PROTEINS: Structure, Function, and Bioinformatics 64:652–664 (2006)

Error Distribution Derived NOE Distance Restraints

Michael Nilges[1,2]* Michael Habeck[1,2] Seán I. O'Donoghue[2,3] and Wolfgang Rieping[1,2]
[1]*Unité de Bio-informatique structurale, CNRS URA 2185, Institut Pasteur, Paris, France*
[2]*Structural and Computational Biology Programme, EMBL, Heidelberg, Germany*

ABSTRACT **Errors and imprecisions in distance restraints derived from NOESY peak volumes are usually accounted for by generous lower and upper bounds on the distances. In this paper, we propose a new form of distance restraints, replacing the subjective bounds by a potential function obtained from the error distribution of the distances. We derived the shape of the potential from molecular dynamics calculations and by comparison of NMR data with X-ray crystal structures. We used complete cross-validation to derive the optimal weight for the data in the calculation. In a model system with synthetic restraints, the accuracy of the structures improved significantly compared to calculations with the usual form of restraints. For experimental data sets, the structures systematically approach the X-ray crystal structures of the same protein. Also standard quality indicators improve compared to standard calculations. The results did not depend critically on the exact shape of the potential. The new approach is less subjective and uses fewer assumptions in the interpretation of NOESY peak volumes as distance restraints than the usual approach. Figures of merit for the structures, such as the RMS difference from the average structure or the RMS difference from the data, are therefore less biased and more meaningful measures of structure quality than with the usual form of restraints. Proteins 2006;64:652–664.**

INTRODUCTION

Parameters characterizing a molecular conformation can be obtained from NMR data only with some error. This applies in particular to NOE derived distances, mostly because of approximations in the theory relating the measured data to the structure. The standard way to account for errors and imprecisions during structure calculations is by distance bounds that are wide enough to account for all sources of error and remove geometrical inconsistencies. This data representation with lower and upper bounds, strongly influenced by the concept of distance geometry,[1,2] has been used almost exclusively for determining NMR structures to date. Over the years, a number of methods to generate molecular conformations (nearly) satisfying the bounds have been developed and extensively reviewed.[3–5]

In contrast, little development has taken place in data treatment during the structure calculation. A typical potential for NOE-derived distance restraints has a flat-bottom-harmonic-wall (FBHW) form, that is, it is flat between the upper and lower bounds U and L, and rises harmonically in the deviation of the distance r ($\mathbf{X}$) calculated from the current structure $\mathbf{X}$ from the bounds. Modifications have been proposed to increase numerical stability, convergence, and tolerance to errors in the data.[6] In recent versions of X-plor[7,8] and CNS,[9] the energy of a single distance restraint is

$$E^i_{NOE} = w^i_{NOE}\begin{cases}\alpha' + \beta'\Delta^i_-(\mathrm{X}) + \gamma'\Delta^i_-(\mathrm{X})^{-1} & \text{if } \mathrm{r}^i(\mathrm{X}) < \mathrm{L}^i - \sigma' \\ \Delta^i_-(\mathrm{X})^2 & \text{if } \mathrm{L}^i - \sigma' \le \mathrm{r}^i(\mathrm{X}) < \mathrm{L}^i \\ 0 & \text{if } \mathrm{L}^i \le \mathrm{r}^i(\mathrm{X}) \le \mathrm{U}^i \\ \Delta^i_+(\mathrm{X})^2 & \text{if } \mathrm{U}^i < \mathrm{r}^i(\mathrm{X}) \le \mathrm{U}^i + \sigma \\ \alpha + \beta\Delta^i_+(\mathrm{X}) + \gamma\Delta^i_+(\mathrm{X})^{-1} & \text{if } \mathrm{r}^i(\mathrm{X}) > \mathrm{U}^i + \sigma \end{cases} \quad (1)$$

where, Δ^i_- ($\mathbf{X}$) = [$\mathrm{L}^i - \mathrm{r}^i(\mathbf{X})$]/Å, Δ^i_+ ($\mathbf{X}$) = ($\mathrm{r}^i(\mathbf{X}) - \mathrm{U}^i$)/Å, and β, β', σ, and σ' are the coefficients of the soft-square potential; α, α', γ, and γ' follow from the conditions of continuity and differentiability at σ and σ'. $r(\mathbf{X})$, U, L, and σ are measured in Å. w^i_{NOE} is the weight on distance restraint i.

FBHW potentials are popular since energy and force are zero between U_i and L_i and, hence, do not influence the structure if the distance restraint is satisfied. In practice, however, the distance bounds representation can be problematic:

- The criteria for the derivation of the bounds are subjective. The quality and the consistency of the data are unknown before the structure calculation and need to be determined empirically.

Abbreviations: BPTI, bovine pancreatic trypsin inhibitor; DRS, distance restraint set; MD, molecular dynamics; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser exchange spectroscopy; PDB, protein data bank; RMS, root mean square; $\mathrm{RMSD}_{\mathrm{X\text{-}ray}}$, mean RMS distance from X-ray crystal structure; $\mathrm{RMSD}_{\mathrm{AVE}}$, mean RMS distance from average structure; PDF, probability density function; Il4, Interleukin 4; ARIA, ambiguous restraints for iterative assignments.

Grant sponsor: DFG; Grant numbers: Ni 499/1–1,2; Grant sponsor: EC; Contract numbers: QLG2-CT-2000-01313, QLG2–CT–2002–00988.

S.I. O'Donoghue's present address is Cay Labs, Untere Neckarstrasse 56, 69117 Heidelberg, Germany.

*Correspondence to: Michael Nilges, Unité de Bio-informatique structurale, CNRS URA 2185, Institut Pasteur, 25–28 rue du docteur Roux, F–75015 Paris, France. E-mail: nilges@pasteur.fr

Received 13 July 2005; Revised 11 October 2005; Accepted 22 October 2005

Published online 25 May 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20985

TABLE I. Data Sets Used to Derive the Restraint Sets for the Calculations

Protein	N_{NOE}[a]	N_{AA}[b]	$\frac{N_{NOE}}{N_{AA}}$	Comments	Reference
Il4	809	133	63	Distance classes + IC[c]	19
BPT1	642	58	115	Distance classes + IC	20
PH	1438[d]	106	13[d]	NOE volumes, ISPA[e]	21
GB1	845	56	153	Distance classes + IC	22
Ubiquitin	1444	75	19	ISPA calibration + IC	23
$BPTI_{MD}$	1543[d]	58	26[d]	MD trajectory	10

[a] N_{NOE}: Number of restraints.
[b] N_{AA}: Number of amino acids.
[c] IC: Individual corrections on some bounds.
[d] Includes all intra–residue distances, also covalently fixed distances.
[e] ISPA: Isolated spin pair approximation.

- The influence of the energy parameters, in particular the nonbonded representation, is large;
- Other data (e.g., coupling constants, residual dipolar couplings) are usually not modeled with FBHW potentials;
- Very wide error bounds may need to be used to obtain low energy structures, reducing the information content. This is problematic in particular with sparse data sets;
- Selective modification of some restraints is usually necessary to obtain low energy structures, adding another source of potential bias.

One of the consequences is that we do not have reliable figures of merit for NMR structures. The bounds representation makes a comparison of the determined structures to the estimated input distances meaningless. The often quoted number of violations or RMS difference to bounds does not reflect the quality of the structure or the fit to the data, since the bounds concept demands that the data have been manipulated to the extent that all violations disappear.

Also the RMS difference of the structure ensemble from the average structure ($RMSD_{ave}$, precision) is heavily influenced by the width of the bounds and the manipulation of individual bounds,[10] and is therefore not an unbiased measure of the structure quality. Recent attempts to improve the estimation of precision by resampling[11] do not get rid of the source of the problem, the subjective and biased choice of the distance bounds.

To date, only very few structure determinations did not use error bounds for distance restraints,[12] among them a model study using noise-free distances.[13] It seems time to reevaluate alternatives to distance bounds with a variety of data sets and state-of-the-art structure calculation algorithms.

In the approach we present in this paper, the FBHW potential is replaced by a potential derived from the error distribution of the distances. Since the error distribution for NOE-derived distances is, a priori, unknown, we followed two different approaches to obtain error distributions and restraint potentials. First, we derived an error distribution from a 6.6-nsec molecular dynamics trajectory of BPTI.[10,14] We used a potential derived from this error distribution by manual fitting for calculating structures with the theoretical restraint set derived from the same MD trajectory, and with several experimental data sets. For one data set, we tested the effect of increased noise in the data by randomizing the target distances. Second, we derived potentials from comparisons of experimental NOE distances and X-ray crystal structures of the same protein.

In all calculations, we studied the influence of the data weight on various parameters. To compare the results to those obtained with standard distance bounds, we recalculated the structures with the original restraint sets and standard calculation parameters, with the following difference: we did not use any data other than NOE-derived distance restraints (such as hydrogen bonds, coupling constants, torsion angles, or residual dipolar couplings). We used complete cross-validation[15] in all calculations, to define an optimal data weight, and as a figure of merit of structure quality. We analyzed all structures in terms of their difference to the X-ray crystal structures, precision, and the fit to the data. Our principal aim is to demonstrate that distance potentials without a flat bottom region have advantages over FBHW potentials, even for very noisy data sets, and produce structures of similar or higher quality. We discuss the usefulness of different criteria (fit to the data, cross-validation, precision) as figures of merit in the new calculation scheme.

THEORY AND METHODS

Derivation of Distance Estimates From Experimental Restraint Tables

The list of experimental data sets is given in Table I.

We obtained the experimental distance bounds from the restraint files submitted to the PDB,[16] apart for the PH domain for which we used NOE volumes.

There are not many cases for which an X-ray crystal structure and a good NOE data set in the appropriate format (NOE intensities or at least uncorrected distance estimates) are available. Some of the experimental bounds (IL4, GB1, BPTI) contained distance corrections that had been necessary in the original structure calculation to account for pseudo atoms.[17] For IL4, GB1, and BPTI, individual distance bounds had been modified for undocumented reasons, to obtain a geometrically consistent dataset.

For the BPTI data set, only upper bounds were reported, which made some guesses necessary as to the values of the corrections. Pseudo atom corrections are not necessary due to the use of sum averaging[18] for methyl and aromatic protons, and our present calculation approach makes also the modification of "outlier" distance restraints superfluous. We removed all these corrections at least approximately and then estimated the original volumes from the inverse sixth or fourth powers of the distance estimate or upper distance bounds. We assumed that a fourth power law had been used to obtain the upper distance limits from the NOEs involving side-chains for BPTI.

Each data set was then calibrated with the known structure, using the automatic calibration in ARIA with the isolated spin pair approximation and sixth power distance dependency of the NOE volumes, to obtain distance estimates.

Derivation of a Restraint Potential From Error Distributions

An experimental protein structure **X** needs to satisfy the experimental data (in the present case, a set of interproton distances) and certain general requirements valid for all proteins (e.g., bond lengths) simultaneously. This is usually achieved by minimizing a weighted sum of a term describing the experimental data and a term describing the prior knowledge about the molecule.[4] The prior knowledge is introduced in the form of an adapted molecular dynamics force field. The relative weight between the terms in the hybrid energy function is usually set empirically.

The distribution of errors of NOE-derived distances is unknown. This is one of the reasons to use the FBHW potential (Equation 1). In the case of a Gaussian error distribution, it would be appropriate to minimize the square of the differences between structure and data, and the restraint potential would therefore be harmonic.[24] In general, once we know the error distribution $P(r - r_{ref})$ in the distances r around the "true" distance r_{ref}, we can construct a restraint potential by taking the negative logarithm of the distribution. Assuming that the individual distance measurements are statistically independent, we obtain as potential E_i^{NOE} for a single restraint i:

$$E_i^{NOE} \propto - \log[P(r_{exp}^i - r^i(X))], \quad (2)$$

where $r^i(\mathbf{X})$ is the distance calculated in the structure **X**, and r_{exp}^i is the measured distance.

In the present paper, we derived a first error distribution and the corresponding potential from a molecular dynamics simulation. A total of 1543 "experimental" distances were calculated from cross-relaxation rates as described.[10] As "true" values we used the arithmetic average over the trajectory, $r_{ref} = \langle r \rangle_t$, where t is the time in the trajectory. To calculate histograms, we binned the distance differences into bins of 0.1 Å.

We then used the following pairs of X-ray crystal structures and NOE data sets to derive additional error distributions: GB1,[22,25] BPTI,[20,26] Il4[19,27] the mouse β spectrin PH domain,[21,28] and Ubiqutin.[23,29] The data sets and protein structures were quite varied in completeness and quality. The data set for Il4 is of intermediate quality and completeness; the GB1 and BPTI data sets are rather complete, with qualitative distance estimates in three (GB1) or five (BPTI) classes. Il4 is mostly helical, whereas the other three proteins contain helix and sheet structure. We calculated the differences in the distances obtained from the calibrated NOEs and those in X-ray crystal structure to obtain error distributions.

Optimal Fitting of Potential Parameters

In order to use the probability-derived potential (Equation 2) in structure calculations, we need to fit a differentiable analytic function to it. We used the functional form already implemented in X-plor[8] and CNS[9] (Equation 1, by setting $r_{exp}^i = U_i = L_i$ for all restraints i and fitting the potential parameters β, β′, σ, and σ′):

$$E_{NOE}^i = w_{NOE}^i \begin{cases} \alpha' + \beta' \Delta^i(X) + \gamma' \Delta^i(X)^{-1} & \text{if } r^i(X) < r_{exp}^i - \sigma' \\ \Delta^i(X)^2 & \text{if } r_{exp}^i - \sigma' \le r^i(X) \le r_{exp}^i + \sigma \\ \alpha + \beta \Delta^i(X) + \gamma \Delta^i(X)^{-1} & \text{if } r^i(X) > r_{exp}^i + \sigma \end{cases} \quad (3)$$

where now $\Delta^i(\mathbf{X}) = |r_{exp}^i - r^i(\mathbf{X})|/\text{Å}$, and r_{exp}^i is the target distance estimated from the data.

Note that α, α′, γ, and γ′ follow from the conditions of continuity and differentiability at σ and σ′. $r(\mathbf{X})$, r_{exp}^i, and σ are measured in Ångstroms. w_{NOE}^i is the weight on distance restraint i.

To use a potential with this analytical form for structure calculations with error distribution derived potentials, we needed to obtain optimal values for the shape parameters β,β′ and σ,σ′. We first derived a potential from the trajectory data by fitting potentials with different values of β,β′ and σ,σ′ to the plotted probability distribution by eye. To be more quantitative, we developed a program using Bayesian parametric density estimation. This approach determines most probable values of the parameters with their respective reliabilities.

We start from a set of distance differences $D = \{d^i\}$, with $d^i = r_{exp}^i - r_{ref}^i$, where r_{exp}^i is the experimental distance and r_{ref}^i the reference distance (either from the trajectory or from an X-ray crystal structure). We then used the exponential of the negative potential defined in Equation 3 with $w_i^{NOE} = 1$ to model the shape of the "experimental" error distribution. Two additional parameters k and δ describe the width of the potential and its shift relative to zero. We employed Bayesian inference to determine probable values for all unknowns. Assuming logical independence of the i), the likelihood function

$$L(\lambda,k,\delta) = [Z(\lambda,k,\delta)]^{-n} \exp\left\{ -\frac{k}{2} \sum_{i=1}^{n} E_i^{NOE}(d^i - \delta;\lambda) \right\} \quad (4)$$

quantifies the likelihood of observing the data D, given values for k, δ, and the shape parameters $\lambda = \{\beta,\beta',\sigma,\sigma'\}$ of the potential. Here, n denotes the size of the data set, and $Z(\cdot)$ is a normalizing function. Using a flat prior distribu-

tion for β, β′, σ, σ′ and γ, and Jeffreys's prior[30] for k, Bayes' theorem[31] yields the probability distribution for all unknowns:

$$Prob(\lambda,k,\delta|D) \propto k^{-1} L(\lambda,k,\delta). \quad (5)$$

We generated a set of stochastic samples drawn from distribution (Eq. 5) using a self-written Markov chain Monte-Carlo sampling program. Expectation values and standard deviations for each parameter were calculated using Monte Carlo estimators.

Determination of the Optimal Weight with Cross-Validation

An unbiased measure of structure quality can be obtained by complete cross-validation.[15] Data are divided into a working set, used to calculate the structure, and a test set, used only to evaluate the structure. Two measures of quality can be obtained: the RMS difference to the working set defines the usual fit to the distance restraints (or distance bounds):

$$RMS_{work} = \sqrt{N_{work}^{-1} \sum_{i=1}^{N_{work}} \Delta^i(X)^2} \quad (6)$$

$$RMS_{CV} = \sqrt{N_{CV}^{-1} \sum_{i=1}^{N_{CV}} \Delta^i(X)^2}, \quad (7)$$

where N_{work} and N_{CV} are the number of data points in the working and test sets, respectively. We used tenfold complete cross-validation, that is, the data were divided into ten partitions of roughly equal size. Each of these partitions was chosen in turn as a test set and ten structures were calculated with the corresponding working set. The values for RMS_{work} and RMS_{CV} for the three lowest energy structures for all working sets were then averaged.

A minimum in RMS_{CV} indicates optimal calculation conditions.[15] We varied the weight on the distance restraint term in factors of 2 from 2 to 512. Since RMS_{CV} rarely showed a minimum, we defined the optimal weight consistently as the point where there was less than 10% change in RMS_{CV}.

For the calculations with FBHW potentials, the RMS differences were evaluated from the bounds:

$$RMS_{bounds} = \sqrt{\frac{1}{N_{data}} \sum_{i=1}^{N_{data}} (\min(0, L_i - \sigma_i(X)) + \min(\sigma_i(X) - U_i))^2}, \quad (8)$$

where N_{data} is the number of data points in the working or test set.

Structure Calculation

Structures were calculated with ARIA1.2/CNS, using the ARIA simulated annealing protocol with torsion angle dynamics and Cartesian dynamics,[6] with the same number of annealing steps for all structures. The distance restraint potential in Figure 1(a) with its very flat asymptotic region would make convergence from random structures slow. The asymptotes for large violations (parameters β and β′) were therefore set to 1.0 and −1.0 in the high temperature phase for all calculations. In this phase, the weights for the NOE distance restraints were set to the same values in all calculations (50), independently of the final weight.

In the first cooling phase, the weights and the asymptotes of the NOE distance potential were then decreased or increased to the final values, and the weight of the vdW repel energy was brought to its final value. We used version 5.3 of the PARALLHDG parameters,[32] with covalent parameters from Engh and Huber,[33] and nonbonded parameters from PROLSQ.[34] Disulfide bonds in BPTI and IL4 were included as covalent bonds. The three lowest energy structures for each test set were selected in each calculation, resulting in 30 structures out of a total of 100 for each weight. Structures were superposed with a modified iterative fitting procedure ("wellordered.inp") where, in contrast to previous versions,[35] ill-fitting residues are iteratively down-weighted rather than excluded from the fit. RMS differences to X-ray crystal structures and average structures were calculated for the regions that were well defined in the originally published NMR structures, that is, residues 1–72 for Ubiqutin, residues 1–56 for GB1, residues 2–56 for BPTI, residues 5–129 for IL4, and residues 2–11, 24–31, 34–38, 41–46, 62–65, 75–79, 85–89 and 93–104 for the PH domain. Results were analyzed and plotted with Mathematica 5.1.[36] "Sausage" plots were generated with MolMol;[37] macros for coloring the structures with different quantities were generated automatically with ARIA/CNS.

RESULTS AND DISCUSSION

Derivation of a Restraint Potential From a Molecular Dynamics Trajectory

The molecular dynamics simulation on BPTI[14] provided a consistent model system, since NOE distance restraints and reference structure had been obtained from the same trajectory. As in a previous paper,[10] we defined the arithmetic average of the inter-proton distance over the trajectory as the true value r_{ref} for a distance, and the distance calculated from the relaxation rate as the experimental value r_{exp}. The negative logarithm of the distribution of the difference $r_{ref} - r_{exp}$ is shown in Figure 1(a). As already discussed previously,[10,14] the distribution is very different from a Gaussian distribution and has a long tail due to many large outliers.

We first fitted the parameters β, β′, σ, and σ′ (and an additional overal scale k, see Table II) manually to the negative logarithm of the probability distribution. We then symmetrized the potential to take into account effects leading to a decrease in the measured cross-relaxation rate (e.g., due to peak broadening) that are not included in the derivation of the data set from a dynamics trajectory. We call this potential PDF_{MDman}. It is very different from a flat-bottom-harmonic-wall potential. It has a small harmonic minimum region close to zero, rises virtually linearly for values smaller than zero, and tails off quickly

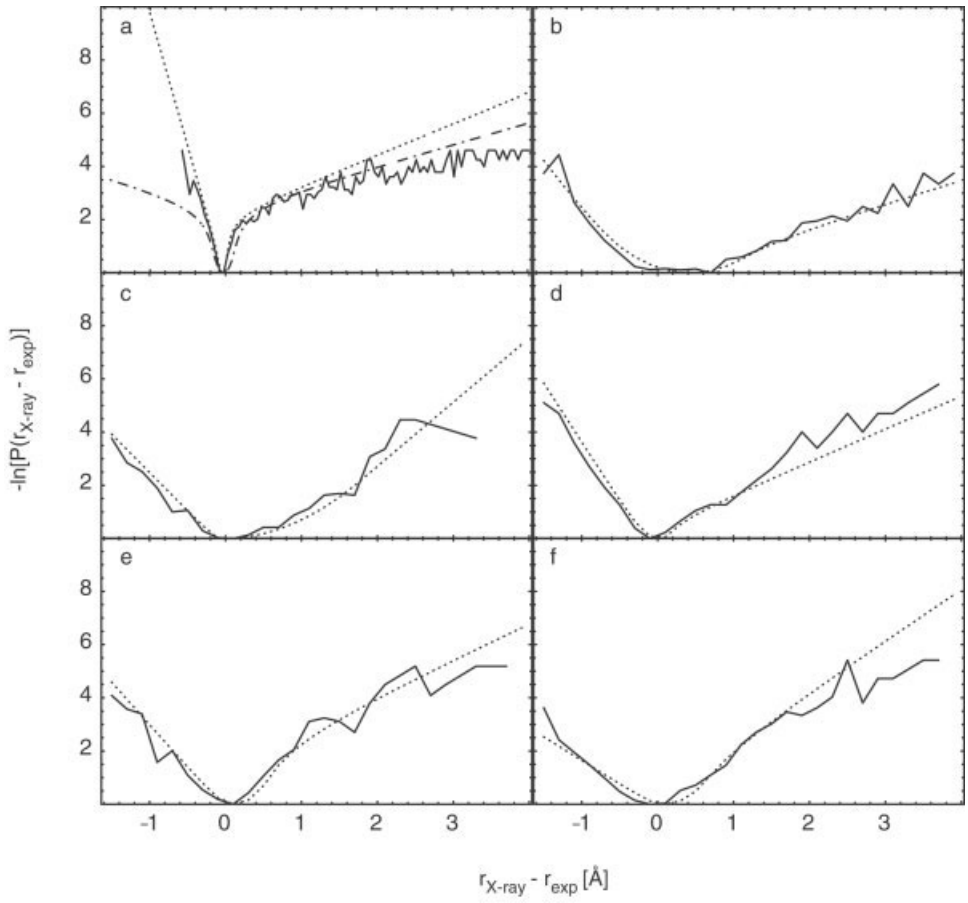

Fig. 1. Negative logarithm of the normalized distribution of $r_{ref} - r_{exp}$ (solid line), a manually fitted symmetric function with the functional form given in Equation (2.3) [dashed line, only in (a)], and optimal fit (dotted line). (**a**) BPTI trajectory data, (**b**) Il4, (**c**) BPTI, (**d**) PH domain, (**e**) GB1 domain, (**f**) Ubiquitin. See Table II for the values of the parameters $\sigma,\sigma',\beta,\beta'$.

TABLE II. Potential Parameters Derived from the MD Trajectory and X–Ray Structures

	β'	σ'	δ	β	σ	k^a
$BPTI_{MDman}$	−2.00E − 2	1.50E − 1	0.00E + 0	2.00E − 2	1.50E − 1	4.00E + 1
$BPTI_{MDfit}$	−6.18E − 2	5.11E − 2	−4.77E − 2	7.20E − 3	6.82E − 2	1.61E + 2
Il4	−1.60E + 0	8.47E − 1	4.29E − 1	7.54E − 1	7.12E − 1	1.14E + 0
BPTI Exp.	−4.32E + 0	5.99E − 2	−2.67E − 2	3.59E + 0	1.81E + 0	6.69E − 1
PH 80ms	−6.40E − 1	2.47E − 1	5.93E − 3	1.76E − 1	2.18E − 1	6.99E + 0
GB1	−6.61E − 1	1.79E − 1	1.23E − 1	2.70E − 1	4.53E − 1	4.89E + 0
Ubiquitin	−5.10E − 1	1.70E − 1	4.92E − 2	9.49E − 1	3.16E − 1	3.33E + 0

[a]β and β' are the asymptotes of the soft–square potential, σ' and σ the switchpoints between the harmonic and asymptotic regions, δ describes the deviation of the minimum of the potential from zero, and k is an overall scale factor necessary for fitting.

toward a small asymptotic value for values larger than zero. We also derived an optimal fit with a computer program (see Methods). The parameters for the manual and optimal fits are listed in Table II (lines BPTI MDman and BPTI MDfit).

Potentials Derived From X-Ray Crystal Structures

The conformational space sampled in a few nanoseconds of molecular dynamics does not include all motions that are relevant for an NOE spectrum. We therefore decided to use, in addition, pairs of NOE data sets and X-ray crystal structures to define error distributions. The definition of the X-ray crystal structure as the "true" structure is obviously not strictly justified since there may be real differences between structures in solution and in the crystal. However, significant differences between NMR and X-ray crystal structures are generally limited to some surface residues and loops directly affected by crystal packing and would not change the statistics of interproton distances within the molecule drastically.

Figure 1 shows the negative logarithms of the histograms of distance differences for the five experimental data sets, together with the fitted potentials. The derived parameters are the most probable parameters, derived with a Bayesian sampling procedure (see Methods). The shape parameters for the derived potentials PDF_{Xray} are rather different but the potentials have common features. They show an almost perfect linear behavior for distances smaller than the target distance r_{exp}. The harmonic region is widest for the potential derived from the experimental BPTI data; the potential derived from the trajectory has the narrowest harmonic well around the minimum and the smallest slope for the asymptotic regions for distances larger than r_{exp}. This is due to the fact that this data set contains many NOE-derived distances close to the average distance, but also a large number of outliers.

The parameters for the soft-square potential are summarized in Table II. The parameters give a concise description of the deviations from a normal distribution in the data set: the larger the switch points σ and σ', the larger is the harmonic region of the potential, and therefore the part of the data set that could be well described by a normal distribution; the smaller the asymptotes, the larger is the fraction of outliers in the data set.

Structure Calculations With the PDF_{MDman} Potential

We first performed calculations for all data sets (see Table I) with the manually fitted symmetrized potential PDF_{MDman}. With FBHW potentials, the actual weight on the distance restraint term has little influence on the result, provided that the bounds are geometrically consistent and that the weight is large enough to ensure convergence.[15,38] In contrast, for a potential without a flat bottom (Equation 3) the energy can only vanish for a completely noise-free distance restraint set, and in absence of competition with the energy parameters. Hence, the weight on the experimental data will be more important. We therefore calculated ensembles of 100 structures with weights on the distance restraint term ranging from 2 to 512 and performed tenfold complete cross-validation[15] to identify an optimal weight (see Methods).

For the experimental data sets, we performed the same series of calculations also with FBHW potentials, using the same NOE distance bounds as in the original structure determination, however with complete cross-validation. For ubiquitin, we used only the 1444 nonredundant distance bounds, choosing the tightest bounds whenever several values were present in the original file.

In all cases, the data for the calculations comprised NOE-derived distances only. We did not use the torsion angle, coupling constant, hydrogen bonding, or residual dipolar coupling data that had been used for the original structure determination. Only the PH domain and the BPTI trajectory data set contained also structurally irrelevant NOEs (e.g., NOEs corresponding to covalently fixed distances). These had been removed from the restraint lists for calculating the originally reported structures since it is usually argued that they do not add any structural information. While one would not expect a strong influence of these restraints on structures calculated with the PDF_{MDman} potential, deviations from covalently fixed distances can be significant with this potential. We think, however, that they should always be kept as an integral part of the data submitted to the data bases since they can serve to assess data quality.

BPTI Model Data

The principal advantage of the molecular dynamics model system is the well-defined unbiased reference structure, which enables us to measure the accuracy of the calculated structures. This reference had been obtained by refinement into an electron density calculated from the same molecular dynamics trajectory that also served to calculate cross-relaxation rates.[10] In order to obtain low energy structures in a previous study[10] using FBHW potentials, about a third of the restraints had to be discarded since they were substantially affected by internal dynamics. Choosing error bounds wide enough to get low-energy structures led to a systematic deviation of the calculated structure from the reference structure of around 1.5 Å.

Figure 2 shows RMS differences for the three best structures for each working set (a total of 30 structures) for different weights on the NOE term. The weight on the NOE distance potential has an influence on all shown measures: the RMS difference from experimental distances (RMS_{work}); the cross-validated RMS difference from experimental distances (RMS_{CV}); the accuracy ($RMSD_{X\text{-}ray}$), measured as the average RMS distance between the calculated structures and the reference structure; the precision ($RMSD_{ave}$), measured as the average RMS distance between the calculated structures and the average structure.

RMS_{work} improves over the whole range of weights (Fig. 2, top panel). As is often observed with cross-validation, the curve for RMS_{CV} does not show a pronounced minimum but reaches a plateau value after an initial decrease. The beginning of the plateau corresponds roughly to the minimum in RMS distance to the reference structure.

Accuracy and precision improve steadily until minimal values are reached at data weights around 16–32, with values of 0.48 Å for accuracy, $RMSD_{X\text{-}ray}$, and 0.26 Å for precision, $RMSD_{ave}$ (Fig. 2, bottom panel). For larger weights, both accuracy and precision deteriorate slightly. This result is significantly better than that obtained in the previous study.[10] In contrast to the previous study, however, we did not have to discard or modify any distance restraint to obtain the present result.

Experimental Data Sets

For the experimental data sets, the distance to the X-ray crystal structure, $RMSD_{X\text{-}ray}$, does not necessarily provide a measure of accuracy. The X-ray crystal structure is, however, a useful reference point. We note that for the examples chosen, the structures get closer to the X-ray structure with further refinement and/or inclusion of additional data (compare, for example, Powers et al.[19,39]).

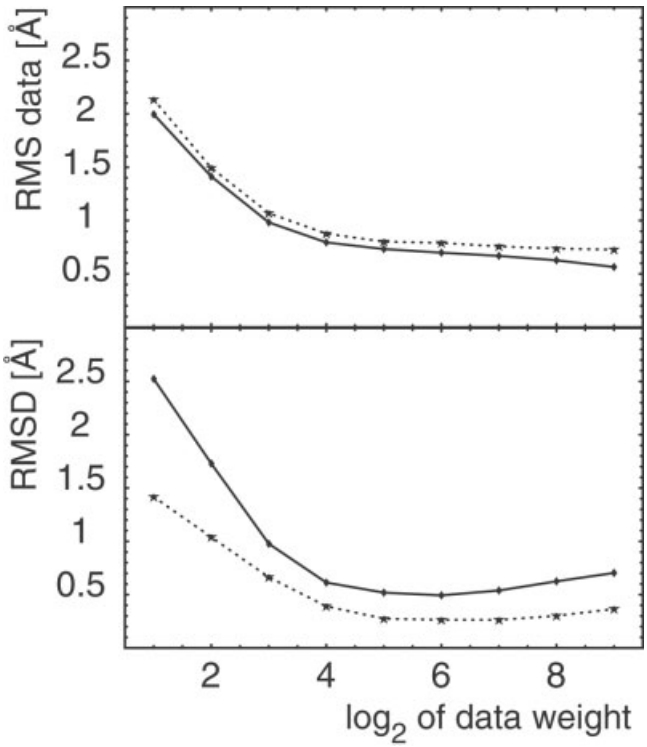

Fig. 2. RMS differences for calculations with the PDF_{MDman} potential. Top panel: RMS difference from experimental distances: (RMS_{work}; solid line), (RMS_{CV}; dotted line). Bottom panel: RMS distances from the reference structure ($RMSD_{X\text{-}ray}$; solid line) and from the average structure ($RMSD_{ave}$; dotted line)

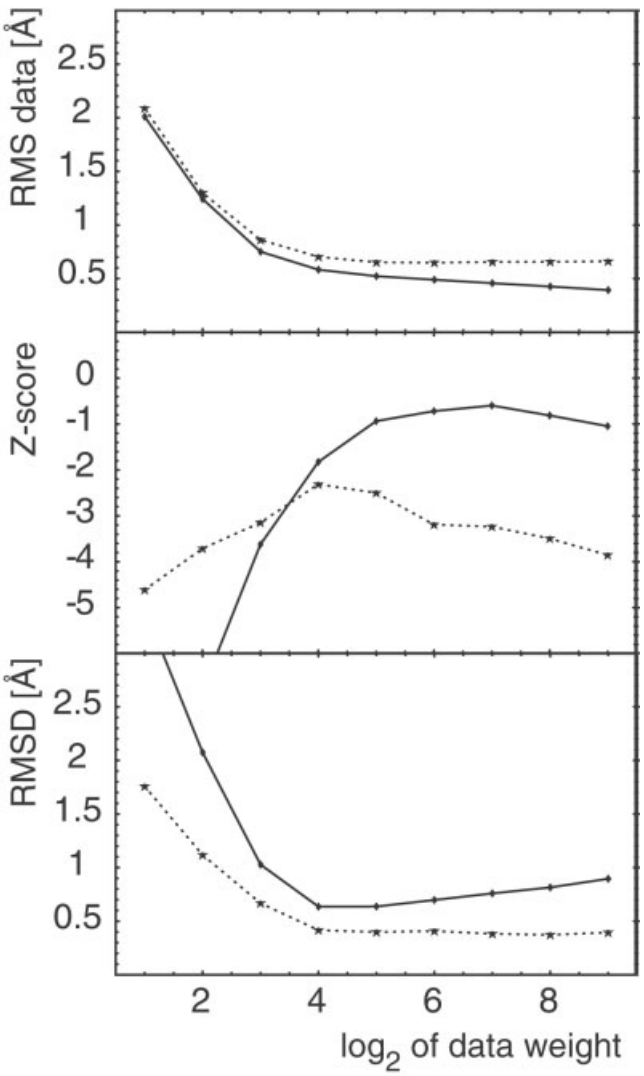

Fig. 3. RMS differences and quality indices for ubiquitin for calculations with the PDF_{MDman} potential. Top panel: RMS difference from data RMS_{work} (solid line), cross-validated RMS difference from data RMS_{CV} (dotted line). Middle panel: QUACHK (solid line) and RAMCHK (dotted line). Bottom panel: RMS distance from the X-ray crystal structure $RMSD_{X\text{-}ray}$ (solid line) and RMS distance from average structure $RMSD_{ave}$ (dotted line).

Figure 3 shows, for Ubiquitin, the RMS differences and the WhatCheck[40] quality indices QUACHK and RAMCHK, as a function of the data weight used in the calculation. Table III gives an overview over a larger selection of quality indices calculated with the programs WhatCheck and Prosa,[41] for the optimal NOE weight determined by cross-validation, compared to FBHW potentials with standard parameters. The structures calculated with PDF_{MDman} are of similar or better quality than those calculated with FBHW potentials. In particular the Ramachandran quality RAMCHK is more than a standard deviation better with the PDF_{MDman} potential. The only criterion that becomes significantly worse is C12CHK, which measures the Z-score of the χ_1/χ_2 correlation. This is not surprising, since dihedral angle potentials are used in the force field,[32] and for a FBHW potential, there is no competition between the distance restraint and the dihedral angle potential. The positive value of the Z-score for C12CHK indicates that the structures are in this respect better than a good X-ray crystal structure. No distortions are observed in the covalent geometry: the RMS Z-scores for bonds, bond angles and planarity indicate a very "tight" geometry, and there are the same number of nonbonded overlaps as with FBHW potentials (Table III).

Table IV summarizes the results for the five experimental data sets. In all cases, $RMSD_{X\text{-}ray}$ is smaller for the ensembles calculated with the PDF_{MDman} potential than for those calculated with the FBHW potential. Whereas the closeness to the X-ray crystal structure cannot as such be taken as proof of increased accuracy obtained with the PDF_{MDman} potential, this is remarkable. Note that less human intervention is necessary for the PDF_{MDman} potential—all corrections to the upper bounds had been removed (see above and Methods).

Also $RMSD_{ave}$ decreases, but the effect is sometimes significantly stronger for $RMSD_{X\text{-}ray}$ than for $RMSD_{ave}$. This is particularly evident in Figure 4. Precision seems to overestimate accuracy to a less extent when using the PDF_{MDman} potential.

Concerning the RMS differences from data, RMS_{work} and RMS_{CV}, one must note that they are necessarily calculated in a very different manner, since for the FBHW potential they measure the differences from the bounds U

TABLE III. Quality Parameters for Ubiquitin

Quality check	PDF_{MDman}	FBHW
QUACHK	−1.37E + 0	−1.62E + 0
NQACHK	−1.94E + 0	−2.03E + 0
RAMCHK	−2.38E + 0	−3.37E + 0
C12CHK	3.39E − 1	1.02E + 0
BBCCHK	6.98E − 2	4.94E − 2
BNDCHK	1.71E − 1	1.86E − 1
ANGCHK	3.11E − 1	3.13E − 1
OMECHK	8.95E − 2	6.18E − 2
PLNCHK	9.31E − 2	6.86E − 2
HNDCHK	1.75E − 1	1.71E − 1
INOCHK	9.93E − 1	1.00E + 0
BMPCHK	7.83E + 0	7.30E + 0
BH2CHK	1.09E + 1	1.31E + 1
BA2CHK	1.13E + 0	1.53E + 0
PROSA	−1.71E + 0	−1.70E + 0

Z–scores: QUACHK, 1st generation side-chain packing quality; NQACHK, 2nd generation side-chain packing quality; RAMCHK, Ramachandran quality; C12CHK, $\chi_1 - \chi_2$ quality; BBCCHK, backbone conformation quality. RMS Z–scores: BNDCHK, covalent bonds; ANGCHK, bond angles; OMECHK, ω planarity; PLNCHK, side–chain planarity; HNDCHK, chirality; INOCHK, inside–outside distribution. Nonbonded interactions: BMPCHK, number of interatomic bumps; BH2CHK, number of unsatisfied H–bond donors; BA2CHK, number of unsatisfied H–bond acceptors; PROSA, mean energy calculated with the program PROSA.41

and L (i.e., RMS_{work} should be zero), whereas for the PDF_{MDman} potential they measure the difference from the estimated target distances r_{exp} (and RMS_{work} could be zero only for a completely noise-free data set). It is interesting that in all cases the differences between RMS_{CV} and RMS_{work} are significantly smaller for PDF_{MDman} potential, indicating the information loss due to the FBHW potential.

Sensitivity to Noise in the Data

To test the influence of larger errors in the experimental distances, we derived data sets for Ubiquitin with noise of random and systematic nature. In one data set ($Ubiquitin_{stereo}$), we used incorrect assignments for all stereospecifically assigned protons. In a second data set ($Ubiquitin_{ran}$), we replaced the experimental distances by random values drawn from a uniform distribution between 1.8 and 4.5 Å, thus removing all distance dependence. A further series of data sets was derived by adding random numbers of increasing value to the experimental distances (random numbers within [−0.2,0.2] for the least noisy set; within [−1.6,1.6] for the noisiest set).

For the Ubiquitin data set with incorrect stereospecific assignments, RMS_{work} and RMS_{CV} increase, directly reflecting the decreasing quality of the data (Table IV). $RMSD_{X\text{-}ray}$ and $RMSD_{ave}$ rise only little, indicating that the incorrect restraints influence the structure only slightly.

For the Ubiquitin data set with randomized distances, the optimal value for the weight with the NOEs determined from complete cross-validation is shifted from 16 to 32 (see Table IV). As can be expected, this much more drastic addition of noise causes larger increases of RMS_{work} and RMS_{CV}. $RMSD_{ave}$ increases substantially, and the distance to the X-ray crystal structure rises.

A better idea of the dependency of quality indices on noise is given in Figure 5, showing a largely linear increase of RMS and RMSD values on the amount of noise added to the experimental distances. These results indicate that the RMS values, in particular RMS_{CV}, are not only indicators of data completeness[15] but also of the data consistency.

The influence of data quality on structure is also evident in the "sausage" plot (Fig. 6), where $RMSD_{ave}$ is indicated by the thickness, and the color indicates local data consistency from blue (lowest RMS_{CV}) to yellow (highest RMS_{CV}).

Recalculation of the Structures with the X-ray derived potentials and influence of potential parameters

We recalculated all structures with a potential derived from the NOE data and the X-ray crystal structure of the same protein (see Table II). This calculation is therefore biased towards the X-ray crystal structure and would be expected to produce the smallest $RMSD_{X\text{-}ray}$.

The results of structure calculations with these potentials are compiled in Table V and are not significantly different from those with the potential estimated "by eye" from the MD trajectory. In general, it is impractical to use an X-ray crystal structure to derive the potential for every new NMR structure. It would also be questionable to bias the structure calculation in such a way toward the X-ray crystal structure.

To test the influence of the exact definition of the potential more systematically, we calculated for bptiW structures with a series of potentials with different shapes. Figure 7 shows the $RMSD_{ave}$ and $RMSD_{X\text{-}ray}$ for nine different potential shapes, all symmetric, with the parameters $\sigma = \sigma'$ set to 0.1, 0.9, and the parameters $\beta = \beta'$ set to $0.01\ 2^n$, $n = 1,\ldots,9$. For each of these potentials, the weights were varied from 0.5 to 128. Figure 7 plots RMS_{CV} against RMS_{work}. With larger harmonic regions and steeper gradients of the asymptotes, RMS_{CV} can be reduced to very low values. In contrast, RMS_{CV} converges to approximately the same value for all potential shapes.

For each of these potential shapes, the optimal weight for the structural statistics shown in Figure 4 was determined approximately, as described above, by choosing a value close to the elbow region. Optimal weights ranged between 1 for the potential with the widest harmonic range and steepest asymptote, and 64 for the one with the narrowest harmonic range and flattest asymptote. The weight of 32 obtained for the PDF_{MDman} potential is only valid for that particular potential shape ($\sigma = \sigma' = 0.15$; $\beta = \beta' = 0.02$).

Again, we find that the exact choice of parameters did not significantly affect RMS_{CV}, $RMSD_{ave}$, and $RMSD_{X\text{-}ray}$. We advocate, however, the use of a potential with a narrow harmonic region and small gradient of the asymptote to reduce the influence of large outliers.

TABLE IV. RMS From Data and RMSD Values

Protein	$RMSD_{ave}$	$RMSD_{X\text{-}ray}$	RMS_{work}	RMS_{CV}	$RMS_{CV} - RMS_{work}$	Relative difference
FBHW potential						
Il4	2.83E + 0	3.57E + 0	5.32E − 2	7.81E − 1	8.72E − 1	7.23E − 1
BPTI	5.59E − 1	9.17E − 1	4.66E − 2	4.06E − 1	3.54E − 1	6.98E − 1
PH	1.31E + 0	1.57E + 0	1.33E − 1	6.46E − 1	5.13E − 1	5.85E − 1
GB1	4.97E − 1	1.09E + 0	1.93E − 2	2.05E − 1	1.86E − 1	8.28E − 1
Ubiquitin	4.36E − 1	6.73E − 1	2.76E − 2	1.85E − 1	1.58E − 1	7.41E − 1
PDF_{MDman} potential						
Il4	1.48E + 0	2.37E + 0	1.17E + 0	1.59E + 0	4.21E − 1	1.53E − 1
BPTI	4.75E − 1	6.83E − 1	8.39E − 1	1.08E + 0	2.42E − 1	1.26E − 1
PH	9.17E − 1	1.34E + 0	1.31E + 0	1.55E + 0	2.43E − 1	8.51E − 2
GB1	4.42E − 1	7.79E − 1	8.36E − 1	9.87E − 1	1.51E − 1	8.29E − 2
Ubiquitin	4.16E − 1	6.35E − 1	5.87E − 1	7.01E − 1	1.14E − 1	8.87E − 2
$Ubiquitin_{stereo}$	5.11E − 1	7.27E − 1	7.62E − 1	9.23E − 1	1.61E − 1	9.56E − 2
$Ubiquitin_{ran}$	7.89E − 1	1.21E + 0	1.55E + 0	1.69E + 0	1.37E − 1	4.23E − 2

For the FBHW potential, a weight of 64 was used for all data sets. For the PDF_{MDman} potential, the optimal weight was 32 for the data sets Il4, BPTI, PH, GB1, and $Ubiquitin_{ran}$, and 16 for Ubiquitin and $Ubiquitin_{stereo}$. Relative difference is defined as $(RMS_{CV} - RMS_{work})/(RMS_{CV} + RMS_{work})$.

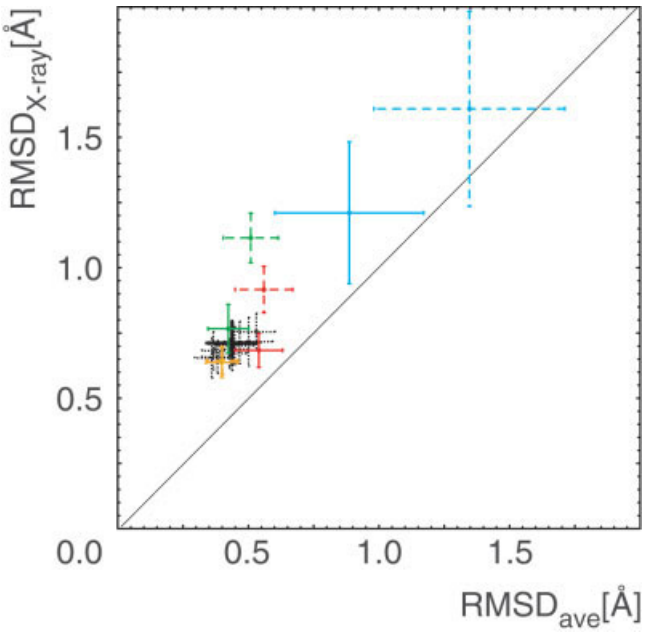

Fig. 4. RMS distance from the X-ray crystal structure against RMS distance from the average structure. The bars indicate one standard deviation. Results for FBHW potentials in dashed lines, for PDF_{MDman} in solid lines. PH domain (cyan), GB1 (green), BPTI (red), and Ubiquitin (orange). For BPTI, the results of a calculation with the parameters derived from the X-ray crystal structure are indicated as a thick black line, and results with a series of different potential parameters (see text) as dotted black lines.

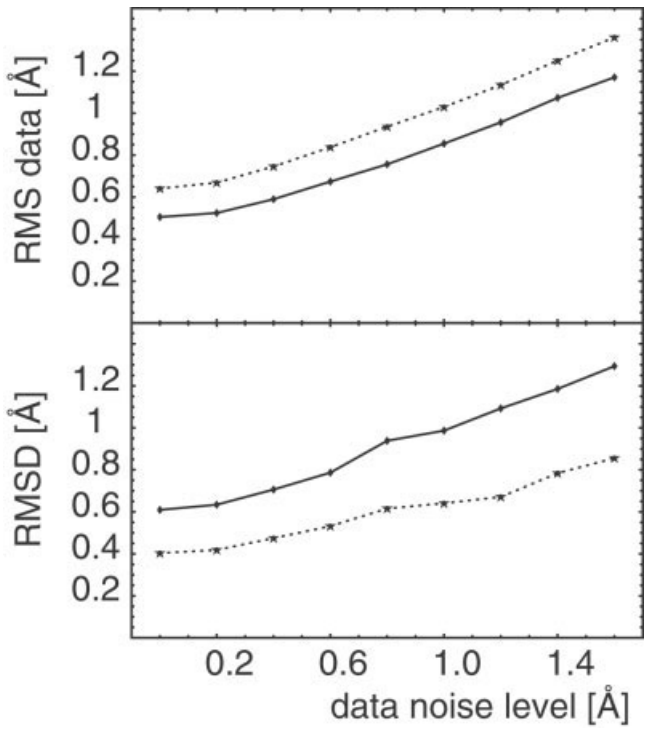

Fig. 5. RMS and RMSD values plotted against the level of added random noise added to the experimental distances for ubiquitin, calculated with the PDF_{MDman} potential and a data weight of 32. The values indicate the interval from which the random number was drawn; 0.2 indicates the interval [−0.2,0.2] and so forth. Top panel: RMS_{work} (solid line) and RMS_{CV} (dotted line). Bottom panel: $RMSD_{X\text{-}ray}$ (solid line) and $RMSD_{ave}$ (dotted line).

Relevance of Figures of Merit for NMR Structures

For the present analysis, we were mostly interested in intrinsic quality indicators (RMS from data and RMSD values). The precision of an NMR structure ensemble is one of the usual indicators of its quality and is also used to predict its accuracy.[42] However, precision is to a large extent influenced by the width of error bounds, which is determined empirically by subjective criteria. In particular, modifying individual error bounds to get rid of a few violations has a profound influence on the precision.[10] This reduces the usefulness of this criterion.

With the restraint potential described in this paper, choosing lower and upper bounds is replaced by choosing an optimal weight for the NOE restraints. For a potential

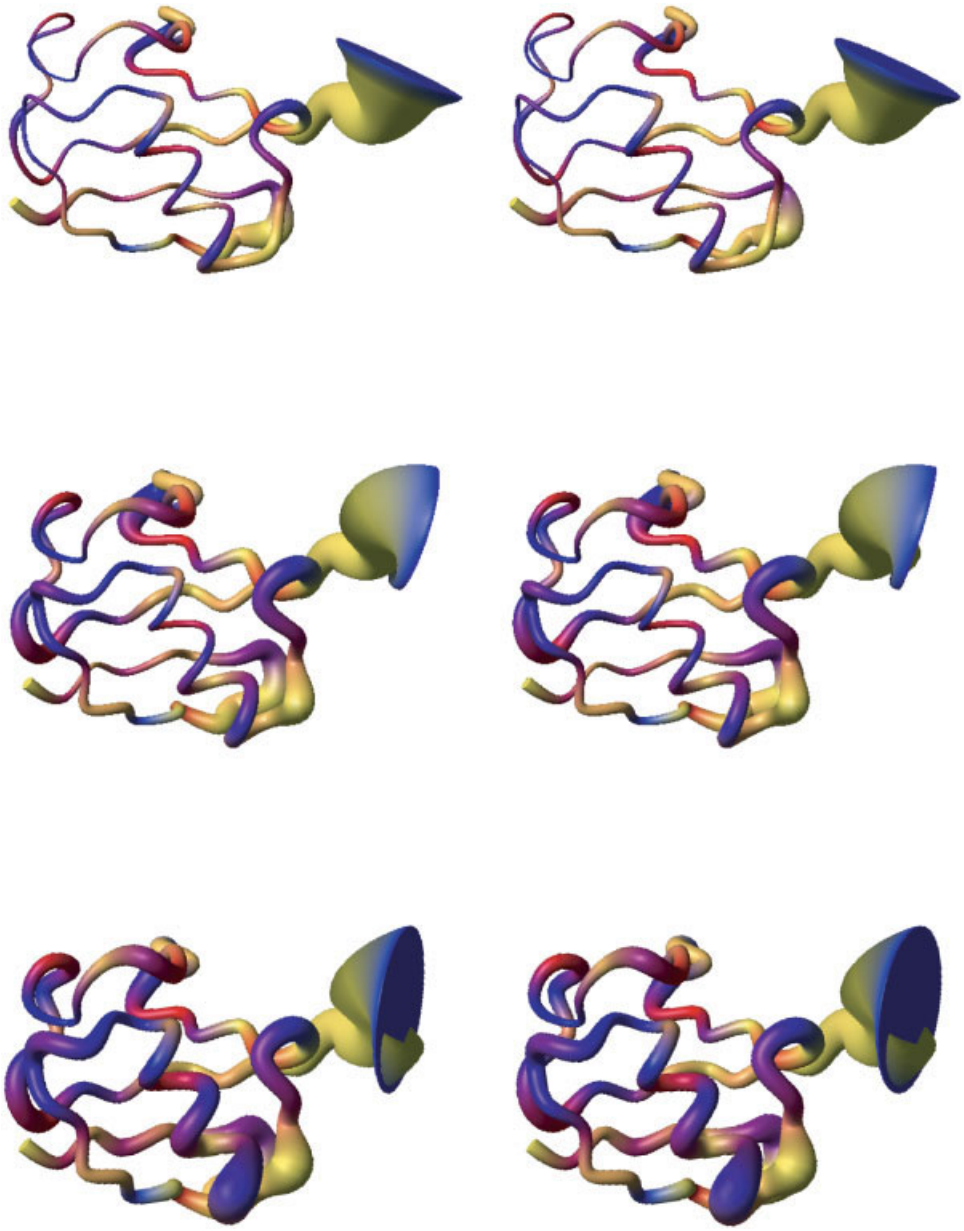

Fig. 6. Stereo views of average structures for ubiquitin calculated with random noise added to the experimental distances (PDF_{MDman} potentialdat and data weight 32). Top panel: no added noise. Middle panel: noise added in the interval [−0.8,0.8]. Bottom panel: noise added in the interval [−1.6,1.6]. The color indicates the local RMS_{CV}. Blue: lowest RMS_{CV}; Yellow: largest RMS_{CV}; Red: average. The plot was generated with MolMol.[37]

TABLE V. RMS From Data and RMSD Values for Structures Calculated with the $PDF_{X\text{-}ray}$ Potentials

Protein	$RMSD_{ave}$	$RMSD_{X\text{-}ray}$	RMS_{work}	RMS_{CV}	$RMS_{CV} - RMS_{work}$	Relative difference
Il4	1.95e + 0	2.63e + 0	9.27e − 1	1.61e + 0	6.89e − 1	2.70e − 1
BPTI	4.38e − 1	7.10e − 1	5.82e − 1	8.66e − 1	2.84e − 1	1.96e − 1
PH	8.90e − 1	1.36e + 0	1.11e + 0	1.43e + 0	3.16e − 1	1.25e − 1
GB1	4.32e − 1	8.31e − 1	7.58e − 1	8.91e − 1	1.32e − 1	8.03e − 2
Ubiquitin	2.89e − 1	6.17e − 1	3.89e − 1	5.82e − 1	1.93e − 1	1.99e − 1
$BPTI_{MD}$	2.69e − 1	5.26e − 1	7.89e − 1	8.62e − 1	7.25e − 2	4.40e − 2

Relative difference is defined as $(RMS_{CV} - RMS_{work})/(RMS_{CV} + RMS_{work})$.

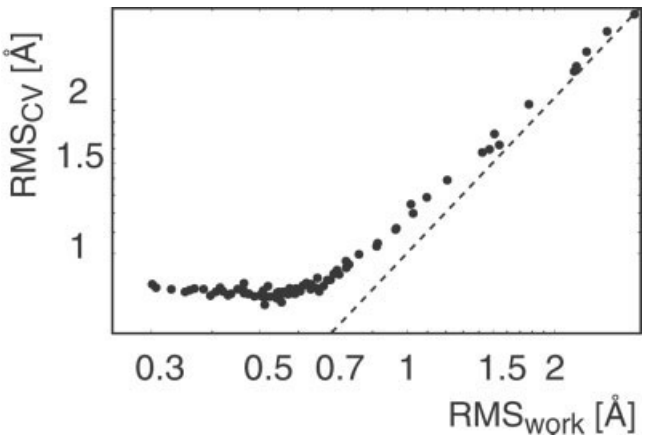

Fig. 7. RMS_{CV} against RMS_{work} for the nine potential shapes tested for all weights for BPTI (from 0.5 to 128). The dotted line indicates $RMS_{CV} = RMS_{work}$

of least-squares type (such as a harmonic potential), the optimal weight depends on another quantity that is unknown before the structure calculation, the data quality.[24,43,44] Cross-validation provides a well defined procedure to define the weight.[15] The problem is thus easier to solve and less biased than the determination of bounds. For the potential PDF_{MDman}, the optimal weight determined by cross-validation is stable and depends only to a small degree on data quality and data density, since the same weight was determined by cross-validation for very different data sets.

In particular RMS_{CV}, with a potential as proposed in this paper, could be a useful quality index. More experience with a larger number of data sets is necessary to establish "normal" values for an NMR structure.

The difference between RMS_{work} and RMS_{CV} is substantially higher for structures calculated with the FBHW potential than for those calculated with the PDF_{MDman} potential (Fig. 3 and Table IV). This is a clear indication of a loss of information when FBHW potentials are used.

Advantages of the PDF Restraint Potential

We could establish several advantages of the novel restraints:

1. We do not need to derive distance bounds to calculate NMR structures, but can use the estimated distances. This removes the arbitrary choices involved in setting these bounds and permits the use of the data in a more direct way.
2. Outliers have a small influence on the derived structures, since the gradient of the potential decreases with increasing violation.
3. Individual bounds do not need to be removed or "recalibrated." The difference of the distance in the structure from the experimentally measured distance becomes an important result, rather than hidden by increasing bound widths.
4. Neither do we have to increase the width of all bounds to obtain low energy structures. This avoids loss of information content in the data.
5. The type of restraint potential is consistent with the potential used for other data (e.g., coupling constants, residual dipolar couplings), which do not employ bounds.
6. The optimal value of the weight on the distance restraint term is not very critical and can be easily determined by complete cross-validation. Around the optimal weight, calculations are not very sensitive to the precise value, and 32 seems a good starting value for the potential PDF_{MDman}.
7. The quality of the structures, in terms of the precision $RMSD_{ave}$, automatically decreases with the data quality, making $RMSD_{ave}$ a meaningful number.

Our study documents the powers and limitations of cross-validation. It allows the determination of the optimal value for the relative weight between data and prior knowledge. However, in most cases we were not able to find a clear minimum and had to use the elbow region of RMS_{CV} to define an optimal value.

Conclusion

In this paper, we have used error-tolerant restraint potentials for distances derived from NOEs. In contrast to practically all previous structure calculations based on NOE derived interproton distances, the potential does not have a flat region between lower and upper bounds but a unique minimum. It is difficult to compare our conclusions to previous work using a similar potential[12] since that work suffered from insufficient convergence of the calculation algorithm used. In our work, the RMS distance from the average structure is similar to the values obtained with an FBHW potential or smaller.

The present study proposes to replace the standard flat-bottom harmonic-walls potential for NOE-derived distance restraints by a potential derived from an error distribution in the distances. We could conclusively demonstrate that the new potential not only has conceptual advantages, but is also easier to use in practice (since there is no need to "recalibrate" individual distance restraints), and it produces structures of similar or better quality. The concept presented in this paper will also have advantages for automated procedures such as ARIA[6,45] or CANDID,[46] since one of the difficulties there is the automated setting of bounds, and the treatment of small consistent violations.

It is common practice to include distance restraints that are difficult to calibrate (because of overlap) as very generous upper bounds. We note that also in the present paper, some experimental data sets contained such upper distance limits (e.g., GB1 and BPTI). We simply converted them into a weak NOEs (by the automatic conversion), and calibrated them to obtain distance estimates. The results we obtained with this procedure indicate that the exact value does not seem to matter very much in terms of structural distortions, but will simply give locally a worse fit to the data. If possible, we suggest using ambiguous distance restraints for overlapped peaks. The original aim

of this study was to treat NOE-derived distance restraints in an optimal way, by using error distributions derived from the data themselves. To fit the error distributions, we propose a potential with a total of six parameters, if one includes the overall weight and the deviation of the minimum position. During the study, we realized that the results were rather tolerant to the exact values of these parameters, and we used a symmetric form with only three parameters (PDF_{MDman}) for most of the calculations. Although it is somewhat unsatisfactory to have adjustable parameters, we note that the standard structure calculation concept using bounds has two adjustable parameters per restraint, the lower and upper bounds. The proposed new approach is therefore less prone to human bias than the standard approach, and should result in more reliable structures from NMR data.

We have recently introduced an entirely probabilistic strategy for structure calculation from NMR data,[47] using stochastic sampling and not minimization. The error model used for distances in that work is the so-called log-normal distribution. The potential corresponding to the log-normal distribution resembles the potentials used in this paper. The present work shows that such errror-tolerant potentials with a unique minimum have advantages also in standard minimization approaches.

ACKNOWLEDGMENTS

M.N. thanks Jeff Hoch for stimulating discussions, and Thérèse Malliavin for a critical reading of the manuscript.

REFERENCES

1. Blumenthal LM. Theory and applications of distance geometry. New York: Chelsea; 1970.
2. Havel T, Wüthrich K. A distance geometry program for determining the structures of small proteins and other macromolecules from nuclear magnetic resonance measurements of intramolecular $^1H1^1H$ proximities in solution. Bull Math Biol 1984;46:673–698.
3. Braun W. Distance geometry and related methods for protein structure determination from NMR data. Quart Rev BioPhys 1987;19:115–157.
4. Brunger AT, Nilges M. Computational challenges for macromolecular structure determination by X–ray crystallography and solution NMR spectroscopy. Quart Rev BioPhys 1993;26:49–125.
5. Güntert P. Structure calculation of biological macromolecules from NMR data. Quart Rev Biophys 1998;31:145–237.
6. Nilges M, O'Donoghue SI. Ambiguous NOEs and automated NOE assignment. Prog Nucl Magn Reson Spectrosc 1998;32:107–139.
7. Brunger A. X–plor. A System for X-ray crystallography and NMR. New Haven, CT: Yale University Press; 1992.
8. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM. The Xplor-NIH NMR molecular structure determination package. J Magn Reson 2003;160:65–73.
9. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54:905–921.
10. Chalaoux FR, O'Donoghue SI, Nilges M. Molecular dynamics and accuracy of NMR structures: effects of error bounds and data removal. Proteins 1999;34:453–463.
11. Spronk CA, Nabuurs SB, Bonvin AM, Krieger E, Vuister GW, Vriend G. The precision of NMR structure ensembles revisited. J Biomol NMR 2003;25:225–234.
12. Kim Y, Prestegard JH. A dynamic model for the structure of acyl carrier protein in solution. Biochemistry 1989;28:8792–8797.
13. Brunger A, Clore GM, Gronenborn AM, Karplus M. Three-dimensional structure of proteins determined by molecular dynamics with interproton distance restraints: application to crambin. Proc Natl Acad Sci USA 1986;83:3801–3805.
14. Schneider TR, Brunger AT, Nilges M. Influence of internal dynamics on accuracy of protein NMR structures: derivation of realistic model distance data from a long molecular dynamics trajectory. J Mol Biol 1999;285:727–740.
15. Brunger AT, Clore GM, Gronenborn AM, Saffrich R, Nilges M. Assessing the quality of solution nuclear magnetic resonance structures by complete cross-validation. Science 1993;261:328–331.
16. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov I, Bourne P. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
17. Wüthrich K, Billeter M, Braun W. Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton–proton distance constraints with nuclear magnetic resonance. J Mol Biol 1983;169:949–961.
18. Nilges M. A calculation strategy for the structure determination of symmetric dimers by ^1H NMR. Proteins 1993;17:297–309.
19. Powers R, Garrett DS, March CJ, Frieden EA, Gronenborn AM, Clore GM. Three-dimensional solution structure of human interleukin-4 by multidimensional heteronuclear magnetic resonance spectroscopy. Science 1992;256:1673–1677.
20. Berndt KD, Güntert P, Orbons LP, Wüthrich K. Determination of a high-quality nuclear magnetic resonance solution structure of the bovine pancreatic trypsin inhibitor and comparison with three crystal structures. J Mol Biol 1992;227:757–775.
21. Nilges M, Macias M, O'Donoghue SI, Oschkinat H. Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β–spectrin. J Mol Biol 1997;269:408–422.
22. Gronenborn AM, Filpula DR, Essig NZ, Achari A, Whitlow M, Wingfield PT, Clore GM. A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein g. Science 1991;253:657–661.
23. Cornilescu G, Marquardt JL, Ottiger M, Bax A. Validation of protein structure from anisotropic carabonyl chemical shifts in a dilute liquid crystalline phase. J Am Chem Soc 1998;120:6836–6837.
24. Hoch JC, Stern AS, Conolly PJ. Fitting protein structures to experimental data: lessons from before your mother was born. In: Jardetzky O, Lefevre J-F, editors. Protein dynamics, function, and design, New York: Plenum; 1998.
25. Gallagher T, Alexander P, Bryan P, Gilliland GL. Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. Biochemistry 1994;33:4721–4729.
26. 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 B Biol Crystallogr 1983;39:480–485.
27. Wlodawer A, Pavlovsky A, G. A. Crystal structure of human recombinant interleukin–4 at 2.25 Å resolution. FEBS Lett 1992;309:59–64.
28. Hyvönen M, Macias MJ, Nilges M, Oschkinat H, Saraste M, Wilmanns M. Structure of the binding site for inositol phosphates in a PH domain. EMBO J 1995;14:4676–4685.
29. Vijay-Kumar S, Bugg CE, Cook WJ. Structure of ubiquitin refined at 1.8 Å resolution. J Mol Biol 1987;194:531–544.
30. Jeffreys H. An invariant form for the prior probability in estimation problems. Proc Roy Soc A 1946;186:453.
31. Jaynes ET. Probability theory: the logic of science. Cambridge UK: Cambridge University Press; 2003.
32. Linge JP, Williams MA, Spronk CA, Bonvin AM, Nilges M. Refinement of protein structures in explicit solvent. Proteins 2003;50:496–506.
33. Engh RA, Huber R. Accurate bond and angle parameters for x-ray structure refinement. Acta Crystallogr A 1991;47:392–400.
34. Hendrickson WA. Stereochemically restrained refinement of macromolecular structures. Meth Enzymol 1985;11:252–270.
35. Nilges M, Clore GM, Gronenborn AM. A simple method for delineating well-defined and variable regions in protein structures determined from interproton distance data. FEBS Lett 1987;219:17–21.
36. Wolfram Research, I. Mathematica. Champaign, Illinois, 5.1 edition, 2004.

37. Koradi R, Billeter M, Wüthrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 1996;14:51–55.
38. Clore GM, Robien MA, Gronenborn AM. Exploring the limits of precision and accuracy of protein structures determined by nuclear magnetic resonance spectroscopy [published erratum appears in J Mol Biol 1994;237:243]. J Mol Biol 1993;231:82–102.
39. Powers R, Garrett DS, March CJ, Frieden EA, Gronenborn AM, Clore GM. The high-resolution, three-dimensional solution structure of human interleukin-4 determined by multidimensional heteronuclear magnetic resonance spectroscopy. Biochemistry 1993;32:6744–6762.
40. Hooft RW, Vriend G, Sander C, Abola EE. Errors in protein structures. Nature 1996;381:272.
41. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. Proteins 1993;17:355–362.
42. Gronenborn AM, Clore GM. Structures of protein complexes by multidimensional heteronuclear magnetic resonance spectroscopy. Crit Rev Biochem Mol Biol 1995;30:351–385.
43. Habeck M, Rieping W, Nilges M. A new principle for macromolecular structure determination. In: Erickson G, Zhai Y, editors. 23rd international workshop on Bayesian inference and maximum entropy methods in science and engineering. Melville, NY: American Institute of Physics; 2004. p 157–166.
44. Rieping W, Habeck M, Nilges M. Structure determination from heterogeneous NMR data. In Fischer R, Press R, von Toussaint U, editors. 24th international workshop on Bayesian inference and maximum entropy methods in science and engineering. Melville, NY: American Institute of Physics; 2004. p 268–275.
45. Habeck M, Rieping W, Linge JP, Nilges M. NOE assignment with ARIA 2.0—the nuts and bolts. In: Downing AK, editor. Protein NMR Techniques. Totowa, NJ: Humana Press; 2004.
46. Herrmann T, Guntert P, Wüthrich K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 2002;319:209–227.
47. Rieping W, Habeck M, Nilges M. Inferential structure determination. Science 2005;309:303–306.