Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

New computational tools for H/D determination in macromolecular structures from neutron data

Dritan Siliqi,* Rocco Caliandro, Benedetta Carrozzini, Giovanni Luca Cascarano and Annamaria Mazzone

Institute of Crystallography, CNR, Via G. Amendola 122/O, 70126 Bari, Italy

Correspondence e-mail: dritan.siliqi@ic.cnr.it

Received 14 April 2010
Accepted 2 July 2010

Two new computational methods dedicated to neutron crystallography, called *n-FreeLunch* and *DNDM-NDM*, have been developed and successfully tested. The aim in developing these methods is to determine hydrogen and deuterium positions in macromolecular structures by using information from neutron density maps. Of particular interest is resolving cases in which the geometrically predicted hydrogen or deuterium positions are ambiguous. The methods are an evolution of approaches that are already applied in X-ray crystallography: extrapolation beyond the observed resolution (known as the *FreeLunch* procedure) and a difference electron-density modification (*DEDM*) technique combined with the electron-density modification (*EDM*) tool (known as *DEDM-EDM*). It is shown that the two methods are complementary to each other and are effective in finding the positions of H and D atoms in neutron density maps.

1. Notation

The following notation has been used in this article.
EDM/*NDM*: electron/neutron density modification.
DEDM/*DNDM*: difference electron/neutron density modification.
ρ, ρ_p: electron/neutron densities of a structure composed of N atoms and of its partial model composed of p atoms, respectively.
$\rho_q = \rho - \rho_p$: ideal difference Fourier synthesis.
F, F_p, F_q: structure factors of ρ, ρ_p and ρ_q, respectively. φ, φ_p and φ_q are their phases.
E, E_p, E_q: normalized structure factors of ρ, ρ_p and ρ_q, respectively.
$\Sigma_N = \sum_{j=1}^{N} f_j^2$, $\Sigma_p = \sum_{j=1}^{p} f_j^2$, where N/p are the number of atoms in the unit cell and f_j are the atomic scattering factors for the structure (thermal factor included).
$D = \langle \cos(2\pi \mathbf{h}\Delta\mathbf{r}) \rangle$: the estimate of the error in a partial structure from the coordinate errors (Luzzati, 1952), where $\Delta\mathbf{r}$ is the error in the positions of the p atoms. The average involves the p atoms and is performed per resolution shell.
$\sigma_A = D(\Sigma_p/\Sigma_N)^{1/2}$: the σ_A parameter.
$\langle |\mu|^2 \rangle$: expected measurement error on $|F|$.
$e = 1 + (\langle |\mu|^2 \rangle / \Sigma_N)$: normalized measurement error.
$I_i(x)$: modified Bessel function of order i.
$m = \langle \cos(\varphi - \varphi_p) \rangle = I_1(x)/I_0(x)$, where $x = 2\sigma_A |E||E_p|/(e - {\sigma_A}^2)$: figure of merit.

2. Introduction

Historically, the use of neutrons in macromolecular crystallography has been small compared with that of X-rays, owing to

the requirement for larger crystals in order to overcome the relatively low flux of neutron sources and the weaker interaction of neutrons with matter. Crystal volumes of 1 mm^3 are typically required in order to investigate the structure of a protein by neutron single-crystal diffraction and such sample volumes are difficult to obtain with biological macromolecules. For biological systems, the interest in neutron radiation for structural studies arises mainly from the relatively strong and positive interaction between neutrons and deuterium nuclei compared with hydrogen nuclei. The situation is now changing owing to improvements in instrumentation, data collection and sample preparation (Blakeley *et al.*, 2008). Furthermore, new and more powerful neutron sources are becoming available that will enable the use of smaller samples (Teixeira *et al.*, 2008).

Neutron diffraction provides a way of determining the positions of polypeptide H atoms in proteins from their negative density. However, a problem with neutron scattering from H atoms is the presence of strong incoherent scattering, which introduces significant noise into the diffraction data. Furthermore, interference between the negative density associated with H atoms and the positive density associated with C, N and O atoms can cause density-cancellation effects. These problems can be circumvented by producing fully deuterated samples or by collecting data to sufficiently high resolution (Afonine *et al.*, 2010; Langan *et al.*, 2008).

For medium- to high-resolution data (2.5–1 Å or higher), the strong scattering of neutrons from D nuclei means that the water molecules may be completely determined by inspection of the nuclear density map obtained using neutron data. Less-ordered water molecules may be still observed in a nuclear density map as single peaks (Niimura & Bau, 2008). With X-rays, complete water molecules may only be observed if they are well ordered and by using data at very high resolution, *i.e.* greater than 1 Å (see, for example, Hazemann *et al.*, 2005; Baker & Hubbard, 1984).

Recently, Adams *et al.* (2009) developed a joint X-ray and neutron structure-refinement procedure, which was combined with recent advances in modern structure-refinement tools to provide a new generalized method for complete (including hydrogen) structure refinement. This method requires the collection of both neutron and X-ray data at the same temperature from crystals grown under similar conditions. The procedure has been included in two crystallographic software packages: *CNS* (Brünger *et al.*, 1998) and *PHENIX* (Adams *et al.*, 2010). The X-ray model is used as the starting model; the H atoms are subsequently added and the model is optimized to obtain the best agreement with the observed neutron data. The initial positions of H atoms can be predicted from known geometries for most of the functional groups. The remaining initial positions can only be determined by inspecting the neutron scattering density map through Fourier and difference Fourier syntheses. Both packages can also manage refinement against neutron data alone.

At the same time, new advanced computational tools have been developed for phase refinement and model completion using X-ray data. In particular, a procedure called *FreeLunch* can be used to improve the phase values of observed reflections and to increase the interpretation of the corresponding electron-density maps by using the extrapolated phases and amplitudes of non-observed reflections (Caliandro *et al.*, 2005*a*), while a procedure called *DEDM* improves the potentialities of the difference Fourier synthesis and, when combined with the standard electron-density modification (*EDM*) technique, allows the calculation of more efficient partial models (Caliandro *et al.*, 2008). Both methods are included in the package *IL MILIONE* (Burla *et al.*, 2007) for protein structure determination.

Starting from these new tools, which were originally developed and successfully used for X-ray data, we have devised dedicated procedures to improve the quality of neutron density maps obtained by Fourier and difference Fourier syntheses. The first, called *n-FreeLunch*, is aimed at extrapolating the neutron data behind and beyond the observed resolution; the second, called the *DNDM* procedure, is aimed at improving the difference Fourier map. Some preliminary tests performed on selected structures from the Protein Data Bank (Bernstein *et al.*, 1977; Berman *et al.*, 2002) for which neutron data sets are available will be presented.

3. *n-FreeLunch*: extrapolation behind and beyond the observed resolution

Limited experimental resolution is a commonly observed feature of macromolecular crystallography: it may hinder or make difficult the determination of the crystal structure. In several papers (Caliandro *et al.*, 2005*a*,*b*, 2007), a novel procedure known as *FreeLunch* has been presented which extrapolates the moduli and phases of nonmeasured reflections beyond and behind the experimental resolution limit from an approximated electron-density map. Applications to a number of test structures have demonstrated that the extrapolation can be successfully accomplished. The phase estimates of the observed reflections are improved and the use of extrapolated values for nonmeasured reflections provides additional features to the map. As a consequence, the interpretability of the electron-density maps increases and they show a resolution higher than the experimental resolution.

The above-described *FreeLunch* procedure has been modified to cope with neutron data. In this case, we suppose that a model without hydrogen (or deuterium) positions obtained from X-ray data is available (for brevity, this model will be called the 'X-ray model'). To check the full power of the procedure, we did not include the unambiguous H or D positions which can be determined geometrically in the X-ray model. Let ρ_{NX} be the map obtained using observed (neutron) amplitudes and phases calculated from the X-ray model (without H/D atoms) by means of neutron scattering factors. This map represents the starting point of our procedure, the most important features of which may be summarized as follows:

(i) the higher resolution limit for the extrapolation is determined automatically according to

$$\mathrm{RES_{ext}} = \mathrm{RES_{obs}}[(1 - \mathrm{MIS})(1 + \mathrm{GEN})]^{-1/3}, \qquad (1)$$

where $\mathrm{RES_{obs}}$ is the resolution limit of the experimental data (in Å), MIS is the percentage of nonmeasured reflections up to $\mathrm{RES_{obs}}$ and GEN is the ratio (number of extrapolated reflections)/(number of measured reflections). The formula is based on geometrical considerations, according to which the number of extrapolated reflections is given by the nonmeasured reflections lying in the reciprocal-space sphere defined by the experimental resolution limit (the so-called missing reflections) plus the reflections lying in a spherical shell outside this sphere, while the number of measured reflections is proportional to the volume of the reciprocal sphere defined by the experimental resolution limit. The size of the spherical shell is arbitrary, but may be fixed by the parameter GEN. The chosen value of GEN is a compromise between two opposing requirements: the necessity of extrapolating as many reflections as possible in order to improve the phases of the observed reflections and the increasing weight of the wrong phase and modulus estimates for the extrapolated reflections when one moves too far away from $\mathrm{RES_{obs}}$ towards the high-resolution zone. We have chosen GEN = 2.0 when $\mathrm{RES_{obs}} < 1.7$ and GEN = 2.2 when $\mathrm{RES_{obs}} \geq 1.7$ (for further details, see Caliandro *et al.*, 2007).

(ii) The algorithm is cyclic and consists of two steps. The number of cycles in each step depends on the $\mathrm{RES_{obs}}$ parameter, which ranges from 10 to 50. Each cycle implements a Fourier transform and an inverse Fourier transform operation. The former is intended to invert a density map (say ρ), *i.e.* to compute structure factors from the map giving values to a set of phases (say φ), and the latter is intended to calculate a new density map by using the previous phase set and observed amplitudes. The cycles may therefore be represented by $\rho \rightarrow \varphi \rightarrow \rho$. In the first operation (namely in the half-cycle $\rho \rightarrow \varphi$) the current map ρ is modified: only 10% of the neutron density (the portion with the largest positive and negative intensities) is preserved and the rest is set to zero. The negative part of the map is considered beside the positive part to account for the signal from H atoms. The map is then inverted and extrapolation takes place, since it is also possible to assign a phase value to unobserved reflections. In the second operation (*i.e.* in the half-cycle $\varphi \rightarrow \rho$) selected extrapolated reflections are used together with the observed reflections to build a new density map. The selection is based on the value of the extrapolated amplitudes and the number of extrapolated reflections added to the measured reflections increases with the cycle number. In the second step the map-modification criteria are changed. The fraction of ρ preserved ranges from 10 to 30% of the neutron density (the portion with the largest positive and negative intensity) according to the cycle number and a mask is applied as follows: a molecular envelope is calculated from the X-ray model using the method of Wang (1985) and a mask is derived by setting the portion of the map covered by the highest part of the envelope (the part above its average value) to 0.5 and the remaining part of the unit cell to 1 and is applied to the current map ρ by multiplying each point of the neutron density map by the corresponding value of the mask. In this way, the current map is reduced in the region covered by the X-ray model and left unvaried outside this region, where H/D atoms (not included in the X-ray model) are expected to be located. Therefore, this map modification will highlight the contribution from H/D atoms not included in the X-ray model.

It is worthwhile noting that this map modification differs from that used in the standard *FreeLunch* procedure, in which the molecular envelope is calculated from the current electron-density map and the mask is used to enhance the contributions from its highest parts. Moreover, in our experience the extrapolated amplitudes cannot be used for refinement owing to large errors (instead of very good phase estimation for these reflections). For this reason, the free set (the reflections used for R_{free} calculation) is not changed.

The *n-FreeLunch* procedure was first applied to the structure 1iu6 (Chatake *et al.*, 2004), which is a rubredoxin mutant containing 53 residues, 290 H atoms and 88 D atoms in the asymmetric unit and which was calculated from neutron data collected to 1.5 Å resolution. Two neutron density maps were considered in the analysis: the first, corresponding to ρ_{NX}, represents the initial map and has the experimental resolution, while the second (denoted by ρ_{FL}) represents the final map obtained after application of *n-FreeLunch* and has resolution extended to 1.11 Å according to (1). These maps are compared in Fig. 1, in which the negative parts of ρ_{NX} contoured at 1σ (blue) and of ρ_{FL} contoured at 2σ (green) are shown superimposed on the published coordinates of the Ala16 residue. It may be noted that the overall interpretability of the neutron map is increased: two of the three H atoms bonded to

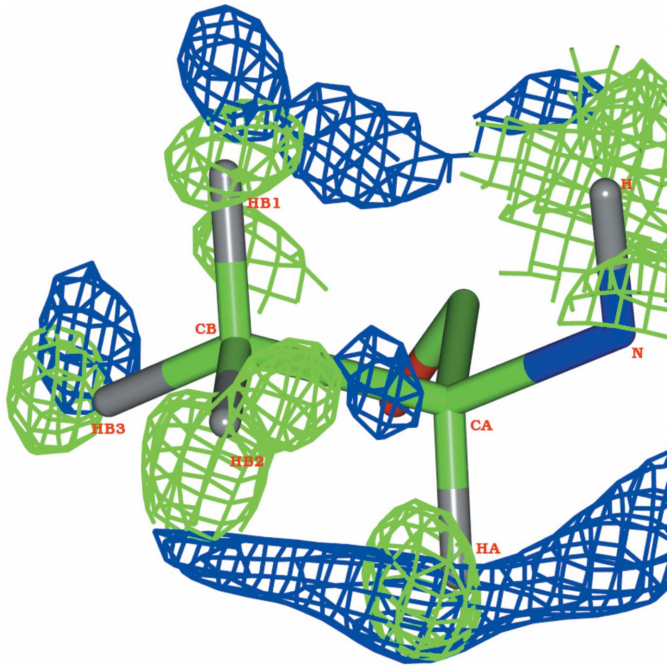

Figure 1
1iu6: negative neutron density maps before (blue) and after (green) *n-FreeLunch* contoured at 1σ and 2σ, respectively, superposed on the published coordinates of the Ala16 residue. The two maps correspond to those denoted ρ_{NX} and ρ_{FL} in the text. The figure was prepared using *CCP4mg* (Potterton *et al.*, 2002).

Table 1
Statistics for the deepest peaks in the negative Fourier neutron density map before and after the application of the *n-FreeLunch* procedure as a function of the σ level at which the map has been cut (see equation 2).

NP is the number of peaks found, NCP is the number of these with a distance of less than 1.0 Å from the published positions of the H atoms and DIST is the mean distance in Å of the NCP peaks from their published positions.

	Initial map (ρ_{NX})				Map after *n-FreeLunch* (ρ_{FL})			
Structure	t	NP	NCP	DIST	t	NP	NCP	DIST
1iu6	−6.0	22	1	0.54	−6.0	140	43	0.32
	−3.0	113	8	0.52	−3.7	300	72	0.50
	−2.3	300	38	0.66				
2mb5	−3.0	86	2	0.91	−3.0	429	60	0.68
	−2.0	854	97	0.81	−2.0	896	157	0.70
	−1.4	1000	65	0.84	−1.9	1000	164	0.70

CB (HB1 and HB3) and the HA atom bonded to CA are better located in the ρ_{FL} map, while the HB2 atom, which was not present in the ρ_{NX} map, has appeared. For the HB atoms, the additional information available in the ρ_{FL} map may be used to unambiguously determine their positions.

The procedure has been further applied to the carbonmonoxymyoglobin structure 2mb5 (Cheng & Schoenborn, 1990), which contains 153 residues, 1004 H atoms and 475 D atoms in the asymmetric unit. In this case, the experimental resolution of the neutron data is 1.81 Å and the *n-FreeLunch* procedure extrapolates the data to 1.52 Å resolution according to (1). The negative parts of the ρ_{NX} map contoured at 1σ (blue) and of the ρ_{FL} map contoured at 2σ (green) superimposed on the published coordinates of the Lys47 residue are shown in Fig. 2. In this case the improvement is substantial: most of the H atoms can only be identified in the ρ_{FL} map.

For both test structures, a more quantitative comparison may be carried out by considering the deepest peaks in the negative parts of ρ_{NX} and ρ_{FL} and relating them to the true hydrogen positions (those of the model deposited in the PDB) according to the following protocol. The neutron density maps have been cut at different levels, with threshold values given by the formula

$$\rho_{th} = \langle\rho\rangle + t\sigma, \qquad (2)$$

where $\langle\rho\rangle$ and σ are the mean and standard deviation values of the neutron density map, respectively. All pixels with values greater than ρ_{th} are set to zero. For each cutoff level, the number of peaks found in the two maps and the number with a distance of less than 1.0 Å from the published positions of the H atoms were determined. These two numbers, denoted NP and NCP, respectively, are reported in Table 1 separately for the ρ_{NX} and ρ_{FL} maps, together with the corresponding values of t and the mean distance of the NCP peaks from their true positions (DIST). The lowest t values have been chosen, so that the number of selected peaks is nearly equal to the total number of H atoms to be found. For the structure 1iu6, only 38 peaks out of 300 were found within 1 Å of their true positions in the ρ_{NX} map, with a mean distance about 0.66 Å, while in the improved map ρ_{FL} 72 peaks were found with a mean distance from the true positions of 0.50 Å. Even better improvements can be seen for the structure 2mb5, for which the NCP among 1000 peaks increases from 65 for the ρ_{NX} map to 164 for the ρ_{FL} map.

Table 2
Statistics for the deepest peaks in the negative Fourier neutron density map corresponding to the ambiguous H positions before and after the application of the *n-FreeLunch* procedure.

N_{amb} is the total number of H atoms whose positions cannot be determined by stereochemistry, NCP is the number of peaks with a distance less than 1.0 Å from the published positions of the N_{amb} atoms and DIST is the mean distance in Å of the NCP peaks from their published positions.

		Initial map (ρ_{NX})		Map after *n-FreeLunch* (ρ_{FL})	
Structure	N_{amb}	NCP	DIST	NCP	DIST
1iu6	51	4	0.69	8	0.46
2mb5	91	6	0.79	29	0.70

A more detailed analysis is reported in Table 2, in which only peaks related to H atoms with ambiguous positions have been considered. In this case the number of correctly located peaks again increases when going from the ρ_{NX} to the ρ_{FL} map, with a greater ratio for the 2mb5 structure than for the 1iu6 structure.

As a further test, in addition to the initial map (ρ_{NX}) and that improved by *n-FreeLunch* (ρ_{FL}) a third map was considered: that obtained by purely extending the resolution of the data by replacing all unobserved reflections up to RES_{ext} with structure factors calculated from the X-ray model (without H/D atoms) by means of neutron scattering factors. The features of the map (which we designate ρ'_{NX}) obtained using this simplified resolution-extension technique have been analyzed in the same way as for the previous maps and the

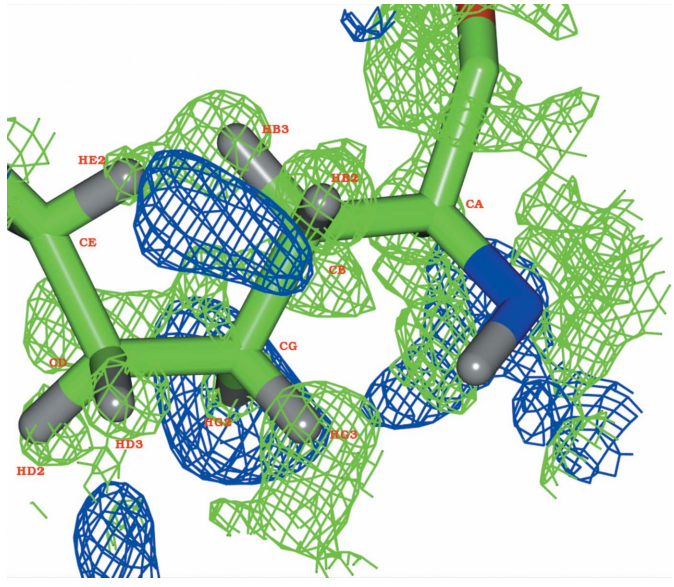

Figure 2
2mb5: negative neutron density map before (blue) and after (green) *n-FreeLunch* contoured at 1σ and 2σ, respectively, superposed on the published coordinates of the Lys47 residue. The two maps correspond to those denoted ρ_{NX} and ρ_{FL} in the text. The figure was prepared using *CCP4mg* (Potterton *et al.*, 2002).

results are reported in Table 3. The NCP for this map is much lower than that of ρ_{FL} and even lower than that of ρ_{NX} for both of the test structures. The DIST parameter is lower in ρ'_{NX} than in ρ_{NX} but much higher than in ρ_{FL}. It may be argued that the observed hierarchy in the quality of the three maps is governed by the amount of information about the missing H/D atoms that they contain: in ρ_{NX} this information only resides in the observed neutron amplitudes, while in ρ'_{NX} the resolution is extended but the information is confined to the set of observed amplitudes, so that its overall quality deteriorates, and in ρ_{FL} the information is instead also contained in the phases of both the observed and extrapolated reflections as a result of the *n-FreeLunch* cycles.

These tests demonstrate that *n-FreeLunch* allows a better location of the H atoms in the negative part of the neutron density map. For practical purposes, this may be used to discriminate between H/D for atoms that are well predicted by the geometry and to determine the positions of H atoms that are not unambiguously defined by the stereochemistry.

When applied to the positive part of the neutron density map to better locate the D atoms, the extrapolation procedure was not as successful. This may be explained by considering the competitive contribution from the other protein atoms, which binds the changes in the density map and hinders the recovery of completely new atoms.

The resolution limit for *n-FreeLunch* will only be set after extensive tests of the procedure, but from our experience with X-ray data the extrapolation procedure should have negligible effects for structures at resolutions lower than 2.5 Å.

4. *DNDM*: improved difference Fourier synthesis

Many Fourier syntheses can be used to recover missing atoms when a partial structural model is available (Ramachandran & Raman, 1959; Srinivasan, 1961; Ramachandran & Srinivasan, 1970). The use of this approach is mostly restricted to the observed Fourier synthesis, the difference Fourier synthesis and their combination. They are currently used in X-ray crystallography in their weighted forms, with the following coefficients:

(i) $m|F|\exp(i\varphi_p)$ for the observed Fourier synthesis (Sim, 1959),

(ii) $\big|m|F| - D|F_p|\big| \exp[i(\varphi_p + s\pi)]$, where $s = 1$ or 0 according to whether the sign of $m|F| - D|F_p|$ is positive or negative, for the difference Fourier synthesis (Henderson & Moffat, 1971; Main, 1979; Read, 1986; Ursby & Bourgeois, 1997; Urzhumtsev *et al.*, 1996) and

(iii) $(2m|F| - |F_p|)\exp(i\varphi_p)$ for the combination of Fourier and difference Fourier synthesis.

All three syntheses suffer from an unavoidable bias generated by the use of prior information. Indeed,

(i) the phases of the observed Fourier synthesis may only assume φ_p values: the supplemental information about missing atoms is only provided by the observed moduli;

(ii) in the difference Fourier synthesis the phase values may be φ_p or $\varphi_p + \pi$; hence, φ and φ_p are assumed to be collinear. In addition, $\big|m|F| - D|F_p|\big|$ is a rough approximation of $|F_q|$ for reflections which have φ_p different from φ (the triangle formed by F, F_p and F_q collapses into a line);

(iii) similar considerations hold for syntheses which are combination of the previous two.

Table 3
Statistics for the deepest peaks in the negative Fourier neutron density map obtained by replacing all unobserved reflections up to RES_{ext} with structure factors calculated from the X-ray model (without H/D atoms) by means of neutron scattering factors as a function of the σ level at which the map has been cut (see equation 2).

NP is the number of peaks found, NCP is the number of peaks with a distance less than 1.0 Å from the published positions of the H atoms and DIST is the mean distance in Å of the NCP peaks from their published positions.

	ρ'_{NX}			
Structure	t	NP	NCP	DIST
1iu6	−2.42	300	10	0.39
2mb5	−1.90	1000	63	0.76

A powerful technique for the Fourier synthesis is the widely used electron-density modification (*EDM*) procedure (Lunin, 1993; Cowtan, 1994, 1999; Abrahams, 1997; Abrahams & Leslie, 1996; Zhang *et al.*, 2001, Refaat & Woolfson, 1993, Giacovazzo & Siliqi, 1997). This is a cyclic process based on sensible modifications of the electron-density maps, aimed at improving the phase estimates by capturing the desired features of the map (*e.g.* its positivity, the molecular envelopes, its statistical description *via* density histograms *etc.*).

Recently, the difference Fourier synthesis has been revisited (Caliandro *et al.*, 2008). The main achievements may be summarized as follows:

(i) the variance of F_q was derived,

$$\sigma_q^2 = (1 - m^2)|F|^2 + \langle|\mu|^2\rangle, \tag{3}$$

and the accuracy of the φ_q estimates is expected to be inversely proportional to σ_q;

(ii) reflections with the same value of $\big(\big|m|F| - D|F_p|\big|\big)$ have different values of σ_q and hence different accuracies of the related φ_q. In particular, reflections with large negative values of $\big(\big|m|F| - D|F_p|\big|\big)$ have the most accurate φ_q estimates, as their σ_q values are small;

(iii) reflections with large positive or negative values of $\big(\big|m|F| - D|F_p|\big|\big)$ are quite useful for defining the difference Fourier synthesis but are of limited use for the Fourier synthesis, since their phase φ_p is a rough approximation of φ (they have $|F|$ very different from $|F_p|$).

The above results are the basis of the *DEDM* (difference electron-density modification) algorithm, which is based on modification of the difference electron density (Caliandro *et al.*, 2009*a*,*b*). It works in a way that is substantially independent of *EDM* and levers on the phases of reflections with small σ_q values (which are left practically unvaried during the process). The *DEDM* procedure breaks down the collinearity between model structure phases and difference structure phase estimates. It consists of repeated cycles which implement the following steps.

(i) An initial estimate of the difference Fourier map is obtained by using coefficients $\left|m|F| - D|F_p|\right| \exp[i(\varphi_p + s\pi)]$. This map is denoted $(\rho_q)_{\rm initial}$.

(ii) The map is modified by squaring its highest positive and deepest negative parts (each corresponding to 10% of the map pixels) and setting the remainder to zero.

(iii) Fourier inversion of the modified map generates new phase values φ_q which differ from the initial estimate $(\varphi_q + s\pi)$. New modulus values are calculated accordingly using the formula

$$|F_q| = -D|F_p|\cos(\Delta\varphi_q) + [m^2|F|^2 - D^2|F_p|^2\sin^2(\Delta\varphi_q)]^{1/2}, \quad (4)$$

where $\Delta\varphi_q = \varphi_q - (\varphi_p + s\pi)$. This formula was derived by applying the Carnot theorem to the triangle formed by mF, DF_p and F_q.

(iv) A new difference Fourier map is calculated with coefficients $|F_q|\exp(i\varphi_q)$, which undergoes a new cycle starting from (ii).

At the end of the procedure, a final difference Fourier map is obtained which is improved with respect to the initial one, *i.e.* it shows maxima corresponding to atoms that were not present in the initial model ρ_p. This map is denoted $(\rho_q)_{\rm final}$.

DEDM is expected to be complementary with respect to *EDM*, since reflections with large $|F_p|$ and $|F|$ values largely contribute to the Fourier synthesis but are of limited use for the difference Fourier synthesis, while reflections with large $|F_p|$ and small $|F|$ values may not be important for the Fourier synthesis but are very informative for the difference Fourier synthesis. These findings led to the development of a combined procedure, called *DEDM-EDM*, which takes advantage of the complementarity of the two techniques (Caliandro *et al.*, 2009*a*,*b*). The procedure may be outlined as follows.

(i) *DEDM* cycles are executed, which, starting from an initial estimate of the difference Fourier map $(\rho_q)_{\rm initial}$, lead to an improved one $(\rho_q)_{\rm final}$.

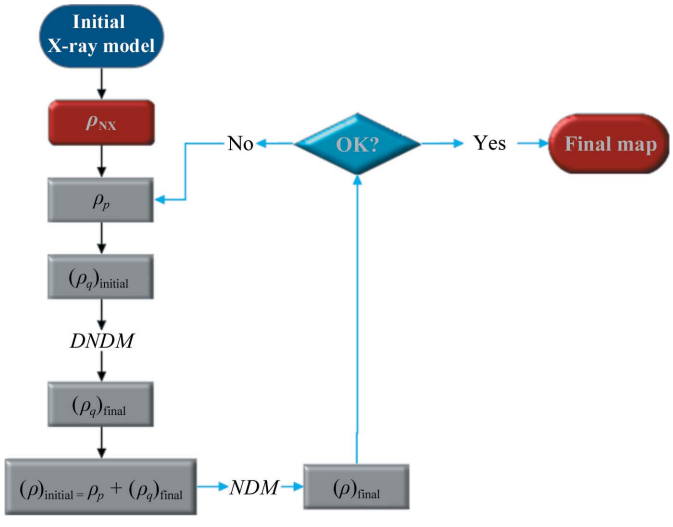

Figure 3
DNDM-NDM iteration procedure flowchart.

Table 4
Statistics for the highest peaks in the positive difference Fourier neutron density map before and after application of the *DNDM-NDM* procedure as a function of the σ level at which the map has been cut (see equation 2).

NP is the number of peaks found, NCP is the number of these with a distance of less than 1.0 Å from the published positions of the H atoms and DIST is the mean distance in Å of the NCP peaks from their published positions.

	Initial map $(\rho_q)_{\rm initial}$				Map after *DNDM-NDM*			
Structure	t	NP	NCP	DIST	t	NP	NCP	DIST
1iu6	5.0	8	8	0.26	5.0	6	6	0.16
	4.0	32	28	0.29	4.0	27	22	0.25
	2.9	90	48	0.32	3.0	90	42	0.30
2mb5	4.0	8	7	0.60	4.0	16	34	0.39
	3.0	110	53	0.58	3.0	101	68	0.49
	2.1	500	134	0.64	2.2	500	183	0.54

(ii) The vectors F_q obtained at the end of the *DEDM* cycles are combined with the initial structure-factor vectors F_p through the equation $F = F_p + F_q$. In direct space, this corresponds to the operation $(\rho)_{\rm initial} = \rho_p + (\rho_q)_{\rm final}$.

(iii) Standard *EDM* cycles are performed which, starting from $(\rho)_{\rm initial}$, lead to an improved density map. In these cycles the molecular envelope is applied.

(iv) A new iteration of the procedure starts by setting $\rho_p = (\rho)_{\rm final}$ and calculating a new $(\rho_q)_{\rm initial}$ with updated coefficients.

(v) The number of *DEDM* and *DEDM-EDM* cycles is not fixed. The numbers are decided by using figures of merit (for details, see Caliandro *et al.*, 2009*a*).

In analogy with the procedures devised for X-rays, a difference neutron density modification (*DNDM*), a neutron density modification (*NDM*) and a combination of the two (*DNDM-NDM*) have been developed. In Fig. 3 the flowchart for the *DNDM-NDM* iterations is shown. Differences from the X-ray implementation are in the starting structural model, which in the case of neutrons is calculated by using phases from the X-ray model and neutron amplitudes (*i.e.* $\rho_p = \rho_{\rm NX}$ as defined in §3) and in the use of the molecular envelope in the *NDM* cycles. The molecular envelope is used in the same way as in the *n-FreeLunch* procedure and is applied to enhance the part of the neutron density map which is not covered by the X-ray model.

DNDM-NDM has been used to improve the positive part of the neutron density maps in order to better locate the D atoms. The results of its application to the 2mb5 structure are reported in Table 4, with the same notation used as in Table 1. We considered the first 500 highest peaks of the positive difference neutron density map. In the initial map ($\rho_{\rm NX}$) only 134 peaks among them are correct (*i.e.* they have a distance less than 1 Å from the published D-atom positions), with a mean distance of 0.59 Å. For the map obtained at the end of *DNDM-NDM* [$(\rho)_{\rm final}$] we found 183 correct peaks with a smaller distance from their true positions (0.54 Å). An analysis as a function of the σ level reveals that the correctness of the peaks (measured by the ratio NCP/NP) is better for the highest part of the neutron density map (at large σ levels).

The improvement of the difference Fourier map may also be useful for determining the protonation states of histidines. Fig. 4 shows selected views of the neutron difference Fourier map before (blue) and after (green) the application of the *DNDM-DNM* procedure, contoured at 1σ and 1.5σ, respectively, and superimposed on the published coordinates of the His64 residue. The ND1 atom may be clearly distinguished from the NE2 atom in the final neutron density map on the basis of the peak occurring at the DD1 deuterium site.

5. Conclusions

Stemming from new tools recently developed for X-ray protein crystallography which are used for phase refinement and model completion, we have devised dedicated algorithms for neutron protein crystallography. Two computational tools for H/D determination have been developed: *n-FreeLunch*, which extrapolates the phases and amplitudes of observed reflections to unobserved reflections falling behind and beyond the experimental resolution limit, and *DNDM-NDM*, which performs a difference Fourier synthesis with increased potentialities. The new methods have been tested on neutron data from two protein structures deposited in the PDB, obtaining an increase in the number of H and D atoms found just by inspection of the neutron density map. In addition, they were successfully used to locate H or D atoms that are ambiguously defined by the stereochemistry. It was found that *n-FreeLunch* is particularly suited to searching for H atoms, while *DNDM-NDM* gives the best results for D atoms. This complementarity suggests that they may be most productively used in combination to achieve the maximum information from the neutron density map. The tools have been tested starting from a structural model completely deprived of H/D atoms, so it is expected that their efficiency would increase if H or D positions that are geometrically predictable are included. It has been suggested that these tools could be exploited in combination with the joint X-ray and neutron refinement recently included in the *PHENIX* package to increase the success rate of protein structure determination. Moreover, they may easily be extended to cope with H/D determination in DNA/RNA molecules.

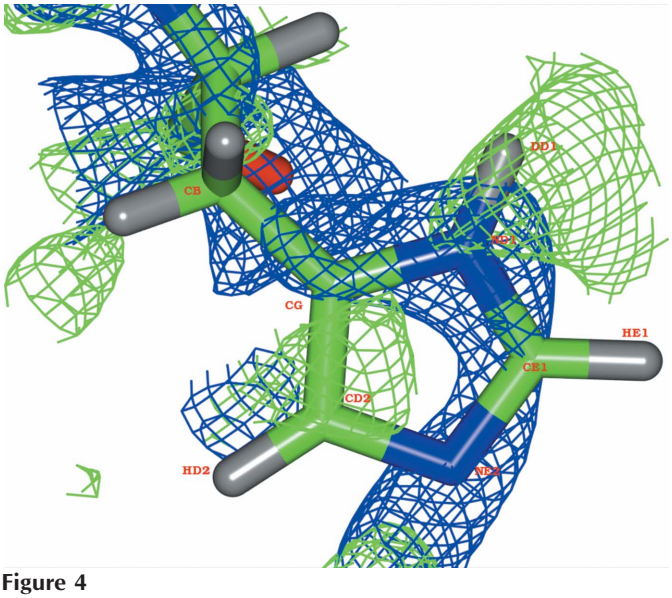

Figure 4
2mb5: positive difference Fourier neutron density map before (blue) and after (green) *DNDM-NDM* contoured at 1σ and 1.5σ, respectively, superposed on the published coordinates of the His64 residue. The figure was prepared using *CCP4mg* (Potterton *et al.*, 2002).

Most of this work was supported by DS's CNR Short Term Mobility Program for the years 2008 and 2009 at Los Alamos National Laboratory (Los Alamos, New Mexico, USA). We would like to thank Dr Paul Langan and Dr Marat Mustyakimov from the Protein Crystallography Station at Los Alamos National Laboratory (Los Alamos, New Mexico, USA) for very important and helpful discussions.

References

Abrahams, J. P. (1997). *Acta Cryst.* D**53**, 371–376.
Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Adams, P. D., Mustyakimov, M., Afonine, P. V. & Langan, P. (2009). *Acta Cryst.* D**65**, 567–573.
Afonine, P. V., Mustyakimov, M., Grosse-Kunstleve, R. W., Moriarty, N. W., Langan, P. & Adams, P. D. (2010). *Acta Cryst.* D**66**, 1153–1163.
Baker, E. N. & Hubbard, R. E. (1984). *Prog. Biophys. Mol. Biol.* **44**, 97–179.
Berman, H. M. *et al.* (2002). *Acta Cryst.* D**58**, 899–907.
Bernstein, F. C., Koetzle, T. F., Willians, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.
Blakeley, M., Langan, P., Niimura, N. & Podjarny, A. (2008). *Curr. Opin. Struct. Biol.* **18**, 593–600.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Burla, M. C., Caliandro, R., Camalli, M., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C., Polidori, G., Siliqi, D. & Spagna, R. (2007). *J. Appl. Cryst.* **40**, 609–613.
Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Siliqi, D. (2005*a*). *Acta Cryst.* D**61**, 556–565.
Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Siliqi, D. (2005*b*). *Acta Cryst.* D**61**, 1080–1087.
Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Siliqi, D. (2007). *J. Appl. Cryst.* **40**, 931–937.
Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Siliqi, D. (2008). *Acta Cryst.* A**64**, 519–528.
Caliandro, R., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Mazzone, A. M. & Siliqi, D. (2009*a*). *Acta Cryst.* D**65**, 249–256.
Caliandro, R., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Mazzone, A. M. & Siliqi, D. (2009*b*). *Acta Cryst.* D**65**, 477–484.
Chatake, T., Kurihara, K., Tanaka, I., Tsyba, I., Bau, R., Jenney, F. E., Adams, M. W. W. & Niimura, N. (2004). *Acta Cryst.* D**60**, 1364–1373.
Cheng, X. & Schoenborn, B. P. (1990). *Acta Cryst.* B**46**, 195–208.
Cowtan, K. (1999). *Acta Cryst.* D**55**, 1555–1567.
Cowtan, K. D. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
Giacovazzo, C. & Siliqi, D. (1997). *Acta Cryst.* A**53**, 789–798.
Hazemann, I., Dauvergne, M. T., Blakeley, M. P., Meilleur, F., Haertlein, M., Van Dorsselaer, A., Mitschler, A., Myles, D. A. A. & Podjarny, A. (2005). *Acta Cryst.* D**61**, 1413–1417.

Henderson, R. & Moffat, J. K. (1971). *Acta Cryst.* B**27**, 1414–1420.
Langan, P., Fisher, Z., Kovalevsky, A., Mustyakimov, M., Sutcliffe Valone, A., Unkefer, C., Waltman, M. J., Coates, L., Adams, P. D., Afonine, P. V., Bennett, B., Dealwis, C. & Schoenborn, B. P. (2008). *J. Synchrotron Rad.* **15**, 215–218.
Lunin, V. Yu. (1993). *Acta Cryst.* D**49**, 90–99.
Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.
Main, P. (1979). *Acta Cryst.* A**35**, 779–785.
Niimura, N. & Bau, R. (2008). *Acta Cryst.* A**64**, 12–22.
Potterton, E., McNicholas, S., Krissinel, E., Cowtan, K. & Noble, M. (2002). *Acta Cryst.* D**58**, 1955–1957.
Ramachandran, G. N. & Raman, S. (1959). *Acta Cryst.* **12**, 957–964.
Ramachandran, G. N. & Srinivasan, R. (1970). *Fourier Methods in Crystallography*. London: Wiley–Interscience.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Refaat, L. S. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 367–371.
Sim, G. A. (1959). *Acta Cryst.* **12**, 813–815.
Srinivasan, R. (1961). *Acta Cryst.* **14**, 607–611.
Teixeira, S. C. M. *et al.* (2008). *Chem. Phys.* **345**, 133–151.
Ursby, T. & Bourgeois, D. (1997). *Acta Cryst.* A**53**, 564–575.
Urzhumtsev, A. G., Skovoroda, T. P. & Lunin, V. Y. (1996). *J. Appl. Cryst.* **29**, 741–744.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Zhang, K. Y. J., Cowtan, K. D. & Main, P. (2001). *International Tables for Crystallography*, Vol. *F*, edited by M. G. Rossmann & E. Arnold, pp. 311–331. Dordrecht: Kluwer Academic Publishers.