Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# Multivariate phase combination improves automated crystallographic model building

**Pavol Skubák,*‡ Willem-Jan Waterreus‡ and Navraj S. Pannu***

Biophysical Structural Chemistry, Leiden University, PO Box 9502, 2300 RA Leiden, The Netherlands

‡ These authors contributed equally to this work.

Correspondence e-mail: p.skubak@chem.leidenuniv.nl, raj@chem.leidenuniv.nl

Received 9 February 2010
Accepted 21 April 2010

Density modification is a standard technique in macromolecular crystallography that can significantly improve an initial electron-density map. To obtain optimal results, the initial and density-modified map are combined. Current methods assume that these two maps are independent and propagate the initial map information and its accuracy indirectly through previously determined coefficients. A multivariate equation has been derived that no longer assumes independence between the initial and density-modified map, considers the observed diffraction data directly and refines the errors that can occur in a single-wavelength anomalous diffraction experiment. The equation has been implemented and tested on over 100 real data sets. The results are dramatic: the method provides significantly improved maps over the current state of the art and leads to many more structures being built automatically.

## 1. Introduction

The single-wavelength anomalous diffraction (SAD) experiment is a widely used technique to provide an estimate of the unknown structure-factor phases and thus experimentally solve a macromolecular structure. To improve an electron-density map obtained by single-wavelength anomalous diffraction or another experimental method providing initial phase estimates, crystallographers employ density-modification techniques. Density modification (DM) relies on prior information. For example, if the experimentally determined map is of sufficient quality to distinguish between regions of ordered macromolecule and disordered solvent, the electron density of the solvent region can be flattened (Wang, 1985; Leslie, 1987). This density-modified map is then combined with the initial density map and the process is iterated, as illustrated in Fig. 1.

'Phase combination' or combining the original and density-modified electron-density map is important for the success of DM. Many currently used density-modification programs [*e.g.* *DM* (Cowtan, 1994), *SOLOMON* (Abrahams & Leslie, 1996), *SHELXE* (Sheldrick, 2008) and *CNS* (Brünger *et al.*, 1998)] employ the *SIGMAA* (Lunin & Urzhumtsev, 1984; Read, 1986) algorithm. In the *SIGMAA* algorithm, the original experimentally determined phases, represented by Hendrickson–Lattman coefficients (Hendrickson & Lattman, 1970), are assumed to be independent of the density-modified structure factors and are combined with them through a heuristic weighting scheme (Read, 1997). However, the density-modified map is obtained from the experimental map, invalidating this assumption of independence. Methods have been developed to reduce the correlation between the original

and the density-modified structure factors *via* the $\gamma$-correction (Abrahams & Leslie, 1996; Abrahams, 1997) or 'statistical density modification' (Terwilliger, 2000; Cowtan, 2000). However, the correlation is not considered explicitly and still remains.

Most density-modification programs obtain prior phase information in the form of Hendrickson–Lattman coefficients. When these coefficients are used, it is assumed that errors in structure factors can be represented by a one-dimensional probability distribution over phases. Furthermore, any computer program that inputs Hendrickson–Lattman coefficients assumes that the program that generated them can provide a reliable and accurate estimate of the phase probability.

We have developed a multivariate probability distribution that no longer assumes independence between the initial and combined electron-density map, but considers the correlation between the observed structure-factor amplitudes, the density-modified structure factor and a heavy-atom substructure for a single-wavelength anomalous diffraction experiment. The equation generates phase information directly from the heavy-atom substructure and the observed diffraction data and thus does not require input of Hendrickson–Lattman coefficients. The equation also models and refines the errors in the density-modified and heavy-atom structure factors to provide an advanced multivariate model for a single-wavelength anomalous diffraction experiment and phase combination. We have implemented this 'SAD-DM' function in a new program called *MULTICOMB*. We were motivated to consider these methods for phase combination since our previous work in applying similar multivariate methods to experimental phasing (Pannu & Read, 2004; Ness *et al.*, 2004) and in model building with iterative refinement (Skubák *et al.*, 2004, 2005) produced better results compared with previous approaches.

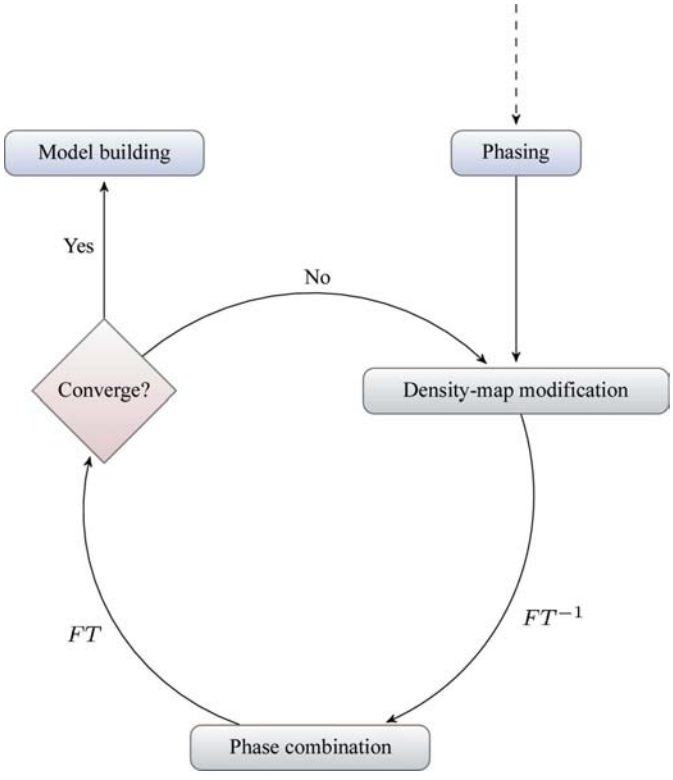


**Figure 1**
Flow chart of the iterative procedure of density modification.

## 2. Methods

The multivariate SAD-DM function used for correlated DM phase combination implemented in *MULTICOMB* is the following:

$$\begin{aligned}
P_{\text{SAD-DM}} &= P(|F^+|, |F^-|; |F_{\text{DM}}|, \alpha_{\text{DM}}, |F_{\text{HA}}|, \alpha_{\text{HA}}) \\
&= \frac{2|F^+||F^-|\det(\Sigma_2)}{\pi\det(\Sigma)}\exp[-a_{11}|F^+|^2 - a_{22}|F^-|^2 \\
&\quad - (a_{33}-c_{33})|F_{\text{DM}}|^2] \\
&\quad \times \exp[-(a_{44}-c_{44})|F_{\text{HA}}|^2 - 2|F_{\text{DM}}||F_{\text{HA}}|(a_{34}-c_{34}) \\
&\quad \times \cos(\alpha_{\text{DM}}-\alpha_{\text{HA}})] \\
&\quad \times \int_0^{2\pi} \exp[-2|F^-||F_{\text{DM}}|a_{23}\cos(\alpha^- - \alpha_{\text{DM}})] \\
&\quad \times \exp[-2|F^-||F_{\text{HA}}|a_{24}\cos(\alpha^- - \alpha_{\text{HA}})] \\
&\quad \times I_0\big(2|F^+|\{[(a_{12}|F^-|\cos(\alpha^-) + a_{13}|F_{\text{DM}}|\cos(\alpha_{\text{DM}}) \\
&\quad + a_{14}|F_{\text{HA}}|\cos(\alpha_{\text{HA}})]^2 \\
&\quad + [a_{12}|F^-|\sin(\alpha^-) + a_{13}|F_{\text{DM}}|\sin(\alpha_{\text{DM}}) \\
&\quad + a_{14}|F_{\text{HA}}|\sin(\alpha_{\text{HA}})]^2\}^{1/2}\big)\,\mathrm{d}\alpha^-.
\end{aligned} \tag{1}$$

In (1), $|F^+|$ and $|F^-|$ denote the observed Bijvoet/Friedel pairs from a SAD experiment and $|F_{\text{DM}}|$, $\alpha_{\text{DM}}$ and $|F_{\text{HA}}|$, $\alpha_{\text{HA}}$ are the amplitude and phase of the density-modified and heavy-atom structure factors, respectively. $\Sigma$ is the covariance matrix, with its elements denoted by $a_{ij}$, and $\Sigma_2$ is the bottom-right submatrix of $\Sigma$, with its elements denoted by $c_{ij}$. This equation has been derived in Appendix *A* of Pannu & Read (2004) and Skubák *et al.* (2004) in the context of SAD heavy-atom and protein refinement.

To compare *MULTICOMB* with the current *SIGMAA* program (Read, 1986), we used *SOLOMON* from *CCP*4 (Collaborative Computational Project, Number 4, 1994). The modular design of *SOLOMON* allowed an easy switch between *MULTICOMB* and *SIGMAA* for testing purposes.

In our tests, we started with merged diffraction data from a wide range of real SAD data sets. Only the intensities or structure-factor amplitudes, together with the sequence of a protein monomer, the number of substructure atoms expected per monomer and the $f'$ and $f''$ values for the substructure atoms, were input into the *CRANK* (Ness *et al.*, 2004) structure-solution suite. *CRANK* performed substructure detection using either *AFRO* (Pannu *et al.*, unpublished work) and *CRUNCH*2 (de Graaff *et al.*, 2001) or *SHELXC*, *SHELXD* and *SHELXE* (Sheldrick, 2008). *BP*3 (Pannu & Read, 2004) was used for substructure phasing and 20 cycles of density modification were performed in *SOLOMON*. Information about noncrystallographic symmetry was not used in density modification. Either three cycles of *Buccaneer* or ten

cycles of *ARP/wARP* both iterated with *REFMAC* (Murshudov *et al.*, 1997) using the SAD refinement target (Skubák *et al.*, 2004) were used for automated model building. A resolution cutoff or specification of the number of disulfide bonds was needed for successful substructure determination for a few data sets. Default values of all other options or parameters were used in all programs. In the results shown below, for any data set the changes shown were caused only by running either *SIGMAA* or *MULTICOMB* with *SOLOMON*.

We used a total of 132 real data sets from several different sources as listed in Appendix *A*. The majority of these structures were originally solved by multiwavelength anomalous diffraction (MAD), single isomorphous replacement with anomalous scattering or molecular replacement. Data sets where we could not determine the substructure or where a program within any pipeline terminated abnormally were excluded from the statistics presented, resulting in 102 data sets. In cases where multiple SAD data sets were available (*i.e.* data collected either at different wavelengths or processed with different redundancies), the SAD data set corresponding to the highest $f''$ value and to the highest data redundancy was used. The data sets provided a wide range of resolution (from 0.94 to 3.29 Å) and anomalous scatterers, including selenium, sulfur, chloride, sulfate, manganese, bromide, calcium and zinc.

## 3. Results

Fig. 2 compares the quality of the electron-density maps obtained by density modification using the *SIGMAA* algorithm with the maps obtained using the multivariate SAD-DM function for the 102 SAD data sets. The quality of a map is measured by its correlation with the map constructed from the deposited model. In general, the graph can be divided into three distinct regions. In the first region the map correlation for both algorithms is less than 0.2. This region shows test cases where the initial experimental phase information is too weak and results in an uninterpretable electron-density map. In the second region, which contains map correlations over 0.8 after DM using both algorithms, all the points lie very close to the diagonal line, indicating that both algorithms performed similarly, producing high-quality phases for these data sets. The remaining region shows the greatest variation between the two methods: in this region there is a 21.4% increase in average map correlation for the multivariate SAD-DM function. The overall increase in average map correlation calculated from all data sets is 16.9% (from 0.539 to 0.630).

In Fig. 3 the data sets are divided into two categories based on the improvement of SAD-DM over *SIGMAA* maps and shown as a function of data resolution and map correlation after experimental phasing. The figure indicates that when the input map correlation is below 0.3 neither method can usually significantly improve the map. Futhermore, when the input map correlation is above 0.4 and the data set has a resolution of 2.0 Å or better both algorithms usually produce equally high-quality maps. A region in which SAD-DM clearly outperforms *SIGMAA* are data sets with a resolution lower than 2.1 Å and an input map correlation higher than 0.35.

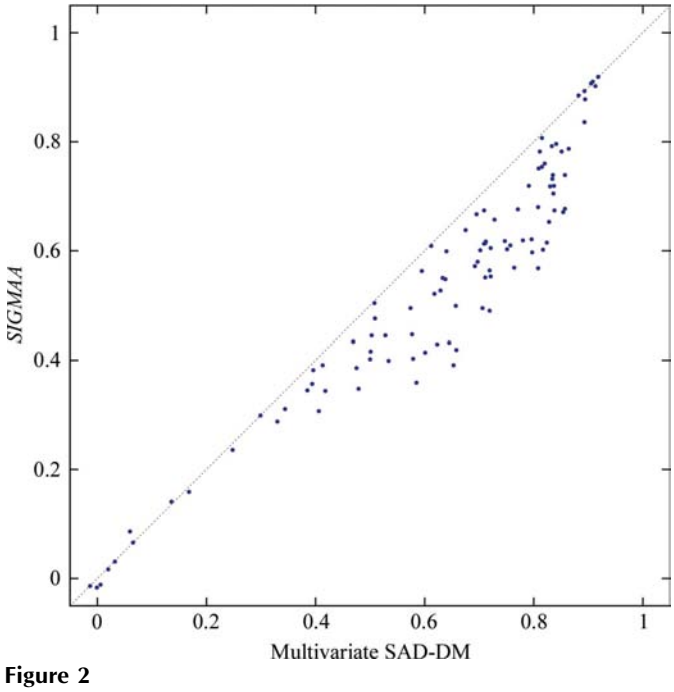


**Figure 2**
Correlation of maps obtained from the multivariate SAD-DM function (*x* axis) and the *SIGMAA* algorithm (*y* axis) with maps generated from the final deposited models.

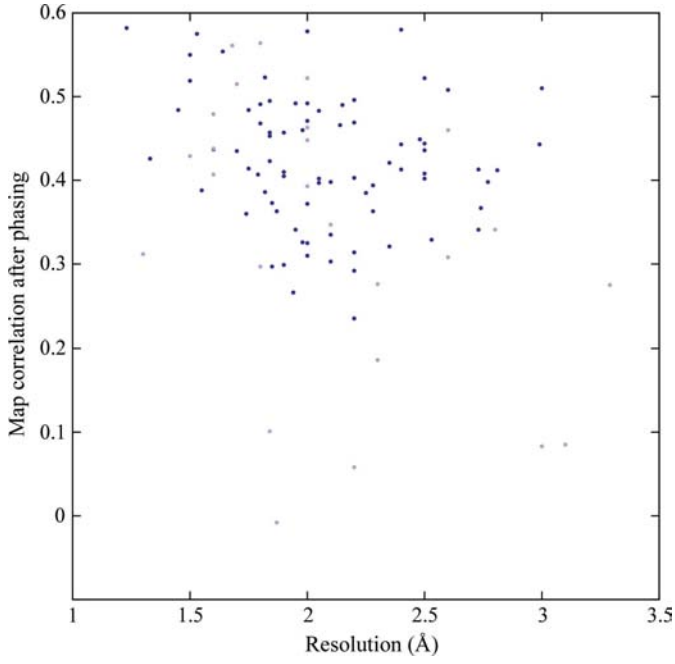


**Figure 3**
Map improvement of the SAD-DM algorithm over *SIGMAA* as a function of resolution and map correlation after experimental phasing. The data sets were divided into two groups based on the difference between the SAD-DM and *SIGMAA* map correlations with the deposited structure: the blue points represent data sets where the SAD-DM map correlation is better than the *SIGMAA* map correlation by 0.03 or higher, while a correlation difference of less than 0.03 is shown in purple.

**Table 1**
Average fraction of the model correctly built by *Buccaneer* and *ARP/wARP*.

| | *SIGMAA* | SAD-DM |
|---|---|---|
| *Buccaneer* | 0.446 | 0.610 |
| *ARP/wARP* | 0.533 | 0.620 |

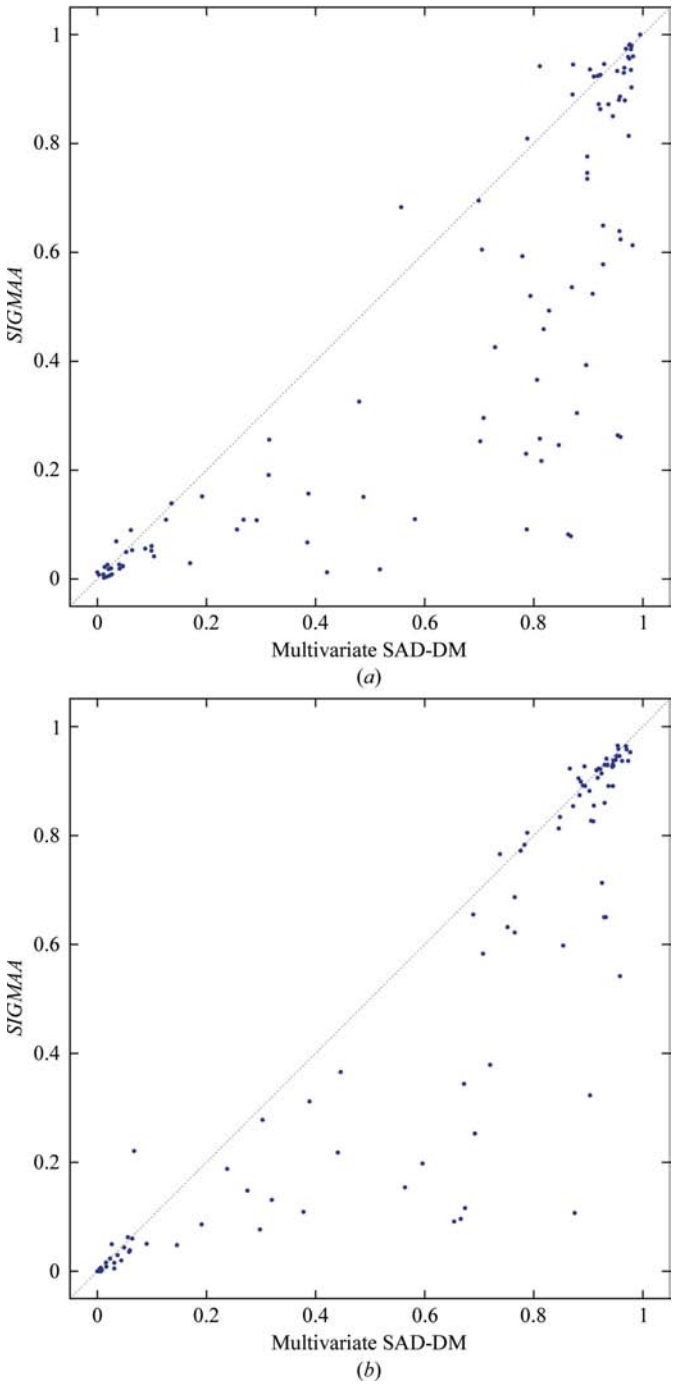


**Figure 4**
Performance of the multivariate SAD-DM function *versus* the *SIGMAA* algorithm in terms of the fraction correctly built by the model-building programs (*a*) *Buccaneer* and (*b*) *ARP/wARP*. The *x* axis and *y* axis are the fraction of the model built after phase combination with *MULTICOMB* and *SIGMAA*, respectively. A residue is regarded as 'correctly built' if it is within 1 Å of the deposited model.

Next, we examined the effect of the different phase-combination algorithms on automated model building with iterative model refinement using *ARP/wARP* (Perrakis *et al.*, 1999) and *Buccaneer* (Cowtan, 2006). Since *ARP/wARP* and *Buccaneer* use different protein-tracing algorithms, examining the behaviour of both programs provides a better comparison of the two different phase-combination algorithms. Fig. 4(*a*) shows the results from *Buccaneer*, while Fig. 4(*b*) shows the results produced by *ARP/wARP*. The two graphs can again be divided into three distinct regions: a region in which the fraction of the model correctly built is under 10%, indicating that model building has failed for both phase-combination algorithms, a region in which both phase-combination algorithms lead to a model that is over 80% correctly built and a remaining region which shows the greatest disparity between the two phase-combination methods. This last region includes 36 data sets for which the maps produced by the multivariate SAD-DM function result in a 59.6% increase in the average fraction of the model built by *ARP/wARP*. The same region contains 50 data sets when using *Buccaneer*, resulting in a 93.0% increase in the fraction of the model built using multivariate methods over that obtained using *SIGMAA*.

## 4. Discussion

The results clearly show that in general the multivariate SAD-DM function significantly outperforms the existing phase-combination method in *SIGMAA*. The improved map correlation provided by the multivariate SAD-DM function led to a 16.4% improvement in the overall average fraction built using *ARP/wARP* and a 36.7% improvement when using *Buccaneer* compared with the state of the art (Table 1). The difference in the degree of improvement between the model-building programs is mainly caused by the multivariate SAD refinement target used with *ARP/wARP*, which enables successful model building after *SIGMAA* for many data sets that would otherwise fail: using the default Rice target the *ARP/wARP* improvement is similar to that of *Buccaneer* (44%).

Although only SAD data were used in the above tests, the multivariate framework laid out is by no means only limited to SAD experiments. Indeed, we have already implemented a multivariate single isomorphous replacement with anomalous scattering function for phase combination and the initial results are very promising (Skubák & Pannu, in preparation). Furthermore, a multivariate MAD function is currently being implemented.

The introduction of the multivariate function to phase combination has important consequences for automated structure solution and manual structure solution at lower resolutions. The same multivariate function described above is used in substructure phasing, density modification and model refinement. Thus, at any step in the process up to or including substructure phasing, if a correct anomalous or even non-anomalous portion of the structure is added significantly improved phases can be generated *via* simultaneously considering phasing, density modification and model refinement together. In essence, the successful implementation of multi-

variate methods to phase combination should lead to a new paradigm in which the crystallographer no longer considers phasing, density modification and model refinement as separate parts of structure determination but as one process using one mathematical function that can lead to combined and significant improvements over considering them separately.

## APPENDIX *A*
## Data sets

A total of 132 SAD data sets were used and were composed of 78 data sets from the Joint Center for Structural Genomics (JCSG; http://www.jcsg.org/; 1vjn, 1vjr, 1vjz, 1vk4, 1vkm, 1vlm, 1vqr, 1z82, 1zy9, 1zyb, 2a2m, 2a2o, 2a3n, 2a6b, 2aml, 2avn, 2b8m, 2etd, 2etj, 2ets, 2etv, 2evr, 2f4p, 2fea, 2ffj, 2fg0, 2fg9, 2fna, 2fqp, 2fur, 2fzt, 2g0t, 2g42, 2gc9, 2nlv, 2nuj, 2nwv, 2o08, 2o1q, 2o2x, 2o2z, 2o3l, 2o62, 2o7t, 2o8q, 2obp, 2oc5, 2od5, 2od6, 2oh3, 2okc, 2okf, 2ooj, 2opk, 2osd, 2otm, 2ozg, 2ozj, 2p10, 2p4o, 2p7h, 2p7i, 2p97, 2pg3, 2pg4, 2pgc, 2pim, 2pn1, 2pnk, 2ppv, 2pr7, 2prr, 2prv, 2prx, 2pv4, 2pw4, 2b78 and 2b79); 23 data sets from Mueller-Dieckmann *et al.* (2007) (2g4h, 2g4i, 2g4j, 2g4k, 2g4p, 2g4q, 2g4l, 2g4n, 2g4o, 2g4r, 2g4s, 2g4t, 2g4u, 2g4v, 2g4w, 2g4x, 2g4y, 2g4z, 2ill, 2g51, 2g52, 2g54 and 2g55); and 31 from various other individual data-set contributors: 1e42 (Owen, Vallis *et al.*, 2000), 1e6i (Owen, Ornaghi *et al.*, 2000), 1hf8 (Ford *et al.*, 2001), 2ahy (Shi *et al.*, 2006), 2hba (J.-H. Cho, S. Sato, E. Y. Kim, H. Schindelin & D. P. Raleigh, unpublished work), 2o0h (Sun *et al.*, 2007), 2rkk (Xiao *et al.*, 2008), 3bpj (L. Nedyalkova, B. Hong, W. Tempel, F. MacKenzie, C. H. Arrowsmith, A. M. Edwards, J. Weigelt, A. Bochkarev & H. Park, unpublished work), 2fdn (Dauter *et al.*, 1997), 1of3 (Boraston *et al.*, 2003), 1i4u (Gordon *et al.*, 2001), 1dw9 (Walsh *et al.*, 2000), 1v0o (Holton *et al.*, 2003), 1fse (Ducros *et al.*, 2001), 1xib (Carrell *et al.*, 1989), 1fj2 (Devedjiev *et al.*, 2000), 1h29 (Matias *et al.*, 2002), 1c8u (Jia *et al.*, 2000), 1lvy (Schiltz *et al.*, 1997), 1lz8 (Dauter *et al.*, 1999), 1e3m (Lamers *et al.*, 2000), 1ga1 (Dauter *et al.*, 2001), 1djl (White *et al.*, 2000), 1dtx (Skarzynski, 1992), 1dpx (Weiss, 2001), 1mso (Smith *et al.*, 2003), 1ocy (Thomassen *et al.*, 2003), 1rju (Calderone, 2004), 1rgg (Sevcik *et al.*, 1996), 1m32 (Chen *et al.*, 2002) and a subtilisin data set (Betzel *et al.*, 1988; Dauter *et al.*, 2002).

We thank all of the authors who kindly provided us with the SAD data sets, including the JCSG (http://www.jcsg.org/), M. Weiss, C. Mueller-Dieckmann and Z. Dauter. We thank R. A. G. de Graaff, J. P. Abrahams and K. Cowtan for useful discussions. Funding for this work was provided by Leiden University, the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) and Cyttron. *MULTICOMB* will be distributed as free open-source software *via* the website http://www.bfsc.leidenuniv.nl/software/crank/ and in future versions of *CCP*4.

## 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.
Betzel, C., Dauter, Z., Dauter, M., Ingelman, M., Papendorf, G., Wilson, K. S. & Branner, S. (1988). *J. Mol. Biol.* **204**, 803–804.
Boraston, A. B., Revett, T. J., Boraston, C. M., Nurizzo, D. & Davies, G. J. (2003). *Structure*, **11**, 665–675.
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.
Calderone, V. (2004). *Acta Cryst.* D**60**, 2150–2155.
Carrell, H. L., Glusker, J. P., Burger, V., Manfre, F., Tritsch, D. & Biellmann, J. F. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 4440–4444.
Chen, C. C. H., Zhang, H., Kim, A. D., Howard, A., Sheldrick, G. M., Mariano-Dunaway, D. & Herzberg, O. (2002). *Biochemistry*, **41**, 13162–13169.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
Cowtan, K. (2000). *Acta Cryst.* D**56**, 1612–1621.
Cowtan, K. (2006). *Acta Cryst.* D**62**, 1002–1011.
Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
Dauter, Z., Dauter, M. & Dodson, E. J. (2002). *Acta Cryst.* D**58**, 494–506.
Dauter, Z., Li, M. & Wlodawer, A. (2001). *Acta Cryst.* D**57**, 239–249.
Dauter, Z., Wilson, K. S., Sieker, L. C., Meyer, J. & Moulis, J. M. (1997). *Biochemistry*, **36**, 16065–16073.
Devedjiev, Y., Dauter, Z., Kuznetsov, S. R., Jones, T. L. & Derewenda, Z. S. (2000). *Structure*, **8**, 1137–1146.
Ducros, V. M., Lewis, R. J., Verma, C. S., Dodson, E. J., Leonard, G., Turkenburg, J. P., Murshudov, G. N., Wilkinson, A. J. & Brannigan, J. A. (2001). *J. Mol. Biol.* **306**, 759–771.
Ford, M. G., Pearse, B. M., Higgins, M. K., Vallis, Y., Owen, D. J., Gibson, A., Hopkins, C. R., Evans, P. R. & McMahon, H. T. (2001). *Science*, **291**, 1051–1055.
Gordon, E. J., Leonard, G. A., McSweeney, S. & Zagalsky, P. F. (2001). *Acta Cryst.* D**57**, 1230–1237.
Graaff, R. A. G. de, Hilge, M., van der Plas, J. L. & Abrahams, J. P. (2001). *Acta Cryst.* D**57**, 1857–1862.
Hendrickson, W. A. & Lattman, E. E. (1970). *Acta Cryst.* B**26**, 136–143.
Holton, S., Merckx, A., Burgess, D., Doerig, C., Noble, M. & Endicott, J. (2003). *Structure*, **11**, 1329–1337.
Jia, L., Derewenda, U., Dauter, Z., Smith, S. & Derewenda, Z. S. (2000). *Nature Struct. Biol.* **7**, 555–559.
Lamers, M. H., Perrakis, A., Enzlin, J. H., Winterwerp, H. H., de Wind, N. & Sixma, T. K. (2000). *Nature (London)*, **407**, 711–717.
Leslie, A. G. W. (1987). *Acta Cryst.* A**43**, 134–136.
Lunin, V. Yu. & Urzhumtsev, A. G. (1984). *Acta Cryst.* A**40**, 269–277.
Matias, P. M., Coelho, A. V., Valente, F. M., Placido, D., LeGall, J., Xavier, A. V., Pereira, I. A. & Carrondo, M. A. (2002). *J. Biol. Chem.* **277**, 47907–47916.
Mueller-Dieckmann, C., Panjikar, S., Schmidt, A., Mueller, S., Kuper, J., Geerlof, A., Wilmanns, M., Singh, R. K., Tucker, P. A. & Weiss, M. S. (2007). *Acta Cryst.* D**63**, 366–380.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Ness, S. R., de Graaff, R. A. G., Abrahams, J. P. & Pannu, N. S. (2004). *Structure*, **12**, 1753–1761.
Owen, D. J., Ornaghi, P., Yang, J. C., Lowe, N., Evans, P. R., Ballario, P., Neuhaus, D., Filetici, P. & Travers, A. A. (2000). *EMBO J.* **19**, 6141–6149.
Owen, D. J., Vallis, Y., Pearse, B. M., McMahon, H. T. & Evans, P. R. (2000). *EMBO J.* **19**, 4216–4227.
Pannu, N. S. & Read, R. J. (2004). *Acta Cryst.* D**60**, 22–27.

Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Read, R. J. (1997). *Methods Enzymol.* **277**, 110–128.
Schiltz, M., Shepard, W., Fourme, R., Prangé, T., de La Fortelle, E. & Bricogne, G. (1997). *Acta Cryst.* D**53**, 78–92.
Sevcik, J., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1996). *Acta Cryst.* D**52**, 327–344.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Shi, N., Ye, S., Alam, A., Chen, L. & Jiang, Y. (2006). *Nature (London)*, **440**, 570–574.
Skarzynski, T. (1992). *J. Mol. Biol.* **224**, 671–683.
Skubák, P., Murshudov, G. N. & Pannu, N. S. (2004). *Acta Cryst.* D**60**, 2196–2201.
Skubák, P., Ness, S. & Pannu, N. S. (2005). *Acta Cryst.* D**61**, 1626–1635.
Smith, G. D., Pangborn, W. A. & Blessing, R. H. (2003). *Acta Cryst.* D**59**, 474–482.
Sun, S., Kondabagil, K., Gentz, P. M., Rossmann, M. G. & Rao, V. B. (2007). *Mol. Cell*, **25**, 943–949.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Thomassen, E., Gielen, G., Schutz, M., Schoehn, G., Abrahams, J. P., Miller, S. & van Raaij, M. J. (2003). *J. Mol. Biol.* **331**, 361–373.
Walsh, M. A., Otwinowski, Z., Perrakis, A., Anderson, P. M. & Joachimiak, A. (2000). *Structure*, **8**, 505–514.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.
White, S. A., Peake, S. J., McSweeney, S., Leonard, G., Cotton, N. P. & Jackson, J. B. (2000). *Structure*, **8**, 1–12.
Xiao, J., Xia, H., Zhou, J., Azmi, I. F., Davies, B. A., Katzmann, D. J. & Xu, Z. (2008). *Dev. Cell*, **14**, 37–49.