Acta Cryst. (1997). D**54**, 461–466

Enhanced electron-density envelopes by extended solvent definition

NICK BLOM* AND JURGEN SYGUSH *at Département de biochemie, Université de Montréal, CP 6128 Station Centre Ville, Montréal, Canada H3C 3J7. E-mail: nick@bch.umontreal.ca*

(*Received 25 October 1996; accepted 29 April 1997*)

Abstract

Extended delineation of water molecules, monitored using R_{free} values, afforded considerable improvement in quality of electron-density maps for structure determination of mammalian class I and *E. coli* class II aldolases. Augmented solvent definition results in an additional decrease in R_{free} values of 3–4% and is reflected in significantly enhanced electron-density envelopes enabling tracing of amino-acid sequences through regions of otherwise discontinuous or weak electron density.

1. Introduction

The molecular architectures of D-fructose 1,6-bisphosphate aldolases from rabbit skeletal muscle (Blom & Sygusch, 1997), rabbit liver (Blom & Sygusch, 1998) and *E. coli* (Blom *et al.*, 1996) have been recently determined to high resolution and corresponding crystal data are summarized in Table 1. Here we report on the method of electron-density enhancement by extended solvent definition that was used successfully in elucidating amino-acid chain traces through regions of otherwise difficult to interpret $2F_o - F_c$ electron densities. In structure determination of rabbit muscle class I aldolase, initial chain tracing of residues 344–363, corresponding to the COOH terminal region, was unsuccessful (Sygusch *et al.*, 1987) and was attributed to conformational flexibility of this region because of its likely role in catalytic activity (Hannappel *et al.*, 1974; Humphreys *et al.*, 1986; Sygusch *et al.*, 1990; Dobeli *et al.*, 1991; Berthiaume *et al.*, 1991, 1993). In the structure determination of rabbit liver aldolase, the COOH terminal region also was not visible in initial $2F_o - F_c$ or $F_o - F_c$ electron-density maps. Unequivocal determination of the trace of the COOH terminal region in both isozyme structures was essential because of persistent electron density in each active site vicinal to the Schiff-base forming Lys229 residue, which plays an essential catalytic role in class I aldolases (Grazi *et al.*, 1962).

In the crystal structure determination of Zn-dependent class II *E. coli* aldolase, two loop regions corresponding to residues 177–196 and 227–234, respectively, could not be discerned from the $2F_o - F_c$ or $F_o - F_c$ electron-density map, even after numerous iterations of phase combination, model rebuilding, simulated annealing and energy refinement. The structure of *E. coli* aldolase displays a novel active site in which the catalytic zinc metal ion exchanges between two mutually exclusive positions. Among the zinc-chelating residues identified in the catalytic sites was a histidine residue positioned at the onset of the smaller loop region. Unequivocal tracing of the loop regions was, therefore, essential for comprehensive structure determination.

Earlier work describing structural studies in which a large number of solvent molecules were identified included the 2Zn pig insulin structure refined to 1.5 Å resolution (Baker *et al.*, 1988). High-resolution refinement delineated a large number of water molecules, particularly localized in crevices, up to 16 Å deep, which were a consequence of hexagonal packing of the insulin molecules. In the present contribution, the addition of large numbers of water molecules enhanced the definition of electron-density envelopes and is described below.

Table 1. *Crystal data summary*

Class I D-fructose 1,6-bisphosphate aldolase from rabbit skeletal muscle

Unit-cell parameters (Å, °)	a = 163.88, b = 57.47, c = 85.03, β = 102.7
Space group	$P2_1$
No. of reflections	103662
Resolution (Å)	1.9
Asymmetric unit	Homotetramer, 1452 amino-acid residues

Class I D-fructose 1,6-bisphosphate aldolase from rabbit liver

Unit-cell parameters (Å, °)	a = 377.25, b = 130.55, c = 80.03
Space group	$C222_1$
No. of reflections	85613
Resolution (Å)	2.1
Asymmetric unit	2 half tetramers, 1452 amino-acid residues

Class I D-fructose 1,6-bisphosphate aldolase from *E. coli*

Unit-cell parameters (Å, °)	a = 90.53, b = 73.38, c = 57.80, β = 106.6
Space group	$P2_1$
No. of reflections	75160
Resolution (Å)	1.65
Asymmetric unit	Dimer, 716 amino-acid residues

2. Materials and methods

In all three aldolase structures, the R_{free} test value (Brünger, 1992*a*) was calculated based on 8% of the data that had been selected prior to any refinement and was excluded throughout all refinement. Structure determination was performed according to proper model-building and refinement practice (Kleywegt & Jones, 1998). Initial interpretation of the $2F_o - F_c$ electron-density envelopes was based on amino-acid sequence and was typically followed by model rebuilding, refinement and monitored using the R_{free} value. Putative water molecules could be readily discerned from $F_o - F_c$ electron-density maps at an R_{free} value of 30% in the course of structure refinement and was used as starting point for inclusion of water molecules in all three structure determinations. Not until a considerable number of solvent molecules had been delineated, monitored by R_{free}, could previously ill-

defined regions in class I and class II aldolases be successfully traced through electron density.

Water molecules were selected as follows. Putative water molecules were selected from respective $F_o - F_c$ electron-density maps corresponding to electron-density peak levels exceeding 2.4σ and within 4 Å of atom positions, obtained from a previous round of refinement, to satisfy potential hydrogen bond or van der Waals distance criteria. Calculation of the value of the standard deviation, σ, for the $2F_o - F_c$ and $F_o - F_c$ electron-density maps was based on a volume derived from the asymmetric unit and surrounded by a 3 Å additional cushion. During refinement using the *X-PLOR* suite of programs (Brünger 1992*b*), water molecules were harmonically restrained to their initial selected positions. In cases where a simulated-annealing protocol involved elevated temperatures (3000–4000 K), a force-restraining constant of 8368 kJ mol^{-1} (2000 kcal mol^{-1}) was applied. Simulated annealing was based on a protocol comprising 500 steps of 0.2 fs simulation followed by a temperature decrement of 250 K after each simulation to a final temperature of 300 K. Energy-refinement protocols, generally consisted of 100–150 cycles of Powell conjugate-gradient minimization (Brünger, 1992*b*). During energy refinement water molecules were harmonically restrained using a force constant of 83.68 kJ mol^{-1} (20 kcal mol^{-1}). Force-field parameters for water molecules during simulated annealing and refinement were based on the TIP3P model (Jorgenson *et al.*, 1983) simplified to include only the non-bonded term of the water O atom. Force-field parameters used to describe protein atoms corresponded to Engh and Huber parameters (Engh & Huber, 1991). After refinement, a validation round was performed testing each water molecule. Using the *rsr-rigid* tool in the program *O* (Jones *et al.*, 1991), each water molecule was repositioned to coincide with a nearest electron-density peak within a 3 Å radius, previously identified from a $2F_o - F_c$ electron-density map where all other known atoms had been previously subtracted (Jones *et al.*, 1991). Subsequently, each water molecule position was validated in terms of hydrogen-bonding geometry and close contacts (the resultant peak position must be > 2.4 Å from neighbouring atoms). Water molecules displaying poor geometry and close contacts were rejected. Next, the protein structure and surviving water molecules were subjected to a round of Powell energy minimization (~50 cycles) to relieve close contacts. Finally, water molecules were evaluated based on their temperature factors and rejected if their B factor exceeded 100 Å^2. The resulting model was used in a subsequent round of model rebuilding.

This procedure was iterated until no further improvement in $R_{\rm free}$ could be obtained from selection of additional water molecules and coincided with successful chain tracing of the problem regions in all three aldolase structures. Structures were finalized by visually inspecting all water molecule positions in the asymmetric unit. Where appropriate, water molecules were either added, deleted and/or recentered. During rebuilding, model integrity was monitored using program *OOPS* (Kleywegt & Jones, 1996) and inspected for multiple side-chain conformations.

To validate the above water molecule definition protocol, final atomic coordinates from rabbit muscle aldolase, rabbit liver aldolase and *E. coli* aldolase were used as starting point for comparative $R_{\rm free}$ calculations (referred to as model *C*). In order to assess the contribution of extended solvent definition all water molecules were first removed from the final model and the resultant atomic coordinate set used for $R_{\rm free}$ and $2F_o - F_c$ electron-density map calculations (structure model *A*). To eliminate model bias, protein structure model *A* was subjected to an *XREF* restraint dynamical simulation protocol (Brünger 1992*b*) at ambient temperature for 500 steps of 0.2 fs (*Shake and Bake*) followed by 50 cycles of Powell minimization and 40 cycles of B-factor refinement, after which $R_{\rm free}$ was calculated (structure model *A'*). Model *A* was also used to calculate an $F_o - F_c$ electron-density map from which putative water molecule coordinates corresponding to peak electron-density levels exceeding 3.0σ were selected, added to model *A*, and checked for geometric integrity. The resulting model was subjected to the same *Shake and Bake* and refinement protocols as previously and followed by an $R_{\rm free}$ calculation (structure model *B*). Bulk-solvent deconvolution protocols described elsewhere (Brünger, 1992*b*) optimized for scale factor and overall B factor were applied to structure model *B* and $R_{\rm free}$ calculated (structure model *B'*).

3. Results and discussion

Results of the proposed method applied to the structures of class I aldolases from rabbit muscle and rabbit liver as well as to class II aldolase from *E. coli* are illustrated in Table 2, where the last column lists the $R_{\rm free}$ values for respective structure models (*c.f.* §2). Also shown are the number of water molecules in each structure model and the r.m.s. deviations of atomic coordinates (non-H atoms) of the protein structure with respect to model *C*. Deletion of all water molecules from model *C* (model *A*) shows a moderate increase in $R_{\rm free}$ of 1.5–2% in the case of rabbit muscle aldolase and *E. coli* aldolase. The significant jump of ~7% in $R_{\rm free}$ for rabbit liver aldolase exemplifies the importance of solvent contribution in model description. By comparison, removal of previous ill-defined regions from model *C* showed only a slight increase in $R_{\rm free}$ of less than 0.2% for all three structures.

Figs. 1 and 2 show $2F_o - F_c$ electron-density maps as function of structure model, corresponding to portions of previous ill-defined regions for rabbit muscle aldolase (residues 349–353) and *E. coli* aldolase (residues 177–182), respectively, and having superimposed the amino-acid trace of the final structure. In both Figs. 1(*a*) and 2(*a*), removal of water molecules (model *A*) deteriorates the quality of the resulting electron-density maps with respect to structure model *C* maps. A similar degradation in map quality was also observed for rabbit liver aldolase (not shown). Further deterioration in map quality is evident after elimination of model bias (model *A'*) and is shown in Figs. 1(*b*) and 2(*b*). From Table 2, comparison of model *B* and *C* structures indicates that the additional water selection in the 3.0–2.4σ range diminishes $R_{\rm free}$ values by 3–4% and is comparable to the reduction in $R_{\rm free}$ values when comparing model *A'* with *B* structures. Improvement $R_{\rm free}$ by model *C* with respect to model *B* is not a consequence of structural changes since the two protein structures exhibit negligible deviation in their atomic coordinates (Table 2 column 2) and display no appreciable conformational differences. The traditional method of bulk-solvent flattening applied to the structure model *B*, model *B'*, has little impact on $R_{\rm free}$ with respect to the present explicit approach. The decrease in $R_{\rm free}$ values resulting from the extended definition of water molecules suggests that considerable information is still present in the

Table 2. *Explicit solvent definition and R_{free} value comparison*

R.m.s. values are based on all atoms (non-H atoms) of the protein molecule with respect to model *C*. Crystallographic *R* factor = $\sum_{hkl} ||F_o(hkl)| - |F_c(hkl)|| / \sum_{hkl} |F_o(hkl)|$. $R_{free} = \sum_{hkl \in T} ||F_o(hkl)| - |F_c(hkl)|| / \sum_{hkl \in T} |F_o(hkl)|$ where *T* is the test data set randomly selected from the observed reflections prior to refinement. The test data set contained 8% of the total data and was not used in the refinement. Model *A* is the original complete and refined structure, without water molecules, exhibiting maximal model bias. Model *A'* is the structure obtained after solvent removal from structure model *C*, *Shake and Bake* algorithm applied, followed by Powell minimization and *B*-factor refinement. Model *B* is the structure containing water-molecule peaks selected from electron-density difference maps whose peak heights exceed 3σ cutoff level, after *Shake and Bake*, Powell minimization and *B*-factor refinement. Model *B'* is the structure containing water-molecule peaks selected from electron-density difference maps whose peak heights exceed 3σ cutoff level, after *Shake and Bake*, Powell minimization and *B*-factor refinement and with additional bulk-solvent correction. Model *C* is the original complete and refined structure.

Class I D-fructose 1,6–bisphosphate aldolase from rabbit skeletal muscle

	R.m.s. (Å)	Water molecules included	R_{free} (R_{conv}) (%)
Model *A*	—	0	22.38 (19.36)
Model *A'*	0.265	0	28.02 (23.89)
Model *B*	0.265	686	24.63 (20.95)
Model *B'*	0.265	686	24.57 (20.95)
Model *C*	—	3327	20.33 (16.09)

Class I D-fructose 1,6–bisphosphate aldolase from rabbit liver

	R.m.s. (Å)	Water molecules included	R_{free} (R_{conv}) (%)
Model *A*	—	0	28.58 (25.47)
Model *A'*	0.342	0	29.51 (24.06)
Model *B*	0.315	767	26.40 (21.64)
Model *B'*	0.315	767	26.22 (21.58)
Model *C*	—	3702	21.98 (16.89)

Class I D-fructose 1,6–bisphosphate aldolase from *E. coli*

	R.m.s. (Å)	Water molecules included	R_{free} (R_{conv}) (%)
Model *A*	—	0	21.74 (19.53)
Model *A'*	0.265	0	26.60 (23.34)
Model *B*	0.243	440	23.05 (20.48)
Model *B'*	0.243	440	22.75 (20.39)
Model *C*	—	1611	20.24 (17.29)

bulk-solvent region. Even though water molecule selection in model *B* was based on the $F_o - F_c$ electron-density difference map of model *A* and, thus, biased towards model *C* and the original set of water molecules, no apparent improvement in map quality was obtained in model *B* (Figs. 1*c* and 2*c*).

Clearly the enhanced electron-density features observed in Figs. 1(*e*) and 2(*e*) are a consequence of the extended solvent-definition protocol including the repetitive validation of water molecule positions. The enhanced quality of the resultant electron-density maps suggests that selection of water molecule positions from peak densities lower than the traditional 3.0σ cutoff (McRee, 1993) should be useful in future structure determinations. Addition of peaks to lower σ levels ($<2.4\sigma$) resulted in the inclusion of many spurious noise peaks and did not decrease the R_{free} value. The extended water selection resulted primarily in identification of additional water molecules in protein crevices, subunit interfaces and first solvation layers on protein surfaces.

The proposed method depends critically on the use of R_{free} value to objectively monitor model improvement at each stage of refinement, because of the inherent danger of over-refinement a consequence of underdetermination of the atomic coordinates. The use of R_{free} where in our case 8% of the data was excluded from refinement, has been shown to objectively monitor phase improvement (Brünger, 1992*a*) and correlates with our improved discrimination of density envelopes from background noise. Success of the proposed method is based on the premise that the water selection protocol selects with higher probability authentic than artifactual water molecules. Since a large portion of the structure is generally well defined at the inception of the water selection protocol, repeated application of a protocol that removes potentially artifactual waters from refinement and electron-density map calculation will contribute to overall structure improvement. Recursive elimination of artifactual water molecules *versus* retention of legitimate water molecules is essential for success of the approach. Support for the method is evident from crystallographic Fourier theory where improved phasing yields improved electron-density maps with concomitant improved interpretation, and *vice versa* (Drenth, 1994). The iterative nature of the above method reflects this theoretical underpinning in which each step reduces potential model bias and resulting in an R_{free} decrease. More importantly, ordered solvent contributes non-negligibly to structure-factor calculations in the low-resolution range (>3.0 Å) and represents typically the range in which structure having temperature factors >60 Å^2 such as disordered and/or conformationally flexible regions also contribute to protein structure factors. By thus including water molecules with *B* factors of up to 100 Å^2 the resultant structure model can render an improved definition of previous difficult to discern continuous regions of electron density shown in Figs. 1 and 2. Final average *B* factor corresponding to the problem region (residues 344–363) in rabbit muscle aldolase was 66±24 Å^2 for the corresponding region in rabbit liver aldolase 84±12 Å^2, and for the two loop regions (residues 177–194 and 227–234) in *E. coli* aldolase 86±9 and 71±11 Å^2, respectively.

Inclusion of water molecules having *B* factors $\simeq$ 100 Å^2 represents a non-negligible scattering contribution, typically greater than 10% at 3.5 Å resolution, with respect to water molecules having a *B* factor of zero. In addition, incorporating scatterers with high *B* factors into the model enables compensation for partially disordered water molecules present in the solvent region. In the solvent region, the energy minimum corresponding to the position of a water molecule vicinal to the protein surface is less well defined, reflected in much shorter average residence times, compared with water molecules in crevices of the protein structure (Levitt & Park, 1993; Karplus & Faerman, 1994). From a physical standpoint, it is, therefore, more appropriate to treat solvent disorder using high *B* factors than to treat it as a multiple of partially occupied solvent sites with separate *B* factors that are individually much lower. Recently, application of multiple anomalous dispersion (MAD) phasing techniques (Hendrickson *et al.*, 1988) have proved useful in determining the site-specific radial distribution function of solvent electron

density and resulted in the identification of well defined solvation shells around both hydrophilic and hydrophobic amino-acid residues (Burling *et al.*, 1996). Experimental validation of the proposed method – addition of water molecules from solvent regions using a reduced σ cutoff for peak selection – could be sought from structures solved by MAD phasing techniques. Statistical aspects of the solvation based on the present method will be dealt with in separate paper.

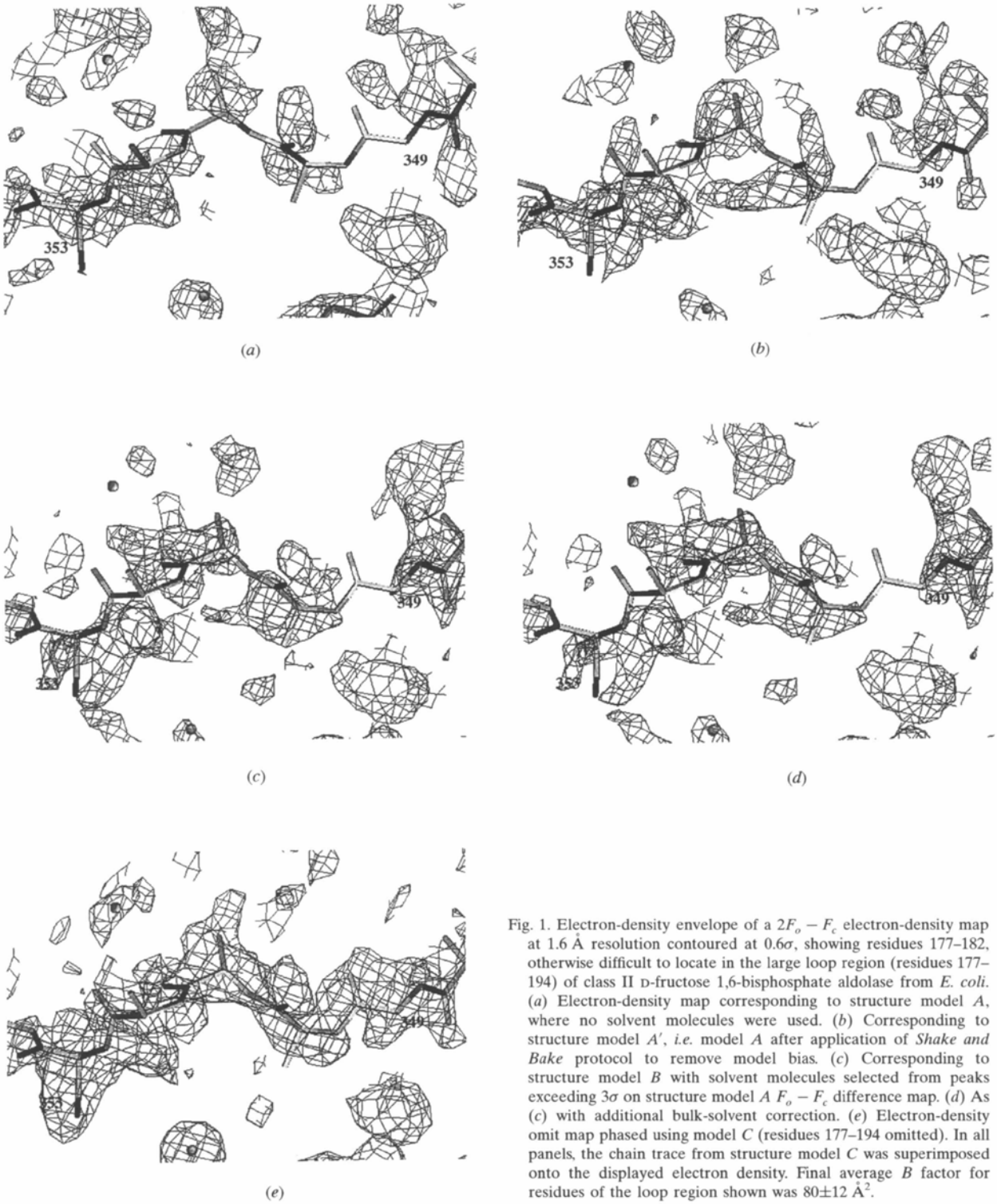
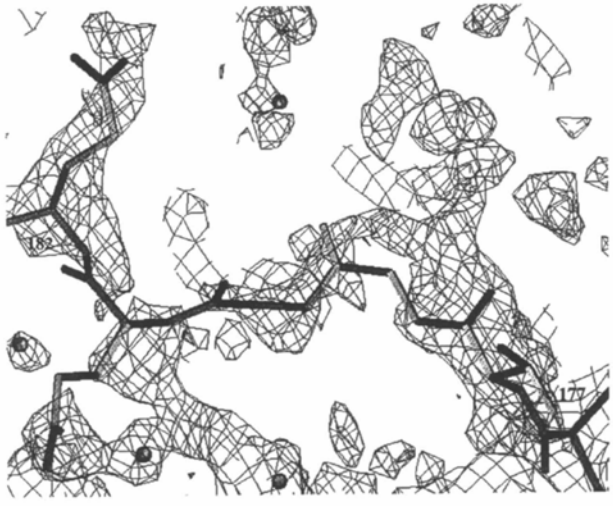
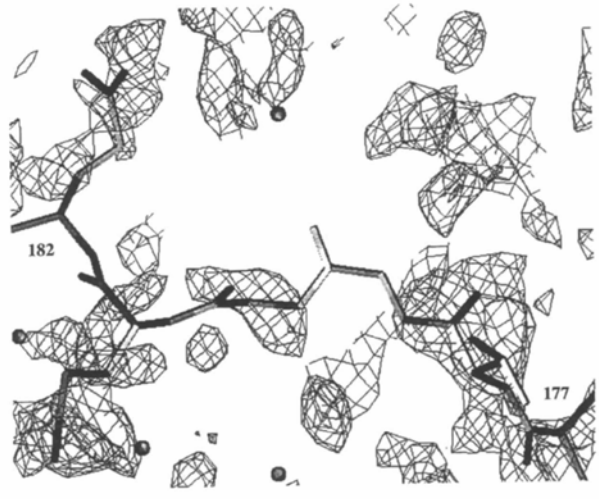


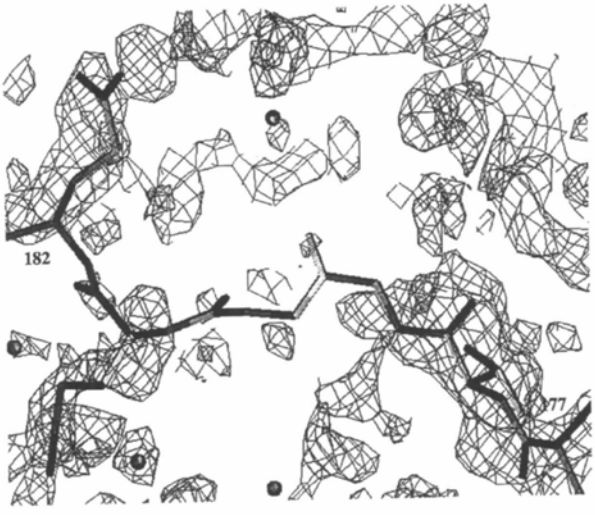
Fig. 1. Electron-density envelope of a $2F_o - F_c$ electron-density map at 1.6 Å resolution contoured at 0.6σ, showing residues 177–182, otherwise difficult to locate in the large loop region (residues 177–194) of class II D-fructose 1,6-bisphosphate aldolase from *E. coli*. (*a*) Electron-density map corresponding to structure model *A*, where no solvent molecules were used. (*b*) Corresponding to structure model *A'*, *i.e.* model *A* after application of *Shake and Bake* protocol to remove model bias. (*c*) Corresponding to structure model *B* with solvent molecules selected from peaks exceeding 3σ on structure model *A* $F_o - F_c$ difference map. (*d*) As (*c*) with additional bulk-solvent correction. (*e*) Electron-density omit map phased using model *C* (residues 177–194 omitted). In all panels, the chain trace from structure model *C* was superimposed onto the displayed electron density. Final average *B* factor for residues of the loop region shown was 80±12 Å^2.



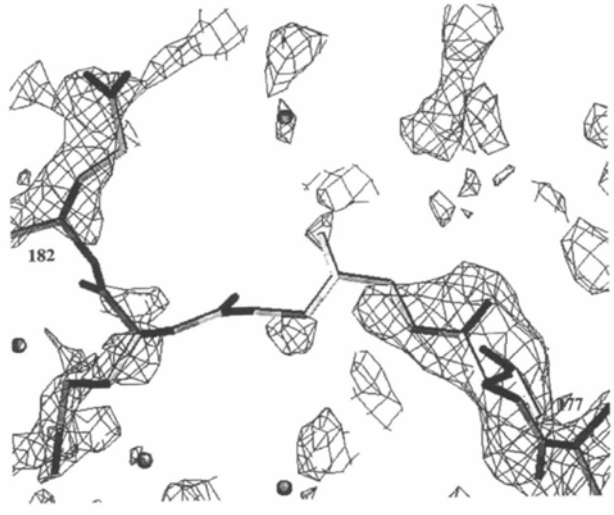
(*a*)



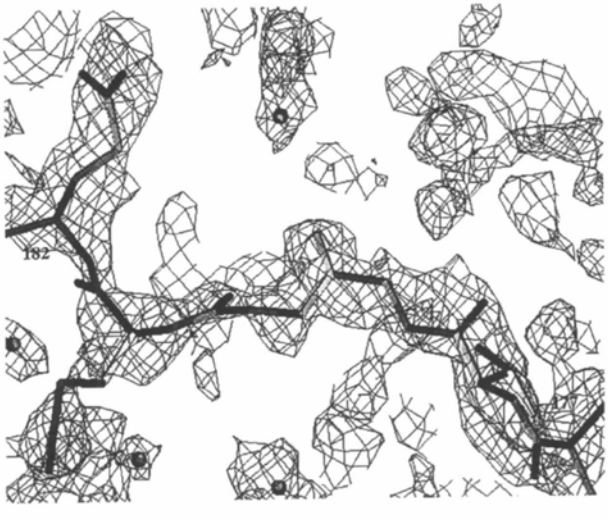
(*b*)



(*c*)



(*d*)



(*e*)

Fig. 2. Electron-density envelope of an $F_o - F_c$ electron-density map at 1.9 Å resolution, contoured at 0.6σ, showing residues 349–353, otherwise difficult to locate in the COOH terminal region (residues 344–363) of class I D-fructose 1,6-bisphosphate aldolase from rabbit muscle. (*a*) Electron-density map corresponding to structure model *A*, where no solvent molecules were used. (*b*) Corresponding to structure model *A′*, model *A* after application of *Shake and Bake* protocol to remove model bias. (*c*) Corresponding to structure model *B* with solvent molecules selected from peaks exceeding 3σ on structure model *A* $F_o - F_c$ difference map. (*d*) As (*c*) with additional bulk-solvent correction. (*e*) Electron-density map phased using model *C* (residues 344–363 omitted). In all panels, the chain trace from structure model *C* was superimposed onto the displayed electron density. Final average *B* factor for residues of COOH terminal region shown was 96±2 Å^2.

References

Baker, E. N., Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D. M. C, Hubbard, R. E., Isaacs, N. W., Reynolds, C. D., Sakabe, K., Sakabe, N. & Vijayan, N. M. (1988). *Philos. Trans. R. Soc. London Ser. B*, **319**, 369–456.

Berthiaume, L., Loisel, T. & Sygusch, J. (1991). *J. Biol. Chem.* **266**, 17092–17105.

Berthiaume, L., Tolan, D. R. & Sygusch, J. (1993). *J. Biol. Chem.* **268**, 10826–10835.

Blom, N. & Sygusch, J. (1997). *Nature Struct. Biol.* **4**(1), 36–39.

Blom, N. & Sygusch, J. (1998). In preparation.

Blom, N., Tétrault, S., Coulombe, R. & Sygusch, J. (1996). *Nature Struct. Biol.* **3**(10), 856–862.

Brünger, A. T. (1992*a*). *Nature (London)*, **355**, 472–474.

Brünger, A. T. (1992*b*). *X-PLOR Version 3.1, A system for X-ray Crystallography and NMR*. Yale University, Connecticut, USA.

Burling, F. T., Weis, W. I., Flaherty, K. M. & Brünger, A. T. (1996). *Science*, **271**, 72–77.

Dobeli, H., Itin, C., Meier, B. & Certa, U. (1991). *Acta Leidensia*, **60**(1), 135–140.

Drenth, J. (1994). *Principles of Protein Crystallography*. New York: Springer-Verlag.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Grazi, E., Cheng, T. & Horecker, B. L. (1962). *Biochem. Biophys. Res. Commun.* **7**, 250–253.

Hannappel, E. MacGregor, J. S., Davoust, S. & Horecker, B. L. (1974). *Arch. Biochem. Biophys.* **214**, 293–298.

Hendrickson, W. A., Smith, J. L., Phizackerley, R. P. & Merritt, E. A. (1988). *Proteins*, **4**, 77–88.

Humphreys, L., Reid, S. & Masters, C. (1986). *Int. J. Biochem.* **18**, 7–13.

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

Jorgensen, W., Chandrasekar, J., Madura, J., Impey, R. & Klein, M. (1983). *J. Chem. Phys.* **79**, 926–935.

Karplus, P. A. & Faerman, C. (1994). *Curr. Opin. Struct. Biol.* **4**, 770–776.

Kleywegt, G. J. & Jones, T. A. (1996). *Acta Cryst.* D**52**, 829–832.

Kleywegt, G. J. & Jones, T. A. (1998). *Methods in Enzymology*, edited by R. M. Sweet & C. W. Carter. In the press.

Levitt, M. & Park, B. H. (1993). *Structure*, **1**(4), 223–226.

McRee, D. E. (1993). *Practical Protein Crystallography*. London: Academic Press.

Sygusch, J., Beaudry, D. & Allaire, M. (1990). *Arch. Biochem. Biophys.* **283**, 227–233.

Sygusch, J., Beaudry, D. & Allaire, M. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 7846–7850.