Acta Cryst. (1997). D**53**, 448–455

wARP: Improvement and Extension of Crystallographic Phases by Weighted Averaging of Multiple-Refined Dummy Atomic Models

ANASTASSIS PERRAKIS,[a] TITIA K. SIXMA,[a] KEITH S. WILSON[b] AND VICTOR S. LAMZIN[c]

[a]*Netherlands Cancer Institute (NKI), Department of Molecular Carcinogenesis, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands,* [b]*Protein Structure Group, Department of Chemistry, University of York, Heslington, York YO1 5DD, England, and* [c]*European Molecular Biology Laboratory (EMBL) Hamburg, c/o DESY, Notkestrasse 85, 22603 Hamburg, Germany.*
E-mail: perrakis@nki.nl

(*Received 27 January 1997; accepted 11 April 1997*)

Abstract

wARP is a procedure that substantially improves crystallographic phases (and subsequently electron-density maps) as an additional step after density-modification methods such as solvent flattening and averaging. The initial phase set is used to create a number of dummy atom models which are subjected to least-squares or maximum-likelihood refinement and iterative model updating in an automated refinement procedure (*ARP*). Averaging of the phase sets calculated from the refined output models and weighting of structure factors by their similarity to an average vector results in a phase set that improves and extends the initial phases substantially. An important requirement is that the native data have a maximum resolution beyond ~2.4 Å. The *wARP* procedure shortens the time-consuming step of model building in crystallographic structure determination and helps to prevent the introduction of errors.

1. Introduction

A major time-consuming and critical step in protein crystallography is the building of a molecular model in the initial electron density. Decisions about connectivity of the polypeptide chain and matching of the known amino-acid sequence to the electron density (chain tracing) are made at this stage. Although automatic approaches are being developed for this purpose (*e.g.* by Zou & Jones, 1996), human intervention is still required as a decisive step. Because the initial phases are often not very accurate, tracing the polypeptide chain can be very difficult. Over the years a number of density modification techniques have been developed to improve starting phases without introducing model bias, *e.g.* solvent flattening (Wang, 1985), histogram matching (Zhang & Main, 1990), density skeletonization (Baker, Bystroff, Fletterick & Agard, 1993), non-crystallographic symmetry averaging (Rossmann & Blow, 1963; Bricogne, 1974) and phase extension (Bricogne, 1984; Sayre, 1972). Here we describe a method, *wARP* (Okuda, Okuda & Mirek, 1992), which stands for 'weighted *ARP*', that is applied to the best initial map after density-modification procedures and substantially improves the phases.

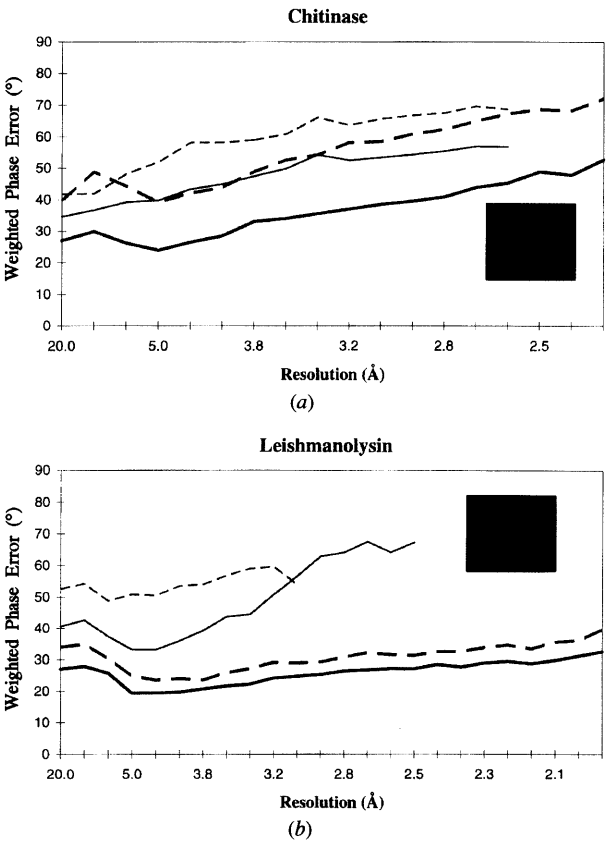

Fig. 1. Weighted phase difference to the phases calculated from the final model in resolution shells is shown for phase sets from MIR, optimal solvent flattening, the best single *ARP* run and the *wARP* phase combination, for chitinase (top) and leishmanolysin (bottom). Map correlation coefficient of the final model with electron-density maps resulting from these data sets were: (*a*) for chitinase: 46.0, 68.4, 70.4 and 81.2% for MIR, *DM*, *ARP* and *wARP* maps, respectively. (*b*) For leishmanolysin: 42.7, 66.6, 88.3 and 92.0% for MIR, *DM*, *ARP* and *wARP* maps, respectively.

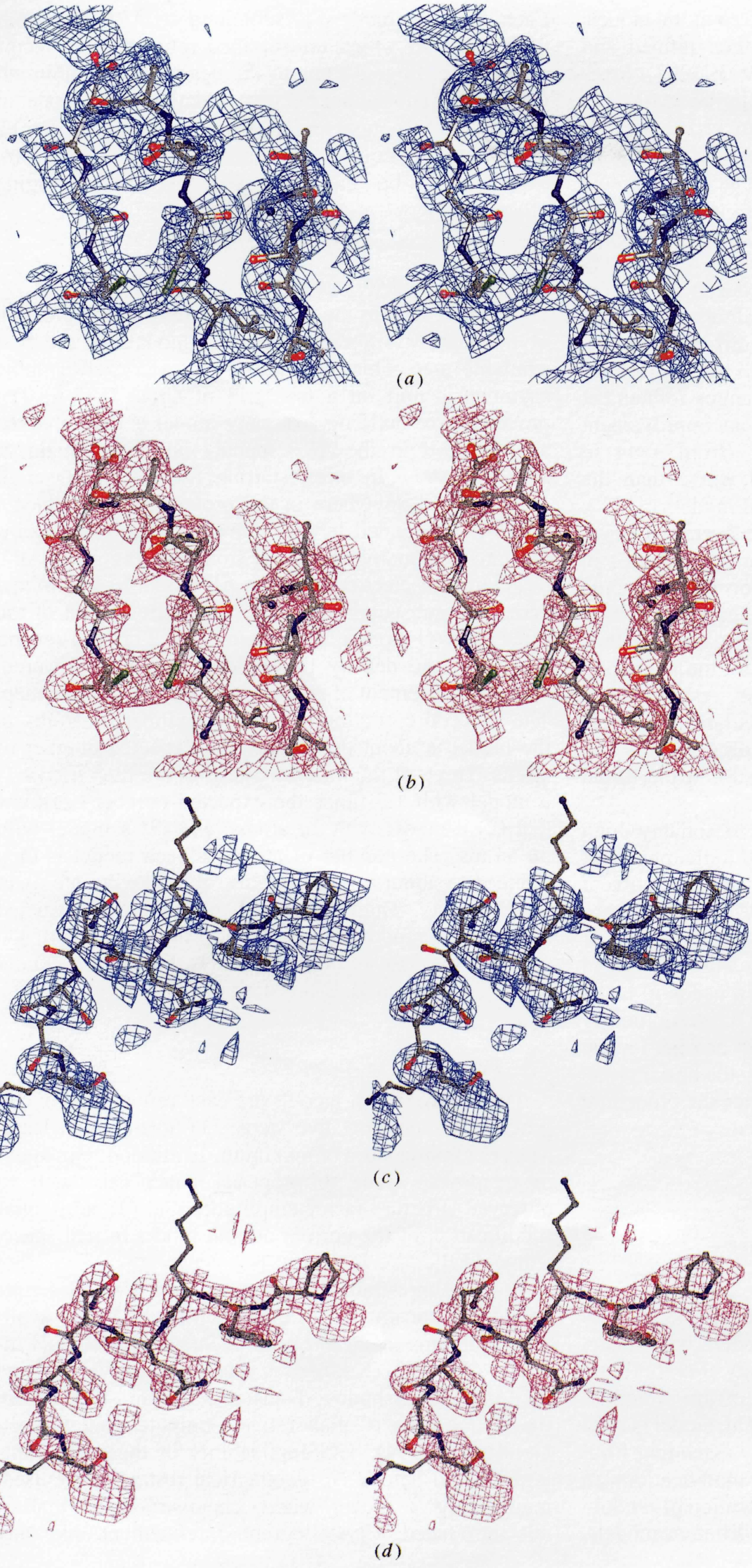

Fig. 2. Representative regions of the solvent flattened (*a*, *c*) and equivalent *wARP*-averaged maps (*b*, *d*) for chitinase A, shown in stereo. Drawn by *O/oplot* (Jones, Zou, Cowan & Kjeldgaard, 1991).

The *wARP* procedure is based on free-atom models built in the best available map, and then refined and updated by *ARP* (Lamzin & Wilson, 1993, 1997). Free-atom refinement has been applied for similar reasons in the past (Agarwal & Isaacs, 1977), but had the major drawback of absence of automatic model rebuilding, as supplied by *ARP*. The structure factors calculated from a set of such models, with small differences to one another, are averaged and proper weights are assigned to each reflection. The idea of averaging has been extensively utilized in protein crystallography in order to reduce the noise introduced by experimental or algorithmic inaccuracies. Structure-factor averaging of different models, without any weighting scheme, has been tried before for improving results from molecular dynamics refinement (Rice & Brünger, 1994) and showed an improvement compared to simple phase averaging (from a 44 to a 41.8° phase error), which was still worse than the phase error of one of the four individual models (41.7°). In *wARP*, the structure factors to be averaged, have uncorrelated errors caused by different inaccuracies in the models. However, these errors are correlated to some extent because all models are refined against the same data, thus limiting the effect of averaging. Despite this, in the weighted-average structure factors the averaging procedure effectively reduces the noise, resulting in a much improved phase set. A map calculated from this phase set will speed up the model-building process and helps to prevent the introduction of errors in the initial protein model.

Requirements for the method are reasonably high resolution native data (at least 2.4 Å) and initial phases of such quality that the contrast between protein and solvent is well defined and some elements of protein stereochemistry are recognizable. Since any measurement of initial map quality is subjective, we cannot define in absolute terms how good the initial map needs to be. It must be noted, however, that the higher the resolution of the native data the worse the initial map can be, for the method to work. The resolution to which the initial phase set extends is relatively unimportant, since the procedure itself provides excellent phase extension.

2. The *wARP* procedure

2.1. *Outline of the method*

In the *wARP* procedure the best available map is used to create a dummy atomic model, with equal atoms placed in regions of high density. For building this model, only the molecular weight of the protein is required and no sequence information. The model is built automatically by the program, gradually extending from a small random set of atoms. Slight modifications in the model-generation procedure or application of random shifts are used to generate six or more different models. Each of these models is submitted to *ARP* free-atom refinement, in which unrestrained refinement in reciprocal space moves atoms to the nearest local minimum followed by substantial iterative updating of the model in real space. Structure factors from these models can be presumed to contain different errors, which are expected to be cancelled out by the averaging procedure.

2.2. *Making dummy models*

The first step in the *wARP* procedure is the creation of moderately different free atom models in the best available map. This map must cover a crystallographic asymmetric unit on a fine grid of about 0.25 Å. The procedure for building a dummy model is then invoked as described in the *ARP* manual and by Lamzin & Wilson (1997). In brief, starting from a small set of atoms placed anywhere in the protein region, a model is slowly expanded by the stepwise addition of atoms that are at bonding distances from existing atoms and in significant electron density. All atoms in this model and in all subsequent steps are considered to be of the same type (O atoms). While geometrical criteria remain the same, the density threshold is gradually lowered, allowing placement of atoms in weaker areas of the map. The procedure continues until the number of atoms in the model is about three times the expected number of atoms. Three of the intermediate files are then used: (1) a model with 1.5 times the expected number of atoms (n); (2) a model with $2n$ atoms; and (3) a model with $3n$ atoms. The number of atoms in each model is then reduced to about $n + 20\%$ atoms, by removing atoms in weak density. Three additional models are constructed by applying random small shifts to the positions of the atoms in the first set of three models. For all subsequent steps these six models are used.

2.3. *Refinement of the models*

The six free-atom models are each refined in a cyclic procedure alternating two steps: (1) unrestrained least-squares minimization or maximum-likelihood refinement in reciprocal space to properly match calculated to observed structure-factor amplitudes and (2) substantial modification of the current atomic model in real space, using *ARP*.

For the unrestrained refinement step, C-shell scripts have been constructed to employ most currently available programs in the procedure. Standard protocols include *PROLSQ* (Konnert & Hendrickson, 1980) and *REFMAC* (Murshudov, Dodson & Vagin, 1996) from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). The applicability of these programs is discussed below. No geometrical restraints are used, maintaining a model where chemistry-based bias is not introduced. Crystallographic refinement, like any

minimization technique, requires an overdetermined system for convergence. The number of experimental observations (reflections measured) must be higher than the number of parameters describing the model (x, y, z, B per atom). The ratio of experimental observations to model parameters in crystallographic refinement becomes greater as the resolution of diffraction data increases. However, for different crystals, this ratio is not only determined by resolution. It also depends on the solvent content of the unit cell and the completeness of the data. Requirements for *wARP* are discussed in more detail below.

ARP updates the model after each refinement step, mimicking human intervention between refinement cycles. It removes atoms based on the density in the $3F_o - 2F_c$ Fourier synthesis, shape criteria (sphericity) and distance criteria (if atoms come too close to or too far from existing atoms). It adds atoms in significant density (the threshold is estimated automatically by the program) in the $F_o - F_c$ Fourier synthesis, provided that they are bonded to existing atoms. Real-space refinement is carried out to optimize atomic positions, before step (1) is iterated.

2.4. *wARP averaging*

In protein crystallography there are generally insufficient data for convergence of free-atom refinement to a global minimum. Thus, slightly different starting models are expected to result in final models with small differences, *i.e.* containing different errors. Averaging between these models can then be utilized to minimize the overall error.

Structure factors are calculated for all models after refinement and scaled to the observed amplitudes. A vector average of the calculated structure factors from the different refined models is then calculated. The phase of the vector average is more accurate than those calculated from any of the individual models. Subsequently, a weighting scheme is applied to enhance the overall quality of phases. A weight, $w_{\rm warp}$, is assigned to each structure factor on the basis of the variance of the two-dimensional distribution of the individual structure factors around the average. For any reflection, let $F_{\rm obs}$ be the observed structure-factor amplitude, $\mathbf{F}_{\rm aver}$ the vector average structure factor, $\mathbf{F}_i$ the vectors used to construct the average and n the number of individual models. Then the corresponding weight is,

$$w_{\rm wARP} = F_{\rm obs}^2/(F_{\rm obs}^2 + \sum |\mathbf{F}_{\rm aver} - \mathbf{F}_i|^2/n)$$

The mean value of $w_{\rm wARP}$ over all reflections and the R factor after averaging can be used to judge the progress of the averaging procedure. For a succesfull use of the method, the mean weight must be above 0.6 and the combined R factor must drop by 5–10% with respect to the R factors for individual models, typically within values ranging from 12 to 16%.

3. Examples

Two examples are presented here, chitinase A (2.3 Å) and *Leishmania* virus coat protein (2.0 Å). Both structures were originally solved by isomorphous replacement. Since the models for these proteins have been refined to high resolution, detailed statistics and comparisons can be given. *wARP* has also been applied to four other projects (three MIR and one molecular replacement) that are still under refinement. Each of these resulted in substantial improvement of the electron-density maps (Harry Tong, Kristina Djinovic, Matti Saraste, Erik van Asselt & Bauke Dijkstra, personal communication).

3.1. *Assessment of map quality*

The following figures are used throughout to characterize the quality of the phase sets or density maps. (1) The weighted mean phase error to the phases calculated from the final model. Weights were derived from averaging as described above or from isomorphous replacement. (2) Correlation coefficients for the density maps computed with or without weights, according to Lunin & Woolfson (1993). (3) Real-space correlation coefficients of the final model to the various maps, as described by Brändén & Jones (1990).

3.2. *Chitinase A*

The structure of chitinase A from *Serratia marcescens* (ChiA) (Perrakis *et al.*, 1994) was initially solved by multiple isomorphous replacement with anomalous signal (MIRAS); with only one derivative contributing to resolution higher than 5.0 Å. The MIRAS map (2.5 Å) was solvent flattened using the *PHASES* package (Furey & Swaminathan, 1990). Model building was not straightforward and much time was spent in tracing the protein chain.

In the *wARP* procedure the solvent-flattened map was used to initiate building of dummy models. *PROLSQ* least-squares minimization against the native 2.3 Å data was used with *ARP*. Refinement of the models resulted in crystallographic R factors ranging between 20.1 and 22.4%. Each of the *ARP* models gave phases marginally worse than the phases already available by solvent flattening, because of the limited resolution of the native data, Fig. 1(*a*). However, the *wARP* average procedure resulted in a reduction of 11.2° in the weighted mean phase error. The map correlation coefficient between the final map and the *wARP* map was 81.2%, better by 12.8% than for the solvent-flattened map. The weighted phase error in resolution shells for all phase sets is shown in Fig. 1(*a*).

Visual inspection of the maps shows great improvement, especially near the protein surface where solvent flattening may have introduced errors caused by slightly incorrect masks, Figs. 2(*a*) and 2(*b*), but also in the protein interior where it resolved a number of ambiguities

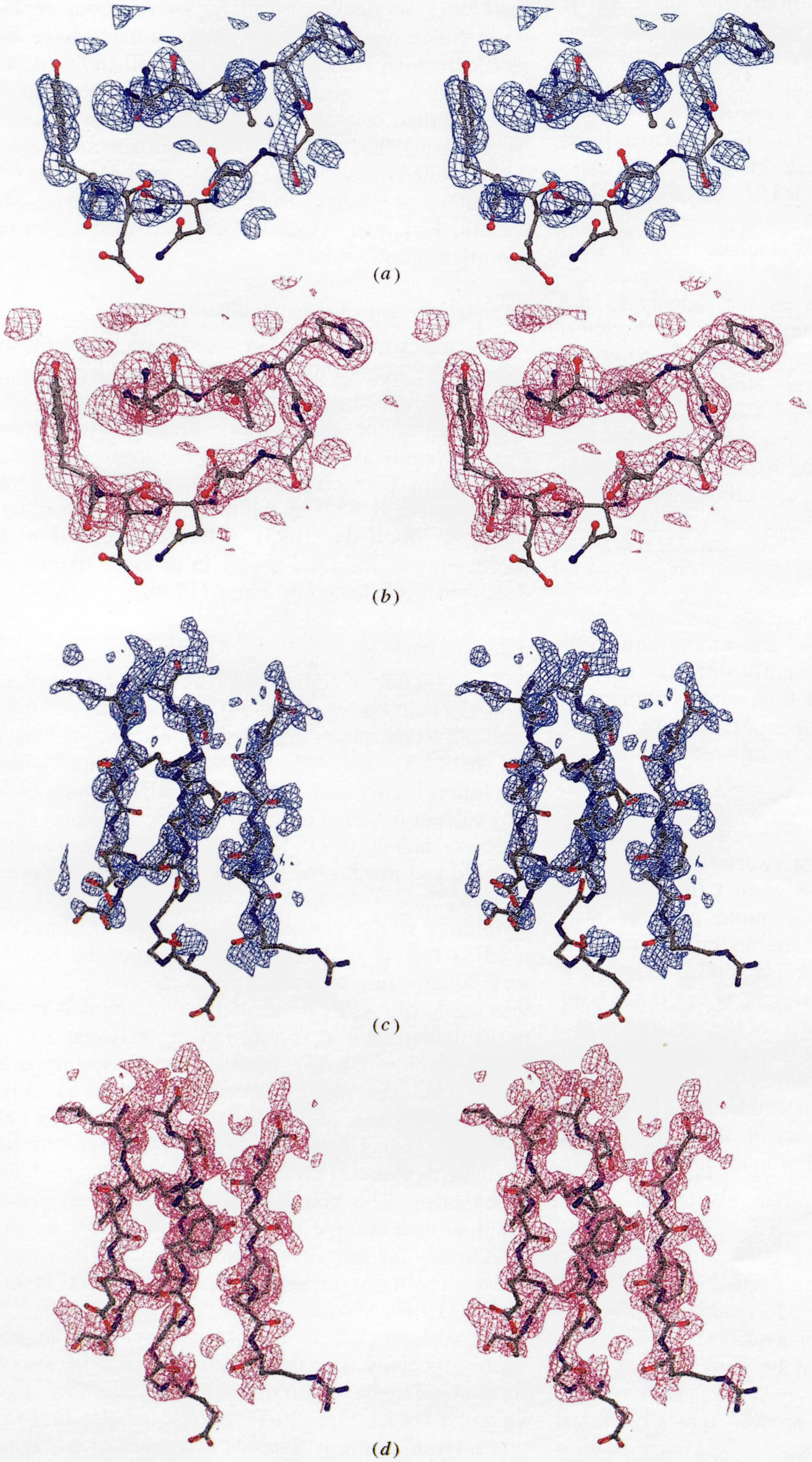

Fig. 3. Representative regions of the solvent flattened (*a*, *c*) and *wARP*-averaged maps (*b*, *d*) for leishmanolysin, shown in stereo. Drawn by *O/oplot* (Jones *et al.*, 1991).

in chain tracing, Figs. 2(*c*) and 2(*d*). In a number of places where the solvent flattened and *wARP* maps are equally continuous, details for side chain and carbonyl O-atom placement were clearly improved. Only in a few places were errors present in the solvent flattened map maintained in the *wARP* map – as judged both by visual inspection and by real space correlation coefficients (data not shown). We could not identify any regions of the map where *wARP* had introduced erroneous features.

3.3. *Leishmanolysin*

The structure of the *Leishmania* coat protein (leishmanolysin, PSP) was solved with a complicated

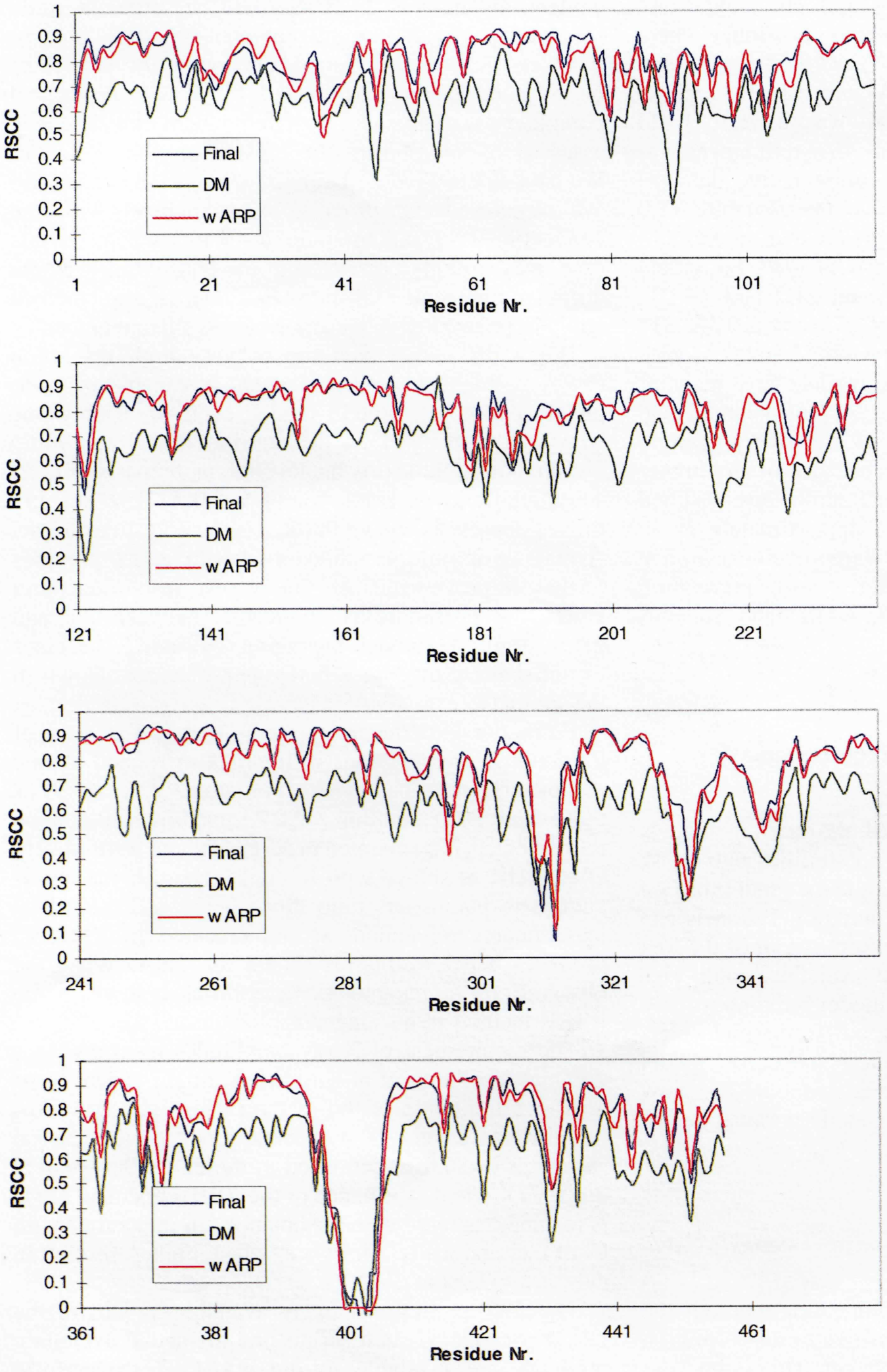

Fig. 4. Residue based real-space map correlation (RSCC) to the solvent-flattened, *wARP* and final maps for leishmanolysin.

protocol involving the use of SIRAS phases for two different crystal forms, averaging between those, solvent flattening and density skeletonization (unpublished data kindly provided by Dr Peter Metcalf). For the *wARP* test one set of SIRAS phases was used, which extends to a resolution of 3.0 Å. These phases were determined for the first crystal form for which native data extending to 2.5 Å were used for solvent flattening and phase extension with the *DM* program (Cowtan, 1994) in *CCP*4. The solvent-flattening and phase-extension procedure improved phases to the resolution range where phases where available (20–3.0 Å) from 54 to 41° and resulted in new phases (3.0–2.5 Å) with 64° mean phase error. This density-modified map was used to build the initial models for *wARP*. The *ARP* refinement was performed against a higher resolution native data set than that originally used to solve the structure (2.0 instead of 2.5 Å). *REFMAC* unrestrained maximum-likelihood minimization was used with *ARP*. Individual models refined to *R* factors between 21.2 and 19.1%, with approximate phase errors of 27° (20–3.0 Å), 31° (3.0–2.5 Å) and resulted in new phases (2.5–2.0 Å) with a phase error of 35°. All of these models gave maps of dramatically better quality than the solvent-flattened map Figs. 3(*a*)–(*d*) . Here the power of the *ARP* procedure itself is larger than for chitinase because of the higher resolution of the data. The *wARP* procedure resulted in an additional improvement of approximately 5° in weighted phase difference or 3.7% improvement in map correlation, Fig. 1(*b*). In Fig. 4 the real-space correlation of the final model to the final, *wARP* and solvent-flattened (*DM*) maps is shown.

4. Applicability and requirements

4.1. *Variation of starting models*

The choice of the variation of the starting models is somewhat arbitrary and other possibilities may work better in particular cases. For example the randomization of atomic positions for the three additional models can be performed after some cycles of refinement of the first three models. A partial model, if available, could also be used to initiate the 'dummy' model building.

When the procedure is started from molecular-replacement solutions, other alternatives for model building become possible. The search model can now be used as a start model after automatic removal and/or addition of a number of atoms with a procedure similar to the one described for MIR maps.

4.2. *Use of different refinement methods and resolution requirements*

In contrast to most density-modification methods the *wARP* procedure is extremely sensitive to the resolution of observed data in the native data set. This is because of the limitations of the unrestrained refinement step and the real-space update of the model by *ARP*.

Below 2.5 Å resolution individual atoms are not resolved, thus making real-space update by *ARP* essentially impossible. The real limitation however, lies in the unrestrained refinement step. That cannot be expressed solely in resolution terms, but as observations/parameters ratio, which is largely dependent on solvent content. Thus, for a crystal with high solvent content, about 70%, 2.5 Å data will be sufficient while for a crystal with low solvent content, 40%, data to 2.0 Å resolution are required. Obviously the data must be of good quality, as judged by $R_{\rm merge}$, $I/\sigma(I)$ and completeness. The success of refinement can be easily assessed by monitoring the crystallographic *R* factor. We have not used $R_{\rm free}$ (Brünger, 1993) in the examples. We monitored the progress of refinement with the conventional crystallographic *R* factor which, in this case, was strongly correlated to the phase improvement during refinement. However, $R_{\rm free}$ use is implemented and can be used to judge the progress of refinement.

When the native data are of very high resolution (1.5 Å or higher), *ARP* refinement is capable of completely building a protein model, even when only one heavy atom is used as the starting model (*ab initio* phasing of rubredoxin, manuscript in preparation). At substantially lower resolution (1.5–2.0 Å) *ARP* still improves phases, as shown in the leishmanolysin example. Averaging of multiple refined models by *wARP* provides further improvement. At the lowest resolution limit where the procedure is applicable, between 2.0 and 2.5 Å, multiple model averaging by *wARP* becomes essential for providing a better phase set, as shown in the chitinase example.

From our experience, it seems that if the ratio of the number of reflections in the data to refined atomic parameters (four parameters per atom, *x*, *y*, *z*, *B*) is more than 2.0 (resolution ~2.0 Å) maximum-likelihood refinement as implemented in *REFMAC* can be used very effectively, as shown with leishmanolysin. If the observations to parameters ratio drops below 2.0 traditional least-squares refinement as implemented in *PROLSQ* produces better results, as shown for ChiA. When the observations to parameters ratio drops below 1.5 the *wARP* method is no longer applicable.

The requirement of X-ray amplitudes measured to a nominal resolution of at least 2.5 Å, might appear to be a serious limitation of the method. However, this resolution range covers the vast majority of protein crystal structures (~80.0%) deposited in the PDB (Bernstein *et al.*, 1977). The performance of the *ARP* refinement alone is (as demonstrated with leishmanolysin in contrast with ChiA) dramatically improved when higher resolution data are available.

The *ARP* program is freely available as part of the *CCP*4 package. C-shell scripts and the actual averaging program, are available to run *wARP*. They perform

the dummy model building, *ARP* refinements and final averaging in an automated manner. They are also capable to split jobs in a 'parallel' manner to different processors which can be located in different computers over a network, thus minimizing the actual required run time to the one needed for a single *ARP* job – provided that enough processors are available. The scripts are tested on several Irix 5.3 based clusters, but should be straightforward to adapt for usage with any Unix-based system.

A WWW *ARP/wARP* home page is now available, at http://den.nki.nl/~perrakis/arp.html from where the complete *ARP/wARP* package can be obtained. A mailing list is also open for questions and discussion for *ARP/wARP* usage. To subscribe, simply go through the WWW page or send a mail with one line 'subscribe arp-users' to majordomo@linde.nki.nl.

AP is an EMBO fellow (ALTF-215 1995). We would like to thank Zbyszek Dauter for his helpful suggestions and Peter Metcalf for providing unpublished data.

References

Agarwal, R. C. & Isaacs, G. (1977). *Proc. Natl Acad. Sci. USA*, **74**, 2835–2839.

Baker, D., Bystroff, C., Fletterick, R. J. & Agard, A. D. (1993). *Acta Cryst.* D**49**, 429–439.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Roger, J. B., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Brändén, C.-I. & Jones, T. A. (1990). *Nature (London)*, **343**, 687–689.

Bricogne, G. (1974). *Acta Cryst.* A**30**, 395–405.

Bricogne, G. (1984). *Acta Cryst.* A**40**, 410–445.

Brünger, A. T. (1993). *Acta Cryst.* D**49**, 24–36.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Cowtan, K. (1994) *Jnt CCP4 and ESF-EACBM Newslett. Protein Crystallogr.* **31**, 34–38.

Furey, W. & Swaminathan, S. (1990). Am. Crystallogr. Assoc. Meet. Abstr. **18**, 73.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991) *Acta Cryst.* A**47**, 110–119.

Konnert, J. H. & Hendrickson, W. A. (1980). *Acta Cryst.* A**36**, 344–350.

Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.

Lamzin, V. S. & Wilson, K. S. (1997). *Methods Enzymol.* In the press.

Lunin, V. Y. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 530–533.

Murshudov, G. N., Dodson, E. S. & Vagin, A. A. (1996). In *Macromolecular Refinement. Proceedings of the CCP4 Study Weekend*, pp. 93–104. Warrington: Daresbury Laboratory.

Okuda, M., Okuda, D. & Mirek, D. (1992). In *The Star Trek Encyclopedia, a Guide to the Future*, pp. 371–372. New York: Pocket Books.

Perrakis, A., Tews, I., Dauter, Z. Oppenheim, A., Chet, I., Wilson, K. S. & Vorgias, C. E. (1994). *Structure*, **2**, 1169–1180.

Rice & Brünger, A. (1994). *Proteins Struct. Funct. Genet.* **19**, 277–290.

Rossmann, M. G. & Blow, D. M. (1963). *Acta Cryst.* **16**, 39–45.

Sayre, D. (1972). *Acta Cryst.* A**28**, 210–212.

Wang, B. C. (1985). *Methods Enzymol.* **115**, 90–112.

Zhang, K. Y. J. & Main, P (1990). *Acta Cryst.* A**46**, 41–46.

Zou, J.-Y. & Jones A. (1996). *Acta Cryst.* D**52**, 833–841