X-ray crystallography is, in effect, an imaging technique because the Fourier transform of the observed structure factor amplitudes (providing their phases are known) yields an electron density map of the molecule under consideration, thereby permitting one to model-build the structure directly in real space and refine it in reciprocal space. The structural information provided by NMR, on the other hand, resides principally in pairwise distance interactions in n-dimensional distance space so that a simple mathematical relationship between the experimental restraints and cartesian coordinates is not available. Indeed, model calculations indicate that, in the case of protein structures determined by NMR, small variations in the values of the van der Waals radii employed for the various atoms can introduce coordinate shifts of the order of 0.2–0.3 Å (Clore et al., 1993).

Another important consideration with regard to currently employed nonbonding potentials is their lack of converging and discriminating power within the context of structure determination calculations, such as high-temperature simulated annealing. For example, it is well known that only certain combinations of backbone ϕ,ψ dihedral angles are populated in proteins (see, for example, Morris et al., 1992), and many years ago, Ramachandran et al. (1966) provided a rationale for this observation by calculating the Lennard-Jones energy of model peptides as a function of ϕ and ψ. Yet, good Ramachandran plots are generally only found in high-resolution NMR structures based on an average of at least 15 experimental restraints per residue and incorporating extensive stereospecific assignments (Clore & Gronenborn, 1991). Typically, lower-resolution NMR structures exhibit large numbers of residues in regions of the Ramachandran plot that are unpopulated in very high-resolution X-ray structures (Morris et al., 1992), indicating that the discriminating power of the nonbonded potential is insufficient to exclude structures that have energetically highly unlikely combinations of ϕ,ψ angles. Indeed, calculations aimed at examining the sampling of various structure determination protocols in the absence of any experimental restraints have shown that the repulsive van der Waals potential only excludes two regions of ϕ,ψ space from $\phi = -30°$ to $+30°$ and $\phi = 90$–$150°$, and that the distribution of ϕ,ψ angles within the remaining regions is essentially uniform (Kuszewski et al., 1992). We have obtained very similar results using a Lennard-Jones potential (J. Kuszewski & G.M. Clore, unpubl. data). Comparable observations also hold for the side-chain conformations where the nonbonded potentials do not discriminate against skewed χ_1 rotamers, despite the fact that it is known that >90% of side-chain χ_1 angles in high-resolution crystal structures are within 15° of the ideal conformations (60°, −60°, or 180°) (Nilges et al., 1990; Morris et al., 1992; Kleywegt & Jones, 1996).

The deficiencies in the current nonbonded potentials within the context of experimental structure determination can probably be attributed to two factors. First, the differences in nonbonded energy within the allowed regions are small relative to the total energy of the system, owing to the presence of multiple minima within the global minimum region. Second, and perhaps more importantly, is that the nonbonded potentials do not act directly on rotatable bonds and consequently lack explicit discriminatory power *vis a vis* the backbone and side-chain torsion angles.

The agreement between the observed and expected distribution of backbone ϕ,ψ angles represents the underlying concept of a number of recent approaches aimed at checking the validity of experimentally determined protein structures. These typically rely on large databases of protein structures that have been solved crystallographically to high resolution (2 Å or better) with R-factors less than 20%, and are therefore believed to be accurate. PROCHECK (Laskowski et al., 1993) is a computer program embodying one such attempt at "quality control." Using

its database of 163 high-resolution nonhomologous X-ray structures, it examines, among other parameters, the dihedral angles in a protein to find any angles in the model that are uncommon, and therefore suspect. Dunbrack and Karplus (1993, 1994) and Vriend and Sander (1993) have examined similar databases of structures to find correlations between backbone and side-chain dihedral angles.

In this paper, we define a new conformational database potential term, based on the relative populations of various combinations of dihedral angles observed in databases of high-resolution X-ray structures, as a means of improving the quality of structures determined by NMR and crystallography. We implement this new potential term in the simulated annealing refinement program XPLOR (Brünger, 1992a) and apply it to the refinement of the NMR structure of reduced human thioredoxin, a protein of 105 amino acids, the structure of which has been determined previously to very high precision by NMR spectroscopy (Qin et al., 1994). Although the new conformational database energy term works directly only on certain rotatable bonds in the protein, it improves the overall packing of the structures, and hence their quality, without violating the experimental restraints. The accuracy of the resulting structures are further verified by cross-validation of the observed and calculated ^{1}H chemical shifts. Given our present knowledge of macromolecular structure afforded by the availability of a very large number of high-resolution protein X-ray structures, it seems only reasonable that the incorporation of such quality control directly into the refinement procedure should be an integral part of any NMR structure determination, no different from the incorporation of the standard a priori restraints relating to covalent geometry and nonbonded interactions.

Theory and implementation

The conformational database potential

The derivation of the conformational database potential involves using the PROCHECK database of high-resolution X-ray structures (Morris et al., 1992; Laskowski et al., 1993) to create one- or two-dimensional matrices of energy values at evenly spaced points along axes that correspond to the various types of dihedral angles found within proteins. The method we used to process the database into energy values is similar to that used by PROCHECK itself for processing it into probability values. For example, in determining the energy of backbone ϕ and ψ angles for all residues (excluding Pro and Gly), we divide ϕ,ψ space into 2,025 $8° \times 8°$ bins. We then examine the PROCHECK database of crystal structures and determine the number of examples of residues whose ϕ and ψ values are within each bin of ϕ,ψ space. The fractional probability P for a residue to appear within each bin of ϕ,ψ space is simply obtained by dividing the number of examples in a given bin by the total number of examples in the database. To avoid instabilities for bins that have only a small number of examples, we require that the number of examples be greater than a minimum cutoff value (10, in this work). If the number of examples is less than the set cutoff value, we add the examples seen in its neighboring bin, starting from the bin's closest neighbors and working outward along each dimension of the grid, until we have added in enough examples to exceed the cutoff value (see Fig. 1). We then divide this new number of examples by the number of bins whose data were included, giving a local average number of examples for those bins that contain very few data points.

From statistical thermodynamics, these probabilities, P_i, are converted into a potential of mean force using the equation:

$$E_{DB}(i) = -k_{DB}(\log P_i), \tag{2}$$

where k_{DB} is a scale factor.

In this manner, we transformed the PROCHECK database into a series of energy values. This provided one-dimensional energy grids (8° per grid) for the χ_1 angles of Arg, Asn, Asp, Cys, Gln, Glu, His, Ile, Leu, Lys, Met, Phe, Ser, Thr, Trp, Tyr, and Val; the χ_3 angles of Arg, Cys (in a disulfide bridge), Gln, Glu, Lys, and Met; and the χ_4 angles of Arg and Lys. This also provided two-dimensional energy grids (8° by 8° per grid square) of ϕ versus ψ for Pro, Gly, and all other residues; and of χ_1 versus χ_2 for Arg, Asn, Asp, Cys (in a disulfide bridge), Gln, Glu, His, Ile, Leu, Lys, Met, Phe, Trp, and Tyr. Although PROCHECK has data on backbone ω angles as well, we did not make use of them because these are already restrained to be *trans* (or *cis* where appropriate) by the improper torsion angle terms included in the covalent geometry restraints.

The backbone-dependent rotamer database of Dunbrack and Karplus (1993, 1994) was processed in an identical manner. Although it provides data relating the number of examples in its database with the values of ϕ, ψ, χ_1, and χ_2 in 10° hypercubes, we only used the first three dimensions in the present work in

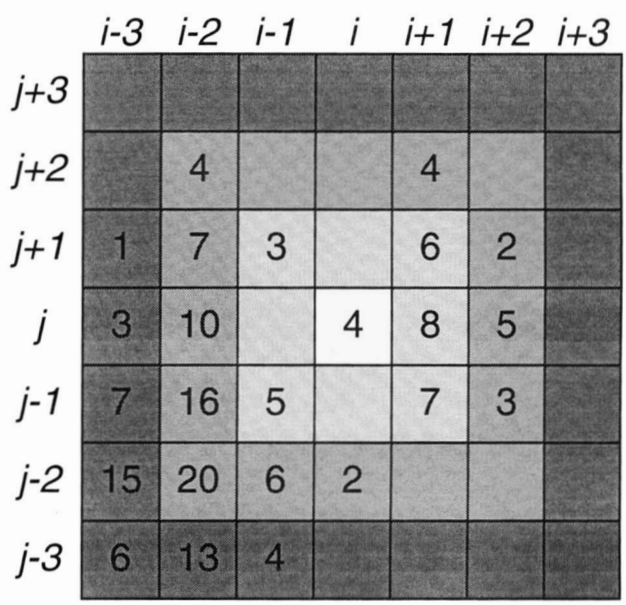

Fig. 1. An illustration of the method used to generate probability scores for each grid square of the structure database. To calculate a probability score for the square at (i, j), the number of examples in the database that are seen in that square, four in this particular example, is compared with the minimum cutoff value *mc* (equal to 10 in this work). If it is less than *mc*, then the examples from neighboring grid squares (light gray) are added to the number from (i, j). This process continues, adding the values from grid squares in increasingly large shells around (i, j), until the total number of examples is greater than or equal to *mc*. At this point, the probability is calculated as described in the text.

order to keep the number of examples in each grid cube reasonably high.

The application of the conformational database potential energy to molecular dynamics requires knowledge of its partial derivatives (i.e., the forces) with respect to the atomic coordinates. Because the conformational database energy is not a continuous function, but rather is known in discrete blocks, these partial derivatives were approximated in a manner analogous to that employed in our previous ^{13}C chemical shift potential term (Kuszewski et al., 1995b). To this end, the energy for every rotatable bond (or set of rotatable bonds) being refined against the conformational database potential, was defined by looking up the value in the grid bin that encompasses the current dihedral angle(s), and the partial derivatives of the energy with respect to the rotatable bond angles were then approximated by the local slope of the energy function, defined by

$$\partial E_{DB}(\phi_i)/\partial\phi \approx -k_{DB}[E_{DB}(\phi_{i-1}) - E_{DB}(\phi_{i+1})]/2, \quad (3)$$

where $E_{DB}(\phi_i)$ is the database energy of bin i along the rotatable bond ϕ_i, and $E_{DB}(\phi_{i-1})$ and $E_{DB}(\phi_{i+1})$ are the database energies of the bins that precede and follow the bin that contains the actual energy value.

The nonbonded van der Waals potential

Because the conformational database energy E_{DB} is only a function of local interactions (i.e., the positions of atoms connected by less than three rotatable bonds), it is an adjunct to the standard nonbonded contact energy terms, not a replacement for them. In this work, we have therefore also evaluated three different models for nonbonded interactions: the standard quartic van der Waals repulsion term used in XPLOR (Nilges et al., 1988a), the CHARMM (Brooks et al., 1983) 6-12 Lennard-Jones potential, and a new van der Waals term, which we call the "attractive-repulsive" potential.

The quartic van der Waals repulsive potential is of the form (Nilges et al., 1988a):

$$\begin{aligned} E_{rep}[R, R_{eff}, k_{rep}] &= k_{rep}[R_{eff}^2 - R^2]^2, && \text{if } R < R_{eff} \\ &= 0, && \text{if } R \geq R_{eff}, \end{aligned} \quad (4)$$

where R is the actual distance (in Å) between the centers of two atoms; R_{eff}, the sum of their effective hard sphere radii, generally given by 0.8 times the value of R_{min}, which is the distance between the two atoms corresponding to the minimum of the Lennard-Jones potential (using the PARAM19/20 CHARMM energy parameters; Brooks et al., 1983; Reiher, 1985); and k_{rep}, a force constant (in kcal·mol^{-1}·Å^{-4}). The 6-12 Lennard-Jones potential is defined as:

$$E_{LJ}(R) = 4\epsilon[(R_{min}\cdot 2^{-1/6}/R)^{12} - (R_{min}\cdot 2^{-1/6}/R)^6], \quad (5)$$

where ϵ is the depth of the minimum of E_{LJ} (and depends on the atom pairs involved).

The "attractive-repulsive" term $E_{att\text{-}rep}$ is designed to combine the attractive well of the Lennard-Jones potential with the overall behavior, computational efficiency and flexibility of the repulsive potential, and is defined as follows:

$$\begin{aligned} E_{att\text{-}rep}(R, R_{eff}, R_{att}) &= k_{rep}(R_{eff}^2 - R^2)^2, && \text{if } R < R_{eff} \\ &= \frac{-\epsilon[-(R - R_{min})^2 + R_{min}^2(1 - R_{eff}/R_{min})^2]^2}{R_{min}^4(1 - R_{eff}/R_{min})^4}, \\ & && \text{if } R_{eff} \leq R \leq R_{min} \\ &= \frac{-\epsilon(R_{att} - R)^2(R + R_{att} - 2R_{min})^2}{(R_{att} - R_{min})^4}, \\ & && \text{if } R_{min} \leq R < R_{att} \\ &= 0, && \text{if } R_{att} \leq R, \end{aligned} \quad (6)$$

where R_{att} is the distance at which the two atoms begin to attract each other (the other terms have the same meaning as in Equations 4 and 5).

The behavior of these three nonbonded potentials as two aliphatic hydrogen atoms approach each other is illustrated in Figure 2. As can be seen in panel A, the overall behavior of the Lennard-Jones potential (dark gray line) is quite different from that of the repulsive or attractive-repulsive terms (dotted light gray and solid black lines, respectively), which are completely superimposed in this overall view. It is clear that atoms connected with the repulsive or attractive-repulsive potentials can sometimes move through each other, because the maximal repulsive energy is finite and small in relation to the energy of two partially overlapped Lennard-Jones atoms. This property of the repulsive (and attractive-repulsive) potential is used in the simulated annealing schedules that are commonly used in NMR structure determination to enhance their conformational sampling properties (Nilges et al., 1998a, 1988b, 1988c). Panel B shows a detail of the region close to the attractive well of the Lennard-Jones potential. In this view, the repulsive potential is identical to the attractive-repulsive potential, until R becomes greater than R_{eff}. At this point, the repulsive potential becomes zero, but the attractive-repulsive potential enters an energy well of the same depth as that of the Lennard-Jones well. All of these functions become zero at internuclear distances greater than 4.5 Å. Thus, the attractive-repulsive potential combines the overall behavior of the repulsive potential with an attractive energy well that is similar to that of the Lennard-Jones potential.

Calculational strategy

All calculations in this work were performed on the reduced form of human thioredoxin, a protein of 105 residues, the structure of which has been determined to very high precision by NMR spectroscopy on the basis of an average of about 30 NMR derived experimental restraints per residue (Qin et al., 1994), including direct refinement against $^3J_{HN\alpha}$ coupling constants (Garrett et al., 1994), $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts (Kuszewski et al., 1995b), and ^1H chemical shifts (Kuszewski et al., 1995a).

Six groups of structures were calculated using a standard slow-cooling simulated annealing protocol (Nilges et al., 1988a) using the program XPLOR (Brünger, 1992a). All structures were calculated using standard covalent geometry restraints (bond lengths, bond angles, and improper torsions) and the previously reported terms (Qin et al., 1994; Kuszewski et al., 1995b) for the 3,128 experimental NMR restraints (comprising 2,571 approx-

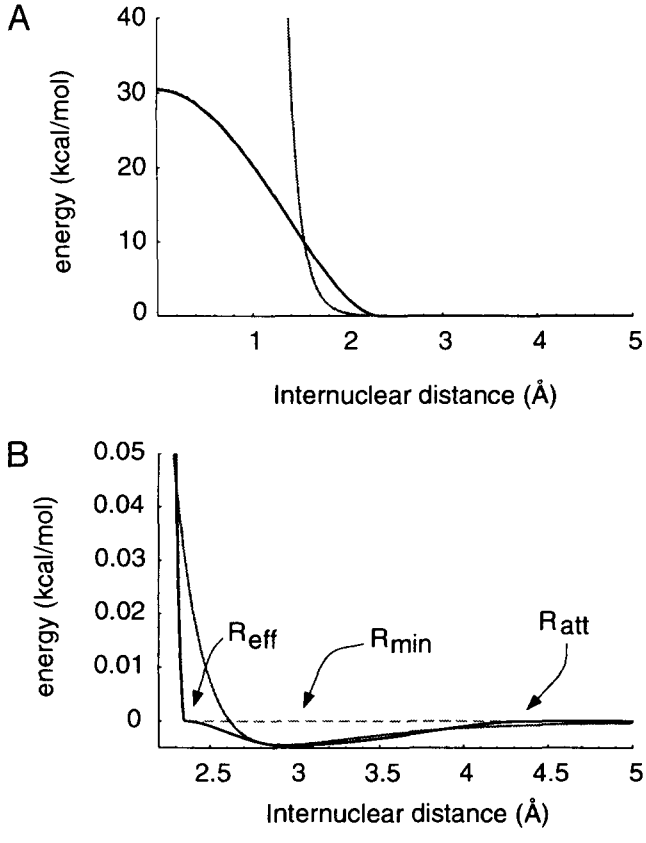

Fig. 2. A comparison of the behavior of the Lennard-Jones (dark gray line), repulsive (dotted light gray line) and attractive-repulsive (black line) van der Waals potentials for two interacting aliphatic hydrogen atoms as a function of internuclear distance. R_{min} is 2.92 Å, R_{eff} and R_{att} are set to 0.8 times (2.34 Å; Nilges et al., 1988a) and 1.5 times (4.4 Å) that value, respectively. In panel A, the overall behavior of the three potentials are compared. The repulsive and attractive-repulsive potentials are nearly identical in this region and are thus superimposed. Note that the Lennard-Jones potential tends to infinity as the internuclear distance tends to zero. Panel B shows a closeup of the attractive region of the Lennard-Jones potential, together with the corresponding regions of the repulsive and attractive-repulsive potentials. Once again, the repulsive and attractive-repulsive potentials are identical at $R < R_{eff}$, which is 2.34 Å in this case. The position and depth of the Lennard-Jones and attractive-repulsive wells are the same, but their shape is somewhat different in the region from R_{eff} to R_{min}.

imate interproton distance restraints, 273 torsion angle restraints, 89 $^3J_{HN\alpha}$ coupling constants, and 100 $^{13}C\alpha$ and 95 $^{13}C\beta$ chemical shifts). No ^1H chemical shift restraints, however, were employed in the present calculations. Thirty structures each were calculated using either the repulsive, attractive-repulsive, or Lennard-Jones van der Waals terms with and without the conformational database potential term. Because of the nature of the annealing schedule employed, the Lennard-Jones energy could not be applied during the high-temperature dynamics or slow-cooling phases of the protocol. Instead, the repulsive van der Waals term was used during these stages, with the Lennard-Jones term replacing it during a long minimization at the end of the refinement. Because the overall behavior of the attractive-repulsive potential is very similar to that of the purely repulsive one, we applied the attractive-repulsive potential during the entire annealing schedule. The conformational database energy was never used without either the repulsive, attractive-repulsive, or Lennard-Jones van der Waals potentials. This is because the conformational database energy only applies to a fraction of the total possible nonbonded interactions in the molecule.

The force constants for the various terms in the target function were chosen so as to ensure that the experimental restraints were satisfied within their errors, while at the same time maintaining very small deviations from idealized covalent geometry and good nonbonded contacts. The final values of the force constants are as follows: 500 kcal·mol^{-1}·Å^{-2} for bond lengths, 500 kcal·mol^{-1}·rad^{-2} for angles and improper torsions, 30 cal·mol^{-1}·Å^{-2} for NOE-derived interproton distance restraints, 200 kcal·mol^{-1}·rad^{-2} for experimental torsion angle restraints, 1 kcal·mol^{-1}·Hz^{-2} for coupling constants, 0.5 kcal·mol^{-1}·ppm^{-2} for ^{13}C chemical shifts, 4 kcal·mol^{-1}·Å^{-4} for the repulsive and repulsive-attractive nonbonded term. The values of ϵ, the depth parameter for the Lennard-Jones and repulsive-attractive van der Waals potentials employed are those of the PARAM19/20 CHARMM energy parameters (Brooks et al., 1983; Reiher, 1985). The values of R_{eff} and R_{att} are set to 0.8 and 1.4 times the value of R_{min}, respectively. The scale factor k_{DB} for the conformational database term was set to 3 for most calculations. In addition, a series of calculations were carried out with k_{DB} values between 0 and 10 to examine the impact of different weightings of the conformational database term.

The quality of the structures generated by these various protocols was examined using several methods. The agreement with expected dihedral angle values, the quality of hydrogen-bonding interactions, the overall quality of the packing, and the number of bad nonbonded contacts were quantified using the programs PROCHECK (Laskowski et al., 1993) and WHAT IF (Vriend & Sander, 1993). The quality control provided by WHAT IF evaluates the packing of structures by examining the distributions of atoms around various residue fragments in the molecule and comparing those distributions to those expected from a database of high-resolution X-ray structures.

Results and discussion

Behavior of the conformational database energy function

Three examples of one-dimensional conformational database energy terms are shown in Figure 3, namely, for χ_1 of Phe, χ_4 of Arg, and χ_3 of Gln. For comparison, the Lennard-Jones and repulsive van der Waals energies of isolated Phe, Arg, and Gln residues are also shown as functions of these torsion angles. These three types of rotatable bonds were chosen to show the various types of angular dependencies observed in the database. The χ_1 angle of Phe has three distinct, equally populated rotamers at $\chi_1 = 60°$, 180°, and 300°; the χ_4 angle of Arg has one region at $\chi_4 = 0°$ that is completely unpopulated, but has other regions at $\chi_4 = 120°$ and 240° that are slightly less populated than the remaining regions; and the χ_3 angle of Gln is unrestricted in its rotation by Lennard-Jones contacts and is populated everywhere equally. It is immediately apparent from the data shown in Figure 3 that the Lennard-Jones, repulsive van der Waals, and conformational database energy functions are consistent with each other, because the locations of energetically favorable and unfavorable regions are the same. It is also apparent, however, that the magnitudes of the peaks and troughs of the conformational database energy function are different

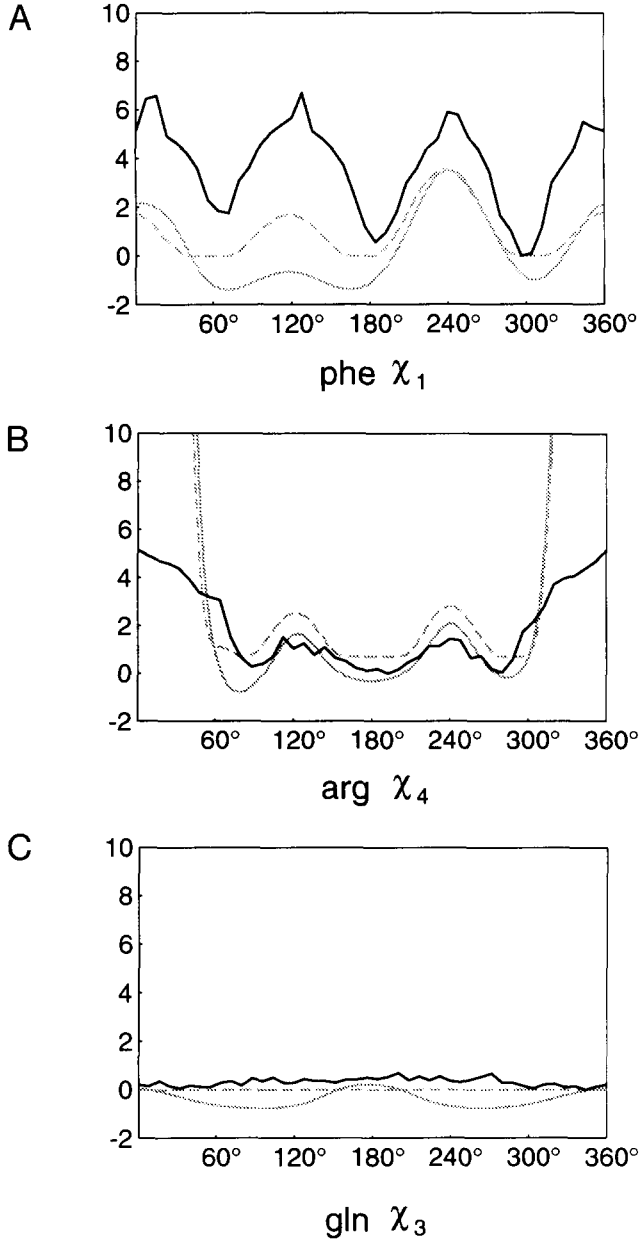

Fig. 3. A comparison of the conformational database, Lennard-Jones and repulsive van der Waals energies for several side-chain dihedral angles. **A:** Conformational database energy (in black), the Lennard-Jones energy (in gray), and the repulsive van der Waals energy (in dashed gray) of an isolated phenylalanine residue as a function of the χ_1 angle. **B,C:** The same energies for an isolated arginine residue as a function of χ_4 and an isolated glutamine residue as a function of χ_3. The conformational database energy is calculated with the scale factor k_{DB} set to 3, and a constant subtracted such that the minimum energy for the χ_1, χ_2, and χ_4 potential functions is zero.

from those observed for the Lennard-Jones and repulsive van der Waals terms. For example, the energy maximum at Phe $\chi_1 = 120°$ in the Lennard-Jones function is approximately six times greater than the peak at 240°, even though these two regions are equally unpopulated in high-resolution X-ray structures, as indicated by the peak heights observed at these two angle values in the conformational database function. This suggests that the Lennard-Jones energy function is not accurate enough to explain the observed distribution of Phe χ_1 values. In addition, the dynamic range of the conformational database energy function is much less than that of the Lennard-Jones or repulsive van der Waals terms. For example, the energy of the peak at Arg $\chi_4 = 0°$ is approximately five times greater than the peaks at 120° and 240° in the conformational database energy term, but it is 10^5 times larger in the Lennard-Jones energy term. The difference in energy, however, between the three minima at 60°, 180°, and 300°, and the peaks at 120° and 240° is 2-3 kcal·mol^{-1} for all three potentials. As a result, the repulsive and Lennard-Jones van der Waals terms provide little discriminatory power for χ_4 angles in the 50-310° range, because the differences in energy within this region are so small compared with that between this region and that in the completely nonallowed region between −50° and +50°.

To ensure that the PROCHECK (Laskowski et al., 1993) and backbone-dependent rotamer (Dunbrack & Karplus, 1993, 1994) databases are consistent with each other, the structure of reduced human thioredoxin was refined including information from these two database in the conformational database energy potential E_{DB}, either one at a time or simultaneously. Simultaneous refinement against the two databases produced structures that agreed with either database equally well as those structures that had been generated by independent refinement (data not shown). Thus, the backbone dependent rotamer and PROCHECK databases were used simultaneously in all further structure refinements.

Effects of conformational database refinement

The effects of including the conformational database potential on the quality, accuracy, and precision of the resulting structures of human thioredoxin are illustrated in Figures 4, 5, 6, and 7, and Tables 1, 2, and 3.

Whenever a new term is introduced into the target function (Equation 1), it is essential to first establish its optimal weighting relative to all the other terms. To this end, we carried out a series of calculations in which the scale factor k_{DB} for the conformational database term was varied between 0 and 10. The results are displayed in Figure 4. Not surprisingly, the introduction of the conformational database potential results in a slight decrease in the agreement with the experimental restraints. The reduction in the agreement with the experimental interproton distance restraints and the improvement in the quality of the backbone (measured as the percentage of residues in the most favorable region of the Ramachandran plot) and packing of the structures (measured by the WHAT IF packing score) follow an approximately asymptotic relationship as a function of increasing k_{DB}, beginning to level off at $k_{DB} \geq 3$. However, the reduction in the agreement between observed and calculated values of the $^3J_{HN\alpha}$ coupling constants that are directly related to ϕ is approximately linear up to $k_{DB} = 10$. Given that the agreement between measured $^3J_{HN\alpha}$ coupling constants in solution and those calculated from high-resolution crystal structures ranges from 0.5 to 0.8 Hz (Kay et al., 1989; Bartik et al., 1993; Vuister & Bax, 1993; Wang & Bax, 1996), we decided to opt for a value of $k_{DB} = 3$ for all subsequent calculations. At this value of k_{DB}, the agreement with the $^3J_{HN\alpha}$ coupling constant data falls within the expected range.

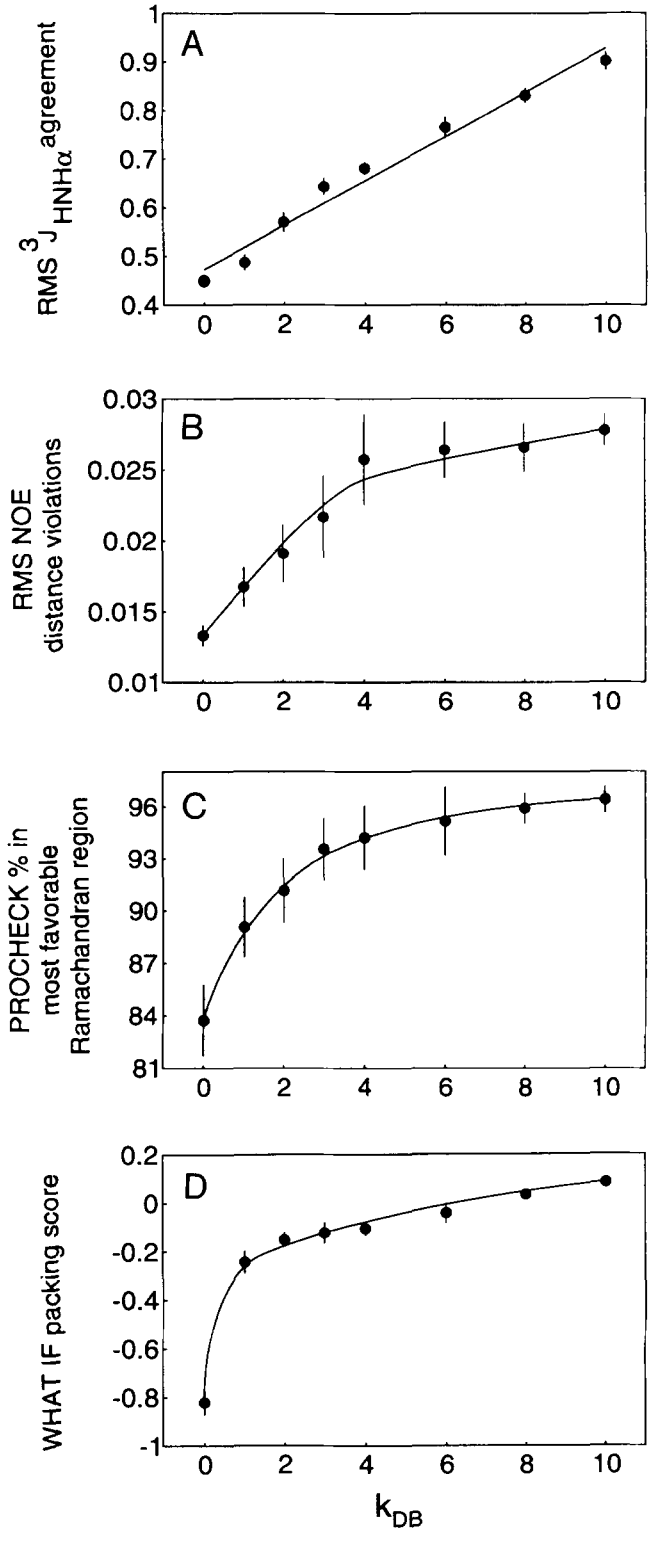

Fig. 4. Effect of the conformational database scale factor k_{DB} (Equation 2) on the agreement with the experimental interproton distance and $^3J_{HN\alpha}$ coupling constant restraints and on the quality of the backbone and packing. The quality of the backbone is expressed as the percentage of residues in the most favorable regions of the Ramachandran plot (Morris et al., 1992; Laskowski et al., 1993), and the packing is measured by the WHATIF packing score (Vriend & Sander, 1993). The filled-in circles represent the averages for 10 simulated annealing structures and the vertical bars are the standard deviations for the values.

In this particular case, including the conformational database potential in the target function with a k_{DB} value of 3 has relatively little impact on the agreement with either the experimental restraints or idealized covalent geometry (Table 1). Specifically, although the agreement between calculated and target values is somewhat worse upon conformational database refinement, the differences remain within the errors of the experimental and covalent geometry restraints. This provides a good indicator of the quality of both the original structures and the experimental restraints. If significant errors were present in the experimental interproton distance restraints (e.g., due to either misassignment of NOEs or severely misclassified interproton distance ranges), this would be reflected in a large increase in the RMS difference between calculated and target restraints, well beyond the expected errors in the experimental data. The precision of the coordinates remains essentially unchanged upon conformational database refinement, and the atomic RMS shifts in the mean coordinate positions are very small and within the scatter of the coordinates of the two ensembles of simulated annealing structures (Table 2). Interestingly, the use of a repulsive (Equation 4), repulsive-attractive (Equation 6), or Lennard-Jones (Equation 5) potential for the nonbonded contacts has no significant effect on the coordinate precision or positions, on the agreement with the experimental and covalent geometry restraints, or on the various measures of structure quality provided by PROCHECK (Laskowski et al., 1993) and WHAT IF (Vriend & Sander, 1993) (data not shown). This is not too surprising because all three potentials make use of the same PARAM19/20 CHARMM van der Waals radii (Brooks et al., 1983; Reiher, 1985) and are designed to display very similar properties with regard to the nonbonded contacts.

Figure 5A shows a typical Ramachandran ϕ,ψ backbone dihedral angle quality assessment for a reduced human thioredoxin structure calculated using the repulsive van der Waals potential for the nonbonded interactions. This plot is also typical of structures refined using the Lennard-Jones or attractive-repulsive van der Waals potentials. Figure 5B shows the same plot for a structure calculated using the conformational database potential in conjunction with the repulsive van der Waals potential. Upon conformational database refinement, the clustering of α helical-like ϕ,ψ angles become much tighter. The β sheet-like ϕ,ψ angles, on the other hand, still have about as broad a distribution after conformational database refinement as before. This reflects the wider variety of β sheet structures, which can be parallel or antiparallel and can have significant twists.

Figure 6 shows a similar assessment for some of the side-chain dihedral angles of the structures analyzed in Figure 5. Extremely poor conformations, such as that of the side chain of Gln 88, are avoided upon conformational database refinement. Unusual side-chain conformations, however, are still sometimes populated in the structure refined with the conformational database potential, reflecting the fact that some unusual conformations are found in high-resolution X-ray structures. In addition, the spread of the dihedral angles within a single rotamer becomes significantly tighter and closer to the expected idealized rotamer value after conformational database refinement.

Various quantitative measures of quality before and after conformational database refinement are provided in Table 3. Not surprisingly, conformational database refinement results in large improvements in the dihedral angle quality of the structures. The percentage of ϕ,ψ backbone dihedral angles in the most favor-

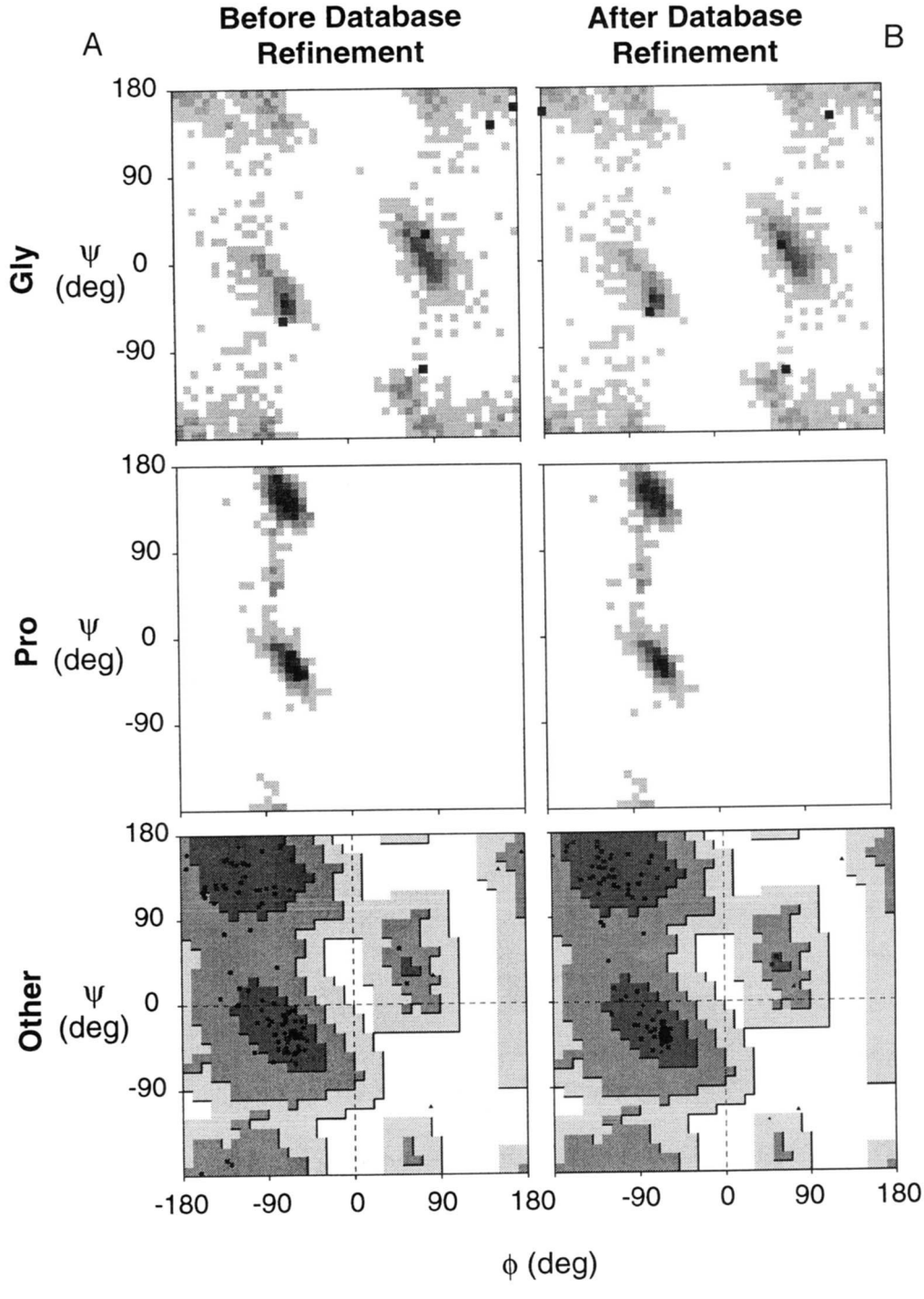

Fig. 5. Typical PROCHECK output showing the improvement in the quality of the Ramachandran ϕ,ψ plot upon conformational database refinement. **A:** Typical structure calculated with the van der Waals repulsive potential. **B:** Typical structure calculated with both the conformational database and van der Waals repulsive potentials. Both calculations include potentials for the 3,128 experimental NMR restraints and for covalent geometry. The intensity of the gray scale for the Gly and Pro ϕ,ψ plots is proportional to the populations in the PROCHECK database. There are four tones of grey from dark grey to white for the ϕ,ψ plot of the other residues, which refer to the most favored regions, the additionally allowed regions, the generously allowed regions, and the disallowed regions (Morris et al., 1992). In the PROCHECK database, the latter two regions are equally unpopulated (Morris et al., 1992).

able regions of the Ramachandran plot is unaffected by the type of potential (repulsive, repulsive-attractive, or Lennard-Jones) used for the nonbonded contacts, but increases from ~84% to ~93% upon conformational database refinement (with k_{DB} set to 3). For reference, a good quality model would be expected to have over 90% of residues in the most favored regions (Morris et al., 1992). However, WHAT IF's calculation of the number of residues with backbone torsion angles in unusual regions

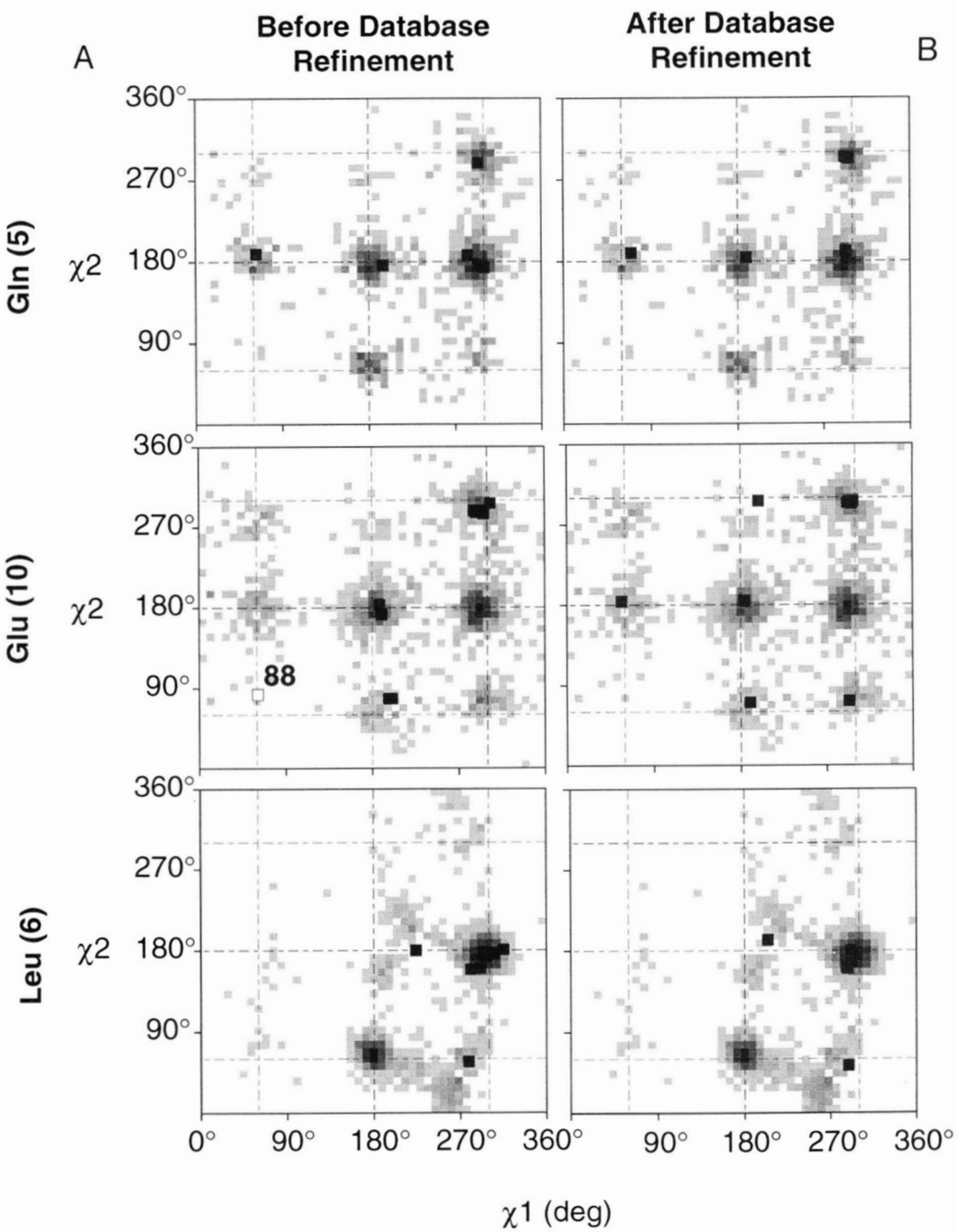

Fig. 6. Typical PROCHECK output showing the improvement in the quality of several side-chain torsion angles upon conformational database refinement. **A:** Typical structure calculated with the van der Waals repulsive potential (note extremely poor conformation of the side chain of Glu 88). **B:** Typical structure calculated with both the conformational database and van der Waals repulsive potentials. The intensity of the grey scale is proportional to the populations found in the PROCHECK database. Both calculations include potentials for the 3,128 experimental NMR restraints and for covalent geometry. Numbers in parentheses next to the residue names indicate the number of times that particular residue type is present in human thioredoxin.

of the Ramachandran plot remains essentially unaltered. WHAT IF can also evaluate the quality of the backbone ϕ,ψ dihedral angles by examining the distances between the actual and expected positions of the carbonyl oxygen atoms. If this distance is greater than 1 Å, a good alternative backbone position exists. The number of residues for which this is true is unaffected by the form of the nonbonded potential, but drops on average from 3 to 1 after conformational database refinement. The mean PROCHECK dihedral angle *G*-factors, for which values below −1.0 indicate the need for further investigation of the structure, are also unaffected by the form of the nonbonded potential, but improve from an average of −0.28 to +0.24 upon conformational database refinement. WHAT IF's average torsion angle quality score, for which a value less than −2 is considered poor,

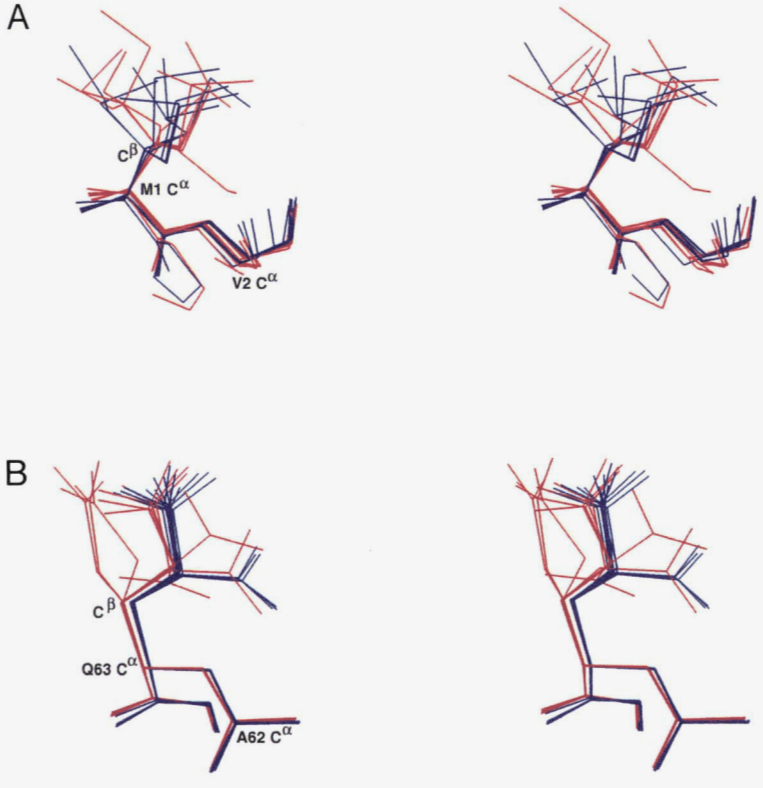

Fig. 7. Structural effects of conformational database energy refinement. **A:** Twenty superimposed structures of Met 1 (the side chain of which is unconstrained by the experimental NMR restraints). **B:** The same for Gln 63 (the side chain of which is partially constrained by the experimental NMR restraints). In each case, 10 structures (shown in red) were calculated with the van der Waals repulsion term alone, and 10 (shown in blue) with both the conformational database and repulsive van der Waals potentials.

improves from −0.25 to +0.31 upon conformational database refinement. Finally, WHAT IF's estimate of the likelihood of the values of the χ_1 angles, given the values of the backbone ϕ,ψ angles, is essentially unchanged upon conformational database refinement. This reflects the very high quality of the original structure, because nearly all of its χ_1 values are constrained by experimental restraints.

The number of bad van der Waals contacts in the structures, measured using either PROCHECK's or WHAT IF's definition of a "bad contact," is essentially unchanged upon conformational database refinement (Table 3).

Both PROCHECK and WHAT IF provide measures of the quality of hydrogen bonds in the structure (Table 3). PROCHECK's estimation of the hydrogen bond energy, calculated

Table 1. *Effect of conformational database refinement on the structural statistics*[a]

	Structures	
	No database refinement ⟨SA⟩	With database refinement ⟨SA_{DB}⟩
RMS deviations from experimental restraints		
Interproton distances (Å) (2,571)	0.014 ± 0.001	0.021 ± 0.002
Torsion angles (°) (273)	0.25 ± 0.04	0.43 ± 0.06
$^3J_{HN\alpha}$ couplings constants (Hz) (89)	0.45 ± 0.01	0.63 ± 0.02
$^{13}C\alpha$ chemical shifts (100)	1.01 ± 0.01	1.03 ± 0.07
$^{13}C\beta$ chemical shifts (95)	0.94 ± 0.007	0.98 ± 0.06
Deviations from idealized covalent geometry		
Bonds (Å) (1,653)	0.003 ± 0.000	0.004 ± 0.006
Angles (°) (2,985)	0.39 ± 0.01	0.45 ± 0.02
Impropers (°) (838)	0.28 ± 0.01	0.41 ± 0.07

[a] The number of restraints for the various terms are given in parentheses. The values reported are the means (±1 SD) for the 30 simulated annealing structures calculated with ($k_{DB} = 3$; cf. Equation 2) and without the conformational database potential using a quartic van der Waals repulsion term (Equation 4) for the nonbonded contacts. The results using the Lennard–Jones (Equation 5) or attractive–repulsive (Equation 6) van der Waals potentials are essentially identical. Full details of the experimental restraints are given in Qin et al. (1994). There are no interproton distance, torsion angle, or $^3J_{HN\alpha}$ coupling constant violations greater than 0.5 Å, 5°, or 2 Hz, respectively.

Table 2. *Coordinate precision and atomic RMS shifts*[a]

	Atomic RMS difference (Å)	
	Backbone	All atoms
Precision		
⟨SA⟩ versus $\overline{SA}$	0.25 ± 0.04	0.64 ± 0.05
⟨SA_{DB}⟩ versus $\overline{SA}_{DB}$	0.24 ± 0.04	0.58 ± 0.04
Atomic RMS shift		
$\overline{SA}_{DB}$ versus $\overline{SA}$	0.16	0.38

[a] The coordinate precision is defined as the average atomic RMS difference between the 30 individual simulated annealing coordinates and the mean coordinates generated by best-fitting the individual structures to residues 1–105. ⟨SA_{DB}⟩ and ⟨SA⟩ are the ensembles of simulated annealing structures (30 each) calculated with (k_{DB} = 3; cf. Equation 2) and without conformational database refinement, respectively; $\overline{SA}_{DB}$ and $\overline{SA}$ are the corresponding mean coordinates calculated from the ensembles. E_{rep}, E_{LJ}, and $E_{att-rep}$ are the repulsive (Equation 4), Lennard-Jones (Equation 5), and attractive-repulsive (Equation 6) van der Waals potentials. R_{eff} and R_{att} in Equations 4 and 6 are set to 0.8 and 1.4 times the value of R_{min}, respectively.

[b] The backbone atoms comprise the N, Cα, C, and O atoms.

using the method of Kabsch and Sander (1983), is largely unaffected by the choice of the nonbonded potential function. WHAT IF's estimate of the hydrogen bonding quality, as expressed by the number of unsatisfied hydrogen bond donors and acceptors in the structure and the number of His, Asn, and Gln side chains that could produce better hydrogen bonds if they were flipped 180°, is not affected significantly either by any of the refinement protocols used.

The packing of the structures, as measured by the Lennard-Jones energy, does not improve with conformational database refinement, but WHAT IF's estimation of the quality of the overall packing of the structures improves dramatically (Table 3). Structures calculated with repulsive van der Waals interactions alone generally have packing scores in the range typical for very good quality homology models, but structures calculated with both repulsive van der Waals and database potentials have packing scores that are equal to those seen in high quality protein crystal structures. This change is remarkable in that the conformational database energy we have described only affects local contacts, that is, those atoms that are connected by a maximum of three rotatable bonds, whereas WHAT IF's packing quality assessment considers interactions among all residues. Thus, improving the accuracy of local dihedral angle values through conformational database refinement greatly improves the overall packing of the resulting structures.

The agreement between observed and calculated ^{1}H chemical shifts provides a sensitive independent experimental probe of the accuracy of local structure in proteins (Williamson et al., 1995). This remains unaffected by either the choice of nonbonded potential or conformational database refinement (Table 3). Thus, the structural changes induced by conformational database refinement do not reduce the accuracy of the structures.

Two details of the sorts of structural changes induced by conformational database refinement are shown in Figure 7. Panel A shows 20 superimposed structures for Met 1 for which there are no side-chain NOE restraints. Ten of these structures, shown in red, were refined with the van der Waals repulsive potential alone, and the remaining 10, shown in blue, were refined both with the van der Waals repulsion potential and the conformational database potential. The side-chain dihedral angles are undetermined, but close inspection of the conformational database-refined structures shows that their χ_1 angles are closer to the three possible rotamers (although all three rotamers are populated). Panel B illustrates another situation for the side chain of Gln 63, which is partially restrained by NOEs. Although several χ_1 rotamers are seen in the original structures, only one is populated in the backbone-dependent rotamer database at that residue's backbone dihedral angles, and thus is the only one populated in the conformational database refined structures. There are still two possible conformations for χ_2, however, and their relative populations are close to those observed in PROCHECK's χ_1,χ_2 database for Gln.

Concluding remarks

The use of information from databases of highly refined, high-resolution protein crystal structures is becoming increasingly popular as a way of building side-chain conformations from Cα coordinates (Dunbrack & Karplus, 1993; Mathiowetz and Goddard, 1995), improving homology models of proteins (Sali & Blundell, 1993; Chinea et al., 1995), examining the intrinsic backbone conformation preferences of residues (Swindells at al., 1995), and de novo prediction of polypeptide structure (Evans et al., 1995). This work shows that these database techniques provide a useful addition for the refinement of NMR structures, leading to improvements in the physicochemical reasonableness of dihedral angles and the overall packing, while at the same time increasing the precision of the structures only slightly. Particularly important is that these improvements are not achieved at the expense of either the quality of the covalent geometry or the agreement with even a large number of structurally sensitive experimental restraints, providing, of course, that the latter do not contain any significant errors (e.g., arising from misassignments of NOEs or severe misclassification of NOE intensities). Indeed, the accuracy of the structures of reduced human thioredoxin is not hurt in any way, as evidenced by the independent measure provided by the agreement between the calculated and observed ^{1}H chemical shifts.

Although the conformational database energy term has the potential to predict reasonable χ_1 values, given the backbone dihedral angles, the χ_1 angles of the internal side chains in thioredoxin were only minimally affected by conformational database refinement because they were already restrained by the very high number of experimental NOE and torsion angle restraints. The conformational database potential, though, is expected to be of great use in the early stages of structure determination by NMR, when relatively few experimental restraints are available, and in cases where relatively few experimental NMR restraints can be extracted from the data, for example, as a result of unfavorable sample properties resulting in poor spectra. Further, because one can safely assume that 90–95% of all residues have a side-chain conformation resembling that of a common rotamer (Kleywegt & Jones, 1996), restricting the search of conformational space to that of the commonly occurring rotamers seems to be a most reasonable strategy. Under these conditions, residues that truly exhibit a skewed rotamer conformation will be spotted by specific discrepancies between

Table 3. *Effects of conformational database refinement on the overall quality of the calculated structures, as measured by PROCHECK (Laskowski et al., 1993), WHAT IF (Vriend & Sander, 1993), the total Lennard-Jones energy, and the agreement between observed and calculated ^1H chemical shifts*[a]

	Structures	
	No database refinement ⟨SA⟩	With database refinement ⟨SA_{DB}⟩
Quality of backbone		
PROCHECK: % in most favorable region[b]	84 ± 1.4	93.4 ± 1.1
WHAT IF: No. of residues with unusual backbone[c]	1.4 ± 0.7	1.4 ± 0.8
WHAT IF: No. of oxygen position violations[d]	3.0 ± 1.0	1.1 ± 0.9
Quality of side-chain torsion angles		
PROCHECK: Mean overall dihedral G-factor[e]	−0.28 ± 0.04	+0.24 ± 0.15
WHAT IF: Torsion angle score[f]	−0.25 ± 0.03	+0.31 ± 0.03
WHAT IF: Position specific rotamer score[g]	0.67 ± 0.008	0.67 ± 0.007
Quality of nonbonded contacts		
PROCHECK: No. of bad contacts per 100 residues[h]	3.0 ± 1.3	3.0 ± 1.2
WHAT IF: Bad contact score[i]	26 ± 4	24 ± 4
Quality of hydrogen bonds		
PROCHECK: H-bond energy[j]	0.68 ± 0.04	0.75 ± 0.05
WHAT IF: No. of unsatisfied H-bond donors[k]	9.6 ± 2	7.5 ± 1.4
WHAT IF: No. of unsatisfied H-bond acceptors[k]	0.6 ± 0.7	0.7 ± 0.7
WHAT IF: No. of flipped side chains[k]	0.9 ± 0.9	0.7 ± 0.6
Quality of packing		
WHAT IF: Packing score[l]	−0.82 ± 0.05	−0.13 ± 0.05
Lennard-Jones van der Waals energy (kcal·mol^{-1})[m]	−509 ± 7	−513 ± 9
Agreement with expt. ^1H chemical shifts		
RMS between observed and calculated (ppm)[n]	0.31 ± 0.005	0.32 ± 0.005

[a] The values reported are the averages for the 30 simulated annealing structures calculated with ($k_{DB} = 3$; cf. Equation 2) and without conformational database refinement using the quartic van der Waals repulsion term (Equation 4) for the nonbonded contacts. Essentially identical results are obtained using the Lennard-Jones (Equation 5) or attractive-repulsive (Equation 6) van der Waals potentials.

[b] For a good quality structure, greater than 90% of residues should occupy the most favorable region of the Ramachandran plot (Morris et al., 1992).

[c] This value reports the number of residues in strange loops or with something wrong with neighboring residues.

[d] The number of residues with an oxygen positional score greater than 1, indicating that a good alternative position exists.

[e] Ideally, the overall dihedral *G*-factor should have a score greater than −0.5. Values less than −1.0 need investigation. Note the value reported does not include the peptide backbone torsion angle ω because this is fixed by the covalent geometry restraints to be 180°, except for the *cis* peptide bond between Thr 74 and Pro 75, where it is fixed to 0°.

[f] A score of less than −2 for any residue is poor.

[g] A score of 1.0 indicates that all rotamers are in preferred orientations, whereas a score of 0.0 indicates that no rotamers are in preferred orientations.

[h] Less than 10 bad contacts per 100 residues are expected for a good quality structure.

[i] The WHAT IF bad contact score reports the number of atom pairs closer than the sum of the two van der Waals radii minus 0.4 Å.

[j] The expected PROCHECK H-bond energy for a good quality structure is between 0.6 and 1.0.

[k] In very good structures, the number of unsatisfied H-bond donors and acceptors, and the number of flipped His, Asn, and Gln side chains that would provide a better hydrogen bonding arrangement will tend toward zero. (Clearly, there will still be unsatisfied H-bond donors and acceptors on the surface of the protein in the absence of modeled water molecules).

[l] The meaning of the WHAT IF packing score is as follows: >−0.5, perfect structure; −0.5, average good structure; −1.0 to −0.5, still good or very good model; −1.5, the model is probably correct, but has many small errors; −2.0, the score that one would expect for an ab initio designed protein or a very poor model of a real protein; −3.0, the model is almost guaranteed to be incorrect.

[m] The Lennard-Jones energy is calculated with the PARAM19/20 CHARMM energy parameters (Brooks et al., 1983; Reiher, 1985).

[n] The expected RMS difference for a perfect structure is 0.23 ppm, and ~0.5 ppm for a random structure.

the model and the experimental restraints, and, in most circumstances, such violations will be accounted for by special structural features of the model. Moreover, one should be especially careful in believing a nonrotamer side-chain conformation in NMR structures in the absence of extensive NOE and coupling constant data relating to that particular residue. Exactly the same arguments can be applied to ϕ,ψ angles located in unfavorable regions of the Ramachandran plot, which likewise should be treated with extreme caution unless there is extensive experimental evidence to account for their unusual values (Kleywegt & Jones, 1996).

One would also expect the conformational database potential to be of particular value in the early stages of X-ray structure refinement, providing a means to avoid unreasonable dihedral angles, thereby reducing the number of iterative steps of modeling and refinement required to complete the structure determination. However, in crystallographic applications of conformational database refinement, it will be all the more important to make use of the free R-factor as a measure of accuracy (Brünger, 1992b) throughout the refinement procedure, because examination of the distribution of dihedral angles will no longer provide an independent measure of quality.

Finally, incorporation of the conformational database energy term in the final stages of refinement should provide a good indicator of the quality of both the model and the experimental restraints. Thus, in the case of an NMR structure determination, for example, the presence of errors in the experimental restraints will be reflected by a large deterioration in the agreement between calculated and target restraints upon conformational database refinement. Similarly, in the case of an X-ray structure determination, a poor quality model should be reflected by a large increase in the R and free-R factors upon conformational database refinement.

Some may regard the introduction of a conformational database energy term as a major step toward empiricism in NMR structure refinement, adding a term with apparently no direct physical counterpart, whose effect will be to make the dihedral angle distributions in NMR refined structures look more like those in crystal structures. We believe that the combined quality and quantity of high-resolution (≤2 Å) protein structures in the crystallographic databases argues strongly against such a viewpoint and makes it very difficult to ignore the available experimental observations relating to dihedral angles in proteins. First, it is invariably the case that high-resolution X-ray structures show significantly better agreement with solution observables such as coupling constants, ^{13}C chemical shifts, and proton chemical shifts, than the corresponding NMR structures, including the very best ones (obtained in the absence of direct coupling constant and chemical shift restraints) (Ösapay & Case, 1991; Williamson & Asakura, 1993; Garrett et al., 1994; Ösapay et al., 1994; Kuszewski et al., 1995a, 1995b; Williamson et al., 1995). Hence, in most cases, a high-resolution (≤2 Å) crystal structure will provide a better description of the structure in solution than the corresponding NMR structure. Second, the probability distributions for the various dihedral angles observed in the crystallographic database are a direct result of the underlying physical chemistry of the system and, as such, provide a perfectly reasonable, albeit empirically derived, measure of the relative energetics of different combinations of dihedral angles. Third, the discriminating and converging power of the conformational database potential with regard to dihedral angles is significantly better than that of the currently available empirical nonbonded potentials. This is hardly surprising because the conformational database potential acts directly on rotatable bonds, whereas the nonbonding potentials do not.

In conclusion, the conformational database refinement strategy presented in this paper is a method of restricting sampling during simulated annealing to conformations that are likely to be energetically possible by limiting the choices of dihedral angles to those that are known to be physically realizable. The variability in the structures produced by this method is therefore more likely to be a function of the experimental restraints, rather than an artifact of an inadequate nonbonded interaction model.

Acknowledgments

We thank Attila Szabo and Roman Laskowski for useful discussions. This work was supported by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (to G.M.C. and A.M.G.). The code for the conformational database potential is available upon request from the authors.

References

Bartik K, Dobson CM, Redfield C. 1993. ^{1}H-NMR analysis of turkey egg-white lysozyme and comparison with hen egg-white lysozyme. *Eur J Biochem 215*:255–266.

Braun W. 1987. Distance geometry and related methods for protein structure determination from NMR data. *Q Rev Biophys 19*:115–157.

Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. 1983. CHARMM: A program for macromolecular energy minimization and dynamics calculations. *J Comput Chem 4*:187–217.

Brünger AT. 1992a. *X-PLOR 3.0 reference manual*. New Haven, Connecticut: Yale Press.

Brünger AT. 1992b. The free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature 355*:472–474.

Brünger AT, Kuriyan J, Karplus M. 1987. Crystallographic R factor refinement by molecular dynamics. *Science 235*:458–460.

Brünger AT, Nilges M. 1993. Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR spectroscopy. *Q Rev Biophys 26*:49–125.

Brunne RM, Liepinsh E, Otting G, Wüthrich K, van Gunsteren WF. 1993. Hydration of proteins: A comparison of experimental residence times of water molecules solvating the bovine pancreatic trypsin inhibitor with theoretical model calculations. *J Mol Biol 231*:1040–1048.

Chandrasekhar I, Clore GM, Szabo A, Gronenborn AM, Brooks BR. 1992. A 500 ps molecular dynamics simulation study of interleukin-1β in water: Correlation with nuclear magnetic resonance spectroscopy and crystallography. *J Mol Biol 226*:239–250.

Chinea G, Pardon G, Hooft RWW, Sander C, Vriend G. 1995. The use of position-specific rotamers in model building by homology. *Proteins Struct Funct Genet 23*:415–421.

Clore GM, Ernst J, Clubb R, Omichinski JG, Kennedy WP, Sakaguchi K, Appella E, Gronenborn AM. 1995. Refined solution structure of the oligomerization domain of the tumour suppressor p53. *Nature Struct Biol 2*:321–33.

Clore GM, Gronenborn AM. 1989. Determination of three dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy. *CRC Crit Rev Biochem Mol Biol 24*:479–564.

Clore GM, Gronenborn AM. 1991. Comparison of the solution nuclear magnetic resonance and X-ray crystal structures of human recombinant interleukin-1β. *J Mol Biol 221*:47–53.

Clore GM, Gronenborn AM, Brünger AT, Karplus M. 1985. The solution conformation of a heptadecapeptide comprising the DNA binding helix F of the cyclic AMP receptor protein of *Escherichia coli*: Combined use of ^{1}H-nuclear magnetic resonance and restrained molecular dynamics. *J Mol Biol 191*:523–551.

Clore GM, Robien MA, Gronenborn AM. 1993. Exploring the limits of precision and accuracy of protein structures determined by nuclear magnetic resonance spectroscopy. *J Mol Biol 231*:82–102.

Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Fergusson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. 1995. A second gen-

eration force field for the simulation of proteins, nucleic acids and organic molecules. *J Am Chem Soc 117*:5179–5197.
Dunbrack RL, Karplus M. 1993. Backbone-dependent rotamer library for proteins: Application to side-chain prediction. *J Mol Biol 230*:543–574.
Dunbrack RL, Karplus M. 1994. Conformational analysis of the backbone-dependent rotamer preferences of protein side chains. *Nature Struct Biol 1*:334–340.
Engh RA, Huber R. 1991. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr A 47*:392–400.
Eriksson MAL, Härd T, Nilsson L. 1995. Molecular dynamics simulations of the glucocorticoid receptor DNA-binding domain complex with DNA and free in solution. *Biophys J 68*:402–426.
Evans JS, Mathiowetz AM, Chan SI, Goddard WA. 1995. De novo prediction of polypeptide conformations using dihedral probability grid Monte Carlo methodology. *Protein Sci 4*:1203–1216.
Garrett DS, Kuszewski J, Hancock TJ, Lodi PJ, Vuister GW, Gronenborn AM, Clore GM. 1994. The impact of direct refinement against three bond HN-C$^{\alpha}$H coupling constants on protein structure determination by NMR. *J Magn Res Series B 104*:99–103.
Gronenborn AM, Clore GM. 1995. Structures of protein complexes by multidimensional heteronuclear magnetic resonance spectroscopy. *CRC Crit Rev Biochem Mol Biol 30*:351–385.
Halgren TA. 1992. Representation of van der Waals (vdW) interactions in molecular mechanics force fields: Potential form, combination rules, and vdW parameters. *J Am Chem Soc 114*:7827–7843.
Havel TF. 1991. An evaluation of computational strategies for use in the determination of protein structure from distance constraints obtained by nuclear magnetic resonance. *Prog Biophys Mol Biol 56*:43–78.
Hendrickson WA. 1985. Stereochemically restrained refinement of macromolecular structures. *Methods Enzymol 11*:252–270.
Jack A, Levitt M. 1978. Refinement of large structures by simultaneous minimization of energy and *R* factor. *Acta Crystallogr A 34*:931–935.
Jorgensen WL, Tirado-Rives. 1988. The OPLS potential functions for proteins: Energy minimizations for crystals of cyclic peptides and crambin. *J Am Chem Soc 110*:1657–1666.
Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*:2577–2637.
Kay LE, Brooks BR, Sparks SW, Torchia DA, Bax A. 1989. Measurement of NH-CaH coupling constants in staphylococcal nuclease by two-dimensional NMR and comparison with X-ray crystallographic results. *J Am Chem Soc 111*:5488–5490.
Kleywegt GJ, Jones TA. 1996. Good model-building and refinement practice. *Methods Enzymol*. Forthcoming.
Konnert JH, Hendrickson WA. 1980. A restrained parameter thermal factor refinement procedure. *Acta Crystallogr A 36*:344–349.
Kuszewski J, Gronenborn AM, Clore GM. 1995a. The impact of direct refinement against proton chemical shifts on protein structure determination by NMR. *J Magn Res Series B 107*:293–297.
Kuszewski J, Nilges M, Brünger AT. 1992. Sampling and efficiency of metrix matrix distance geometry: A novel partial metrization algorithm. *J Biomol NMR 2*:33–56.
Kuszewski J, Qin J, Gronenborn AM, Clore GM. 1995b. The impact of direct refinement against $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts on protein structure determination by NMR. *J Magn Res Series B 106*:92–96.
Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr 26*:283–291.
Loncharich RB, Brooks BR. 1990. Temperature dependence of dynamics of hydrated myoglobin. *J Mol Biol 215*:439–455.
Mathiowetz AM, Goddard WA. 1995. Building proteins from Cα coordinates using the dihedral probability grid Monte Carlo method. *Protein Sci 4*:1217–1232.
Momany FA, Carruthers LM, McGuire RF, Scheraga HA. 1974. Intermolecular potentials from crystal data III: Determination of empirical potentials and application to the packing configurations and lattice energies in crystals of hydrocarbons, carboxylic acids, and amides. *J Phys Chem 78*:1595–1620.
Morris AL, MacArthur MW, Hutchinson EG, Thronton AM. 1992. Stereochemical quality of protein structure coordinates. *Proteins Struct Funct Genet 12*:345–364.
Nilges M, Clore GM, Gronenborn AM. 1988a. Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett 229*:317–324.
Nilges M, Clore GM, Gronenborn AM. 1988b. Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms. *FEBS Lett 239*:129–136.
Nilges M, Clore GM, Gronenborn AM. 1990. ^{1}H-NMR stereospecific assignments by conformational database searches. *Biopolymers 29*:813–822.
Nilges M, Gronenborn AM, Brünger AT, Clore GM. 1988c. Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints: Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. *Protein Eng 2*:27–28.
Ösapay K, Case DA. 1991. A new analysis of proton chemical shifts in proteins. *J Am Chem Soc 113*:9436–9444.
Ösapay K, Theriault Y, Wright PE, Case DA. 1994. Solution structure of carbonmonoxy myoglobin determined from nuclear magnetic resonance distance and chemical shift constraints. *J Mol Biol 244*:183–197.
Qin J, Clore GM, Gronenborn AM. 1994. The high-resolution three-dimensional solution structures of the oxidized and reduced states of human thioredoxin. *Structure 2*:503–522.
Ramachandran G, Venkatachalam C, Krimm S. 1966. Stereochemical criteria for polypeptide and protein chain conformations. *Biophys J 6*:849–872.
Reiher WE. 1985. Theoretical studies of hydrogen bonding. [thesis]. Cambridge, Massachusetts: Harvard University.
Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol 234*:779–815.
Swindells MB, MacArthur MW, Thornton JM. 1995. Intrinsic ϕ,ψ propensities of amino acids derived from the coil regions of known structures. *Nature Struct Biol 2*:596–603.
van Gunsteren WF, Lugue FJ, Timms D, Torda AE. 1994. Molecular mechanics in biology: From structure to function taking into account solvation. *Ann Rev Biophys Biomol Struct 238*:847–863.
Vriend G, Sander C. 1993. Quality control of protein models: Directional atomic contact analysis. *J Appl Crystallogr 26*:47–60.
Vuister GW, Bax A. 1993. Quantitative J correlation: A new approach for measuring homonuclear three-bond J($H^{N}H^{\alpha}$) coupling constants in ^{15}N enriched proteins. *J Am Chem Soc 115*:7772–7777.
Wang AC, Bax A. 1996. Determination of backbone dihedral angles ϕ in human ubiquitin from reparametrized empirical Karplus equations. *J Am Chem Soc*. Forthcoming.
Weiner SJ, Kollman PA, Nguyen DT, Case DA. 1986. An all-atom force field for simulations of proteins and nucleic acids. *J Comput Chem 7*:230–252.
Williamson MP, Asakura T. 1993. Empirical comparisons of models for chemical-shift calculations in proteins. *J Magn Res Series B 101*:63–71.
Williamson MP, Kikuchi J, Asakura T. 1995. Application of ^{1}H NMR chemical shifts to measure the quality of protein structures. *J Mol Biol 247*:541–546.