

## Experimental assessment of differences between related protein crystal structures

Gerard J. Kleywegt

Department of Cell and Molecular Biology,  
Uppsala University, Biomedical Centre, Box  
596, SE-751 24 Uppsala, Sweden

Correspondence e-mail: gerard@xray.bmc.uu.se

Received 20 April 1999

Accepted 30 July 1999

Prior to attaching any biological significance to differences between two related protein crystal structures, it must be established that such differences are genuine, rather than artefacts of the structure-determination protocol. This will be all the more important as more and more related protein structures are solved and comparative structural biology attempts to correlate structural differences with variations in biological function, activity or affinity. A method has been developed which enables unbiased assessment of differences between the structures of related biomacromolecules using experimental crystallographic information alone. It is based on the use of local density-correlation maps, which contain information regarding the similarity of the experimental electron density for corresponding parts of different copies of a molecule. The method can be used to assess *a priori* which parts of two or more molecules are likely to be structurally similar; this information can then be employed during structure refinement. Alternatively, the method can be used *a posteriori* to verify that differences observed in two or more models are supported by the experimental information. Several examples are discussed which validate the notion that local conformational variability is highly correlated to differences in the local experimental electron density.

### 1. Introduction

An important aspect of structural biology is the study of the relationship between structure and function and of the effect of structural changes on the function, activity or affinity of biomacromolecules. Usually, such studies involve a structural comparison of related forms of a biomacromolecule in different biological or physical states or conditions. Examples include comparisons of mutant *versus* wild-type proteins, of free proteins *versus* their complexes with inhibitors, substrate analogues, co-factors or other biomolecules, of proteins under different pH, temperature or solvent conditions and of multiple independent copies of a protein located in the asymmetric unit of a crystal (so-called non-crystallographic symmetry or NCS). However, before any conclusions can be drawn regarding the possible biological relevance of structural differences, it is imperative that the veracity or significance of such differences be established first. In the past few years, protein crystallographers have come to realise that 'observed' structural differences between related copies of a molecule are sometimes the result of the use of inappropriate model-refinement protocols rather than a reflection of genuine structural dissimilarity (Kleywegt & Jones, 1995, 1997; Kleywegt, 1996; Kleywegt & Brünger, 1996; Kleywegt *et al.*, 1996). In 1995, Kleywegt & Jones argued, based on experiments

involving the use of the independent free  $R$  value (Brünger, 1992, 1993; Kleywegt & Brünger, 1996), that ignoring or restraining NCS during refinement might well be a major determinant of 'observed' structural dissimilarity (Kleywegt & Jones, 1995). Subsequently, it was found that the magnitude of structural differences between NCS-related molecules found in the Protein Data Bank (PDB), when measured in terms of main-chain torsion-angle variation, increased essentially linearly with deteriorating resolution (Kleywegt, 1996). In one case, it was demonstrated that inappropriate refinement protocols had disguised serious errors in a protein model and had introduced structural differences that were entirely artefactual (Kleywegt *et al.*, 1996). All these observations indicate that the prevailing practice in the protein crystallographic community of equating large structural differences with significant ones is highly inappropriate, in particular when medium- and low-resolution structures are involved. Obviously, a more objective method is highly desirable.

Assessing structural differences between refined models is an *a posteriori* validation problem. However, crystallographers often face *a priori* dilemmas, the resolution of which will greatly impact on the final result. Such dilemmas occur when NCS is present (*e.g.* should NCS-related molecules be treated independently or should they be restrained to be similar or constrained to be identical?) or when a structure is to be refined that is very similar to a known structure (*e.g.* should a 3.0 Å mutant or complex structure be restrained to be similar to a 1.8 Å wild-type or native structure?).

NCS occurs in roughly half of all low-resolution (worse than 2.5 Å) protein crystallographic studies (Kleywegt, 1996). Most protein crystallographers realise that NCS provides a powerful mechanism of increasing the effective data-to-parameter ratio during crystallographic refinement, if it is employed using NCS constraints or restraints. In addition, difference refinement techniques, as pioneered by Terwilliger & Berendzen (1995, 1996), emphasize the need to employ structural similarities during refinement of related models. In the future, it will be possible to refine native, complex and mutant structures of a macromolecule simultaneously, thereby exploiting the geometric redundancies to the full (Bricogne, 1974). At present, however, the crystallographer is faced with a problem when deciding, for instance, whether NCS should be constrained or restrained during refinement. If NCS is restrained, there is a constellation of possible ways to implement this, *e.g.* different weights for different parts of the structure, different weights for main-chain and side-chain atoms, no restraints for 'floppy' loops or exposed side chains *etc.* At present, there is no method to obtain *a priori* information concerning which parts of a structure can be expected to display substantial structural differences and which parts can be expected to be essentially identical. A practical solution which many crystallographers use is to inspect NCS-averaged electron density and to use strong NCS-restraints for those parts of the model that have good averaged density.

Here, a method is presented which uses experimental information (present in a so-called 'local density-correlation map'; Vellieux *et al.*, 1995) alone to assess structural differ-

ences between related protein crystal structures. The method is general and potentially quantitative and can be used both *a priori* and *a posteriori*. In addition, it is completely unbiased, provided that experimental phase information is available for all (crystal) forms of the molecule under study. Although many methods exist to quantify structural differences between models based on coordinates and temperature factors (*e.g.* Korn & Rose, 1994; Kleywegt & Jones, 1995; Stroud & Fauman, 1995; Kleywegt, 1996; Abagyan & Totrov, 1997), to my knowledge this is the first method which uses the experimental crystallographic data alone to investigate the validity of any observed differences.

## 2. Local density-correlation maps

Read and co-workers (Vellieux *et al.*, 1995) have introduced local density-correlation maps as a tool to generate appropriate molecular envelopes (or masks) for use in electron-density averaging procedures (Vellieux & Read, 1997; Kleywegt & Read, 1997). Such maps are calculated on a coarse grid which encompasses the space inside the crystal that is expected to be occupied by the molecule or domain for which a (set of) Cartesian NCS rotation/translation operator(s) is assumed to be valid. The only input it requires is an electron-density map (calculated using experimental phase information) and the NCS operator(s). The latter can also be obtained using experimental information alone: for instance, from superpositioning of heavy-atom or selenium sites followed by operator refinement through global density-correlation optimization or from molecular-replacement calculations using the density of one copy of the molecule or domain (Vellieux & Read, 1997; Kleywegt & Read, 1997). For every point on the grid, the density within a sphere around it is extracted and this is repeated for each NCS-related copy of the grid point. The density values within the spheres are then used to calculate correlation coefficients, which are subsequently combined into  $r_1$  or  $r_2$  scores (Vellieux *et al.*, 1995). The linear correlation coefficient,  $CC(i, j)$ , of two sets of electron-density values  $\rho_i$  and  $\rho_j$  is defined as

$$CC(i, j) = \frac{\langle \rho_i \rho_j \rangle - \langle \rho_i \rangle \langle \rho_j \rangle}{(\langle \rho_i^2 \rangle - \langle \rho_i \rangle^2)^{1/2} (\langle \rho_j^2 \rangle - \langle \rho_j \rangle^2)^{1/2}}, \quad (1)$$

where  $\langle \rangle$  denotes an average value taken over all grid points inside the current sphere. The  $r_1$  and  $r_2$  scores (Vellieux *et al.*, 1995) are defined as

$$r_1(i) = [\sum_{j \neq i} CC(i, j)] / (N - 1), \quad (2)$$

$$r_2 = [\sum_i \sum_{j \neq i} CC(i, j)] / [N(N - 1)/2]. \quad (3)$$

In other words,  $r_1$  is the average correlation coefficient of one copy of the density and each of the other copies, whereas  $r_2$  is the average correlation coefficient of all possible pairs [if  $N = 2$ ,  $r_1 = r_2 = CC(1,2)$ ]. If the grid point concerned is part of the molecular envelope for which the NCS operators are valid, this will result in very similar spheres of density around it and around all of the NCS-related grid points and hence the

correlation of the spherical densities will be high. If, on the other hand, the grid point is situated in the solvent region (or in any part of space for which the NCS relationships are not valid), the correlation between related spheres of density is expected to be low (correlation coefficient close to zero).

The local density-correlation (LDC) map calculated as outlined above thus contains information regarding the local similarity of the electron density for each grid point and its associated NCS-related grid points. The density correlation is determined entirely by the electron-density map and by the NCS operator(s). If the map is calculated using purely experimental phases (*e.g.* MIR, MAD) and the operators were also derived from the data alone, it follows that the correlation map then reflects the information present in the experimental data alone regarding the local similarity of the electron density. In other words, regions of high LDC values are regions where the models of the various NCS-related molecules ought to be very similar (since their electron density is very similar) and *vice versa*.

In its original form (Vellieux *et al.*, 1995), the LDC map is calculated for NCS-related molecules, but the method can easily be extended to correlate the density of two related molecules in different crystals or crystal forms (Kleywegt & Jones, 1999). Hence, LDC maps can also be calculated for different crystal forms of a particular molecule (or for identical crystal forms determined independently in different laboratories), for a wild-type protein and a mutant or for an unliganded and a complexed form of a protein *etc.* This means that provided experimental phase information is available for both forms, unbiased information regarding regions which are similar or dissimilar in the two forms can be extracted using the experimental data alone.

LDC maps are therefore extremely powerful tools, both for the *a priori* design of appropriate restraint schemes and for the *a posteriori* validation of structural differences. For instance, one could retrieve the average value of the LDC map around each atom, group of atoms or residue and use these to classify them as strongly, moderately or weakly similar or as dissimilar, and use this information to restrain them appropriately during subsequent refinement. Alternatively, if one has refined a wild-type and a mutant structure, for instance, and one finds that there are seemingly large structural differences between the models, an LDC map can be calculated using the densities of the two molecules. Inspection of this map will reveal whether or not the regions of structural dissimilarity in the models coincide with regions of space where the electron density is poorly correlated. If this is not the case, the observed structural differences may be a consequence of an inappropriate refinement protocol for one or both models.

One important exception arises if part of a structure does not obey the Cartesian rotation/translation relationships which were used to calculate the LDC map. For instance, for a two-domain protein with a flexible hinge region, it may be more appropriate to define separate Cartesian operators for each domain and to relate local structural differences only to the correlation map calculated with the appropriate operator. The fact that a fixed set of operators is used also means that, in

this particular context, structural differences are best assessed in terms of (r.m.s.) distances, rather than torsion-angle differences (Kleywegt, 1996).

If the LDC map is to be a reflection of experimental information regarding structural similarity, the map(s) from which it is calculated must be based on purely experimental phase information. However, if this is not available (*e.g.* if a structure was solved by molecular-replacement methods), the technique can probably still be used (albeit with caution), in particular in the presence of non-crystallographic symmetry. In this case, there is one extra source of information which enters into the calculation, namely the molecular envelope that is used during averaging. One should be aware of the fact that the definition of such an envelope is subjective and that it will introduce some bias (but note that this is not the case when experimental phase information is available). For example, if a loop or side chain is not covered by the molecular envelope, its density will be flattened, which will lead to low values for the local density correlation. However, if the mask is of sufficiently high quality, this is probably a minor problem. Real-space electron-density averaging has been demonstrated to be able to converge from almost random phases to an essentially correct phase set, in the process removing model bias (*e.g.* Braig *et al.*, 1994; Kleywegt & Read, 1997). Hence, phases obtained after averaging are often as good as (if not better than) experimental phases, so that the LDC method can probably be used in such cases as well (*caveat emptor*). The phase set obtained after the final averaging cycle should be used for map calculation. Parts of the molecular envelope or model for which the NCS relationships do not hold are not expected to display well defined and correlated density in the various NCS-related copies. If neither experimental phase information nor NCS is available, the reliability of the present method is compromised by model bias. Although the effect of model bias can be suppressed to some extent through the calculation of  $\sigma_A$  maps (Read, 1986) or simulated-annealing omit maps (Hodel *et al.*, 1992), it remains to be established if these methods are sufficiently powerful for the present purpose. Further development of this method for cases where experimental phases are not available will presumably require maximum-likelihood approaches.

### 3. Results and discussion

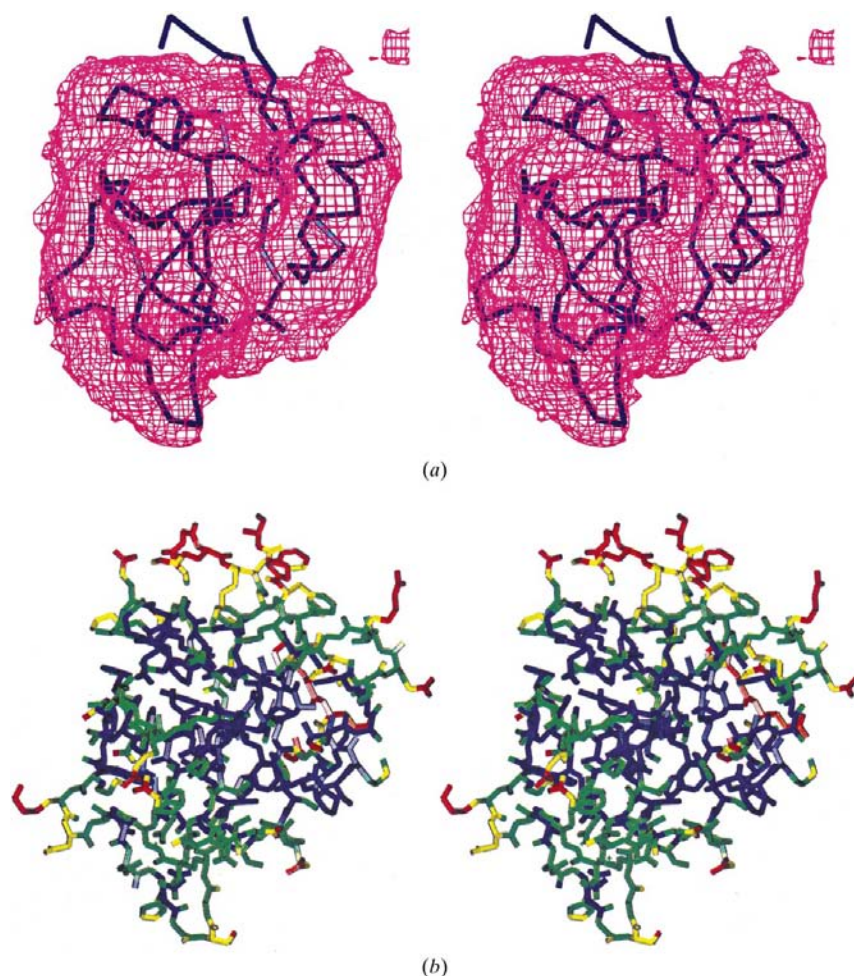
In the examples described below, local software (available to the community) was used for all computations. Local density correlation maps were calculated with *COMA* (in the case of NCS) or *MASKIT* (for different crystals or crystal form; Kleywegt & Jones, 1999). Local average correlation values for individual atoms were extracted from the correlation maps using *MAPMAN* (Kleywegt & Jones, 1996; 'PEek SPHERE' command). Structural superpositioning and comparison was performed with *LSQMAN* (Kleywegt, 1996). An illustrated tutorial showing how to carry out the calculations is available at [http://alpha2.bmc.uu.se/usf/dens\\_corr.html](http://alpha2.bmc.uu.se/usf/dens_corr.html).

It is important to note that although the applications below are limited to proteins, the LDC method itself is entirely

general and can be applied to any type of molecule (nucleic acids, ligands, inhibitors, covalent attachments, water molecules, metals, inorganic ions *etc.*).

### 3.1. Mannose-binding protein

For a 230-residue dimeric fragment of rat mannose-binding protein, a 1.8 Å resolution MAD-phased electron-density map is available (Burling *et al.*, 1996) and this provides an excellent case for testing the assumption that the LDC map provides experimental information concerning the similarity or dissimilarity of NCS-related molecules. The experimental data set is provided by Dr Axel Brunger (Yale University) as part of the distribution of the software package *CNS* (Brunger *et al.*, 1998). Coordinates of the final model (refined without any NCS constraints or restraints; Burling *et al.*, 1996) were retrieved from the PDB (Bernstein *et al.*, 1977; entry 1ytt).



**Figure 1**

(a) Local density-correlation map for molecule *A* of mannose-binding protein, calculated from the MAD-phased experimental electron-density map (Burling *et al.*, 1996). The  $C^\alpha$  trace of the final refined model (PDB code 1ytt, chain *A*) is also shown. The correlation values for the atoms of this model range from  $-0.26$  to  $+0.93$ , with an average value of  $+0.65$  ( $\sigma$  0.24). (b) All-atom representation of the final refined model. The atoms have been colour coded depending on their average local density-correlation value: red, negative correlation; orange, between 0 and  $+0.25$ ; yellow, between  $+0.25$  and  $+0.5$ ; green, between  $+0.5$  and  $+0.75$ ; blue, correlation  $> +0.75$ . Figure created with *O* (Jones *et al.*, 1991).

Fig. 1(a) shows the LDC map calculated from the MAD-phased electron density, contoured at a level of  $+0.5$ . It also shows the  $C^\alpha$  trace of the *A* molecule of the final refined model, revealing that, in particular, the termini of the molecule 'stick out' of the correlation map, indicating that these may not obey the NCS. Fig. 1(b) shows an all-atom model of the same molecule, in which the atoms have been colour-coded based on their LDC values. Fig. 2(a) shows a combined plot of the  $C^\alpha-C^\alpha$  distance of corresponding residues in the two NCS-related molecules (after least-squares superpositioning of the two molecules; values have been cut off at 1.0 Å for clarity) and the average LDC value for the main-chain atoms of each residue. It is immediately obvious that troughs in the correlation density curve are associated with peaks in the  $C^\alpha-C^\alpha$  distance curve, supporting the basic assumption put forward in this paper. This relationship becomes even clearer in a scatter plot of the average LDC

values of the main-chain atoms *versus* the  $C^\alpha-C^\alpha$  distance (Fig. 2b). In this case, the structural differences between the two NCS-related molecules are clearly supported by the experimental data, since there are no regions where large structural differences are accompanied by high LDC values (which would indicate over-fitting or model errors).

Note that although there appears to be a linear relationship between the  $C^\alpha-C^\alpha$  distance and the average main-chain LDC (their linear correlation coefficient is  $-0.83$  for all 112 residues), this is unlikely to be the case. As the distance gets larger, the density correlation will become smaller and will eventually become approximately zero (*i.e.* uncorrelated density), no matter whether the distance is 5 or 50 Å. In this particular case, distances between 0 and 1 Å yield density-correlation values of 0.5–0.9, distances between 1 and 4 Å yield density-correlation values of 0.2–0.7 and distances exceeding 4 Å are accompanied by density-correlation values between  $-0.1$  and  $+0.1$ .

It could be argued that refined atomic temperature factors might provide similar information to the LDC method. However, there are several arguments against this supposition. In the case of mannose-binding protein, the average main-chain temperature factors display a much weaker correlation with the structural differences (correlation coefficient  $+0.48$ ) than do the average main-chain LDC values (correlation coefficient  $-0.83$ ; the correlation coefficient of the temperature factors and the LDC values is  $-0.55$ ). This implies that LDC values are more reliable indicators of genuine structural differences than temperature factors. Moreover, temperature

factors are a product of the experimental data, the refinement algorithm, the atomic model and any errors which the latter contains. LDC values, on the other hand, are derived from

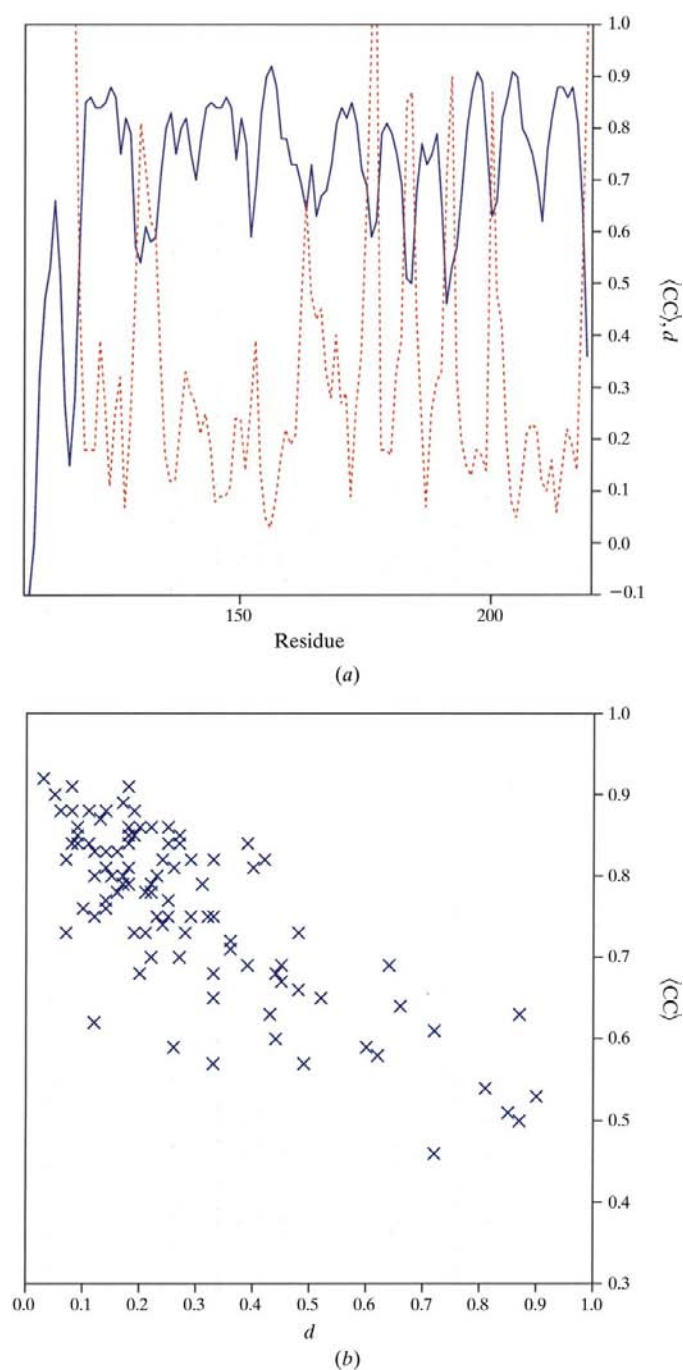
experimental information alone and are therefore to be preferred.

### 3.2. *Candida antarctica* lipase B

*C. antarctica* lipase B has been crystallized in several different crystal forms (Uppenberg, 1994; Uppenberg *et al.*, 1994, 1995). Two of these, for which experimental phase information was available, were used here to investigate whether observed structural differences between the final models coincide with regions of low local density correlation. The structure of a monoclinic crystal form was solved first, using MIR methods and molecular averaging. A crude model was used to solve an orthorhombic crystal form using molecular-replacement techniques, but this solution was only used to calculate difference Fourier maps to find the heavy-atom positions of the three derivatives (Uppenberg, 1994). Hence, two independent MIR maps were obtained. MIR density for molecule *A* of the monoclinic crystal form (PDB entry 1tcc; 2.5 Å resolution) and for the only molecule of the orthorhombic crystal form (PDB entry 1tca; 1.55 Å resolution) were used to calculate an LDC map. Fig. 3 shows the combined plots of  $C^\alpha-C^\alpha$  distance and average LDC value for the main-chain atoms of each residue, both as a function of residue number and as a scatter plot. As was the case for mannose-binding protein, there is an excellent agreement between poor local density correlation and relatively large structural differences, although there are a few discrepancies. In general, residues with low density correlation and small  $C^\alpha-C^\alpha$  distance are not a concern. The density could be poor in such regions, but the models are similar in the absence of experimental evidence to the contrary. However, residues with high density correlation which nevertheless display large structural differences should be treated with caution, since these differences contradict the experimental information. Again, the conclusion is that, *grosso modo*, the differences between the refined models reflect genuine dissimilarities between the structures that are supported by the experimental data. This case is an example of two closely related structures at rather different levels of resolution. In such cases, the danger of introducing artefactual differences in the low-resolution structure looms ominously (Kleywegt & Jones, 1995, 1997).

### 3.3. P2 myelin protein

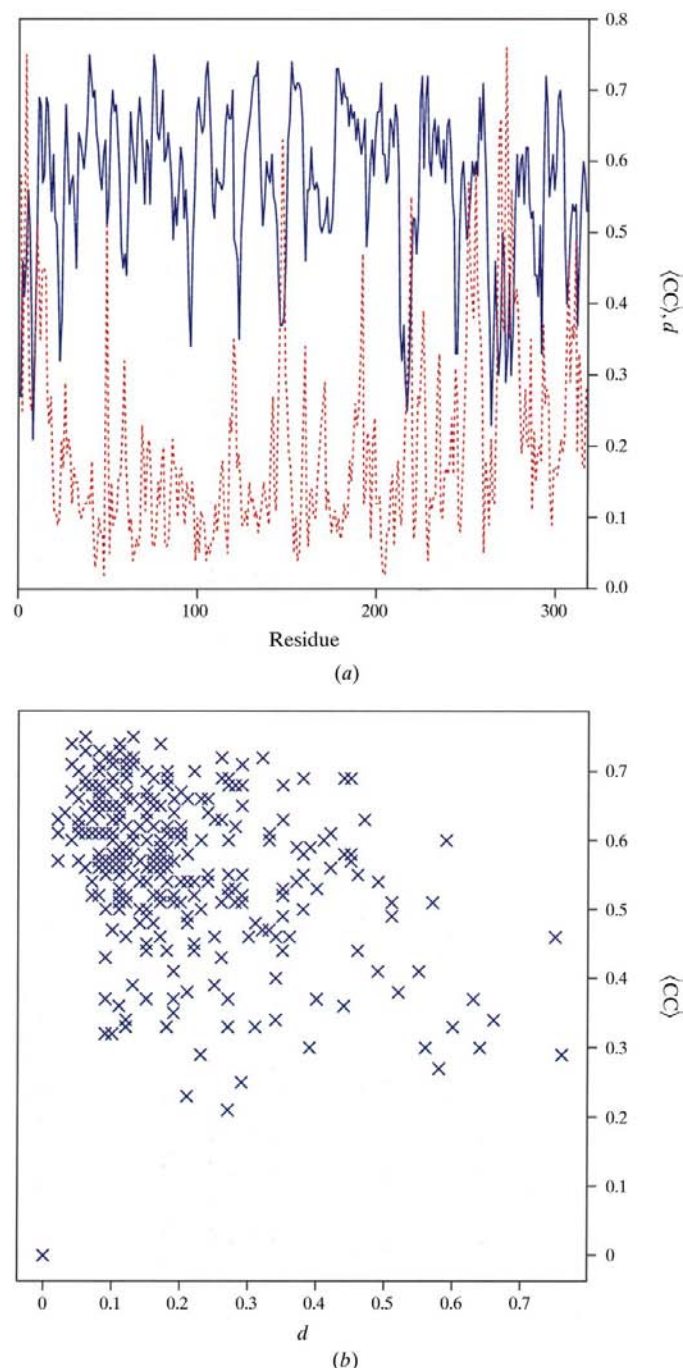
For P2 myelin protein, experimental MIR phases are available to a resolution of 2.7 Å (Jones *et al.*, 1988). The orthorhombic crystals contain three molecules per asymmetric unit, which have been refined using strict NCS constraints and, subsequently, tight NCS restraints (Cowan *et al.*, 1993). An  $r_2$ -type LDC map (Vellieux *et al.*, 1995) was calculated and the average LDC value for the main-chain atoms of each residue of the final model (PDB entry 1pmp, chain *A*) is shown in Fig. 4. Subsequently, the experimental map was subjected to six cycles of real-space averaging and a new map was calculated with the phases obtained after this procedure. A new  $r_2$ -type LDC map was calculated and the average correlation density value for the main-chain atoms of each residue of the



**Figure 2**

(a) The blue solid curve shows the average local density correlation for the main-chain atoms of all residues of the mannose-binding protein fragment; the red dashed curve shows the  $C^\alpha-C^\alpha$  distance between corresponding residues of the two NCS-related molecules after superpositioning (cut-off at 1.0 Å for clarity). (b) Scatter plot of the average local density correlation for the main-chain atoms of all residues of the mannose-binding protein fragment and the  $C^\alpha-C^\alpha$  distance between corresponding residues of the two NCS-related molecules after superpositioning. Residues for which the  $C^\alpha-C^\alpha$  distance exceeds 1.0 Å have been omitted for clarity.

final model in this map is also shown in Fig. 4. Comparison of both curves shows that the correlation curve has shifted upward considerably, but also that the trend in both curves is very similar (their correlation coefficient is +0.74). Hence,



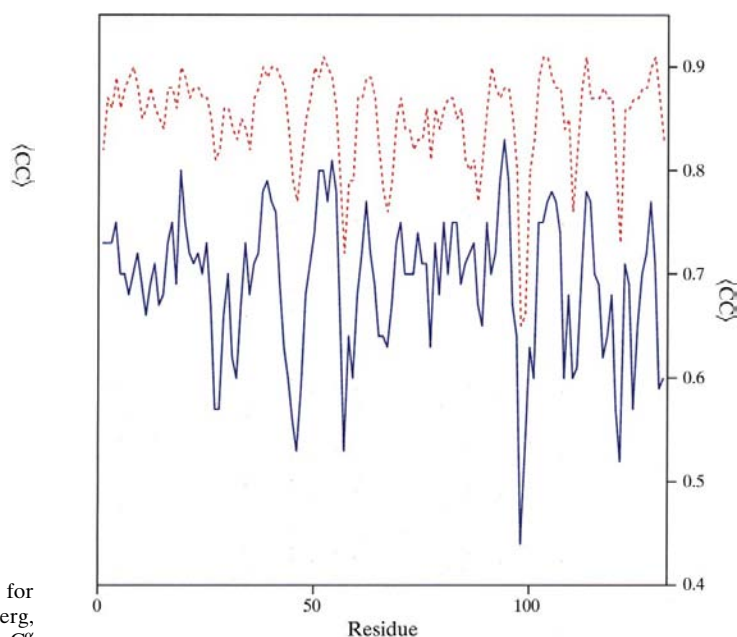
**Figure 3**  
(a) The blue solid curve shows the average local density correlation for the main-chain atoms of all residues of *C. antarctica* lipase B (Uppenberg, 1994; Uppenberg *et al.*, 1994); the red dashed curve shows the  $C^\alpha-C^\alpha$  distance between corresponding residues of two lipase molecules in different crystal forms after superpositioning. (b) Scatter plot of the average local density correlation for the main-chain atoms of all residues of the lipase and the  $C^\alpha-C^\alpha$  distance between corresponding residues of the two molecules after superpositioning (the correlation coefficient of the two is  $-0.42$  for 317 residues).

both curves could be used to validate the structural differences which might result from refinement without NCS constraints or could be used to design an appropriate NCS restraint scheme for such a refinement protocol. This observation supports the assumption that in the absence of purely experimental phase information, phase sets obtained after NCS-averaging may also be used (albeit with the caveat that the resulting LDC map will not be entirely unbiased).

#### 4. Concluding remarks

As the number of solved protein structures continues to increase, comparative structural biology (which attempts to correlate structural differences to variation in biological function, activity or affinity) will become increasingly important to deepen our understanding of the interplay between protein sequence, structure and function. However, the validity of conclusions drawn from any structural comparison is critically dependent on a proper assessment of the reliability of the individual structural models which are being compared and of the degree to which any observed differences are supported by experimental data.

It is shown here how experimental crystallographic information alone can be used to assess whether differences between two models of a protein (or of two related proteins) are likely to be significant or not. To my knowledge, this is the first such method which only uses experimental information and which is completely unbiased (by the model or the crystallographer). Hitherto, crystallographers have tended to equate large structural differences with significant ones, even



**Figure 4**  
The blue solid curve shows the average local density correlation for the main-chain atoms of all residues of P2 myelin protein, as calculated from the MIR map (Jones *et al.*, 1988). The red dashed curve shows the same values obtained from a map calculated with phases after six cycles of real-space electron-density averaging.

though it has been shown more than once that inappropriate refinement protocols are quite likely to be a major source of 'observed' structural dissimilarity. Hence, a method with which structural differences can be assessed without any bias is extremely useful, in particular if any biological relevance is to be attached to such differences.

It is shown that in cases where (relatively) unbiased phase information is available, local electron-density correlation maps contain information regarding the structural similarity of multiple copies of (chemically) similar or identical molecules. This knowledge can be employed to devise appropriate model refinement protocols, for instance by using local density-correlation values to define suitable weights for NCS restraints. Alternatively, structural differences obtained after model refinement can be validated by relating them to local density-correlation values. The method is simple yet powerful. It is also completely general and can therefore be used for macromolecules, ligands *etc.* The only requirements are the availability of one or more electron-density maps (with minimal or no model bias) and appropriate Cartesian rotation/translation operators which relate the various copies of the molecule or molecular envelope. The results of the assessment are conveniently presented in plots such as those shown in Fig. 2.

Since NCS occurs in roughly half of all low-resolution protein crystallographic studies (Kleywegt, 1996), the present method will probably be applied mostly for such cases. The number of cases where experimental phase information is available for two or more related structures (complexes, mutants *etc.*) is small at present. However, if one encountered a case with surprisingly large structural differences, the present method offers a route to experimental validation of these differences. A conscientious crystallographer may attempt to obtain (possibly crude) experimental phases for the possibly aberrant structure. Finally, in cases where related or identical structures are determined independently in different laboratories, their comparison and the analysis of their differences will greatly benefit from application of the present method. In general, of course, the importance of validation methods such as the one described here increases as the resolution of the study decreases.

Although the present method is powerful, it critically depends on the availability of experimental information. The entire structural biology community would therefore benefit from routine data bank deposition of experimental crystallographic information, including experimental phases.

### 5. Availability

Manuals for *COMA*, *MASKIT*, *MAPMAN* and *LSQMAN* are available on the WWW (<http://alpha2.bmc.uu.se/usf/>). All these programs are available free of charge to the academic and not-for-profit community (<ftp://alpha2.bmc.uu.se/pub/gerard/>); commercial users may contact the author (mailto:gerard@xray.bmc.uu.se) for licensing details. An illustrated

tutorial, showing how to carry out the calculations and how to produce pictures like those shown here, is available on the WWW at URL [http://alpha2.bmc.uu.se/usf/dens\\_corr.html](http://alpha2.bmc.uu.se/usf/dens_corr.html).

The author wishes to thank Dr Jonas Uppenberg for providing the MIR maps of *C. antarctica* lipase B and Professor T. Alwyn Jones for providing the MIR data of P2 myelin protein as well as for commenting on this manuscript. Constructive comments from an anonymous referee are gratefully acknowledged. This work was supported by the Swedish Foundation for Strategic Research and its Structural Biology Network (SBNet).

### References

- Abagyan, R. A. & Totrov, M. M. (1997). *J. Mol. Biol.* **268**, 678–685.
- Bernstein, F. C., Koetzle, T. F., Williams, 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.
- Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L. & Sigler, P. B. (1994). *Nature (London)*, **371**, 578–586.
- Bricogne, G. (1974). *Acta Cryst.* **A30**, 395–405.
- Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
- Brünger, A. T. (1993). *Acta Cryst.* **D49**, 24–36.
- Brunger, 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.* **D54**, 905–921.
- Burling, F. T., Weiss, W. I., Flaherty, K. M. & Brünger, A. T. (1996). *Science*, **271**, 72–77.
- Cowan, S. W., Newcomer, M. E. & Jones, T. A. (1993). *J. Mol. Biol.* **230**, 1225–1246.
- Hodel, A., Kim, S.-H. & Brünger, A. T. (1992). *Acta Cryst.* **A48**, 851–858.
- Jones, T. A., Bergfors, T., Sedzik, J. & Unge, T. (1988). *EMBO J.* **7**, 1597–1604.
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* **A47**, 110–119.
- Kleywegt, G. J. (1996). *Acta Cryst.* **D52**, 842–857.
- Kleywegt, G. J. & Brünger, A. T. (1996). *Structure*, **4**, 897–904.
- Kleywegt, G. J., Hoier, H. & Jones, T. A. (1996). *Acta Cryst.* **D52**, 858–863.
- Kleywegt, G. J. & Jones, T. A. (1995). *Structure*, **3**, 535–540.
- Kleywegt, G. J. & Jones, T. A. (1996). *Acta Cryst.* **D52**, 826–828.
- Kleywegt, G. J. & Jones, T. A. (1997). *Methods Enzymol.* **277**, 208–230.
- Kleywegt, G. J. & Jones, T. A. (1999). *Acta Cryst.* **D55**, 941–944.
- Kleywegt, G. J. & Read, R. J. (1997). *Structure*, **5**, 1557–1569.
- Korn, A. P. & Rose, D. R. (1994). *Protein Eng.* **7**, 961–967.
- Read, R. J. (1986). *Acta Cryst.* **A42**, 140–149.
- Stroud, R. M. & Fauman, E. B. (1995). *Protein Sci.* **4**, 2392–2404.
- Terwilliger, T. C. & Berendzen, J. (1995). *Acta Cryst.* **D51**, 609–618.
- Terwilliger, T. C. & Berendzen, J. (1996). *Acta Cryst.* **D52**, 1004–1011.
- Uppenberg, J. (1994). PhD thesis, Uppsala University, Sweden.
- Uppenberg, J., Öhrner, N., Norin, M., Hult, K., Kleywegt, G. J., Patkar, S., Waagen, V., Anthonson, T. & Jones, T. A. (1995). *Biochemistry*, **34**, 16838–16851.
- Uppenberg, J., Trier Hansen, M., Patkar, S. & Jones, T. A. (1994). *Structure*, **2**, 293–308.
- Vellieux, F. M. D. A. P., Hunt, J. F., Roy, S. & Read, R. J. (1995). *J. Appl. Cryst.* **28**, 347–351.
- Vellieux, F. M. D. & Read, R. J. (1997). *Methods Enzymol.* **277**, 18–53.