Acta Cryst. (1995). D51, 626–628

On the Application of Phase Relationships to Complex Structures. XXXV. Some Experiments with 2-Zn Insulin

By M. Mukherjee

Department of Solid State Physics, Indian Association for the Cultivation of Science, Jadavpur, Calcutta-700 032, India

and M. M. Woolfson

Department of Physics, University of York, York YO1 5DD, England

(*Received* 4 *August* 1994; *accepted* 22 *November* 1994)

Abstract

The direct-methods program *SAYTAN* has been applied to the known structure of 2-Zn insulin with 806 atoms, excluding solvent, in the asymmetric unit. Useful sets of phases can be obtained and selected by figures of merit for data resolutions between 1.5 and 2.25 Å and these can be extended by *SAYTAN* to give mean phase errors of 68° for more than 2000 reflections. A feature of the phases so found is that the phase errors decrease with increasing resolution – which is the opposite of the situation when phases are found by isomorphous-replacement techniques.

Introduction

It was shown by Woolfson & Yao (1990) that the direct-methods program *SAYTAN* (Debaerdemaeker, Tate & Woolfson, 1988) could be successfully applied to the solution of the small protein avian pancreatic polypeptide (aPP). This structure has 36 amino acids, 80 water molecules and a Zn atom in the asymmetric unit, space group $C2$ with $a = 34.18$, $b = 32.92$, $c = 28.44$ Å, $\beta = 105.3°$ and data to 0.98 Å resolution (Glover *et al.*, 1983). By artificial truncation of the data Mukherjee & Woolfson (1993) showed that useful sets of phases for aPP could be found even with a data resolution as low as 3.0 Å although figures of merit were incapable of recognizing the better sets below 2.0 Å resolution.

We now report on some further experiments with the larger protein 2-Zn insulin (INS) which has 806 non-H atoms, excluding solvent, in the asymmetric unit, space group $R3$ with $a = 82.5$, $c = 34.0$ Å and with data to 1.5 Å resolution (Baker *et al.*, 1988).

Experimental procedure

Our first experiments consisted of applying *SAYTAN* to INS data in the normal way, generating 1000 trial sets of starting phases by the magic-integer approach (White & Woolfson, 1975). At all resolutions, even at 1.5 Å, no solution was found for any of the 1000 trials with a mean phase error (MPE) less than 80°. We then adopted two procedures which are not a normal part of *SAYTAN*. The first of these, proposed by Woolfson & Yao (1990), was to keep fixed the phases of the 50 reflections with the largest $|E|$ values until the last cycle of refinement when they were allowed to relax to fit in with other phases. The second procedure was to use five cycles of the parameter-shift method (Debaerdemaeker & Woolfson, 1983) to improve the initial phases prior to starting the *SAYTAN* process, as described by Mukherjee & Woolfson (1993). In the *SAYTAN* process the weighting-scheme option used with the tangent formula was that proposed by Hull & Irwin (1978).

Results

Table 1 gives a summary of our results for INS at various resolutions. We have designated 'good' sets as those with an MPE less than 71° for experience has shown that this is about the critical value of initial MPE for phase extension and refinement for structures of this complexity although a larger value can sometimes give a successful outcome (*e.g.* Refaat & Woolfson, 1993).

It is clear that the two procedures we apply prior to *SAYTAN* application make a difference and useful phase sets are obtained even at 2.25 Å resolution. There are 35 good solutions at 1.5 Å resolution and the best of these gives a 61° MPE.

In order to apply *SAYTAN* to an unknown protein structure it is necessary to have a figure of merit (FOM) which will discriminate in favour of the better phase sets. Conventional *MULTAN* FOM's are ineffective for larger structures and we have suggested some alternative ones – ABSM, PSIM and RESM (Mukherjee & Woolfson, 1993). These are,

$$\mathrm{ABSM} = s/s_{\mathrm{exp}}, \tag{1}$$

$$\mathrm{PSIM} = \sum_{\mathbf{l}} \left| \sum_{\mathbf{k}} E(\mathbf{k})E(\mathbf{h}-\mathbf{k}) \right| / s, \tag{2}$$

© 1995 International Union of Crystallography
Printed in Great Britain - all rights reserved

Table 1. *The results of applying SAYTAN to INS at different resolutions*

NREF is the number of large $|E|$ reflections used in *SAYTAN*. NREL the number of three-phase relationships. MAXSET the number of trial sets of initially random phases. NG the number of sets generated with MPE <71°. LMPE is the lowest MPE for the NG good sets.

Resolution (Å)	NREF	NERL	MAXSET	NG	LMPE (°)
1.5	800	13766	1000	35	61
1.77	600	11426	1000	17	70
2.0	500	9823	1000	5	69
2.25	500	11957	700	10	71

and

$$\mathrm{RESM} = \sum_{\mathbf{h}} |\{[\alpha(\mathbf{h})]/s\} - \{[\alpha(\mathbf{h})_{\mathrm{exp}}]/s_{\mathrm{exp}}\}| \times 100, \quad (3)$$

where $\alpha(\mathbf{h}) = |\sum_{\mathbf{k}} E(\mathbf{k})E(\mathbf{h}-\mathbf{k})|$, $s = \sum_{\mathbf{h}} \alpha(\mathbf{h})$, the subscript exp corresponds to the value with true phases and where summations over **h** are for large $|E|$'s and summations over **l** are for small $|E|$'s.

The figure of merit PSIM is a substitute for the conventional *MULTAN* figure of merit PSIZERO and, from general theoretical considerations, was expected to be small. It was noted by Mukherjee & Woolfson (1993) that PSIM was not a reliable indicator of good phase sets and, indeed, for the structure aPP at resolutions below 2.0 Å it was larger values of PSIM which indicated good phase sets. Thus, at 1.5 Å correct phases gave a value of PSIM = 0.41 with other phase sets giving between 0.49 and 0.52 with little obvious correlation with MPE. On the other hand at 2.0 Å resolution correct phases gave 0.31, good phase sets gave 0.38–0.40 while poorer phase sets gave 0.16–0.23. The same pattern was found at 2.25 Å with values of 0.33, 0.44–0.46 and 0.17–0.26, respectively.

Accepting the aPP results, that larger values of PSIM indicate better phase sets, we calculated a combined figure of merit,

$$\begin{aligned}\mathrm{CFOM2} = {} & w_1 \frac{(\mathrm{ABSM} - \mathrm{ABSM}_{\mathrm{min}})}{(\mathrm{ABSM}_{\mathrm{max}} - \mathrm{ABSM}_{\mathrm{min}})} \\ & + w_2 \frac{(\mathrm{PSIM} - \mathrm{PSIM}_{\mathrm{min}})}{(\mathrm{PSIM}_{\mathrm{max}} - \mathrm{PSIM}_{\mathrm{min}})} \qquad (4) \\ & + w_3 \frac{(\mathrm{RESM}_{\mathrm{max}} - \mathrm{RESM})}{(\mathrm{RESM}_{\mathrm{max}} - \mathrm{RESM}_{\mathrm{min}})},\end{aligned}$$

which also discriminates in favour of large values of ABSM and small values of RESM. In this application we have taken weights $w_1 = w_2 = w_3 = 1.0$.

Table 2. *A selection of results from an application of SAYTAN at* 1.5 Å *resolution*

ISET, the number of the trial. ABSM, PSIM & RESM, the figures of merit given in (1), (2) and (3). MPE, the mean phase error. D, the mean deviation of three-phase invariants from zero or π. CFOM2, the figure of merit given in (4).

ISET	ABSM	PSIM	RESM	D (°)	CFOM2	MPR (°)
32	1.23	2.44	37.5	29.2	2.37	65.1
52	0.31	0.64	48.5	48.5	0.01	84.6
257	1.58	1.99	31.7	26.6	2.74	64.7
357	0.41	0.68	39.1	43.9	0.72	84.3
418	1.34	2.25	34.0	34.0	2.56	64.2
517	0.53	0.63	38.2	43.9	0.77	84.0
728	0.37	0.65	37.6	43.9	0.69	84.0
932	0.97	2.29	37.1	36.0	2.10	61.2

Table 3. *Phase extension and refinement by SAYTAN to* 1.5 Å *resolution*

M is the final number of reflections with determined phases. NOT is the number of trials with $M - m$ initially random phases. NG is the number of good sets of M phases. LMPE is the lowest MPE for the NG good sets.

(*a*) A selection of m phases from solution 932 (Table 2) corresponding to the largest values of α

		Initial resolution 1.5 Å		
m	M	NOT	NG	LMPE (°)
300	1364	37	27	69
400	1818	29	29	67
500	2273	7	7	68

(*b*) A selection of m phases from a good solution with 2.0 Å data

		Initial resolution 2.0 Å		
m	M	NOT	NG	LMPE (°)
300	1364	75	21	70
400	1818	1	1	69

Table 2 shows a selection of results for INS at 1.5 Å resolution, one half corresponding to better phases and the other half to poor phase sets. The better phase sets are clearly indicated by CFOM2 at 1.5 Å resolution and the same pattern is found at higher resolutions. For the better phase sets there is also shown the value of the figure of merit, D, whose value is the mean deviation of three-phase invariants from the nearer of zero or π (Mukherjee & Woolfson, 1993). A value greater than about 15° indicates that the phase set should give good enantiomorph discrimination.

The next experiments involved phase extension and refinement and here we followed the general procedure described by Woolfson & Yao (1990). We started with the phases of set 932 (Table 2), the 'good' set with the largest value of D. The m phases with the largest values of α are selected as a basis for the refinement and extension and another $M - m$ reflections with large values of $|E|$ are given random values prior to *SAYTAN* refinement. Several trials were run with different initial random phases for the $M - m$ additional reflections. Table 3(*a*) shows the results of trials with different values of m and M. It is clear that the phase extension and refinement process is quite successful and useful phase sets are obtained for more than 2000 reflections.

The phase extension process associated with Table 3(*a*) involves extending the number of determined phases within the initial resolution limit. Table 3(*b*) shows the results of more conventional extension from phases determined at 2.0 to 1.5 Å. The final MPE's for the extended sets are only a little worse than those obtained starting with phases at 1.5 Å resolution.

Table 4. *The mean phase error as a function of* $(\sin\theta)/\lambda$

The reflections were ranked in order of resolution. In (*a*) the first entry shows the mean value of $(\sin\theta)/\lambda$ and mean phase error for 1–100 in the list, the second entry the same quantities for 51–150 in the list and so on. In (*b*) the first entry shows the mean value of $(\sin\theta)/\lambda$ and mean phase error for 1–200 in the list, the second entry the same quantities for 101–300 and so on.

(*a*) For 700 reflections of set 932 (see Table 2)

$\langle(\sin\theta)/\lambda\rangle$	MPE (°)
0.112	68.5
0.148	66.8
0.172	69.8
0.192	64.9
0.209	69.3
0.224	72.8
0.240	64.5
0.255	64.1
0.269	59.1
0.280	50.4
0.291	50.8
0.301	51.0
0.312	48.7

(*b*) For 1800 reflections of the final set in Table 3(*b*)

$\langle(\sin\theta)/\lambda\rangle$	MPE (°)
0.113	75.1
0.146	74.58
0.168	67.0
0.185	66.3
0.201	66.1
0.214	69.2
0.227	69.2
0.238	67.0
0.250	63.4
0.261	59.7
0.271	64.0
0.288	67.2
0.289	61.8
0.297	57.5
0.306	60.0
0.322	61.1

An interesting feature of the phases so found is that in general, although with some statistical fluctuation, the mean phase error increases with decreasing resolution, as can be seen in Table 4. This is the opposite of the trend found for phases determined from isomorphous-replacement methods where at higher resolution, because of a breakdown of isomorphism, the estimated phases tend to be less reliable. The implication of this pattern has yet to be determined but it may give a less favourable starting condition for the usual techniques of phase extension and refinement.

Discussion

The best maps obtained from the direct-methods approaches will have conventional map correlation coefficients of the order 0.45 (Lunin & Woolfson, 1993) and so should benefit from further phase refinement and extension by processes such as *SQUASH* (Zhang & Main, 1990*a*,*b*) or low density elimination (Shiono & Woolfson, 1992; Refaat & Woolfson, 1993) before model fitting is attempted. Both the applications of direct methods to proteins we have reported, to aPP and INS, have been on structures containing Zn and the question we have to answer is whether structures of similar size, without heavy atoms, would also give useful results. We have also wondered whether the unexpected behaviour of the figure of merit PSIM has also been due to the presence of the heavy atoms. This will be the subject of our next investigation.

Nevertheless, on the basis of our experiments so far it can be claimed that, for structures with up to 100 atoms in the asymmetric unit and containing some heavier atoms, the approach we have described can possibly give useful initial sets of phases for further development as long as the data resolution is not much below 2.0 Å.

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**, 469–456.

Debaerdemaeker, T., Tate, C. & Woolfson, M. M. (1988). *Acta Cryst.* A**44**, 353–357.

Debaerdemaeker, T. & Woolfson, M. M. (1983). *Acta Cryst.* A**39**, 193–196.

Glover, I., Haneef, I., Pitts, J., Wood, S., Moss, T., Tickle, I. & Blundell, T. (1983). *Biopolymers*, **22**, 293–304.

Hull, S. E. & Irwin, M. J. (1978). *Acta Cryst.* A**34**, 863–870.

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

Mukherjee, M. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 9–12.

Refaat, L. S. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 367–371.

Shiono, M. & Woolfson, M. M. (1992). *Acta Cryst.* A**48**, 451–456.

White, P. & Woolfson, M. M. (1975). *Acta Cryst.* A**31**, 53–56.

Woolfson, M. M. & Yao, J.-X. (1990). *Acta Cryst.* A**46**, 409–413.

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

Zhang, K. Y.-J. & Main, P. (1990*b*). *Acta Cryst.* A**46**, 377–381.