Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

A quick solution: *ab initio* structure determination of a 19 kDa metalloproteinase using *ACORN*

Katherine E. McAuley,[a] Yao Jia-Xing,[a] Eleanor J. Dodson,[a] Jan Lehmbeck,[b] Peter R. Østergaard[b] and Keith S. Wilson[a]*

[a]York Structural Biology Laboratory, Chemistry Department, University of York, Heslington, York YO10 5DD, England, and [b]RandD, Novozymes, DK-2880 Bagsvaerd, Denmark

Correspondence e-mail: keith@ysbl.york.ac.uk

Received 16 May 2001
Accepted 10 August 2001

PDB Reference: deuterolysin, 1eb6.

A data set from the metalloproteinase deuterolysin was collected at atomic resolution (1.0 Å) with synchrotron radiation. The high resolution allowed the structure to be solved with the new direct-methods program *ACORN* using the coordinates of the Zn atom as a starting point. The phases obtained from *ACORN* were of sufficient quality to allow automated building to be carried out in *ARP*/*wARP*. Minimal manual rebuilding of the model was required and the structure determination was completed using the maximum-likelihood refinement program *REFMAC*. The whole process, starting from the processed and merged data and ending with a refined model, required less than 6 h of computational time.

1. Introduction

1.1. *ACORN*

Protein structures have mainly been solved at resolutions less than atomic (roughly defined for this purpose as 1.2 Å or better) and this has necessitated determination of initial phases either from experimental techniques such as multiple isomorphous replacement or multiwavelength anomalous diffraction, or by molecular replacement using a homologous model. These phases were required to be of sufficient quality to construct an atomic model, which is subsequently refined with imposition of stereochemical restraints during least-squares or maximum-likelihood-based minimization. This was in marked contrast to the situation for small molecules, for which truly atomic resolution data, usually around 0.8 Å spacing, allowed the *ab initio* solution of the structure either by the heavy-atom method, where a substructure is defined from a Patterson synthesis, or more commonly by direct methods based on statistical relationships between the phases once the structure-factor amplitudes are known.

The last decade has seen a formidable advance in the data-collection facilities and techniques available to the macromolecular crystallographer. These have involved the use of more intense X-ray sources, in particular dedicated synchrotron beamlines, highly efficient two-dimensional detectors in the form of imaging plates, and more recently charge-coupled devices, and of cryogenic vitrification to alleviate the effects of radiation damage and extend the resolution of data accessible. Taken together, these advances have led to the realisation that a significant number of protein crystals indeed diffract to atomic resolution. For the first set of such structures, the major gain was the ability to carry out more complete refinement of the structures, typically with an anisotropic atomic model and riding H atoms, and effectively to release the non-experimental stereochemical restraints. Restraints were of course still needed for the poorly ordered regions and some

Table 1
Data statistics.

Values in brackets refer to the highest resolution shell *i.e.* 1.02–1.00 Å.

Temperature (K)	100
Synchrotron	ESRF
Beamline	ID14-2
Wavelength (Å)	0.933
Space group	$P2_1$
Unit-cell parameters	
a (Å)	38.4
b (Å)	34.8
c (Å)	60.3
β (°)	106.0
V_M (Å^3 Da^{-1})	2.0
Solvent content (%)	39.2
Resolution (Å)	30–1.0
R_{merge} (%)	5.2 (24.0)
R_{meas}† (%)	6.1 (29.3)
$I/\sigma(I)$	12.8 (4.8)
Unique reflections	82199
Redundancy	3.5 (3.4)
Completeness	99.6 (99.8)

† As defined in Diederichs & Karplus (1997).

restraints on the atomic displacement factors were still applied comparable to those used in small-molecule analyses.

The structures still posed problems in terms of solution: *ab initio* phasing from a single data set seemed an almost insurmountable task for most proteins. In addition, the refinement procedure was relatively tedious, requiring considerable user input to exploit fully the available data. Recent advances in computer power and in algorithms for phasing have started to transform the field. The *Shake-and-Bake* method (Weeks *et al.*, 1994; Weeks & Miller, 1999; Xu *et al.*, 2000) and the *Half-Baked* approach with a wealth of associated techniques such as *E*-map peak enhancement (Sheldrick & Schneider, 1997) have been applied to increasingly large structures. Methods based on more conventional direct-methods schemes have also had some success, again especially for metalloproteins (Mukherjee, 1999; Mukherjee *et al.*, 1999).

ACORN is a new program designed to solve protein structures *ab initio* at atomic resolution (Foadi *et al.*, 2000), providing some limited initial phasing from a substructure is available. A small – or indeed large – fragment can be used as a search model and placed in the *P*1 cell using a molecular-replacement approach. The fragment can be as small as a single α-helix. Alternatively, an experimental substructure, such as a metal or a set of S, Se or similar atoms, can be located from anomalous scattering measurements. *ACORN* then uses a combination of approaches, most importantly dynamic density modification, to develop a refined set of phases. Key to the procedure is the use of a correlation factor for the weak amplitudes as a criterion of phase quality. The program has been shown to be effective for a variety of test cases, ranging from haem-containing proteins and small proteins such as lysozyme and RNAse AP1 to larger proteins, *e.g.* penicillopepsin with 323 residues and a molecular weight of 34 kDa (Foadi *et al.*, 2000). The advantage of *ACORN* over other methods for phasing protein structures is that it can be readily applied to larger proteins and that it is extremely fast. In addition, it gives an objective electron-density map with no selection of atoms on the way. The deviation of the *ACORN*-derived phases from those of the refined model can be as low as 15–20° and the resulting density maps closely resemble those for the complete model, even in the ordered solvent regions.

In this work, we describe the application of *ACORN* to solve the structure of a zinc metalloproteinase, deuterolysin from *Aspergillus oryzae*, for which atomic resolution data could be measured but for which no homology model was available. This was the first determination of the structure of an unknown protein using *ACORN*.

1.2. Deuterolysin

Metalloproteinases have been classified into more than 30 different families, many of which contain a HE*xx*H motif that forms part of the metal-binding site (Rawlings & Barrett, 1995). The two histidine residues in the motif coordinate a Zn atom and the glutamate is the catalytic residue. The families containing this HE*xx*H motif previously belonged to either clan MA, MB or MX based on evolutionary relationships and the identity of the third residue coordinating the zinc. The gluzincins have a glutamate as the third zinc ligand (HE*xx*H + E) and belonged to clan MA. Clan MB contained the metzincins, which have HE*xx*H + H and a conserved Met-turn. However, as more three-dimensional structures became available, it was observed that there were common structural motifs in both the gluzincins and the metzincins and so clans MA and MB were merged to become clan MA. The motif for clan MA is now HE*xx*H + E, H or D (http://www.merops.co.uk; Rawlings & Barrett, 2000). In clan MX, the identity of the third ligand to the zinc is unknown.

Deuterolysin from *A. oryzae* (previously known as neutral proteinase II) is a zinc metalloproteinase consisting of 177 amino acids with a molecular weight of approximately 19 kDa and three disulfide bridges (Tatsumi *et al.*, 1994). Deuterolysins belong to clan MX and family M35. They contain the HE*xx*H motif, but until recently the third residue coordinating the zinc was unknown. Site-directed mutagenesis studies on the enzyme from *A. oryzae* confirmed that His128 and His132 are zinc ligands and that Glu129 is a catalytically crucial residue (Fushimi *et al.*, 1999). These studies also identified an aspartic acid residue (Asp164) as the third zinc ligand and, as a consequence, a new family name for deuterolysins containing the zinc-binding motif HE*xx*H + D was proposed: the 'aspzincins'.

2. Experimental and computational methods

2.1. Crystallization

Crystals of deuterolysin were grown at 291 K by vapour diffusion from hanging drops. 1 µl of a 30 mg ml^{-1} protein solution was mixed with 1 µl of a precipitant solution containing 1.8 *M* sodium/potassium phosphate, 0.1 *M* HEPES buffer pH 7.5. The drop was equilibrated against 1 ml of the precipitant solution. Large needle-shaped crystals grew within

3–4 d. The crystals belong to space group $P2_1$, with unit-cell parameters a = 38.4, b = 34.8, c = 60.3 Å, β = 106.0°. There is one molecule in the asymmetric unit, giving rise to a Matthews number of 2.0 Å^3 Da^{-1}, corresponding to an approximate solvent content of 39% (Matthews, 1968).

2.2. Data collection

The crystals were soaked briefly in 2.0 *M* sodium/potassium phosphate, 0.1 *M* HEPES pH 7.5, 25% ethylene glycol before vitrification in liquid nitrogen. The data were collected at 100 K using an ADSC Q4 CCD detector on beamline ID14-2 at the ESRF. Two different crystal-to-detector distances (60 and 290 mm for high- and low-resolution data, respectively) were used with oscillation steps of 0.5 and 1.5°, respectively. The data were processed and scaled using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997) (Table 1). Subsequent crystallographic calculations mainly used programs from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994).

2.3. Phasing

ACORN requires a small initial fragment of the structure to be positioned and the size of the fragment can be less than 5% of the scattering matter of the unit cell. For a metalloprotein, the position of the metal alone makes a suitable starting point for phase refinement. The coordinates of the Zn atom of deuterolysin were determined by calculating an anomalous difference Patterson over the resolution range 30–1 Å. The zinc position needed to be given with an error of less than 0.3 Å for the procedure to be successful: use of a coarse interpolation grid resulted in a sufficient error in the estimated coordinate to prevent its successful use as a starting point. The strongest peak in the Harker section of the anomalous difference Patterson was observed at (u, v, w) = (0.115, 0.5, 0.540). This peak was also clearly dominant in the sharpened native Patterson map calculated using normalized structure factors (Fig. 1), indicating that the anomalous information was not in fact required.

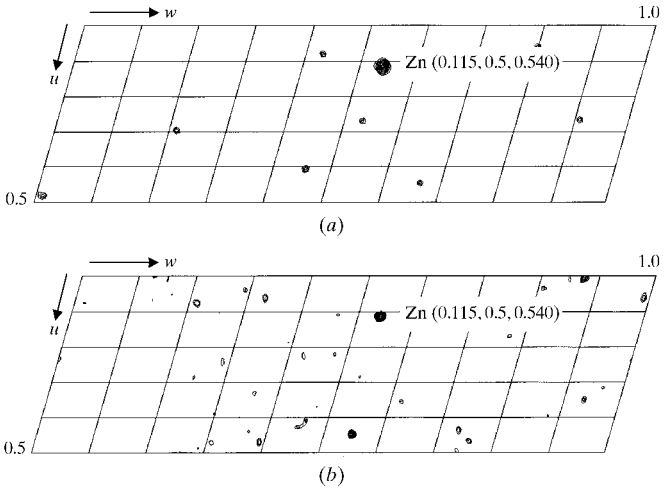

Figure 1
v = 0.5 Harker sections of (*a*) the anomalous difference and (*b*) the sharpened native Patterson functions.

Table 2
Overview of crystallographic stages.

Stage	Time (min)
Phasing with *ACORN*	
From Zn coordinate	2.8†
From 5000 random atoms	8.2†
Model-building with *ARP*	
WarpNtrace: first time	100†
WarpNtrace: optimized	20‡
Side dock	5†
Model-building in *XtalView*	∼10†
Refinement and solvent building in *REFMAC5*/*ARP*	225†

† Silicon Graphics O2. ‡ Pentium III 800 MHz PC.

Normalized structure factors (E) were calculated from the native protein data using the program *ECALC* (Collaborative Computational Project, Number 4, 1994). The coordinates of the Zn atom were input into *ACORN* and phase development was carried out by dynamic density modification (DDM) for 38 cycles using the 18 566 largest E values (E > 1.2). The correlation coefficient (CC_s) for the reflections with small E values (E < 1.2) increased from 0.00942 to 0.45529, indicating that a solution had been obtained (Foadi *et al.*, 2000). The time required was 3 min (Table 2).

Indeed, the structure could also be solved without knowledge of the zinc position: a solution was obtained from 5000 trials starting from a single atom in a random position. In this case, 49 cycles of DDM were required and the CC_s for the smallest E values increased from 0.00846 to 0.45318. The situation is of course simplified by the nature of the space group: in $P2_1$ only the x and z parameters need to be varied for the single-atom starting 'fragment'. The time required for this computation was still only 8 min, probably rather faster than computing, displaying and solving the Patterson in practice.

2.4. Building and refinement

The electron-density maps obtained at this stage were of excellent quality and it would have been simple to build the protein model using typical protein-tracing methods. However, it was preferred to build the model in an automated fashion using *REFMAC* combined with *ARP*/*wARP* (Perrakis *et al.*, 1999). The phases from *ACORN* were input into the warpNtrace mode of *ARP*/*wARP*. Initially, 1509 water molecules were located in the map calculated with the phases from *ACORN*; these free atoms were refined in *REFMAC* (Murshudov *et al.*, 1997, 1999) for one cycle and then building of the protein main chain commenced. The initial trace consisted of 169 residues in four protein chains; this increased to 173 residues in two chains after the second building cycle. In between building cycles, four cycles of *REFMAC* refinement were carried out. Further building steps failed to build the missing residues and so the program was therefore terminated after the fourth building cycle. The R factor and free R factor were 37.6 and 39.5%, respectively, at the beginning of the *ARP*/*wARP* procedure and decreased to 19.2 and 23.1% at the

end. The initial building of the main chain required 100 min of computational time, but this time could be reduced by decreasing the number of refinement cycles between *ARP* updates and by performing just two building steps (Table 2).

The side chains were added to the model using the side-dock script of *ARP*/*wARP*, which relates the protein fragments to the given amino-acid sequence and then fits the side chains using real-space torsional refinement. The missing residues and the zinc ion were added manually to the model using *XtalView* (McRee, 1999).

Anisotropic maximum-likelihood refinement of the resulting model was continued using *REFMAC*5 (Murshudov *et al.*, 1997, 1999) combined with the addition of water molecules by *ARP*/*wARP*. Solvent molecules were added to positive density peaks greater than 3σ in the *REFMAC*-weighted $mF_o - DF_c$ map if there were acceptable hydrogen-bonding distances to neighbouring atoms. Waters were removed from the model if their density was less than 1.0σ in the $2mF_o - DF_c$ map. Visual inspection and manual adjustments to the model were carried out using *XtalView*.

The final model consists of 1380 non-H protein atoms, 260 water molecules, three molecules of ethylene glycol and a zinc ion. The geometry is good and the protein model was verified using the programs *PROCHECK* (Laskowski *et al.*, 1993) and

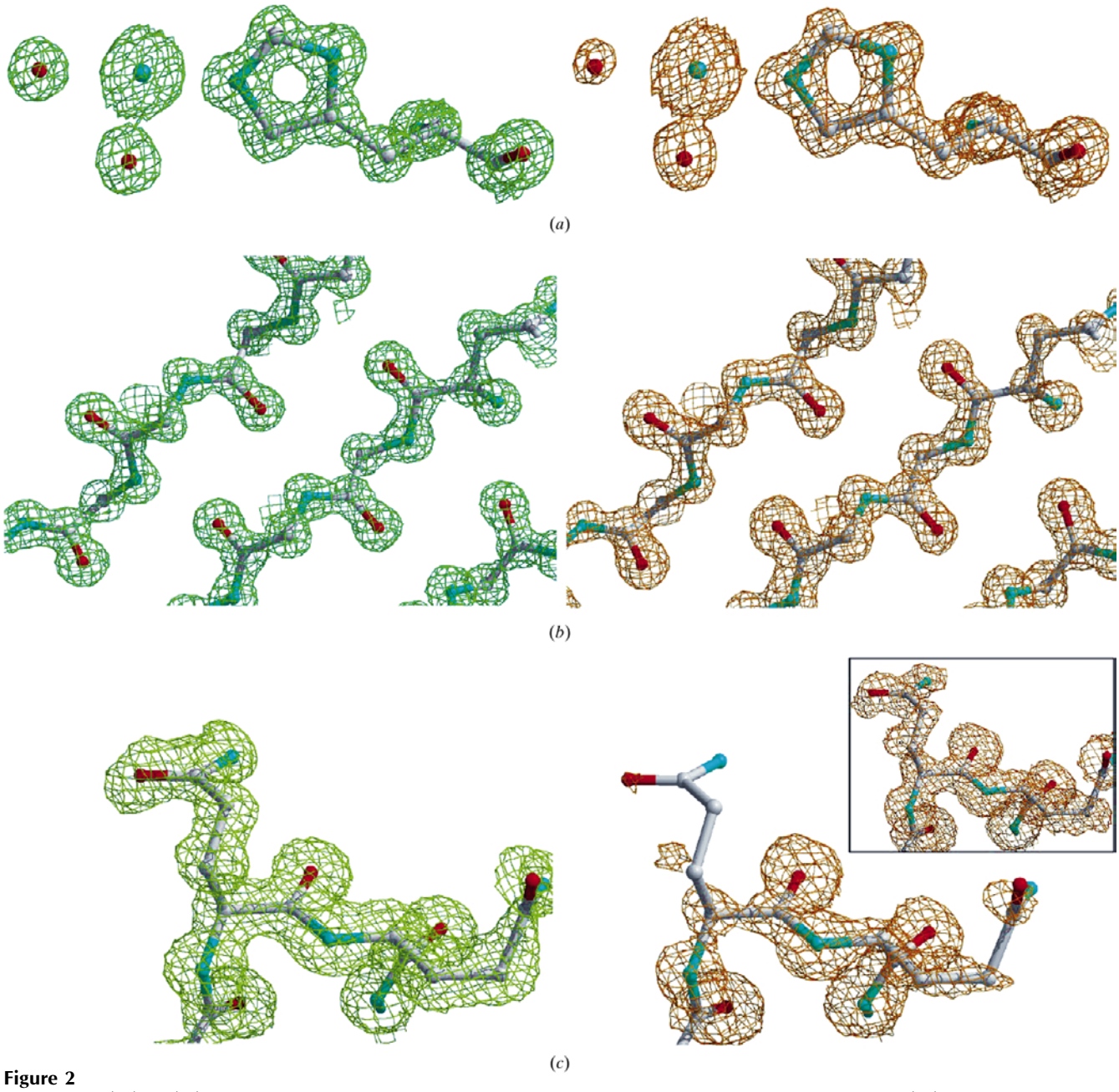

Figure 2
REFMAC $2m|F_o| - D|F_c|$ density map for refined model (in green) and *ACORN* map [calculated with the coefficients $|F_o|\exp(i\varphi_{acorn})$ in orange] for (*a*) His132 and the zinc ion contoured at 2σ, (*b*) a region of β-sheet contoured at 2σ and (*c*) a glutamine residue on the surface of the protein contoured at 1σ, with the inset diagram contoured at 0.5σ. Figure produced using *XtalView* (McRee, 1999).

WHAT CHECK (Hooft *et al.*, 1996). All residues fall within allowed regions of the Ramachandran plot.

3. Results and discussion

3.1. Phasing and automated building of the model

Metalloproteins are easier to solve by direct methods as the early cycles locate the heavy atoms and the rest of the structure can then be developed from that. For deuterolysin, the zinc was input into *ACORN* and a phase set with an excellent correlation coefficient was very quickly obtained. *ACORN* has previously shown to be an effective tool for phasing of proteins structures using trial data, but this work is the first example of its use in determining the structure of an unknown protein. The solution of the structure required only 3 min CPU time on a Silicon Graphics O2. Even without inputting the coordinates of the zinc, it was possible to obtain a solution by starting from 5000 random atom fragments and this increased the CPU time to just 8 min, still a very time-efficient method. The maps calculated at this stage were of excellent quality (Fig. 2) and the mean phase discrepancy compared with the final model was only 14°.

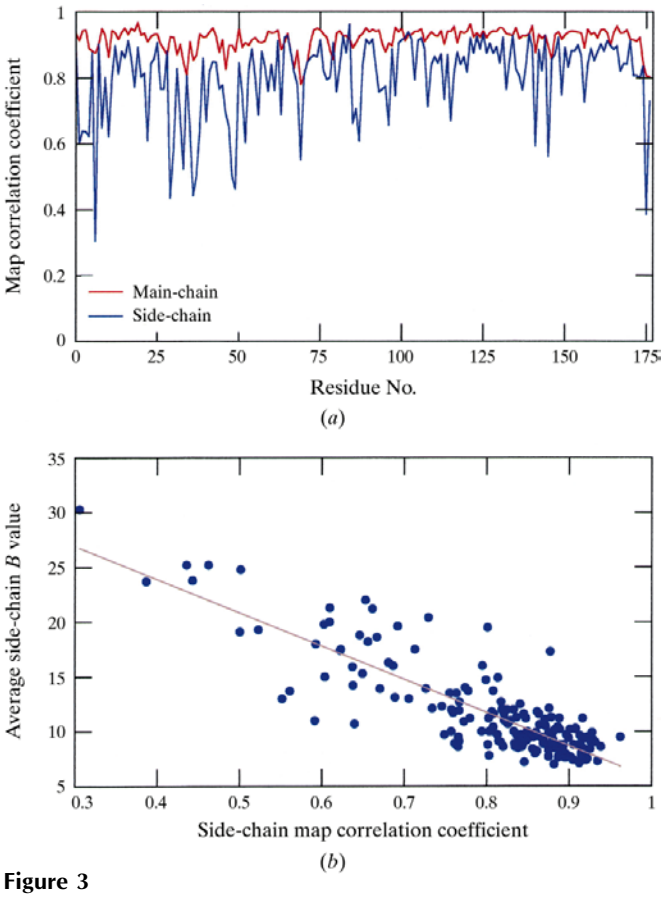

Figure 3
(*a*) Map correlation for main chain (in red) and side chains (in blue), showing the quality of the phases obtained from *ACORN* and (*b*) plot of the average side-chain *B*-factor *versus* map correlation coefficient for the side chains. Note the relationship between residues with high *B* values and residues with lower side-chain correlation coefficients.

Table 3
Refinement statistics.

R factor (%)	10.3
R_{free} (%)	12.5
Ramachandran plot, non-glycine residues in	
Most favoured region (%)	93.7
Allowed region (%)	6.3
Generously allowed region (%)	0
Disallowed region (%)	0
No. of atoms/molecules	
Protein	1380
Zn	1
Ethylene glycol	3
Solvent	260
Mean *B* factor (Å^2)	
Protein	10.7
Main chain	9.3
Side chains	12.2
Zn	8.2
Ethylene glycol	14.4
Solvent	27.5
Root-mean-square deviations	
Bond lengths (Å)	0.018
Bond angles (°)	2.4
PDB code	1eb6

The quality of the *ACORN* phases was examined further by calculating the correlation coefficient between the *ACORN* map and the final map as a function of residue number using the program *OVERLAPMA*P (Collaborative Computational Project, Number 4, 1994; Fig. 3*a*). The correlation for the protein main chain is very high, ranging from 0.78 to 0.97. The correlations for the side chains are also generally high, with the majority of residues in the range 0.6–0.97. However, some

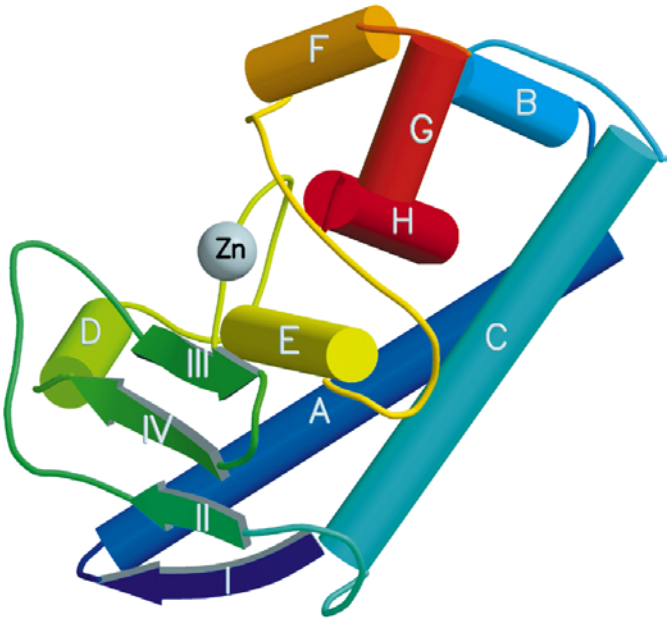

Figure 4
Secondary structure of deuterolysin, also showing the zinc ion in grey. The helices are labelled from *A* to *H* and the β-sheets from I to IV. The protein is coloured according to sequence by a rainbow colour ramp from red at the C-terminus to blue at the N-terminus. Figs. 4, 5 and 6 were drawn using *MOLSCRIPT* (Kraulis, 1991) and *Raster3D* (Merritt & Bacon, 1997).

side chains have lower correlation coefficients, *i.e.* between 0.6 and 0.3. A comparison of the side-chain correlation coefficients with the average side-chain *B* values was made (Fig. 3*b*) and it is clear that low correlation coefficients correlate strongly with higher *B* values. In all cases, the residues with poor correlation coefficients and high *B* values are located on the surface of the protein. Fig. 2(*c*) shows the differences in electron density between the final map and the *ACORN* map for one of these surface residues.

Almost the complete protein main chain was automatically built in *ARP*/*wARP* 5.1 after only a few building cycles. The four missing amino acids were the N- and C- terminal residues and a *cis*-proline with its neighbouring residue. These are places where the warpNtrace algorithm can be expected to have difficulties in its present implementation. The four missing residues could easily be fitted to the density, requiring only a few minutes on the graphics computer. The progress and more details of the building and refinement steps are given in Table 3.

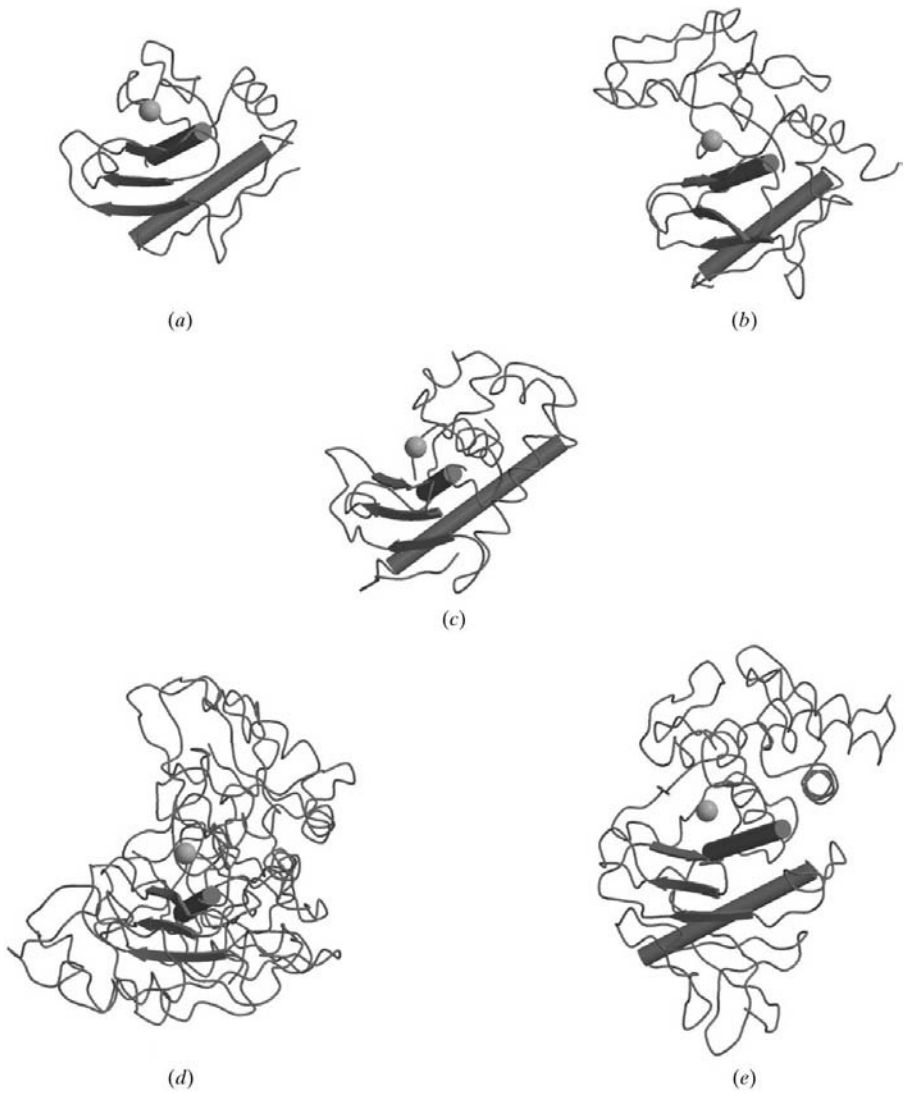

Figure 5
Comparison of the metalloproteinases (*a*) snapalysin (PDB code 1kuh), (*b*) astacin (PDB code 1iab), (*c*) deuterolysin, (*d*) leishmanolysin (PDB code 1lml) and (*e*) thermolysin (PDB code 2tmn). The matching secondary-structure elements are shown as either cylinders (for helices) or arrows (for sheets) and the remainder of the protein is shown as a smoothed C^{α} trace.

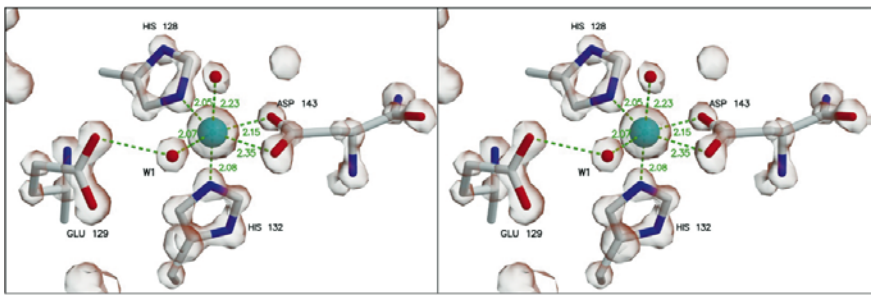

Figure 6
Active-site residues of deuterolysin with *REFMAC* $2m|F_o| - D|F_c|$ density contoured at 1.6σ. The nucleophilic water is labelled W1. Distances (Å) are shown in green. Density was drawn using *CONSCRIPT* (Lawrence & Bourke, 2000).

3.2. Structure of deuterolysin

Deuterolysin is an $\alpha\beta$ protein with a zincin-like fold. It comprises eight α-helices and four strands of β-sheet. There are three disulfide bridges between residues *A*6–*A*78, *A*85–*A*103 and *A*117–*A*177; the last of these involves the C-terminal residue, which bridges to a cysteine residue located on a loop region between helices *D* and *E*. The secondary structure of deuterolysin is shown in Fig. 4.

A *BLAST* search (Altschul *et al.*, 1990) had shown that deuterolysin had no significant sequence homology to any other protein in the Protein Data Bank (Berman *et al.*, 2000). To look for structural homology with other proteins in the PDB, the program *TOP* was used (Lu, 2000). *TOP* outputs a list of homologous proteins ranked according to their topological similarities. A topology search using *TOP* revealed that the most closely structurally related enzyme, at the time this work was carried out, was the zinc protease from *Streptomyces caespitosus* (also known as snapalysin; PDB code 1kuh), giving a topological diversity score of 20.0 and with six matching secondary-structure

elements. Although these proteins are similar structurally, the sequence homology is very low.

The other proteins with low values of topological diversity in the list are the other metalloproteinases belonging to clan MA, *e.g.* leishmanolysin, thermolysin and astacin. Although these proteins belong to clan MA, they are divided into different families within the clan. The structural similarity relating these metalloproteins lies in the arrangement of β-sheets and α-helices that form the active site. A comparison of the structures of deuterolysin, snapalysin, leishmanolysin, astacin and thermolysin is shown in Fig. 5.

The structure of another aspzincin with 21% sequence homology to deuterolysin, the metalloproteinase from *Grifola frondosa*, has recently been determined (Hori *et al.*, 2001). This enzyme is closely related in structure to deuterolysin, with an r.m.s. difference in C^{α} positions of 1.2 Å and a topological diversity score of 8.4, yet it is still a poor molecular replacement search model.

3.3. Active-site residues

The active site is located in a large cleft on the protein surface formed by β-sheets and an α-helix. The histidine residues coordinating the zinc are located on this helix (αE). The Zn atom is liganded by two histidine residues (His128 and His132), an aspartic acid residue (Asp143) and two water molecules, forming a distorted octahedral geometry as shown in Fig. 6. The metal–ligand distances range from 2.05 to 2.35 Å, falling into the expected range for zinc metal geometry (Harding, 1999). Several of the angles within the octahedron deviate by more than 10° from ideal values and this distortion is a consequence of accommodation of the bidentate carboxylate group of Asp143.

The catalytic residue which promotes the nucleophilic attack of a water molecule on the carbonyl C atom of the substrate is Glu129.

The structure confirms that deuterolysin is an aspzincin as proposed by Fushimi *et al.* (1999), but there is a disagreement in the identity of the aspartic acid that ligands the zinc. Fushimi and coworkers deduced this was Asp164 after their studies showed that mutating this to an asparagine residue led to loss of the zinc ion, whereas the Asp143→Asn mutant retained the zinc. It is not clear why the Asp143 mutant retained the zinc; however, a suggestion as to why mutating Asp164 led to loss of the zinc can be made. Asp164 is located close to His128, with a distance of 2.86 Å between the carboxylate oxygen OD1 of Asp164 and nitrogen ND1 of the histidine. The loss of this interaction after mutation to asparagine may lead to a change in the orientation or position of this histidine, resulting in loss of zinc-binding ability.

3.4. Conclusions

The power of *ACORN* has been demonstrated by the *ab initio* phasing of a 19 kDa metalloproteinase at 1.0 Å spacing. The *ACORN* phases were very close to those of the refined model, with a mean difference of only 14°. Phase determination took only 3 min. The extraction of an atomic model using the present version of warpNtrace took approximately 20 min of computing and, since the algorithm was developed for structures at substantially lower resolution, this can almost certainly be substantially reduced in the future for maps of this quality. Perhaps the most important message from this study is that *ACORN* has made it extremely easy to compute in a fast and automatic way a set of phases which are close to correct for an atomic resolution structure, avoiding extensive and iterative computations for refinement and in particular long periods at a molecular graphics terminal. In addition, the electron-density map obtained with the *ACORN* phases is devoid of subjective input with regard to an atomic model, relying solely on the dynamic density-modification algorithm. This makes *ACORN* ideal for deriving unbiased phase sets when ligands are introduced to a protein and the configuration of the complex is uncertain. The crystallographer can instead spend his/her time analysing those details of the structure relevant to the functional questions addressed, or indeed the real structural detail of the molecule.

In spite of a lack of significant sequence homology, deuterolysin has been shown to be an aspzincin, related in three-dimensional fold to the thermolysin family of metalloproteinases. However, the aspartate which ligates the zinc is different from that expected from biochemical studies. None of the homologous proteins were sufficiently similar to deuterolysin to act as effective search models for molecular replacement. This demonstrates the need for more than one member of homology families. Even if through structural genomics studies we can solve a representative structure of each 'fold family', this will not be sufficient to solve the structures of highly varying members of a family by molecular replacement, let alone homology modelling with our present algorithms.

We thank the BBSRC for support through the award of a Structural Biology Centre Grant, 87/SB09829. YJ-X was supported by the BBSRC grant B08494 and EJD by the MRC grant G9708625. We thank the ESRF for provision of synchrotron facilities and we thank Lisa Wright and Koen Verschueren for data collection. We thank Miroslawa Dauter who grew the original batch of crystals for deuterolysin and Zbigniew Dauter who initially characterized these crystals at room temperature.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403–410.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.
Foadi, J., Woolfson, M., Dodson, E. J., Wilson, K. S., Jia-xing, Y. & Chao-de, Z. (2000). *Acta Cryst.* D**56**, 1137–1147.
Fushimi, N., Ee, C. E., Nakajima, T. & Ichishima, E. (1999). *J. Biol. Chem.* **274**, 24195–24201.
Harding, M. M. (1999). *Acta Cryst.* D**55**, 1432–1443.

Hooft, R. W. W., Vriend, G., Sander, C. & Abola, E. E. (1996). *Nature (London)*, **381**, 272.
Hori, T., Kumasaka, T., Yamamoto, M., Nonaka, T., Tanaka, N., Hashimoto, Y., Ueki, T. & Takio, K. (2001). *Acta Cryst.* D**57**, 361–368.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Laskowski, R. A, MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lawrence, M. C. & Bourke, P. (2000). *J. Appl. Cryst.* **33**, 990–991.
Lu, G. (2000). *J. Appl. Cryst.* **33**, 176–183.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Mukherjee, M. (1999). *Acta Cryst.* D**55**, 820–825.
Mukherjee, M., Maitia, S. & Woolfson, M. M. (1999). *Acta Cryst.* D**56**, 1132–1136.
Murshudov, G. N., Lebedev, A., Vagin, A. A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Morris, R. M. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Rawlings, N. D. & Barrett, A. J. (1995). *Methods Enzymol.* **248**, 183–228.
Rawlings, N. D. & Barrett, A. J. (2000). *Nucleic Acids Res.* **28**, 323–325, 2000.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Tatsumi, H., Ikegawa, K., Murakami, S., Kawabe, H., Nakano, E. & Motai, H. (1994). *Biochim. Biophys. Acta*, **1208**, 179–185.
Weeks, C. M., De Titta, G. T., Hauptman, H. A., Thuman, P. & Miller, R. (1994). *Acta Cryst.* A**50**, 210–220.
Weeks, C. M. & Miller, R. (1999). *Acta Cryst.* D**55**, 492–500.
Xu, H., Hauptman, H. A., Weeks. C. M. & R. Miller, R. (2000). *Acta Cryst.* D**56**, 238–240.