Time-resolved and static-ensemble structural chemistry of hydroxymethylbilane synthase

John R. Helliwell,*†[a] Yeu Perng Nieh,[a] Jarjis Habash,[a] Paul F. Faulder,[a] James Raftery,[a] Michele Cianci,[a] Michael Wulff[b] and Alfons Hädener[c]

[a] *Laboratory of Structural Chemistry, Department of Chemistry, University of Manchester, Manchester, UK M13 9PL*
[b] *ESRF, BP 220 , Grenoble CEDEX, France*
[c] *Department of Chemistry, Universität Basel, CH-4051 , Basel, Switzerland*

Received 6th February 2002, Accepted 9th April 2002
First published as an Advance Article on the web 5th August 2002

The enzyme hydroxymethylbilane synthase (HMBS, EC 4.3.1.8), 313 amino acid residues and MW 34 kDa, also known as porphobilinogen deaminase (PBGD), catalyses the stepwise polymerization of four molecules of porphobilinogen (PBG) to the linear tetrapyrrole 1-hydroxymethylbilane. Several crystallographic structures of HMBS have been previously determined, most recently including by time-resolved Laue protein crystallography of the Lys59Gln mutant form with reaction initiation undertaken by use of a flow cell carrying the substrate PBG. In this paper we review these structures and add new molecular graphics representations and analyses. Moreover we present a new structure refined at 1.66 Å resolution using diffraction data recorded at cryo-temperature (100 K) in an attempt at trapping the polypeptide loop (residues 47 to 58) in the vicinity of the enzyme active site, missing in all previous structure determinations. This loop still has not appeared in the electron density maps, in spite of the advantage of cryo-temperature, but nevertheless the 1.66 Å cryo-structure extends the ensemble of known HMBS structures. The cryo-model of protein, cofactor and 320 bound water molecules has been refined to a final *R*-factor and *R*-free of 0.198 and 0.247 respectively; the PDB deposition codes, coordinates and structure factors are 1GTK and R1GTKSF respectively. Finally a protein comparison study is presented of the *Mycobacterium tuberculosis* (MTb) HMBS, with the *E. coli* HMBS. This has been done as preparation for future structural studies on the MTb HMBS from this important disease bearing organism. The overall amino acid sequence identity is 41%. Most interestingly there is a two-residue reduction in length of the loop referred to above (Asp 50 and Gly 58 being missing in the MTb form). This gives the hope that this loop will be less flexible and thus might become visible to crystallographic analysis.

1 Introduction

The enzyme HMBS is found in all organisms, except viruses. Hydroxymethylbilane (HMB) formed by the enzyme (Fig. 1) is subsequently transformed into uroporphyrinogen III by

† Present address: CCLRC, Director of SR Science, Daresbury Labratory, Warrington WA4 4AD, Cheshire, UK. Email: j.r.helliwell@dl.ac.uk.

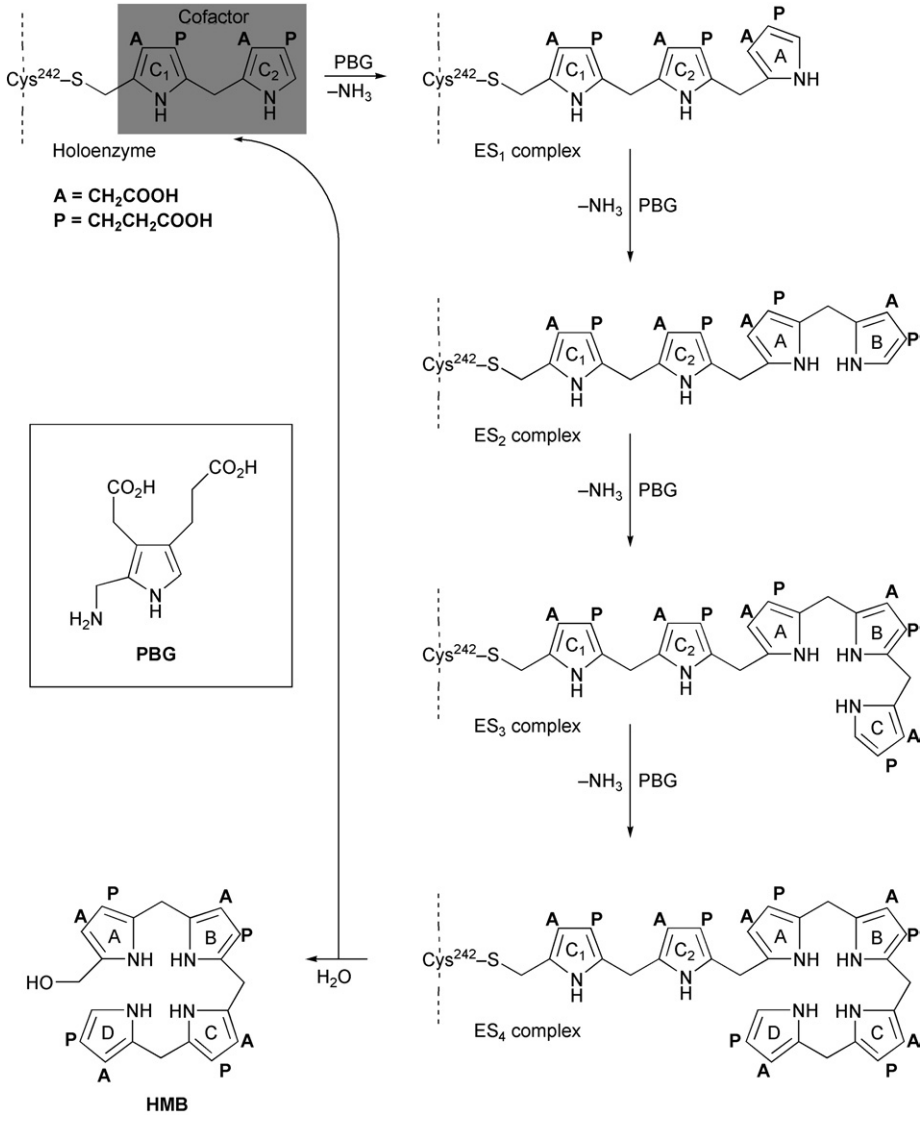

Fig. 1 Overall course of the reaction catalysed by HMBS.

uroporphyrinogen III synthase in a reaction involving rearrangement of ring D and cyclisation. Uroporphyrinogen III is the common precusor of natural tetrapyrroles such as haem, chlorophyll and vitamin B_{12}. *In vitro*, and in the absence of uroporphyrinogen III synthase, however, HMBS cyclises spontaneously (with a half life of *ca.* 4 min at 37 °C and pH 8.25) without rearrangement to give uroporphyrinogen I.[1] In the presence of air, the latter is readily oxidised to uroporphyrin I, which is red in colour. To accomplish the assembly of the bilane, HMBS uses a dipyrromethane cofactor to which the growing chain remains attached during oligomerisation. The cofactor is covalently bound to the enzyme from *Escherichia coli via* the sulfur atom of Cys 242.[2] The catalytic cycle of the HMBS reaction involves the stepwise, head-to-tail polymerisation of four molecules of PBG with expulsion of ammonia at each of these steps and a reaction with water when HMB is finally cleaved from the enzyme.[3,4] Reduction in the activity of this enzyme in humans, due to natural induced mutations, causes the hereditary disease 'acute intermittent porphyria'[5] (the 'madness of King George III').

There are two forms for the enzyme; the reduced 'active' form and the oxidised 'inactive' form.[6] Both crystal structures, at room temperature, of both the catalytically inactive enzyme, carrying the cofactor in an oxidised state, and the catalytically active enzyme, carrying the cofactor in the reduced state, have been determined, respectively, by MIR (1.9 Å room temperature, PDB code 1PDA[7]—in a later study the structure is expanded to 1.76 Å[8]) and selenomethionine MAD X-ray crystallography (2.4 Å room temperature, PDB code 1AH5[9]). The latter study also determined the positions of the C-terminal residues 308–313 not determined by the former. The two dipyrromethane cofactor rings, C1 and C2, are 11° from coplanarity in the oxidised form but 61° apart in the reduced form.

In addition the wild-type active protein structure has been refined at 2.3 Å resolution against Laue protein crystal data (2.3 Å room temperature, PDB code 2YPN[10]). These experimentally determined structures established the structure of the enzyme. Moreover the MIR structure report[7] proposed that the oxidised cofactor ring 2 position was a likely binding site for the first substrate. This is in keeping with earlier evidence that the C2 ring atom C4B is the atom to which the first PBG unit is attached during the catalytic cycle.[11,12] Louie *et al.*[7] also proposed that there would be relative movement of the three domains of the protein to accomplish subsequent, successive, ring coupling steps. The enzyme active site cavity was estimated as being sufficient in any case as being capable of accommodating three and a half PBG molecules.[6]

2 The 3-dimensional structure

The secondary structure is a combination of α-helices, β-strands and loops and the tertiary structure forms three domains[7] (Fig. 2). The first includes residues 4–99 and 200–217, the second 105–193 and the third 222–313.

There is a mobile loop in the polypeptide chain involving residues 42–63 which connects α-helix and β-strand and is important due to its proximity to the active-site cleft. Likewise the cofactor

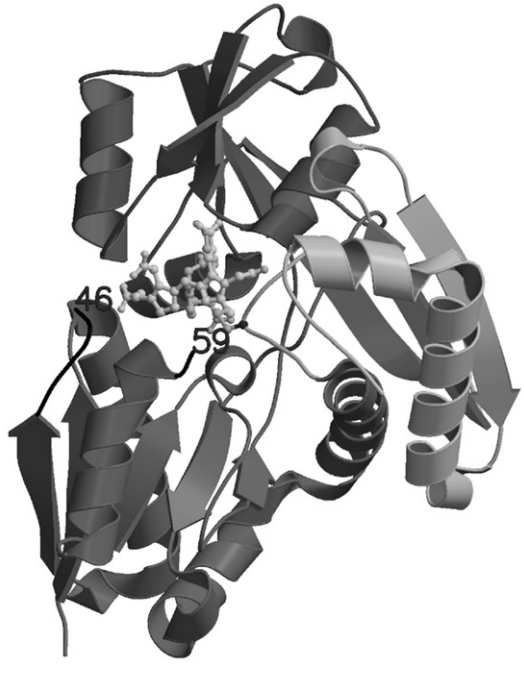

Fig. 2 Ribbon diagram of wild-type HMBS with the cofactor in the reduced catalytically active state (PDB code 1AH5) and the cofactor in the oxidised state (PDB code 1PDA). The lighter shade cofactor (ring C2 is at left and face on) is the oxidised inactive state.

dipyrromethane lies proximal to the cleft between domain 1 and domain 2. Its C1 ring is covalently linked directly to the side chain of Cys 242 which is in a loop on domain 3. The C2 ring atom C4B is the reactive atom towards PBG as substrate. The acetate and propionate side chain groups (marked A and P in Fig. 1) of the cofactor are involved in an extensive network of interactions (salt-bridges and hydrogen bonds) with the polypeptide chain, and which thus cross-links the three domains.[7] Residues Arg 131, Arg 132 and Arg 155, Lys 83, Ser 129 and Thr 127 (largely from domain 2) form the cofactor site C (*i.e.* C1 and C2). Arg 11, Arg 149, Arg 155 and Ser 13 (largely domain 1) form the substrate-binding site S.[6] The carboxyl side chain of catalytic Asp 84 (an invariant residue across species) forms hydrogen bonds with both pyrrole NH groups and is within hydrogen bonding with substrate thus implicated in stabilising intermediate states throughout the reaction cycle.

3 Resume of the time-resolved Laue structural studies

HMBS exhibits several features that made this enzyme suitable for time-resolved crystallography.[13] It is a slow enzyme with $k_{cat} = 0.1\ s^{-1}$ at 310 K and pH 7.5 for the saturated enzyme from *E. coli*,[14] compared to $k_{cat} = 1–10^7\ s^{-1}$ for most enzymes. An even slower reaction rate can be achieved by using a mutant, lower pH or lower temperature.

Thus an ensemble of refined time-resolved Laue protein crystal structures of HMBS K59Q mutant as substrate flowed over the crystal was determined.[15] These were at time points of 1 min, 2 min and 8 min, using data merged from several crystals, but each at essentially identical time-slices, and then, using one crystal each, also at 25 (+/−10) min and 2 h (+/−10 min). In addition a 'static' crystal structure was refined using monochromatic data recorded at 12 h (+/−30 min) *i.e.* 10 h after flow of substrate over a crystal in the flow cell had ceased. These various structures are described in detail.[15,16] A gradual build up of extended electron density running from the position of the oxidised C2 cofactor ring position was observed in the enzyme active site cavity. The elongated electron density was most pronounced at 2 h and this crystal structure was refined and examined in detail and coordinates deposited in the Protein Data Bank (2.3 Å room temperature, PDB code 1YPN[15]). The elongated density fits within the large cavity that forms the active site (Fig. 3). The elongated density at one end sat in the C2 ring oxidised position (Figs. 4 and 5) thus experimentally supporting the proposal of Louie *et al.*[7] that this place was most likely the substrate first PBG molecule binding site. The density commences adjacent to and above residue Asp 84. It then extends past residue Arg 149, then past residue Arg 155 (residues whose mutation causes build up of ES_1 and ES_4 intermediate enzyme–substrate complexes, respectively[17,18]), and out towards the open solvent channel of the crystal. There is also density enhancement around both ends of the 42–63 polypeptide loop *i.e.* residues 42–47 and 58–63 show increased ordering *versus* the wild-type

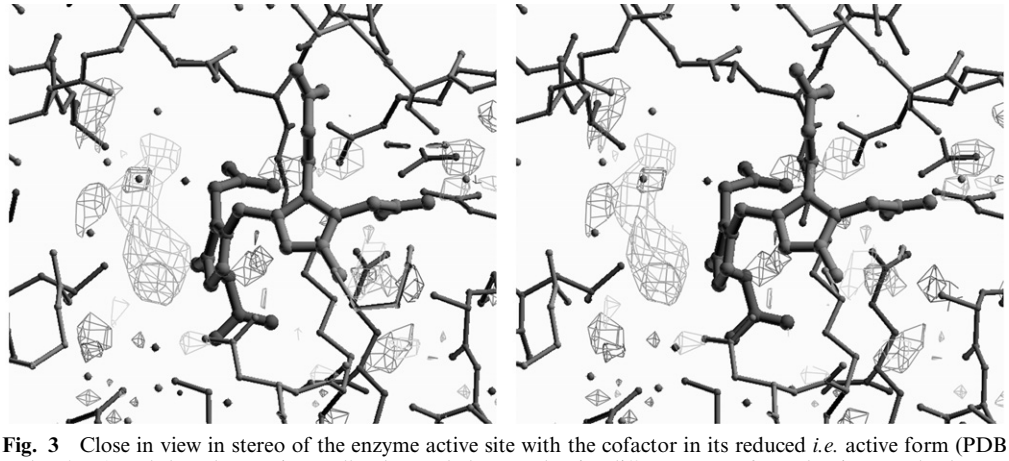

Fig. 3 Close in view in stereo of the enzyme active site with the cofactor in its reduced *i.e.* active form (PDB code 1AH5) and the 2 h experimentally observed electron density difference map from the time-resolved Laue diffraction studies included.

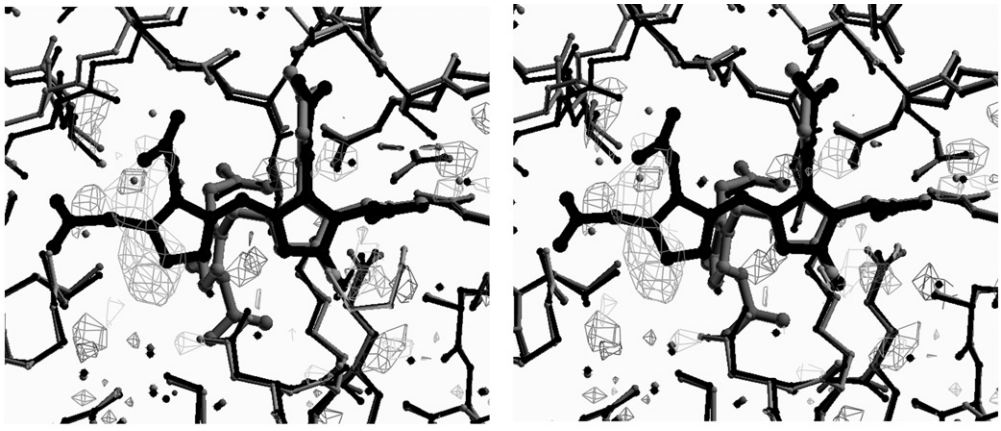

Fig. 4 As Fig. 3 but with the model of oxidised cofactor (in black) added. The start of the elongated density overlaps nicely with the C2 oxidised ring position (PDB code 1PDA).

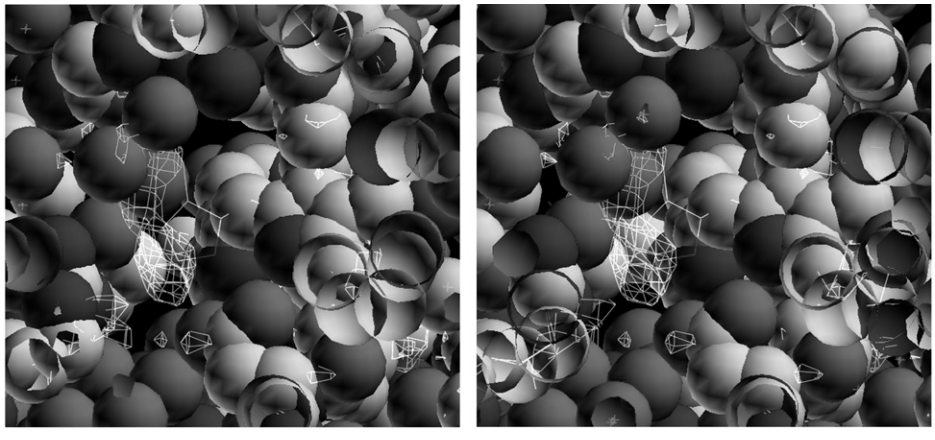

Fig. 5 As Figs. 3 and 4 but with the protein (PDB code 1AH5), and active cofactor, in space filling mode. Carbons in light grey, oxygens in darker grey and nitrogens as darkest. The elongated density and oxidised C2 ring (shown in 'stick representation'; PDB code 1PDA) sit nicely in the cavity void that is the enzyme active site.

reduced form structure (2YPN). However residues 48–57 do not obviously become visible. There are also electron density peaks sitting on some key residues near the active site (*e.g.* Cys 242, Arg 149, Arg 11 and Arg 132) and on the cofactor showing that the active site becomes more rigid 'at 2 h'. After flow of substrate had ceased, a situation represented by the data set at 12 h, the elongated electron density feature had disappeared, as had the density enhancement of the loop region. By this time, and for the later Laue data sets, the crystal had gone a pronounced red colour characteristic of HMB fully released and cyclised to uroporphyrin.

4 Structure of wild-type active form HMBS at cryo-temperature in order to try to trap the 'missing polypeptide loop (48–57)'

4.1 Crystallization

Crystallization[9,10,14] was carried out at pH 5.3 by the use of sitting drops of 50 μl volume. Each drop contained 6–7 mg ml^{-1} of protein, 0.3 mM EDTA, 15 mM dithiothreitol, 10% (w/v) polyethylene glycol 6000 and 0.01% NaN_3 in 0.1 M sodium acetate. This was equilibrated by vapour

diffusion at 293 K against a reservoir containing 10–20 mg of solid dithiothreitol. Crystallization was undertaken afresh at the ESRF/EMBL laboratories in France immediately in the weeks prior to data collection. Crystals obtained in this way contained active, reduced, HMBS and were colourless as expected,[14] whereas crystals of inactive HMBS containing the oxidised cofactor are yellow in colour.[8]

4.2 Data collection and analysis

Data were collected at the ESRF, beamline 3 (ID09). A single crystal at cryo-temperature was exposed to X-rays of wavelength 0.8611 Å and data recorded with a high read out image intensifier/CCD detector system. Two subsets of data were collected. The first included 135° of low resolution data images recorded in 3° oscillations per image, at a long distance of the detector from the crystal so as to capture the reflection spots bordering the ice rings that unfortunately were present due to lack of an ideal cryo-protectant. [This latter also led to diffraction resolution (and *B* factors) not superior to the previous best room temperature cases.] The second data set comprised 135° of high resolution data images in 1° oscillations per image. The data were processed in DENZO and merged in SCALEPACK.[19] The space group was $P2_12_12$ and the refined unit cell parameters were $a = 87.5$, $b = 75.9$ and $c = 50.1$ Å. The statistics for data reduction for the two subsets are shown in Table 1. Comparison of Table 1a and 1b confirms that the low resolution ice ring around 3.5 Å does not affect completeness in the 'low resolution' data set. A final composite data set for use in the refinement of the structure was made by combining the two subsets. Table 2 shows details for this final data set; the completeness breakdown is reasonable although some reflections have still been lost at low resolution presumably due to spots saturating the detector.

4.3 Structure refinement

The starting model was that of the wild-type native structure determined at room temperature at 2.3 Å resolution (2YPN). This was subjected to rigid body refinement along with overall temperature factor refinement (5 cycles) using CCP4 'Refmac5'.[20] Assessment of phase accuracy was made by cross validation and was applied throughout the refinement.[21] The overall *R*-factor, *R*-free and figure of merit (FoM) reached 0.262, 0.290 and 0.727 respectively. The input and output models in this rigid body refinement exhibited high temperature factors for the atoms of the residues 3, 42, 60, 61, 62 and 305–313. The model then was subjected to restrained isotropic temperature factor refinement (5 cycles). The overall *R*-factor and *R*-free dropped to 0.212 and 0.252 respectively and overall FoM rose to 0.761. By generating Fo–Fc and 2Fo–Fc maps *via* the FFT programme[20] and inspecting the 2Fo–Fc map using the O molecular graphics system[22] it became evident that at 1 r.m.s. these same residues represented poorly defined density of the model structure. Careful inspection of further Fo–Fc and 2Fo–Fc maps was made to check for any linking electron density between residues 42 and 60 that would define the missing loop *i.e.* residues 43 to 59, but no meaningful density was found. For the third iteration of the model refinement, TLS *B*-factor refinement (10 cycles) was made, and the overall *R*-factor, *R*-free and overall FoM converged at 0.208, 0.248 and 0.771 respectively. At this stage Lys 26, Met 41, Ser 128 and Arg 184 were partially modelled according to Fo–Fc and 2Fo–Fc densities.

The refinement was carried out for a further 10 iterations essentially to define the bound water structure using the ARP/wARP programme.[20] The details of each iteration of the refinement are shown in Table 3 and the final parameters in Table 4. The final model contained 320 bound water molecules and the overall *R*-factor, *R*-free and overall FoM were 0.198, 0.247 and 0.767 respectively. Fig. 6 shows the variation of the reliability index *R* as a function of each iteration in the cryo-refinement.

4.4 Results and discussion of the cryo-structure

No meaningful density (either in Fo–Fc or 2Fo–Fc) was observed for the final model linking the missing loop between residues 42 and 60. The cofactor dipyrromethane generally shows continuous 2Fo–Fc density but the pyrromethane C1 ring is more resolved than the C2 ring, as is reflected in the temperature factor values. The average temperature factor for the C1 pyrromethane atoms is 27.0 $Å^2$, but 30.4 $Å^2$ for the C2 ring. Cys 242 is within a mobile region but is well resolved despite

Table 1 Cryo-data reduction statistics

(a) Low resolution subset

Resolution shells/Å	*R*-merge	Number of reflections	Completeness (%)
100.00–6.79	0.041	609	89.4
6.79–5.39	0.039	571	91.9
5.39–4.71	0.035	573	92.7
4.71–4.27	0.033	559	93.6
4.27–3.97	0.056	554	93.1
3.97–3.73	0.041	564	94.0
3.73–3.55	0.058	538	91.5
3.55–3.39	0.052	548	93.8
3.39–3.26	0.061	566	94.8
3.26–3.15	0.064	560	96.1
3.15–3.05	0.067	552	93.9
3.05–2.96	0.082	568	96.4
2.96–2.89	0.206	541	92.6
2.89–2.82	0.183	507	88.8
2.82–2.75	0.114	418	71.1
2.75–2.69	0.050	30	5.1
All shells	0.052	8258	86.4

(b) High resolution subset

Resolution shells/Å	*R*-merge	Number of reflections	Completeness %
100.00–4.48	0.062	2012	89.9
4.48–3.55	0.058	766	36.2
3.55–3.11	0.052	1835	88.3
3.11–2.82	0.052	1952	95.1
2.82–2.62	0.052	1974	96.1
2.62–2.46	0.057	1958	95.9
2.46–2.34	0.063	1986	96.5
2.34–2.24	0.082	685	33.7
2.24–2.15	0.113	1286	63.0
2.15–2.08	0.138	1968	97.8
2.08–2.01	0.167	1983	97.7
2.01–1.96	0.211	1953	96.3
1.96–1.90	0.239	1278	63.7
1.90–1.86	0.332	1899	94.1
1.86–1.82	0.320	1974	97.4
1.82–1.78	0.324	1950	97.2
1.78–1.74	0.359	1884	92.6
1.74–1.71	0.360	1710	85.3
1.71–1.68	0.344	1593	80.7
1.68–1.65	0.364	492	24.3
All shells	0.077	33 138	81.1

exhibiting a rather high temperature factor and is covalently bonded to the CH atom of the C1 ring of the cofactor (CH–SG bond length 1.87 Å). The COO^- group of the catalytically important residue Asp 84 is within hydrogen bonding distances of the two NH groups of the dipyrromethane cofactors, and is sitting in a nice well-formed electron density (Fig. 7). The hydrogen bond distances OD1–N1 (C1 ring) and OD2–N1 (C2 ring) are 2.88 and 3.20 Å respectively. The Asp 84 side chain is also within hydrogen bonding distance to the putative substrate binding site.

The bound water structure comprises 320 water molecules mostly with well defined densities. The average temperature factor for the waters is 35.6 $Å^2$ with a minimum and maximum *B* of 18.5 and 61.8 $Å^2$ respectively. The bulk of the waters lie within a *B* value range of 20–50 $Å^2$.

An Fo–Fc omit density map was generated with the cofactor, Asp 84 and Cys 242 removed from the model. These residues play key roles in the catalytic process for the enzyme and are shown in the omit map (Fig. 7).

Table 2 Statistics for the combined dataset

Resolution shells/Å	Number of reflections	Completeness (%)
57.74–5.23	1300	90.7
5.23–3.70	2258	93.3
3.70–3.03	2884	94.1
3.03–2.62	3393	94.5
2.62–2.35	3911	96.3
2.35–2.14	2337	52.4
2.14–1.98	4708	97.5
1.98–1.85	4262	82.9
1.85–1.75	5273	95.9
1.75–1.66	4091	70.7
All shells	34 417	87

4.5 Alignment of *E. coli* HMBS and *Mycobacterium tuberculosis* HMBS amino acid sequences

The homology modelling for *Mycobacterium tuberculosis* HMBS was done with Swiss-Model and PDBViewer[23,24] and BLAST2.0.[25] The amino acid sequence alignment for *M. tuberculosis* HMBS with *E. coli* HMBS was optimized with BLAST2.0. The overall sequence identity is 41%. (Table 5 illustrates the results.) This information was used, together with coordinates of the related crystal structure (PDB code 1AH5,[9]), to derive with Swiss-Model and PDBViewer a theoretical model for *Mycobacterium tuberculosis* HMBS. The core parts of the structure appear to be highly conserved. The catalytic residues Arg 11, Asp 84, Arg 132, Arg 149, Cys 242 (using 1AH5 sequence numbering) have conserved positions. There are two residue deletions (His 31 and Pro 32) for the loop region 29–36 and nine residue insertions in the loop region 255–256. The most interesting feature, however is in the loop region 46–59 that is shorter by two residues (Asp 50 and Gly 58). This aspect could result in the loop region 46–59 being more rigid and then reducing the turn-over rate in changing its conformation.

Mycobacterium tuberculosis HMBS could then be a better target for time-resolved crystallographic studies and crystallographic studies in order to understand the role of the loop region 46–59 in the reaction mechanism.

5 Discussion and concluding remarks

The time-resolved Laue crystallographic studies of Lys59Gln HMBS have opened up a completely new window on the structural chemistry of the HMBS enzyme system. Since the monitoring of the enzyme substrate intermediate complexes (ES_1, ES_2, ES_3 and ES_4, which are not coloured) individually by microspectrophotometry is not possible, the time evolution of this enzyme system functioning in the crystalline state has been established by Laue diffraction. Some success has been achieved by this approach. By 2 h *i.e.* much later than anticipated (2 min for the build up of ES_2 and 10 min for ES_3 had been expected[3] based on kinetic studies of immobilised Lys59Gln HMBS) elongated electron density of substantial volume appeared in the active site.[15] The slow build-up could be due to the pH of the crystal mother liquor being non-optimal, or the crystalline state itself. The density is located adjacent to and above the critical Asp 84 side chain and at the oxidised C2 ring position of the cofactor previously proposed to indeed be the PBG binding site position. It then extended towards Arg 149 and past Arg 155, amino acid residues whose mutation causes build up of ES_1 and ES_4 respectively, and then out towards the crystal open solvent channel. As was discussed in ref. 15 previously, whilst the first pyrrole ring and its side chains overlapped the start of the elongated density in some detail, the subsequent pyrrole rings do not have density for their side chains. It was noted then that the elongated density could possibly be a mixture of bound PBG and ordering of the missing loop *i.e.* residues 48–57. Nevertheless the construction of a Beevers molecular model (not shown here but available at the FD122 meeting itself) allowed the exploration of the various positions of the mobile loop whilst tethered at its known end points (residues 46 and 58), which was generally distant from the position of the elongated electron density. This, and the juxtaposition of the residues Arg 149 and Arg 155, favoured the

Table 3 Iterations of the cryo-model refinement

Iteration	Refinement	*R*-factor	*R*-free	FoM	Molecular graphics
1	Rigid body, B = overall	0.262	0.290	0.727	–
2	Restrained, B = isotropic	0.212	0.252	0.761	–
3	Restrained, B = TLS, B = isotropic	0.208	0.248	0.771	–
4	Restrained, B = TLS, B = isotropic	0.207	0.245	0.770	Fitting Lys 26, Met 41, Ser 128, Arg 184. Deletion 66 waters.
5	Restrained, B = TLS, B = isotropic	0.222	0.255	0.764	Deletion all waters. Input 129 waters calculated by DDQ.[26]
6	Restrained, B = TLS, B = isotropic	0.221	0.257	0.763	Fitting Glu 305. Total waters 132 from DDQ.
7	Restrained, B = TLS, B = isotropic	0.208	0.250	0.769	ARP/wARP determines 55 waters, removes 1. Total waters 186.
8	Restrained, B = TLS, B = isotropic	0.203	0.246	0.768	ARP/wARP determines 55 waters, removes 4. Total waters 237.
9	Restrained, B = TLS, B = isotropic	0.199	0.245	0.769	Fitting Glu 305. ARP/wARP determines 50 waters, removes 3. Total waters 284.
10	Restrained, B = TLS, B = isotropic	0.196	0.243	0.768	ARP/wARP determines 40 waters, removes 3. Total waters 321.
11	Restrained, B = TLS, B = isotropic	0.196	0.245	0.766	ARP/wARP determines 25 waters, removes 5. Total waters 341.
12	Restrained, B = TLS, B = isotropic	0.196	0.246	0.767	ARP/wARP determines 5 waters, removes 2. Deletion 17 waters. Total waters 327.
13	Restrained, B = TLS, B = isotropic	0.198	0.247	0.767	ARP/wARP determines 1 waters, removes 2. Deletion 6 waters. Total waters 320.

Table 4 Final iteration cryo-model refinement parameters

Resolution range high	1.66 Å
Resolution range low	43.85 Å
Data cutoff	None
Completeness for range	87%
Number of reflections in refined set	32 708
Free *R* value test set size	5%
Free *R* value test set count	1709
Number of non-hydrogen atoms used in refinement	2578
Mean *B* value	22.2 Å^2
Estimated overall coordinate error	
Coordinate error based on *R* value	0.118 Å
Coordinate error based on free *R* value	0.122 Å
Coordinate error based on maximum likelihood	0.066 Å
Error for *B* values based on maximum likelihood	1.946 Å^2
Correlation coefficient Fo–Fc	0.961
Correlation coefficient Fo–Fc free	0.937
Bond lengths refined atoms (r.m.s.)	0.021 Å
Bond angles refined atoms (r.m.s.)	2.0°
Torsion angles (r.m.s.)	5.0°
Non-bonded contacts refined atoms (r.m.s.)	0.281 Å
Isotropic thermal factor restraints (r.m.s.)	
Main-chain bond refined atoms	1.17 Å^2
Main-chain angle refined atoms	1.98 Å^2
Side-chain bond refined atoms	3.14 Å^2
Side-chain angle refined atoms	4.70 Å^2
Ramachandran plot statistics[a]	
Residues in most favoured regions	91.6%
Residues in additional allowed regions	7.2%
Residues in generously allowed regions	0.8%
Residues in disallowed regions	0.4%

[a] Calculated *via* Procheck. The rest of the values were determined by Refmac5 (both of the CCP4 programme suite[20]).

interpretation of the elongated density as not only bound PBG but also bound HMB. The lack of bridging density to the C4B position of ring C2 and steric arguments (bond angles) preclude interpretations in terms of any of the four covalent enzyme–substrate complexes ES_1, ES_2, ES_3 or ES_4, but allow interpretations in terms of ES and EP.[3] In particular, the red colour of the crystals at 2 h, due to uroporphyrin I, presumes the presence of HMB within the crystals. The structure at 2 h (PDB code 1YPN) may therefore allow a glimpse of a Michaelis-type enzyme–substrate and/or enzyme–product complex of HMBS.

A 'static' cryo-temperature HMBS native form crystal structure has been presented here. Again, unfortunately, the 'missing loop' is still not determined. The ensemble of HMBS structures has been extended however in the process (Fig. 8 shows the ensemble of positions of the side chain of Arg 149 which shows quite some variation, and is in the vicinity of the elongated 2 h Laue electron density at the end remote from the C2 position), and the coordinates deposited in the Protein Data Bank (1.66 Å cryo-temperature, PDB code 1GTK). In addition to this new experiment we also have presented a new space filling rendering of the HMBS structure with the "2 h" elongated electron density and the C2 ring oxidised position included (Fig. 5). This then both confirms the Beevers molecular model work referred to above and extends it *via* space filling rendering of the structure on the graphics. The elongated density sits nicely in the void that is the active site region. As discussed at the CIBA Foundation Symposium[6] the void is big enough in volume without interdomain movement to accommodate up to three and a half pyrroles. Presumably the final step of the full HMB being synthesised must stimulate interdomain movement to release the HMB after the final EP complex (enzyme product complex) is formed. However, as emphasised by Louie *et al.*,[7] interdomain movement has also been suggested in the formation of ES_2 onwards due to the presence of only one conserved Asp (Asp 84) in the active site region. Such studies of interdomain movement, as remarked by Helliwell *et al.*,[15] would have to involve a quite different experimental

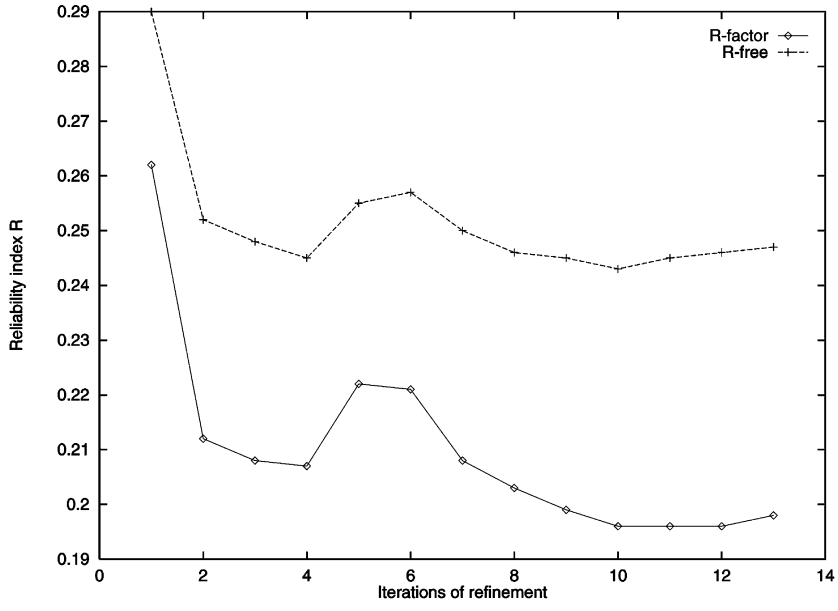

Fig. 6 Reliability index *R versus* iterations of the cryo-refinement.

approach such as time-resolved solution X-ray scattering *i.e.* so as to avoid the constraints of the crystal lattice.

Finally we suggest two further approaches to seek understanding of the structural chemistry of this fascinating, important, molecular system whose complexity verges on that of a molecular machine. Firstly, as NMR continues to expand its scope in molecular weight capability, perhaps NMR solution studies could be harnessed to study this 34 kDa protein, although the time needed to record a 3D NMR pulse sequence would prevent perhaps the necessary fast glimpse of the EP complex that would be required. Secondly, genomic sequencing data are accumulating. Thus the knowledge of which amino acid residues are conserved across species continues to expand. A putative HMBS has been identified for example in the *Mycobacterium tuberculosis* genome and its

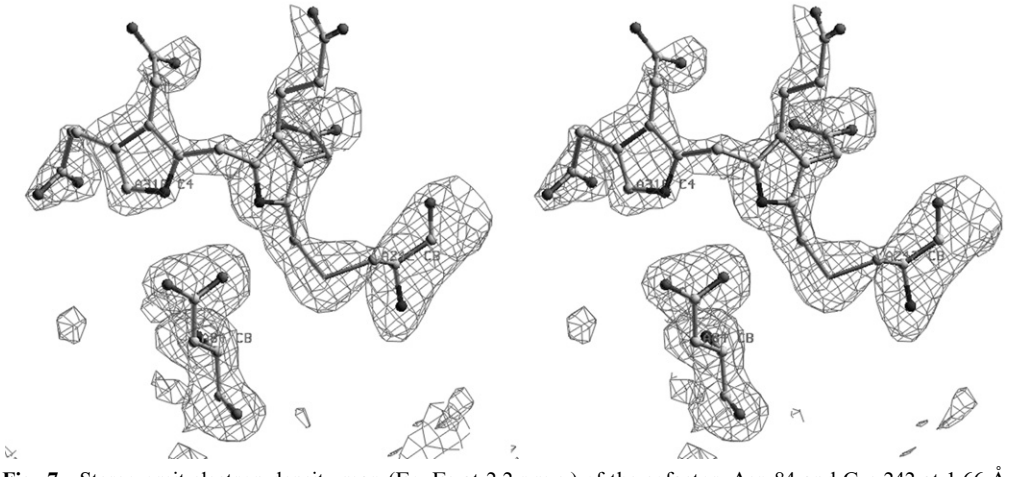

Fig. 7 Stereo omit electron density map (Fo–Fc at 2.2 r.m.s.) of the cofactor, Asp 84 and Cys 242 at 1.66 Å resolution *i.e.* for the 'cryo-structure'.

Table 5 Alignment of HMBS amino acid sequences from *E. coli* and *Mycobacterium tuberculosis* (SWISS-PROT entries P06983 and Q11173, respectively). Vertical bars denote identical amino acids

E. coli	1	MLDNVLRIAT RQSPLALWQA HYVKDKLMAS HPGLVVELVP MVTRGDVILD TPLAKVGGKG
M. tuberculosis	1	----MIRIGT RGSLLATTQA ATVRDALIAG --GHSAELVT ISTEGDRSM- APIASLG-VG
E. coli	61	LFVKELEVAL LENRADIAVH SMKDVPVEFP QGLGLVTICE REDPRDAFVS NNYDSLDALP
M. tuberculosis	53	VFTTALREAM EAGLVDAAVH SYKDLPTAAD PRFTVAAIPP RNDPRDAVVA RDGLTLGELP
E. coli	121	AGSIVGTSSL RRQCQLAERR PDLIIRSLRG NVGTRLSKLD NGEYDAIILA VAGLKRLGLE
M. tuberculosis	113	VGSLVGTSSP RRAAQLRALG LGLEIRPLRG NLDTRLNKVS SGDLDAIVVA RAGLARLGRL
E. coli	181	SRIRAALPPE ISLPAVGQGA VGIECRLDDS RTRELLAALN HHETALRVTA ERAMNTRLEG
M. tuberculosis	173	DDVTETLEPV QMLPAPAQGA LAVECRAGDS RLVAVLAELD DADTRAAVTA ERALLADLEA
E. coli	241	GCQVPIGSYA ELIDG----- ----EIWLRA LVGAPDGSQI IRGERRGAPQ DAEQMGISLA
M. tuberculosis	233	GCSAPVGAIA EVVESIDEDG RVFEELSLRG CVAALDGSDV IRASGIGSCG RARELGLSVA
E. coli	292	EELLNNGARE ILAEVYNGDA PA
M. tuberculosis	293	AELFELGARE LMWGVRH--- --

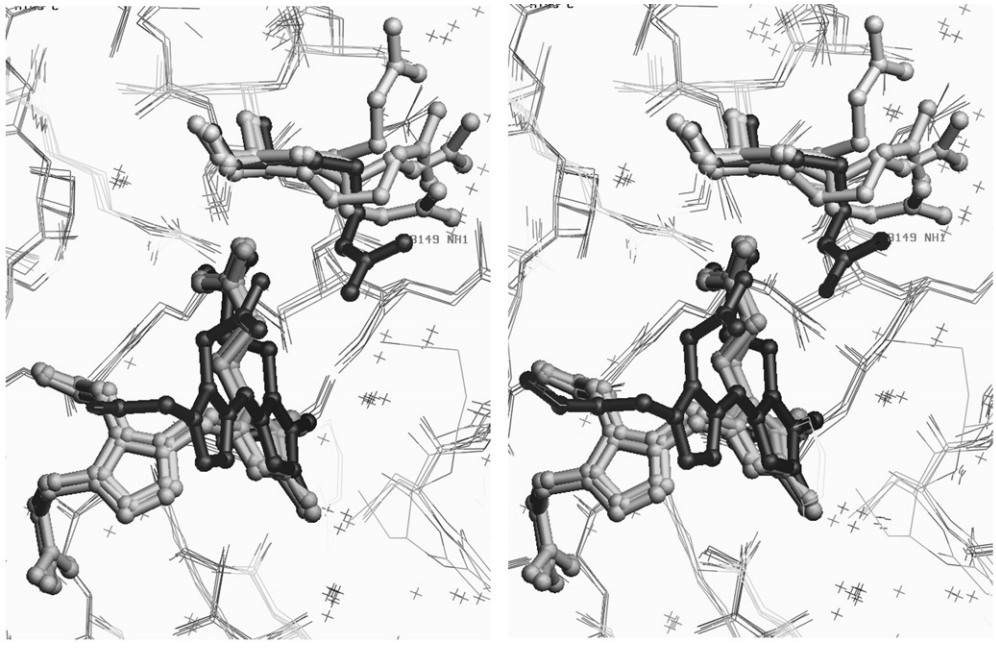

Fig. 8 The ensemble of known HMBS enzyme crystal structures. Again the cofactor conformers sit essentially in the centre of the graphics picture with those atoms highlighted (in black = the oxidised inactive form PDB code 1PDA). Also highlighted in solid stick (with atoms as spheres) is the side chain of residue 149 (as labelled) which displays a range of positions (PDB codes 1AH5, 2YPN, 1YPN, 1GTK and 1PDA, the latter is in black).

3D structure targeted for study (see http://www.nwsgc.ac.uk/), with a comparison study presented here for the first time between MTb and *E. coli* HMBS cases.

Acknowledgements

JRH thanks The Leverhulme Trust for a grant award (partial salary support of JH) and EPSRC for a studentship award to PFF. The BBSRC and The Wellcome Trust are thanked for a research grant to JRH for the Manchester Structural Chemistry computing suite and partial salary support of JR (BBSRC). AH gratefully acknowledges financial support from the Swiss National Science Foundation and the Ciba-Geigy-Jubiläums-Stiftung. Thanks are also extended to ESRF in Grenoble for the provision of synchrotron radiation at beamline ID09 'monochromatic mode' for the cryo-temperature data set collection, and the time-resolved Laue work was undertaken also at ESRF ID09 'Laue mode', both under beamtime awards to AH and JRH. The time-resolved data analyses, which have been reviewed here, were undertaken mainly by Dr Y. P. Nieh in Manchester whilst a PhD student with JRH, and funded by the ORS and the Nieh family. YPN is now based at the Fred Hutchinson Research Center in Seattle.

References

1 A. R. Battersby, C. J. R. Fookes, K. E. Gustafson-Potter, E. McDonald and G. W. J. Matcham, *J. Chem. Soc., Perkin Trans. 1*, 1982, 2427.
2 G. J. Hart, A. D. Miller and A. R. Battersby, *Biochem. J.*, 1988, **252**, 909.
3 A. C. Niemann, P. K. Matzinger and A. Hädener, *Helv. Chim. Acta*, 1994, **77**, 1791.
4 P. M. Anderson and R. J. Desnick, *J. Biol. Chem.*, 1980, **255**, 1993.
5 A. Kappas, S. Sassa, R. A. Galbraith and Y. Nordmann, in *The porphyrias, in the metabolic and molecular bases of inherited disease*, ed. C. R. Scriver, A. L. Beaudet, W. S. Sly and D. Valle, vol. II, McGraw-Hill Inc, New York, 1995, p. 2103.

6 *The Biosynthesis of the Tetrapyrrole Pigments*, Ciba Foundation Symposium 180, John Wiley & Sons, 1994, p. 70 and p. 97.
7 G. V. Louie, P. D. Brownlie, R. Lambert, J. B. Cooper, T. L. Blundell, S. P. Wood, M. J. Warren, S. C. Woodcock and P. M. Jordan, *Nature (London)*, 1992, **359**, 33.
8 G. V. Louie, P. D. Brownlie, R. Lambert, J. B. Cooper, T. L. Blundell, S. P. Wood, V. N. Malashkevich, A. Hädener, M. J. Warren and P. M. Shoolingin-Jordan, *Proteins: Struct., Genet.*, 1996, **25**, 48.
9 A. Hädener, P. K. Matzinger, A. R. Battersby, S. McSweeney, A. W. Thompson, A. P. Hammersley, S. J. Harrop, A. Cassetta, A. Deacon, W. N. Hunter, Y. P. Nieh, J. Raftery, N. Hunter and J. R. Helliwell, *Acta Crystallogr.*, 1999, **D55**, 631.
10 Y. P. Nieh, J. Raftery, S. Weisgerber, J. Habash, F. Schotte, T. Ursby, M. Wulff, A. Hädener, J. W. Campbell, Q. Hao and J. R. Helliwell, *J. Synchrotron Rad.*, 1999, **6**, 995.
11 G. J. Hart, A. D. Miller, F. J. Leeper and A. R. Battersby, *J. Chem. Soc., Chem. Commun.*, 1987, 1762.
12 P. M. Jordan and M. J. Warren, *FEBS Lett.*, 1987, **225**, 87.
13 J. R. Helliwell, Y. P. Nieh, A. Cassetta, J. Raftery, A. Hädener, A. C. Niemann, A. R. Battersby, P. D. Carr, M. Wulff, T. Ursby, J. P. Moy and A. W. Thompson, in *Time-Resolved Diffraction*, ed.. J. R. Helliwell and P. M. Rentzepis, Oxford University Press, 1997, p. 187.
14 A. Hädener, P. K. Matzinger, V. N. Malashkevich, G. V. Louie, S. P. Wood, P. Oliver, P. R. Alefounder, A. R. Pitt, C. Abell and A. R. Battersby, *Eur. J. Biochem.*, 1993, **211**, 615.
15 J. R. Helliwell, Y.-P. Nieh, J. Raftery, A. Cassetta, J. Habash, P. D. Carr, T. Ursby, M. Wulff, A. W. Thompson, A. C. Niemann and A. Hädener, *J. Chem. Soc., Faraday Trans.*, 1998, **94**, 2615.
16 Y. P. Nieh, PhD Thesis, University of Manchester, UK, 1997.
17 M. Lander, A. R. Pitt, P. R. Alefounder, D. Bardy, C. Abell and A. R. Battersby, *Biochem. J.*, 1991, **275**, 447.
18 P. M. Jordan and S. C. Woodcock, *Biochem. J.*, 1991, **280**, 445.
19 Z. Otwinowski and W. Minor, *Methods Enzymol.*, 1997, **276**, 307.
20 CCP4 Collaborative Computational Project, Number 4, *Acta Crystallogr.*, 1994, **D50**, 760.
21 A. T. Brünger, *Acta Crystallogr.*, 1993, **D49**, 24.
22 T. A. Jones, J. Y. Zou, S. W. Cowan and M. Kjeldgaard, *Acta Crystallogr.*, 1991, **A47**, 110.
23 N. Guex and M. C. Peitsch, *Electrophoresis*, 1997, **18**, 2714.
24 M. C. Peitsch, *Biochem. Soc. Trans.*, 1996, **24**, 274.
25 S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman, *Nucleic Acids Res.*, 1997, **25**, 3389.
26 F. van den Akker and W. G. J. Hol, *Acta Crystallogr.*, 1999, **D55**, 206.