

Account / Revue

## Probabilistic structure calculation

Michael Nilges<sup>a,\*</sup>, Michael Habeck<sup>b,c</sup>, Wolfgang Rieping<sup>d</sup>

<sup>a</sup> *Unité de bio-informatique structurale, CNRS URA 2185, Institut Pasteur, 25–28 rue du Docteur-Roux, F-75015 Paris, France*

<sup>b</sup> *Max-Planck-Institute for Developmental Biology, Spemannstr. 35, D-72076 Tübingen, Germany*

<sup>c</sup> *Max-Planck-Institute for Biological Cybernetics, Spemannstr. 38, D-72076 Tübingen, Germany*

<sup>d</sup> *Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK*

Received 29 April 2007; accepted after revision 13 November 2007

Available online 21 February 2008

### Abstract

Molecular structures are usually calculated from experimental data with some method of energy minimisation or non-linear optimisation. Key aims of a structure calculation are to estimate the coordinate uncertainty, and to provide a meaningful measure of the quality of the fit to the data. We discuss approaches to optimally combine prior information and experimental data and the connection to probability theory. We analyse the appropriate statistics for NOEs and NOE-derived distances, and the related question of restraint potentials. Finally, we will discuss approaches to determine the appropriate weight on the experimental evidence and to obtain in this way an estimate of the data quality from the structure calculation. Whereas objective estimates of coordinates and their uncertainties can only be obtained by a full Bayesian treatment of the problem, standard structure calculation methods continue to play an important role. To obtain the full benefit of these methods, they should be founded on a rigorous Bayesian analysis. **To cite this article:** *M. Nilges et al., C. R. Chimie 11 (2008).*

© 2007 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

### Résumé

Pour calculer des structures moléculaires à partir des données expérimentales, on utilise habituellement une méthode de minimisation d'énergie, ou d'optimisation. Les objectifs majeurs d'un calcul de structure sont l'estimation de l'incertitude des coordonnées, et l'estimation d'une mesure de la qualité de l'ajustement aux données. Nous discutons des approches pour combiner d'une façon optimale l'information préalable et les données expérimentales, et le lien avec la théorie des probabilités. Nous analysons les statistiques appropriées pour les NOE et les distances dérivées, et les potentiels de contrainte. Finalement, nous discutons des approches pour un choix optimal du poids sur l'évidence expérimentale, pour obtenir de cette façon une évaluation de la qualité de données à partir du calcul de la structure. Des estimations objectives de coordonnées et de leurs incertitudes peuvent seulement être obtenues par un traitement complètement bayésien du problème. Néanmoins, les méthodes standard de calcul de structure

*Abbreviations:* BPTI, bovine pancreatic Trypsin inhibitor; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; PH, pleckstrin homology; PDB, protein data bank; RMS, root mean square; RMSD, mean RMS distance;  $\text{RMSD}_{\text{X-ray}}$ , mean RMS distance from X-ray crystal structure;  $\text{RMSD}_{\text{aver}}$ , mean RMS distance from average structure; I14, Interleukin 4; ARIA, ambiguous restraints for iterative assignments; CNS, crystallography and NMR system; SH3, src homology 3.

\* Corresponding author.

*E-mail address:* [nilges@pasteur.fr](mailto:nilges@pasteur.fr) (M. Nilges).

continuent à jouer un rôle important. Pour obtenir le plein avantage de ces méthodes, elles devraient être fondées sur une analyse Bayésienne rigoureuse. *Pour citer cet article : M. Nilges et al., C. R. Chimie 11 (2008).*

© 2007 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

**Keywords:** Bayes's theorem; NOE; Sampling; Optimisation

**Mots-clés :** théorème de Bayes ; NOE ; Echantillonnage ; Optimisation

## 1. Introduction: structures from noisy and incomplete data

Experimental data are rarely sufficient to determine the three-dimensional structure of a macromolecule by themselves but need to be complemented with prior physical information. Therefore, structure calculation is typically a search for conformations that simultaneously have a low physical energy  $E_{\text{phys}}(X)$  and minimise a cost function  $E_{\text{data}}(X)$  quantifying the disagreement between a structural model  $X$  and the data. This approach was first implemented for X-ray crystal structure determination as the minimisation of a hybrid energy [1,2]:

$$E_{\text{hybrid}}(X) = E_{\text{phys}}(X) + w_{\text{data}}E_{\text{data}}(X), \quad (1)$$

where the force field  $E_{\text{phys}}$  compensates the lack of data by imposing physical constraints on the structure. A target function of this form is widely used in macromolecular structure determination, notably from NMR data [3,4]. The weight  $w_{\text{data}}$  controls the contribution of the data relative to the force field. Already when introducing the hybrid energy concept, Jack and Levitt remarked that correct weighting of the data “is something of a problem” [1]. Its value can be critical. If it is too large, the contribution of the force field might be too small (overfitting of the data); if the weight is too small, the data contributes too little to define the structure (underfitting of the data).

Structure determination is an example of fitting of parameters (principally, the coordinates) to experimental data. “*To be genuinely useful, a fitting procedure should provide (i) parameters, (ii) error estimates of the parameters, and (iii) a statistical measure of a goodness of fit.*” [5].

Since reviews exist on the most used methods for obtaining the coordinates [2,6–10], we will in this paper analyse to which extent different approaches satisfy these three conditions. First, we will look at the relationship of the terms in Eq. (1) to probability theory. We will then discuss the restraint potential, which is related to the likelihood in Bayesian probability and plays a fundamental role in obtaining coordinates, their precision, and evaluating the goodness of fit. Finally, we will

look at the primordial question of correct weighting of experimental evidence in a structure calculation.

## 2. Bayes's rule and the hybrid energy

The hybrid energy function is motivated by maximising the posterior probability of a structure when prior information and experimental data are available. If we restrict the analysis to the molecular coordinates  $X$ , Bayes's theorem [11] yields the posterior probability distribution for the unknown coordinates:

$$p(X|D, I) \propto \pi(X|D, I)L(X|D, I). \quad (2)$$

The posterior  $p(X|D, I)$  factorises into two natural components (Fig. 1).

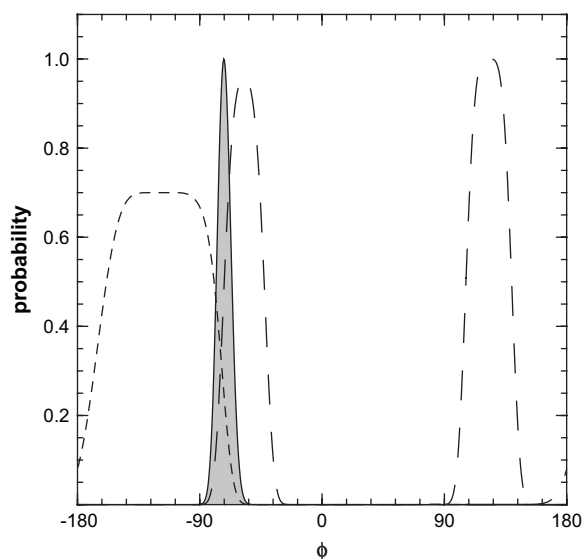


Fig. 1. Prior knowledge is incorporated in a natural way using the laws of probability theory. In the illustrated case, the prior knowledge (dotted line) is the probability to observe a particular torsion angle before any data are measured (for example, we know that the protein backbone torsion angle  $\phi$  is in most cases negative). The likelihood (dashed line) adds the knowledge obtained from the data: in our case, there are two peaks in the likelihood. The posterior probability (solid line) is obtained by multiplication of the prior probability and the likelihood and represents the total knowledge we have about the conformation.

1. The prior  $p(X|I)$  describes knowledge about general properties of biomolecular structures. At a given temperature of the system, the Boltzmann distribution  $\pi(X) \propto \exp(-\beta E_{\text{phys}}(X))$  is the least biasing prior distribution [11];  $\beta$  is  $(k_{\text{B}}T)^{-1}$ , with the Boltzmann constant  $k_{\text{B}}$  and temperature  $T$ .  $E_{\text{phys}}(X)$  is in general related to a molecular dynamics force field, although the analytical form and the weighting of different energy terms in the force field is often adapted to the structure calculation task. Molecular dynamics force fields are derived to match observed *dynamical* properties of a molecule, whereas a refinement force field should provide the probability of a distortion [12].
2. In the Bayesian context, the second term on the right hand side,  $L(D|X, I)$ , is called the likelihood. This is the probability of the data  $D$ , given the molecular structure  $X$ . To evaluate this probability, we need to have a model or a theory allowing us to calculate the data from a structure. It is important to note that these theories introduce further parameters (such as the parameters of the Karplus relationship between torsion angles and scalar coupling constants) that we cannot measure directly. We also need an error distribution, that is, a distribution of deviation of the measurements from our theory. If we repeat an experiment many times, the experimental values scatter around an average value that will approach the “true” value with the number of experiments. To derive the likelihood, we usually assume that the experimental data are distributed around the average value in the same way as they are distributed around the value predicted from the theory, i.e., that the model does not introduce a systematic bias. This does not mean that the true value and the predicted value are identical, just that the distributions are similar enough. For example, the data may follow the distribution:

$$L(D|X, I) \propto \exp\left\{-\frac{1}{2\sigma^2} \chi^2(X)\right\}, \quad (3)$$

where the function  $\chi^2(X)$  quantitates the average discrepancy between the experimental measurements  $y_i$  and the data predicted from the structure  $X$  by our a theory,  $y_i(X)$ . For a Gaussian distribution, this is

$$\chi^2(X) = \sum_{i=1}^n [y_i - y_i(X)]^2. \quad (4)$$

$\sigma$  is the mean deviation of the measurements from the theoretical value. This is another, important, parameter

we generally cannot measure but need to introduce for the modelling.

Bayes’s rule combines the two components (the prior and the likelihood; quantities we can calculate) to derive the probability of a particular structure  $X$  (the quantity we are interested in). In the absence of prior information  $I$ , Bayes’s rule reduces to  $p(X|D) \propto L(D|X)$ , i.e., we directly identify the probability of the data, given the structure, with the probability of the structure, given the data. This is the basis of the maximum likelihood method.

If we take the negative logarithm of both sides of Eq. (2), we obtain an equation of the form of Eq. (1), the hybrid energy:

$$-\log[p(X|D, I)] = -\log[\pi(X|I)] - \log[L(D|X, I)] + \text{const.} \quad (5)$$

We can identify the hybrid energy function  $E_{\text{hybrid}}(X)$  with  $-\beta \log[p(X|D)] - \text{const.}$  and the pseudo energy  $E_{\text{data}}(X)$  with the negative logarithm of the likelihood,  $-\log[L(D|X, I)]$ . If we assume a known and constant  $\sigma$ , we obtain for the likelihood in Eq. (3):

$$E_{\text{hybrid}}(X) = \beta E_{\text{phys}}(X) + \frac{1}{2\sigma^2} \chi^2(X), \quad (6)$$

where the factor  $\beta$  defines the energy scale; it is 1 if we measure the energies in units of  $k_{\text{B}}T$ . Two conclusions follow from this analysis: first, if we know  $\sigma$ , it determines the weight on the data in the minimisation,  $w_{\text{data}}$ . It should ideally be set to reflect the quality (consistency) of the data: the larger  $\sigma$ , the lower the weight. Second, minimising a hybrid energy  $E_{\text{hybrid}}(X)$  is a way to maximise the probability of a structure  $X$ . In the absence of prior information, and with a Gaussian error distribution, this illustrates the relationship between least-squares and maximum likelihood methods.

We use this here mostly as a motivation, since we will see further down that the analysis is incomplete. In particular, knowledge of  $\sigma$  (or of any of the parameters necessary to formulate the likelihood  $L(D|X, I)$ ) is not necessary in a truly probabilistic analysis. The assumption of the prior knowledge of  $\sigma$  is unrealistic:  $\sigma$  varies from data type to data type, and also from experiment to experiment. In NMR, in contrast to X-ray crystallography, we cannot easily obtain an estimate from repeatedly doing the same experiment, since, in particular for the NOE,  $\sigma$  is dominated by discrepancies between theory and experiment and not experimental noise. Approaches based on minimising  $E_{\text{hybrid}}$  (or maximising the posterior probability  $p(X|D, I)$ ) are

much less powerful than a full Bayesian treatment, since all additional unknown parameters need to be assumed as known and are generally fixed during the calculation.

### 3. Obtaining coordinates and their precision

#### 3.1. Estimation from repeated structure calculation

In structure determination by NMR (and sometimes also in X-ray crystallography [13,14]) one tries to obtain a measure of coordinate uncertainty by repeated, independent minimisations of Eq. (1). Estimating uncertainties in coordinates in this way has been a pre-occupation in NMR structure determination since its beginning [15,16]. The results differ from calculation to calculation with identical data, since all standardly employed minimisation approaches contain a random element. In approaches using standard minimisation (such as DIANA [17]), the initial coordinates are set to random values; in simulated annealing approaches [18,19], initial coordinates and initial velocities are chosen randomly; and in metric matrix distance geometry [20], a definite distance needs to be chosen for each pair of atoms before the embedding step, again using a random number generator.

Since no algorithm ever locates “the” global minimum but gets invariably trapped in one of the many local minima, repeated calculation with exactly the same data will result in different structures. A binary criterion (“no violations above 0.5 Å”) is then employed to distinguish between accepted and rejected structures. The accepted structures are considered equivalent solutions, and the distribution of the structures in space is a measure of the uniqueness and is often called “sampling” of conformational space. The acceptance criterion has been of some concern, since it is inherently subjective. Energy-ordered RMSD plots [21], for example, give some better insights into the convergence of the structure ensemble and can be used to rationalise the choice of a particular cutoff; however, they fail to offer a truly objective criterion.

The expectation is that the distribution of structures is influenced by the data quality. If the weight  $w_{\text{data}}$  depends on the standard deviation  $\sigma$  of the data, the influence of the data is reduced for low quality data and the result of repeated structure calculations can be expected to show larger variation. Most structure calculations employ lower and upper bounds with error tolerances that should be set according to  $\sigma$ . Then, the expectation is that the wider the bounds, the larger the difference between individual structures [16].

The resulting structure ensemble of  $m$  structures can be characterized by its average structure  $X_{\text{ave}}$ :

$$X_{\text{ave}} = \langle X \rangle = \frac{1}{m} \sum_{k=1}^m X_i, \quad (7)$$

and the covariance matrix  $C(X)$ , with diagonal elements containing the variances of each element of  $X$ , and off-diagonal elements containing the covariances between the elements of  $X$ .

$$C(X)_{i,j} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \quad (8)$$

where  $i, j = 1, 1/4, \dots, 3n_{\text{atom}}$ .

Although the result of such a procedure is useful as a rough guide, there are two fundamental problems with using independent solutions of an optimisation to characterise the distribution of structures. First, optimisation algorithms are neither guaranteed to find all important regions of the distribution nor to reproduce the correct populations of the different regions: optimisation methods will have the tendency to end up in the region that is easiest to reach for the particular algorithm. Thus the “sampling” provided by optimisation methods, starting from randomly varying initial points, will mostly depend on algorithmic properties. Second, many of the parameters that are necessary for calculating structures (such as the weight  $w_{\text{data}}$  in Eq. (1) need to be fixed before the calculation and the influence of their value and of its variation on the coordinate precision cannot be assessed.

#### 3.2. The probabilistic least-squares approach

Altman and Jardetzky [22] introduced an entirely different approach to the problem, which directly obtains coordinate uncertainty by the use of the Kalman filter. The Kalman filter is a set of equations that was originally developed to extract a time-dependent signal from a set of noisy observations. The aim of its application to NMR structure determination is to transform a collection of observations into optimal mean atomic positions and associated uncertainties implied by the data set. Given an initial state, the algorithm sequentially introduces the observations on the molecule. The result is a series of iteratively improved structures and uncertainty estimates. The algorithm directly optimises mean positions  $X_{\text{ave}}$ , and the covariance matrix  $C(X)$ .

Initial values must be assigned to  $X$  and  $C(X)$  before the introduction of data. They can be set to random values or obtained from an exhaustive sampling of conformational states with simplified models.

Restraints derived from data have generally the form:

$$z = h(X) + v, \quad (9)$$

where  $z$  is the measured value (e.g., a distance derived from NOE measurements),  $h(X)$  is the function that calculates the data from  $X$  (e.g., the distance between two points in the structure  $X$ ), and  $v$  is a random variable (white noise) that models the experimental errors in the system. Values for  $v$  are drawn from a Gaussian distribution with zero mean and with a variance that depends on the data precision (e.g., 0.1 Å for bond lengths and between 1 and 5 Å for NMR-derived distances).

The coordinates and the covariance matrix are then updated iteratively based on the values of the restraint function  $h(X)$ . For an  $h(X)$  linear in  $X$ :

$$X(\text{new}) = X(\text{old}) + K[z - h(X(\text{old}))], \quad (10)$$

where  $K$  is a matrix depending on the covariance matrix  $C(X)$  and the derivative of the restraint function  $h(X)$ . Simply stated, a new estimate of mean position is based on the current estimate of mean position, corrected by a weighted difference between the observed value of the measurement  $z$ , and the value predicted from the current  $X$ . The weight depends *via* the matrix  $K$  directly on the derivative of the restraints with respect to the coordinates, the covariance matrix, and the estimated data precision. The covariance matrix has a critical role in guiding the search for the optimum: the elements of  $X$  that have a large variance are more “flexible” than those with a smaller variance and are therefore more affected by the introduction of data. The covariance matrix itself is updated in parallel in a similar manner. The restraints are usually introduced sequentially since the introduction of all restraints simultaneously would lead to large matrix inversions.

The basic equations for the Kalman filter, Eq. (10), presume a linear dependence of  $h(X)$  on the coordinates  $X$ . In the (usual) case of a non-linear dependence, an extended form of the Kalman filter needs to be used. The algorithm may run into convergence problems (called “divergence”) due to the fact that in the course of the iterations, the elements of the covariance matrix may get so small that the newly introduced data carry almost no weight.

The Kalman filter is related to least squares estimation. It has been pointed out [23] that it represents a recursive solution to the standard least squares problem, in particular for a stationary solution and in the absence of prior information. In structure determination, we are

looking for stationary solutions. The advantage over standard least-squares estimation is the more efficient calculation due to the recursive nature of the algorithm, and the straightforward introduction of prior information in the form of initial estimates for  $X_{\text{ave}}$  and  $C(X)$ . We note that the prior information here is of a different nature and is handled differently from the Bayesian treatment.

The method has the merit to try to directly address the problem of obtaining an estimate of the uncertainty of the structure determination. The result obtained for  $C(X)$  depends directly on the estimated quality of the data [24,25]. The results can be displayed in the form of ellipsoids describing the extent of motion in the three spatial directions.

However, the method makes strong and unjustified assumptions that influence the final result: (i) the distribution of structures is assumed to be uni-modal and Gaussian (a mean position with a standard deviation), and (ii) it requires prior knowledge of data uncertainties. The mean position obtained by the algorithm does not necessarily correspond to a physical structure. Despite the introduction of the algorithm nearly 20 years ago, there is not much experience that would allow an evaluation of its reliability and convergence properties.

### 3.3. The inferential structure determination method

In order to rigorously address the problem of obtaining unbiased coordinate precision with the full dependency on all unknowns, we need to abandon the idea of minimising a hybrid energy or maximising the probability. Rather, we need to evaluate a probability  $P_i$  for all possible structures  $X_i$ . Generally, a continuum of  $P_i$  values is distributed over conformational space. Only if all but one  $P_i$  vanish, we can invert the data uniquely, obtaining exactly one structure from the data. On the other hand, in the case of uniform  $P_i$ , the data are completely uninformative with respect to the structure. In the case of a continuous parametrisation (Cartesian coordinates, dihedral angles)  $P_i$  is a density  $p(X|D, I)$ ; the integral  $\int_R dX p(X|D, I)$  evaluates the probability that region  $R$  of conformational space contains the true structure.

For a single—or very few—unknown (coordinates and other parameters), one could calculate the probability of every conformation, for example, by a grid search. For the large number of unknowns typical for the structure determination of a macromolecule this is unfeasible and the space of possible conformations has to be explored by a suitable sampling algorithm. The recently developed inferential structure determination method ISD [26] is therefore based on Monte

Carlo sampling to explore the probability distribution over conformational space. Monte Carlo sampling is not used as a means to find the maximum of a probability but to evaluate the integrals over parameters that appear in the use of Bayes's rule, Eq. (2) or (11).

Once the model to describe the data (i.e., the likelihood function) has been chosen, the rules of probability theory, Eq. (2) or (11), uniquely determines the posterior distribution. The appropriate statistics for modelling distances and NOEs are discussed further below. No additional assumptions need to be made.

### 3.3.1. Nuisance parameters

The full power of a full Bayesian treatment of the problem becomes apparent if there are additional unknown, auxiliary parameters. It is basically always necessary to introduce such auxiliary parameters in order to describe the problem adequately. For example, the parameters  $A$ ,  $B$ ,  $C$  of the Karplus relationship are, strictly speaking, unknown for the particular protein that one is investigating. Also, the data quality  $\sigma$  is an unknown parameter, as is the calibration factor  $\gamma$  for NOE volumes.

In Bayesian theory, these additional parameters are called “nuisance parameters”. In ISD, all additional unknown parameters of the error model and of the theory are estimated along with the structure. They are treated in the same way as the coordinates. To add the unknown  $\sigma$ , we simply replace  $X$  with  $(X, \sigma)$  in Eq. (2), and the full posterior becomes:

$$p(X, \sigma | D, I) \propto \pi(X | I) \pi(\sigma | I) L(D | X, \sigma, I). \quad (11)$$

Here, we *a priori* assume independence of  $X$  and the nuisance parameters—the prior for the coordinates does not depend on the values of the  $\sigma$  and *vice versa*—and we introduce the additional prior  $\pi(\sigma | I)$  (Jeffreys prior [27]) expressing our ignorance on this parameter. Other nuisance parameters (the calibration factor  $\gamma$ , Karplus parameters, tensor parameters, ...) are treated in exactly the same way and simply lead to additional terms in the equation.

The posterior density for the coordinates by themselves is formally obtained by integration over nuisance parameters (also called marginalisation [11]):

$$p(X | D, I) = \int d\sigma p(X, \sigma | D, I) \propto p(X | I) \int d\sigma p(D | X, \sigma, I) \pi(\sigma | I). \quad (12)$$

That is, in order to account for our ignorance regarding nuisance parameter  $\sigma$ , we have to replace  $p(D | X, I)$  in

Eq. (2) with a *weighted average* over the likelihood conditioned on all possible values of  $\sigma$ . This is in marked difference to standard structure determination by minimisation, where the value of any unknown parameter needs to be fixed before the structure calculation, and therefore only one single value is used. In contrast, the result of a structure calculation by inference directly contains the influence of the uncertainty in the additional parameters.

### 3.3.2. Sampling

A good sampling algorithm will produce samples with the correct probability. That is, the probability can be directly calculated from the number of times a particular region is visited. In contrast to an optimisation algorithm, it is designed to visit all regions of high probability, and not to locate efficiently one of the maxima.

Sampling of the probability distribution of protein conformations is difficult, due to a number of factors: there are many degrees of freedom (essentially the main chain torsion angles of a protein); the degrees of freedom are highly coupled; islands of high probability are separated by large stretches of low probability. To address this problem, we proposed an extended replica-exchange Monte Carlo scheme for simulating the posterior densities [26,28]. The extended scheme uses a combination of different concepts to sample the coordinates and nuisance parameters: a replica-exchange scheme using a hybrid Monte Carlo algorithm to sample the coordinates. The nuisance parameters are sampled with a standard Monte Carlo method.

In the replica-exchange Monte Carlo method, several non-interacting copies of the system, so-called replicas, are simulated in parallel at different temperatures. Exchanges of configurations between neighboring replicas are accepted according to the Metropolis criterion. In this way, configurations diffuse between high temperature and low temperature, which effectively reduces the risk of being trapped in a particular conformation. Each replica is simulated using the Hybrid Monte Carlo (HMC) [29] method. This consists of running a short dynamics trajectory of 250 integration steps starting with randomly assigned momenta to generate a new proposal state, which is then accepted or rejected according to the Metropolis criterion.

### 3.3.3. Analysing the results

The result of a Bayesian structure calculation is a large ensemble of structures sampled at many different values of the nuisance parameters. The Bayesian procedure derives statistically meaningful, objective

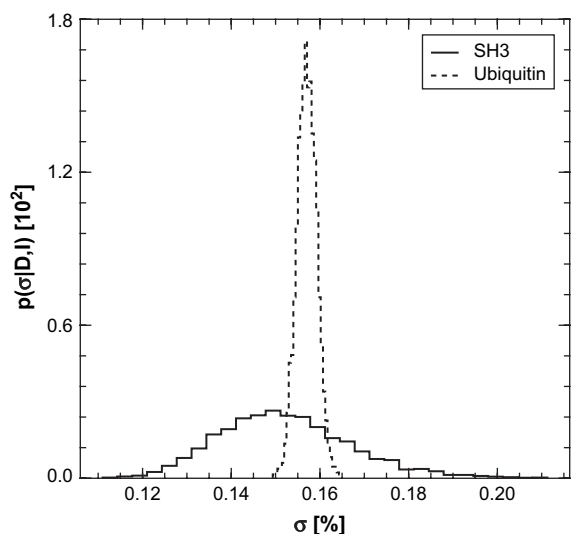


Fig. 2. Estimation of auxiliary quantities. Marginal posterior  $p(\sigma|D, I)$  distribution for the nuisance parameter  $\sigma$  (error of the lognormal model) obtained from Monte Carlo samples. The data set for Ubiquitin consisted of 1444 non-redundant distances taken from the restraint file, PDB code 1d3z; the data set for SH3 was for a perdeuterated sample and contained 150 distances between exchangeable protons.

error bars for all parameters, not only the atom positions but also the nuisance parameters. An example showing the distribution of  $\sigma$  for two different data sets is given in Fig. 2. Note that the average value, i.e., the quality, for the two very different data sets is approximately similar, but the width of the distribution is very different. The reason is the drastically different number of constraints available for the two proteins.

The variance of the structures automatically contains the influence of the nuisance parameters on the structures. The Bayesian structure calculation method is the only one that provides estimates for all unknown

parameters, error estimates of the parameters, and a statistical measure of a goodness of fit. Average structure and the covariance matrix calculated from a structural ensemble calculated with ISD are therefore more meaningful than in other structure determination approaches. Fig. 3 shows the result for two structure calculations, SH3 derived from a very sparse data set, and Ubiquitin from a much more complete data set.

Since the method has no free parameters that need to be fixed before the calculation, user intervention is not necessary, and structure determination becomes more objective.

#### 4. Data statistics and restraint potentials

The likelihood function  $L(D|X, I)$  (or the related potential  $E_{\text{data}}(X)$ ) needs to be known for any structure calculation. It has an important influence on the resulting distribution of structures. The role of  $L(D|X, I)$  or  $E_{\text{data}}(X)$  is to introduce our knowledge about expected deviations between measured and calculated data, and to evaluate the importance of these deviations. For certain data types, a Gaussian distribution can be assumed to a good approximation, e.g., for scalar or residual couplings. In contrast, NOEs and derived distances have too many large deviations to be well represented by a Gaussian.

##### 4.1. The standard representation using lower and upper bounds

The standard way to account for errors and imprecisions in NOE-derived distances 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

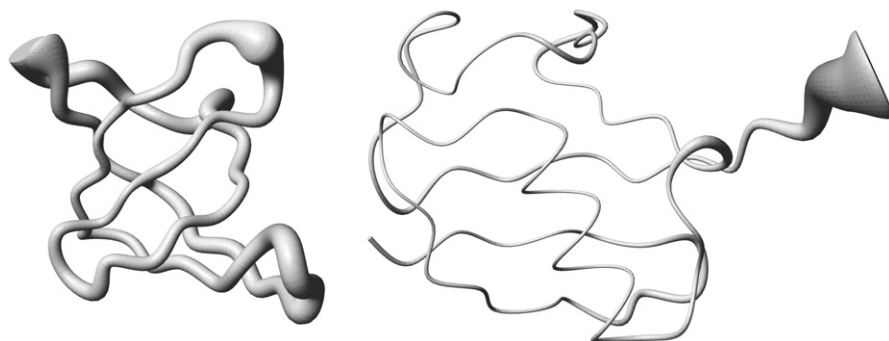


Fig. 3. Ensembles of most probable structures for the SH3 (left) domain and Ubiquitin (right), with the same data sets as in Fig. 2. The width of the “sausage” is proportional to the RMSD around the average structure. The distribution of structures contains the uncertainty due to the unknown nuisance parameters. Note that, whereas the structure ensembles are shown as an isotropic distribution for the sake of simplicity, distributions calculated with ISD are not isotropic or uni-modal.

influenced by the concept of distance geometry [20,30], has been used almost exclusively for determining NMR structures to date. The corresponding 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  $l$ , which are derived from the NOE, and rises harmonically in the deviation of the distance  $d(X)$  from  $u$  and  $l$ .

FBHW potentials are popular since energy and force are zero between  $u$  and  $l$  and, hence, do not influence the structure if the distance restraint is satisfied. In practice, however, the distance bounds representation can be problematic [31,32], and adds a strong subjective element to structure determination from NMR data. Figures of merit for NMR structures are therefore not very informative. In particular, the bounds representation makes a comparison of the determined structures to the estimated input distances meaningless. Also the RMS difference of the structure ensemble from the average structure ( $\text{RMSD}_{\text{ave}}$ , “precision”) is heavily influenced by the width of the bounds and the manipulation of individual bounds [33], and is therefore not an unbiased measure of the structure quality. In the following, we discuss some possibilities to replace the FBHW potential by potentials derived from an error distribution of the distances.

#### 4.2. Derivation of a restraint potential from error distributions

The distribution of errors of NOE-derived distances is *a priori* unknown. If we knew the error distribution  $g(d, d_0)$  in the distances  $d$  around the “true” distance  $d_0$ , we could 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  $E_i^{\text{NOE}}$  for a single restraint  $i$ :

$$E_i^{\text{NOE}} \propto -\log[g(d_{\text{exp}}^i, d^i(X))], \quad (13)$$

where  $d^i(X)$  is the distance calculated in the structure  $X$ , and  $d_{\text{exp}}^i$  is the measured distance.

To derive the error distribution and the corresponding potential, we need to know the “true” value of the distance. This can only be warranted by a model study. We proposed therefore to derive a distance distribution from a molecular dynamics trajectory [31,33]. As “true” values  $d_0$  we used the arithmetic average over the trajectory. To calculate histograms, we binned the distance differences  $d_0 - d$  into bins of 0.1 Å. Another possibility is to use the distances in an X-ray crystal

structure as a true value. An error distribution can then be derived by comparing pairs of X-ray crystal structures and NOE data sets of the same protein. Here, we simply assume that the shape of the distributions around the true distance and around the distance in the X-ray crystal structure are similar, not that the distances themselves are the same. A probability-derived potential for structure calculations is then derived by fitting a differentiable analytic function to the negative logarithm of the “raw” distribution.

We showed for several examples that structures calculated with the potential derived from a molecular dynamics trajectory on BPTI (Fig. 4(a)) were of better quality than those calculated with the FBHW potential, and that they were systematically closer to the X-ray crystal structure of the same protein [31]. In addition, we found that the distribution of structures around the average automatically depends on the data quality. This is evidenced in Fig. 5, where increasing amounts of random noise was added to the experimental data. Even with a constant weight on the distance restraints, the RMS difference of the structures to the average increases linearly with the noise in the restraints.

#### 4.3. The lognormal distribution and a derived potential

Despite the robustness of the results to the exact shape of the potential, it may seem problematic to use a potential with several free parameters in addition to the overall weight. We discuss therefore another approach to the derivation, an error distribution and a restraint potential, from more fundamental properties of NOEs and derived distances. To this end, we analyse the expected deviation of a measurement from the ideal value. Although the original analysis [46] was performed for NOE intensities  $I$ , it is valid also for distances.

NOE intensities and derived distances are inherently positive. A calibration factor  $\gamma$  need to be introduced in order to relate the intensity scale to a distance scale. Changing the units does not affect the information content of the data. Hence, the distribution  $g(I_{\text{obs}}, I_{\text{calc}})$  of the deviations between observed and calculated intensities must be invariant under scaling, i.e.,  $g(I_{\text{obs}}, I_{\text{calc}}) = \alpha g(\alpha I_{\text{obs}}, \alpha I_{\text{calc}})$ , which follows from the transformation rule of probability densities. A distribution that shows this scale invariance is the lognormal distribution:

$$g(I_{\text{obs}}, I_{\text{calc}}) = \frac{1}{\sqrt{2\pi\sigma^2}I_{\text{obs}}} \exp\left[-\frac{1}{2\sigma^2} \log^2\left(\frac{I_{\text{obs}}}{I_{\text{calc}}}\right)\right]. \quad (14)$$



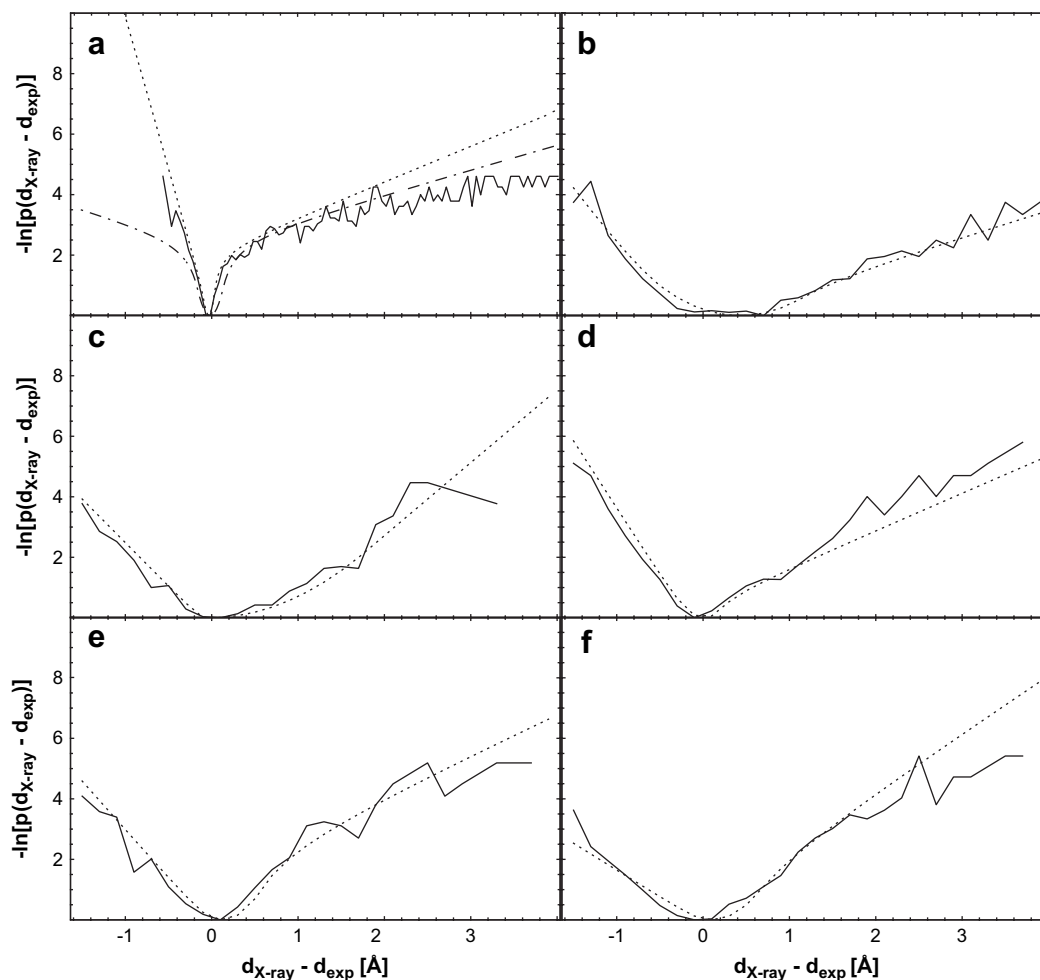


Fig. 4. Negative logarithm of the normalized distribution of  $r_{\text{ref}} - r_{\text{exp}}$  (solid line), a manually fitted symmetric function, (dashed line, only in (a)), and optimal fit (dotted line); (a) BPTI trajectory data, (b) I14 [34,35], (c) BPTI [36,37], (d) pH domain [38,39], (e) GB1 domain [40,41], (f) Ubiquitin [42,43]. The dotted line represents a fit of the functional form (the “soft-square” potential) already implemented in X-plor [44] and CNS [45] to the negative logarithm of the distribution. Note the general similarity of the potentials, despite the differences in the used data sets with respect to completeness and quality. From Ref. [31], with permission.

This distribution is restricted to the positive axis and is asymmetric around its median  $I_{\text{calc}}$ . Measurements are incorporated without bias in the sense that the probability of over- or underestimating the true intensity is both 1/2. This is not the case for error distributions defined on the entire axis, such as a Gaussian, which assign a non-vanishing probability to unobservable negative intensities. The parameter  $\sigma$  quantifies the relative deviation of the observed from the calculated intensity, provided that their difference is sufficiently small. Experimental NOE data follow this distribution quite well (see Fig. 6). This indicates that the assumption that the shapes of the distributions around the mean and around the value calculated by the theory is indeed similar, even for the simple ISPA approximation.

Fig. 6(c) and (d) shows distributions for distance differences  $d_{\text{obs}} - d_{\text{calc}}$  fitted to a Gaussian distribution. The distance error distributions are asymmetric and long-tailed and a Gaussian can significantly underestimate the probability of large deviations. Both properties are much better accounted for by the lognormal distribution. From a practical point of view, the lognormal model has several favorable properties. Unlike a probability distribution corresponding to a flat-bottom potential, it has a unique maximum. Hence, measurements are not weighted equally between bounds but are always penalized depending on the degree of disagreement with the structure. Furthermore, the lognormal distribution is invariant under power law transformations. If we raise the intensity

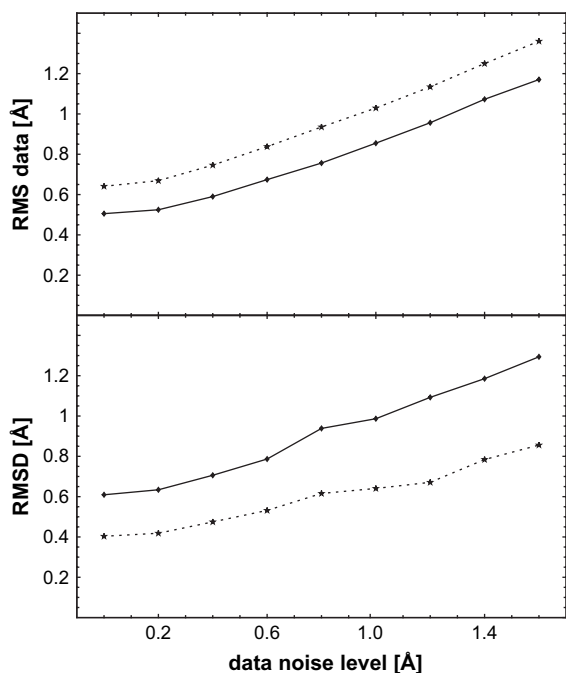


Fig. 5. RMS and RMSD values plotted against the level of added random noise added to the experimental distances for Ubiquitin, calculated with the potential from Fig. 4(a), and a data weight of 32. The values indicate the interval from which the random number was drawn; 0.2 indicates the interval  $[-0.2, 0.2]$  and so forth. Top panel:  $\text{RMS}_{\text{work}}$  (solid line) and  $\text{RMS}_{\text{cv}}$  (dotted line). Bottom panel:  $\text{RMSD}_{\text{X-ray}}$  (solid line) and  $\text{RMSD}_{\text{ave}}$  (dotted line). Note that all the calculations were performed with exactly the same conditions, that is, in marked difference to other studies [16,24], values for the expected error did not have to be known before the calculation. From Ref. [31], with permission.

to a power the transformed intensity still follows a lognormal law, with transformed median and error parameters. Since the ISPA is a special case of a power transformation, we obtain identical distributions for intensities and distances, provided that  $\sigma$  has been transformed appropriately. The lognormal model is implemented in our software for probabilistic structure determination (ISD) for calculating the likelihood based on NOE intensities or NOE-derived distances.

The negative logarithm of the distribution in Eq. (14) is the corresponding restraint potential:

$$E_i^{\text{NOE}} \propto -\log[g(I_{\text{obs}}, I_{\text{calc}})] = \frac{1}{2\sigma^2} \log^2\left(\frac{I_{\text{obs}}}{I_{\text{calc}}}\right), \quad (15)$$

that is, it is harmonic in the difference of the logarithm of calculated and experimental intensities. Note that this “log-harmonic” potential has only one single parameter, the weight depending on  $\sigma$ . Due to the

property of the logarithm, the corresponding potential for distances has exactly the same functional form, if one assumes the ISPA. The exponent  $-1/6$  necessary to convert the intensities to distances becomes a multiplicative factor of the logarithm and simply modifies the force constant. We have recently implemented this potential in the program CNS [45], and initial experiences for structure calculations give similar results as the potentials described in the previous section (unpublished).

## 5. Data quality and the weight on $E_{\text{data}}(X)$

As already mentioned in Section 1, the weight plays a fundamental role in calculating structures from experimental data. Within ISD, the weight is estimated along with all the other unknown parameters (see above). In a standard structure calculation by minimisation, the experimental data are weighted empirically:  $w_{\text{data}}$  is set ad hoc and held constant during structure calculation. In their original paper, Jack and Levitt [1] proposed to adjust the weight so as to equalise  $E_{\text{phys}}$  and  $w_{\text{data}}E_{\text{data}}(X)$ ; this has been later refined to equalise average gradients [47]. In NMR, one usually uses fixed empirical values. In particular for the single minimum potentials discussed in the previous section, the weight has a profound influence on the results.

### 5.1. Setting the weight by empirical means: cross-validation

An unbiased empirical method to determine the optimal weight is cross-validation [48,49]. 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), and the RMS difference to the test set defines a cross-validated measure that is not directly influenced by the data. If we measure the deviations from the distance (not from the bounds):

$$\text{RMS}_{\text{work}} = \sqrt{\frac{\chi_{\text{work}}^2}{n_{\text{work}}}} \quad (16)$$

$$\text{RMS}_{\text{cv}} = \sqrt{\frac{\chi_{\text{cv}}^2}{n_{\text{cv}}}} \quad (17)$$

where the  $\chi^2$  functions are restricted to the working and test sets, and  $n_{\text{work}}$  and  $n_{\text{cv}}$  are the number of data points in the working and test sets, respectively. Due to the properties of NMR data (in contrast to X-ray

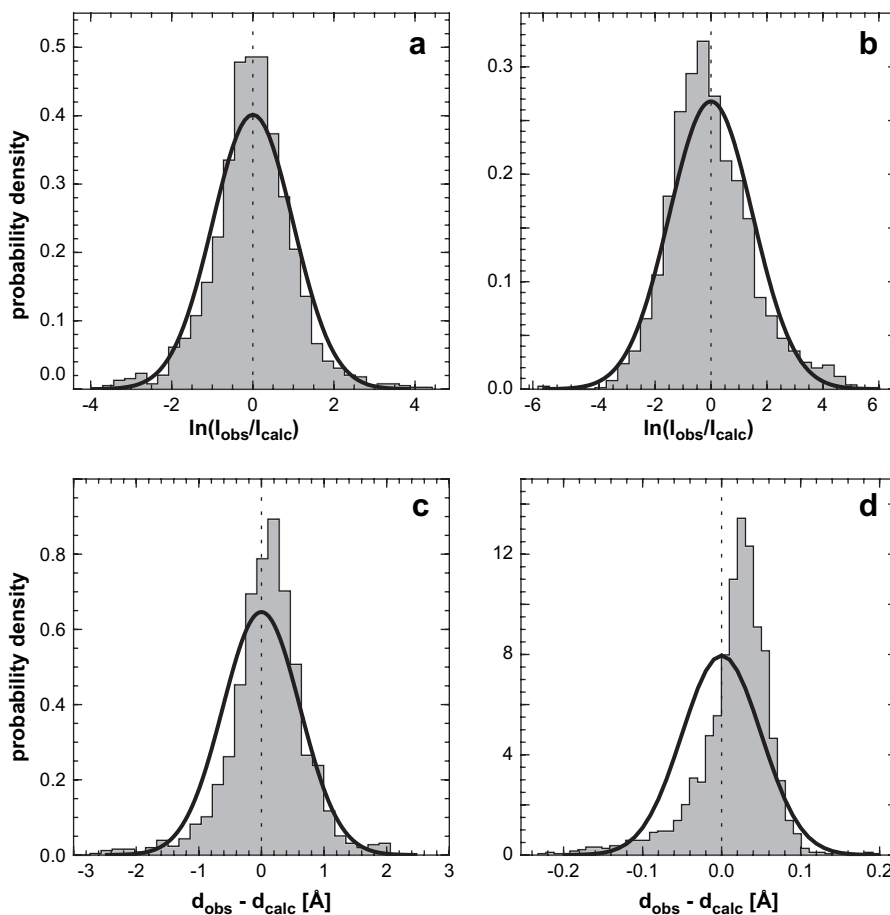


Fig. 6. Experimental intensity error distributions for (a) Ubiquitin and (b) the Tudor domain. Solid lines indicate fitted lognormal distributions; (c) and (d) corresponding distance error distributions and fitted Gaussians. Because intensities were unavailable for Ubiquitin, we used the ISPA to convert the published distance data into intensities (1444 non-redundant distances taken from the restraint file, PDB code 1d3z). The Tudor data (1875 intensities from two  $^{13}\text{C}$  and  $^{15}\text{N}$  edited spectra) were calibrated such that observed and calculated values have the same geometric mean. From Ref. [46], with permission.

crystallographic data, they are local and sparse), more stable results are obtained by using ten-fold complete cross-validation [49], that is, the data is divided into ten partitions of roughly equal size, each used in turn as a test set, and the results for  $\text{RMS}_{\text{work}}$  and  $\text{RMS}_{\text{cv}}$  are averaged. A minimum in  $\text{RMS}_{\text{cv}}$  indicates optimal calculation conditions.

To determine optimal conditions for a structure calculation, we need to vary these conditions (in our case, the weight on the distance restraint term) and repeat structure calculations with otherwise unchanged conditions (around 100 structures for each weight, i.e., ten per test set). Often,  $\text{RMS}_{\text{cv}}$  does not show a clear minimum but an elbow, and it is convenient to define an optimal weight using the elbow region (e.g., as the point where there was less than 10% change in  $\text{RMS}_{\text{cv}}$ ).

Fig. 7 shows, for Ubiquitin, the RMS differences and the WhatCheck [50] quality indices NQACHK and RAMCHK, as a function of the data weight used in the calculation. Around the optimal weight, calculations are not very sensitive to the precise value, and the elbow region is usually around values between 16 and 32, coinciding with the optimal values for RMS differences from the X-ray crystal structure, and for RAMCHK.

While cross-validation is an unbiased approach to obtain optimal parameters for a structure calculation, it is a rather insensitive to the parameter values. If many weights for different data terms need to be determined, its use becomes cumbersome, since many combinations of weights need to be evaluated. Another disadvantage is the necessity to remove data from the calculation, which, in the case of very sparse data, may reduce convergence.

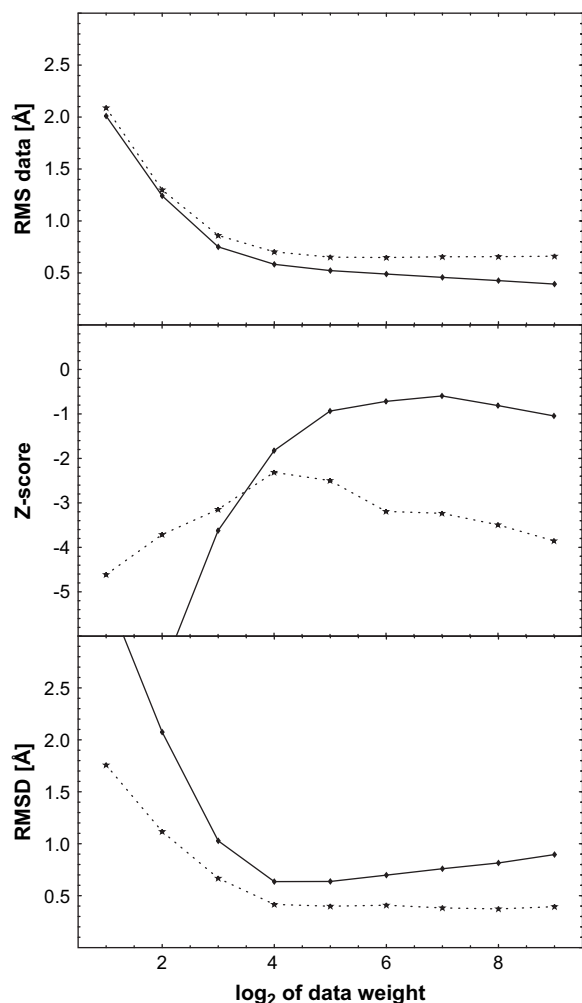


Fig. 7. RMS differences and quality indices for Ubiquitin for calculations with the PDF<sub>MDman</sub> potential. Top panel: RMS difference from data RMS<sub>work</sub> (solid line), cross-validated RMS difference from data RMS<sub>cv</sub> (dotted line). Middle panel: NQACHK (solid line) and RAMCHK (dotted line). Bottom panel: RMS distance from the X-ray crystal structure RMSD<sub>X-ray</sub> (solid line) and RMS distance from average structure RMSD<sub>ave</sub> (dotted line). Structures were calculated with the distance restraint potential shown in Fig. 4(a). From Ref. [31], with permission.

## 5.2. Probabilistic approaches to set the weight

Probability theory gives us other possibilities to weight experimental data in an objective way: (i) the unknown weight parameter can be sampled along with the other unknown parameters, as in the ISD approach [26] described above; (ii) the weight can be included into the restraint function [51]; or (iii) the weight can be eliminated by marginalisation.

We focus on the second method, inclusion of the weight into a restraint function, since it is of most

practical importance for minimisation. The negative logarithm of Eq. (11), the full posterior probability including the nuisance parameter  $\sigma$  is

$$E_{\text{joint}}(X, \sigma) = \frac{1}{2\sigma^2}\chi^2(X) + \beta E_{\text{phys}}(X) + \log[Z(\sigma)/\pi(\sigma)], \quad (18)$$

where, for the lognormal model,  $\chi^2(X) = \sum_{i=1}^n \log^2[y_i/y_i(X)]$ .

The term  $\log[Z(\sigma)/\pi(\sigma)]$  in the joint hybrid energy, Eq. (18), is not included in the standard target function  $E_{\text{hybrid}}$ , Eq. (1). However, it is precisely this additional term which allows us to determine the error.  $Z(\sigma)$  originates in the normalisation of  $P(D|X, I)$ ,  $\pi(\sigma)$  is required by Bayes's theorem. Both terms are missing in purely optimisation-based approaches, where normalisation constants and prior probabilities are usually not incorporated; this shows that a probabilistic framework is necessary for the correct treatment of the problem.

One might think that including the weight directly into a restraint energy would favour large values for  $\sigma$  with the corresponding weight approaching zero, since this would automatically minimise the restraint energy. However, the Bayesian analysis shows [51] that in the joint target function  $E_{\text{joint}}$ , Eq. (18), two contributions counterbalance each other:  $\chi^2/\sigma^2$  decreases when  $\sigma$  increases, thus preferring large values for the error when  $E_{\text{hybrid}}$  is minimized with respect to the error. In contrast, the term  $\log[Z(\sigma)/\pi(\sigma)]$  is monotonically increasing with  $\sigma$  (for the lognormal distribution, Eq. (14), it is  $\propto \sigma^{n+1}$ , where  $n$  is the number of data points). The ratio of the two terms therefore shows a finite minimum, which can be used to calculate the error, and, correspondingly, the optimal weight.

Minimisation of the resulting joint hybrid energy  $E_{\text{joint}}(X, \sigma)$  yields the most probable structure  $X_{\text{max}}$  and the most probable error  $\sigma_{\text{max}}$ . In case of the lognormal model, Eq. (14), we obtain  $\sigma_{\text{max}} = \sqrt{\chi^2(X_{\text{max}})/(n+1)}$ . Further analysis yields for the average weight  $\langle w_{\text{data}} \rangle = n/\chi^2(X)$  as an estimate. These estimates concur with common sense: the average weight quantifies, in good approximation, how well the structure fits the data, independent of the size of the data set. The precision of the estimate, i.e., the width of the weight distribution, in contrast, decreases with the square root of the number of data points [51]. Therefore, when sampling the weight, we obtain sharp distributions for  $\sigma$  for typical NMR data (see Fig. 2). This estimate is much more precise than what we can obtain from cross-validation.

To apply this estimate in the context of structure determination by minimisation, we can iteratively update

the current weight to  $w_{\text{data}} = n/\chi^2(X^2) = \text{RMS}_{\text{work}}^{-2}$ . This weight is a conservative estimate since it is always smaller than  $\chi^2(X_{\text{max}})/(n + 1)$ , the most probable weight derived from the most probable structure. We are at present testing implementations of this simple rule within CNS and ARIA.

## 6. Conclusions

In this paper, we reviewed approaches to obtain molecular coordinates from experimental data, with a special emphasis on treating it as a data-analysis problem of parameter fitting. Only the inferential structure determination method ISD satisfies the three conditions for a good fitting procedure mentioned in Section 1 in that it provides fitted parameters, error estimates of the parameters, and a statistical measure of a goodness of fit. It is the only method that can derive these estimates for the coordinates taking into account uncertainties in all additional parameters that are necessary for the structural modelling. It does not require prior estimates of data quality, but only information on the general form of the error distribution of the data. It can therefore be regarded as the “king’s way” to structure calculation.

Least-squares type methods have the disadvantages that one must define experimental errors *a priori*, and that they impose a uni-modal Gaussian distribution on the resulting structures. In contrast, the standard way of estimating structural uncertainty by repeated structure calculations does not impose any distribution on the structures but suffers from the problem that the structural ensembles are largely influenced by algorithmic properties.

Despite the success of ISD, standard structure determination by minimisation will undoubtedly continue to play a dominant role in practical applications. We argue that, to obtain the maximum benefit, these approaches should derive from a rigorous probabilistic treatment. This has implications on the restraint potential and on the weighting of experimental evidence. For example, the necessity of defining the data quality *a priori* can be removed. Whereas we cannot get statistically meaningful coordinate uncertainties from the minimisation approach, our experience shows that the uncertainties in the structure determination automatically depend on the data quality if an appropriate restraint potential is used. We also foresee that the analysis of differences between experimental and calculated data on a per-restraint or per-residue basis will be more informative than with the presently used fixed weights and FBHW potentials.

## References

- [1] A. Jack, M. Levitt, *Acta Crystallogr.* A34 (1978) 931.
- [2] A.T. Brunger, M. Nilges, *Quart. Rev. Biophys.* 26 (1993) 49.
- [3] R. Kaptein, E.R.P. Zuiderweg, R.M. Scheek, R. Boelens, W.F. van Gunsteren, *J. Mol. Biol.* 182 (1985) 179.
- [4] A.T. Brunger, G.M. Clore, A.M. Gronenborn, M. Karplus, *Proc. Natl Acad. Sci. USA* 83 (1986) 3801.
- [5] W. Press, B. Flannery, A. Teukolsky, W. Vetterling, *Numerical Recipes: The art of scientific computing*, Cambridge University Press, Cambridge, 1986.
- [6] W. Braun, *Quart. Rev. Biophys.* 19 (1987) 115.
- [7] T.F. Havel, *Prog. Biophys. Mol. Biol.* 56 (1991) 43.
- [8] A.T. Brunger, P.D. Adams, L.M. Rice, *Structure* 5 (1997) 325.
- [9] M. Nilges, S.I. O’Donoghue, *Prog. Nucl. Magn. Reson. Spectrosc.* 32 (1998) 107.
- [10] P. Güntert, *Quart. Rev. Biophys.* 31 (1998) 145.
- [11] E.T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, UK, 2003.
- [12] R.A. Engh, R. Huber, *Acta Crystallogr.* A47 (1991) 392.
- [13] A.T. Brunger, *J. Mol. Biol.* 203 (1988) 803.
- [14] M.A. Depristo, P.I. de Bakker, R.J. Johnson, T.L. Blundell, *Structure* 13 (2005) 1236.
- [15] K. Wüthrich, *NMR of Proteins and Nucleic Acids*, John Wiley, New York, 1986.
- [16] T.F. Havel, K. Wüthrich, *J. Mol. Biol.* 182 (1985) 281.
- [17] P. Güntert, W. Braun, K. Wüthrich, *J. Mol. Biol.* 217 (1991) 517.
- [18] M. Nilges, G.M. Clore, A.M. Gronenborn, *FEBS Lett.* 239 (1988) 129.
- [19] P. Güntert, C. Mumenthaler, K. Wüthrich, *J. Mol. Biol.* 273 (1987) 283.
- [20] T.F. Havel, I.D. Kuntz, G.M. Crippen, *Bull. Math. Biol.* 45 (1983) 665.
- [21] H. Widmer, A. Widmer, W. Braun, *J. Biomol. NMR* 3 (1993) 307.
- [22] R.B. Altman, O. Jardetzky, *Methods Enzymol.* 177 (1989) 218.
- [23] H.W. Sorenson, *IEEE Spectrum* 7 (1970) 63.
- [24] R. Pachter, R.B. Altman, O. Jardetzky, *J. Magn. Reson.* 89 (1990) 578.
- [25] Y. Liu, D. Zhao, R. Altman, O. Jardetzky, *J. Biomol. NMR* 2 (1992) 273.
- [26] W. Rieping, M. Habeck, M. Nilges, *Science* 309 (2005) 303.
- [27] H. Jeffreys, *Proc. R. Soc. A*186 (1946) 453.
- [28] M. Habeck, W. Rieping, M. Nilges, *Phys. Rev. Lett.* 94 (2005) 18105.
- [29] S. Duane, A.D. Kennedy, B. Pendleton, D. Roweth, *Phys. Lett. B* 195 (1987) 216.
- [30] L.M. Blumenthal, *Theory and Applications of Distance Geometry*, Chelsea, Bronx, New York, 1970.
- [31] M. Nilges, M. Habeck, S.I. O’Donoghue, Wolfgang Rieping, *Proteins* 64 (2006) 652.
- [32] J.C. Hoch, A.S. Stern, P.J. Conolly, in: O. Jardetzky, J.-F. Lefevre (Eds.), *Protein dynamics, function, design*, Plenum, New York, 1998.
- [33] F.R. Chalaoux, S.I. O’Donoghue, M. Nilges, *Proteins* 34 (1999) 453.
- [34] R. Powers, D.S. Garrett, C.J. March, E.A. Frieden, A.M. Gronenborn, G.M. Clore, *Science* 256 (1992) 1673.
- [35] A. Gustchina, A. Wlodawer, A. Pavlovsky, *FEBS Lett.* 309 (1992) 59.

- [36] K.D. Berndt, P. Güntert, L.P. Orbons, K. Wüthrich, *J. Mol. Biol.* 227 (1992) 757.
- [37] M. Marquart, J. Walter, J. Deisenhofer, W. Bode, R. Huber, *Acta Crystallogr. B* 39 (1983) 480.
- [38] M. Nilges, M. Macias, S.I. O'Donoghue, H. Oschkinat, *J. Mol. Biol.* 269 (1997) 408.
- [39] M. Hyvönen, M.J. Macias, M. Nilges, H. Oschkinat, M. Saraste, M. Wilmanns, *EMBO J.* 14 (1995) 4676.
- [40] A.M. Gronenborn, D.R. Filpula, N.Z. Essig, A. Achari, M. Whitlow, P.T. Wingfield, G.M. Clore, *Science* 253 (1991) 657.
- [41] T. Gallagher, P. Alexander, P. Bryan, G.L. Gilliland, *Biochemistry* 33 (1994) 4721.
- [42] G. Cornilescu, J.L. Marquardt, M. Ottiger, A. Bax, *J. Am. Chem. Soc.* 120 (1998) 6836.
- [43] S. Vijay-Kumar, C.E. Bugg, W.J. Cook, *J. Mol. Biol.* 194 (1987) 531.
- [44] C.D. Schwieters, J.J. Kuszewski, N. Tjandra, G.M. Clore, *J. Magn. Reson.* 160 (2003) 65.
- [45] A.T. Brunger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, J.S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, *Acta Crystallogr. D* 54 (1998) 905.
- [46] W. Rieping, M. Habeck, M. Nilges, *J. Am. Chem. Soc.* 127 (2005) 16026.
- [47] A.T. Brunger, A. Krukowski, J.W. Erickson, *Acta Crystallogr. A* 46 (1990) 585.
- [48] A.T. Brunger, *Nature* 355 (1992) 472.
- [49] A.T. Brunger, G.M. Clore, A.M. Gronenborn, R. Saffrich, M. Nilges, *Science* 26 (1993) 328.
- [50] R.W. Hoof, G. Vriend, C. Sander, E.E. Abola, *Nature* 381 (1996) 272.
- [51] M. Habeck, W. Rieping, M. Nilges, *Proc. Natl Acad. Sci. USA* 103 (2006) 1756.