

Version 1.2 of the Crystallography and NMR system

Axel T Brunger

The Howard Hughes Medical Institute and Departments of Molecular and Cellular Physiology, Neurology and Neurological Sciences, Structural Biology, and Photon Science, Stanford University, J.H. Clark Center E300C, 318 Campus Dr., California 94305, USA. Correspondence should be addressed to A.T.B. (brunger@stanford.edu).

Published online 25 October 2007; doi:10.1038/nprot.2007.406

Version 1.2 of the software system, termed Crystallography and NMR system (CNS), for crystallographic and NMR structure determination has been released. Since its first release, the goals of CNS have been (i) to create a flexible computational framework for exploration of new approaches to structure determination, (ii) to provide tools for structure solution of difficult or large structures, (iii) to develop models for analyzing structural and dynamical properties of macromolecules and (iv) to integrate all sources of information into all stages of the structure determination process. Version 1.2 includes an improved model for the treatment of disordered solvent for crystallographic refinement that employs a combined grid search and least-squares optimization of the bulk solvent model parameters. The method is more robust than previous implementations, especially at lower resolution, generally resulting in lower *R* values. Other advances include the ability to apply thermal factor sharpening to electron density maps. Consistent with the modular design of CNS, these additions and changes were implemented in the high-level computing language of CNS.

INTRODUCTION

The program CNS was the first advanced software system in structural biology that made use of a modular, multilevel approach to computing, utilizing a high-level symbolic structure-determination language¹. Since its first dissemination in 1998, CNS has become one of the most widely used systems to determine structures based on X-ray diffraction or nuclear magnetic resonance data or both. Its key advantages are the flexibility and generality of the system, allowing ‘computational experimentation’ with new algorithms or applications to different types of experimental data without tedious software development.

In CNS, many algorithms were moved from the source code into this symbolic language. The high-level CNS computing language allows definition of symbolic target functions, data structures, procedures and modules. The compiled CNS program, written in Fortran77, acts as an interpreter for the high-level CNS language and includes hard-wired functions for the efficient processing of computing-intensive tasks. Methods and algorithms are therefore more clearly defined and easier to adapt to new and challenging problems. The result is a multilevel system that provides maximum flexibility to the user. The CNS language provides a common framework for nearly all the computational procedures required for structure determination. A comprehensive set of crystallographic procedures for phasing, density modification and refinement has been implemented in this language. The CNS language permits the design and execution of nearly any numerical task in structure determination, using a minimal set of ‘hard-wired’ functions and routines. Task files consist of CNS language statements and module invocations. The task files, which can also be accessed, modified or viewed through an HTML graphical interface, are available to carry out these procedures.

Most operations within a structure-determination algorithm are defined through modules and task files. This allows the development of new algorithms and for existing algorithms to be precisely defined and easily modified without the need for source code modifications. This hierarchical structure of CNS allows extensive testing at each level. For example, once the source code and CNS basic commands have been tested, testing of the modules and task

files can be performed. A test suite consisting of hundreds of test cases is available to detect and correct programming errors. This testing scheme makes CNS highly reliable. It also makes it easier to modify the program and add new features.

The source codes of CNS and its predecessor X-PLOR (see ref. 2) are available, allowing users to easily interface their algorithms with these programs. Examples of such interfaces include tools developed for NMR structure determination³, the protein–protein docking method HADDOCK based on biochemical and/or biophysical information⁴, the ARIA method for automated NOE assignment and NMR structure calculation⁵, incorporation of electrostatics and continuum dielectric methods in refinement⁶, a database of interatomic distance probabilities⁷, time-averaged molecular dynamics refinement against X-ray diffraction data⁸ and molecular dynamics in refinement against fiber diffraction data⁹.

The major new features of version 1.2 of CNS include an improved bulk solvent model for crystallographic refinement and the ability to manipulate electron density maps by thermal factor sharpening. These and most other changes in the system were restricted to the high-level task and module files (**Supplementary Note** online).

Bulk solvent modeling in X-ray crystallography

The correct modeling of the barrier between the bulk solvent in the crystal lattice and the protein itself is an important part of macromolecular structure refinement. The structure factor F_{calc} of a macromolecular crystal structure is expressed in CNS as

$$F_{\text{calc}} = \exp[-2\pi^2 \vec{h}^t U^* \vec{h}] \{F_{\text{macro}} + F_{\text{bound}} + F_{\text{bulk}}\} \quad (1)$$

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) and the symmetric second rank tensor U describes overall mean-square displacements of the crystal lattice (dimensionless



anisotropic mean-square displacements (ADPs)). 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} (see ref. 10). 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} , a mask is created to distinguish between macromolecular and solvent regions. This problem is closely related to the computation of accessible and molecular surface areas¹¹. A three-dimensional map *mask* is defined on a grid that covers an asymmetric unit of the crystal. The values of *mask* are restricted to 0 and 1. The grid size is chosen to be small enough to avoid Fourier series truncation errors. By trial and error, we set the grid size to one-third of the high-resolution limit with the additional condition that the grid must be in the range between 0.57 and 0.9 Å.

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} . The van der Waals radius is defined as half the distance at which the Lennard-Jones potential energy function reaches its minimum.

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. The resulting boundary between solvent and macromolecule is a combination of contact and reentrant surface areas¹². 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 widely used ‘flat’ solvent model assumes that solvent regions outside the molecular surface show relatively little variation in density as compared to the macromolecule¹³. The structure factor of the solvent F_{bulk} is then simply computed by Fourier transformation of *mask*. 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} :

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

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. Thus, for a well-behaved solvent model, k_{sol} should be close to $0.3 \text{ e} \text{ \AA}^{-3}$ and B_{sol} reasonably close (within a factor 2) to the average thermal factor of the macromolecular model.

The optimum solvent model is obtained by minimizing the expression

$$(F_{\text{obs}} - F_{\text{calc}}[k_{\text{sol}}, B_{\text{sol}}, U])^2 \quad (3)$$

where F_{obs} is the observed structure factor. Due to the implementation of fundamental features in the CNS source code, the optimization is broken up into optimization of isotropic and anisotropic parameters (Table 1). The bulk solvent procedure in the earlier versions of CNS often resulted in numerical instabilities for the refinement of the solvent parameters k_{sol} and B_{sol} for structures

TABLE 1 | Schematic of the bulk solvent optimization in CNS 1.2.

Loop over $R_{\text{shrink}}=R_{\text{probe}}$ (at lower than 3 Å resolution, otherwise fixed at 1 Å)
Calculate <i>mask</i>
Loop over k_{sol}
Least squares optimization of $(F_{\text{obs}} - F_{\text{calc}}[k_{\text{sol}}, B_{\text{sol}}, U])^2$ with U enforced to be isotropic
Expansion to $P1$
Least squares optimization of $(F_{\text{obs}} - F_{\text{calc}}[k_{\text{sol}}, B_{\text{sol}}(\text{optimum}), U])^2$ with anisotropic U
Reduction to space group of crystal
Calculate R
End loop
End loop
Take solvent model with optimum R

determined at low to moderate resolution (i.e., lower than 3 Å resolution). The procedure was therefore modified by introduction of a grid search for k_{sol} . It was found to be sufficient to perform the grid search only for k_{sol} , while letting B_{sol} and the other scale factors being determined by least-squares optimization for each selected value of k_{sol} (Table 1). Others have found a similar solution to this problem involving a grid search of both k_{sol} and B_{sol} (see ref. 14). On a rare occasion, the CNS 1.2 procedure produces a non-converging solution for a particular value of k_{sol} , which is then excluded by choosing the minimum R -value solution for the entire search.

It should be noted that proper implementation of the overall anisotropic thermal factor refinement is complicated by restrictions imposed by crystallographic symmetry on the individual components of the thermal factor tensor¹⁰. To simplify the calculations, the diffraction data are temporarily expanded to space group $P1$, the overall anisotropic thermal factor refinement is performed in $P1$, followed by reduction to the particular space group. This procedure ensures the proper symmetry restrictions on the thermal factor tensor while the computational overhead in $P1$ is not of particular concern on modern computing platforms.

For refinements at medium to high resolution (up to around 3 Å resolution), $R_{\text{probe}} = R_{\text{shrink}} = 1$ is the optimum choice¹⁵. However, for the refinement of the ATPase p97/VCP at 4.5 Å resolution, it was necessary at the early stages of the refinement to adjust R_{probe} and R_{shrink} for computation of the bulk solvent mask to obtain optimum R and R_{free} values^{16,17}. Initially, optimization showed that different values for each of the R_{probe} and R_{shrink} parameters could produce a slightly lower R_{free} . However, in the final refinement stages, equal values of R_{probe} and R_{shrink} produced optimum results. Thus, in CNS 1.2, both parameters are changed simultaneously with $R_{\text{shrink}} = R_{\text{probe}}$. Furthermore, both values tend to approach the value of 1 as the atomic model is being completed and improved.

This entire procedure was implemented with new modules (‘scale_and_solvent’ and ‘scale_and_solvent_grid_search’) and suitably modified task files without any source code modification. Extensive testing indicated that the CNS 1.2 bulk solvent model is robust for structures solved at both high and low resolution. The new solvent model may require significantly more computing time in the start-up phase of refinement tasks compared to the previous version. However, compared to the overall computing time for refinement, this additional time is usually insignificant. The

new method often yields lower R values than the previous version of CNS.

Thermal factor sharpening of electron density maps

Thermal (B')-factor sharpening is a useful tool for enhancement of low-resolution maps^{16–19}. Thermal factor sharpening entails the use of a negative B_{sharp} value in a resolution-dependent weighting scheme applied to a particular electron density map:

$$F_{\text{sharpened_map}} = \exp(-B_{\text{sharp}} \sin^2 \theta / \lambda^2) \times F_{\text{map}} \quad (4)$$

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. Applying a negative B_{sharp} value effectively up-weights higher resolution terms. The result of this weighting scheme is increased

detail for higher resolution features such as side-chain conformations. However, the cost of the increased detail can be increased noise throughout the electron density map. Sometimes, the noise can coincide with regions of backbone or side-chain electron density, producing potential artifacts. Thus, thermal factor sharpening is a density-modification technique that is only as good as the diffraction data and phases that are available, and therefore, the original unweighted electron density maps should always be considered. Furthermore, little improvement is observed on electron density maps that are computed with phases derived solely from molecular replacement, so experimental phase information appears to be important to get the most benefit from thermal factor sharpening¹⁶.

Thermal factor sharpening can be viewed as a simple weighting function applied to the observed amplitudes and consequently the electron density maps by virtue of relative scaling between atomic

The screenshot shows a web-based interface for editing a CNS task file. The browser address bar shows 'http://cns.csb.yale.edu - Edit refine.inp'. The interface is organized into several sections:

- HL coefficients:** A table with columns A, B, C, and D.
- reciprocal space array containing weighting scheme for observed amplitudes:** optional, only used for the "residual" and "vector" targets - this will default to a constant value of 1 if array is not present.
- complex reciprocal space array containing experimental phases:** optional, required for the "mixed" and "vector" targets.
- reciprocal space array containing experimental figures of merit:** optional, required for the "mixed" target.
- resolution limits to be used in refinement:** the full resolution range of observed data should be used in refinement. A bulk solvent correction should be applied to allow the use of low resolution terms. If no bulk solvent correction is applied, data must be truncated at a lower resolution limit of between 8 and 6 Angstrom.

	lowest	highest
resolution	500.0	2.0
- apply rejection criteria to amplitudes or intensities:** amplitude intensity
- Observed data cutoff criteria:** applied to amplitudes or intensities, reflections with magnitude(Obs)/sigma < cutoff are rejected. Value: 0.0
- rms outlier cutoff:** applied to amplitudes or intensities, reflections with magnitude(Obs) > cutoff*rms(Obs) will be rejected. Value: 10000
- non-crystallographic symmetry:**
 - NCS-restraints/constraints file: see auxiliary/ncs.def
- overall B-factor and bulk solvent corrections:**
 - overall B-factor correction: no isotropic anisotropic
 - bulk solvent correction: a mask is required around the molecule(s). The region outside this mask is the solvent region. true false
 - bulk solvent mask file: mask will be read from O type mask file if a name is given otherwise calculated from coordinates of selected atoms.
 - automatic bulk solvent parameter search: true false
 - optional file with a listing of the results of the automatic bulk solvent grid search:
- fixed solvent mask parameters if the automatic option is not used:**

	probe radius (A)	shrink radius (A)
bulk solvent	1.0	1.0
- fixed solvent parameters if the automatic option is not used:**

	e-density level (e/A^3)	B-factor (A^2)
bulk solvent	-1	-1
- atom selection:** select atoms to be included in refinement, this should include all conformations if multiple conformations are used. List: known and not hydrogen.

Figure 1 | A screenshot of a section of the task file relevant to bulk solvent modeling.



model and observed amplitudes. In the case of the crystal structure of the ATPase p97/VCP, the B_{sharp} value that produced the most useful electron density map coincided with the smallest absolute value of B_{sharp} that results in a Wilson plot that is positive in all resolution bins¹⁹. The new task file 'bsharp.inp' implements this approach for determining B_{sharp} . However, it remains to be seen if this empirical rule applies to the general case. Thermal factor sharpening can also be viewed as a pseudo Wilson scaling of the diffraction data, so another reasonable choice for B_{sharp} would be to set it to the negative Wilson B value of the diffraction data.

The implementation in version 1.2 of CNS applies equation (4) at the step just before application of the fast Fourier transformation to compute the actual three-dimensional map. In other words, the Fourier coefficients for the particular electron density map

are computed and then equation (4) is applied to them. Thus, this procedure works for any type of electron density map based on experimental or model phases and with any type of weighting or phase-combination scheme. The only required modifications involved changes of the CNS task and module files related to electron density map calculations. No changes in the FORTRAN source code were required. The thermal factor sharpening feature does not impose any particular additional computational time compared to the previous version of CNS. One could envision a generalization to anisotropic thermal factor sharpening that could be particularly useful for highly anisotropic crystals.

The procedure contains two examples of how the new features of CNS can be used. The first example is for crystallographic refinement and the second example is for computing an electron density map.

MATERIALS

EQUIPMENT

A computer with access to the Internet and a web browser. Implementations for some of the most common operating systems are available, including Linux and Mac OS X.

Data All CNS data formats are text files. For coordinates, the format of the Protein Data Bank (PDB)²⁰ is supported. For diffraction data, conversion programs are available in CNS and the CCP4 suite²¹ to import files generated by commonly used data reduction programs.

maps are calculated $u \cdot F_o - v \cdot F_c$
eg. 2fo-fc map -> u=2 and v=1 or fo-fc map -> u=1 and v=1

specify u	2
specify v	1
type of map	sigmaaa
refinement target - only used for gradient maps	mlf
number of bins for refinement target	-1
use model amplitudes for unmeasured data	<input type="radio"/> true <input checked="" type="radio"/> false
scale map by dividing by the rms sigma of the map	<input checked="" type="radio"/> true <input type="radio"/> false
map format	cns
b-factor sharpening (A ²), for example, -100	0
map grid size: dmin*grid	0.33
memory allocation for FFT calculation	-1
extent of map	molecule

*sigmaaa: (u m/Fo/ - v D/Fc/)^exp(i phi_calc)
m and D calculated from sigmaaa
unweighted: (u |Fo/ - v k/Fc/)^exp(i phi_calc)
no figure-of-merit weighting
combined: (u m/Fo/ ^exp(i phi_comb) - v D/Fc/ ^exp(i phi_calc))
model and experimental phases combined, m and D from sigmaaa
observed: (u m/Fo/ ^exp(i phi_obs) - v k m/Fc/ ^exp(i phi_calc))
observed phases and fom from phase probability distribution
gradient: d(target)/dFc
gradient of the current crystallographic target wrt Fc
NB. experimental phases must be supplied as a phase
probability distribution in the Hendrickson-Lattman arrays*

*mlf: maximum likelihood target using amplitudes
mli: maximum likelihood target using intensities
mhl: maximum likelihood target using amplitudes
and phase probability distribution
residual: standard crystallographic residual
vector: vector residual
mixed: (1-fom)*residual + fom*vector
e2e2: correlation coefficient using normalized E²
e1e1: correlation coefficient using normalized E
f2f2: correlation coefficient using F²
f1f1: correlation coefficient using F*

this will be determined automatically if a negative value is given otherwise the specified number of bins will be used

this will not be applied to gradient or difference maps

otherwise map will be on an absolute fobs scale

choice: "cns" "ezd"

this will be determined automatically if a negative value is given otherwise the specified number of words will be allocated

Figure 2 | A screenshot of a section of the task file for computation of an electron density map.



Programs The CNS website <http://cns-online.org> contains all task files ('Input Files'), parameter and library files for the program ('Libraries'), the module files ('Modules'), information about specific hardware implementations ('Installation'), a syntax manual ('Syntax Manual') and a tutorial for the most common tasks ('Tutorial'). This website can also be used to modify the task files and assign default values to certain parameters, such as the unit cell dimensions

and space group symmetry. The website provides information on how to obtain the CNS program.

The user performs the calculations on local computers using task files downloaded or modified by the CNS website. To execute a task file on a local computer, it is advisable to create a directory that contains all required input files.

In a UNIX environment, CNS is executed by the following command:
CNS <input_file> output_file

PROCEDURE

- 1| The starting coordinate file must be in the PDB format. Convert the diffraction data file to the CNS format.
- 2| Go to the CNS website's 'Input Files' section.
- 3| Set the default parameters (space group, cell dimensions and optionally, anomalous form factors) by going to the 'Set' menu, and then click on 'Start editing files'.
- 4| If the diffraction data file already contains a test set for cross-validation, skip this step. Otherwise, select the task file 'make_cv.inp' in the 'Refinement' section. Enter any changes that may be required. Save the file to a local computer and execute CNS. The task file creates a new diffraction data file that contains the test set information in addition to the observed data.
- 5| If you would like to continue with the general crystallographic refinement task file ('refine.inp') or compute an electron density map, follow option A or B respectively.

(A) General crystallographic refinement

- (i) Select the task file 'refine.inp' in the 'Refinement' section and enter information that is required, such as the name of the coordinate file, the diffraction data file with test set (e.g., generated by the previous step) or the name of the output coordinate file. If the structure contains ligands or other non-standard residues, the 'generate.inp' task file in the 'General' section has to be run first to generate a 'molecular topology' file. If the protein contains *cis* peptide bonds, except for proline, a special parameter file needs to be generated by the task file 'cis_peptide.inp' in the 'General' section and the resulting parameter file read in all subsequent refinement tasks.
- (ii) Read additional restraints, such as non-crystallographic symmetry restraints or constraints.
- (iii) Change options and parameters of the 'refine.inp' task file as required. The default values should be reasonable for most cases. The following is a screenshot of a section of the task file relevant to bulk solvent modeling (Fig. 1).
- (iv) Save the file to your local computer and execute CNS. Examine the output file for errors. Usually, it is sufficient to go to the bottom of the output file and make sure that it is saved normally without an abort message.
- (v) The coordinate file that is generated by 'refine.inp' contains useful information at the top of the file (see ANTICIPATED RESULTS). The resulting coordinate file can be viewed or further manipulated with any molecular graphics program that supports the PDB format.

(B) Computation of an electron density map

- (i) Select the 'model_map' task file in the 'Refinement' section. Change options and parameters of the task file, as required. The default values should be reasonable for most cases. The following is a screenshot of a section of the task file (Fig. 2). Note the line that specifies the thermal ('B') factor sharpening.
- (ii) Save the file to a local computer and execute CNS. Examine the output file for errors. Usually, it is sufficient to go to the bottom of the output file and make sure that it is saved normally without an abort message.

```
REMARK disulphide added: from 11 to 27
REMARK disulphide added: from 45 to 73
REMARK coordinates from minimization and B - factor refinement
REMARK refinement resolution: 500.0- 2.0 A
REMARK starting r= 0.3842 free_r= 0.3612
REMARK final r= 0.2526 free_r= 0.3005
REMARK rmsd bonds= 0.006775 rmsd angles= 1.48826
REMARK B rmsd for bonded mainchain atoms= 1.318 target= 1.5
REMARK B rmsd for bonded sidechain atoms= 2.036 target= 2.0
REMARK B rmsd for angle mainchain atoms= 2.093 target= 2.0
REMARK B rmsd for angle sidechain atoms= 2.838 target= 2.5
REMARK target= mlf final wa= 2.1395
REMARK final rweight= 0.0813 (with wa= 2.1395)
REMARK md-method= torsion annealing schedule= constant
REMARK starting temperature= 2000 total md steps= 1 * 100
REMARK cycles= 3 coordinate steps= 20 B-factor steps= 10
REMARK sq= P2(1)2(1)2(1) a= 61.76 b= 40.73 c= 26.74 alpha= 90 beta= 90 gamma= 90
REMARK topology file 1 : CNS_TOPPAR:protein.top
REMARK topology file 2 : CNS_TOPPAR:dna-ma.to p
REMARK topology file 3 : CNS_TOPPAR:water.top
REMARK topology file 4 : CNS_TOPPAR:ion.top
REMARK parameter file 1 : CNS_TOPPAR:protein_rep.param
REMARK parameter file 2 : CNS_TOPPAR:dna-ma_rep.param
REMARK parameter file 3 : CNS_TOPPAR:water_rep.param
REMARK parameter file 4 : CNS_TOPPAR:ion.param
REMARK molecular structure file: automatic
REMARK input coordinates: amy.pdb
REMARK reflection file= amy.cv
REMARK ncs= none
REMARK Anisotropic B -factor tensor Ucart of atomic model without isotropic component :
REMARK B11= -3.889 B22= 0.031 B33= 3.858
REMARK B12= 0.000 B13= 0.000 B23= 0.000
REMARK Isotropic component added to coordinate array B: -0.381
REMARK bulk solvent: probe radius=1, shrink value=1
REMARK bulk solvent: density level= 0.4 e/A^3, B -factor= 41.1087 A^2
REMARK reflections with IFobs/sigma_F < 0.0 rejected
REMARK reflections with IFobs > 10000 * rms(Fobs) rejected
REMARK fft gridding factor = 0.3333, B factor offset = 4.41002 A^2, Elimit = 8
REMARK theoretical total numb er of refl. in resol. range: 4905 ( 100.0 % )
REMARK number of unobserved reflections (no entry or IF|=0): 1549 ( 31.6 % )
REMARK number of reflections rejected: 0 ( 0.0 % )
REMARK total number of reflections used: 3356 ( 68.4 % )
REMARK number of reflections in working set: 3016 ( 61.5 % )
REMARK number of reflections in test set: 340 ( 6.9 % )
CRYST1 61.760 40.730 26.740 90.00 90.00 P 21 21 21
REMARK FILENAME=?Users/axel/cns_solve_1.2/inputs/xtal_refine/run/refine.pdb*
REMARK DATE:27-Jun-2007 21:37:15 created by user: brunger
REMARK VERSION:1.2
ATOM 1 CB ASP 1 4.960 41.839 12.138 1.00 36.91
ATOM 2 CG ASP 1 5.092 41.294 13.548 1.00 38.14
ATOM 3 OD1 ASP 1 4.070 40.845 14.119 1.00 38.21
```

Figure 3 | The top of the refined PDB file produced by the task file 'refine.inp'.



ANTICIPATED RESULTS

Figure 3 shows the top of the refined PDB file produced by the task file 'refine.inp'.

Examine the reduction in *R* values and the values of the bulk solvent model. A significant increase of the free *R* value indicates problems with the model. The B_{sol} parameter should be within a factor of 3 of the average *B* value of the atomic model, and the k_{sol} parameter should be reasonably close to the expected value of water electron density ($\sim 0.3 \text{ e } \text{\AA}^{-3}$). The coordinate file also contains information about the local geometry of the refined model (average deviation of bond lengths and bond angles from ideal values) and the variation of thermal factors for atoms connected through one or two bonds.

A thermal factor sharpened map should always be compared with the original map that was computed without thermal factor sharpening, as artifacts can be introduced by thermal factor sharpening.

Note: Supplementary information is available via the HTML version of this article.

ACKNOWLEDGMENTS I am grateful to Paul Adams and Byron DeLaBarre for stimulating discussions.

Published online at <http://www.natureprotocols.com>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions>

1. Brunger, A.T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr.* **54** (Part 5): 905–921 (1998).
2. Brunger, A.T. *X-PLOR, version 3.1. A System for X-ray Crystallography and NMR* (Yale University Press, New Haven, Connecticut, USA, 1992).
3. Schwieters, C.D., Kuszewski, J.J., Tjandra, N. & Clore, G.M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65–73 (2003).
4. Dominguez, C., Boelens, R. & Bonvin, A.M. HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737 (2003).
5. Linge, J.P., Habeck, M., Rieping, W. & Nilges, M. ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics (Oxford, England)* **19**, 315–316 (2003).
6. Moulinier, L., Case, D.A. & Simonson, T. Reintroducing electrostatics into protein X-ray structure refinement: bulk solvent treated as a dielectric continuum. *Acta Crystallogr.* **59**, 2094–2103 (2003).
7. Wall, M.E., Subramaniam, S. & Phillips, G.N. Jr. Protein structure determination using a database of interatomic distance probabilities. *Protein Sci.* **8**, 2720–2727 (1999).
8. Clarage, J.B. & Phillips, G.N. Jr. Cross-validation tests of time-averaged molecular dynamics refinements for determination of protein structures by X-ray crystallography. *Acta Crystallogr.* **50**, 24–36 (1994).
9. Wang, H. & Stubbs, G. Molecular dynamics in refinement against fiber diffraction data. *Acta Crystallogr. A* **49**, 504–513 (1993).
10. Grosse-Kunstleve, R. & Adams, P. On the handling of atomic anisotropic displacement parameters. *J. Appl. Crystallogr.* **35**, 477–480 (2002).
11. Lee, B. & Richards, F.M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400 (1971).
12. Richards, F.M. Calculation of molecular volumes and areas for structures of known geometry. *Methods Enzymol.* **115**, 440–464 (1985).
13. Phillips, S.E. Structure and refinement of oxymyoglobin at 1.6 Å resolution. *J. Mol. Biol.* **142**, 531–554 (1980).
14. Afonine, P.V., Grosse-Kunstleve, R.W. & Adams, P.D. A robust bulk-solvent correction and anisotropic scaling procedure. *Acta Crystallogr.* **61**, 850–855 (2005).
15. Jiang, J.S. & Brunger, A.T. Protein hydration observed by X-ray diffraction. Solvation properties of penicillopepsin and neuraminidase crystal structures. *J. Mol. Biol.* **243**, 100–115 (1994).
16. DeLaBarre, B. & Brunger, A.T. Complete structure of p97/valosin-containing protein reveals communication between nucleotide domains. *Nat. Struct. Biol.* **10**, 856–863 (2003).
17. DeLaBarre, B. & Brunger, A.T. Nucleotide dependent motion and mechanism of action of p97/VCP. *J. Mol. Biol.* **347**, 437–452 (2005).
18. Bass, R.B., Strop, P., Barclay, M. & Rees, D.C. Crystal structure of *Escherichia coli* MscS, a voltage-modulated and mechanosensitive channel. *Science* **298**, 1582–1587 (2002).
19. DeLaBarre, B. & Brunger, A.T. Considerations for the refinement of low-resolution crystal structures. *Acta Crystallogr.* **62**, 923–932 (2006).
20. Westbrook, J.D. & Fitzgerald, P.M. The PDB format, mmCIF, and other data formats. *Methods Biochem. Anal.* **44**, 161–179 (2003).
21. CCP. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.* **50**, 760–763 (1994).

Supplementary Note

List of Changes for version 1.2

Program:

- default number of atoms set to 200,000
- atom number can go up to 999999 in the PDB files produced by CNS by using some of the 'ATOM' and 'HETATM' record if the number goes above 99999
- fixed bug in treatment of HL coefficients for remapping of reflections for some spacegroups
- installation files for x86_64 linux platforms (using gfortran or ifort)
- installation files for Intel Mac systems (using ifort)
- installation files for Mac PPC systems
- Intel Mac OSX binary distribution
- all known dynamic memory pointer problems fixed
- Windows binary installation removed
- use of sort function updated in cns_solve_env file
- use of sed updated in cns_solve_env file
- consolidated distribution into one tar file
- updated output files in the test directory
- generic CPU_TIME subroutine call for Intel Mac OSX, Intel X86_64, and Intel i686
- replaced -tpp7 option with -axP for Intel X86_64 and Intel i686
- removed "norm" function from atom property do command
- enabled DISP=KEEP for CLOSe statement on linux and mac intel versions

Task files:

- updated bulk solvent model and scale factor calculations (developed for robust refinement and low resolution applications):
 - all refinement and model map task files:
 - new solvent model grid search and solrad/shrink search
 - absolute grid size range for solvent mask calculation is forced to be between $[\sqrt{3} \cdot 0.577 = 1] 0.57$ and 0.9 \AA
 - removed low_res_bscale parameter, the low resolution limit is used
 - reported anisotropic tensor in coordinate files is now Ucart (not Ucif as before)
 - fixed anisotropic B-scaling symmetry restraints in all modules (scalef,
 - removed "anisotropic_fixed_isotropic" option throughout all script files
 - fixed method to remove isotropic component from anisotropic B-tensor
- new bulk solvent and scaling routines in structure deposition files:
 - xtal_pdbsubmission/xtal_pdbsubmission.inp
 - xtal_mmcif/deposit_mmcif.inp

- fixed bug when a structure has a lower resolution than the default Wilson limit:
xtal_pdbsubmission/xtal_pdbsubmission.inp
- automatic and fixed bulk solvent parameter options:
all refinement and model map task files
- new automatic sharpening B-factor determination script file:
xtal_util/bsharp.inp
- B-sharpening parameter introduced into various map files:
xtal_util/fourier_map.inp
xtal_refine/model_map.inp
sa_omit_map.inp
composite_omit_map.inp
fo-fo_map.inp
- corrected local real space correlation coefficient computation:
xtal_refine/model_map.inp
composite_omit_map.inp
water_delete.inp
xtal_twin/model_map_twin.inp
- fixed Wilson calculation for very large structures
xtal_util/bsharp.inp
analyse.inp
scale.inp
- fixed bug in water picking if no sites are found:
xtal_refine/water_pick.inp
xtal_twin/water_pick_twin.inp
- duplicate water parameters no longer read in water picking:
xtal_refine/water_pick.inp
xtal_twin/water_pick_twin.inp
- introduced optional B-factor reset option:
xtal_refine/refine.inp
- increased the default number of cycles to 3:
xtal_refine/refine.inp
- removed multiple coordinate file reading option:
xtal_refine/composite_omit_map.inp
sa_omit_map.inp
model_map.inp

xtal_util/model_phase.inp
xtal_twin/model_map_twin.inp

- removed option to write separate bulk solvent model structure factors:
xtal_util/model_phase.inp
- increased the default number of macrocycles to 15:
xtal_phase/mad_phase.inp
- changed the default output array name to fmodel and added warning if fcalc is used as the name:
xtal_util/model_fcalc.inp
- increased maxtree to 20:
in all torsion angle input files
- Ucart reported instead of Ucif:
heavy atom site refinement scripts
- modified Patterson map task files to use ascii map files instead of binary map files:
xtal_patterson/predict_patterson.inp
patterson_map.inp
- updated density_modify.inp task file:
mask completion
warning about absolute scaling of input amplitudes
- new input file to calculate structure factors from an EM reconstruction map:
xtal_util/em_map_to_hkl.inp

HTML Documentation:

- updated installation documentation

Modules:

- fixed bug in siterefine module: FOMs for centrics were non-zero for SAD phasing
- removed "anisotropic_fixed_isotropic" option throughout all modules:
scaletarget, scaletarget2, scale_and_solvent
- new module: scale_solvent
- new module for automatic determination of FFT gridding parameters
fft_parameter_check:
 - determines optimum B-factor increment
 - ELIM parameter is automatically increased for atomic resolution structures (starting at 1.4 Å and higher)

- default ELIM value is set to 8
- grid size is set to the min of (1, $d_{\min} * 0.3333$)

Libraries:

- conformational database library files incorporated into distribution
libraries/confdb/*

Utilities:

- modified Patterson map programs to use ascii map files instead of binary
map files:
utils/PSmap*