

1.7 Refinement of X-ray Crystal Structures

AT Brunger, Stanford University, Palo Alto, CA, USA

PD Adams, University of California Berkeley, Berkeley, CA, USA

© 2012 Elsevier B.V. All rights reserved.

1.7.1	Introduction	105
1.7.2	Target Functions for Refinement	106
1.7.2.1	Maximum Likelihood	106
1.7.3	The Model	107
1.7.3.1	Bulk Solvent Models	107
1.7.3.2	A Priori Chemical Information	108
1.7.3.3	Atomic Displacement Parameters	108
1.7.3.3.1	Atomic displacement parameter restraints	109
1.7.3.3.2	Translation-libration-screw refinement	109
1.7.3.4	Other Restraints	109
1.7.3.4.1	Noncrystallographic symmetry	109
1.7.3.5	Ensemble Models	110
1.7.4	Cross-Validation	110
1.7.5	Optimization Methods	111
1.7.5.1	Gradient Descent Methods	111
1.7.5.2	Searching Conformational Space with Simulated Annealing	111
1.7.5.2.1	Molecular dynamics	111
1.7.5.2.2	Temperature control	111
1.7.5.2.3	Torsion-angle parameterization	112
1.7.5.2.4	Multi-start simulated annealing refinement	112
1.7.6	Special Considerations at Low Resolution	112
1.7.6.1	Treatment of Weak Intensities	112
1.7.6.2	Thermal Factor Sharpening of Electron Density Maps	112
1.7.6.3	Deformable Elastic Network Refinement	113
1.7.7	Conclusions and Outlook	113
	Acknowledgments	113
	References	113

Abbreviations

ADP atomic displacement parameter
CCD charge-coupled device
CNS Crystallography & NMR system
DEN Deformable Elastic Network
MAD multiwavelength anomalous dispersion

NCS noncrystallographic symmetry
SA simulated annealing
SAD single-wavelength anomalous dispersion
SIR single-isomorphous replacement
TLS Translation-Libration-Screw

Glossary

Least squares (LSQ) A classical method for formulating an optimization problem, where the function minimizes the squared difference between a target value and value calculated from the model.

Maximum likelihood (ML) A statistical approach to defining the target for an optimization problem, which

often reduces to a least-squares formalism if certain conditions are met.

Noncrystallographic symmetry (NCS) Occurs when there are symmetric relationships between molecules in the crystallographic asymmetric unit that do not obey pure crystallographic symmetry.

1.7.1 Introduction

Over the last decade, developments in molecular biology, X-ray diffraction instrumentation, and computational methods have allowed a nearly exponential growth of macro-

molecular structural studies. In particular, cryo-protection to extend crystal life,¹ the availability of tunable synchrotron sources,² high-speed charge-coupled device (CCD) data collection devices, and the ability to incorporate anomalously scattering selenium atoms into proteins have all made

structure solution much more efficient.³ The multiple anomalous diffraction (MAD) method⁴ often allows high-quality experimental electron density maps to be obtained. The analysis of the experimental data generally requires sophisticated computational procedures that culminate in refinement and structure validation. This refinement procedure can be formulated as the chemically constrained or restrained nonlinear optimization of a target function, which usually measures the agreement between observed data and data computed from an atomic model. The ultimate goal is to optimize the simultaneous agreement of an atomic model with observed data and with a priori chemical information.⁵

The target function used for this optimization normally depends on several atomic parameters, but most importantly on atomic coordinates. The large number of adjustable parameters (typically at least three times the number of atoms in the model) gives rise to a very complicated target function. For crystallographic refinement, the introduction of cross-validation (the 'free' R -value) has significantly reduced the danger of overfitting the diffraction data.⁶ The complexity of the conformational space has been reduced by the introduction of torsion-angle molecular dynamics,⁷ which decreases the number of adjustable parameters that describe a model approximately tenfold. Similar reductions in the number of adjustable atomic displacement parameters (ADPs) have also been realized using the Translation–Libration–Screw (TLS) method.⁸ The target function has been improved by incorporating the concept of maximum-likelihood, which takes into account model error, model incompleteness, and errors in the experimental data.^{9–11} Finally, the sampling power of simulated annealing can be used for exploring the molecule's conformational space in cases where the molecule undergoes dynamic motion or static disorder through multiconformer models.^{12–14} More recently, information from a reference model has been incorporated into refinement at low to medium resolution, greatly increasing the radius of convergence and accuracy of refinement for such difficult cases.¹⁵

1.7.2 Target Functions for Refinement

In essence, macromolecular structure calculation and refinement is a search for the global minimum of a target function

$$E = E_{\text{chem}} + w_{\text{data}}E_{\text{data}} \quad [1]$$

as a function of the parameters of an atomic model, in particular atomic coordinates. E_{chem} comprises empirical information about chemical interactions; it is a function of all atomic positions, describing covalent (bond lengths, bond angles, torsion angles, chiral centers, and planarity of aromatic rings) and nonbonded (intra-molecular as well as inter-molecular and symmetry-related) interactions. E_{data} describes the difference between observed and calculated diffraction data, and w_{data} is a weight appropriately chosen to balance the gradients (with respect to atomic parameters) arising from the two terms. The optimum weight can be obtained empirically by performing a series of refinements with different weights and selecting the weight that produces the lowest free R -value.⁶

For many years, the conventional form of E_{xray} consisted of the crystallographic residual E^{LSQ} , defined as the sum over the squared differences between the observed F_o and calculated F_c structure-factor amplitudes for a particular atomic model:

$$E_{\text{xray}} = E^{\text{LSQ}} = \sum_{hkl} (|F_o| - k|F_c|)^2 \quad [2]$$

where hkl are the indices of the reciprocal lattice points of the crystal and F_o and F_c are the observed and calculated structure-factor amplitudes, and k is a relative scale factor.

1.7.2.1 Maximum Likelihood

Reduction of E^{LSQ} can not only result from improvement in the atomic model, but also from an accumulation of systematic errors in the model or fitting noise in the data.¹⁶ The least-squares residual is therefore poorly justified when the model is far away from the correct one or incomplete.¹⁷ An improved target for macromolecular refinement can be obtained using a maximum-likelihood formulation.^{9,17–19} The goal of the maximum-likelihood method is to determine the probability of making a measurement, provided the model is given, and estimates of the model's errors and those of the measured intensities. The effects of model errors (incorrectly placed and missing atoms) on the calculated structure factors are first quantified with σ_A values, which correspond roughly to the fraction of each structure factor that is expected to be correct. However, overfitting of the diffraction data causes the model bias to be underestimated and under-corrected in the σ_A values. The effect of this overfitting can be reduced by cross-validating σ_A values, that is, by computing them from a randomly selected test set that is excluded from the summation^{6,20} on the right-hand side of eqn [2]. The expected values of $\langle F_o \rangle$ and the corresponding variance (σ_{ML}^2) are derived from σ_A , the observed F_o , and calculated F_c . These quantities can be readily incorporated into a maximum-likelihood target function⁹

$$E_{\text{xray}} = E_{\text{ML}} = \sum_{hkl \in \text{working set}} \left(\frac{1}{\sigma_{\text{ML}}^2} \right) (|F_o| - \langle |F_o| \rangle)^2 \quad [3]$$

In order to achieve an improvement over the least-squares residual (eqn [2]), cross-validation was found to be essential¹⁰ for the computation of σ_A and its derived quantities in eqn [3].

For many crystal structures, some initial experimental phase information is available from either isomorphous heavy atom replacement or single- or multiwavelength anomalous diffraction methods. These phases represent additional observations that can be incorporated in the refinement target. The maximum-likelihood formulation naturally extends itself to incorporation of this information.^{11,21} Tests have shown that the addition of experimental phase information, including single-isomorphous replacement (SIR) or single-wavelength anomalous dispersion (SAD), greatly improves the results of refinement.^{11,20} More recently, SAD phasing target functions have been incorporated directly into structure refinement,²² and show improved results particularly when used in automated model-building applications. It should be noted that with advances in synchrotron X-ray radiation

instrumentation and density modification it is now possible in a large number of cases to use SAD phasing to solve the crystal structure.²³

Pannu and Read⁹ have developed an efficient Gaussian approximation for the case of structure-factor amplitudes with no prior phase information, termed the Maximum-Likelihood Function (MLF) target function. In the limit of a perfect model, MLF reduces to the traditional least-squares residual (eqn [2]) with $1/\sigma^2$ weighing. In the case where prior phase information is included, the integration over the phase angles is carried out numerically, and is termed the Maximum Likelihood with Hendrickson–Lattman coefficients (MLHL) target function.¹¹ A maximum-likelihood function that expresses the probability distributions in terms of observed intensities has also been developed, and is termed Maximum-Likelihood function with Intensities (MLI).⁹

1.7.3 The Model

The contents of the macromolecular crystal are modeled and this model is optimized with respect to the experimental data and prior chemical knowledge. The model consists of two principal components: atoms that model the macromolecule in question, and a term that accounts for the disordered solvent that surrounds the macromolecule in the crystal. This solvent typically accounts for 50% or more of the crystal contents. The atoms describing the macromolecule are represented as isotropic or anisotropic scatterers of X-rays with a defined occupancy within the crystal. The solvent term is typically represented as a set of partial structure factors arising from the scattering of X-rays by the disordered solvent (see below). By adding structure factors calculated from the atomic model to those of the solvent, it is possible to calculate reciprocal-space structure factors from the atomic model and compare them to what was observed experimentally. The atomic model is restrained by the introduction of prior chemical knowledge (see below) in order to maintain appropriate distances between the atoms. The process of structure refinement changes the position of the atoms in space, and modifies the displacement parameters and occupancies in order to best fit the experimental data.

1.7.3.1 Bulk Solvent Models

The correct modeling of the disordered solvent in the crystal lattice is an important part of macromolecular structure refinement, and it becomes especially important for structures determined at low to medium resolution. The structure factor F_{calc} of a macromolecular crystal structure is

$$F_{\text{calc}} = k \exp[-2\pi^2 \vec{h}^t U^* \vec{h}] \{F_{\text{macro}} + F_{\text{bound}} + F_{\text{bulk}}\} \quad [4]$$

where the structure factor F_{macro} is obtained from the atomic model of the macromolecule, F_{bound} is computed from all bound water molecules, F_{bulk} is obtained from an appropriate model for disordered solvent, \vec{h} is a column vector with the Miller indices of a Bragg reflection, \vec{h}^t denotes the transpose of it (i.e., a row vector), k is a scale factor, and the symmetric second rank tensor U describes overall mean-square displacements of

the crystal lattice (dimensionless anisotropic mean-square displacements, atomic displacement parameter (ADP)). The isotropic component of the ADPs is usually separated from U and applied directly to F_{macro} , F_{bound} , and F_{bulk} . To do this, the U tensor is converted into Cartesian coordinate space U_{cart} .²⁴ One third of its trace (i.e., $(U_{\text{cart}}[11] + U_{\text{cart}}[22] + U_{\text{cart}}[33])/3$) is the isotropic thermal factor contribution.

To compute F_{bulk} , one approach is to create a mask in order to distinguish between macromolecular and solvent regions.^{25–27} All grid points of mask are initially set to 1. Grid points of mask within a distance of r_i around any atom i of the atomic model and its symmetry mates are then set to 0. The atomic model includes the macromolecule and any bound water molecules or ligands. r_i is defined as the sum of the van der Waals radius r_{vdw} of atom i and the probe radius r_{probe} . All grid points of mask marked 0 are tested to see if they fall within a distance r_{shrink} from a grid point set to 1. If this is the case, the tested grid point is set to 1. This procedure effectively ‘shrinks’ the accessible surface area. Generally, $R_{\text{probe}} = R_{\text{shrink}} = 1$ is the optimum choice,²⁶ although there are cases where the optimum values can be different than 1.^{28,29} The grid points of mask marked 1 comprise the solvent regions whereas those marked 0 are associated with the atomic model and its symmetry mates.

The structure factor of the solvent F_{bulk} is then simply computed by Fourier transformation of mask. In order to blur the sharp boundary between macromolecule and solvent as imposed by the mask, resolution-dependent scaling in reciprocal space is applied using an isotropic ‘thermal’ factor B_{sol}

$$F_{\text{bulk}}(k_{\text{sol}}, B_{\text{sol}}, r_{\text{probe}}, r_{\text{shrink}}) \\ = k_{\text{sol}} \exp(-B_{\text{sol}} \sin^2 \theta / \lambda^2) FT(\text{mask}[r_{\text{probe}}, r_{\text{shrink}}]) \quad [5]$$

where FT denotes the three-dimensional Fourier transformation, and k_{sol} is a scale factor that defines the mean electron density in the solvent region. For a well-behaved aqueous solvent model, k_{sol} is generally in the range 0.3–0.4 $e/\text{\AA}^3$, and B_{sol} is typically close (within a factor two) to the average thermal factor of the macromolecular model.

The optimum solvent model is obtained by minimizing the expression:

$$(F_{\text{obs}} - F_{\text{calc}}[k, k_{\text{sol}}, B_{\text{sol}}, U])^2 \quad [6]$$

as a function of the anisotropic thermal factor U , scale factor k , and the bulk solvent parameters k_{sol} and B_{sol} , where F_{obs} is the observed structure factor. A straightforward application of least-squares optimization to determine the minimum of this expression results in numerical instabilities for structures determined at lower than 3 \AA resolution. To avoid this problem, grid search optimization has been used.^{25,30} An implementation in CNS uses a one-dimensional grid search for k_{sol} while letting B_{sol} and the other adjustable parameters be determined by least-squares optimization for each selected value of k_{sol} .²⁵ Another implementation in Phenix³¹ utilizes a two-dimensional grid search with both k_{sol} and B_{sol} .³⁰ Both implementations (which are available in the latest versions of CNS and Phenix) are robust over a wide range of minimum Bragg spacings of the diffraction data, including at low resolution.

Instead of the solvent mask, it is possible to represent bulk solvent through an analytical function involving the atomic coordinates of the macromolecule.³² There are advantages and disadvantages to both approaches: the mask-based methods allow more flexibility in terms of adjusting the molecular surface with the 'shrink' parameter, while the analytical function allows the computation of atomic derivatives of the bulk solvent model.

1.7.3.2 A Priori Chemical Information

The geometric energy function E_{chem} consists of terms for covalent bonds, bond angles, chirality, planarity, and non-bonded repulsion.⁵ The parameters for the covalent terms can be derived from average geometry and root-mean-square (r.m.s.) deviations observed in a small-molecule database. Extensive statistical analyses were undertaken for the chemical moieties of proteins³³ and of polynucleotides³⁴ using the Cambridge crystallographic database.³⁵ Analysis of the ever-increasing number of atomic resolution macromolecular crystal structures has led to modifications of these parameters.³⁶

The first implementation of crystallographic refinement used a full force field with electrostatics and a van der Waals potential.³⁷ Benefits of including electrostatics in crystallographic refinement were observed during the refinement of influenza virus hemagglutinin,³⁸ although formation of incorrect hydrogen bonds were observed during high-temperature annealing, especially for charged groups such as the headgroups of arginine residues. These problems with charged groups may be related to the lack of continuum electrostatics in the force field that was used at the time. For simplicity, it has become the practice in most modern refinement programs and protocols^{19,25,31,39} to exclude electrostatics during refinement¹⁰ and to use a purely repulsive quartic function ($E_{\text{repulsive}}$) for the nonbonded interactions⁵ that are included in E_{chem}

$$E_{\text{repulsive}} = \sum_{ij} \left((cR_{ij}^{\text{min}})^n - R_{ij}^n \right)^m \quad [7]$$

where R_{ij} is the distance between two atoms i and j , R_{ij}^n is the van der Waals radius for a particular atom pair ij , $c \leq 1$ is a constant that is sometimes used to reduce the radii, and $n = 2$, $m = 2$ or $n = 1$, $m = 4$. These simplifications are valid since the experimental data contains information that is able to produce atomic conformations consistent with actual nonbonded interactions. In fact, atomic resolution crystal structures can be used to derive parameters for electrostatic energies.⁴⁰

Purely repulsive nonbonded interactions are used partly because the calculation is simplified, and therefore computationally faster. However, the main motivation is to avoid biasing the structure calculation to artifacts that may be present in the force field. In particular, the electrostatic terms are often difficult to parameterize. If the experimental diffraction information is insufficient to fully determine the macromolecular structure, use of electrostatic, attractive van der Waals, and simulated solvent interactions can bias the structure toward the theoretical nonbonded model. In this instance, it is preferable that the atoms do not attract one

another but rather are moved to points of minimal interaction as a result of repulsion.

Geometric energy functions are related to force fields that were developed for energy-minimization and molecular-dynamics studies of macromolecules.⁴¹ These force fields were not designed for structure determination, and therefore required some modification for use in macromolecular structure refinement by using covalent geometries derived from small-molecule crystal structures as described above and further modifications as described in References 38,42–44. More recently, there have been some attempts to re-introduce electrostatics in the refinement of NMR structures⁴⁵ and crystal structures,^{46,47} especially in combination with all hydrogen atoms in the refinement.

1.7.3.3 Atomic Displacement Parameters

The crystallographic experiment is a space and time average over the volume of the crystal illuminated by X-rays and the duration of the experiment. At room temperature, the atoms in the crystal will be subject to thermal motion. At cryo-cooled temperatures, the static conformations captured in each unit cell of the crystal typically show deviations from one another, on the same order of magnitude as thermal motion at room temperature. In both cases, the mean displacement of each atom around its centroid position must be modeled. These ADPs are typically modeled either isotropically or anisotropically. The former requires only one parameter per atom to define the displacement and is thus practical at the typical data resolutions encountered in macromolecular crystallography ($\sim 2 \text{ \AA}$ resolution). The use of anisotropic displacements requires six parameters per atom and is therefore only practical at high resolution ($\sim 1.4 \text{ \AA}$ resolution and better).

The mean-square displacements that define the probability density functions of atomic displacements are commonly parameterized as a trivariate Gaussian. The effect of the atomic displacements enters into the structure-factor calculation as the Debye–Waller factor $T(h)$, where h is a column vector with the Miller indices of a Bragg reflection. The fundamental expression for $T(h)$ is

$$T(h) = e^{-2\pi^2 \langle (hu)^2 \rangle} \quad [8]$$

where u is a row vector with the components of the displacement vector.⁴⁸ Note that there are a number of different conventions and notations used in connection with ADPs (\mathbf{B} , \mathbf{U} , \mathbf{U}^* , \mathbf{U}_{carb} , \mathbf{U}_{CIF}); see Grosse-Kunstleve and Adams²⁴ for a comprehensive review. In the following, authors refer to \mathbf{U} with its attendant conventions.

In practice, the atomic displacement on each atom is a superposition of a number of contributions,^{8,49–51} such as local atomic vibration, motion due to a rotational degree of freedom (e.g., libration around a torsion bond), loop or domain movement, whole molecule movement, and crystal lattice vibrations. Therefore, algorithms typically approximate the total ADP of each atom using multiple terms, with the total displacement $\mathbf{U}_{\text{cryst}}$ arising from the overall mean-square displacements of the crystal lattice described in Section 1.7.3.1 above. $\mathbf{U}_{\text{total}}$ can be split into three components: $\mathbf{U}_{\text{cryst}} + \mathbf{U}_{\text{group}} + \mathbf{U}_{\text{local}}$. $\mathbf{U}_{\text{local}}$ can be modeled using the less detailed

isotropic model that uses only one parameter per atom. As mentioned above, a more detailed (and accurate) anisotropic parameterization using six parameters requires more experimental observations to be practical. Group atomic displacement, $\mathbf{U}_{\text{group}}$, can be modeled using the Translation-Libration-Screw (see below) parameterization (\mathbf{U}_{TLS}) or just one parameter per group of atoms ($\mathbf{U}_{\text{subgroup}}$). Typically, $\mathbf{U}_{\text{subgroup}}$ is defined so as to refine one or two $\mathbf{U}_{\text{subgroup}}$ per residue. An arbitrary selected set of atoms can also be defined as a group. The choice usually depends on the data resolution and is typically made by using the R_{free} as a criterion to determine if the revised parameterization is appropriate. The CNS program allows for restraints to be applied to neighboring groups such that their displacements are similar. The current state-of-the-art in medium to high-resolution ADP refinement combines TLS and restrained residual isotropic displacements.

1.7.3.3.1 Atomic displacement parameter restraints

$\mathbf{U}_{\text{local}}$ models harmonic displacements around a mean position. If a displacement is purely a result of vibration, then it should obey Hirshfeld's rigid bond postulate,⁵² which states that atoms sharing a covalent bond have similar displacements. At the very high resolutions encountered in small-molecule crystallography, this is typically observed. However, at the data to parameter ratios seen in macromolecular crystallography, there is usually insufficient data to define the displacements accurately. Therefore, the rigid bond postulate is often used as a restraint to maintain chemically reasonable displacements for bonded atoms. In the general case of anisotropic displacements, the following restraint term is added to the optimization target for ADPs:

$$E_{\text{adp}} = \sum_{\text{bonded atoms } (i,j)} \sum_{k=1}^6 \left(U_{\text{local},i}^k - U_{\text{local},j}^k \right)^2 \quad [9]$$

The effect of this restraint is to generate similar displacements for bonded atoms. This more general equation can be easily reduced to a form suitable for isotropic displacements. It has been observed that the bond-based ADP restraints, while effective, can still lead to relatively large differences in displacements for atoms close together in space. Additionally, they do not provide any restraints for atoms without bonds such as water molecules at the surface of the macromolecule. Therefore, other approaches to restraining ADPs have been developed,^{53,54} for example:

$$E_{\text{adp}} = \sum_{i=1}^{N_{\text{atoms}}} \left[\sum_{j=1}^{M_{\text{atoms}}} \frac{1}{r_{ij}^p} \frac{(U_{\text{local},i} - U_{\text{local},j})^2}{(U_{\text{local},i} + U_{\text{local},j})^q} \right] \quad [10]$$

where N_{atoms} is the total number of atoms in the model, M_{atoms} is the atoms in the sphere of radius R around an atom i , r_{ij} is a distance between two atoms i and j , $U_{\text{local},i}$ and $U_{\text{local},j}$ are the corresponding isotropic ADPs, and p and q are empirical constants defined by refinement of several hundred macromolecular structures. This function has the benefit of including atoms in the restraint regardless of covalent bonding, and also permitting larger deviations in displacements for atoms close together in space that have large displacements.

1.7.3.3.2 Translation-libration-screw refinement

Many of the displacements in macromolecules are strongly coupled and often rigid body in nature. This observation was made for small molecules many years ago.^{49,55} This has led to the development of very successful rigid-body displacement refinement methods, such as Translation-Libration-Screw (TLS).⁸ In this formalism, the anisotropic displacements of groups of atoms are constrained.

$$\mathbf{U}_{\text{TLS}} = \mathbf{T} + \mathbf{A}\mathbf{L}\mathbf{A}^t + \mathbf{A}\mathbf{S} + \mathbf{S}^t\mathbf{A}^t \quad [11]$$

where the \mathbf{T} , \mathbf{L} , and \mathbf{S} matrix elements are refinable and describe translations (\mathbf{T}), librations (rotation) (\mathbf{L}), and the correlations between \mathbf{T} and \mathbf{L} (\mathbf{S}) for the rigid group. \mathbf{A} is the set of points corresponding to the rest positions of the atoms in the rigid bodies. The translations and librations are defined with respect to a set of orthogonal axes, with translations along the three axes and librations around them. A simultaneous translation and libration results in a corkscrew motion, hence the screw component. The origin of the axes is typically defined as the center of mass of the rigid group, and the directions of the axes are refined.

The choice of rigid groups is a key part of the TLS method. If rigid bodies are chosen inappropriately, it will not be possible to fit the underlying atomic displacements with a constrained model. Currently, rigid groups are assigned either using manual analysis of the structure to identify potentially rigid-body moving domains by visual inspection or by fitting the TLS parameters to the partially refined B-factors of the current model.^{56,57} The latter method is attractive as it provides a numerical method for derivation of the groupings, and is readily available as a web service. However, it still leaves some aspects of subjectivity as the user is still required to choose the number of rigid groups.

1.7.3.4 Other Restraints

Additional constraints or restraints may be used to effectively improve the ratio of observations to parameters. For example, atoms can be grouped so that they move as rigid bodies during refinement, or bond lengths and bond angles can be kept fixed.^{7,58,59} The existence of noncrystallographic symmetry can be used to average over equivalent molecules and thereby reduces noise in the diffraction data.³⁸

1.7.3.4.1 Noncrystallographic symmetry

Two different methods for treating noncrystallographic symmetry are typically used, which are termed 'strict' and 'restrained.' Strict Noncrystallographic symmetry (NCS) assumes that NCS related monomers (or, more generally, protomers) are strictly identical, which permits refinement of a single protomer. In addition to reducing the amount of model inspection and adjustment required during refinement, strict NCS reduces the empirical energy calculation by a factor of approximately n times for n -fold NCS. Moreover, imposition of strict NCS permits averaging of the structure-factor derivatives with respect to the NCS, which improves the signal-to-noise ratio by averaging out noise in the data. Strict NCS is implemented in CNS.

The concept of restraining molecules by NCS symmetry was first introduced by Hendrickson.⁵ The entire crystallographic asymmetric unit is refined and NCS related atoms are restrained to their average positions after least-squares superposition⁶⁰ of $N - 1$ protomers onto a reference protomer by adding an effective energy term

$$E_{\text{NCS}} = k_{\text{NCS}} \sum_{n \in \text{NCS}} \sum_i (r_{i,n} - \bar{r}_i)^2 \quad [12]$$

to the target function, where k_{NCS} is an effective force constant used to weight this term. Isotropic temperature factors can be treated similarly with the restraint term

$$\frac{1}{\sigma_{\text{NCS}}^2} \sum_{n \in \text{NCS}} \sum_i (B_{i,n} - \bar{B}_i)^2 \quad [13]$$

where σ_{NCS} is a target standard deviation for this term.⁵ Different groups of NCS related atoms can be weighed separately, or not included at all. This allows the imposition of NCS only on part of the structure.

The default NCS restraints in programs such as Phenix and CNS are typically very tight, with targets of 0.05 Å rms. At resolutions lower than about 2.5 Å, these tight restraints on NCS (or strict NCS) should usually be applied. At higher resolutions, it may be appropriate to use looser restraints or to remove them altogether. Additionally, if there are segments of the chains that clearly do not obey the NCS relationships, they should be excluded from the NCS restraints. The Phenix and REFMAC programs automate some aspects of NCS restraint generation by performing sequence alignment to identify corresponding chains and superimposing matching chains to calculate the level of structural similarity; those atoms that superpose well are automatically selected for inclusion in the NCS restraints.

1.7.3.5 Ensemble Models

In cases of conformational variability or discrete disorder, there is no single correct solution to the optimization problem eqn [1]. Rather, the X-ray diffraction data represent a spatial and temporal average over all conformations that are assumed by the molecule. Ensembles of structures, which are simultaneously refined against the observed data, may thus be a more appropriate description of the data. This has been used for some time in X-ray crystallography when alternate conformations were modeled locally. Alternate conformations can be generalized to global conformations,^{12,13,61} that is, the model is duplicated n -fold, the corresponding calculated structure factors are added and refined simultaneously against the observed X-ray diffraction data, and each member of the family is chemically 'invisible' to all other members. The number n can be determined by cross-validation.^{13,14}

An advantage of a multiconformer model is that it directly incorporates many possible types of disorders and motions (global disorder, local sidechain disorder, local wagging and rocking motions, etc.). Furthermore, it can be used to automatically detect the most variable regions of the molecule by inspecting the atomic root-mean-square difference around the mean as a function of residue number. Thermal factors of single-conformer models may sometimes be misleading by

underestimating the degree of motion or disorder⁶² and, thus, the multiple-conformer model is a more faithful representation of the diffraction data. A disadvantage of the multiconformer model is that it introduces many more degrees of freedom. However, cross-validated maximum-likelihood refinement can address this problem. For example, the R_{free} and R -values were 0.239 and 0.237 for a single-conformer refinement and 0.231 and 0.230, respectively, for a four-conformer refinement at 50–1.7 Å resolution data of a fragment of mannose-binding protein A¹⁴ illustrated that introduction of multiple conformers did not increase the amount of overfitting compared to the single-conformer case (unpublished results).

Although there are some similarities between averaging individually refined structures and multiconformer models, there are also fundamental differences. For example, in the case of X-ray crystallography, averaging seeks to improve the calculated electron density map by averaging out the noise present in the individual models (see Section 1.7.3.4.1). In contrast, multiconformer refinement seeks to create an ensemble of structures at the final stages of refinement which, taken together, best represent the data. It should be noted that each individual conformer of the ensemble does not necessarily remain a good description of the data since the whole ensemble is refined against the data. Clearly, this method requires high-quality data and a high observation-to-parameter ratio.

1.7.4 Cross-Validation

Cross-validation⁶ plays a fundamental role for validation and in the maximum-likelihood target functions described below. A few remarks about this method are therefore warranted here (for reviews, see References 63,64). For cross-validation, the diffraction data are divided into two sets: a large working set (usually comprising 90% of the data), and a complementary test set (comprising the remaining 10%). The diffraction data in the working set are used in the normal crystallographic refinement process, whereas the test data are not. The cross-validated (or 'free') R -value computed with the test set data is a more faithful indicator of model quality. It provides a more objective guide during the model building and refinement process than the conventional R -value. It also ensures that introduction of additional parameters (e.g., water molecules, relaxation of noncrystallographic symmetry restraints, or multi-conformer models) improves the quality of the model rather than increasing overfitting.

Since the conventional R -value shows little correlation with the accuracy of a model, coordinate-error estimates derived from the Luzzati⁶⁵ or σ_A ¹⁷ methods are unrealistically low. Kleywegt and Brunger⁶³ showed that more reliable coordinate errors can be obtained by cross-validation of the Luzzati or σ_A coordinate-error estimates. The conventional R -value improves as the resolution decreases and the quality of the model worsens. Consequently, coordinate-error estimates do not display the correct behavior either: the error estimates are approximately constant regardless of the resolution and actual coordinate error of the models. However, when cross-validation is used (i.e., the test reflections are used to compute

the estimated coordinate errors), the results are much better: the cross-validated errors are close to the actual coordinate error, and they show the correct trend as a function of resolution.⁶⁶

1.7.5 Optimization Methods

The high-dimensionality of the parameter space of the atomic model (typically three times the number of atoms) introduces many local minima of the target function and, thus, gradient descent methods, such as conjugate-gradient minimization or least-squares methods,⁶⁷ normally do not achieve shifts of atomic positions large enough to fully refine the structure. In a sense, the difficulty of refinement arises from the crystallographic phase problem.

Electron density maps computed by a combination of native crystal amplitudes and experimentally observed phases are sometimes insufficient to allow a complete and unambiguous tracing of the macromolecule. Furthermore, electron density maps for macromolecules are usually obtained at lower than atomic resolution and, thus, are prone to human errors when interpreting the maps. Thus, initial atomic models are likely to contain (partially) incorrect regions and require refinement with a large radius of convergence.

Several algorithms to refine macromolecular crystal structures have been developed over the past 20 years.⁶⁸ These algorithms can be generally classified into constrained or restrained least-squares optimization,^{5,59,69} conjugate-gradient minimization,^{70,71} and simulated annealing (SA) refinement.³⁷

1.7.5.1 Gradient Descent Methods

Restrained least-squares refinement techniques were reviewed by Hendrickson.⁵ An improved algorithm for conjugate-gradient minimization was described by Tronrud,⁷² which employs information about curvature in order to speed convergence and to enable simultaneous refinement of positions and B-factors. In any case, gradient-driven refinement of coordinates can only move atoms within a convergence radius, which is approximately 1.0 Å.⁷³

1.7.5.2 Searching Conformational Space with Simulated Annealing

Annealing denotes a physical process wherein a solid is heated until all particles randomly arrange themselves in a liquid phase, and then is cooled slowly so that all particles arrange themselves in the lowest energy state. By formally defining the target E (eqn [1]) to be the equivalent of the potential energy of the system, one can simulate the annealing process.⁷⁴ There is no guarantee that simulated annealing will find the global minimum, except in the case of an infinitely long search.⁷⁵ Compared to conjugate-gradient minimization, where search directions must follow the gradient, simulated annealing achieves more optimal solutions by allowing motion against the gradient.⁷⁴ The likelihood of uphill motion is determined by a control parameter referred to as temperature. The higher the temperature, the more likely it is that simulated annealing

will overcome barriers. It should be noted that the simulated annealing temperature normally has no physical meaning and merely determines the likelihood of overcoming barriers of the target function.

The simulated annealing algorithm requires a generation mechanism to create a Boltzmann distribution at a given temperature T . Simulated annealing also requires an annealing schedule, that is, a sequence of temperatures $T_1 > T_2 > T_n$ at which the Boltzmann distribution is computed. Implementations of the generation mechanism differ in the way they transit from one set of parameters to another that is consistent with the Boltzmann distribution at a given temperature. The two most widely used generation mechanisms are Metropolis Monte Carlo⁷⁶ and molecular-dynamics⁷⁷ simulations. For X-ray crystallographic refinement, molecular dynamics has proved extremely successful,³⁷ whereas Monte Carlo methods have yet to be shown to be effective.

1.7.5.2.1 Molecular dynamics

A suitably chosen set of atomic parameters can be viewed as generalized coordinates that are propagated in time by the classical (Hamilton) equations of motion.⁷⁸ If the generalized coordinates represent the x, y, z positions of the atoms of a molecule, the Hamilton equations of motion reduce to the more familiar Newton's second law:

$$m_i \left(\frac{\partial^2 r_i}{\partial t^2} \right) = -\nabla_i E \quad [14]$$

The quantities m_i and r_i are respectively the mass and coordinates of atom i , and E is given by eqn [1]. The solution of the partial differential equations (eqn [13]) is achieved numerically using finite-difference methods.^{77,79} This approach is referred to as molecular dynamics.

Initial velocities for the integration of eqn [13] are usually assigned randomly from a Maxwell distribution at the appropriate temperature. Assignment of different initial velocities will produce a somewhat different structure after simulated annealing. By performing several refinements with different initial velocities, one can therefore improve the chances of success of simulated annealing refinement. Furthermore, this improved sampling can be used to determine discrete disorder and conformational variability.

1.7.5.2.2 Temperature control

Simulated annealing requires the control of the temperature during molecular dynamics. The current temperature of the simulation (T_{current}) is computed from the kinetic energy

$$E_{\text{kin}} = \sum_{i=1}^{n \text{ atoms}} \frac{1}{2} m_i \left(\frac{\partial r_i}{\partial t} \right)^2 \quad [15]$$

of the molecular-dynamics simulation,

$$T_{\text{current}} = \frac{2E_{\text{kin}}}{3nk_b} \quad [16]$$

Here n is the number of degrees of freedom and k_b is Boltzmann's constant. One commonly used approach to

control the temperature of the simulation consists of coupling the equations of motion to a heat bath. A friction term (\mathbf{g}_i) to control the temperature

$$-m_i\gamma_i v_i \left(1 - \left(\frac{T}{T_{\text{current}}}\right)\right) \quad [17]$$

can be added to right-hand side of eqn [13], where v_i are the velocities of the atoms.⁸⁰ This method generalizes the concept of friction by allowing a negative friction coefficient and by determining the friction coefficient and its sign by the ratio of the current simulation temperature to the target temperature T_{current} .

1.7.5.2.3 Torsion-angle parameterization

Although Cartesian (i.e., flexible bond lengths and bond angles) minimization and molecular dynamics place restraints on bond lengths and bond angles, one might want to implement these restrictions as constraints, that is, fixed bond lengths and bond angles.⁵⁸ This is supported by the observation that the deviations from ideal bond lengths and bond angles are usually small in macromolecular X-ray crystal structures. Indeed, fixed-length constraints have been applied to crystallographic refinement by least-squares minimization.⁵⁸ However, it took some time until efficient and robust algorithms were developed for molecular dynamics in torsion-angle space.^{81–83} The first implementation of an exact torsion-angle integrator for macromolecules was described in 1994⁷ and is implemented in the CNS package.^{25,84} This approach retains the Cartesian coordinate formulation of the target function and its derivatives with respect to atomic coordinates so that the calculation remains relatively straightforward and can be applied to any macromolecule or their complexes.⁷ In this formulation, the expression for the acceleration becomes a function of positions and velocities. Iterative equations of motion for constrained dynamics in this formulation can be derived and solved by finite-difference methods.⁷⁹ This method is numerically very robust and has a significantly increased radius of convergence in crystallographic refinement compared to Cartesian molecular dynamics.^{7,10,20}

1.7.5.2.4 Multi-start simulated annealing refinement

Multiple simulated annealing refinements will generally produce somewhat different structures, some of which may be better (as assessed, for example, in terms of the free R -value) than others. This approach offers several advantages. First, a more optimum structure can be obtained from multiple trials as opposed to a single simulated annealing calculation. Second, each member of the family of refined structures may be better in different regions of the molecule. Thus, by examining the ensemble during model-building, one may gain insights into possible local conformations of the molecule. Third, the structure factors of all structures of the family may be averaged. This averaging will reduce the effect of local errors (noise) that are presumably different in each member of the family.

Torsion-angle molecular-dynamics simulated annealing with the maximum-likelihood target (eqn [4]) performed on human heterogeneous ribonucleoprotein A1, hnRNP⁸⁵ showed that averaging produced the least model-biased map (as indicated by the lowest free R -value and the lowest $R_{\text{free}} - R$

difference) with the polypeptide backbone being completely connected.⁸⁶ This example is another demonstration that cross-validation of the R -value is essential for assessing model correctness⁶ since the normal R -value decreases with increasing model-bias of the electron density maps whereas the free R -value shows the correct behavior.

1.7.6 Special Considerations at Low Resolution

1.7.6.1 Treatment of Weak Intensities

With the emergence of maximum likelihood-based refinement methods,^{10,11} it is possible to include all weak diffraction data in refinement. Clearly, this is especially important when analyzing crystals that diffract only to low resolution. Weak reflections with large experimental error estimates are automatically down-weighted in the likelihood-based target function. R. J. Read suggested usage of the resolution-dependence of σ_A as a guide to determine the effective resolution limit.⁸⁷ This approach was applied to set the resolution limit for p97/VCP in complex with ADP; the suggested resolution limit corresponded to a conventional I/σ_1 cut-off of 1.2. For the ADP·AlF_x and AMP-PNP ligated structures, this approach resulted in I/σ_1 cut-offs as low as 0.8.²⁹ Slight improvements in electron density maps are observed on inclusion of all weak diffraction data in the refinement and map calculations. A possible generalization of this approach would be to take into account anisotropic diffraction since this is a common place for crystals of large macromolecular assemblies.

1.7.6.2 Thermal Factor Sharpening of Electron Density Maps

Thermal (B)-factor sharpening is a useful tool for the enhancement of low-resolution electron density maps.^{28,29,88,89} Thermal factor sharpening entails the use of a negative B_{sharp} value in a resolution-dependent weighing scheme applied to a particular electron density map:

$$F_{\text{sharpened_map}} = \exp(-B_{\text{sharp}} \sin^2 \theta / \lambda^2) \times F_{\text{map}} \quad [18]$$

where F_{map} is the structure factor of the particular electron density map, $F_{\text{sharpened_map}}$ is the structure factor of the sharpened map, θ is the reflecting angle, and λ is the wavelength of the X-ray radiation. A reasonable choice for B_{sharp} is the negative Wilson B value of the diffraction data. Since the customary procedure to obtain the Wilson B value requires high-resolution diffraction data, a maximum likelihood-based method should be used for low-resolution data sets as described by Popov and Bourenkov⁹⁰ and implemented in Phenix.³¹

Applying a negative B_{sharp} value effectively up-weights higher resolution terms. The result of this weighing scheme is more detailed higher resolution features such as sidechain conformations. However, the cost of this can be increased noise throughout the electron density map. Thus, thermal factor sharpening is a density modification technique that is only as good as the diffraction data and phases that are available, and, so, the original un-weighted electron density maps should always be considered. B-factor sharpening

provided some utility for the refinement of the original p97/VCP models, but it proved even more useful for the re-refined structures due to improved model phase accuracy.⁹¹

1.7.6.3 Deformable Elastic Network Refinement

Low-resolution X-ray diffraction data at 5 Å contains, in principle, sufficient information to determine the true structure (the 'target structure') since the number of observable diffracted intensities exceeds the number of torsion-angle degrees of freedom of a macromolecule (W. A. Hendrickson, personal communication). Although an exhaustive conformational search in torsion-angle space against the diffraction data should lead to an accurate structure at 5 Å resolution, such a search is computationally intractable. A recent approach aids the search by adding known information to the observed data at low resolution. Instead of just adding generic information about macromolecular stereochemistry (idealized chemical bond lengths, bond angles, and atom sizes that heralded the era of reciprocal-space restrained refinement),^{5,70} *specific* information for the particular macromolecule(s) or complex is added in the form of Deformable Elastic Network (DEN) restraints, this information derived from known structures of homologous proteins or domains (the 'reference model').^{15,92}

The target structure often differs from the reference model by large-scale deformations related to the approximate conservation of local polypeptide geometry as sequence and function evolve. How can such deformations be mathematically described? An early approach⁹³ used low-frequency normal modes, shown to reproduce large-scale collective changes in structures with very few degrees of freedom;⁹⁴ it has been used to refine protein structures with low-resolution X-ray or cryo-electron microscopy data.^{95,96} DEN defines springs between selected atom pairs using the reference model as the template. The equilibrium distance of each spring (distance at which its potential energy is minimum) is initially set to the distance between these atoms in the starting structure for refinement. As torsion-angle molecular dynamics against a combined target function (comprising diffraction data, DEN, and empirical energy function) proceeds, the equilibrium lengths of the DEN network are adjusted to incorporate the distance information from the reference model which can be equal to the starting model of refinement or an external model. The degree of this adjustment is controlled by a parameter, γ . The optimum value for γ and for the weight of the DEN energy term can be obtained empirically by performing a series of DEN-refinements with different combinations of these parameters, and then selecting the DEN-refinement with the best free *R*-value; to further increase the chances of obtaining an optimum solution, multiple trials should be performed for each parameter pair.¹⁵

1.7.7 Conclusions and Outlook

Crystallographic refinement is now a well-established technique and routinely used for crystal structures that diffract to at least 4 Å resolution. In this review, authors discussed most of the commonly used techniques to refine such crystal

structures. However, there are situations where refinement remains a challenge, especially at very low resolution. As an example, authors mention the crystal structures of the ATPase p97/VCP, consisting of an N-terminal domain followed by a tandem pair of ATPase domains (D1, D2).²⁸ The structures were originally solved by molecular replacement with the high-resolution structure of the N-D1 fragment of p97/VCP, whereas the D2 domain was manually built using homology to the D1 domain as a guide. The structure of the D2 domain alone was subsequently solved at 3 Å resolution.⁹¹ The refined model of D2 and the high-resolution structure of the N-D1 fragment were then used as starting models for re-refinement against the low-resolution diffraction data for full-length p97. The re-refined full-length models showed significant improvement in both secondary structure and *R* values. The free *R* values dropped as much as 5% compared to the original structure refinements, indicating that refinement is meaningful at low resolution and that there is information in the diffraction data even at ~4 Å resolution that objectively assesses the quality of the model. Thus, *de novo* model building which is problematic at low resolution and refinement should start from high-resolution crystal structures whenever possible.

In retrospect, the recently developed DEN approach that includes information from a reference model,¹⁵ would have enabled an improved refinement of the p97 structures than the original structures without knowledge of the high-resolution D2 domain structure (see online tutorial for CNS v1.3). As this example shows, knowledge of the known protein structure 'universe' can assist refinement when the available diffraction data are insufficient to fully determine the structure. The DEN method is just a first step to make use of this vast structural knowledge. Methods will be developed that incorporate more prior chemical knowledge into the refinement and automated building of macromolecular structures. We anticipate that further developments in this area will allow for routine refinement of structures at even lower resolution, a timely topic considering the emergence of an entirely new class of X-ray sources and free-electron lasers, that will enable the structure determination of nanocrystals and, eventually, single molecules.

Acknowledgments

PDA would like to thank the NIH (grant GM063210) and the Phenix Industrial Consortium for support of the Phenix project. ATB acknowledges support by HHMI. This work was supported in part by the US Department of Energy under Contract No. DE-AC02-05CH11231.

References

- [1] Garman, E. F. Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta. Crystallogr. D: Biol. Crystallogr.* **2010**, *66*, 339–351.
- [2] Dauter, Z. Efficient use of synchrotron radiation for macromolecular diffraction data collection. *Prog. Biophys. Mol. Biol.* **2005**, *89*, 153–172.
- [3] Walden, H. Selenium incorporation using recombinant techniques. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 352–357.

- [4] Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **1991**, *254*, 51–58.
- [5] Hendrickson, W. A. Stereochemically restrained refinement of macromolecular structures. *Meth. Enzymol.* **1985**, *115*, 252–270.
- [6] Brunger, A. T. The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **1992**, *355*, 472–474.
- [7] Rice, L. M.; Brunger, A. T. Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins: Struct. Function Genet.* **1994**, *19*, 277–290.
- [8] Winn, M. D.; Isupov, M. N.; Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D: Biol. Crystallogr.* **2001**, *57*, 122–123.
- [9] Pannu, N. S.; Read, R. J. Improved structure refinement through maximum likelihood. *Acta Cryst.* **1996**, *A52*, 659–668.
- [10] Adams, P. D.; Pannu, N. S.; Read, R. J.; Brunger, A. T. Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proc. Natl. Acad. Sci. USA.* **1997**, *94*, 5018–5023.
- [11] Pannu, N. S.; Murshudov, G. N.; Dodson, E. J.; Read, R. J. Incorporation of prior phase information strengthens maximum likelihood structural refinement. *Acta Cryst.* **1998**, *D54*, 1285–1294.
- [12] Kuriyan, J.; Ösapay, K.; Burley, S. K.; Brunger, A. T.; Hendrickson, W. A.; Karplus, M. Exploration of disorder in protein structures by X-ray restrained molecular dynamics. *Proteins* **1991**, *10*, 340–358.
- [13] Burling, F. T.; Brunger, A. T. Thermal motion and conformational disorder in protein crystal structures: comparison of multi-conformer and time-averaging models. *Israel J. Chem.* **1994**, *34*, 165–175.
- [14] Burling, F. T.; Weis, W. I.; Flaherty, K. M.; Brunger, A. T. Direct observation of protein solvation and discrete disorder with experimental crystallographic phases. *Science* **1996**, *271*, 72–77.
- [15] Schröder, G. F.; Levitt, M.; Brunger, A. T. Super-resolution biomolecular crystallography with low-resolution data. *Nature* **2010**, *464*, 1218–1222.
- [16] Silva, A. M.; Rossmann, M. G. The refinement of southern bean mosaic virus in reciprocal space. *Acta Cryst.* **1985**, *B41*, 147–157.
- [17] Read, R. J. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst.* **1986**, *A42*, 140–149.
- [18] Bricogne, G. A multisolution method of phase determination by combined maximization of entropy and likelihood. III. Extension to powder diffraction data. *Acta Cryst.* **1991**, *A47*, 803–829.
- [19] Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.* **1997**, *D53*, 240–255.
- [20] Adams, P. D.; Pannu, N. S.; Read, R. J.; Brunger, A. T. Extending the limits of molecular replacement through combined simulated annealing and maximum likelihood refinement. *Acta Cryst.* **1999**, *D55*, 181–190.
- [21] Bricogne, G. Bayesian statistical viewpoints on structure determination: basic concepts and examples. *Meth. in Enzymol.* **1997**, *276*, 361–423.
- [22] Skubák, P.; Murshudov, G. N.; Pannu, N. S. Direct incorporation of experimental phase information in model refinement. *Acta Crystallogr. D: Biol. Crystallogr.* **2004**, *60*, 2196–2201.
- [23] Rice, L. M.; Earnest, T. N.; Brunger, A. T. Single wavelength anomalous diffraction phasing revisited. *Acta Cryst.* **2000**, *D56*, 1413–1420.
- [24] Grosse-Kunstleve, R. W.; Adams, P. D. On the handling of atomic anisotropic displacement parameters. *J. Appl. Cryst.* **2002**, *35*, 477–480.
- [25] Brunger, A. T. Version 1.2 of the crystallography and NMR system. *Nat. Protoc.* **2007**, *2*, 2728–2733.
- [26] Jiang, J. S.; Brünger, A. T. Protein hydration observed by X-ray diffraction. Solvation properties of penicillopepsin and neuraminidase crystal structures. *J. Mol. Biol.* **1994**, *243*, 100–115.
- [27] Phillips, S. E. Structure and refinement of oxymyoglobin at 1.6 Å resolution. *J. Mol. Biol.* **1980**, *142*, 531–554.
- [28] DeLaBarre, B.; Brunger, A. T. Complete structure of p97/valosin-containing protein reveals communication between nucleotide domains. *Nat. Struct. Biol.* **2003**, *10*, 856–863.
- [29] DeLaBarre, B.; Brunger, A. T. Nucleotide dependent motion and mechanism of action of p97/VCP. *J. Mol. Biol.* **2005**, *347*, 437–452.
- [30] Afonine, P. V.; Grosse-Kunstleve, R. W.; Adams, P. D. A robust bulk-solvent correction and anisotropic scaling procedure. *Acta Cryst.* **2005**, *D61*, 850–855.
- [31] Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst.* **2010**, *D66*, 213–221.
- [32] Fenn, T. D.; Schnieders, M. J.; Brunger, A. T. A smooth and differentiable bulk-solvent model for macromolecular diffraction. *Acta Crystallogr. D: Biol. Crystallogr.* **2010**, *66*, 1024–1031.
- [33] Engh, R. A.; Huber, R. Accurate bond and angle parameters for X-ray structure refinement. *Acta Cryst.* **1991**, *A47*, 392–400.
- [34] Parkinson, G.; Vojtechovsky, J.; Clowney, L.; Brunger, A. T.; Berman, H. M. New parameters for the refinement of nucleic acid containing structures. *Acta Cryst.* **1996**, *D52*, 57–64.
- [35] Allen, F. H.; Kennard, O.; Taylor, R. Systematic analysis of structural data as a research technique in organic chemistry. *Acc. Chem. Res.* **1983**, *16*, 146–153.
- [36] Tronrud, D. E.; Berkholz, D. S.; Karplus, P. A. Using a conformation-dependent stereochemical library improves crystallographic refinement of proteins. *Acta Crystallogr. D: Biol. Crystallogr.* **2010**, *66*, 834–842.
- [37] Brunger, A. T.; Kuriyan, J.; Karplus, M. Crystallographic R factor refinement by molecular dynamics. *Science* **1987**, *235*, 458–460.
- [38] Weis, W. I.; Brünger, A. T.; Skehel, J. J.; Wiley, D. C. Refinement of the influenza virus hemagglutinin by simulated annealing. *J. Mol. Biol.* **1990**, *212*, 737–761.
- [39] Blanc, E.; Roversi, P.; Vornrhein, C.; Flensburg, C.; Lea, S. M.; Bricogne, G. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D: Biol. Crystallogr.* **2004**, *60*, 2210–2221.
- [40] Pearlman, D. A.; Kim, S.-H. Atomic charges for DNA constituents derived from single-crystal X-ray diffraction data. *J. Mol. Biol.* **1990**, *211*, 171–187.
- [41] Karplus, M.; Petsko, G. A. Molecular dynamics simulations in biology. *Nature* **1990**, *347*, 631–639.
- [42] Brunger, A. T.; Karplus, M.; Petsko, G. A. Crystallographic refinement by simulated annealing: application to a 1.5 Å resolution structure of Crambin. *Acta Cryst.* **1989**, *A45*, 50–61.
- [43] Brunger, A. T.; Krukowski, A.; Erickson, J. Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Cryst.* **1990**, *A46*, 585–593.
- [44] Fujinaga, M.; Gros, P.; van Gunsteren, W. F. Testing the method of crystallographic refinement using molecular dynamics. *J. Appl. Cryst.* **1989**, *22*, 1–8.
- [45] Linge, J. P.; Williams, M. A.; Spronk, C. A.; Bonvin, A. M.; Nilges, M. Refinement of protein structures in explicit solvent. *Proteins* **2003**, *50*, 496–506.
- [46] Moulinier, L.; Case, D. A.; Simonson, T. Reintroducing electrostatics into protein X-ray structure refinement: bulk solvent treated as a dielectric continuum. *Acta Crystallogr. D: Biol. Crystallogr.* **2003**, *59*, 2094–2103.
- [47] Fenn, T. D.; Schnieders, M. J.; Mustyakimov, M.; Langan, P.; Pande, V. S.; Brunger, A. T. Reintroducing electrostatics into macromolecular crystallographic refinement: application to neutron crystallography and DNA Hydration. *Structure* **2011**, *19*, 523–533.
- [48] Trueblood, K. N.; Bürgi, H. B.; Burzlaff, H.; Dunitz, J. D.; Gramaccioni, C. M.; Schulz, H. H.; Shmueli, U.; Abrahams, S. C. Atomic displacement parameter nomenclature. Report of a subcommittee on atomic displacement parameter nomenclature. *Acta Cryst.* **1996**, *A52*, 770–781.
- [49] Dunitz, J. D.; White, D. N. J. Non-rigid-body thermal-motion analysis. *Acta Cryst.* **1973**, *A29*, 93–94.
- [50] Prince, E.; Finger, L. W. Use of constraints on thermal motion in structure refinement of molecules with librating side groups. *Acta Cryst.* **1973**, *B29*, 179–183.
- [51] Sheriff, S.; Hendrickson, W. A. Description of overall anisotropy in diffraction from macromolecular crystals. *Acta Cryst.* **1987**, *A43*, 118–121.
- [52] Hirshfeld, F. L. Can X-ray data distinguish bonding effects from vibrational smearing? *Acta Cryst.* **1976**, *A32*, 239–244.
- [53] Afonine P. V., Grosse-Kunstleve R. W., Adams P. D. (2005b) The Phenix refinement framework. CCP4 newsletter, July 2005; Contribution 8.
- [54] Terwilliger, T. C.; Grosse-Kunstleve, R. W.; Afonine, P. V.; Moriarty, N. W.; Zwart, P. H.; Hung, L. W.; Read, R. J.; Adams, P. D. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Cryst.* **2008**, *D64*, 61–69.
- [55] Schomaker, V.; Trueblood, K. N. On the rigid-body motion of molecules in crystals. *Acta Cryst.* **1968**, *B24*, 63–76.
- [56] Painter, J.; Merritt, E. A. A molecular viewer for the analysis of TLS rigid-body motion in macromolecules. *Acta Crystallogr. D: Biol. Crystallogr.* **2005**, *61*, 465–471.

- [57] Painter, J.; Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D: Biol. Crystallogr.* **2006**, *62*, 439–450.
- [58] Diamond, R. A real-space refinement procedure for proteins. *Acta Cryst.* **1971**, *A27*, 436–452.
- [59] Sussman, J. L.; Holbrook, S. R.; Church, G. M.; Kim, S.-H. Structure-factor least-squares refinement procedure for macromolecular structure using constrained and restrained parameters. *Acta Cryst.* **1977**, *A33*, 800–804.
- [60] Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Cryst.* **1976**, *A32*, 922–923.
- [61] Gros, P.; van Gunsteren, W. F.; Hol, W. G. J. Inclusion of thermal motion in crystallographic structures by restrained molecular dynamics. *Science* **1990**, *249*, 1149–1152.
- [62] Kuriyan, J.; Petsko, G. A.; Levy, R. M.; Karplus, M. Effect of anisotropy and anharmonicity on protein crystallographic refinement. *J. Mol. Biol.* **1986**, *190*, 227–254.
- [63] Kleywegt, G. J.; Brunger, A. T. Checking your imagination: applications of the free R value. *Structure* **1996**, *4*, 897–904.
- [64] Brunger, A. T. Free R value: cross-validation in crystallography. *Methods Enzym.* **1997**, *277*, 366–396.
- [65] Luzzati, V. Traitement statistique des erreurs dans la détermination des structures cristallines. *Acta Cryst.* **1952**, *5*, 802–881.
- [66] Brunger, A. T.; Adams, P. D.; Rice, L. M. Enhanced Macromolecular Refinement by Simulated Annealing. In *International Tables for Crystallography, Volume F: Macromolecular Crystallography*; Rossmann, M. G.; Arnold, E., Eds.; Kluwer Academic Publishers: Dordrecht, 2001; pp 375–381.
- [67] Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes*; Cambridge University Press: Cambridge, 1986; 498–546.
- [68] Stout, G. H.; Jensen, L. H. *X-ray Structure Determination*; Wiley: New York, 1989.
- [69] Konnert, J. H.; Hendrickson, W. A. A restrained-parameter thermal-factor refinement procedure. *Acta Cryst.* **1980**, *A36*, 344–350.
- [70] Jack, A.; Levitt, M. Refinement of large structures by simultaneous minimization of energy and R factor. *Acta Crystallogr.* **1983**, *A34*, 931–935.
- [71] Tronrud, D. E.; Ten Eyck, L. F.; Matthews, B. W. An efficient general-purpose least-squares refinement program for macromolecular structures. *Acta Cryst.* **1987**, *A43*, 489–501.
- [72] Tronrud, D. E. Conjugate-direction minimization: an improved method for the refinement of macromolecules. *Acta Crystallogr. A* **1992**, *48*, 912–916.
- [73] Agarwal, R. C. A new least-squares refinement technique based on the fast Fourier transform algorithm. *Acta Cryst.* **1978**, *A34*, 791–809.
- [74] Kirkpatrick, S.; Gelatt, C. D.; Vecchi, Jr. M. P. Optimization by simulated annealing. *Science* **1983**, *220*, 671–680.
- [75] Laarhoven, P. J. M.; Aarts, E. H. L. *Simulated Annealing: Theory and Applications*; D. Reidel Publishing Company: Dordrecht, 1987.
- [76] Metropolis, N.; Rosenbluth, M.; Rosenbluth, A.; Teller, A.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
- [77] Verlet, L. Computer 'experiments' on classical fluids. I. Thermodynamical properties of Lennard–Jones molecules. *Phys. Rev.* **1967**, *159*, 98–103.
- [78] Goldstein, H. *Classical Mechanics*, 2nd ed.; Addison-Wesley Pub. Co.: Reading, Massachusetts, 1980.
- [79] Abramowitz, M.; Stegun, I. *Handbook of Mathematical Functions*; Dover Publications: New York, 1968; 896.
- [80] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- [81] Bae, D.-S.; Haug, E. J. A recursive formulation for constrained mechanical system dynamics: Part I. Open loop systems. *Mech. Struct. Mach.* **1987**, *15*, 359–382.
- [82] Bae, D.-S.; Haug, E. J. A recursive formulation for constrained mechanical system dynamics: Part II. Closed loop systems. *Mech. Struct. Mach.* **1988**, *15*, 481–506.
- [83] Jain, A.; Vaidehi, N.; Rodriguez, G. A fast recursive algorithm for molecular dynamics simulation. *J. Comp. Phys.* **1983**, *106*, 258–268.
- [84] Brunger, A. T.; Adams, P. D.; Clore, G. M.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J.-S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. Crystallography & NMR system (CNS): a new software system for macromolecular structure determination. *Acta Cryst.* **1998**, *D54*, 905–921.
- [85] Shamoo, Y.; Krueger, U.; Rice, L. M.; Williams, K. R.; Steitz, T. A. Crystal structure of the two RNA-binding domains of human hnRNP A1 at 1.75 Å resolution. *Nature Struct. Biol.* **1997**, *3*, 215–222.
- [86] Rice, L. M.; Shamoo, Y.; Brunger, A. T. Phase improvement by multi start simulated annealing refinement and structure factor averaging. *J. Appl. Cryst.* **1998**, *31*, 798–805.
- [87] Ling, H.; Boodhoo, A.; Hazes, B.; Cummings, M. D.; Armstrong, G. D.; Brunton, J. L.; Read, R. J. Structure of the shiga-like toxin I B-pentamer complexed with an analogue of its receptor Gb3. *Biochemistry* **1998**, *37*, 1777–1788.
- [88] Bass, R. B.; Strop, P.; Barclay, M.; Rees, D. C. Crystal structure of *Escherichia coli* MscS, a voltage-modulated and mechanosensitive channel. *Science* **2002**, *298*, 1582–1587.
- [89] DeLaBarre, B.; Brunger, A. T. Considerations for the refinement of low-resolution crystal structures. *Acta Crystallogr. D: Biol. Crystallogr.* **2006**, *62*, 923–932.
- [90] Popov, A. N.; Bourenkov, G. P. Choice of data-collection parameters based on statistic modeling. *Acta Crystallogr.* **2003**, *D59*, 1145–1153.
- [91] Davies, J. M.; Brunger, A. T.; Weis, W. I. Improved structures of full-length p97, an AAA ATPase: implications for mechanisms of nucleotide-dependent conformational change. *Structure* **2008**, *16*, 715–726.
- [92] Schröder, G. F.; Brunger, A. T.; Levitt, M. Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution. *Structure* **2007**, *15*, 1630–1641.
- [93] Diamond, R. On the use of normal modes in thermal parameter refinement: theory and application to the bovine pancreatic trypsin inhibitor. *Acta Crystallogr. A* **1990**, *46*, 425–435.
- [94] Levitt, M.; Sander, C.; Stern, P. S. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Biol.* **1985**, *181*, 423–447.
- [95] Delarue, M.; Dumas, P. On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 6957–6962.
- [96] Tama, F.; Miyashita, O.; Brooks, 3rd C. L. Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. *J. Mol. Biol.* **2004**, *337*, 985–999.