Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

An investigation into the protonation states of the C1 domain of cardiac myosin-binding protein C

S. J. Fisher,[a,b] J. R. Helliwell,[a,c]* S. Khurshid,[d] L. Govada,[d] C. Redwood,[d]‡ J. M. Squire[d,e] and N. E. Chayen[d]

[a]School of Chemistry, Brunswick Street, The University of Manchester, Manchester M13 9PL, England, [b]Institut Laue Langevin, BP 156, 6 Rue Jules Horowitz, 38042 Grenoble CEDEX 9, France, [c]STFC Daresbury Laboratory, Warrington, Cheshire WA4 4AD, England, [d]Department of Biomolecular Medicine, Division of SORA, Faculty of Medicine, Imperial College London, Sir Alexander Fleming Building, London SW7 2AZ, England, and [e]Muscle Contraction Group, Department of Physiology and Pharmacology, University of Bristol, Bristol BS8 1TD, England

‡ Current address: Department of Cardiovascular Medicine, University of Oxford, West Wing Level 6, John Radcliffe Hospital, Oxford OX3 9DU, UK.

Correspondence e-mail: john.helliwell@manchester.ac.uk

Received 19 February 2008
Accepted 1 April 2008

PDB Reference: cMyBP-C C1 domain, 3cx2, r3cx2sf.

Myosin-binding protein C (MyBP-C) is a myofibril-associated protein found in cardiac and skeletal muscle. The cardiac isoform (cMyBP-C) is subject to reversible phosphorylation and the surface-charge state of the protein is of keen interest with regard to understanding the inter-protein interactions that are implicated in its function. Diffraction data from the C1 domain of cMyBP-C were extended to 1.30 Å resolution, where the $\langle I/\sigma(I)\rangle$ of the diffraction data crosses 2.0, using intense synchrotron radiation. The protonation-state determinations were not above 2σ (the best was 1.81σ) and therefore an extrapolation is given, based on 100% data completeness and the average DPI, that a 3σ determination could be possible if X-ray data could be measured to 1.02 Å resolution. This might be possible *via* improved crystallization or multiple sample evaluation, *e.g.* using robotics or a yet more intense/collimated X-ray beam or combinations thereof. An alternative would be neutron protein crystallography at 2 Å resolution, where it is estimated that for the unit-cell volume of the cMyBP-C C1 domain crystal a crystal volume of 0.10 mm^3 would be needed with fully deuterated protein on LADI III. These efforts would optimally be combined in a joint X-ray and neutron model refinement.

1. Introduction

Myosin-binding protein C (MyBP-C) is a myofibril-associated protein that is found in cardiac and skeletal muscle. The cardiac muscle isoform (cMyBP-C) is subject to reversible phosphorylation. cMyBP-C is a substrate *in vivo* and *in vitro* of cAMP-dependent protein kinase (PKA) and calcium/phospholipid-dependent protein kinase (PKC) (Mohamed *et al.*, 1998). Mutations of *MYBP3*, which encodes cMyBP-C, can result in the formation of truncated proteins that lack the C-terminal domains necessary for binding to the thick filament backbone or to mis-sense mutations (Govada *et al.*, 2008). Together, these mutations are a major cause of hypertrophic cardiomyopathy (HCM) and account for at least a quarter of affected individuals.

The surface-charge state of the protein is of keen interest with regard to understanding the inter-protein interactions that are implicated in its function. Govada *et al.* (2008) have reported the X-ray crystal structure of the C1 domain of cMyBP-C at 1.55 Å resolution; a protein charge-surface representation was given based on the assumption that each Asp and Glu is deprotonated (Fig. 1).

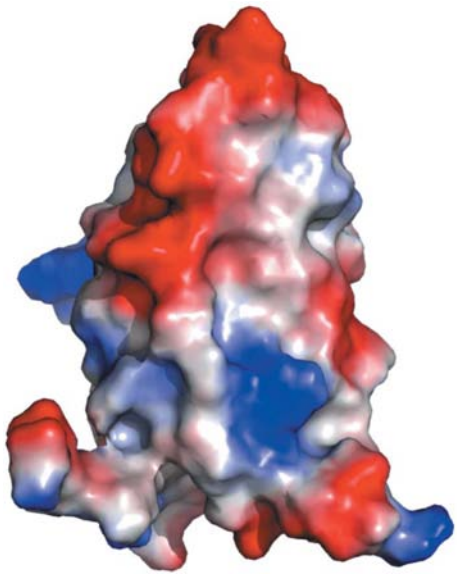

Figure 1
Surface-charge representation of the C1 domain of cMyBP-C. The C-terminal end is highly acidic (red). Reproduced from Govada *et al.* (2008) with permission of Elsevier.

2. The protein, crystallization and data collection

The source of protein was human cMyBP-C from the C0–C1 sequence (residues 1–258), which was amplified from full-length human cMyBP-C. The purified protein was dialysed into 10 m*M* imidazole, 50 m*M* sodium chloride, 0.5 m*M* dithiothreitol, 50 m*M* Tris pH 7.0 and concentrated to at least 20 mg ml^{-1} (further details of the preparation and purification are given in Govada *et al.*, 2008). The crystals were grown at 293 K by the microbatch-under-oil technique in 2 µl drops in Nunc 72-well plates that were incubated in an inverse position. The crystallization conditions (final concentrations of the ingredients in the drops) were 7.5 mg ml^{-1} protein, 8.5% PEG 3350 and 0.05 *M* HEPES pH 6.8. The crystals were taken out of the drops with a loop and flash-frozen without cryoprotectant[1].

For this study of the experimentally determined protonation state, data were collected on beamline 10.1 (Cianci *et al.*, 2005) at the Daresbury Synchrotron Radiation Source (SRS) at a wavelength of 1.283 Å. The crystal diffracted to a nominal resolution of 1.20 Å (*i.e.* into the corners of the detector) and 180° of data were collected. The diffraction images were indexed and integrated using *HKL*-2000 (Otwinowski & Minor, 1997), confirming the space group as $I4_1$ with unit-cell parameters a = 48.718, b = 48.718, c = 95.14 Å. The resulting data were then merged using *HKL*-2000, giving an overall $R_{\rm merge}$ of 4.6%. Ice rings were present in the diffraction images and the data affected by these were automatically removed using *HKL*-2000, resulting in a slightly lower completeness.

Table 1
Summary of data-collection and processing statistics for cMyBP-C C1 domain to 1.2 Å resolution.

Values in parentheses are for the outer shell resolution bin (1.24–1.20 Å).

Resolution (Å)	1.20†
Reflections	380249
Unique reflections	26342
Completeness (%)	79.9 (33.3)
Multiplicity	3.9 (1.2)
$R_{\rm merge}$ (%)	4.6 (64.0)
$\langle I\rangle/\langle\sigma(I)\rangle$	14.0 (1.3)

† The resolution where $\langle I\rangle/\langle\sigma(I)\rangle$ crossed 2.0 was 1.30 Å (see supplementary figures).

Table 2
Summary of diffraction and refinement data statistics to 1.3 Å resolution.

Values in parentheses are for the outer shell resolution bin and are interpolated values.

PDB code	3cx2
Space group	$I4_1$
Unit-cell parameters (Å)	
a	48.848
b	48.848
c	95.132
Resolution (Å)	1.3
$R_{\rm merge}$ (%)	4.6 (50.0)
Completeness (%)	88.4 (93.0)
$\langle I/\sigma(I)\rangle$	14.0 (2.0)
Multiplicity	3.9 (2.7)
No. of unique reflections	22822
R factor (%)	16.4
$R_{\rm free}$ (%)	20.9
Ramachandran core/additional (%)	96.7/3.3
Cruickshank's DPI (Å)	0.0675
Cruickshank's DPI, $R_{\rm free}$ (Å)	0.0628
No. of atoms in protein	866
Mean *B* factor of protein (Å^2)	18.2
Bond-distance r.m.s. deviation (Å)	0.025
Bond-angle r.m.s. deviation (°)	2.043
No. of water molecules	176
Mean *B* factor for water molecules (Å^2)	28.4
Mean *B* factor for all side-chain atoms (Å^2)	24.4
Mean *B* factor for all main-chain atoms (Å^2)	16.6
Mean *B* factor for all side chains and waters (Å^2)	24.4
Total No. of atoms	1042
Total No. of parameters for the final refinement	9403
Mean *B* factor for all atoms (Å^2)	21.3
Largest residual difference electron-density peak (*SHELX*)	0.28
Largest residual difference electron-density hole (*SHELX*)	−0.26

Full X-ray data-integration statistics are shown in Table 1. Supplementary Figs. 1–3[2] show plots of diffraction data $R_{\rm merge}$, $\langle I\rangle/\langle\sigma(I)\rangle$ and completeness. A full summary of diffraction data and protein model refinement statistics to 1.3 Å is given in Table 2.

3. Structure solution

A previously solved model of the C1 domain of cMyBP-C at a resolution of 1.55 Å (Govada *et al.*, 2008; PDB code 2v6h) was used with these extended resolution data. Firstly, rigid-body

[1] Cryoprotectants were tried and included 5, 10, 15 and 20% glycerol, 10, 12 and 14% PEG 3350, and 10 and 20% PEG 400. In all cases the diffraction resolution of the crystals was considerably lowered to about 2–2.3 Å. Therefore, after this experience, the crystal used for data collection was used without cryoprotectant.

[2] Supplementary material has been deposited in the IUCr electronic archive (Reference: FW5170). Services for accessing this material are described at the back of the journal.

Table 3
Structure-refinement statistics for the MyBP-C C1 domain.

N_{ref} is the number of unique reflections, N_w is the number of water molecules and the R factor is defined as $R = \sum ||F_{obs}| - |F_{calc}|| / \sum |F_{obs}|$.

	Resolution (Å)	N_{ref}, working set (N_{ref}, R_{free} set)	R (%)	R_{free} (%)	N_w	Comments
1	∞–2.50	2807 (287)	25.7	26.2	162	*REFMAC*: rigid-body refinement to 2.50 Å
2	∞–1.20	26339 (1364)	24.3	25.6	162	*REFMAC*: restrained refinement to 1.20 Å
3	∞–1.20	26339 (1364)	23.5	25.1	172	*REFMAC*: add multiple conformations for seven disordered side chains, update waters, mutation of Glu248 to Asp
4	∞–1.20	26339 (1364)	23.1	25.2	174	*REFMAC*: model three residues in the disordered region between residues 180 and 184
5	∞–1.20	26339 (1364)	19.4	24.4	177	*REFMAC*: anisotropic refinement, adjust weighting factors
6	10–1.20	26304 (1362)	17.8	22.3	177	*SHELX*-97: CGLS refinement, multiple-conformation occupancy for five side chains, add H atoms at calculated positions
7	10–1.30	22822 (1184)	16.4	20.9	176	*SHELX*-97: truncate to 1.30 Å

refinement was performed to 2.50 Å resolution using *REFMAC*5 (Collaborative Computational Project 4, Number 4, 1994), yielding an initial R factor of 25.7% and an R_{free} of 26.2%. The starting model was initially random-atom shifted using the program *MOLEMAN*2 (Kleywegt, 2003) in order to remove any possible model bias. A restrained refinement was then carried out using *REFMAC*5 to 1.2 Å; the resulting $F_o - F_c$ difference electron-density map was used to locate a series of water molecules and to assign several multiple conformations as well as to model a series of disordered residues located at amino-acid sequence positions 181–184. Another stage of refinement was then carried out; from inspection of the resulting $F_o - F_c$ difference map contoured at 3σ, a mutation from a glutamic acid to an aspartic acid side chain (see Fig. 2) was located at residue 248.

After final refinement using *REFMAC*5, the input file was converted to *SHELX*-97 format and refinement was carried out using *SHELX*-97 (Sheldrick, 2008) in order to refine the multiple conformation occupancies. Refinement proceeded using the standard CGLS refinement parameters and H atoms were added in riding positions, resulting in a final R factor of 17.8% and an R_{free} of 20.0%. As a test the data were also truncated to 1.30 Å and another cycle of refinement was carried out using *SHELX*-97, resulting in an R factor of 16.4% and an R_{free} of 18.6%. The full refinement procedure is shown in Table 3. Fig. 3 shows the structure of the C1 domain of cMyBP-C at 1.30 Å resolution in simplified ribbon view with the determined multiple-occupancy side chains indicated in cyan.

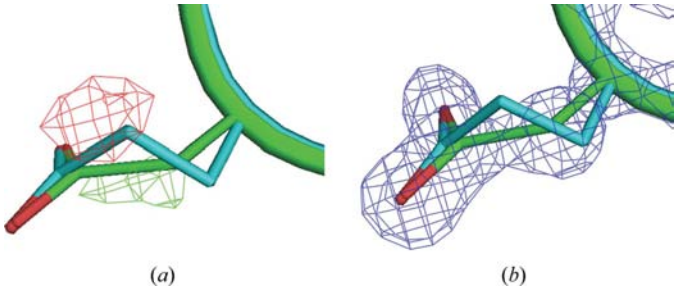

Figure 2
Mutation of residue 248 in cMyBP-C from a glutamic acid to an aspartic acid residue. (*a*) $F_o - F_c$ difference map contoured at 4σ (negative; red) and 3.2σ (positive; green) when refined as a Glu side chain, (*b*) $2F_o - F_c$ map contoured to 2 r.m.s. for the Asp assignment. This figure was generated using *PyMOL* (DeLano, 2002).

4. Results and discussion

The 1.3 Å resolution structure is very similar to the 1.55 Å resolution structure (Govada *et al.*, 2008), with r.m.s. deviations of 0.069 for C^{α} atoms and 0.180 Å for all atoms excluding multiple occupancy side chains. There is a change in conformation of the Gln205 and Lys190 side chains. The five multiple-occupancy side chains differ slightly (Thr167, Ser236, Val241, Ser242 and Phe247 *versus* Lys195, Ser236, Val241, Ser242 and Phe247); leaving these in the r.m.s. calculation increases the r.m.s. to 0.427 Å.

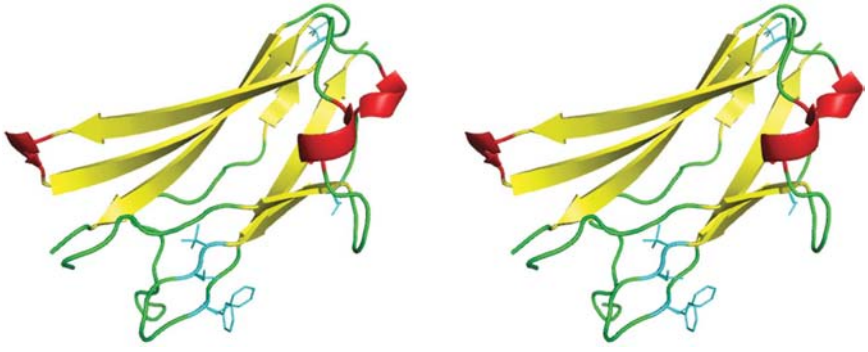

Figure 3
Stereoview (cross-eye viewing) of the structure of the C1 domain of cMyBP-C at 1.30 Å resoution with multiple-occupancy side chains indicated in cyan. This figure was generated using *PyMOL*.

4.1. Determination of protonation states

After the final stage of refinement, the C═O and C—OH bonds of the aspartic and glutamic acid residues were unrestrained and another cycle of CGLS refinement was carried out in order to determine 'real' C═O and C—OH bond lengths. None of the Asp or Glu residues were of multiple occupancy. The *SHELX*-97 refinement parameter L.S.1 was then used in conjunction with the ACTA keyword in order to perform a least-squares full-matrix inversion, which calculates the standard uncertainty for each bond length. It was not

Table 4
Table of C—O and C═O bond lengths with σ levels for the cMyBP-C C1 domain at 1.2 Å resolution.

The standard uncertainties are plotted in Figs. 4 and 5 and are typically 0.09 Å.

	C═O (Å)	*B* factor (Å^2)	C—OH (Å)	*B* factor (Å^2)	Δ(C—O) (Å)	Δ(C—O)/σ[Δ(C—O)]
Asp						
152	1.209	27.31	1.321	30.18	0.11	0.72
163	1.190	17.91	1.280	22.03	0.09	0.91
198	1.221	21.43	1.272	29.31	0.05	0.45
211	1.171	19.35	1.292	21.63	0.12	1.22
214	1.125	25.95	1.356	28.04	0.23	1.81
228	1.227	30.80	1.262	33.09	0.03	0.25
245	1.104	42.77	1.438	50.67	0.33	1.43
248	1.196	17.25	1.215	17.20	0.02	0.16
Glu						
165	1.230	29.74	1.237	38.09	0.01	0.05
223	1.250	18.35	1.263	18.65	0.01	0.13
240	1.235	19.82	1.246	20.84	0.01	0.11
258	1.198	45.21	1.367	49.91	0.17	0.63

Table 5
Table of C—O and C═O bond lengths with σ levels for the cMyBP-C C1 domain at 1.3 Å resolution.

The standard uncertainties are plotted in Figs. 6 and 7 and are typically 0.10 Å.

	C═O (Å)	*B* factor (Å^2)	C—OH (Å)	*B* factor (Å^2)	Δ(C—O) (Å)	Δ(C—O)/σ[Δ(C—O)]
Asp						
152	1.206	26.24	1.319	29.20	0.11	0.73
163	1.181	17.48	1.289	21.57	0.11	1.09
198	1.241	20.46	1.259	29.11	0.02	0.16
211	1.151	19.03	1.295	21.09	0.14	1.45
214	1.135	25.52	1.352	27.38	0.22	1.61
228	1.228	30.64	1.265	32.66	0.04	0.26
245	1.098	42.83	1.451	49.90	0.35	1.51
248	1.198	16.72	1.218	17.28	0.02	0.16
Glu						
165	1.227	28.79	1.229	37.67	0.00	0.01
223	1.247	17.85	1.260	18.21	0.01	0.12
240	1.237	19.37	1.244	20.02	0.01	0.07
258	1.203	44.60	1.367	49.33	0.16	0.58

necessary to include the $R_{\rm free}$ reflections in the full-matrix inversion. These values were then used to calculate the standard uncertainty in the bond-distance difference between the C═O and C—OH bond lengths for each of the Asp and Glu residues (Ahmed *et al.*, 2007). In order to positively infer the protonation state where Δ(C—O) is 0.1 Å, a σ[Δ(C—O)] of less than 0.033 is needed, which represents a 3σ difference. These quantities were calculated using both the original 1.2 Å resolution data and also the 1.3 Å resolution truncated data for comparison in order to judge the effective resolution from this point of view. Dictionary (Engh & Huber, 1991) bond-length values are taken to be 1.208 Å for the C═O bond, 1.304 Å for the C—OH bond and 1.249 Å for the deprotonated C—O$^-$ bond. However, as Ahmed *et al.* (2007) discuss in detail, the value for the C—OH bond distance can be affected by the micro-environment of the hydrogen.

Tables 4 and 5 show the carbonyl bond lengths, *B* factors and σ level for each residue studied at 1.2 Å and 1.3 Å resolution, respectively.

Fig. 4 shows the results for aspartic acid residues at 1.2 Å resolution with standard deviations plotted as error bars. Most of the residues appear to be deprotonated, except for residue 214, which shows some evidence of protonation with a standard uncertainty of 1.81σ; however, these bond lengths lie outside the dictionary values. Residue 245 also shows some evidence of protonation; however, it has very large thermal motion parameters, ultimately resulting in large standard deviations on the bond lengths and also has values outside dictionary values.

Fig. 5 shows the results for glutamic acid residues at 1.2 Å resolution with standard deviations plotted as error bars. All of the residues except residue 258 appear to be deprotonated; however, residue 258 suffers from the same problem as Asp245 in that it has large *B* factors and hence large standard deviations, meaning the protonation state of this residue cannot be accurately inferred.

Fig. 6 shows the results for the same aspartic acid residues but at 1.3 Å resolution with standard deviations plotted as error bars. Most of the residues once again seem to be deprotonated as in the 1.2 Å results. Asp214 still shows signs of being protonated, but still has a relatively low standard precision of 1.61σ. It is interesting that the precision has

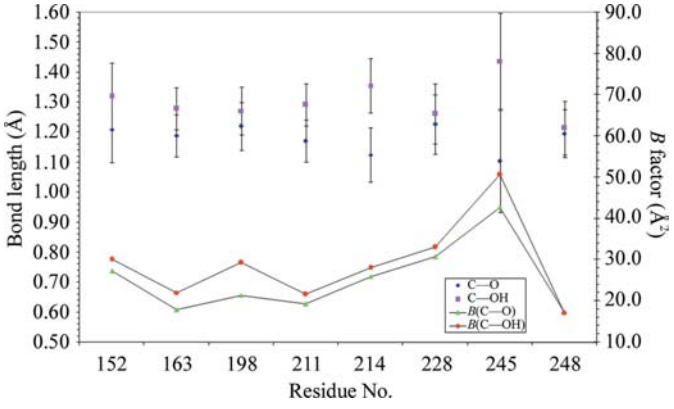

Figure 4
C—O bond lengths for Asp residues in the cMyBP-C C1 domain at 1.2 Å resolution.

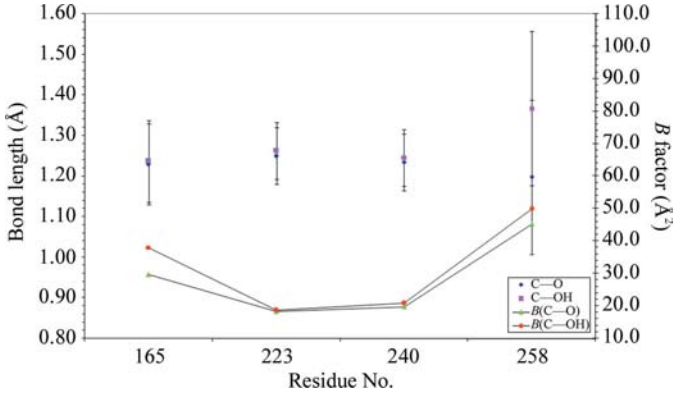

Figure 5
C—O bond lengths for Glu residues in the cMyBP-C C1 domain at 1.2 Å resolution.

become worse even though the data-quality indicators show that data beyond 1.26 Å should not be used.

Fig. 7 shows the results for glutamic acid residues at 1.3 Å resolution with standard deviations plotted as error bars; once again, most residues seem to be deprotonated. Glu258 again shows significant signs of protonation, but has large standard deviations and *B* factors. This is most likely to be a consequence of the fact that residue 258 is the terminal residue of the protein and hence is not as static as other residues in the protein.

Fig. 8 shows Asp214 with electron-density maps contoured at 2 and 3 r.m.s.; this residue has the highest probability of being protonated, with a σ level of 1.61. It is interesting to note that the electron density is concentrated on the shorter of the two C—O bonds, again indicating that this is a C═O bond and not a C—OH bond.

Despite the availability of high-resolution data and reasonable *B* factors (<30 Å^2) for many of the atoms, the standard uncertainties associated with each of the Asp and Glu C═O and C—OH bond lengths are relatively high, with an average of 0.09 Å. This is partly a consequence of the completeness of the data being low. The ice rings present in the diffraction images, as shown in Fig. 9, have resulted in sections of data being removed and thus have impacted on the completeness. One strategy for filling in the missing data is to use a targeted measurement of these reflections *via* a four-circle diffractometer (Emmerich *et al.*, 1994). Clearly, a better approach in terms of overall efficiency is to find a suitable cryoprotectant, possibly combined with annealing, and the list of available cryoprotectants fortunately continues to expand.

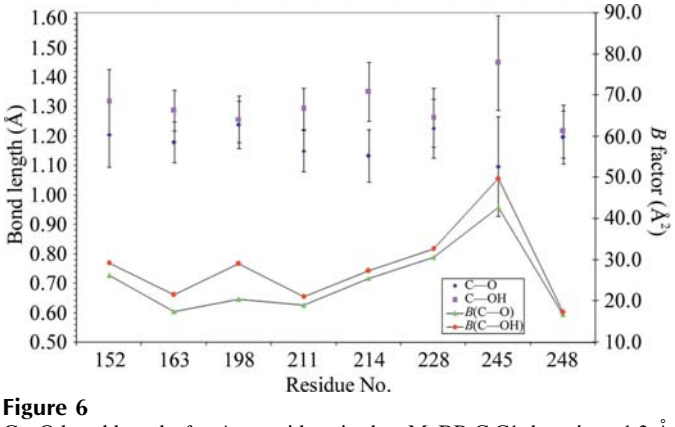

Figure 6
C—O bond lengths for Asp residues in the cMyBP-C C1 domain at 1.3 Å resolution.

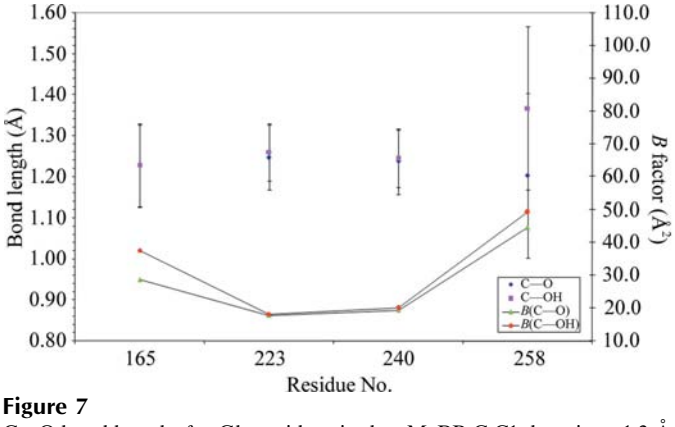

Figure 7
C—O bond lengths for Glu residues in the cMyBP-C C1 domain at 1.3 Å resolution.

4.2. Cruickshank DPI considerations

Cruickshank introduced the diffraction-component precision index (DPI; Cruickshank, 1999*a*,*b*) as an indicator of the average uncertainty of the position (or coordinate) of the atoms within a protein structure as an alternative to full-matrix inversion. It is based on a rough approximation of the least-squares method used by the full-matrix inversion and, as an approximation, is significantly much less computationally intensive. The DPI uses the *R* factor, the diffraction data completeness, the diffraction resolution, the overall (*i.e.* average) *B* factor and the data-to-parameter ratio in order to

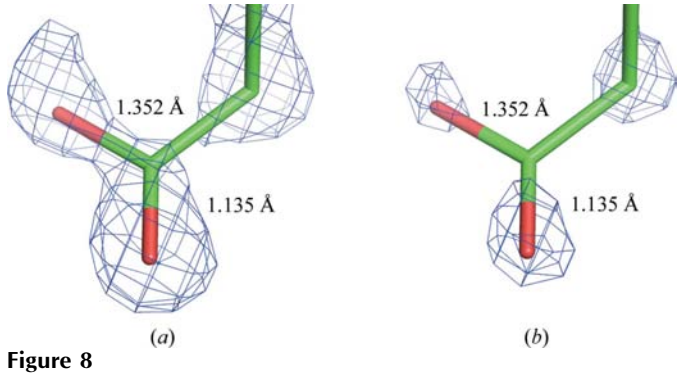

Figure 8
Asp214 with $2F_o - F_c$ electron-density maps contoured at (*a*) 2 r.m.s. and (*b*) 3 r.m.s. This figure was generated using *PyMOL*.

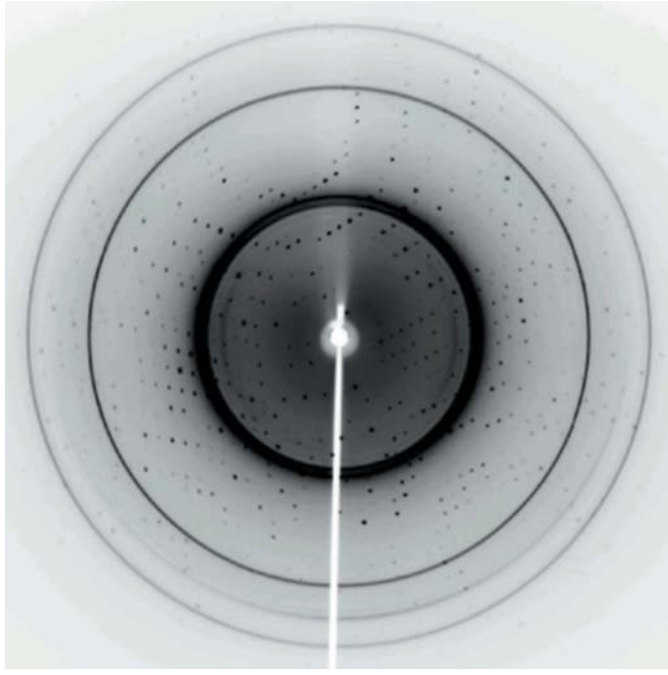

Figure 9
Ice rings present in one of the cMyBP-C C1 domain diffraction images.

estimate an overall position or coordinate uncertainty. It is critical that the differentiation between the average coordinate error (σ_x) and the average bond-length error (σ_l) is made. The average coordinate error is based on single coordinates and the average bond-length error is based on the two average coordinate errors in the direction of the bond and thus for isotropic atoms can be assumed to be $2^{1/2}$ times the coordinate error,

$$\sigma_l = (\sigma_a^2 + \sigma_b^2)^{1/2} = (2\sigma_x^2)^{1/2} = 2^{1/2}\sigma_x. \quad (1)$$

This differentiation is important as *REFMAC*5 calculates DPI coordinate errors during refinement, whereas a full-matrix inversion using *SHELX*-97 calculates bond-length errors. In order for comparisons to be made, the errors must be converted to a consistent type.

Cruickshank's form of the DPI is

$$\sigma(x, B_{\rm avg}) = 1.0(N_i/p)^{1/2} C^{-1/3} R d_{\rm min},$$
$$\sigma(x, B_{\rm avg}) = (N_i/n_{\rm obs})^{1/2} C^{-1/3} R_{\rm free} d_{\rm min}, \quad (2)$$

where $\sigma(x, B_{\rm avg})$ is the DPI for an atom with an average B factor, N_i is the number of fully occupied atoms of type i, $p = n_{\rm obs} - n_{\rm params}$, C is the completeness of the data, R is the R factor and $d_{\rm min}$ is the diffraction resolution. However, in this form it is difficult to assess how the DPI is related to resolution or completeness owing to the interdependencies of $n_{\rm obs}$, C and $d_{\rm min}$. Blow (2002) subsequently rearranged these formulae in order to relate them to parameters that are readily available to an experimentalist,

$$\sigma(x, B_{\rm avg}) = 0.18(1 + s)^{1/2} V_{\rm M}^{-1/2} C^{-5/6} R_{\rm free} d_{\rm min}^{5/2}. \quad (3)$$

The DPI now depends on the solvent content s (= $N_{\rm solv}/N_{\rm atoms}$), the Matthews volume $V_{\rm M}$ (Matthews, 1968), the completeness, $R_{\rm free}$ and resolution. Direct comparisons against resolution and completeness can now be made as all of the variables are independent of each other.

Using Blow's rearrangement of the Cruickshank DPI formula, it is possible to extrapolate the completeness and resolution required to bring the coordinate error down below 0.03 Å. The coordinate (σ_x) DPI values calculated using *REFMAC*5 are 0.063 Å (based on the R factor) and 0.065 Å (based on $R_{\rm free}$) at 1.2 Å and 0.068 Å (R factor) and 0.068 Å ($R_{\rm free}$) at 1.3 Å. These values are reasonably consistent with the calculated bond-length errors from the full-matrix inversion, which average 0.10 Å (the DPI can vary by about 15% from the full-matrix inversion; Blow, 2002), equivalent to σ_x = 0.071 Å.

Fig. 10 shows σ_x *versus* completeness calculated for 1.2 and 1.3 Å resolution data. It can be seen that even at a completeness of 100%, it is still not possible to reach a σ_x of <0.03 Å. Fig. 11(*a*), however, shows σ_x *versus* resolution and it can be seen that a resolution of roughly 0.94 Å would be required for the current data completeness. However, as Fig. 11(*b*) shows, for 100% data completeness a resolution of 1.02 Å would be needed.

The values from *REFMAC*5 show larger uncertainties than those calculated from equation (3) at 1.2 Å (shown in Figs. 10 and 11) thus confirming that our current data do not in fact run to 1.2 Å resolution.

Extrapolated values for the resolution using Fig. 10 indicate that the expected DPI would be ∼0.055 Å at 1.2 Å compared with the value of ∼0.064 Å calculated using *REFMAC*5, showing that the data do not really run to 1.2 Å. However, it is interesting that the effect of the data-to-parameter ratio (2.92 for the 1.2 Å data set and 2.53 for the 1.3 Å data set) leads to

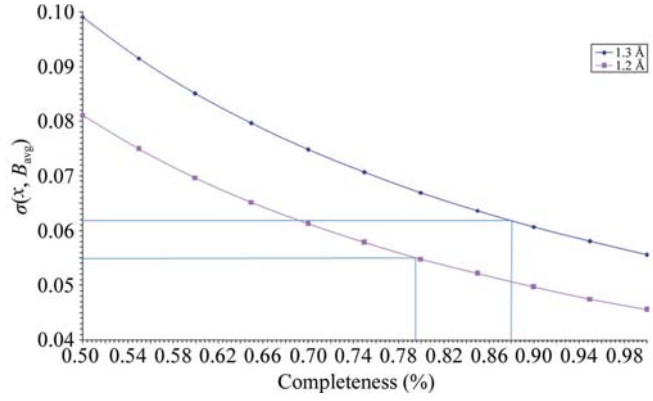

Figure 10
Extrapolated $\sigma(x, B_{\rm avg})$ *versus* completeness from equation (3) for the cMyBP-C C1 domain at 1.2 and 1.3 Å resolution.

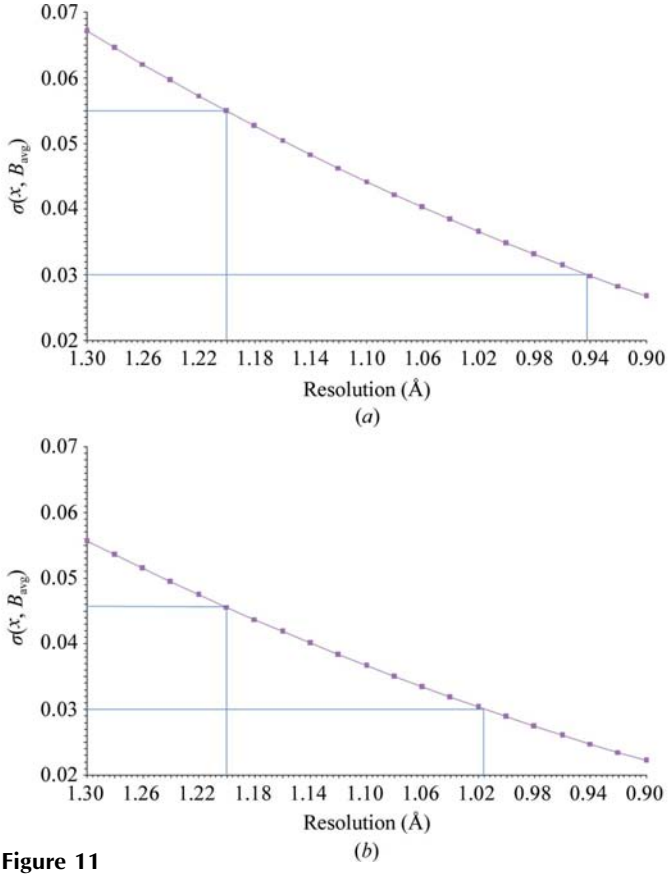

Figure 11
Extrapolated $\sigma(x, B_{\rm avg})$ *versus* resolution from equation (3) for the cMyBP-C C1 domain. (*a*) For the experimental completeness of this study (79.9% at 1.2 Å resolution); (*b*) if 100% data completeness could be obtained.

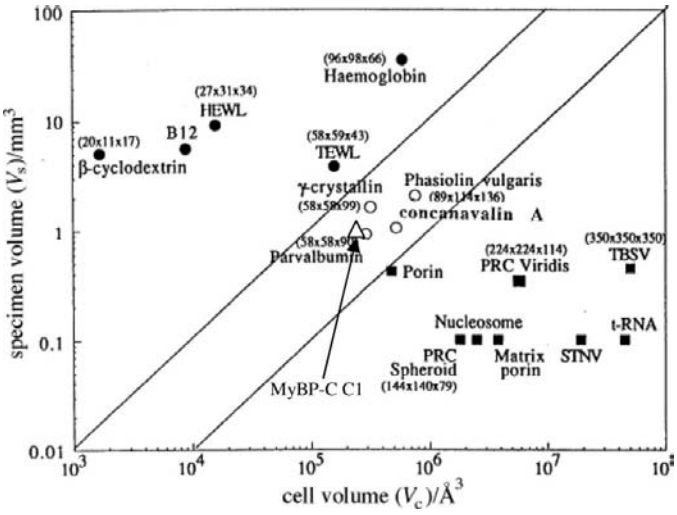

Figure 12
Scatter plot of examples of protein crystals investigated at ILL on the monochromatic neutron instruments D19 (filled circles) and DB21 (filled squares). The line through the origin shows the empirical limit and thus the open circles show examples where the LADI I instrument brought into range the more challenging projects of the time. Reproduced from Habash *et al.* (1997) by permission of The Royal Society of Chemistry. Added here and marked with an arrow is the cMyBP-C C1 domain protein crystal case, unit-cell volume 230 000 Å^3 and an indicative specimen crystal volume of 1 mm^3.

the estimated standard deviations from the full-matrix inversion being beneficially improved (1.6σ at 1.3 Å *versus* 1.81σ at 1.2 Å for Asp214; see Tables 4 and 5) for some residues, but not for others (1.51σ at 1.3 Å *versus* 1.43σ at 1.2 Å for Asp245).

4.3. Extrapolation for a neutron protein crystallography experiment

Fig. 12 is a plot of feasible neutron protein crystallography projects over many years up to 1997 (Habash *et al.*, 1997). Since then, two major developments have occurred. Firstly, the deuteration laboratory (DLAB; Haertlein, 2007) for producing fully deuterated protein had led to enhanced neutron diffraction and consequently radically improved resolution for a given specimen volume, *e.g.* 0.15 mm^3 for fully deuterated aldose reductase (Hazemann *et al.*, 2005). Such a radically smaller crystal volume than the previous typical >1 mm^3 would likewise benefit a project on the cMyBP-C C1 domain. Secondly, LADI I has been superseded by LADI III, with a threefold enhancement in detector sensitivity, which translates into a threefold further reduction in typical specimen volume (to about 0.10 mm^3).

5. Conclusions

The protonation-state determinations of the cMyBP-C C1 domain were not even above 2σ (the best was 1.81σ) and therefore an extrapolation is given, based on 100% data completeness and the average DPI, that a 3σ determination could be possible if 100% complete X-ray data to 1.02 Å could be measured. This imporved resolution might be possible *via* improved crystallization or multiple sample evaluation, *e.g.* using robotics or a yet more intense/collimated X-ray beam or combinations thereof. An alternative would be neutron protein crystallography at ∼2 Å diffraction resolution, where it is estimated that for the unit-cell volume of the cMyBP-C C1 domain a minimum crystal volume of 0.10 mm^3 would be needed with fully deuterated protein. These efforts would optimally be combined in a joint X-ray and neutron model refinement.

We thank The University of Manchester and the Institut Laue Langevin for PhD studentship support (SF). We thank Dr M. J. Ellis for station support of our use of SRS beamline 10 which was funded by BBSRC grants B15474 and REI20571 to S. S. Hasnain PI and J. R. Helliwell Co-PI with North West Structural Genomics Consortium partners (http://www.nwsgc.ac.uk) and subsequently from BBSRC grant BBE0019711 to S. S. Hasnain and R. Strange as well as an NWDA project award (N0002170) to S. S. Hasnain. NEC acknowledges the European Commission OptiCryst Project LSHG-CT-2006-037793 and the UK Engineering and Physical Sciences Research Council EP/D501113/1 for financial support.

References

Ahmed, H. U., Blakeley, M. P., Cianci, M., Cruickshank, D. W. J., Hubbard, J. A. & Helliwell, J. R. (2007). *Acta Cryst.* D**63**, 906–922.
Blow, D. M. (2002). *Acta Cryst.* D**58**, 792–797.
Cianci, M. *et al.* (2005). *J. Synchrotron Rad.* **12**, 455–466.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cruickshank, D. W. J. (1999*a*). *Acta Cryst.* D**55**, 583–601.
Cruickshank, D. W. J. (1999*b*). *Acta Cryst.* D**55**, 1108.
DeLano, W. L. (2002). *PyMOL.* http://www.pymol.org.
Emmerich, C., Helliwell, J. R., Redshaw, M., Naismith, J. H., Harrop, S. J., Raftery, J., Kalb (Gilboa), A. J., Yariv, J., Dauter, Z. & Wilson, K. S. (1994). *Acta Cryst.* D**50**, 749–756.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Govada, L., Carpenter, L., Helliwell, J. R., Fonseca, P., Flashman, E., Chayen, N., Redwood, C. & Squire, J. (2008). *J. Mol. Biol.* **378**, 387–397.
Habash, J., Raftery, J., Weisgerber, S., Cassetta, A., Lehmann, M. S., Hoghoj, P., Wilkinson, C., Campbell, J. W. & Helliwell, J. R. (1997). *J. Chem. Soc. Faraday Trans.* **93**, 4305.
Haertlein, M. (2007). *The Deuteration Laboratory.* http://www.ill.fr/YellowBook/deuteration/.
Hazemann, I., Dauvergne, M. T., Blakeley, M. P., Meilleur, F., Haertlein, M., Van Dorsselaer, A., Mitschler, A., Myles, D. A. A. & Podjarny, A. (2005). *Acta Cryst.* D**61**, 1413–1417.
Kleywegt, G. (2003). *Uppsala Software Factory.* http://xray.bmc.uu.se/usf/.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Mohamed, A. S., Dignam, J. D. & Schlender, K. K. (1998). *Arch. Biochem. Biophys.* **358**, 313–319.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.