Glu46 carboxylate and guanine N1, N2, and by the long and consequently weak guanine N2H···O Asn98 hydrogen bond of 3·3 Å (Table 4).

This work was supported by the Deutsche Forschungsgemeinschaft (Schwerpunkt 'Biophysik der Zelle') and Fonds der Chemischen Industrie. The authors are grateful to Luitpold-Werke, München, for generously providing the extract of *A. oryzae* used in this work.

### References

ALTONA, C. & SUNDARALINGAM, M. (1972). *J. Am. Chem. Soc.* **94**, 8205–8212.
ARNI, R., HEINEMANN, U., TOKUOKA, R. & SAENGER, W. (1988). *J. Biol. Chem.* **263**, 15358–15368.
Biosym Technologies (1988). *DISCOVER User Manual.* Version 2.5. Biosym Technologies, San Diego, CA, USA.
Chemical Design Ltd (1990). *CHEMX.* Chemical Design Ltd, Oxford, England.
DATTAGUPTA, J. K., FUJIWARA, T., GRISHIN, E. V., LINDNER, K., MANOR, P. C., PIENIAZEK, N. J., SAENGER, W. & SUCK, D. (1975). *J. Mol. Biol.* **97**, 267–271.
FINZEL, B. C. (1987). *J. Appl. Cryst.* **20**, 53–55.
FITZGERALD, P. M. G. (1988). *J. Appl. Cryst.* **21**, 273–278.
FÜLLING, R. & RÜTERJANS, H. (1978). *FEBS Lett.* **88**, 279–282.
HEINEMANN, U. (1982). PhD Thesis, Univ. of Göttingen, Germany.
HEINEMANN, U. & HAHN, U. (1989). *Protein–Nucleic Acid Interaction*, edited by W. SAENGER & U. HEINEMANN, pp. 111–141. London: Macmillan.
HEINEMANN, U. & SAENGER, W. (1982). *Nature (London)*, **299**, 27–31.
HENDRICKSON, W. A. (1985). *Methods Enzymol.* **115**, 252–270.
HENDRICKSON, W. A. & KONNERT, J. H. (1980). *Computing in Crystallography*, edited by R. DIAMOND, S. RAMASESHAN & K. VENKATESAN, pp. 13.01–13.23. Bangalore: Indian Academy of Sciences.
HIRSCHFELD, F. L. (1968). *Acta Cryst.* A**24**, 301–311.
HOFFMANN, E., SCHMIDT, J., SIMON, J. & RÜTERJANS, H. (1988). *Nucleosides Nucleotides*, **7**, 757–761.
INAGAKI, F., SHIMADA, I. & MIYAZAWA, T. (1985). *Biochemistry*, **24**, 1013–1020.
JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
KENNARD, O. & HUNTER, W. N. (1989). *Q. Rev. Biophys.* **22**, 327–379.
KOEPKE, J., MASLOWSKA, M., HEINEMANN, U. & SAENGER, W. (1989). *J. Mol. Biol.* **206**, 475–488.
KOSTREWA, D., CHOE, H.-W., HEINEMANN, U. & SAENGER, W. (1989). *Biochemistry*, **28**, 7592–7600.
LESK, A. M. & HARDMAN, K. D. (1982). *Science*, **216**, 539–540.
MALIN, R., ZIELENKIEWICZ, P. & SAENGER, W. (1991). *J. Biol. Chem.* In the press.
MARTINEZ-OYANEDEL, J., CHOE, H.-W., HEINEMANN, U. & SAENGER, W. (1991). Submitted.
NORTH, A. C. T., PHILLIPS, D. C. & MATHEWS, F. S. (1968). *Acta Cryst.* A**24**, 351–359.
OSHIMA, T. & IMAHORI, K. (1972). *J. Biochem. (Tokyo)*, **70**, 197–199.
PAVLOVSKY, A. G. & KARPEISKY, M. Y. (1989). *Structure and Chemistry of Ribonucleases*, edited by A. PAVLOVSKY & K. POLYAKOV, pp. 303–314. Moscow: Academy of Sciences of the USSR.
RABINOVICH, D. & SHAKKED, Z. (1984). *Acta Cryst.* A**40**, 195–200.
ROSSMANN, M. G. (1972). Editor. *The Molecular Replacement Method.* New York: Gordon and Breach.
SAENGER, W. (1984). *Principles of Nucleic Acid Structure.* New York: Springer.
SHIMADA, I. & INAGAKI, F. (1990). *Biochemistry*, **29**, 757–764.
SUGIO, S., OKA, K.-I., OHISHI, H., TOMITA, K.-I. & SAENGER, W. (1985). *FEBS Lett.* **183**, 115–118.
TAKAHASHI, K. & MOORE, S. (1982). *The Enzymes*, **15**, 435–468.

*Acta Cryst.* (1991). B**47**, 527–535

# Solution of the Structure of *Aspergillus niger* Acid α-Amylase by Combined Molecular Replacement and Multiple Isomorphous Replacement Methods

BY R. L. BRADY, A. M. BRZOZOWSKI,* Z. S. DEREWENDA,† E. J. DODSON AND G. G. DODSON

*Department of Chemistry, University of York, Heslington, York YO1 5DD, England*

(*Received* 11 *September* 1990; *accepted* 6 *February* 1991)

## Abstract

The crystal structure of *Aspergillus niger* acid α-amylase was solved by a combination of multiple isomorphous replacement and molecular replacement methods. The atomic coordinates of *Aspergillus oryzae* (TAKA) α-amylase (entry 2TAA in the Protein Data Bank) and experimental diffraction data from a new monoclinic crystal form of TAKA α-amylase, were used during the procedure. Sequence identity between the two proteins is approximately 80%. The atomic parameters derived from the molecular replacement solution were too inaccurate to initiate least-squares crystallographic refinement. The molecular model was extensively revised against the experimental electron density map calculated at 3 Å resolution. Subsequent crystallographic refinement of this model using synchrotron

* Permanent address: Department of Crystallography, Institute of Chemistry, University of Łódź, Łódź, Poland.

† Present address to which all correspondence should be sent: MRC Group in Protein Structure and Function, Department of Biochemistry, 474 Medical Sciences Building, University of Alberta, Edmonton, Canada T6G 2H7.

0108-7681/91/040527-09$03.00 

data to 2·1 Å resolution led to a conventional *R* factor of 16·8%. The structure conforms well to expected stereochemistry with bond lengths deviating from target values by 0·031 Å, and planar groups showing a root-mean-square deviation from ideal planes of 0·025 Å.

## Introduction

α-Amylases (α-1,4-glucan-4-glucanohydrolase, EC 3.2.1.1) constitute a widely distributed group of enzymes found in bacteria, plants and animal secretions. They hydrolyse the α(1–4) glycosidic linkage in amylose, amylopectin, glycogen and other polysaccharides. Amino-acid sequences of amylases from different sources were recently compared (MacGregor, 1988) and demonstrated pronounced variability. Although successful crystallization was reported for several amylases, only two X-ray diffraction studies have been reported: an enzyme from *Aspergillus oryzae* (TAKA amylase) at 3 Å resolution (Matsuura, Kunusoki, Harada & Kakudo, 1984) and porcine pancreatic amylase (PPA) at 2·9 Å resolution (Buisson, Duee, Haser & Payan, 1987). Owing to the relatively low resolution neither of the atomic models was subjected to full crystallographic refinement. The molecules of both enzymes exhibit the $(\alpha/\beta)_8$ folding pattern, which was first identified in triose phosphate isomerase (TIM) (Phillips, Sternberg, Thornton & Wilson, 1978) and has since been referred to as the 'TIM barrel'. This fold has also been observed in 17 other, frequently unrelated enzymes, and has been recognized as one of the important patterns of protein folding (Farber & Petsko, 1990).

Although the location of the active site in the TAKA amylase molecule has been identified by low-resolution X-ray studies (Matsuura *et al.*, 1984), detailed studies of the structural basis of the enzymatic activity and of the molecular architecture necessitate crystallographic refinement of the atomic model at high resolution. In order to achieve this goal we have undertaken the study of an acid α-amylase from *Aspergillus niger* (NOVO amylase). It shows *ca* 80% sequence identity with TAKA amylase and the availability of the TAKA atomic model justified an attempt to solve the structure by molecular replacement methods. However, because of the relative inaccuracy of the atomic coordinates (the TAKA model was not refined against X-ray data) we decided to use diffraction data from a new and well-diffracting crystal form of TAKA amylase, grown in our laboratory. Heavy-atom derivatives of NOVO amylase were also prepared in case the molecular replacement did not work, and to ensure a genuinely independent structure solution.

In this paper we describe the crystallographic details of structure solution using both multiple isomorphous replacement (MIR) and molecular replacement methods, as well as the details of phase combination, phase extension and least-squares crystallographic refinement at 2·1 Å resolution. An atomic model of high accuracy was obtained, as illustrated by the conventional crystallographic *R*-factor value of 16·8%. Some structural features including the amino-acid sequence of NOVO amylase and the calcium-binding stereochemistry, have been described elsewhere (Boel *et al.*, 1990).

## Methods and results

### *Crystals*

The purification of NOVO and TAKA amylases has been described elsewhere [Boel *et al.*, 1990; see also the following paper in this volume (Swift *et al.*, 1991)]. Single crystals of the NOVO amylase were grown in batch from solutions containing 14–15% (*w*/*v*) polyethylene glycol ($M_r$ = 8000) and a protein concentration of 50 mg ml$^{-1}$ in 1 *M* sodium acetate (pH 3·0). The diffraction pattern extended to approximately 2·5 Å on a conventional source, and beyond 1·7 Å on the wiggler synchrotron station ($\lambda$ = 0·88 Å) in the SERC Daresbury Laboratory, England (Helliwell *et al.*, 1982). The crystals exhibit the symmetry of space group $C222_1$ with unit-cell dimensions $a$ = 81·1, $b$ = 98·3, $c$ = 138·0 Å. There is one molecule per asymmetric unit.

The crystallization of TAKA amylase has been achieved through hanging-drop experiments. Monoclinic crystals were grown using the hanging-drop technique (McPherson, 1989); 50 μl volumes of protein solution (30 mg ml$^{-1}$) in 50 m*M* sodium acetate, pH = 6·0, 2 m*M* $CaCl_2$, 8%(*w*/*v*) polyethylene glycol 8000 were equilibrated against a reservoir solution of 16% polyethylene glycol 8000 in the same buffer. The crystals took *ca* 14 days to grow, with one single crystal (2·0 × 0·3 × 0·3 mm) suitable for further X-ray studies (space group $P2_1$, with two molecules in the asymmetric unit, $a$ = 75·0, $b$ = 104·3, $c$ = 67·4 Å, $\beta$ = 104·5°). In an attempt to obtain more crystals suitable for high-resolution X-ray study batch crystallization was set up. Although conditions were analogous to those used in the hanging-drop experiments, a different, orthorhombic crystal form was observed. The high-resolution (2·1 Å) X-ray analysis of this form is now completed (see the following paper).

### *Heavy-atom derivatives*

Four useful heavy derivatives were obtained for NOVO amylase. A mercurial derivative was prepared by cocrystallization in the presence of a 2 m*M* solution of $HgCl_2$. Others were prepared by soaking native crystals in 20%(*w*/*v*) polyethylene glycol solu-

Table 1. *Xentronics area-detector medium-resolution diffraction data details*

A scaled-tube conventional source operated at 50 kV and 35 mA was used in this experiment. Each data set was collected from a single crystal.

| Data set | Resolution (Å) | No. of reflections | $R_{merge}$ (%) | No. of observations | $\langle I/\sigma(I)\rangle$ |
|---|---|---|---|---|---|
| Native | 3·0 | 11010 | 4·27 | 34547 | 39·5 |
| TAKA | 4·0 | 7491 | 4·82 | 20491 | 53·8 |
| Hg derivative | 3·0 | 10846 | 6·66 | 33555 | 22·1 |
| Hg* derivative | 6·0 | 1686 | 4·37 | 23000 | 264·1 |
| Pb derivative | 3·0 | 11060 | 6·88 | 37189 | 46·4 |
| Sm derivative | 2·8 | 13207 | 5·54 | 39818 | 32·4 |
| Pt derivative | 5·4 | 1709 | 5·03 | 6713 | 60·7 |

* Denotes a data set used to obtain accurate anomalous differences.

Table 2. *High-resolution synchrotron photographic data details*

A total of five crystals were used to collect the full data set.

| | |
|---|---|
| Total No. of packs | 80 |
| Observations | |
| Total input | 210327 |
| Total used | 167449 |
| Fully recorded | 143963 |
| Partially recorded | 23486 |
| No. of independent reflections | 31225 |
| No. of reflections with negative intensities (after merging) | 256 |
| $\langle I/\sigma(I)\rangle$ | 9·2 |
| Overall $R_{merge}$ for all symmetry-related reflections (on intensities) | 0·055 |
| $R_{merge}$ for weak intensities [$I < 5{\cdot}4\sigma(I)$] | 0·122 |
| $R_{merge}$ for reflections beyond 2·2 Å | 0·108 |
| Percentage of reflections > $3\sigma$ | 95·2 |

tions containing, respectively, $Sm(OOCCH_3).3H_2O$, $K_2PtCl_6$ and $Pb(OOCH_3).3H_2O$, in 10 m*M* concentrations. In each case the crystals were soaked for 48 h prior to mounting them in the usual way for X-ray data collection.

*Data collection and processing*

Two techniques of X-ray data collection were employed. Low- and medium-resolution (< 3 Å) X-ray data intended for MIR or molecular replacement calculations were collected for NOVO amylase native and derivative crystals as well as TAKA native monoclinic form (4 Å) using the Nicolet/Xentronics imaging proportional chamber (IPC) and the *XENGEN* suite of data-processing programs (Howard, Gilliland, Finzel, Poulos, Ohlendorf & Salemme, 1987). The use of this system mounted on a conventional source has been described (Derewenda & Helliwell, 1989) along with a detailed description of NOVO amylase data collection. Table 1 summarizes these results.

High-resolution (2·1 Å) X-ray data (subsequently used for the least-squares crystallographic refinement of the atomic model of the NOVO amylase) were recorded on film using the Arndt–Wonacott camera at the wiggler station (9.6) of the SERC Daresbury Laboratory, England, using 0·88 Å wavelength (Helliwell *et al.*, 1982). The crystals were mounted with the *c* axis along the capillary. Two data sets were collected. A medium-resolution data set (3 Å) was recorded at a distance of 20 cm using oscillations of 3° per exposure and 25 s exposure times. One crystal only was used and no attempt was made to fill in the blind region (which at 3 Å resolution is very small). A high-resolution (2·1 Å) data set was recorded at a distance of 13 cm using 2° oscillations and exposure times of 100 s. Three crystals were used, and an additional subset was collected from a crystal mounted along the *a* axis to cover the blind region. Three films per pack were used in all cases. This approach allowed the strong intensities (particularly at low and medium resolution), which were frequently saturated in the high-resolution data set, to be measured accurately using the low-resolution films. The complete set of data was processed in the usual way. The films were scanned using a Joyce–Loebl Scandig-3 densitometer and the data integrated using the *MOSPRF* package (written by A. J. Wonacott, modified by F. Korber) incorporating profile-fitting. Oblique incidence, absorption and Lorentz–polarization corrections were applied and the data scaled. Table 2 gives some additional data.

*Application of the molecular replacement method*

The atomic coordinates deposited for the TAKA amylase in the Protein Data Bank (entry 2TAA) were used in an attempt to solve the structure of the NOVO amylase by molecular replacement procedures. At the time only a partial amino-acid sequence was known for the latter enzyme, but it appeared very homologous (approximately 80% identity) to TAKA. Numerous calculations were performed with both complete and modified atomic models of TAKA. The fast-rotation function (Crowther, 1972) and an *R*-factor minimization search function (Nixon & North, 1976) were used throughout. None of the results were convincing. A prevailing pattern was that any change in the input parameters (most notably the resolution range of the data) would cause drastic change in the appearance of the rotation-functon map. *R*-factor search maps calculated for several possible orientations of the model molecule exhibited shallow minima close to the position expected from packing considerations ($X = 0{\cdot}25$, $Y = 0{\cdot}25$, $Z = 0{\cdot}5$), but none were convincing. It was therefore necessary to obtain clear evidence for the correct solution. To this end a 4·0 Å resolution X-ray data set collected from a monoclinic crystal of TAKA amylase was used. The volume of the asymmetric unit clearly pointed to the presence of two molecules. A self-rotation function confirmed this and identified the position of the diad as parallel to *a*. It should follow, that if the two molecules are related by a matrix [*S*], then cross rotations between TAKA model and observed data, as well as between

TAKA and NOVO observed diffraction data, should each give two solutions related by [*S*]. Fig. 1 illustrates this argument. A set of transformations satisfying the above argument was likely to be the true solution.

Two model structure-factor sets were calculated: one from a complete atomic model of TAKA and the other using a molecule in which all the surface side chains and the C-terminal domain were deleted. Each was placed in a *P*1 cell with axial lengths $a = b = c = 80$ Å. Cross-rotation-function maps were calculated for resolution ranges 20·0–7·0, 8·0–5·0 and 6·0–4·0 Å; there was little consistency in the results obtained for the last shell.

Table 3 summarizes the results. One tranformation of a model molecule into the experimental NOVO amylase unit cell satisfying all the requirements was identified (in Eulerian space: $\alpha = 14$, $\beta = 10$, $\gamma = 10°$). The truncated model was rotated accordingly and placed in a *P*1 unit cell with observed axial lengths; an *R*-factor search function close to the expected position was calculated in the usual way. Fig. 2 shows the result, and the location of the minimum *R*-factor value. The model molecule was translated accordingly. Unfortunately subsequent least-squares refinement did not converge satisfactorily and no improvement in the quality of the electron density map (which was fairly ambiguous) was achieved.

### *Multiple isomorphous derivative phasing at* 3·0 Å *resolution*

Because there were some difficulties with heavy-atom derivative preparations resulting from the low pH of the crystals, a single-site Hg derivative was used in an initial attempt to solve the structure by single isomorphous replacement with anomalous scattering (SIRAS) phasing. The single Hg site was identified from an isomorphous difference Patterson

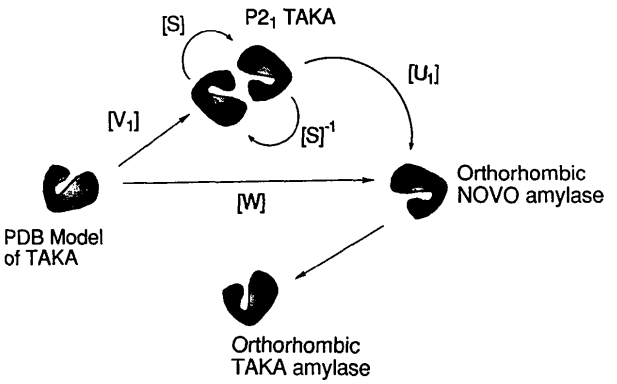


Fig. 1. A diagrammatic representation of the correlations between the rotation matrices relating the three crystal forms investigated in this study, and the TAKA amylase coordinates.

Table 3. *Results of the rotation-function calculations*

The results are given as Eulerian angles (°). For other details refer to Fig. 3 and the text.

| Type of operation | Resolution shell (Å) | | α | β | γ | R | Comments |
|---|---|---|---|---|---|---|---|
| (*P*2₁)–self | 20–7, 8–5 | | 0 | 180 | 0 | | [*S*] |
| | | | 0 | 180 | 180 | | $[S]^{-1}$ |
| (*P*2₁) × (PDB) | 20–7 | (1) | 30 | 145 | 115 | (97·9) | [$V_1$] |
| | | (2) | 150 | −35 | 115 | (81·7) | $[V_2] = [S][V_1]$ |
| | 8–5 | (1) | 25 | 145 | 105 | (92·9) | [$V_1$] |
| | | (2) | 155 | −35 | 120 | (84·7) | $[V_2] = [S][V_1]$ |
| NOVO × (PDB) | 20–7 | | 35 | 5 | −10 | (51·0) | Note: $\alpha + \gamma = 25°$ [*W*] |
| | 8–5 | | −50 | 15 | 80 | (80·9) | Note: $\alpha + \gamma = 30$ [*W*] |
| (*P*2₁) × NOVO | 20–7 | (1) | 150 | 155 | −80 | (60·1) | [$U_1$] |
| | | (2) | 20 | −155 | 100 | (60·2) | $[U_2] = [S][U_1]$ |
| | 8–5 | (1) | 170 | 145 | −70 | (121·2) | [$U_1$] |
| | | (2) | 10 | −145 | 70 | (104·1) | $[U_2] = [S][U_2]$ |

* For small values of β the sum of α + β, and not their individual values, is of essence. Crosscheck: $[W]_{calc} = [U_2]^{-1}[V_1]$; in Eulerian angles this gives $\alpha = 14$, $\beta = -9$, $\gamma = -171$; by symmetry $(\alpha, -\beta, \pi - \gamma)$ this equals 14, 9, 9 with $\alpha + \beta = 23$.

synthesis and refined using the $F_{hle}$ (heavy-atom structure factor, lower estimate) method (Matthews, 1966). Two data sets were used to obtain phases to 3·0 Å resolution (for details of data collection see Derewenda & Helliwell, 1989). Fourier syntheses were calculated for both of the two possible solutions and solvent flattening (Wang, 1985) was used subsequently to improve the quality of the phasing. One map clearly revealed the position of the protein molecule. When the three additional derivatives were finally prepared by soaking techniques, the SIRAS phases were successfully used in difference Fourier syntheses ($F_{derivative} - F_{native}$) through which all the positions of the heavy atoms were revealed. As the

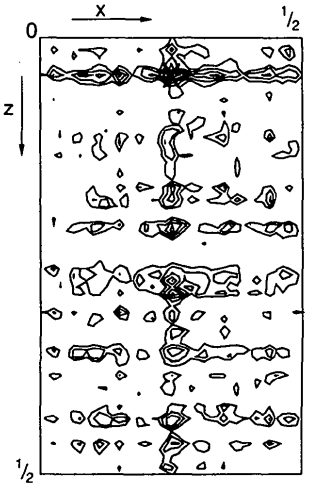


Fig. 2. The *R*-factor search function. Only the *y* section containing the main minimum (section 73 of 400) is shown. The grid in the *xz* plane is close to 1 Å. 505 strongest terms were used. The minimum (0·374) is denoted by the cross, and the grid points are: $X = 20$, $Z = 6$ (which corresponds to 20·3 and 5·9 Å along the crystallographic axes. The map is contoured at equal arbitrary levels.

Table 4. *Heavy-atom parameters and the results of the least-squares refinement*

| | x | y | z | $R_{iso}$ | $R_c$ | B (Å²) | Occupancy |
|---|---|---|---|---|---|---|---|
| Hg | −0·1409 | 0·2343 | 0·0281 | 0·58 | 0·68 | 15·0 | 0·30 |
| Hg* | −0·1384 | 0·2305 | 0·0289 | 0·45 | – | 15·0 | 0·25 |
| Pb | 0·0827 | 0·2115 | 0·0459 | | | 15·0 | 0·19 |
| | 0·8290 | 0·3743 | 0·1237 | | | 15·0 | 0·15 |
| | 0·0011 | 0·1414 | 0·1264 | 0·58 | 0·50 | 15·0 | 0·03 |
| Pt | 0·1258 | 0·0733 | 0·1310 | | | 20·0 | 0·20 |
| | 0·5440 | 0·4848 | 0·0401 | | | 25·0 | 0·08 |
| | 0·5893 | 0·6188 | 0·1580 | | | 25·0 | 0·08 |
| | 0·0804 | 0·1955 | 0·0341 | 0·63 | 0·67 | 25·0 | 0·03 |

* Low-resolution data set collected specifically with anomalous measurements in mind (see Derewenda & Helliwell, 1989); the isotropic temperature factors were held constant; the occupancy is on an arbitrary scale.

Sm derivative sites were identical to those found in the Pb derivative only the latter was used. The Pt derivative was found to be of poor quality (as judged by its phasing power); it was, nonetheless, used for phasing out to approximately 5·4 Å resolution. The final atomic positions of heavy atoms and the results of their crystallographic refinement are given in Table 4. Fig. 3 illustrates the phasing power of the derivatives as a function of resolution. As was expected, the highest figure of merit was obtained when all the derivatives were used. Anomalous contributions, however small, were also included.

*Phase combination, phase extension and least-squares refinement*

An electron density Fourier map, calculated to 3·0 Å resolution using $F_{obs}$ and MIR phase angles weighted by the figures of merit, was found to be interpretable. It was, however, taken through four cycles of solvent flattening, each followed by phase combination. This increased the value of the mean figure of merit, $\langle m \rangle$, to 0·74 (acentric terms) and significantly reduced the noise in the solvent region.

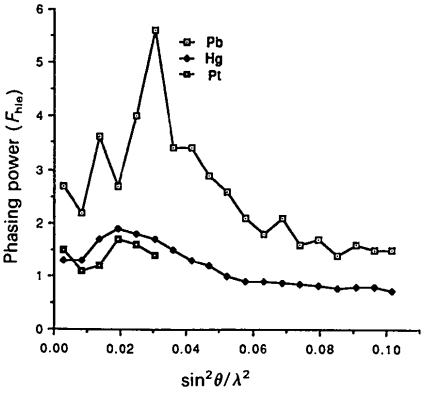


Fig. 3. The phasing power, shown as a function of resolution, of the three principal heavy-atom derivatives used in this study. The visible superiority of the Pb derivative can only be explained in terms of its high isomorphism with the native structure.

Both the raw MIR and the solvent-flattened electron density maps were displayed using an Evans & Sutherland PS300 graphics system and the *FRODO* program (Jones, 1978). When the atomic model obtained by molecular replacement calculations was displayed against this background it was found that it was essentially correct. This made model building relatively easy. The arrival at this stage of the complete amino-acid sequence of NOVO amylase allowed the model to be altered appropriately. The procedure adopted for the interactive building and adjustment of the atomic model may be summarized as follows.

The model was displayed against the map as single residues with no connectivity between the amino acids. Whole residues were shifted into the density sequentially, with bulky aromatic side chains and disulfide bridges providing checkpoints. Only the obvious corrections were made, while doubtful regions were completely deleted. In the early stages care was taken only to position the model within experimental density, with little attention to the detail. During subsequent stages it became clear that some parts were seriously in error; they were deleted and sequence corrections were introduced. Immediately after the first round of model building the 3·0 Å area-detector data were replaced with the 2·1 Å synchrotron data, and the Konnert–Hendrickson (KH) version of geometrically restrained crystallographic refinement (Hendrickson, 1985) with incorporated fast Fourier transform was undertaken. As soon as convergence was approached the atomic model was used to calculate a set of structure factors and phases; phase combination with the original MIR phases was carried out and a new electron density map was computed. Combined phases were used to 3·0 Å resolution and calculated phases weighted by the method of Sim (1959) to 2·8 Å were also included at this stage. The second round of this procedure extended the phasing to 2·6 Å. The subsequent round of model building and refinement reached sufficient accuracy to make it possible to abandon the MIR phases completely in favour of the calculated phases. An electron density map calculated with coefficients $2F_{obs} - F_{calc}$ was calculated in place of the earlier maps. All data to 2·1 Å resolution were used in this calculation. Table 5 gives additional details of this protocol. When the crystallographic $R$ factor fell below approximately 0·25 simultaneous KH refinement of both positional and thermal parameters was introduced. Table 6 shows the stereochemical details of the final model.

## Discussion

Once a structure is solved and refined it is possible to reflect on the nature of the difficulties encountered

Table 5. *Progress of crystallographic least-squares refinement*

Phase angles $\alpha_{comb}$, $\alpha_{calc}$ and $\alpha_{mir}$ refer to combined, calculated and experimental values, respectively. *B* values refer to isotropic temperature factors.

Round 1: starting model based on the molecular replacement solution using TAKA amylase coordinates, extensively reviewed for gross errors *versus* the MIR electron density map.

| | |
|---|---|
| Total No. of atoms | 3576 |
| Solvent atoms | 0 |
| Initial $R_{cryst}$ | 0·570 |
| $R_{cryst}$ after 6 cycles of restrained *xyz* refinement | 0·461 |
| Phase-combination details (acentric terms) | |
| $\langle\lvert\alpha_{calc} - \alpha_{mir}\rvert\rangle$ | 64° |
| $\langle\lvert\alpha_{comb} - \alpha_{mir}\rvert\rangle$ | 28° |
| $\langle\lvert\alpha_{comb} - \alpha_{calc}\rvert\rangle$ | 9° |

Round 2: $F_{obs}\alpha_{comb}$ weighted according to Sim map calculated using combined phases to 3 Å and extended phases to 2·8 Å; manual rebuilding using *FRODO*.

| | |
|---|---|
| Total No. of atoms | 2938 |
| Solvent atoms | 0 |
| Initial $R_{cryst}$ | 0·520 |
| $R_{cryst}$ after 7 cycles of restrained *xyz* refinement | 0·409 |
| Phase-combination details (acentric terms) | |
| $\langle\lvert\alpha_{comb} - \alpha_{mir}\rvert\rangle$ | 61° |
| $\langle\lvert\alpha_{comb} - \alpha_{mir}\rvert\rangle$ | 26° |
| $\langle\lvert\alpha_{comb} - \alpha_{calc}\rvert\rangle$ | 8° |

Round 3: $(2F_{obs} - F_{calc})\alpha_{comb}$ weighted according to Sim map calculated using combined phases to 3 Å and extended phases to 2·6 Å; manual rebuilding using *FRODO*.

| | |
|---|---|
| Total No. of atoms | 3008 |
| Solvent atoms | 0 |
| Initial $R_{cryst}$ | 0·488 |
| $R_{cryst}$ after 7 cycles of restrained *xyz* refinement | 0·378 |
| after 2 cycles of unrestrained *B* refinement | 0·352 |
| Phase-combination details (acentric terms) | |
| $\langle\lvert\alpha_{calc} - \alpha_{mir}\rvert\rangle$ | 57° |
| $\langle\lvert\alpha_{comb} - \alpha_{mir}\rvert\rangle$ | 26° |
| $\langle\lvert\alpha_{comb} - \alpha_{calc}\rvert\rangle$ | 7° |

Round 4: $(2F_{obs} - F_{calc})\alpha_{comb}$ weighted according to Sim map calculated using combined phases to 3 Å and extended phases to 2·1 Å; manual rebuilding using *FRODO*; all *B* values truncated to 12 Å$^2$ prior to further refinement.

| | |
|---|---|
| Total No. of atoms | 3429 |
| Solvent atoms | 53 |
| Initial $R_{cryst}$ | 0·424 |
| $R_{cryst}$ after 6 cycles of restrained *xyz* refinement | 0·334 |
| after 2 cycles of unrestrained *B* refinement | 0·312 |

No phase combination was employed from this stage onwards.

Round 5: $(2F_{obs} - F_{calc})\alpha_{calc}$ was calculated using all data to 2·1 Å; manual rebuilding using *FRODO*; all *B* values truncated to 12 Å$^2$ prior to further refinement.

| | |
|---|---|
| Total No. of atoms | 3694 |
| Solvent atoms | 182 |
| *Initial* $R_{cryst}$ | *0·335* |
| $R_{cryst}$ after 6 cycles of restrained *xyz* refinement | 0·274 |
| after 2 cycles of unrestrained *B* refinement | 0·248 |

Round 6: $(2F_{obs} - F_{calc})\alpha_{calc}$ was calculated using all data to 2·1 Å; manual rebuilding using *FRODO*; *B* values were not truncated.

| | |
|---|---|
| Total No. of atoms | 3986 |
| Solvent atoms | 327 |
| Initial $R_{cryst}$ | 0·271 |
| $R_{cryst}$ after 1 cycle of restrained *xyz* refinement | 0·245 |
| Resolution limits were reset to 7·5–2·1 Å | |
| after 2 cycles of restrained *xyz* refinement | 0·226 |
| after 3 cycles of unrestrained *B* refinement | 0·205 |
| after 3 cycles of restrained *xyz* refinement | 0·195 |
| after 2 cycles of unrestrained *B* refinement | 0·188 |
| after 2 cycles of restrained *xyz* refinement | 0·185 |

Table 5 (*cont.*)

Round 7: $(2F_{obs} - F_{calc})\alpha_{calc}$ was calculated using all data to 2·1 Å; manual rebuilding using *FRODO*; *B* values were not truncated. Simultaneous restrained refinement of positional and thermal parameters was employed.

| | |
|---|---|
| Total No. of atoms | 4029 |
| Solvent atoms | 341 |
| Initial $R_{cryst}$ | 0·186 |
| $R_{cryst}$ after 6 cycles of restrained refinement (*xyz* + *B*) | 0·169 |

Table 6. *Weighting scheme, results and standard deviations for the atomic model after the last cycle of refinement*

| | σ | Standard deviation | No. of parameters |
|---|---|---|---|
| Bond lengths (1–2) (Å) | 0·02 | 0·031 | 3777 |
| Angle distances (1–3) (Å) | 0·04 | 0·078 | 5163 |
| Planar distances (1–4) (Å) | 0·06 | 0·088 | 1442 |
| R.m.s. deviation from planes (Å) | 0·02 | 0·025 | 647 |
| Chiral volume (Å) | 0·12 | 0·209 | |
| Van der Waals contacts (Å) | | | |
| Single-torsion contact | 0·5 | 0·210 | 1167 |
| Multiple-torsion contact | 0·5 | 0·286 | 1644 |
| Possible hydrogen bonds | 0·5 | 0·281 | 331 |
| Torsion angles (°) | | | |
| Peptide plane (ω) | 20·0 | 6·4 | 486 |
| Staggered (60/120°) | 20·0 | 22·5 | 564 |
| Orthonormal (± 90°) | 20·0 | 33·3 | 66 |
| Isotropic temperature factors, *B* (Å$^2$) | | | |
| Main chain (1–2) | 2·0 | 2·48 | 1983 |
| (1–3) | 2·5 | 3·15 | 2518 |
| Side chain (1–2) | 2·5 | 4·40 | 1795 |
| (1–3) | 3·0 | 5·56 | 2652 |

during the study. The MIR phasing coupled with the knowledge of the fundamental architecture of the molecule, and its correct orientation and position in the unit cell (as established by molecular replacement) ensured quick progress of the analysis. The use of phase combination resulted in a very significant improvement of electron density maps in the early stages.

When the original 3·0 Å MIR phases were compared with the final values obtained from atomic model coordinates (Fig. 4) it became clear that they were considerably more accurate than could have been predicted on the basis of the value of $\langle m \rangle$ (0·60). Fig. 5 illustrates the relationship between the phase error, mean intensities and figures of merit. It shows that the mean phase error varies with intensity, but not as dramatically as might be expected. It confirms yet again that the traditional figures of merit are not an accurate measure of the true distribution of error, seriously underestimating the correctly phased reflections while still putting considerable weight on those phased with an error of over 90°. Our subsequent discussion refers to the results of various phasing experiments, and to avoid any bias we use the absolute final calculated phase angles as the target values. The question of proper weights may be addressed separately.

Serious difficulties were encountered when molecular replacement was utilized in an effort to solve the structure using the model of the TAKA amylase. Although the introduction of the additional $P2_1$ form diffraction data made it possible to differentiate between the false maxima and the true solution, the resulting model failed as a starting point for least-squares refinement. The reason for this became clear when this model was compared with the final refined set of coordinates (main-chain atoms only). Over 75% of the atoms were further than 0·7 Å from their target position. The randomly distributed differences between the TAKA model and the NOVO amylase structure yielded the atomic model unrefinable against the NOVO amylase diffraction data, using conventional techniques. It is interesting to note the outcome of the KH refinement of the TAKA model undertaken using the molecular replacement solution as a starting point: although the $R$ factor fell sharply from 0·57 to below 0·40, the mean phase error with respect to the final phases (77° for the initial model calculation) remained practically unchanged (75°). This is a dramatic illustration of the danger of commencing refinement using an inaccurate model.

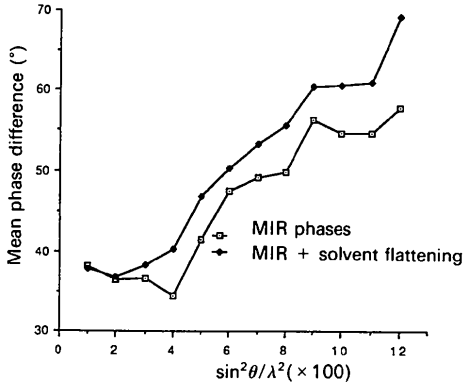


Fig. 4. Comparison of the accuracy of the raw MIR and solvent-flattened MIR absolute phases, with the final values of $\alpha_{calc}$ based on the refined structure. In this, and in the subsequent comparisons, only fully phased (*i.e.* for which native and all derivative observations were available) and observed reflections are shown. This figure illustrates that although solvent flattening results in a clearer and interpretable map, at medium levels of solvent content it generally does not improve the accuracy of phasing.

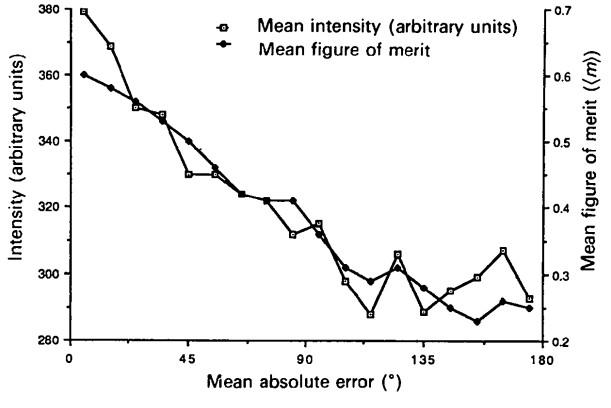


Fig. 5. The relationship between the absolute error of the phase (as determined by the MIR method), the mean intensities of reflections, and the corresponding mean figures of merit. Although there are clear correlations, they are less dramatic than expected. The mean intensity of the reflections phased with an error of 180° is only 20% lower than the intensity of the ideally phased reflections. On the other hand, the mean figure of merit clearly underestimates the accuracy of the phasing, predicting an error of 53° for the ideally phased reflections, and of only 75° for those phased with an error of 180°.

It is nonetheless instructive to investigate the quality of the molecular replacement phases and to reflect on their possible use. Fig. 6 shows the mean difference between the molecular replacement phases and the final values obtained from the refined model, as a function of resolution, and allows a comparison with a similar plot for the MIR phases (Fig. 4). It should be noted that for the low-resolution shell (10–8 Å) the molecular replacement phases are better than those obtained by MIR. Their quality deteriorates rapidly with resolution and at approximately 4 Å the mean difference approaches the random value of 90°. As was already noted, the accuracy of the MIR phases does not seem to change significantly with resolution. Fig. 7 shows a section of a difference Fourier map calculated with coefficients $[F_{h(Pb)} - F_{native}]$ and phase angles calculated from the molecular replacement model, calculated for two different resolution shells of data. As was expected, at low resolution the phasing is accurate enough to reveal the position of the heavy atom with a good signal to noise ratio. This suggests the possibility of using a combination of molecular replacement and SIR methods, in which the molecular replacement

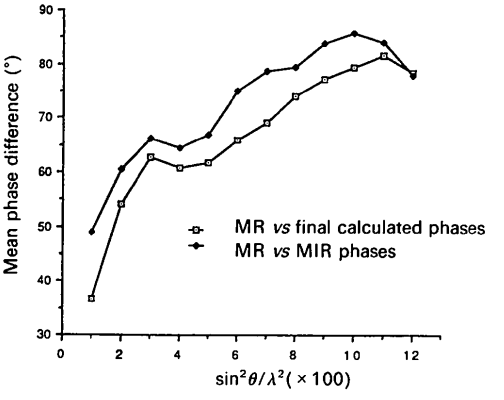


Fig. 6. The accuracy of the molecular replacement phases (MR) (as seen against the values of $\alpha_{calc}$ based on the refined structure) and the comparison of the molecular replacement phases with those obtained by MIR.

phases would be used in a fashion similar to that employed during solvent flattening (Wang, 1985). It seemed that such an approach would have an additional advantage of being able to use a Fourier difference synthesis between the heavy-atom derivative and the native data as a useful check identifying the correct molecular-replacement solution. We chose the Pb-derivative SIR phases as a test case. Fig. 8 shows the analysis of the phasing achieved using raw SIR, combined SIR and solvent flattening, and combined molecular replacement and SIR. It is clear that, at low resolution, molecular replacement plus SIR outperforms the other methods considerably, while deteriorating beyond 4 Å.

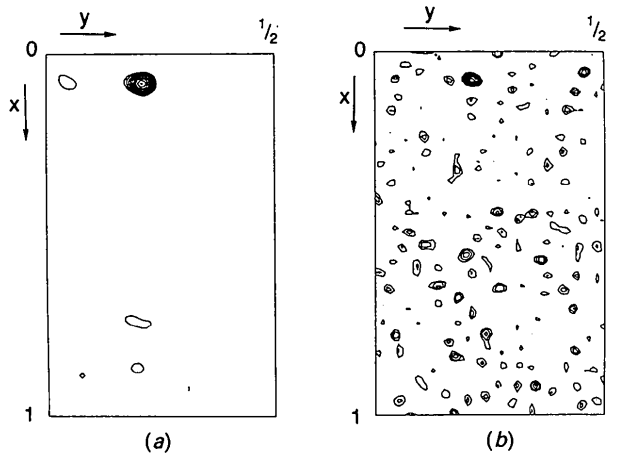


Fig. 7. Difference Fourier maps calculated with the coefficients $[F_{h(Pb)} - F_{native}]$ and phased using phases calculated from the molecular replacement solution: (*a*) in the 10–5 Å resolution range; (*b*) in the 5–3 Å resolution range. Both maps are contoured in the same way, by taking the value of the highest peak as 10, and omitting levels 0 and 1.

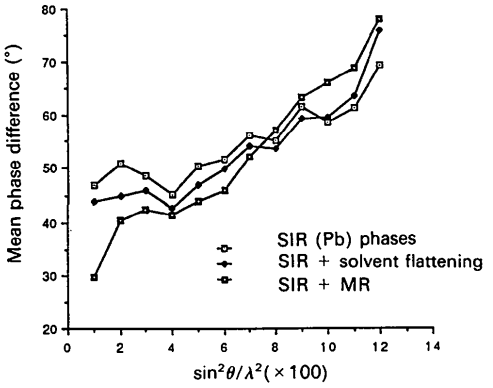


Fig. 8. The accuracy of phases obtained by single isomorphous replacement (SIR) (using the Pb derivative), solvent flattening with SIR, and a combination of SIR with molecular replacement phases. The performance of each method is resolution dependent (for discussion see text).

## Concluding remarks

We have demonstrated that it is practical to use a molecular replacement solution to resolve the ambiguity in the phase determination by the SIR method. The additional advantage of this technique is the possible use of the difference Fourier phased on the molecular replacement model (at low resolution!!) to locate the heavy atoms in the derivative; this eliminates the problem of the correct choice of hand and provides additional confirmation of the correctness of the molecular replacement solution. Better definition of the molecular boundaries in the molecular replacement model results in improved phasing over traditional solvent flattening. The approach we propose was shown to be successful even if the molecular replacement solution is unrefinable by the usual least-squares methods, due to excessive coordinate errors. The Cα coordinates are available from the Protein Data Bank* (Bernstein *et al.*, 1977), while a complete set of atomic coordinates will be released upon publication of the results of structural analysis of the model, which is currently in preparation.

Unless otherwise specified, all the programs referred to in this paper have been used in their VAX VMS versions of the *CCP*4 suite of crystallographic programs (SERC Daresbury Laboratory, 1986).

The amylase structural project at York is supported by Novo–Nordisk Research Laboratories, Copenhagen, Denmark. The protein structure group at York is also supported by a consolidated grant from the Science and Engineering Research Council, England. We are also grateful to the staff of the Daresbury Laboratory for their assistance in data collection.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 2AAA, R2AAASF), and are available in machine-readable form from the Protein Data Bank at Brookhaven. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 37041 (as microfiche). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England. At the request of the authors, the list of structure factors will remain privileged until 1 January 1993.

## References

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F., BRICE, M. D., ROGERS, J. R., KENNARD, O., SHIMANOUCHI, T. & TASUMI, M. J. (1977). *J. Mol. Biol.* **112**, 535–542.

BOEL, E., BRADY, L., BRZOZOWSKI, A. M., DEREWENDA, Z., DODSON, G. G., JENSEN, V. J., PETERSEN, S. B., SWIFT, H., THIM, L. & WOLDIKE, H. F. (1990). *Biochemistry*, **29**, 6244–6249.

BUISSON, G., DUEE, E., HASER, R. & PAYAN, F. (1987). *EMBO J.* **6**, 3909–3916.

CROWTHER, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. ROSSMANN, pp. 173–178. New York: Gordon & Breach.
DEREWENDA, Z. S. & HELLIWELL, J. R. (1989). *J. Appl. Cryst.* **22**, 123–137.
FARBER, G. K. & PETSKO, G. A. (1990). *TIBS*, **15**, 228–234.
HELLIWELL, J. R., GREENHOUGH, T. J., CARR, P. D., RULE, S. A., MOORE, P. R., THOMPSON, A. W. & MORGAN, T. S. (1982). *J. Phys. E*, **15**, 1363–1372.
HENDRICKSON, W. A. (1985). *Methods Enzymol.* **115**, 252–270.
HOWARD, A. J., GILLILAND, G. L., FINZEL, B. C., POULOS, T. L., OHLENDORF, D. H. & SALEMME, F. (1987). *J. Appl. Cryst.* **20**, 383–387.
JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
MACGREGOR, A. (1988). *J. Protein Chem.* **7**, 399–415.
MCPHERSON, A. (1989). *Preparation and Analysis of Protein Crystals*, p. 96. Florida: Robert E. Krieger.
MATSUURA, Y., KUNUSOKI, M., HARADA, W. & KAKUDO, M. (1984). *J. Biochem.* **95**, 697–702.
MATTHEWS, B. W. (1966). *Acta Cryst.* **20**, 230–239.
NIXON, P. E. & NORTH, A. C. T. (1976). *Acta Cryst.* A**32**, 320–325.
PHILLIPS, D. C., STERNBERG, M. J. E., THORNTON, J. M. & WILSON, I. A. (1978). *J. Mol. Biol.* **119**, 329–351.
SERC Daresbury Laboratory (1986). *CCP4. A Suite of Programs for Protein Crystallography*. SERC Daresbury Laboratory, Warrington, England.
SIM, G. A. (1959). *Acta Cryst.* **12**, 813–815.
SWIFT, H. J., BRADY, L., DEREWENDA, Z. S., DODSON, E. J., DODSON, G. G., TURKENBURG, J. P. & WILKINSON, A. J. (1991). *Acta Cryst.* B**47**, 535–544.
WANG, B. C. (1985). *Methods Enzymol.* **115**, 90–112.

*Acta Cryst.* (1991). B**47**, 535–544

# Structure and Molecular Model Refinement of *Aspergillus oryzae* (TAKA) α-Amylase: an Application of the Simulated-Annealing Method

BY HELEN J. SWIFT, LEO BRADY, ZYGMUNT S. DEREWENDA,* ELEANOR J. DODSON, GUY G. DODSON, JOHAN P. TURKENBURG AND ANTHONY J. WILKINSON

*Department of Chemistry, University of York, Heslington, York YO1 5DD, England*

(*Received* 11 *September* 1990; *accepted* 7 *February* 1991)

## Abstract

Monoclinic crystals of a neutral α-amylase from *Aspergillus oryzae*, containing three molecules in the asymmetric unit, have been reported previously and studied at 3 Å resolution [Matsuura, Kunusoki, Harada & Kakudo (1984). *J. Biochem.* **95**, 697–702]. Here we report the solution of the structure of this enzyme in a different crystal form (space group $P2_12_12_1$, $a = 50{\cdot}9$, $b = 67{\cdot}2$, $c = 132{\cdot}7$ Å), with only one molecule in the asymmetric unit. The structure was solved by the molecular replacement method, using a model of acid α-amylase from a related fungus *A. niger* [Brady, Brzozowski, Derewenda, Dodson & Dodson (1991). *Acta Cryst.* B**47**, 527–535]. Conventional least-squares crystallographic refinement failed to converge in a satisfactory manner, and the technique of molecular dynamics in the form of the *XPLOR* package [Brunger (1988). *XPLOR Manual.* Yale Univ., USA] was used to overcome the problem. A large rigid-body type movement of the C-terminal domain was identified and accounted for. The final round of restrained least-squares refinement (at 2·1 Å resolution) including 3675 protein atoms and 247 water molecules resulted in a conventional crystallographic *R* factor of 0·183 and an atomic model which conforms well to standard stereochemical parameters (standard deviation of bond lengths from their expected values is 0·028 Å, while that for planar groups is 0·029 Å).

## Introduction

Neutral α-amylase (α-1,4-glucan-4-glucanohydrolase, EC 3.2.1.1) from *Aspergillus oryzae* (TAKA amylase) is a glycoprotein which catalyses the hydrolysis of internal α(1–4) glycosidic linkages in various polysaccharides. The enzyme consists of a single polypeptide chain of 478 amino-acid residues for which the sequence has been determined (Toda, Konda & Narita, 1982; Tada, Imura, Gomi, Takaahaski, Hara & Yoshizawa, 1989).

The structure of TAKA α-amylase has been investigated by X-ray crystallography at 3·0 Å resolution (Matsuura, Kunusoki, Harada & Kakudo, 1984). The molecule was shown to consist of two main domains (see Fig. 1), with the N-terminal domain (residues 1 to 380) further subdivided into two smaller subdomains *A* and *B*. Domain *A* (residues 1 to 121 and 177 to 380) is folded into an $(\alpha/\beta)_8$ barrel, a pattern first noted in triose phosphate isomerase (Phillips, Sternberg, Thornton & Wilson,

* Present address to which correspondence should be sent: MRC Group in Protein Structure and Function, Department of Biochemistry, 474 Medical Sciences Building, University of Alberta, Edmonton, Canada T6G 2H7.

0108-7681/91/040535-10$03.00 