



ARBOR Ciencia, Pensamiento y Cultura

Vol. 191-772, marzo-abril 2015, a219 | ISSN-L: 0210-1963

doi: <http://dx.doi.org/10.3989/arbor.2015.772n2005>

CELEBRATING 100 YEARS OF MODERN CRYSTALLOGRAPHY / CIEN AÑOS DE CRISTALOGRAFÍA MODERNA

MACROMOLECULAR CRYSTALLOGRAPHIC STRUCTURE REFINEMENT

Pavel V. AfonineLawrence Berkeley National Laboratory
PAfonine@lbl.gov**Alexandre Urzhumtsev**Centre for Integrative Biology, IGBMC, CNRS-INSERM-UdS
Université de Lorraine
sacha@igbmc.fr**Paul D. Adams**Lawrence Berkeley National Laboratory
University of California Berkeley
pdadams@lbl.gov

REFINAMIENTO DE ESTRUCTURAS MACROMOLECULARES CRISTALOGRÁFICAS

Citation/Cómo citar este artículo: Afonine, P. V.; Urzhumtsev, A. and Adams, P. D. (2015). "Macromolecular Crystallographic Structure Refinement". *Arbor*, 191 (772): a219. doi: <http://dx.doi.org/10.3989/arbor.2015.772n2005>

Copyright: © 2015 CSIC. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial (by-nc) Spain 3.0 License.

Received: September 12, 2014. Accepted: February 13, 2015.

ABSTRACT: Model refinement is a key step in crystallographic structure determination that ensures final atomic structure of macromolecule represents measured diffraction data as good as possible. Several decades have been put into developing methods and computational tools to streamline this step. In this manuscript we provide a brief overview of major milestones of crystallographic computing and methods development pertinent to structure refinement.

RESUMEN: El refinamiento es un paso clave en el proceso de determinación de una estructura cristalográfica al garantizar que la estructura atómica de la macromolécula final represente de la mejor manera posible los datos de difracción. Han hecho falta varias décadas para poder desarrollar nuevos métodos y herramientas computacionales dirigidas a dinamizar esta etapa. En este artículo ofrecemos un breve resumen de los principales hitos en la computación cristalográfica y de los nuevos métodos relevantes para el refinamiento de estructuras.

KEYWORDS: bulk-solvent; constraints; fast gradient calculation; Fourier maps; maximum-likelihood; minimization; neutrons; optimization; refinement; restraints; structure factors; X-rays.

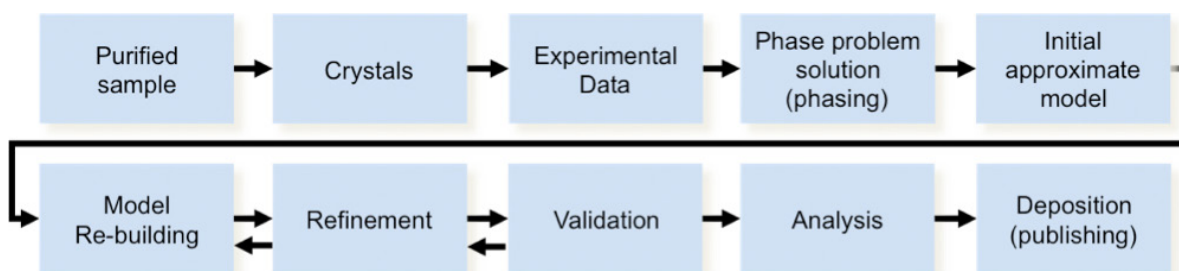
PALABRAS CLAVE: cálculos rápidos de gradiente; constricciones; factores de estructura; mapas de Fourier; máxima verosimilitud; medio acuoso; minimización; neutrones; optimización; rayos-X; refinamiento; restricciones.

INTRODUCTION

Crystallographic structure determination is a complex procedure that involves a number of very diverse steps, shown in Figure 1. It begins with identifying an object of interest (bio-macromolecule), and extracting a sample of it that is sufficiently pure such it can be crystallized. These crystals (that are required to increase the signal from a single molecule) are then used to carry out a diffraction experiment that results in the recorded intensities of diffracted beams (that can be, for example, X-rays, neutrons, electrons). The fundamental problem is that the diffraction experiment does not directly provide an image of the contents of the crystal but instead only provides the intensities of light scattered from the crystal, while the corresponding phases necessary to reconstruct the image are not measured and therefore lost. This con-

stitutes the phase problem – the most fundamental problem in crystallography. Luckily, the phase problem can be solved with a number of different methods that provide approximate phases. These phases can be used to calculate a Fourier image of the molecule in question. In turn this image can be used to build an atomic model of the macromolecule. This atomic model along with its image can be improved iteratively by means of the procedure called refinement. Once no further improvement can be obtained the final structure is subject to thorough checks in order to assess its correctness physically, chemically and crystallographically. Finally, the validated structure is typically deposited at the Protein Data Bank (Bernstein *et al.*, 1977; Berman *et al.*, 2000) as well as analyzed of its biological function. In this manuscript we focus on just one step of this procedure – structure refinement.

Figure 1. Typical crystallographic structure determination workflow



WHAT IS CRYSTALLOGRAPHIC MODEL REFINEMENT AND WHY IT IS NEEDED?

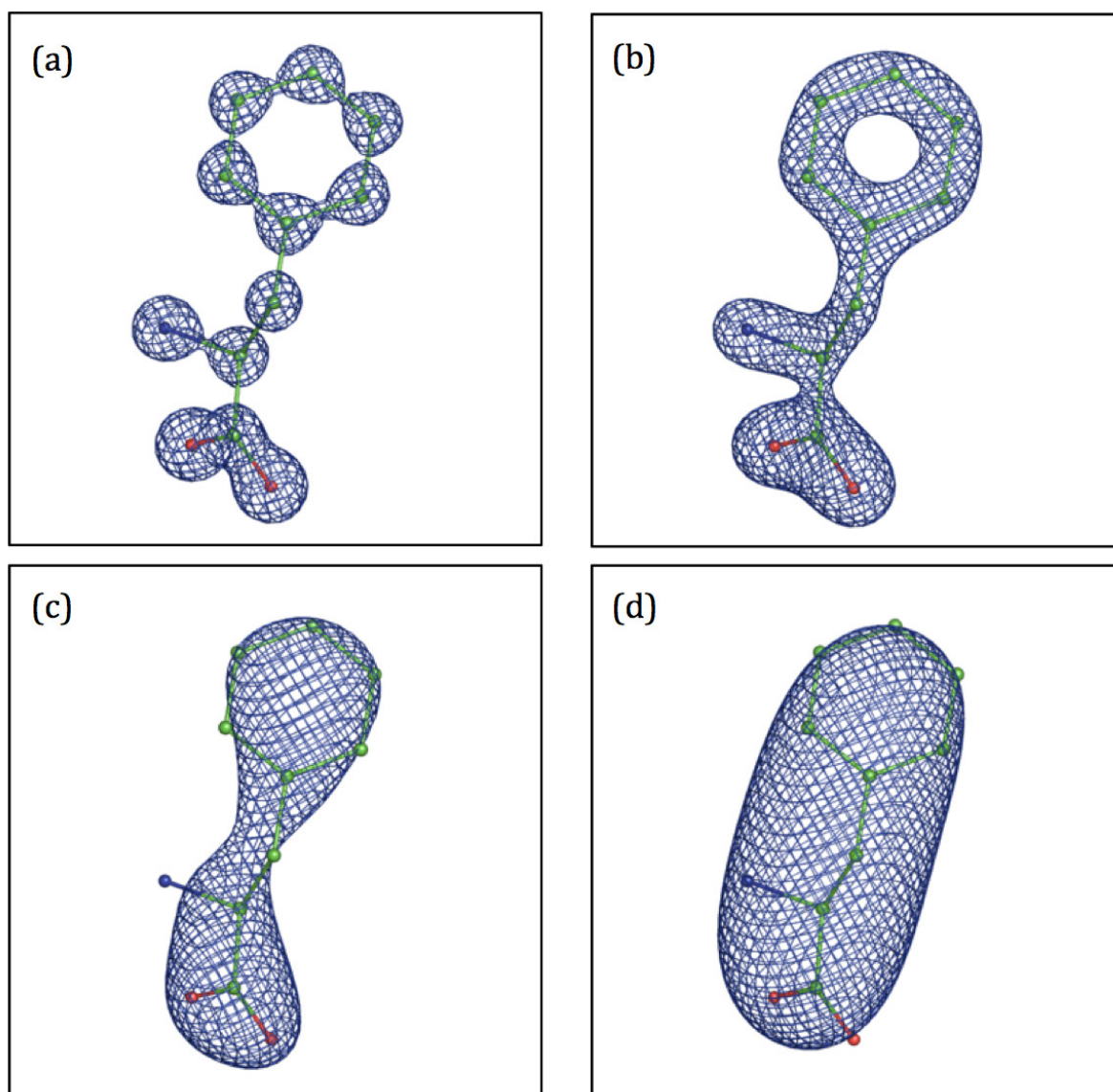
Crystallographic models are constructed as a solution to an inverse problem (notion formally introduced by Ambartsumian, 1929): knowing the results of the experiment (having experimental data), one seeks to obtain the model that reproduces these data. In fact, an atomic model is an interpretation of a physical entity, a distribution of the electron density in a crystal (nuclear density in case of neutron diffraction). This density is generated by atoms that are not point scatterers but are in motion so their positions slightly vary from one unit cell of a crystal to another. In spite of this difference between unit cells the density distribution in a crystal is considered to be a three-dimensional periodic function. In turn this makes it possible to represent the electron density distribution as an infinite three-dimensional set of Fourier

coefficients that are complex numbers identified by three integer indices. A diffraction experiment does not capture the electron density distribution directly but measures the diffracted intensities of the light source scattered from the periodic, regular lattice of the crystal. These intensities result in a three-dimensional finite subset of amplitudes (but not phases!) of complex Fourier coefficients that describes the crystal content. In crystallography, these Fourier coefficients, which we note $F_{obs}(h,k,l)e^{i\varphi(h,k,l)}$, with integer numbers h,k,l , $\{(h,k,l)\}=S$, are called structure factors. Note that phases $\varphi(h,k,l)$ are not available experimentally and need to be determined somehow. The amount and quality of measured intensities defines the quality of final crystal structure model. The better the crystal and more accurate the experiment, the larger this subset S of amplitudes (a crystallographer says: “the higher the resolution of the data set”).

The goal of crystallographic studies is to recover the electron density distribution from measured intensities and interpret this in terms of individual atoms. For small molecules and high-resolution diffraction data, it is possible to recover the atomic information directly from amplitudes. For large molecules and less well diffracting crystals, one first tries to solve an intermediate problem of obtaining phases corresponding to

the measured amplitudes. Measured amplitudes and recovered approximate phases are then used to calculate the corresponding Fourier synthesis, which is a finite-resolution image of the electron density. This image of the electron density is the subject of interpretation in terms of an atomic model. Depending on the data resolution and quality of initial phases the quality of this image may vary substantially (Figure 2).

Figure 2. Illustrative example of an exact electron density distribution (a) and its Fourier images at 2, 3 and 6 Å resolutions (panels b, c, and d, correspondingly)



This means that models obtained from building atoms into maps calculated using initial approximate phases are often inexact and in most cases are insufficient to derive the required structural conclusions: "Structure determination by the methods of X-ray crystallography may be divided into two classes: those in which the object of the investigation is to determine the positions of the atoms with sufficient accuracy to give a general picture of the crystal structure, and those in which the object is to measure as accurately as possible the bond lengths and bond angles between the atoms." (Cochran, 1948). A procedure called refinement needs to be applied to improve the quality of models obtained by original interpretation of diffraction data. In what follows we focus on the refinement procedure, the history of its development, principal programs (Table 1) and their main features.

ture, and those in which the object is to measure as accurately as possible the bond lengths and bond angles between the atoms." (Cochran, 1948). A procedure called refinement needs to be applied to improve the quality of models obtained by original interpretation of diffraction data. In what follows we focus on the refinement procedure, the history of its development, principal programs (Table 1) and their main features.

Table 1. Macromolecular refinement programs using reciprocal-space refinement targets. Programs limited to pure geometry idealization or real-space refinement only are not included. All programs except MOPRO/MOLLY use a spherical atom model with isotropic or anisotropic scattering factors. MOPRO/MOLLY uses a multipolar atom description. Some information may be approximate since program descriptions in literature do not always correspond to the actual program versions and technical documentation is not available for all programs. Here: LS – least-square (amplitudes, F, or intensities, I), ML - maximum likelihood, ML* - maximum likelihood via LS (Adams et al., 1997; Afonine, Lunin and Urzhumtsev, 2003), P ϕ - phase target, (Δd)² - geometric target expressed via quadratic distances, R_E – energy distortion (bond lengths, bond angles, etc), wR_E – weighted energy terms, R_{ref} – different kinds of reference models, RG – rigid groups, TA – torsion angles, ∇ - various gradient methods, PGD – preconditioned conjugate directions, MD – molecular dynamics, ∇^2 - second-matrix derivatives methods; FFT_x – FFT used to calculate structure factors, FFT_v – FFT used to calculate the gradient. Non-geometric restraints include those on ADPs, on parameters of rigid groups, various potentials, etc.

Program	Date	Authors	R _{diff}	R _{map}	R _{geom}	non-geom restr.	constr.	optimization	FFT _x	FFT _v
SHELX, SHELXL	1970, 1997	Sheldrick, Schneider	LS _I	no	(Δd) ² other	no	RG, H _{riding}	∇^2	no	no
PROLSQ	1976, 1980	Konnert, Hendrickson	LS _F	no	(Δd) ²	no	no	∇	no	no
CORELS	1977	Sussman <i>et al.</i>	LS _F	no	(Δd) ²	no	RG	∇	no	no
	1978	Agarwal	LS _F	no	no	no	no	∇	yes	yes
EREF	1978	Jack, Levitt	LS _F	no	R _E	no	no	∇	yes	yes
RESTRAIN	1985, 1989	Haneef <i>et al.</i> , Driessen	LS _F , P ϕ	no	R _E other	no	NCS	lin.eq.	no	no
PROFFT	1987	Finzel	LS _F	no	(Δd) ²	no	no	∇	yes	yes
NUCLSQ	1988	Westhof <i>et al.</i>	LS _F	no	(Δd) ²	no	no	∇	no	no
FROG	1985, 1989	Urzhumtsev <i>et al.</i>	LS _F , LS _I , ML*, P ϕ	yes	wR _E other	yes	RG, NCS	PGD	yes	yes
TNT	1987	Tronrud <i>et al.</i>	LS _F	yes	R _E	no	no	PGD	yes	yes
XPLOR	1987	Brunger <i>et al.</i>	LS _F , ML*, P ϕ	no	R _E other	yes	RG, TA NCS	MD, ∇ , ∇_{LBFGS}	yes	yes
REFMAC	1997	Murshudov <i>et al.</i>	ML, P ϕ	no	R _E other	yes	RG	∇	yes	yes
MOLLY/ MOPRO	1978, 2001, 2005	Hansen & Coppens, Guillot, Jelsch	LSF	no	no	yes	no	sparse ∇^2	no/ yes	not reported
MAIN	1998	Turk	LS _F , ML	yes	R _E	no	yes	∇	yes	yes
CNS	1998	Brunger <i>et al.</i> other authors	LS _F , ML*, P ϕ	yes	R _E other	yes	RG, TA NCS	MD, ∇ , ∇_{LBFGS}	yes	yes
primeX	2012	Bell <i>et al.</i>	ML	not reported	yes	yes	not reported	not reported	yes	yes
BUSTER-TNT	2004	Blanc, Bricogne <i>et al.</i> Tronrud,	ML	yes	R _E other	yes	yes	PGD	yes	yes
phenix.refine	2004, 2012	Adams, Afonine, <i>et al.</i>	LS _F , ML, P ϕ	yes	R _E other	yes	RG, TA H _{riding}	∇_{LBFGS} , MD	yes	yes

HOW REFINEMENT CAN BE PERFORMED AND WHY IT IS DIFFICULT

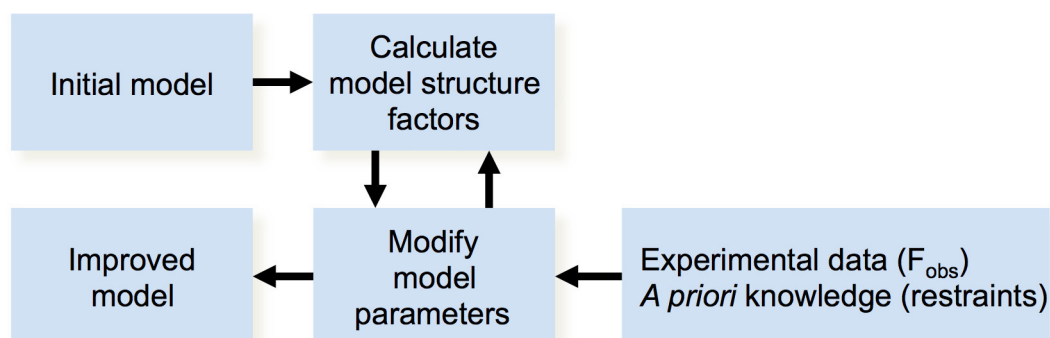
Improving a model means modification of its parameters, resulting in another model that better describes the experimental data. A way to link the parameters describing the model to the available experimental data is to define some function (target) such that its value consistently decreases (or increases) as model improves. Thus refinement of atomic models (simply “refinement” in what follows) can be thought of as an optimization problem (Hughes, 1941; Booth, 1947).

Since refinement can be formulated as an optimization problem the following needs to be defined: a) the *model and parameters* describing the model, b) a *function* that relates the model parameters to experimental data (target or goal function) and c) an *optimization method* that will be used to optimize (typically

minimize) the target function with respect to model parameters (see for example, Afonine *et al.*, 2012).

Model parameters are the variables that describe the crystal and its content. For example, these may be coordinates of the atoms, parameters describing atomic vibrations, disorder, descriptors of the solvent continuum and so on. Once these parameters are defined they can be used to calculate structure factors from the model. The amplitudes of calculated structure factors are then matched against the experimentally measured structure factor amplitudes and the target function is evaluated. An optimization method is then used to decide how the model parameters can be changed such that the target function value decreases (in case of minimization). Once this decision is made, the new set of structure factors is calculated from the updated model and matched against the measured values again. This is repeated several times until convergence (Figure 3).

Figure 3. Schematic representation of the crystallographic structure refinement workflow



Diffraction theory (Ewald, 1913) shows that each atom contributes to each structure factor. The simplest comparison of diffraction data calculated from the model with experimental data can be performed using the standard least-square target for structure factor amplitudes (Hughes, 1941; Booth, 1947). This means that if the structure is composed of K independent atoms (for macromolecules, K is of order 10^4 or larger) and we have M Fourier coefficients (structure factors; for macromolecules, M is also of order 10^4 or larger), then the number of computer operations to calculate a single set of structure factors from a model and compare their amplitudes with the experimental values would require an order of 10^{10} computer operations.

Steepest descent is a simple and powerful optimization method that can be employed to minimize a refinement target. Using this method requires a vector of partial derivatives (gradient) of the refinement target with respect to the model parameters. These derivatives can be calculated either by formulae or by finite difference methods. It's worthwhile to note that the calculation of each partial derivative is as computationally expensive as calculation of the target function itself. In both cases, the number of operations required to calculate a single gradient value is proportional to the number of model parameters (with respect to which this gradient is calculated) giving a number of operations larger than $10^{15} \sim 10 \cdot 10^4 \cdot 10^{10}$

(each atom being characterized by 5-10 parameters). Since the optimization of the target typically requires many iterations to converge to a local minimum of the minimization target, the number of operations may easily rise to 10^{20} .

As a quick illustration we cite Hughes (1941):

"The reduction of the observational equations to normal equations can be carried through a reasonable time, say, three days or less, by an ordinary electrically powered calculator when the number of parameters is about twelve or less... In the case of the h0l data from melanine with over one hundred observations and 18 parameters such a calculation ... might take over a week of extremely tiresome work. The calculations were actually carried through on an International Business Machines Co tabulator using the Hollerith punched card system. ... 2 days to punch the cards and to check them and 4 hours of computations."

Obviously, computer power has increased dramatically since that time but the same is true for the size of the structural problems.

These computational difficulties (computational cost) are convoluted with methodological obstacles. For example, the optimization methods have limited convergence radius: the steepest descent method converges to the closest local minimum that may be far from the global minimum we are interested in. Even if the global minimum of the least-squares target is achieved, it may correspond to an incorrect structure since it does not take into account model incompleteness. Indeed, since structure factors depend on a whole set of atoms, missing atoms in the model at intermediate refinement steps makes the direct comparison of calculated and measured structure factors by a least-squares target inappropriate and minimization of the least-square target may make the structure worse (Lunin, Afonine and Urzhumtsev, 2002). Since diffraction data may be of limited quality (finite resolution, experimental errors in measured intensities) this in turn may have implications for the refined models: limited quality diffraction data will result in erroneous models. In fact, macromolecular diffraction data alone is almost always insufficient to obtain atomic models of acceptable quality. This highlights the need for the introduction of additional information in refinement. This information may be prior knowledge about the chemistry and physical properties of the molecule and can be used as restraints or constraints in refinement. For example, if the length between covalently bonded atoms is known, restraints will enable the refinement

to maintain this length approximately close to the known value, while constraints will enforce that the model value exactly match the known value.

AN HISTORICAL PERSPECTIVE: THE FIRST REFINEMENT PROGRAMS AND RESULTS

The development of refinement programs closely followed the progress of macromolecular structure solution. By the middle of 1960s the first macromolecular atomic structures were reported and computers became accessible to crystallographers. In 1971 R. Diamond reported the first common-use refinement program that employed a number of methodological advances. First, to reduce the number of independent parameters and to avoid distortion of covalent bonding geometry due to insufficient experimental data resolution, Diamond used torsion angles as the variable atomic model parameters. While requiring less variables, the parameterization in torsion angle space may limit refinement convergence as any changes in atomic coordinates need to be propagated along the chain. Second, to avoid the time-consuming calculation of structure factors from the model, Diamond suggested using Fourier maps calculated with the experimental amplitudes and the best available approximate phases as the target for fitting the atomic model parameters. Phases, being approximate, may be inaccurate enough to lead to an incorrectly refined model.

The availability of Diamond's refinement program designed specifically for macromolecules, and progress in macromolecular structure solution in general (Watenpugh *et al.*, 1973) prompted the active development of new methods and tools for crystallographic structure refinement. New refinement programs emerged during the latter half of the 1970s used amplitude based least-squares target functions combined with terms introducing some prior knowledge about molecular geometry (Steigemann, 1974; Konnert, 1976; Sussman *et al.*, 1977; Jack and Levitt, 1978; Konnert and Hendrickson, 1980). These terms, called geometry restraints, are typically quadratic functions that are penalties for deviations of the model geometry (such as bond lengths, bond angles) from 'ideal' (postulated as a library) values and calculated through interatomic distances. For example, a restraint on covalent angles was defined as a sum of restraints on distances between the three atoms defining that angle. At that time a quadratic form of the target was believed to be important for easy and fast calculation of the gradients. However, Jack and Levitt (1978) suggested that a general form of potential-energy function could be used.

Some of these programs (Sussman *et al.*, 1977) used the idea of *geometric constraints* (Scheringer, 1963) with some atomic groups considered rigid and therefore parameterized by their position and orientation. The use of constraints decreases the number of refinable parameters and may increase the radius of convergence of minimization methods. Overall, the introduction of constrained-restrained refinement obsoleted a previously used practice of unrestrained refinement followed by model idealization (for example see Freer *et al.*, 1975; Dodson *et al.*, 1976; Ten Eyck *et al.*, 1976). At that time each structure refinement was an event. Several discoveries in computing and crystallographic methods at the end of 1970s and beginning of the 1980s prepared the ground for the development of a new generation of refinement programs.

IMPORTANT ADVANCEMENTS IN COMPUTATIONAL MATHEMATICS AND CRYSTALLOGRAPHIC METHODS THAT HELPED IMPROVE CRYSTALLOGRAPHIC REFINEMENT

The first major advance was a result of the intuition of D. Sayre (1951). He proposed that while structure factors can be expressed by relatively simple functions of the atomic parameters, their numeric calculation through the Fourier transformation of the electron density could be made more efficient. If the electron density is calculated on a regular grid composed of N grid nodes, and we calculate M Fourier coefficients (usually M has the same order of magnitude as N), then the total number of operations is proportional to NM . When the Fourier transformation is applied to a periodic function calculated in a regular grid, many coefficients of this linear transformation are similar. This allows a dramatic reduction in the number of operations making it much closer to linear instead of quadratic, $N \ln(M)$ instead of NM . Since each atom contributes to the electron density only locally, the calculation of an electron density from an atomic model is proportional to the number of atoms K , where K is much less than N . Thus, introduction of an additional step in the process “atoms – electron density – structure factors” made the calculations not slower but much faster!

The practically useful algorithm for performing Fourier transforms efficiently is known as the Fast Fourier Transformation (FFT) had been suggested by Cooley and Tukey (1965) with other important prerequisite works dating back to the beginning of the 20th century (Runge and Koëning, 1924) and even a hundred years earlier than that. It is interesting to note that Cooley, at the time of the development of the FFT with Tukey, shared a laboratory with Sayre. The FFT

was introduced into crystallography by L. F. Ten Eyck (1973) who finalized the algorithm of the fast and accurate calculation of structure factors using an intermediate electron density step (Ten Eyck, 1977).

The rapid progress of computer hardware stimulated the development of other numerical methods including those of optimization. In particular, Hestenes and Stiefel (1952) and Lanczos (1952) proposed the method of conjugate directions that is frequently and erroneously referred to as the method of conjugate gradients. Practically equivalent in runtime to the steepest descent method (per iteration), the new method largely improved the convergence of the minimization process and became the method of choice for majority of refinement programs. At that point the most time-consuming step, the calculation of the gradients of the refinement target, was yet to be addressed. In 1978 Agarwal noted that for the particular least-squares crystallographic target the gradient can be calculated much faster. Similarly to Sayre’s idea of performing calculations through an intermediate electron density, Agarwal’s procedure used four intermediate density-like functions each calculated from structure factors using the FFT. Lunin (personal communication) and Lifchitz (in Agarwal, 1981) suggested how to reduce this to a single FFT. An inconvenience of Agarwal’s procedure was that it could not use restraints and required repetitive geometry idealizations between refinement cycles and it was limited to the particular least-squares target.

The problem of fast gradient calculation was not specific to crystallographic refinement. Baur and Strassen (1983) and Kim, Nesterov and Cherkassky, (1984) demonstrated a general approach for any function calculated with a computer. The simple idea that computation of a function value is a chain of four arithmetic operations made it possible to reach this important conclusion:

If a function value is computed for a time T , then it is possible to build an algorithm that calculates its exact gradient for a time less than $4T$ regardless the number of parameters. There is a simple constructive way to build such an algorithm.

The principal idea is that for steps involved in the calculation of the target function the calculation of the gradient involves the same steps but in a backward direction. This means that as soon as a fast algorithm for calculation of an arbitrary target function is available, the fast algorithm for calculation of its gradients is guaranteed to be available too.

This important result had a number of implications for the development of refinement programs and methods. First, this showed that the Agarwal's algorithm was a particular case of a general approach. Second, this indicated that there was no need for the refinement targets to be quadratic; the fast gradients can be calculated for any function and therefore the crystallographers could focus on the best choice of the targets from a structural rather than computational point of view. Third, this showed that the crystallographic target can include any type of restraint, and not be limited to quadratic functions of coordinates or distances. Overall, this principle allowed deconvolution of the three basic components of the optimization problem: the choice of the model parameters, choice of the target, and model optimization method.

FURTHER PROGRESS IN REFINEMENT METHODOLOGY

A next important question is whether it is possible to propose a general way of development refinement programs given varieties of models and refinement targets. The considerations above suggest that the determining step is the calculation of the target from initial independent parameters.

The key step of the refinement is generating structure factors from a crystal model (here we assume atomic model) and comparing them with the experimental data via evaluation of the target function. Using constraints means atomic parameters (coordinates and/or scattering parameters) are not independent and are obtained from some other parameters that are varied (refined) independently. An example is a rigid body refinement where groups of atoms are considered rigid (Scheringer, 1963). In this case the position and orientation of the rigid group are defined by 3 angular and 3 positional parameters. Knowing the atomic coordinates and the rigid-groups parameters one can recalculate the coordinates of all atoms for any position of this group. Another example is a torsion angle description (Diamond, 1971; Abagyan, Totrov and Kuznetsov, 1994; Rice and Brunger, 1994). Other independent parameters may be characteristics of individual librations and vibrations describing atomic displacement parameters through the TLS model (Schomaker and Trueblood, 1968). One more example is riding model for hydrogens (Sheldrick and Schneider, 1997) where coordinates and displacement parameter values are generated by purely geometric considerations knowing the parameters of their neighboring atoms. The module of a refinement program corresponding to such a step is specific for a given choice of constraints (in other

words, for the *type of the model*) and it converts these independent parameters into atomic parameters. Obviously, if no constraints used atomic parameters themselves are independent variables. If a program uses different constraints, several such moduli can be used simultaneously.

Once parameters of all atoms are known, the density map in the crystal is generated using spherical or multipolar (Hansen and Coppens, 1978) atoms or other kinds of scatterers such as simple geometric objects (for example, Kalinin, 1980). At this step some non-atomic components such as bulk solvent can be added. Depending on diffraction experiment, density map may be electron or neutron, for example.

The next step is also common for most refinement programs: density map is converted into a set of its Fourier coefficients (structure factors). For this any efficient Fourier transform algorithm, and not necessarily FFT (Cooley and Tukey, 1965), can be used. At this step an extra contribution can be added to the calculated values of structure factors; for example this may be the contribution from a fixed part of the model (Urzhumtsev, Lunin and Vernoslova, 1989) or that from bulk solvent (for example, Afonine *et al.*, 2013).

If a real space target is used, for example the target comparing a model map with a known cryoEM map point by point, one more step is needed to calculate a model map with a subset of structure factors obtained at the previous step, those in the sphere of a limited resolution. Since this requires more calculations, no one of the known programs does this in this strict way. One possibility to avoid this calculation is to convert the experimental map into its Fourier coefficients and use them for comparison with the model structure factors. Another possibility is to estimate a shape of individual atoms in the maps of the same resolution as the experimental one (Diamond, 1971; Lunin and Urzhumtsev, 1984; Chapman, 1995, and references therein; strictly speaking, one must take into account also the weighting scheme and the reflections missed) and consider the experimental map as a 'density distribution' for such kind of atomic shapes. Nevertheless, it is clear that these targets are not fully equivalent to the correct one and are only approximations to it.

As each step shown above we have different kinds of crystal description, *i.e.* its parameterization in terms of different parameters: *independent* (that correspond to certain constraints), *atomic* (or more generally – parameters of geometric models including

their scattering factors), electron (neutron, etc), *density distribution*, *structure factors*, *Fourier maps* (that may be different from the density generated from a model). It is straightforward to move from one kind of parameters to the next one. However the inverse step may be not trivial (determine atomic parameters from a density distribution; determine a full set of structure factors from a map of a limited resolution, etc), so that these transitions are naturally ordered.

Following the steps above that describe the *model*, the *target* to be minimized is traditionally expressed as a weighted sum of various targets where each one is naturally expressed through the corresponding kind of parameters. The most of the modern refinement programs use the diffraction targets (least-squares, maximum-likelihood, phase target) and the targets expressed through the atomic parameters - geometric restraints and restraints on displacement parameters. A few programs use real-space targets.

The overall calculation algorithm is a chain of transitions between different kinds of crystal descriptions; each transition depends neither on the previous steps nor on further steps. Obviously, a transition may pass through its internal intermediate steps. For each kind of parameters various targets can be introduced that are fully independent of the parameters of other kinds. For example, one can envision targets (restraints) on the independent parameters in case of constraints (Urzhumtsev, Lunin and Vernoslova, 1989): for example an interdiction of rigid groups to move far from the original position, etc. If a program is constructed from such blocks (Lunin and Urzhumtsev, 1983, 1985; Tronrud, Ten Eyck and Matthews, 1987; Urzhumtsev, Lunin and Vernoslova, 1989), many of them are common for different purposes and can be reused in different contexts.

The fact that the global target to be optimized is a sum of the composited targets allows an independent calculation of their gradients with respect to the independent parameters. The algorithms to calculate the gradient for each of them with respect to their own variables are obtained by inverting each transition one by one. Then using the chain rule these gradients are recalculated to the gradients with respect to the independent parameters (Lunin and Urzhumtsev, 1985). The sum of these individual gradients gives the total gradient, that along with the target function value are the inputs to an *optimizer*. The choice of the optimizer is independent on the choice of the model and the target.

There are very many advances in macromolecular refinement due to computational progress and methodological understanding; we give only a few examples. Disconnection of a choice of the model, targets and optimization procedure for crystallographic refinement allowed Brünger, Kuriyan and Karplus (1987) to introduce a powerful molecular dynamics based approach to minimize the “energy” of the macromolecule as a sum of geometric and diffraction targets. This method greatly increased the radius of convergence of refinement, avoiding a great deal of extremely time consuming manual model building for much of a structure. Other minimization methods have been also discussed, for example by Tronrud (1992).

This also allowed for the straightforward introduction of a number of new diffraction targets such as a maximum-likelihood (Pannu and Read, 1996; Brucogne and Irwin, 1996; Murshudov, Vagin and Dodson, 1997; Adams *et al.*, 1997; Lunin, Afonine and Urzhumtsev 2002) or a probability-based phase target (Lunin and Urzhumtsev, 1985; Pannu *et al.*, 1998). The ML targets are extremely important since they use a better error model and take into account model incompleteness.

Finally, new parameters could be introduced without the need to reformulate either the target calculation, or the minimization procedure. In particular, this concerns non-atomic parameters such as new bulk-solvent models (for example, Jiang and Brunger, 1994; Fenn, Schnieders and Brünger, 2010; Afonine *et al.*, 2013).

CURRENT STATE OF THE ART

Table 1 traces the history of refinement program development from today back to the beginning of the 1970s. Today’s software for carrying out crystallographic structure refinement of macromolecules is highly sophisticated. Many of the decision-making steps that used to be the responsibility of the researcher are now performed automatically by the refinement program. Programs allow a broad range of model parameterizations depending on data quality. The choice of refinement target is almost exclusively maximum-likelihood, as this target assumes a more appropriate error model and statistically accounts for model incompleteness. Optimization tools are not limited to gradient-driven methods, and other techniques such as local grid (systematic) searches and simulated annealing are in use.

As can be seen from the preceding sections, a refinement program is typically a large suite composed of many modules each one designed to perform a specific task. While older programs are mostly written using FORTRAN as a programming language (one of principal developers of which was D. Sayre), most recently developed tools such as *Phenix* (Adams *et al.*, 2010) use modern concepts of software development using object oriented languages such as C++ and Python. This allows a high degree of extensibility, easier maintenance and promotes collaboration between scientific groups.

In spite of the effort required to develop an efficient and robust refinement program, nowadays the crystallographic community has access to a large number. Most popular are SHELXL (Sheldrick and Schneider, 1997), CNS (Brünger *et al.*, 1998), BUSTER-TNT (Blanc *et al.*, 2004; Tronrud, Ten Eyck and Matthews, 1987), MAIN (Turk, 1992), REFMAC (Murshudov, Vagin and Dodson, 1997), and phenix.refine (Afonine *et al.*, 2012). These programs share the same basic principles but differ in technical implementation making different choices for different circumstances. For example, SHELXL and REFMAC make use of minimizers based on the second derivatives of the target and are typically better suited for refinement with high-resolution data. BUSTER uses advanced statistical approaches both in the description of the model and in the target function. REFMAC and phenix.refine have advanced tools for refinement at low resolution. phenix.refine also permits efficient refinement at ultra-high resolution (Afonine *et al.*, 2007), highly automated rigid-body refinement (Afonine *et al.*, 2009), employs advanced scaling and bulk-solvent modeling protocols (Afonine *et al.*, 2013) and also permits refinement against two diffraction data sets, X-ray and neutron, simultaneously. Using such complementary information about electrons and nuclei results in models that are better than those obtained by refinement against individual each data set (Adams *et al.*, 2009).

A refinement run for a moderately-sized macromolecular model nowadays takes from a few minutes (for a small size protein) to several hours (for structures as large as a ribosome). This acceleration is obviously due to both the availability of new powerful computers and efficient algorithms implemented in refinement programs.

SOME CURRENT CHALLENGES AND FUTURE GOALS

Steps from phasing to final structure report (Figure 1) are now typically highly automated (e.g. Adams *et al.*, 2010). However, structure refinement remains the

least automated step. This stems from the fact that data quality can be very diverse: from high to low resolution, it may be incomplete or affected by various crystal growth disorders such as twinning, quality limiting factors during data collection or limitations of data processing tools. In turn this generates an array of possible model parameterizations that may need to be employed in order to adequately describe these data. All together, diversity of data and model parameterizations are convoluted with the fact that atomic models are an approximate representation of the true crystal content. Depending on how far the current model is from the true structure different optimization methods and tools are needed in order to bring the current model as close as possible to the true one. While some decision-making steps are automated and performed exclusively by software there is still great opportunity for a researcher to intervene in the process and make decisions that may, in the end, determine whether structure solution is successful. Further efforts put into the automation of refinement workflow are therefore critical for streamlining this step and ensuring that the resulting refined models are of high quality.

Improvements in crystallization and data collection techniques have increased the number of low-resolution datasets being collected. Typically this data corresponds to crystals of large molecules that may have substantial mobility. Low-resolution maps combined with the size of the problem (large models result in a large amount of data) make model building and refinement extremely challenging. First, low resolution maps do not readily permit the accurate building of models, so initial models often possess poor geometry and may have gross stereochemical imperfections. Given unfavorable data-to-parameter ratios subsequent refinement often may not yield significant improvement. The lack of experimental data at these resolutions means that successful refinement is highly dependent on prior knowledge – the restraints. While the traditional stereochemistry restraints used in refinement are sufficient at medium to high resolution, they do not provide enough additional information at low resolution (Headd *et al.*, 2012, 2014). Therefore, more prior information is needed to make low-resolution refinement feasible. Such information might include secondary structure organization (helices and sheets in proteins or specific arrangements of nucleobases in DNA/RNA) or extra symmetry arising from specific crystal packing (non-crystallographic symmetry, NCS). Extracting and using this information correctly can be challenging.

As mentioned above, the geometry restraints that are used in refinement programs can be simple and relatively naïve, mostly designed to preserve basic model geometry and prevent a model from deterioration in the case of insufficient quality data, which is almost always the case for macromolecules. As a result these restraints tend to generate unrealistic models if data resolution is limited. An alternative to extending these restraints with additional information is to design better potential functions, which may be not as sophisticated as those used in the molecular simulation field but more tailored to the context of structure refinement. Another approach is the use of QM/MM (quantum mechanics/molecular mechanic) methods to generate accurate structures of small molecules in macromolecular structures or even whole macromolecular structures (Canfield *et al.*, 2006; Reimers, 2011; Falköf, Collyer and Reimers, 2012).

Hydrogen is a weak X-ray scatterer and therefore it is barely observable in maps derived from X-ray diffraction experiments. Historically this has prompted macromolecular crystallographers to generate models without H atoms. Only in ultra-high resolution X-ray diffraction experiments is it possible to visualize some, but typically not all, hydrogen atoms. These ultra-high resolution structures constitute only 0.002% of all structures in PDB. At the same time H atoms constitute nearly half of the atoms in a protein structure, they mediate most of interatomic contacts, often play key roles in catalytic activities of enzymes, and participate in ligand binding. While most hydrogen positions can be inferred from the local geometry, there are still 10-15% of H atoms that have rotational degrees of freedom and thus cannot be predicted based on local stereochemistry alone. Neutron diffraction is therefore a technique of some importance (for review, see for example Afonine *et al.*, 2010). Hydrogen, and deuterium, atoms diffract neutrons almost as well as other typical protein atoms (C, N and O). Therefore, in principle, a neutron diffraction experiment can yield a complete model that contains hydrogen and non-hydrogen atoms. However, this is not without difficulty. First, the neutron scattering length of a H atom is negative: this results in H atoms having negative density in Fourier maps. The consequence of this is cancellation effects where the local density map arising from H atoms cancels out the positive density from heavier atoms in the vicinity. To minimize this problem it is necessary to deuterate samples, where molecules have their H atoms fully or partially replaced with deuterium (D atoms scatter neutrons as strongly as C or N). Another challenge is that until recently neutron

data collection required very large crystals, which are usually challenging to obtain for macromolecules. Recent advances in instrumentation are poised to reduce this bottleneck. The combination of X-ray and neutron data from isomorphous crystals of a macromolecule is a rich source of information from which it is possible to derive a complete atomic model. Although, appropriate use of both data sets simultaneously in refinement is another challenging task

While it is rather rare that crystals of macromolecules diffract to ultra-high resolution, better than approximately 0.9Å, there are ~500 structures in PDB solved at this resolution. The typical Gaussian model parameterization used at lower resolution is insufficient in these extreme cases. Instead, more complex models are needed such as multipolar representations for electron density distributions. However, this approach approximately triples the number of parameters per atom. This poses some fundamental problems. One is that the FFT based method of structure factor and gradient calculation cannot be readily used for a non-gaussian (multipolar) parameterization while a progress has been reported (Schnieders *et al.*, 2009). This makes the calculation times for macromolecular structures prohibitively slow even on today's fastest computers. The other problem is a numerical one, and is due to the fact that multipolar parameters are very diverse in scale and some of them are highly correlated with each other and other atomic parameters that are non-multipole specific, such as occupancy and displacement parameters. This requires special care when developing methods for optimization of these parameters.

Recent improvements in the field of cryo-electron microscopy (cryo-EM) have made it possible to generate structural information at resolutions approaching low-resolution X-ray crystallography (3.5 Å and lower). The result of the cryo-EM experiment is a map that can be used to build and refine an atomic model. Most refinement tools available today were designed for X-ray or neutron crystallography, and therefore designed to perform complete model refinement against diffraction data (amplitudes or intensities of measured structure factors) and not maps. Also, typically these are very large structures. Since the resolution is low the maps are challenging to interpret and provide limited information for model refinement. Therefore, new methods need to be developed, such as real space refinement approaches that can efficiently perform rapid refinement of large macromolecules to generate models of high chemical quality.

Structure validation is a process that aims to perform thorough assessment of model quality. Traditionally structure validation was performed at the very end of structure determination. However, it is now accepted that this is suboptimal because errors created and unnoticed at the beginning of structure determination may propagate and become very difficult to detect and address later on. Therefore active structure validation should be performed constantly through the entire process of structure determination and not only at the very end. This changes the paradigm of the structure determination workflow and thus requires significant changes in the corresponding software.

REFERENCES

- Abagyan, R. A., Totrov, M. M. and Kuznetsov, D. A. (1994). ICM – a new method for protein modeling and design – Applications to docking and structure prediction from the distorted native conformation. *Journal of Computational Chemistry*, 15, pp. 488-506. <http://dx.doi.org/10.1002/jcc.540150503>
- Adams, P. D., Mustyakimov, M., Afonine, P. V. and Langan, P. (2009). Generalized X-ray and neutron crystallographic analysis: more accurate and complete structures for biological macromolecules. *Acta Crystallographica*, D65, pp. 567-573. <http://dx.doi.org/10.1107/S0907444909011548>
- Adams, P. D., Pannu, N. S., Read, R. J. and Brünger, A. T. (1997). Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proceedings of National Academy of Science*, 94, pp. 5018-5023. <http://dx.doi.org/10.1073/pnas.94.10.5018>
- Adams, P. D., Afonine, P. V., Bunkóczy, 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. and Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica*, D66, pp. 213-221. <http://dx.doi.org/10.1107/S0907444909052925>
- Afonine, P., Lunin, V. Y. and Urzhumtsev, A. (2003). MLMF: least-squares approximation of likelihood-based refinement criteria. *Journal of Applied Crystallography*, 36, pp. 158-159. <http://dx.doi.org/10.1107/S0021889802021738>
- Afonine, P. V., Grosse-Kunstleve, R. W., Adams, P. D., Lunin, V. Y. and Urzhumtsev, A. (2007). On macromolecular refinement at subatomic resolution with interatomic scatterers. *Acta Crystallographica*, D63, pp. 1194-1197. <http://dx.doi.org/10.1107/S0907444907046148>
- Afonine, P. V., Grosse-Kunstleve, R. W., Urzhumtsev, A. and Adams, P. D. (2009). Automatic multiple-zone rigid-body refinement with a large convergence radius. *Journal of Applied Crystallography*, 42, pp. 607-615. <http://dx.doi.org/10.1107/S0021889809023528>
- Afonine, P. V., Mustyakimov, M., Grosse-Kunstleve, R. W., Moriarty, N. W., Langan, P. and Adams, P. D. (2010). Joint X-ray and neutron refinement with *phenix.refine*. *Acta Crystallographica*, D66, pp. 1153-1163. <http://dx.doi.org/10.1107/S0907444910026582>
- Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W.; Mustyakimov, M., Tenwilliger, T. C. Urzhumtsev, A. and Zwart, P. H. (2012). Towards automated crystallographic structure refinement with *phenix.refine*. *Acta Crystallographica*, D68, pp. 352-367. <http://dx.doi.org/10.1107/S0907444912001308>
- Afonine, P. V., Grosse-Kunstleve, R. W., Adams, P. D. and Urzhumtsev, A. (2013). Bulk-solvent and overall scaling revisited: faster calculations, improved results. *Acta Crystallographica*, D69, pp. 625-634. <http://dx.doi.org/10.1107/S0907444913000462>
- Agarwal, R. C. (1978). A new least-squares refinement technique based on the fast Fourier transform algorithm. *Acta Crystallographica*, A34, pp. 791-809. <http://dx.doi.org/10.1107/S0567739478001618>
- Agarwal, R. C. (1981). New results on fast Fourier least-squares refinement technique. In Machin, P. A., Campbell, J. W. and Elder, M. (comps.). *Refinement of Protein Structures. Proceedings of the Daresbury Study Weekend, 15-16 November 1980*. Daresbury, Warrington: Science and Engineering Research Council, Daresbury Laboratory, pp. 24-28.
- Ambartsumian, V. A. (1929). On the Relationship between the Solution and the Resolvent of the Integral Equation of the Radiative Balance. *Zeitschrift für Physik*, 52, pp. 263-267.
- Baur, W. and Strassen, V. (1983). The complexity of partial derivatives. *Theoretical Computer Science*, 22, pp. 317-330. [http://dx.doi.org/10.1016/0304-3975\(83\)90110-X](http://dx.doi.org/10.1016/0304-3975(83)90110-X)
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28, pp. 235-242. <http://dx.doi.org/10.1093/nar/28.1.235>
- Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *Journal of Molecular Biology*, 112, pp. 535-542. [http://dx.doi.org/10.1016/S0022-2836\(77\)80200-3](http://dx.doi.org/10.1016/S0022-2836(77)80200-3)

ACKNOWLEDGEMENTS

This work was supported by the NIH (Project 1P01 GM063210) and the Phenix Industrial Consortium. This work was supported in part by the US Department of Energy under Contract No. DE-AC02-05CH11231. AU thanks the French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INSB-05-01 and Instruct, part of the European Strategy Forum on Research Infrastructures (ESFRI).

- Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S. M. and Bricogne, G. (2004). Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallographica*, D60, pp. 2210-2221. <http://dx.doi.org/10.1107/S0907444904016427>
- Booth, A. D. (1947). Application of the method of steepest descents to X-ray structure analysis. *Nature*, 160, p. 196. <http://dx.doi.org/10.1038/160196a0>
- Bricogne, G. and Irwin, J. (1996). Maximum likelihood structure refinement: theory and implementation within BUSTER + TNT. In Dodson, E. J., Moore, M., Ralph, A. and Bailey, S. (eds.). *Macromolecular Refinement: Proceedings of the CCP4 Study Weekend*. Daresbury, Warrington: Science and Engineering Research Council, Daresbury Laboratory, pp. 85-92.
- Brünger, A. T., Kuriyan, J. and Karplus, M. (1987). Crystallographic *R* factor refinement by molecular dynamics. *Science*, 235, pp. 458-460. <http://dx.doi.org/10.1126/science.235.4787.458>
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilques, M., Pannu, N. S., Read, R. J., Rice, L. M., Somonson, T. and Warren, G. L. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallographica*, D54, pp. 905-921. <http://dx.doi.org/10.1107/S0907444998003254>
- Canfield, P., Dahlbom, M. G., Reimers, J. R. and Hush, N. S. (2006). Density-functional geometry optimization of the 150,000-atom photosystem-I trimer. *Journal of Chemical Physics*, 124, pp. 024301-024315. <http://dx.doi.org/10.1063/1.2148956>
- Chapman, M. (1995). Restrained real-space macromolecular atomic refinement using a new resolution-dependent electron-density function. *Acta Crystallographica*, A51, pp. 69-80 <http://dx.doi.org/10.1107/S0108767394007130>
- Cochran, W. (1948). The Fourier method for crystal-structure analysis. *Acta Crystallographica*, 1, pp. 138-142. <http://dx.doi.org/10.1107/S0365110X48000375>
- Cooley, J. W. and Tukey, J. W. (1965). An algorithm for machine calculation of complex Fourier series. *Mathematics of Computation*, 19, pp. 297-301. <http://dx.doi.org/10.1090/S0025-5718-1965-0178586-1>
- Diamond, R. (1971). A real-space refinement procedure for proteins. *Acta Crystallographica*, A27, pp. 436-452. <http://dx.doi.org/10.1107/S0567739471000986>
- Dodson, E. J., Isaacs, N. W. and Rollett, J. S. (1976). A method for fitting satisfactory models to sets of atomic positions in protein structure refinements. *Acta Crystallographica*, A32, pp. 311-315. <http://dx.doi.org/10.1107/S0567739476000685>
- Diessen, H., Haneef, M. I. J., Harris, G. W., Howlin, B., Khan, G. and Moss, D. S. (1989). TLSANL - TLS parameter-analysis program for segmented anisotropic refinement of macromolecular structures. *Journal of Applied Crystallography*, 22, pp. 510-516. <http://dx.doi.org/10.1107/S0021889889004097>
- Ewald, P. P. (1913). About the theory of the interference of X-rays in crystals (Zur Theorie der interferenzen der Röntgenstrahlen in kristallen). *Physikalische Zeitschrift*, 14, pp. 465-472.
- Falköf, O., Collyer, C. A. and Reimers, J. R. (2012). Toward ab initio refinement of protein X-ray crystal structures: interpreting and correlating structural fluctuations. *Theoretical Chemistry Accounts*, 131, 1076. <http://dx.doi.org/10.1007/s00214-011-1076-8>
- Fenn, T. D., Schnieders, M. J. and Brünger, A. T. (2010). A smooth and differentiable bulk-solvent model for macromolecular diffraction. *Acta Crystallographica*, D66, pp. 1024-1031. <http://dx.doi.org/10.1107/S0907444910031045>
- Finzel, B. C. (1987). Incorporation of fast Fourier-transforms to speed restrained least-squares refinement of protein structures. *Journal of Applied Crystallography*, 20, pp. 53-55. <http://dx.doi.org/10.1107/S0021889887087144>
- Freer, S. T., Alden, R. A., Carter, W. C. Jr. and Kraut, J. (1975). Crystallographic structure refinement of chromatin high potential iron protein at two Angstroms resolution. *Journal of Biological Chemistry*, 250, pp. 46-54.
- Guillot, B., Viry, L., Guillot, R., Lecomte, C. and Jelsch, C. (2001). Refinement of proteins at subatomic resolution with MOPRO. *Journal of Applied Crystallography*, 34, pp. 214-223. <http://dx.doi.org/10.1107/S0021889801001753>
- Haneef, I., Moss, D. S., Stanford, M. J. and Borkakoti, N. (1985). Restrained structure-factor least-squares refinement of protein structures using a vector processing computer. *Acta Crystallographica*, A41, pp. 426-433. <http://dx.doi.org/10.1107/S0108767385000915>
- Hansen, N. K. and Coppens, P. (1978). Testing aspherical atom refinements on small-molecule data sets. *Acta Crystallographica*, A34, pp. 909-921. <http://dx.doi.org/10.1107/S0567739478001886>
- Headd, J. J., Echols, N., Afonine, P. V., Grosse-Kunstleve, R. W., Chen, V. B., Moriarty, N. W., Richardson, D. C., Richardson, J. S. and Adams, P. D. (2012). Use of knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution. *Acta Crystallographica*, D68, pp. 381-390. <http://dx.doi.org/10.1107/S0907444911047834>
- Headd, J. J., Echols, N., Afonine, P. V., Moriarty, N. W., Gildea, R. J. and Adams, P. D. (2014). Flexible torsion-angle non-crystallographic symmetry restraints for improved macromolecular structure refinement. *Acta Crystallographica*, D70, pp. 1346-1356. <http://dx.doi.org/10.1107/S1399004714003277>
- Hestenes, M. R. and Stiefel, E. (1952). Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49, pp. 409-436. <http://dx.doi.org/10.6028/jres.049.044>
- Hendrickson, W. A. and Konnert, J. H. (1980). In Srinivasan, R., Subramanian, E. and Yathindra, N. (eds.). *Biomolecular Structure, Conformation, Function, and Evolution* (vol. 1). New York: Pergamon, pp. 43-57.
- Hughes, E. W. (1941). The crystal structure of melanine. *Journal of the American Chemical Society*, 63, pp. 1737-1752. <http://dx.doi.org/10.1021/ja01851a069>
- Jack, A. and Levitt, M. (1978). Refinement of large structures by simultaneous minimization of energy and *R* factor. *Acta Crystallographica*, A34, pp. 931-935. <http://dx.doi.org/10.1107/S0567739478001904>
- Jelsch, C., Guillot, B., Lagoutte, A. and Lecomte, C. (2005). Advances in protein and small-molecule charge-density refinement methods using *MoPro*. *Journal of Applied Crystallography*, 38, pp. 38-54. <http://dx.doi.org/10.1107/S0021889804025518>

- Jiang, J.-S. and Brünger, A. T. (1994). Protein hydration observed by X-ray diffraction. Solvation properties of penicillopepsin and neuroaminidase crystal structures. *Journal of Molecular Biology*, 243, pp. 100-115. <http://dx.doi.org/10.1006/jmbi.1994.1633>
- Kalinin, D. I. (1980). Use of a cylindrical model of a protein to determine the spatial structure of the rhombic modification of leghaemoglobin. *Soviet Physics. Crystallography*, 25, pp. 307-313.
- Kim, K. M., Nesterov, Yu. E. and Cherkassky, B. V. (1984). Ocenka trudoemkosti vyčislenija gradienta. *Doklady Akademii Nauk SSSR*, 275, pp. 1306-1309.
- Konnert, J. H. (1976). A restrained-parameter structure-factor least-squares refinement procedure for large asymmetric units. *Acta Crystallographica*, A32, pp. 614-617. <http://dx.doi.org/10.1107/S0567739476001289>
- Konnert, J. H. and Hendrickson, W. A. (1980). A restrained-parameter thermal-factor refinement procedure. *Acta Crystallographica*, A36, pp. 344-350. <http://dx.doi.org/10.1107/S0567739480000794>
- Lanczos, C. (1952). Solution of systems of linear equations by minimized iterations. *Journal of Research of the National Bureau of Standards*, 49, pp. 33-53. <http://dx.doi.org/10.6028/jres.049.006>
- Lunin, V. Y. and Urzhumtsev, A. (1983). *Program construction for refinement of macromolecular atomic structures on the base of Fast Fourier transformation and Fast differentiation algorithms*. Preprint, Pushchino, ONTI NCBI.
- Lunin, V. Y. and Urzhumtsev, A. (1984). Improvement of protein phases by coarse model modification. *Acta Crystallographica*, A40, pp. 269-277. <http://dx.doi.org/10.1107/S0108767384000544>
- Lunin, V. Y. and Urzhumtsev, A. (1985). Program construction for macromolecule atomic model refinement based on the fast Fourier transform and fast differentiation algorithms. *Acta Crystallographica*, A41, pp. 327-333. <http://dx.doi.org/10.1107/S010876738500071X>
- Lunin, V. Y., Afonine, P. V. and Urzhumtsev, A. (2002). Likelihood-based refinement. I. Irremovable model errors. *Acta Crystallographica*, A58, pp. 270-282. <http://dx.doi.org/10.1107/S0108767302001046>
- Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica*, D53, pp. 240-255. <http://dx.doi.org/10.1107/S0907444996012255>
- Pannu, N. S. and Read, R. J. (1996). Improved Structure Refinement Through Maximum Likelihood. *Acta Crystallographica*, A52, pp. 659-668. <http://dx.doi.org/10.1107/S0108767396004370>
- Pannu, N. S., Murshudov, G. N., Dodson, E. J. and Read, R. J. (1998). Incorporation of Prior Phase Information Strengthens Maximum-Likelihood Structure Refinement. *Acta Crystallographica*, D54, pp. 1285-1294. <http://dx.doi.org/10.1107/S09074449980004119>
- Reimers, J. R. (2011). *Computational Methods for Large Systems: Electronic Structure Approaches for Biotechnology and Nanotechnology*. Hoboken, New Jersey: Wiley. <http://dx.doi.org/10.1002/9780470930779>
- Rice, L. M. and Brunger, A. T. (1994). Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins: Structure, Function and Genetics*, 19, pp. 277-290. <http://dx.doi.org/10.1002/prot.340190403>
- Runge, C. and König, D. (1924). *Die Grundlehren der mathematischen Wissenschaften* (vol. II). Berlin: Springer.
- Sayre, D. (1951). The calculation of structure factors by Fourier summation. *Acta Crystallographica*, 4, pp. 362-367. <http://dx.doi.org/10.1107/S0365110X51001124>
- Scheringer, C. (1963). Least-squares refinement with the minimum number of parameters for structures containing rigid-body groups of atoms. *Acta Crystallographica*, 16, pp. 546-550.
- Schnieders, M. J., Fenn, T. D., Pande, V. S. and Brunger, A. T. (2009). Polarizable atomic multipole X-ray refinement: application to peptide crystals. *Acta Crystallographica*, D65, pp. 952-965.
- Schomaker, V. and Trueblood, K. N. (1968). On rigid-body motion of molecules in crystals. *Acta Crystallographica*, B24, pp. 63-76. <http://dx.doi.org/10.1107/S0567740868001718>
- Sheldrick, G. M. and Schneider, T. R. (1997). SHELXL: High-resolution refinement. *Methods in Enzymology*, 277B, pp. 319-343. [http://dx.doi.org/10.1016/S0076-6879\(97\)77018-6](http://dx.doi.org/10.1016/S0076-6879(97)77018-6)
- Steigemann, W. (1974). Ph.D. thesis, Technische Universität, München.
- Sussman, J. L., Holbrook, S. R., Church, G. M. and Kim, S.-H. (1977). Structure-factor least-squares refinement procedure for macromolecular structures using constrained and restrained parameters. *Acta Crystallographica*, A33, pp. 800-804. <http://dx.doi.org/10.1107/S0567739477001958>
- Ten Eyck, L. F. (1973). Crystallographic fast Fourier transforms. *Acta Crystallographica*, A29, pp. 183-191. <http://dx.doi.org/10.1107/S0567739473000458>
- Ten Eyck, L. F. (1977). Efficient structure-factor calculation for large molecules by the fast Fourier transform. *Acta Crystallographica*, A33, pp. 486-492. <http://dx.doi.org/10.1107/S0567739477001211>
- Tronrud, D. E., Ten Eyck, L. F. and Matthews, B. W. (1987). An efficient general-purpose least-squares refinement program for macromolecular structures. *Acta Crystallographica*, A43, pp. 489-501. <http://dx.doi.org/10.1107/S0108767387099124>
- Tronrud, D. E. (1992). Conjugate-direction minimization: an improved method of the refinement of macromolecules. *Acta Crystallographica*, A48, pp. 912-916. <http://dx.doi.org/10.1107/S0108767392005415>
- Turk, D. (1992). PhD thesis. Technische Universität München, Germany.
- Urzhumtsev, A. G., Lunin, V. Yu. and Vernoslova, E. A. (1989). FROG - high-speed restraint-constraint refinement program for macromolecular structure. *Journal of Applied Crystallography*, 22, pp. 500-506. <http://dx.doi.org/10.1107/S0021889889004905>
- Watenpaugh, K. D., Sieker, L. C., Herriott, J. R. and Jensen, L. H. (1973). Refinement of model of a protein - rubredoxin at 1.5 Å resolution. *Acta Crystallographica*, B29, pp. 943-956. <http://dx.doi.org/10.1107/S0567740873003675>
- Westhof, E., Dumas, Ph. and Moras, D. (1988). Restraint refinement of two crystalline forms of yeast aspartic acid and phenylalanine transfer RNA crystals. *Acta Crystallographica*, A44, pp. 112-123. <http://dx.doi.org/10.1107/S010876738700446X>