Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure determination of the cancer-associated *Mycoplasma hyorhinis* protein Mh-p37

Katherine H. Sippel,[a] Arthur H. Robbins,[a] Robbie Reutzel,[a]‡ John Domsic,[a] Susan K. Boehlein,[b] Lakshmanan Govindasamy,[a] Mavis Agbandje-McKenna,[a] Charles J. Rosser[c] and Robert McKenna[a]*

[a]Department of Biochemistry and Molecular Biology, University of Florida, Gainesville, FL 32610, USA, [b]Program in Plant Molecular and Cellular Biology and Horticultural Sciences, University of Florida, Gainesville, FL 32610, USA, and [c]Department of Urology, University of Florida, Gainesville, FL 32610, USA

‡ Present address: Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02454, USA.

Correspondence e-mail: rmckenna@ufl.edu

Received 18 August 2008
Accepted 18 September 2008

PDB References: Mh-p37, native, 3e78, 3e78sf; I3C complex, 3e79, 3e79sf.

The crystal structure of the *Mycoplasma hyorhinis* protein Mh-p37 has been solved and refined to 1.9 Å resolution. This is the first *de novo* structure to be determined using the recently described heavy-atom reagent [Beck *et al.* (2008), *Acta Cryst.* D**64**, 1179–1182] 5-amino-2,4,6-triiodoisophthalic acid (I3C), which contains three I atoms arranged in an equilateral triangle, by SIRAS methods. Data collection was performed in-house at room temperature. *SHELXD* and *SHELXE* were used to determine the I-atom positions and phase the native protein and *PHENIX AutoBuild* software was used to automatically fit the amino-acid sequence to the electron-density map. The structure was refined using *SHELX*97 to an $R_{\rm cryst}$ of 18.6% and an $R_{\rm free}$ of 24.0%. Mh-p37 is an α/β protein with two well defined domains which are separated by a deep cleft. An unanticipated ligand bound in the center of the molecule at the base of the cleft has been modeled as thiamine pyrophosphate or vitamin B_1. Retrospective attempts to solve the crystal structure by Patterson search methods using either isomorphous or anomalous differences failed. Additionally, attempts to use proteins with the highest structural homology in the Protein Data Bank to phase the data by molecular replacement were unsuccessful, most likely in hindsight because of their poor structural agreement. Therefore, the I3C reagent offers an alternative, quick and inexpensive method for in-house phasing of *de novo* structures where other methods may not be successful.

1. Introduction

Mh-p37 is a protein present on the outer membrane of the bacterium *Mycoplasma hyorhinis*. The protein was originally identified in 1984 as an antigen present on the plasma membrane of mouse sarcoma cells that contributes to their invasiveness (Steinemann *et al.*, 1984). A subsequent study of the antibodies present in cancer patients that had been immunized with their own surgically debulked tumors revealed high titers against a 38 kDa protein named p37 from *M. hyorhinis* (Fareed *et al.*, 1988), referred to here as Mh-p37. In subsequent years several studies have indicated a correlation between the bacterium *M. hyorhinis* and various cancers including, but not limited to, gastric carcinoma, colon carcinoma, esophageal cancer and lung cancer (Sasaki *et al.*, 1995; Huang *et al.*, 2001). Additionally, studies have shown that Mh-p37 alone is sufficient to increase the invasivity of cancer cells (Ketcham *et al.*, 2005; Steinemann *et al.*, 1984). It has also been shown that Mh-p37 may also be associated with cancer malignancy and metastasis (Liu *et al.*, 2007; Gong *et al.*, 2008).

Although the function of Mh-p37 remains unknown, it has been hypothesized that it is part of a periplasmic binding-protein-dependent transport system that is commonly found in Gram-negative bacteria (Dudler *et al.*, 1988; Gilson *et al.*, 1988). It is expected that knowledge of the structure of Mh-p37 may provide insight into its function and provide a target for the development of new small-molecule cancer therapies.

Crystallization and preliminary X-ray data of Mh-p37 have been reported previously (Reutzel *et al.*, 2002). However, its low (∼15%) sequence identity to previously solved structures made molecular-replacement phasing methods unfeasible and most likely impossible (Rossmann, 1990). The use of SeMet to phase the structure was not considered as only two methionines were present in the 400-amino-acid sequence (not including the N-Met, which has a high probability of being disordered). It has previously been shown that one SeMet is required per ∼75–100 residues to assure success using this method and it was thought not be a good strategy to engineer other SeMet sites within Mh-p37 without any prior structural information (Hendrickson & Ogata, 1997). Additionally, the low pH of the crystallization conditions (pH 3) interfered with the binding of heavy atoms and thus the structure remained unsolved for 6 y despite much effort. The recent report of a heavy-atom derivative, 5-amino-2,4,6-triiodoisophthalic acid (I3C), at the American Crystallography Association meeting in Knoxville in 2008 (Beck & Sheldrick, 2008*a*,*b*) presented a straightforward and inexpensive opportunity to phase the structure of Mh-p37. I3C proved to be uniquely suited for the solution of the Mh-p37 structure as it allowed the binding of a heavy-atom compound at low pH. An in-house room-temperature (RT) X-ray diffraction data collection from a single Mh-p37 crystal, quick-soaked in I3C, was sufficient to identify the positions of five I atoms using *SHELXD*.

Reported here is the structure determination of Mh-p37 in space group $P2_1$ using I3C as a heavy-atom derivative for SIRAS phasing and a postmortem of previous attempts to phase the structure to determine why they failed.

2. Experimental procedures

2.1. Expression and purification

An N-terminal truncated form of Mh-p37 (residues 24–403) was expressed and purified using the methods previously described by Ketcham *et al.* (2005).

2.2. Crystallization and diffraction data collection

The crystallization conditions for Mh-p37 were slightly modified from those previously reported by Reutzel *et al.* (2002). The crystals were grown using the batch method in microbridges under paraffin-oil immersion at RT. Crystallization drops were prepared by mixing 3 µl protein solution at a concentration of ∼10 mg ml^{-1} in 50 m*M* Tris–HCl pH 7.5 with 3 µl precipitant solution (100 m*M* citric acid pH 3.0 containing 40% PEG 4000 and 100 m*M* ammonium bromide) and 0.6 µl *N*-decyl-β-D-maltoside detergent (Hampton Research). Useful crystals appeared within 7 d. The heavy-atom reagent I3C (Sigma) was soaked into the native crystals by the addition of 0.5 µl 0.5 *M* I3C solution directly to the crystallization drop. The crystals acquired a pale yellow color within 5 min and were immediately mounted in a quartz capillary for RT data collection.

Table 1
X-ray diffraction data-collection statistics for native and I3C-soaked Mh-p37.

Values in parentheses are for the highest resolution shell.

	Native	IC3 soak
Space group	$P2_1$	$P2_1$
Unit-cell parameters		
a (Å)	50.4	50.1
b (Å)	69.8	69.2
c (Å)	60.0	60.3
β (°)	108.0	106.8
Molecules per ASU	1	1
V_M (Å^3 Da^{-1})	2.32	2.30
Solvent content (%)	47.0	46.5
Data collection		
Resolution (Å)	25.0–1.9 (1.98–1.90)	50.0–1.9 (1.97–1.90)
No. of frames (°)	170	196
Total reflections collected	265806	426490
Unique reflections	29894	29985
Completeness (%)	96.2 (93.2)	92.2 (87.0)
Mean multiplicity	3.5 (3.5)	4.1 (4.1)
R_{merge}† (%)	6.7 (37.6)	9.8 (37.9)
$I/\sigma(I)$	16.3 (5.0)	10.8 (3.7)

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$.

All X-ray diffraction data were collected using an in-house R-AXIS IV^{++} image-plate system with Osmic mirrors and a Rigaku RU-H3R Cu rotating-anode generator operating at 50 kV and 100 mA. The data were collected with a crystal-to-detector distance of 100 mm using 1° oscillation steps with 8 min exposure per image. One native crystal data set of 170 frames to 1.9 Å resolution and one I3C-soaked crystal data set of 200 frames to 1.9 Å resolution were collected. The last four frames of the I3C-soaked data were not used as they exhibited substantial radiation damage. The data frames were indexed, integrated and scaled with *HKL*-2000 (Otwinowski & Minor, 1997). Diffraction data statistics are presented in Table 1.

2.3. Structure solution

The crystal structure of Mh-p37 was solved using SIRAS methods based upon the location of five iodine positions and their associated anomalous differences. The iodine positions were determined using the *SHELXC* and *SHELXD* software (Sheldrick, 2008; Collaborative Computational Project, Number 4, 1994). Protein phases were calculated for the monoclinic space groups $P2$ and $P2_1$ since the diffraction data did not include sufficient $0k0$ reflections for an unambiguous space-group assignment and both enantiomorphs were also checked. Iodine positions from *SHELXD* were displayed in *Coot* (Emsley & Cowtan, 2004) and appeared in two clusters: one was an equilateral triangle and the other was an isolated pair of sites. All iodine–iodine intracluster distances were in the range 5.6–5.9 Å, as expected from the structure of the I3C

Table 2
Refinement statistics for native and I3C-soaked Mh-p37 structures.

	Native	IC3 soak
R_{cryst}† (%)	18.6	18.7
R_{free}‡ (%)	24.0	24.4
$R_{all\ data}$§ (%)	18.7	18.9
Model statistics		
R.m.s. deviations		
Bond distances (Å)	0.006	0.006
Angles (Å)	0.022	0.021
Ramachandran plot (%)		
Most favored	89.1	90.6
Allowed	10.6	9.1
Generously allowed	0.3	0.3
Protein atoms, including alternate conformations	2975	2960
No. of I3C atoms	n/a	32
No. of TPP atoms	26	26
No. of Ca^{2+}/Cl^{-} atoms	1/1	1/1
No. of water molecules	148	133
Average temperature factors (Å^2)		
Protein main chain	28.3	23.1
Protein side chain	36.9	31.2
I3C (site *A*/*B*)	n/a	29.9/35.8
TPP	23.4	20.6
Ca^{2+}	36.0	20.4
Cl^{-}	43.6	32.6
Water	40.8	31.8
PDB code	3e78	3e79

† $R_{cryst} = \sum |F_o| - |F_c| / \sum |F_o| \times 100$. ‡ R_{free} is calculated in the same way as R_{cryst} except that it uses 5% of reflection data omitted from refinement. § $R_{all\ data}$: final cycle of refinement using all reflections.

reagent (Beck & Sheldrick, 2008*a*). *SHELXE* was then used to compute the protein phases and the resulting electron-density maps were examined for reasonable protein density. The best map resulted from phases in space group $P2_1$ and the inverse enatiomer of the original heavy-atom constellation. This electron-density map was used to automatically fit the sequence using the *PHENIX AutoBuild* software (Adams *et al.*, 2002). Within six cycles of fitting, over 120 amino acids were fitted to the electron density. The final *AutoBuild* model contained 311 placed residues, of which 250 were assigned amino-acid sequences. Thus, ∼66% of the 379 amino acids in the truncated sequence were assigned, yielding an initial R_{work} of 29% and an R_{free} of 32%.

2.4. Completion of the Mh-p37 molecular model

The Mh-p37 structure determination was completed by the use of iterative rounds of model building in *Coot* (Emsley & Cowtan, 2004) followed by cycles of refinement in *SHELXL* from the *SHELX*97 software package (Sheldrick, 2008). The model resulting from the *PHENIX AutoBuild* process contained 11 peptides, of which three long stretches of 40, 66 and 147 amino acids had the correct amino-acid sequence assigned. Many of the early model readjustments involved linking the shorter unassigned peptides together and then mutating the amino acids and fitting the side chains into the electron density. The final section of the main chain to be built was a surface loop between residues Asn56 and Asn65. The side-chain electron density and local environment of Phe256 were inconsistent with the presence of a phenylalanine at that position and it was replaced by a serine. This assignment must be confirmed by resequencing of the plasmid used to express the protein. Subsequently, waters were added in a stepwise process using the automated water-divining option in

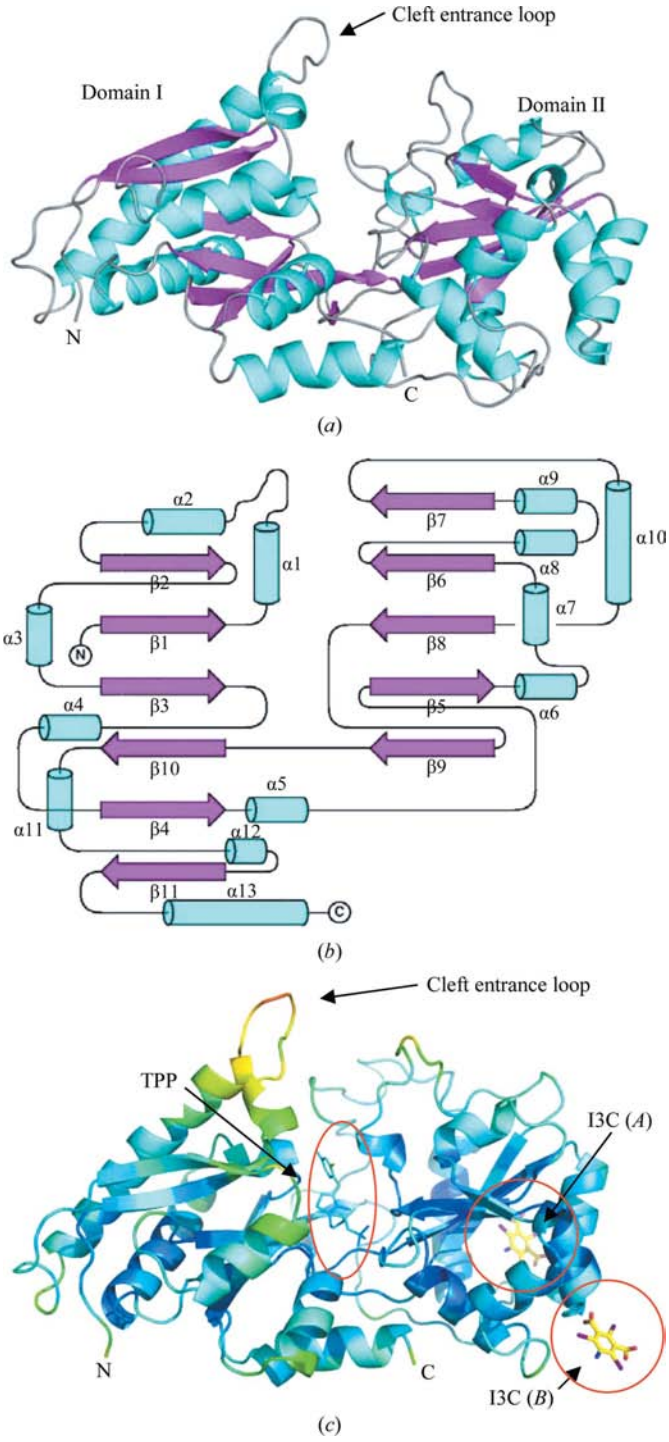

Figure 1
Overall structure of Mh-p37. (*a*) C^{α} ribbon trace showing the secondary-structural elements (α-helices, cyan; β-strands, magenta). (*b*) Structural topology of Mh-p37, defining domain I and II, colored as above. (*c*) C^{α} ribbon trace colored from low to high (blue to red) temperature factor. Also shown (open red circles) are the binding sites for TPP and I3C (sites *A* and *B*). (*a*) and (*c*) were generated using *PyMOL* (DeLano, 2002); (*b*) was generated using *TopDraw* (Bond, 2003).

SHELXL. Waters were deleted from the model if their isotropic temperature factors exceeded 50.0 Å^2. The coordinates of this model (with solvent removed) were then used to calculate phases for the 1.9 Å resolution native data, followed by standard refinement-cycling methods. No significant differences were observed between the native and I3C-soaked structures.

2.5. Identification of ligand densities

In a cleft between two distinct domains of the Mh-p37 structure, well defined electron density was found that suggested a pyrophosphate linked to an undefined density that terminated in a planar entity. After several attempts (discussed in §3.2) had been made, the ligand was eventually identified as thiamine pyrophosphate (TPP), also known as vitamin B_1. A TPP cofactor was retrieved from the *HIC-UP* server (Kleywegt, 2007) and was fitted into the electron density. Molecular bond and angle restraints were generated for TPP using the *SHELXPRO* utility in the *SHELX*97 software. The resulting electron-density map following further cycles of refinement showed no significant $|F_o| - |F_c|$ difference electron density.

The final Mh-p37 model contains residues 39–403 and refined to 1.9 Å resolution with an R_{cryst} of 18.6% and an R_{free} of 24.0%. The Ramachandran plot has 89.1% of the residues in the most favored region and 0.3% in the generously allowed region (Table 2).

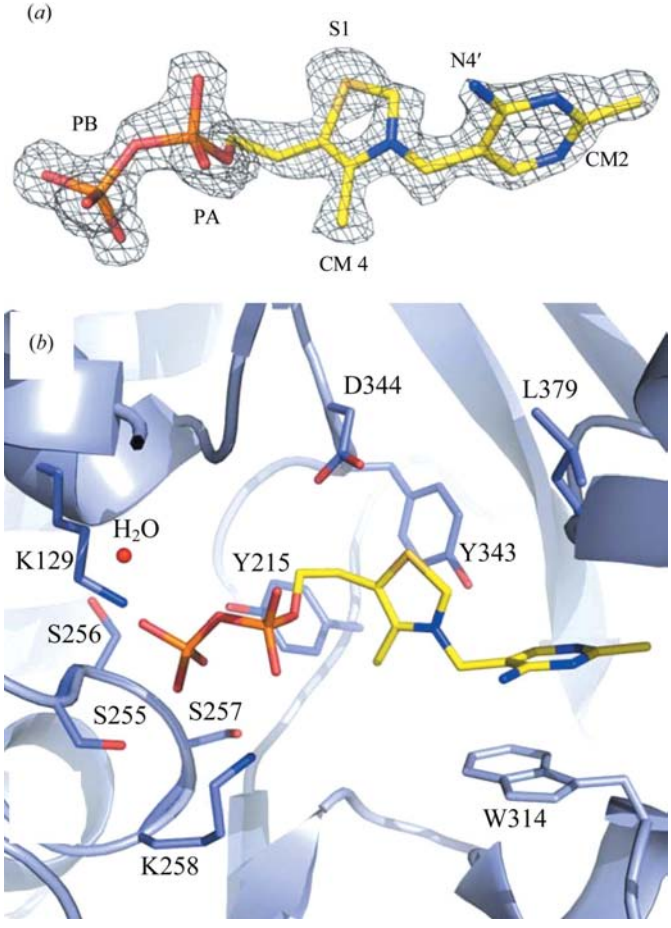

Figure 2
Thiamine pyrophosphate-binding site. (*a*) $2|F_o - F_c|$ electron density of TPP, contoured at 2σ. (*b*) Binding interactions of TPP within the Mh-p37 cleft. Waters are colored red. Amino acids are as labeled. The atom numbering for TPP is as labeled in (*a*). This figure was generated using *PyMOL* (DeLano, 2002).

3. Results and discussion

3.1. Overall fold of Mh-p37

The structural organization of Mh-p37 corresponds to an α/β-class protein arranged into two approximately equal-sized compact domains (domains I and II) separated by a deep cleft at the interface (Fig. 1*a*). The overall shape of Mh-p37 is that of an irregular prolate ellipsoid, with approximate dimensions $70 \times 50 \times 40$ Å. Domain I consists of a six-stranded β-sheet flanked by eight helices and domain II contains five β-strands and five helices (Fig. 1*b*). A short loosely folded loop borders the domain I entrance to the central cleft. Amino-acid residues 39–140 lie within domain I, domain II contains residues 141–342 and residues 342–403 cross back into domain I, adding two strands and two helices to the domain before ending with a helix which spans the interface between domains and flanks the two crossover β-strands (Figs. 1*a* and 1*b*). The first 15 residues of the N-terminus were not observed in the crystal structure, indicating high flexibility in this region. Another additional flexible region includes the cleft entrance loop (residues 55–63). This is readily demonstrated by the higher temperature factors of this region (Fig. 1*c*).

3.2. Binding of thiamine pyrophosphate

The identification of the TPP cofactor proved to be a significant challenge to solving this structure (Fig. 1*c*). Because of the pyrophosphate-like and planar electron densities, the initial guess was ADP, which was placed into the residual $F_o - F_c$ electron density and subjected to additional cycles of refinement. However, the fit of ADP was unsatisfactory and the resulting electron-density map was inconsistent with a ribose or deoxyribose nucleotide, so the sugar was removed from the model. After subsequent refinement and refitting cycles, a large electron-dense portion of the central ring suggested a methyl thiophene, which was added to the model. The thiophene gave a better fit to the electron density than a five-carbon sugar, but the electron density at the adenine was still ambiguous. A search of nucleotide and cofactor structures suggested thiamine pyrophosphate (TPP), which produced an excellent fit to the unknown density (Fig. 2*a*).

Thiamine pyrophosphate is bound between the two domains and makes numerous interactions with both the main chain and the side chains of the adjacent amino acids. The methyl amino pyrimidine ring is sandwiched between the Trp314 indole plane and the peptide-bond plane between residues Leu379 and Gly380. No interaction between the pyrimidine and amino N atoms and water molecules is apparent, although they are exposed to solvent in the cleft. C2

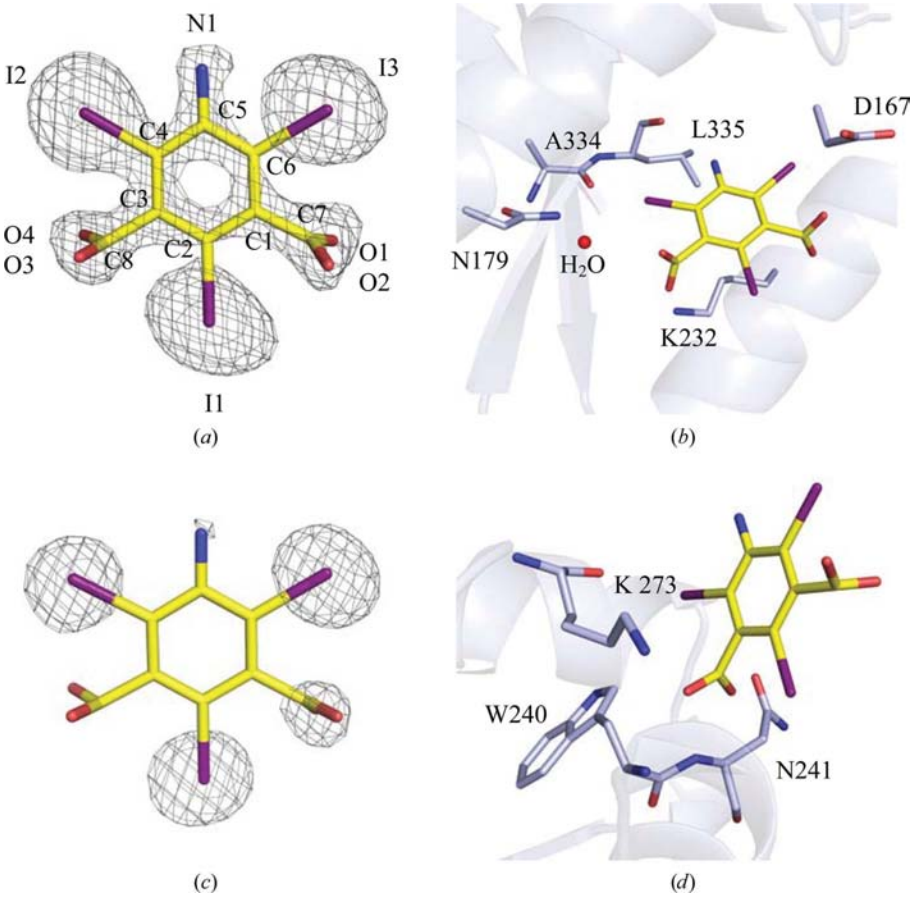

Figure 3
5-Amino-2,4,6-triiodoisophthalic acid binding sites. (*a*) Site *A*: $2|F_o - F_c|$ electron density contoured at 1.7σ. (*b*) Binding interactions of I3C in site *A*. (*c*) Site *B*: $2|F_o - F_c|$ electron density contoured at 1.7σ. (*d*) Binding interactions of I3C in site *B*. Waters are colored red. Amino acids are as labeled. Atom numbering for I3C is as labeled in (*a*). These figures were generated in *PyMOL* (DeLano, 2002).

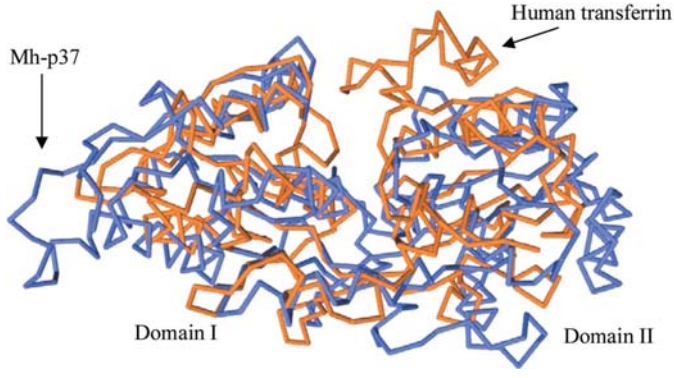

Figure 4
Superimposition of Mh-p37 and human transferrin (PDB code 1jqf; Baker *et al.*, 2001). Mh-p37 is colored blue and tranferrin is colored orange. Viewed as orientated in Fig. 1. This figure was generated using *PyMOL* (DeLano, 2002).

of the thiazole ring interacts with the protein *via* the carbonyl O atom of Leu379 (3.0 Å) and the phenolic O atom of Tyr343 (3.2 Å). A close interaction of Asp344 OD2 with the thiazole S atom of 3.2 Å completes the interactions between the ring atoms and the protein. The side-chain atoms of Tyr215 and Asp344 OD1 are in van der Waals contact with the ethyl C atoms of the TPP (Fig. 2*b*).

Many of the interactions with the protein are made with the pyrophosphate end of the TPP, which is near the base of a helix. The charges on the pyrophosphate are neutralized by two lysine side chains, Lys258 and Lys129. Hydrogen bonds from two waters to the O atoms of the PA phosphate complete the interactions of this phosphate. PB, the terminal phosphate, is involved in many more hydrogen-bonding interactions, including interactions with Lys129 (3.1 Å), the phenolic hydroxyl of Tyr215 (2.6 Å), Ser255 OG (2.6 Å) and the amide N and hydroxyl O atoms of both Ser256 (2.8 Å) and Ser257 (2.7 Å). Ser256 OG bridges to the phosphate O atom *via* a water molecule. It should be noted that Ser256 was originally a phenylalanine in the published amino-acid sequence. In a final note, the pyrophosphate lies adjacent to the positive end of the dipole moment of helix α7 (Fig. 2*b*).

The presence of TPP was unanticipated and its purpose is unknown. The most likely source of the TPP would be the *Escherichia coli* expression system. This leads to the conclusion that TPP must be tightly bound, given that it was not introduced during either purification or crystallization.

3.3. Refinement of the coordinates of the heavy-atom reagent

Using model phases from the complete Mh-p37 structure and the iodine heavy-atom derivative data, the locations of the two I3C compounds (sites *A* and *B*; Fig. 1*c*) were examined to see how they interacted with the protein. The energy-minimized structure for I3C was generated using the *PRODRG*2 website (Schüttelkopf & van Aalten, 2004) and the molecular bond and angle restraints for the heavy-atom compound were generated in *SHELXPRO*.

Both of the I3C molecules were associated with solvent-exposed surfaces on domain II of Mh-p37 (Fig. 1*c*). The site *A* molecule was refined with full occupancy; because of the weaker electron density the site *B* molecule was refined at 0.5 occupancy (Table 2). The full-occupancy I3C molecule (Figs. 3*a* and 3*b*) is located in a cleft defined by two helices, α4 and α5, and β-strand β10 (site *A*; Figs. 1*c* and 3*b*). Hydrogen-bond interactions were observed between N1 of I3C and the carbonyl O atom of Leu335 (3.1 Å), between O4 and

Lys232 NZ (2.5 Å) and between O3 of one carboxyl group through a water to Asn179 ND2. The other carboxyl group induces a rotamer shift of the side chain of Asp167 (compared with the native structure), thus avoiding an unfavorable steric interaction. Of note was an unusually close contact between I2 and Ala334 O (3.0 Å; Fig. 3*b*).

The 0.5 occupancy I3C molecule (site *B*; Figs. 1*c* and 3*d*) was located in a more exposed surface region and was sandwiched between two symmetry-related Mh-p37 molecules (Figs. 3*c* and 3*d*). However, the electron densities of the iodines were still well defined (Fig. 3*c*). The O3 atoms share a hydrogen-bond interaction with the amide N atom of Trp240 (2.9 Å) and Lys273 NZ (3.3 Å) and O4 forms a hydrogen bond to the amide N atom of Asn241 (2.9 Å). In addition, the O2 of the other carboxylate group forms a hydrogen bond to NZ′ of the symmetry-related Lys273 (3.0 Å; Fig. 3*d*).

3.4. Attempts to solve the Mh-p37 crystal structure in retrospect

With the Mh-p37 structure solved, it seemed of interest to address whether it was improvements in software or the use of I3C that led to the successful solution of the Mh-p37 structure. To assess whether alternative heavy-atom phasing methods would have located the I3C in the structure, several alternatives were attempted. *PHENIX AutoSol* was tried (Adams *et al.*, 2002), but the *HASSP* algorithm failed to locate more than two sites. Similarly, the heavy-atom search routine in *CNS* (Brünger *et al.*, 1998) failed to find the correct cluster of sites. It may be that the high temperature factors/low occupancies of the iodine positions caused these Patterson search routines to fail (Table 2). This may have been because both I3C molecules were in similar orientations, *i.e.* orientated in the same planar direction, and therefore the Patterson vectors overlap (Fig. 1*c*).

Once it had been determined that *SHELXD* was ideal for identifying the coordinates of the heavy atoms, the same routine of *SHELXC*, *SHELXD* and *SHELXE* followed by *AutoBuild* was applied to previously collected data from a bromide halide-soaked (Dauter *et al.*, 2000) Mh-p37 crystal (data not shown). Although *SHELXD* was able to identify seven bromines in the structure, the phases generated by *SHELXE* did not result in interpretable electron-density maps and *AutoBuild* failed after 11 cycles. High temperature factors and poor occupancy caused the routine to break down.

As mentioned previously, there is very little sequence identity between Mh-p37 and known protein structures in the PDB (Bernstein *et al.*, 1977), which prevented the use of molecular-replacement phasing techniques (Rossmann, 1990). Once solved, the Mh-p37 structure was put through the *DALI* server (Holm & Sander, 1996) in order to identify proteins that exhibited structural (as opposed to sequence) homology. This search resulted in a multitude of matches (the implications of which will be discussed in a subsequent publication on the theoretical function of Mh-p37) and the closest hit was selected for use in molecular replacement using routines in *CNS* (Brünger *et al.*, 1998). This was human transferrin N-lobe mutant H249Q chain *A* (PDB code 1jqf; Baker *et al.*, 2001), which had the lowest r.m.s.d. value (4.2 Å) for 246 C^{α} positions and 13% sequence identity (Fig. 4). Using the transferrin domain as separate chains failed to give a correct solution to phase the Mh-p37 data. While the secondary-structural topology of transferrin is similar, the overall structure of either domain probably deviated too greatly from that of the Mh-p37 domains for molecular replacement to be successful.

4. Conclusion

Presented here is the high-resolution structure of Mh-p37 as solved using the new heavy-atom derivative I3C. This method was 'magic', as coined by the authors, in its use to phase known structures (Beck *et al.*, 2008). In contrast to the data presented in their paper, the Mh-p37 data collection was at RT from one crystal after a 5 min soak, with relatively low data redundancy and high thermal parameters of the heavy atoms. To our knowledge this is the first *de novo* protein structure to be determined using this method.

The biological implications of this structure and its role in tumorigenesis and cancer progression will be discussed elsewhere.

The authors greatly appreciate Dr Sheldon M. Schuster for introducing us to this fascinating protein, Tobias Beck for his useful guidance in using I3C and the Center of Structural Biology at the University of Florida for its support of the in-house X-ray facility. This work was funded in part by an American Cancer Society grant (No. 58229) to CJR. KS was also funded by the University of Florida, College of Medicine (Alumni Fellowship).

References

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1948–1954.

Baker, H. M., Mason, A. B., He, Q. Y., MacGillivray, R. T. & Baker, E. N. (2001). *Biochemistry*, **40**, 11670–11675.

Beck, T. & Sheldrick, G. M. (2008*a*). *Acta Cryst.* E**64**, o1286.

Beck, T. & Sheldrick, G. M. (2008*b*). *American Crystallographic Association Annual Meeting, Knoxville, TN, USA*. Abstract SP056.

Beck, T., Krasauskas, A., Gruene, T. & Sheldrick, G. M. (2008). *Acta Cryst.* D**64**, 1179–1182.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Bond, C. S. (2003). *Bioinformatics*, **19**, 311–312.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst.* D**56**, 232–237.

DeLano, W. L. (2002). *The PyMOL Molecular Graphics System*. http://www.pymol.org.

Dudler, R., Schmidhauser, C., Parish, R. W., Wettenhall, R. E. & Schmidt, T. (1988). *EMBO J.* **7**, 3963–3970.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Fareed, G. C., Mendiaz, E., Sen, A., Juillard, G. J., Weisenburger, T. H. & Totanes, T. (1988). *J. Biol. Response Mod.* **7**, 11–23.
Gilson, E., Alloing, G., Schmidt, T., Claverys, J. P., Dudler, R. & Hofnung, M. (1988). *EMBO J.* **7**, 3971–3974.
Gong, M., Meng, L., Jiang, B., Zhang, J., Yang, H., Wu, J. & Shou, C. (2008). *Mol. Cancer Ther.* **7**, 530–537.
Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.
Holm, L. & Sander, C. (1996). *Methods Enzymol.* **266**, 653–662.
Huang, S., Li, J. Y., Wu, J., Meng, L. & Shou, C. C. (2001). *World J. Gastroenterol.* **7**, 266–269.
Ketcham, C. M., Anai, S., Reutzel, R., Sheng, S., Schuster, S. M., Brenes, R. B., Agbandje-McKenna, M., McKenna, R., Rosser, C. J. & Boehlein, S. K. (2005). *Mol. Cancer Ther.* **4**, 1031–1038.
Kleywegt, G. J. (2007). *Acta Cryst.* D**63**, 94–100.
Liu, W., Ren, T., Jiang, B., Gong, M. & Shou, C. (2007). *Can. J. Microbiol.* **53**, 270–276.
Otwinowski, Z. & Minor, W. (1997). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
Reutzel, R., Boehlein, S. K., Govindasamy, L., Brenes, R. B., Agbandje-McKenna, M., Schuster, S. M. & McKenna, R. (2002). *Acta Cryst.* D**58**, 2141–2144.
Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 73–82.
Sasaki, H., Igaki, H., Ishizuka, T., Kogoma, Y., Sugimura, T. & Terada, M. (1995). *Jpn J. Cancer Res.* **86**, 791–794.
Schüttelkopf, A. W. & van Aalten, D. M. F. (2004). *Acta Cryst.* D**60**, 1355–1363.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Steinemann, C., Fenner, M., Binz, H. & Parish, R. W. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 3747–3750.