research papers

Elastic and inelastic diffraction changes upon variation of the relative humidity environment of PurE crystals

R. Oliete,[a,b] **J. Pous,**[a,c] **S. Rodríguez-Puente,**[a,b] **C. Abad-Zapatero**[a,d]* **and A. Guasch**[a,b,e]*

[a]Plataforma Automatitzada de Cristal.lografía, Barcelona, Spain, [b]Parc Científic Barcelona, Barcelona, Spain, [c]Institute for Research in Biomedicine, Barcelona, Spain, [d]Center for Pharmaceutical Biotechnology, University of Illinois at Chicago (UIC), Chicago, Illinois, USA, and [e]Institut de Biologia Molecular de Barcelona, Barcelona, Spain

Correspondence e-mail: caz@uic.edu, agmcri@ibmb.csic.es

The different changes observed in the diffraction patterns of three different crystal forms (hexagonal, trigonal and monoclinic) of PurE (EC 4.1.1.21), an enzyme from the purine-biosynthesis pathway of *Bacillus anthracis*, upon a wide range of changes in the relative humidity environment of the crystals are documented. In addition, the changes in the unit-cell parameters, volume and bulk solvent in the three different crystal forms were systematically followed. In an attempt to explain the elastic ($P6_522$) and inelastic ($P3_121$) changes in the diffraction pattern, refined structures of the three different crystal forms determined at 100 K are presented, with particular emphasis on the tertiary and quaternary structural differences, crystal packing, intermolecular and intra-molecular interactions and solvent structure. The refined structures show that the precipitant salts, solvent structure (both ordered and bulk) and conformation of the C-termini all play a role in creating a unique cement at both the intramolecular and intermolecular contacts of the different crystal forms. It is suggested that it is the combination of polyethylene glycol and the structure of the ordered water molecules (first and second layers) as well as the structure of the bulk solvent that are the critical factors in the plasticity of the hexagonal crystal packing as opposed to the inelastic responses of the lower symmetry forms.

1. Introduction

Macromolecular crystallography has shown that most proteins incorporate water molecules as an intrinsic component of their three-dimensional structures. The importance of the presence of significant amounts of solvent in protein crystals was recognized from the early days of the field, and this insight played a critical role in its development (Bernal & Crowfoot, 1934). It was also systematically studied by Perutz in his early attempts to solve the phase problem by studying the shrinkage of haemoglobin crystals in different salt solutions (Perutz, 1985).

Protein crystals contain more than 30% water (Matthews, 1968; Kantardjieff & Rupp, 2003), with an interconnected network of large channels filled with a solution in equilibrium with the mother solution from which the crystal grew. Some of these solvent molecules are bound to the protein and contribute to stabilizing the protein fold and the crystal packing, but others are actually part of the bulk solution. The latter can move freely along the interconnected network of channels and exchange with the medium surrounding the crystal (Blevins & Tulinsky, 1985).

Table 1
Crystal forms observed for PurE.

Structural and crystallographic data entries related to PurE from different species. PDB entries 4ay3, 4ay4 and 4b4k for *B. anthracis* correspond to this work.

PDB entry	Space group, resolution (Å)	Unit-cell parameters (Å, °)	Crystallization conditions		T (K)	Subunits	Organism
			Precipitants	Buffer			
1xmp	$C2$, 1.8	$a = 168.26$, $b = 76.48$, $c = 102.68$, $\beta = 168.20$	15% PEG 4000, 0.8 M sodium formate	0.1 M Tris pH 7.5	291	8	*Bacillus anthracis*
3oow	$P2_12_12_1$, 1.75	$a = 88.56$, $b = 96.30$, $c = 128.62$	0.2 M ammonium dihydrogen phosphate, 6% MPD	0.1 M Tris pH 8.5	295	8	*Francisella tularensis*
1qcz	$I422$, 1.5	$a = b = 113.04$, $c = 49.41$	24% PEG 400, 0.2 M MgCl$_2$	0.1 M Tris pH 8.0	291	1	*Escherichia coli*
1u11	$I422$, 1.55	$a = b = 99.25$, $c = 164.81$	22% PEG 4000, 0.19 M ammonium acetate	0.09 M citrate pH 5.5	293	2	*Acetobacter aceti*
1o4v	$I422$, 1.77	$a = b = 103.25$, $c = 65.44$	2 M ammonium sulfate, 0.2 M potassium sodium tartrate tetrahydrate	0.1 M trisodium citrate dihydrate pH 5.6	293	1	*Thermotoga maritima*
3lp6	$P4_32_12$, 1.7	$a = b = 145.70$, $c = 58.38$	1.7 M sodium formate	0.1 M sodium acetate pH 5.4	298	4	*Mycobacterium tuberculosis*
3trh	$P2_1$, 2.2	$a = 87.27$, $b = 96.29$, $c = 152.72$, $\beta = 87.27$	14% PEG 3350	0.2 M ammonium formate pH 7.5	295	16	*Coxiella burnetii*
3rg8	$P2_1$, 1.74	$a = 83.47$, $b = 87.92$, $c = 86.70$, $\beta = 83.47$	15% PEG 1000, 0.1 M MgCl$_2$	0.1 M imidazole pH 8	298	8	*Treponema denticola*
4ay4	$P6_522$, 2.0	$a = b = 87.00$, $c = 270.00$	0.3 M sodium acetate, 15% PEG 4000	0.1 M Tris pH 8.5	293	4	*Bacillus anthracis*
4ay3	$P3_121$, 1.76	$a = b = 86.88$, $c = 131.37$	0.75 M sodium acetate	0.1 M cacodylate pH 6.5	293	4	*Bacillus anthracis*
4b4k	$C2$, 2.5	$a = 87.56$, $b = 151.90$, $c = 134.85$, $\beta = 98.33$	0.3–0.4 M sodium acetate	0.1 M cacodylate pH 6.5	293	12	*Bacillus anthracis*

The crystallographic water molecules are typically organized into discrete networks. A distinctive feature of these clusters is the frequent occurrence of three layers (Nakasako, 2004). Water molecules in the 'first layer' are defined as those that interact directly with atoms of the protein surface through hydrogen bonds (2.75 Å). Waters in the 'second layer' do not interact directly with the protein and are approximately 3.65 Å from the protein surface (Chen *et al.*, 2008). Moving away from the surface, the experimental data on the influence of water on the protein structures becomes less definite. The number of water molecules observed in a protein crystal structure varies primarily with the resolution of the diffraction data, and a regression equation has been proposed by Carugo & Bordo (1999).

Little is known in atomic detail of what happens when a protein crystal is exposed to drier conditions or about the reverse rehydration processes (Huxley & Kendrew, 1953; Berthou *et al.*, 1972; Madhusudan *et al.*, 1993). Presumably, the water molecules in the solvent channels migrate and eventually escape through vapour diffusion at the crystal surface. The loss of these molecules probably induces lattice shrinkage that accompanies the rearrangements and conformational changes of the protein molecules in the crystal lattices in which they are embedded. Similarly, the atomic details of the rehydration process in the crystalline environment are not known.

In recent years, the availability of devices that permit fine control of the relative humidity (RH) of the crystals [free-mounting systems (FMS) or humidity-control devices (HC1b); Kiefersauer *et al.*, 2000; Sanchez-Weatherby *et al.*, 2009] has made it possible to improve the resolution (in some cases dramatically) of protein crystals whose diffraction properties are suboptimal. The current status of these developments in macromolecular crystallography have recently been reviewed (Russi *et al.*, 2011). The importance of controlling the RH when soaking ligands into pre-grown apoenzyme crystals has been discussed in relation to fragment-based drug-discovery approaches (Bottcher *et al.*, 2011).

In previous work (Abad-Zapatero *et al.*, 2011), we reported that humidity control of the environment of a crystal can be used to compensate for the detrimental effect of alien (non-mother-liquor) solutions in the crystals. We documented these effects using three different crystal forms (hexagonal, trigonal and monoclinic) of the enzyme PurE, a critical enzyme of the purine-biosynthetic pathway in *Bacillus anthracis*. PurE has been found to play an essential role in the growth of bacteria in human blood (Samant *et al.*, 2008); therefore, the structures of enzymes from the purine-biosynthesis pathway have been extensively studied (Zhang *et al.*, 2008). The structure of this enzyme from *Escherichia coli* was first solved at 1.5 Å resolution (PDB entry 1qcz; Mathews *et al.*, 1999), and the structure of a PurE–mononucleotide complex (PDB entry 1d7a) and a high-resolution structure of PurE from *B. anthracis* at 1.8 Å resolution have also been reported (PDB entry 1xmp; Boyle *et al.*, 2005). Three-dimensional structures of PurE from several important pathogens have recently been reported (PDB entries 3rg8, 3rgg, 3oow, 3lp6, 3kuu and 3k5h; Tranchimand *et al.*, 2011; Midwest Center for Structural Genomics, unpublished work; Integrated Center for Structure and Function Innovation/TB Structural Genomics Consortium, unpublished work; Center for Structural Genomics of Infectious Diseases, unpublished work; Thoden *et al.*, 2010). Structural and crystallographic details for the most relevant PurE structures are shown in Table 1 and the amino-acid sequences, including purification tags and overall topological elements of the secondary structure, are presented in Fig. 1(*a*).

Fig. 1(b) shows a superposition of the available PurE structures most relevant to this work, highlighting the differences among them.

Here, we document the different responses of these three different crystal forms of PurE to changes in the RH of the crystals. These crystal forms were characterized in a previous publication (Abad-Zapatero *et al.*, 2011) and have not been reported before. The effect that the changes in the RH had on various crystallographic parameters was systematically studied. Although the packing arrangement of the PurE octamers is rather similar in the three forms, the hexagonal form responded in a flexible way to the RH changes, while the others did not. In order to provide some rationale for these observations, we present the refined structures of the three different crystal forms at resolutions of 1.75, 2.0 and 2.5 Å for

the trigonal ($P3_121$), hexagonal ($P6_522$) and monoclinic ($C2$) forms, respectively (Abad-Zapatero *et al.*, 2011), obtained from crystals cooled to 100 K.

We briefly compare the quaternary structure of the PurE protein and discuss in detail the related packing in the three different crystal forms presented here and in others published elsewhere, with a special focus on intermolecular and intramolecular contacts. In addition, we document in detail the role that the solvent and precipitant salts play in cementing the 422 oligomer and in supporting the crystal contacts so that they allow the elastic and inelastic changes that we observe in the quality and extent of the diffraction pattern upon alteration of the RH of the crystal environment.

Our results with this enzyme system suggest that the factors determining the plasticity of a crystal lattice to changes in the
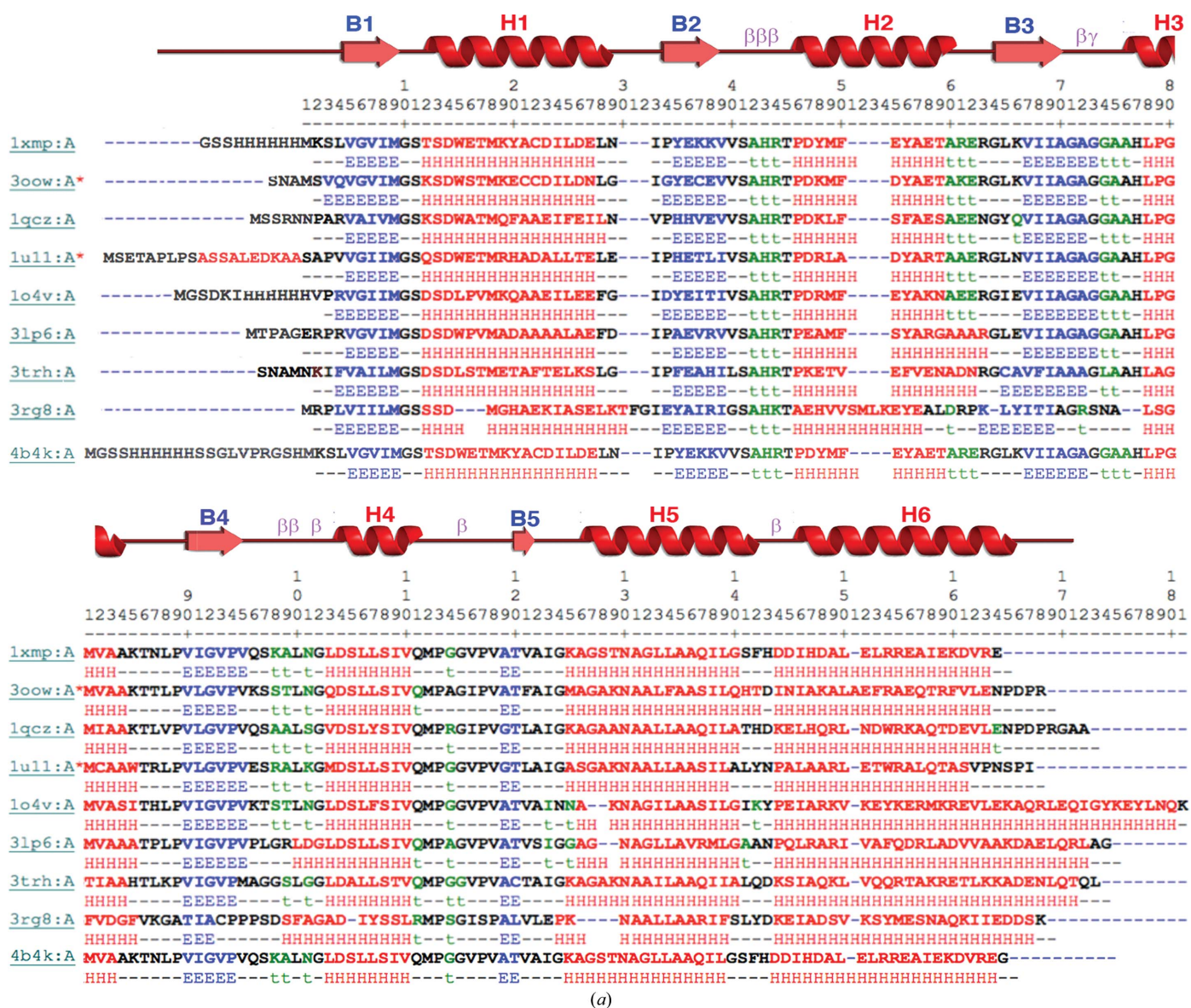

Figure 1
Structures and sequences of PurE from different organisms. (*a*) Sequence and secondary-structure alignment of PurE from *B. anthracis* (PDB entries 4b4k, 4ay4 and 4ay3; this work) with eight other homologous proteins. The software tool *Sequence Annotated by Structure* (*SAS*) from the EBI website was used. The organisms corresponding to the PDB codes are given in Table 1.

RH are dependent on the solvent environment (both ordered and bulk) as well as the protein molecules involved, including the structures of the residues involved in the packing, particularly the C-termini. Whether or not these initial results can be extended or generalized to other proteins has still not been proven. These results add to the body of knowledge that has recently accumulated regarding the effects that changing the RH of the crystals can have on the quality of the diffraction

pattern of macromolecular crystals and might help in developing predictive models in the future.

2. Materials and methods

2.1. Sample preparation

2.1.1. Protein expression, purification and crystallization. Details of the protein expression and purification have been
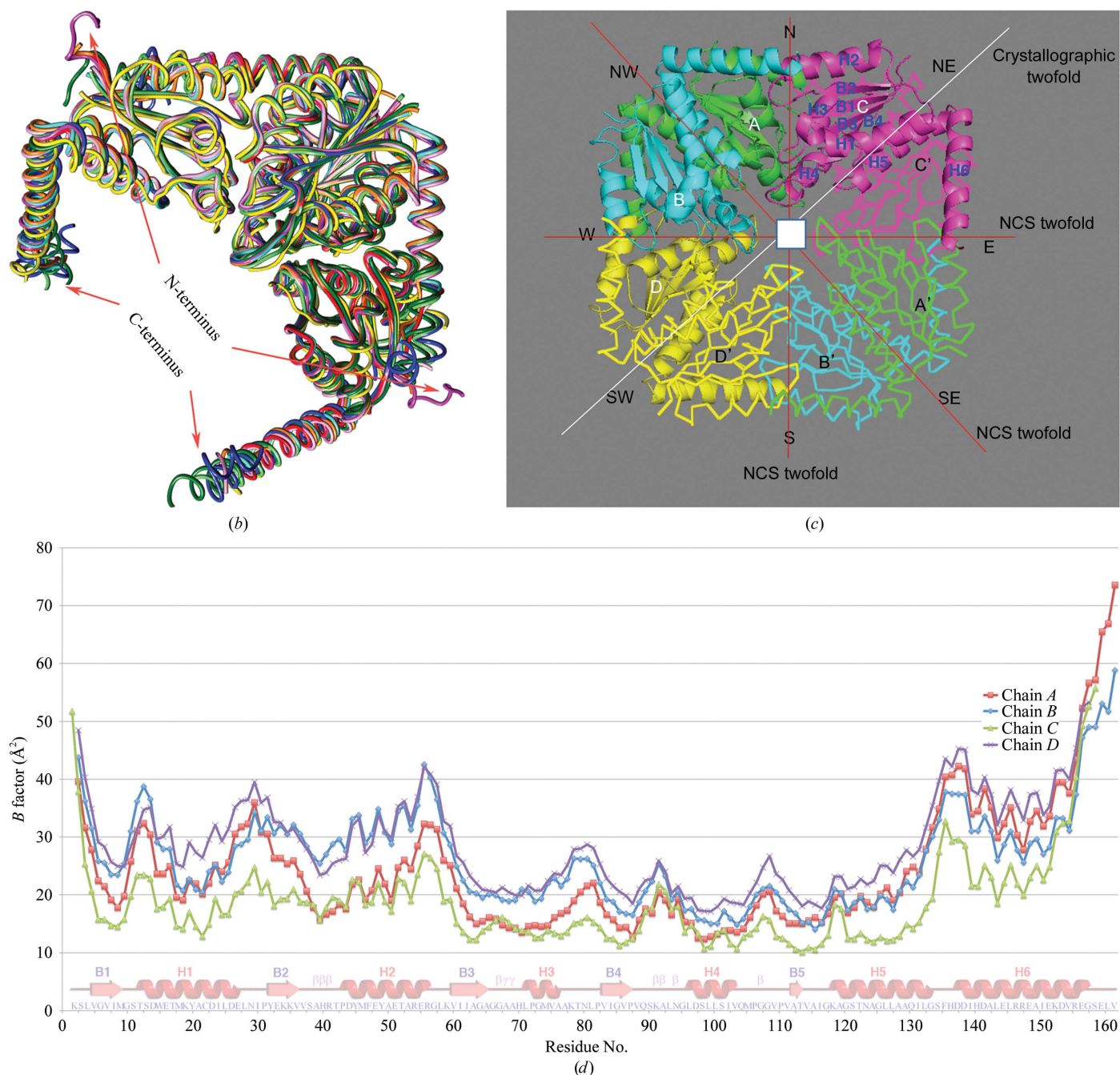
(b)

(c)

(d)

Figure 1 (continued)
(b) Secondary-structure superposition of 4ay3 (red) with PDB entries 3lp6 (dark green), 1xmp (magenta), 3oow (cyan), 3rg8 (yellow), 3trh (light green), 1o4v (blue), 1u11 (orange) and 1qcz (violet). Note the distinct conformation of the N-terminus in PDB entry 1xmp, where the His-tag extension participates in the crystal contacts, and also the long C-terminus in PDB entry 3lp6. (c) Schematic ribbon depiction of the tertiary and quaternary structure of PurE. N, S, E and W correspond to the main geographic directions. (d) B-factor profile of the four different subunits in the asymmetric unit of the trigonal form of PurE. The existence of two flexible subdomains at the N-terminus (Met1–Glu58) and C-terminus (Leu133–Thr161) is apparent.

reported previously (Abad-Zapatero *et al.*, 2011). Details of the crystallization conditions are summarized here for self-consistency. In this study, we focused on three crystal forms: form *A* ($P3_121$; 0.1 *M* cacodylate pH 6.5, 0.75 *M* sodium acetate, which we refer to as the 'salt' condition), form *B* ($P6_522$; 0.1 *M* Tris pH 8.5, 15% PEG 4000, 0.35 *M* sodium acetate, which we refer to as the 'PEG' condition) and form *E* (*C*2; 0.1 *M* cacodylate pH 6.5, 0.3–0.4 *M* sodium acetate). The details of these crystal forms in comparison to structures of PurE from different species are summarized for convenience in Table 1 with regard to crystallographic conditions, unit-cell parameters and unit-cell content.

2.2. Data collection and crystallographic analysis of different crystal forms

2.2.1. Humidity-control devices. Two different devices were used to monitor the effect of the RH on the different crystal forms: trigonal, hexagonal and monoclinic. The first was the instrument referred to as a free-mounting system (FMS; Kiefersauer *et al.*, 2000). This device achieves dehydration by using two airstreams of 0 and 100% RH that are mixed to achieve the desired RH. A feedback mechanism based on dew-point measurement is used to determine the actual RH, which depends on the temperature of the sample. Variations in RH are achieved by software control of the two independent streams. This instrument was installed on a Rigaku MicroMax-007 X-ray generator on an in-house installation. The second instrument, which is described as a humidity-control device (HC1b), is based on the nozzle of a standard cryostream (Sanchez-Weatherby *et al.*, 2009) and was located at the ESRF. The instrument was transportable and the various experiments were performed on various beamlines at the ESRF: BM14, ID14-1 and ID14-2 (Abad-Zapatero *et al.*, 2011).

2.2.2. Stepwise humidity-control experiments. The initial observations of the elastic response of the hexagonal form were performed using the FMS system. The protocol for these experiments was as follows. Changes in the RH of the crystal were specified using the FMS instrument software and the crystal was allowed to

equilibrate for approximately 2 min. After equilibration, diffraction images were collected at the indicated RH points using the following experimental conditions: 5 min exposure time and a crystal-to-detector distance of 250 mm at room temperature (*T* = 293 K) using the in-house installation described above. In these early experiments, only one diffraction image was collected at each RH point (Figs. 2*a* and 2*b*).
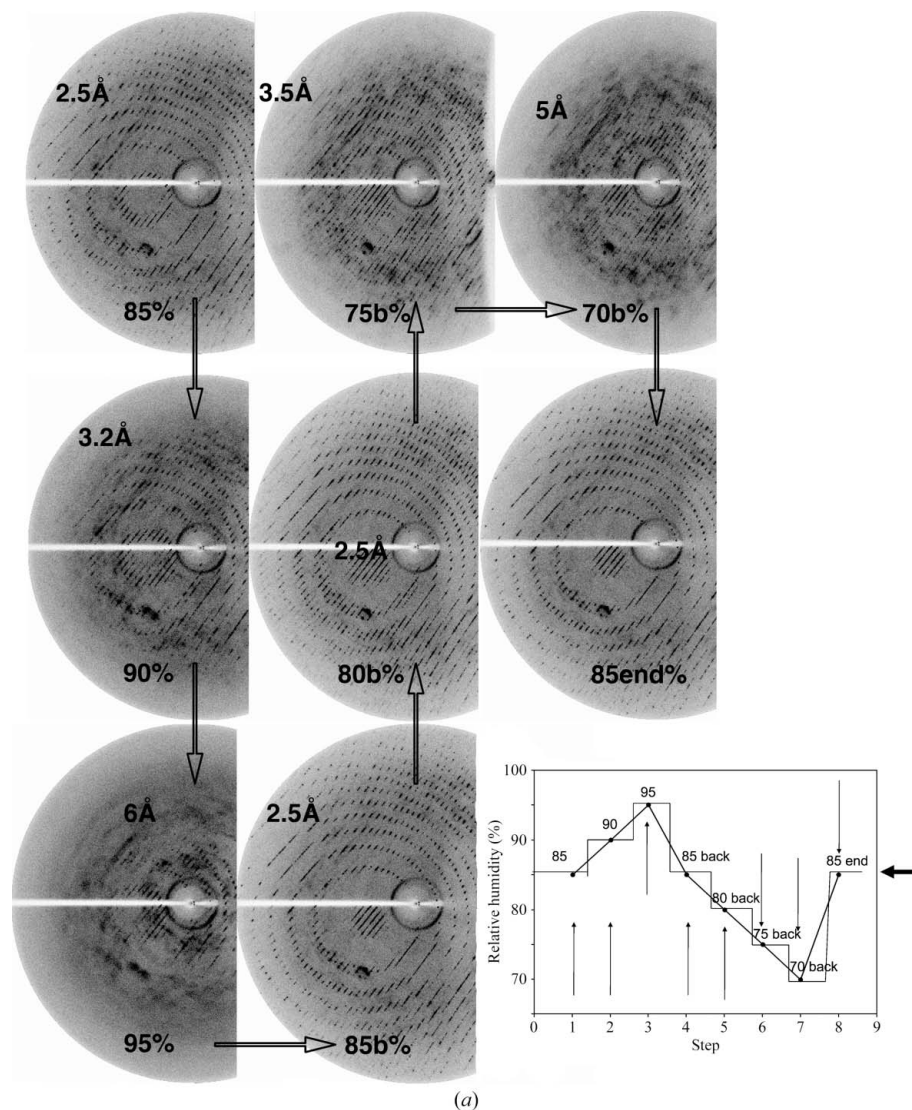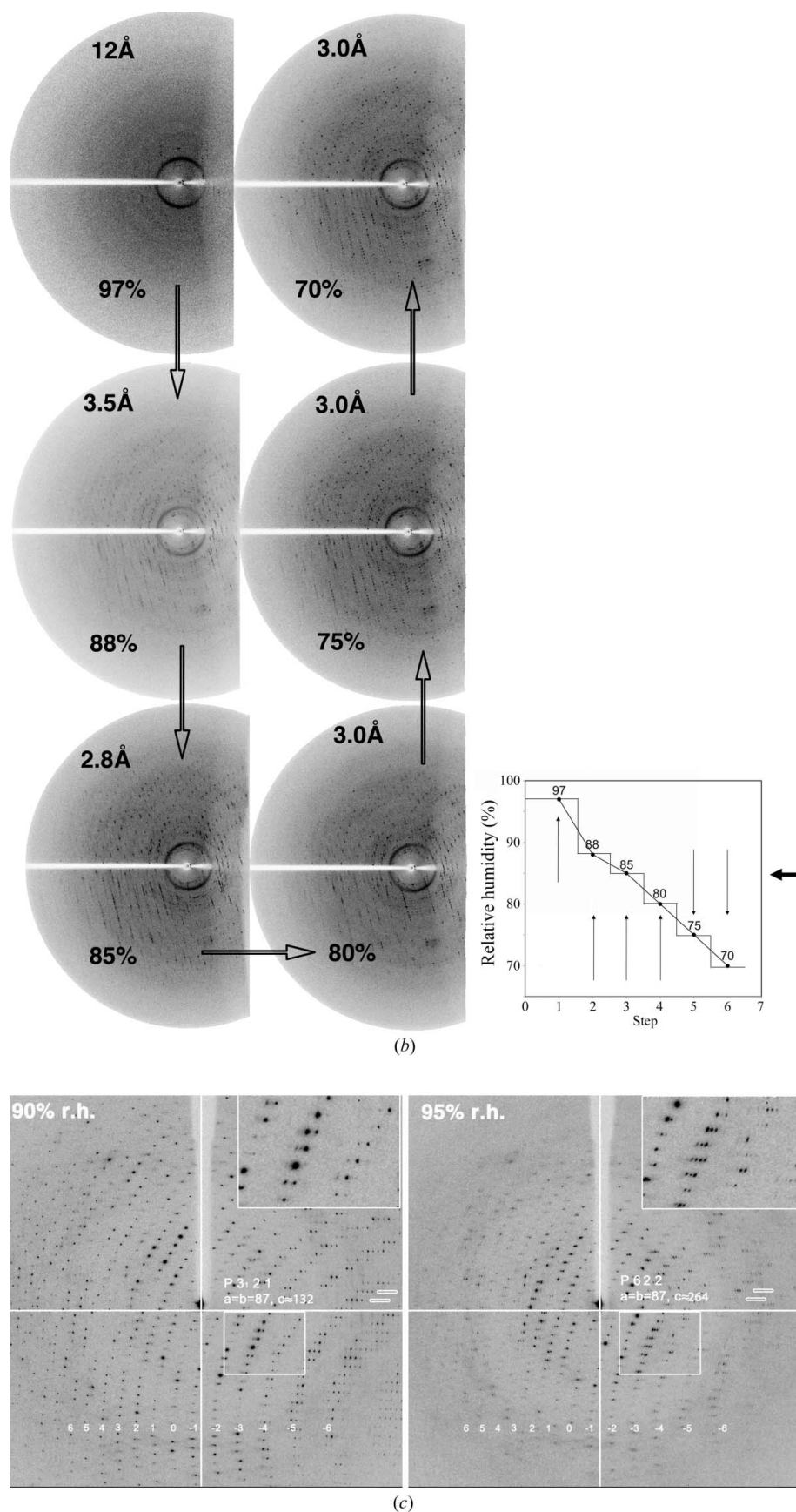

Figure 2
Effect on the diffraction pattern of changes in the RH of the crystals of PurE: (*a*) $P6_522$, (*b*) $P3_121$. Sequential diffraction patterns upon changes in the RH as indicated in the lower right panel for each experiment. Arrows indicate the sequence of the RH changes. The experiments were conducted with the FMS installed in the in-house rotating anode as described in §2.2.2. The collage is made up of images directly obtained from the FMS, in which the right-hand side portion of the detector is partly shaded by the mechanical jacket containing the airflow hoses (Kiefersauer *et al.*, 2000). (*c*) Distortion of the PurE diffraction pattern upon changes in the RH. The diffraction pattern of the $P3_121$ form (left, 90% RH) changes upon raising the RH to 95% (right) and the appearance of the longer (264 Å) spacing is apparent, indicating the transition from the trigonal to a hexagonal lattice in which the *a* and *b* axes are conserved and the *c* axis approximately doubles. It was impossible to determine any systematic absences to establish the presence of screw-axis symmetry. The numbers below refer to the labelling of the different 'lunes' of the diffraction patterns relative to the origin of the reciprocal lattice. Inset: close up of the indicated white rectangles in (*c*). This experiment was performed on ESRF beamline ID14-2.

12Å 97%

3.0Å 70%

3.5Å 88%

3.0Å 75%

2.8Å 85%

3.0Å 80%

(b)

90% r.h. P 3₂ 2 1 a=b=87, c≈132 6 5 4 3 2 1 0 -1 -2 -3 -4 -5 -6

95% r.h. P 6 2 2 a=b=87, c≈264 6 5 4 3 2 1 0 -1 -2 -3 -4 -5 -6

(c)

Figure 2 (continued)

A more detailed characterization of the response of the different crystal forms to changes in the RH was performed at the ESRF using the HC1b unit. The RH protocols for these experiments were similar to those described for the FMS, except for the specific control of the RH in each instrument (Greenspan, 1977). Diffraction images for the hexagonal form ($P6_522$) were collected on BM14 ($\lambda = 0.97625$ Å) with an exposure time of 0.5 s. The initial experiments were focused on establishing the optimal RH for the hexagonal form as a control and as a consistency check between the RH values of the FMS and HC1b instruments. Within the experimental limits of the two instruments, the RH for optimal diffraction of the hexagonal form was ~85% using both instruments and the changes observed in the diffraction pattern upon varying the RH were comparable, with no appreciable difference that could be interpreted as radiation damage.

For the detailed study of the cell changes and the amount of ordered solvent in the different hydration stages of the hexagonal form at different RH values, the experiment was performed collecting two frames (90° apart) for each humidity step. The experiment began at a humidity of 95% RH (relative to the instrument) and it was decreased in 5% steps until it reached 70% (Fig. 3). For a comparable study, diffraction images for the monoclinic and trigonal forms ($C2$ and $P3_121$) were collected on beamline ID14-1 (ESRF) using the HCb1 instrument. Two frames of data (1°) for each humidity value were collected separated by 90°. Attempts were made to collect full data sets at high resolution at distinct RH values at room temperature for all three forms, but these were unsuccessful owing to severe radiation damage.

2.2.3. Synchrotron data collection. Full data sets were collected under cryogenic conditions ($T = 100$ K) for the three different forms at their optimal RH, as described previously (Abad-Zapatero et al., 2011), using 1° oscillation and exposure times ranging from 0.1 to 0.5 s over a maximum of 120°. Full data sets were collected for the trigonal,

Table 2
Refinement statistics for the three crystal forms of PurE.

Refinement statistics for the three different crystallographic forms, $P3_121$, $P6_522$ and $C2$, with details of the solvent: the number of acetate and water atoms. Details of the data-collection statistics for the three crystal forms can be found in Abad-Zapatero et al. (2011). Values in parentheses are for the last resolution shell.

PDB entry	4ay3	4ay4	4b4k
Resolution range (Å)	49.5–1.76 (1.80–1.76)	43.9–2.0 (2.07–2.00)	47.1–2.5 (2.65–2.50)
Space group	$P3_121$	$P6_522$	$C2$
R_{work}	0.15 (0.17)	0.15 (0.16)	0.18 (0.22)
R_{free}	0.18 (0.23)	0.18 (0.18)	0.21 (0.26)
R.m.s.d. bond lengths (Å)	0.011	0.012	0.008
R.m.s.d. bond angles (°)	1.40	1.41	1.31
No. of protein atoms	4755	4704	13945
No. of solvent atoms	497	352	483
No. of ligand atoms	12	20	0
Protein mean B factor (Å^2)	20.6	26.2	45.3
Water mean B factor (Å^2)	36.0	33.1	37.7
Solvent content† (%)	37.8	39.6	39.8

† Estimated using the Matthews coefficient (Matthews, 1968).

hexagonal and monoclinic forms at high resolution on beam-lines ID14-1 ($\lambda = 0.9334$ Å), ID14-2 ($\lambda = 0.9330$ Å) and ID14-1 ($\lambda = 0.9334$ Å), respectively. Further details are given in the previous publication and in the PDB depositions (see Table 1).

2.2.4. Data indexing and processing. Indexing and characterization of the different crystal forms upon changes in the RH was performed using the interactive *iMOSFLM* package (Battye et al., 2011) and *HKL*-2000 (Otwinowski & Minor, 1997), and these packages were also used to process the complete high-resolution data sets for the three crystal forms. Indexing of the transition from the trigonal form to the hexagonal form was performed with two 1° frames (90° apart) using the *iMOSFLM* package (Fig. 2c).

2.2.5. Protein structure refinement. The structures of PurE in the high-symmetry forms (trigonal and hexagonal) were solved using the coordinates of *B. anthracis* PurE as a search model (PDB entry 1xmp) and were partially refined as described in the previous paper (Abad-Zapatero et al., 2011). Further refinement took place with special attention to the crystal contacts, the conformation of the N- and C-termini and the solvent structure (see below) for all of the chains in the different crystal forms. The final statistics for the refinement of these two crystal forms are summarized in Table 2.

In brief, the refined structure of the hexagonal form contained 4704 protein atoms, 352 water molecules and five well defined acetate molecules, with final refinement statistics of $R_{work} = 0.15$ and $R_{free} = 0.18$. The ordered residues in chains A, B, C and D were Lys2–Gly157, Lys2–Leu160, Lys2–Gly157 and Met1–Val161, respectively. The refined trigonal form consisted of 4755 protein atoms, 497 water molecules and three well defined acetate molecules, with final refinement statistics of $R_{work} = 0.15$ and $R_{free} = 0.18$. The ordered residues in chains A, B, C and D were Lys2–Val161, Lys2–Val161, Met1–Ser158 and Lys2–Gly157, respectively.

The monoclinic $C2$ crystal form (form E) was solved using a similar strategy and a definitive solution was found with three independent tetramers in the asymmetric unit. This structure

was also partially refined, and details can be found in the previous publication (Abad-Zapatero et al., 2011). The 12 chains in the asymmetric unit (three tetramers) correspond to a full octamer in a general position and a tetramer at a crystallographic twofold. The final refinement statistics for this form were $R_{work} = 0.18$ and $R_{free} = 0.21$. The model consisted of 13 945 protein atoms and 483 water molecules (Table 2). The number of independent chains (corresponding to one tetramer and a full octamer) was 12 (chains A–L), and the ordered residues were Lys2–Gly157 (chain A), Lys2–Gly157 (chain B), Lys2–Val161 (chain C), Lys2–Gly157 (chain D), Lys2–Val161 (chain E), Lys2–Gly157 (chain F), Lys2–Val161 (chain G), Lys2–Glu156 (chain H), Lys2–Gly157 (chain I), Lys2–Glu156 (chain J), Lys2–Gly157 (chain K) and Lys2–Gly157 (chain L). Chains E–L form the first octamer and chains A–D form the additional tetramer.

2.2.6. Solvent-structure refinement and analysis. During the final stages of refinement of the three forms, water molecules were automatically assigned using *Coot* (Emsley & Cowtan, 2004) and were visually checked for positive density in both $2F_o - F_c$ (1σ) and $F_o - F_c$ (2σ) maps from *REFMAC*5 (Murshudov et al., 2011) refinement. Water molecules with consistent hydrogen bonding were kept.

An analysis of the solvent structures associated with each specific asymmetric unit and of the interstitial surface waters (ISWs) was performed with the *Water Polygon Finder* (Lee & Kim, 2009) and the results obtained for these crystal forms of PurE were compared with some of the results obtained by these authors for an extensive set of protein structures refined at high resolution and deposited in the PDB. Water polygons were modelled within the volume enclosed from 0.2 to 1.2 in fractional cell coordinates for all three axes. The PurE crystal forms studied in this work were $P6_522$, $P3_121$ and $C2$. It should be noted, however, that in their initial analysis Lee and Kim did not make any distinction between different crystallization media, in particular the presence or absence of PEG, certain organic precipitants or different salts (high and low concentrations).

The bulk-solvent contents and unit-cell volumes for each crystal form at different RH stages were calculated from the corresponding unit-cell parameters using the *Matthews Probability Calculator* (Kantardjieff & Rupp, 2003). The number of waters in the bulk solvent was obtained by dividing the total solvent volume by 10.9 Å^3, which corresponds to the volume of a water molecule (1.375 Å radius). Finally, the number of waters divided by the number of ordered residues in the asymmetric unit for the corresponding chains was plotted. All of these data are summarized in Fig. 3(a).

2.3. Structure analysis and comparisons

The comparison of the fully refined structures of the three crystallographic forms of PurE was performed using the *LSQKAB* routine (for least-squares fitting) from the *CCP*4 package (Winn et al., 2011) and the *LSQMAN* routine (for r.m.s.d. calculations) from the Uppsala Software Factory (Kleywegt, 1996) as the core of several Unix scripts. The

superposition was also performed against the two octameric particles present in the $C2$ crystal form: one with all subunits unconstrained and the other formed by a crystallographic dyad running through the vertex direction (upper right to lower left or NE–SW; Fig. 1c). Moreover, the emphasis of the structural comparisons was placed at the level of tertiary and quaternary changes with respect to the previously published structure of *B. anthracis* PurE at 1.8 Å resolution (PDB entry 1xmp; Table 1). The least-squares superposition and the final r.m.s.d. calculation were performed for the residues between the beginning of the ordered chain (Lys2) and Leu160. A full comparison of the tertiary structures was performed among the different crystal forms and for the different subunits in the asymmetric unit.

3. Results

In the sections that follow, we first present the changes in the diffraction patterns observed for three crystal forms of PurE (hexagonal, trigonal and monoclinic) caused by systematic variation in the RH of their environment. We also examine the changes observed in the three different crystal lattices upon systematic changes in the RH. Subsequently, we discuss the results of a detailed analysis of the refined structures of the three crystal forms obtained at cryogenic temperatures under conditions of optimal diffraction, seeking to understand the factors that could explain the differing behaviour of the hexagonal and trigonal forms. The detailed structural analysis includes comparison of the monomeric and oligomeric PurE structures, their packing arrangements, solvent structure (bulk and ordered) and distribution of crystal contacts.

3.1. Diffraction-pattern changes

3.1.1. Elastic diffraction-pattern changes ($P6_522$). One of the most striking observations of the work with this form was the resilience of the crystal lattice to alterations in the RH of the medium surrounding the crystals. Initial studies had shown that the hexagonal form had optimal diffraction at ~85% RH in the FMS in-house installation (Abad-Zapatero *et al.*, 2011). During this work, we observed that the quality and extent of the diffraction pattern was reasonably reproducible even upon large changes in the RH. Fig. 2(a) (inset) depicts a typical sequence of elastic changes in the diffraction pattern of this crystal form following sequential changes in the RH beginning at the optimum value of 85%, ramping up to 95% RH (in steps of 5%), overshooting the initial RH to reach a lowest RH of 70% and returning in one step to the initial RH (85%). Fig. 2(a) illustrates how the corresponding diffraction patterns decreased in quality (sharpness) and extent (lower resolution) when the humidity reached 95%, but that the quality was restored at the initial RH even after reaching a lowest RH of 70%, at which the crystals only diffracted to approximately 5 Å resolution.

The changes in the unit-cell parameters are shown in Fig. 3(b) (maximum of <3% in the c axis) and the lattices could be considered to be isomorphous. Indeed, although not fully

appreciable in the images, the $P6_5$ systematic extinctions appeared to be conserved, suggesting that the space group did not change during the entire process. Similar behaviour was observed on the synchrotron beamlines using the HC1b instrument with different crystals. It is this remarkable elasticity of the crystal lattice properties that we would like to explain by analyzing the refined structure of the hydrated PurE protein found in this crystal form.

3.1.2. Inelastic diffraction-pattern changes ($P3_121$). In contrast, changes in the RH of the crystals of the $P3_121$ form produced dramatic changes in the diffraction pattern in that they (i) were not elastic and (ii) resulted in a duplication of the c-axis spacing, making it more similar to a hexagonal form.

The overall effect on the diffraction quality of sequential changes in RH from 97% to 88, 85, 80, 75% and a minimum of 70% can be observed in the sequence presented in Fig. 2(b) (inset). A dramatic increase in the quality of the diffraction pattern arose from the initial change from 97 to 88% RH, but the changes from here to 85% RH and beyond were characterized more by the appearance of additional reflections in what appeared to be a major disruption of the lattice (Fig. 2b). Exploration with a finer control of the RH of the environment and looking at specific regions of the reciprocal lattice, we found a duplication of the periodicity along the c axis of the trigonal form (Fig. 2c). This transition was most clearly observed within a narrow RH range of 5% using the HC1b device and was best characterized on the synchrotron beamlines. Although changes were observed using the in-house installation (Fig. 2b), it was impossible to characterize the transition as distinctly as was performed in the experiment depicted in Fig. 2(c). Other experiments designed to emulate the behaviour observed in the hexagonal form (cyclic changes in RH: 95–90–85–90–95%) using the trigonal crystal resulted in inelastic diffraction-pattern changes (data not shown). Unfortunately, we have not yet been able to confirm unambiguously whether the new crystal lattice has fully converted to the hexagonal $P6_522$ form described above or whether it belongs to a different space group in the hexagonal class.

3.2. Crystallographic unit-cell changes

The striking differences observed in the behaviour of the hexagonal and trigonal lattices upon changes in the RH of the crystal environment prompted us to perform a more systematic study of the changes in the three different lattices. A series of experiments were planned and performed at synchrotron beamlines as indicated. The results are summarized in Fig. 3.

Figs. 3(b), 3(c) and 3(d) show the changes in the unit-cell parameters of the three crystal forms (hexagonal, trigonal and monoclinic, respectively) upon gradual changes in the RH. The corresponding least-squares regressions for the three forms are also presented. Fig. 3(b) show how the a and c axes of the hexagonal unit cell change concurrently, suggesting that the expansion of the cell is isotropic. Although predominantly linear in the higher RH ranges, the observations have been modelled as parabolas to suggest a different rate of cell

expansion at lower RH values; the two parabolas are essentially parallel. Fig. 3(c) shows an analogous plot for the trigonal form, also with a quadratic regression fit for the observations. The coefficients of the regression analysis (not shown) for the rates of expansion of the a axes in the hexagonal and trigonal forms are significantly larger. In addition, the crossing over of the two parabolas for the trigonal form would suggest different expansion rates for the a and b axes, and this in turn could result in the doubling of the c axis as the RH increases, as we have observed.

In contrast, Fig. 3(d) shows a small and distinctly linear increase in the unit-cell parameters of the monoclinic $C2$ cell. The rate of increase of the axial parameters is lower than the values observed for the higher symmetry forms and the angle β has the lowest rate of increase. This behaviour suggests a very slow response to the RH change that is concurrent in all of the distance parameters, allowing the angular parameter to remain almost unchanged.

Fig. 3(a) illustrates the results of using these changes in the unit-cell parameters to estimate the changes in the corresponding volumes and in the number of water molecules per residue in the asymmetric unit of the different crystal cells as a function of the RH. These results provided additional support for the hypothesis that the trigonal form shows a distinctive behaviour at high RH values, responding more strongly to the changes and 'absorbing' more water molecules per residue as its crystal unit cell expands. The structural framework that supports these observations and this hypothesis is discussed in the following.

3.3. Molecular structural comparisons

The structure of the PurE octamer has been described previously in the initial report of the *E. coli* enzyme (Mathews *et al.*, 1999) and subsequently in the initial report of the *B. anthracis* structure (Boyle *et al.*, 2005). The three structures reported here are fully consistent with the previous results. In the high-symmetry forms (trigonal and hexagonal), the 422 symmetrical oligomer is constructed by a crystallographic dyad running diagonally (NE–SW direction; Fig. 1c) through the vertices in a direction perpendicular to the higher symmetry axes (3_1 or 6_5). The octamer is generated in a similar
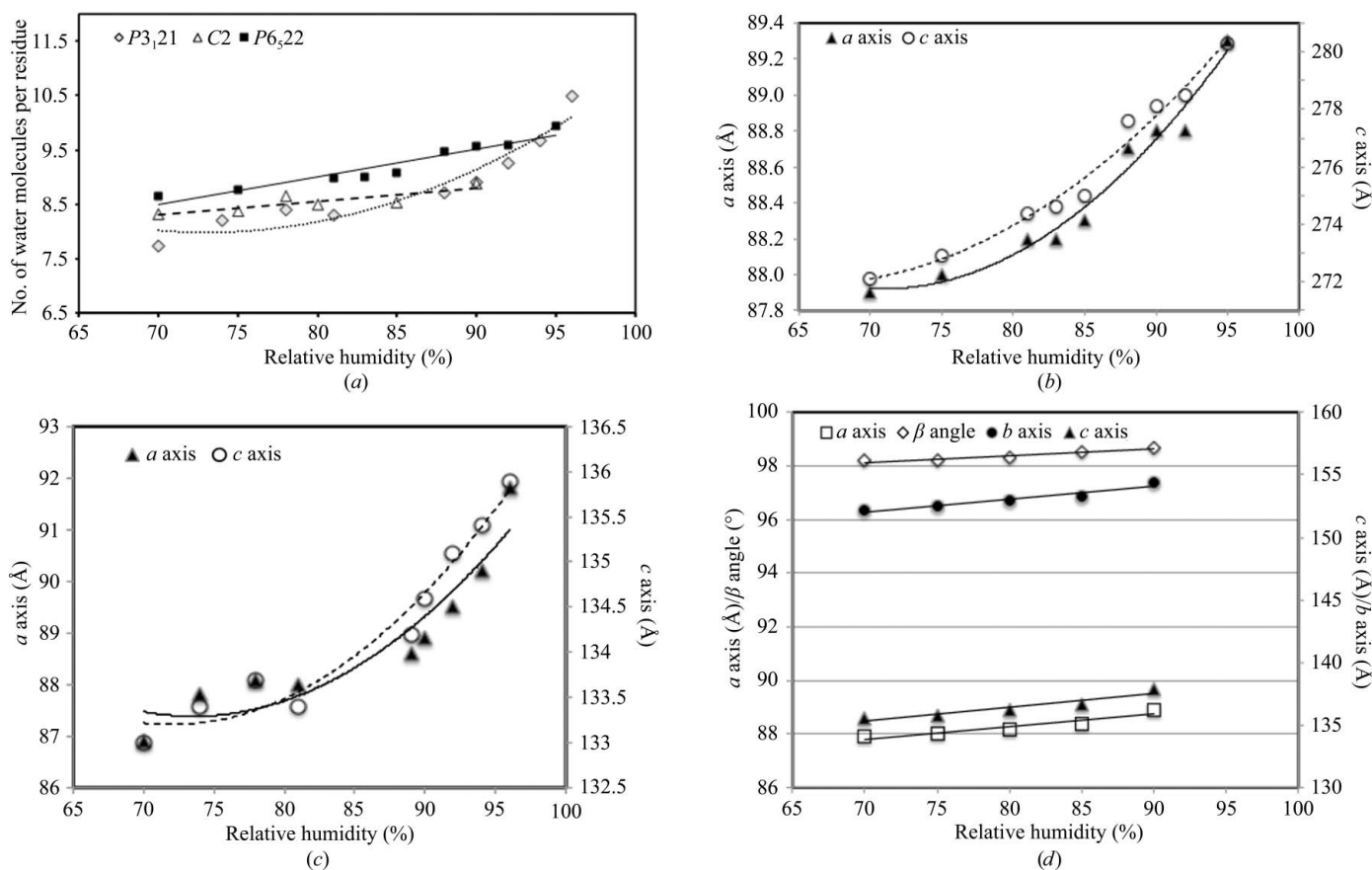

Figure 3
Crystallographic unit-cell changes upon changes in the RH. (a) Number of solvation water molecules per residue. Graph of the number of solvation waters as a function of the RH in the three different crystal forms of PurE. The calculations were performed using the unit-cell parameters for each RH as described in §2.2.6. The monoclinic and hexagonal forms were fitted by a regression line. The trigonal form was fitted by a parabola. (b) Variation of the unit-cell parameters for the hexagonal crystal form upon changes in the RH. The variations in the a and c axes were fitted by parabolas. (c) Variation of the unit-cell parameters for the trigonal crystal form upon changes in the RH. (d) Variation of the unit-cell parameters for the monoclinic crystal form upon changes in the RH. All of these experiments were performed using the HC1b device at the ESRF on beamlines BM14 (hexagonal) and ID14-1, using one crystal of each form for each experiment involving sequential variation of the RH.

way in the structure of *Mycobacterium tuberculosis* PurE (PDB entry 3lp6), although the crystal symmetry and packing are different (Table 1). The most extensive contacts (covering approximately 1800 Å^2) take place among PurE subunits along the diagonal direction and at the centre of the contact area, including a duplication of the $\beta2$–$\beta1$–$\beta3$–$\beta4$–$\beta5$ β-sheet by the perpendicular twofold (Fig. 1, top right quadrant). The contact area extends from the edge of the octamer to the hydrophobic contact between Phe135 (subunits *A–B*) that adopts a slightly different conformation in the twofold-
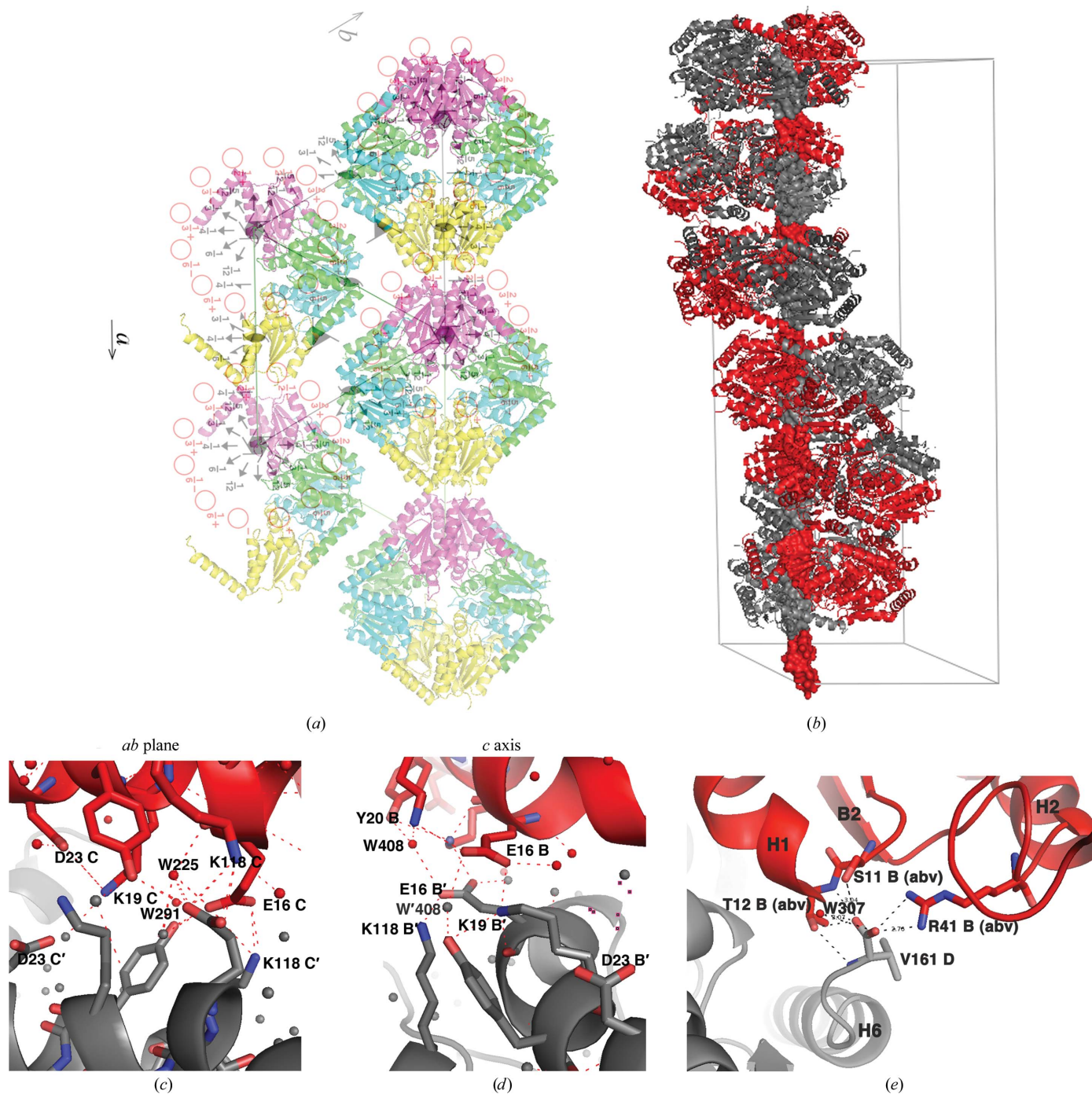

Figure 4
Crystal packing for *P*6$_5$22. Comparative views of the hexagonal form. (*a*) Packing view of the *ab* plane. The colours correspond to the different chains, with the same code as used in Fig. 1(*c*). (*b*) View of the unit cell perpendicular to the 6$_5$ screw axis with the contact residues that are within a distance of 5 Å from the *c* axis, suggesting a 'pivot' point. The colours red and black correspond to the two halves of the PurE octamer formed by the dyad perpendicular to the high-symmetry axes. (*c*) Closer view of the 6$_5$ screw-axis contacts (subunits *C–C'*) and (*d*) *B–B'* 60° away. In both cases residues within H1 mediate most of the contacts in two different orientations: parallel (*c*) and cross-helical (*d*). This part of the structure protrudes from the ideal flat appearance of the PurE octamer (Fig. 1*c*). (*e*) Critical crystal contact of the ordered C-terminus of subunit *D* with residues of H1 and H2 of subunit *B* of the particle above. Dashed lines correspond to hydrogen bonds.

symmetrical contact (D–D'), which is crystallographic in the trigonal and hexagonal forms. Surprisingly, the central contact is made by the shortest of the β-strands ($\beta5$; Thr113-Val114-Ala115). The high-symmetry axis does not coincide with the particle axis, being displaced by a significant distance. In the hexagonal form, the particle fourfold is 20 Å away from the 6_5 axis and in the trigonal form the separation is 12.4 Å. This lack of coincidence prevents the arrangement of the particles along the high-symmetry axis from being enclosed in a column-like space as is observed in the tetragonal structure of *M. tuberculosis* PurE (PDB entry 3lp6). The last helix of the monomeric particle (H6) extends along the edge of the square in an arm-like fashion and makes numerous contacts with H2 and strand $\beta2$ of the neighbouring external subunit, thus contributing extensively to the contacts in the NS and EW directions that configure the edge of the PurE particle.

An analysis of the main-chain temperature factors (B factors) of the refined structures of the three forms reveals an intriguing subdomain structure within the PurE subunit that it is appropriate to discuss in the dynamical description of the structure. The initial depiction of the tertiary fold of PurE was performed in terms of a central domain containing five parallel β-strands ($\beta2$–$\beta1$–$\beta3$–$\beta4$–$\beta5$) flanked by helices on either side, as indicated above (Mathews *et al.*, 1999). However, from the dynamic standpoint a different picture emerges. Fig. 1(*d*) suggests the existence of two flexible subdomains within each subunit, one at the N-terminus comprising $\beta1$–H1–$\beta2$–H2 (residues Met1–Leu58) and a smaller one at the C-terminus containing only H6 (Phe135–Val161). These two smaller subdomains extend from the flat disc-like core of the octamer ($\beta3$–H3–$\beta4$–H4–$\beta5$–H5) and provide contacts with the molecules above and below in the crystallographic assembly.

In PurE from *B. anthracis*, the sequence (Glu151-Lys152-X-X-Arg155-Glu156) at the carboxy end of H6 plays a very important role in stabilizing the helical conformation. We have observed supplementary ionic interactions between Glu151 and Arg155 and between Lys152 and Glu156 along the helical axis in all of the different chains that strengthen the helical configuration (one further turn). Beyond Glu156 the chain adopts different conformations and is often disordered. Also, significant variations among the different subunits can be observed at the N-terminus of the different structures, which in some cases are rather significant owing to ordering of the His-tag extension, as observed in the *C*2 form of *B. anthracis* PurE (PDB entry 1xmp), and where it plays a significant role in crystal contacts (Figs. 1*b* and 1*c*).

3.4. Crystal-packing comparisons

Given its biological importance in the survival of human pathogens such as *M. tuberculosis*, *B. anthracis* and others (see Table 1), the three-dimensional crystal structure of PurE has been the object of intensive study. Three-dimensional structures of PurE from organisms ranging from *E. coli* (PDB entry 1qcz) to *M. tuberculosis* (PDB entry 3lp6) have been deposited in the PDB. There is a crystal structure of *B. anthracis* PurE at 1.8 Å resolution (PDB entry 1xmp) that is directly related to this work but was obtained using a different His-tag sequence at the N-terminus. Although this form also belonged to space group *C*2, the unit-cell parameters are not directly related to the *C*2 form that we present here and should be considered to be different. The earlier structure contains one full octamer in the asymmetric unit, while our *C*2 crystal form has a larger unit cell and contains the equivalent of a full octamer and an additional tetramer in the asymmetric unit. This large discrepancy is probably a consequence of the fact that the affinity His-tag sequence extension introduced at the N-terminus differs in the two crystal forms (Figs. 1*b* and 1*c*). Since the most stable aggregate appears to be an octamer, based on the previously reported structures, the different crystallographic forms differ in the arrangement of the octamers in the crystal lattices. Several structures pack the 422 particle into tetragonal lattices [PDB entries 1u11 (Settembre *et al.*, 2004), 1qcz, 1o4v (Schwarzenbacher *et al.*, 2004) and 3lp6], while the remaining structures use low symmetry to crystallize the entire octamer or even two full octamers in the asymmetric unit (PDB entry 3rh; J. Cheung, M. Franklin, M. Rudolph, M. Cassidy, E. Gary, F. Burshteyn & J. Love, unpublished work).

The crystal structures discussed here are unique in that they basically pack the 422 particle in a hexagonal crystal packing. Moreover, our lowest symmetry form (*C*2) retains essentially the same packing in the *ab* plane and all of the dimensions of the three crystallographic cells are naturally related (Abad-Zapatero *et al.*, 2011). Here, we present a molecular-packing analysis based on the atomic crystal contacts and attempt to relate them to the corresponding crystal contacts responsible for the three different crystal forms. References are made to other crystal structures only when necessary.

3.4.1. The hexagonal *P*6$_5$22 form. The packing in the *ab* plane can be described as a hexagonal close packing of the 422 octameric aggregate. Thus, each octamer is surrounded by six neighbouring octamers in the *ab* plane, but only two of them are close enough to interact. There are two intermolecular contacts between subunits *D* and *C* that relate the N-terminus of subunit *D* or *C* of the tetramer in the asymmetric unit to the N-terminus of subunit *D'* or *C'* of the symmetric octamer along the *a* axis and correspond to vertex–vertex contacts (Fig. 4*a*).

Along the 6_5 screw axis, the quadrangular particles stack around a centre located between residues Thr113 and Ala115 of subunit *C* in the central part, which is the most rigid core of the protomer (Fig. 1*d*). This 'pivot' point is approximately 20 Å from the origin of the octamer along the diagonal direction (Figs. 1*b*, 1*c* and 4*b*). The contacts in the vicinity (<5 Å) and along the high-symmetry axis are shown in Fig. 4(*b*), and all of the contacts are summarized in Table 3.

The *c* axis is localized between the five-stranded parallel β-sheet of subunit *C* and the antiparallel five-stranded β-sheet of subunit *C'* of the same oligomer. The residues involved in this intra-octamer contact are Val111, Thr113 and Ala115 of $\beta5$ together with the same residues from the other subunit. There is a hydrogen bond between the carbonyl O atom of Thr113 and the amine N atom of Thr113 from subunit *C'* mediated by two water molecules (W13 and W15). There is

Table 3
Summary of packing interactions observed in the three different crystal forms ($P3_121$, $P6_522$ and $C2$).

Packing interactions were determined using the *PISA* software (Krissinel & Henrick, 2007). Different interactions and interfaces were visualized using *PyMOL* (Schrödinger LLC) or *Coot* (Emsley & Cowtan, 2004). Polar interactions are highlighted in bold; all other interactions are hydrogen bonds. ref, reference molecule; abv, molecule above; blw, molecule below.

(*a*) $P3_121$.

Intermolecular, threefold screw, cross-helical	Intermolecular, parallel helical	C-terminal	Intramolecular	Intramolecular (antiparallel β-sheet)
Glu16C'(blw)···Lys19C(ref)	Glu16A'(ref)···Tyr20A(blw)	Val161B(ref)···Arg41C'(blw)	Gln89C···Lys91A'	Val111C
Glu16C'(blw)···W407···Tyr20C(ref)	Glu16A'(ref)···Lys118A(blw)	Val161B(ref)···Thr12C'(blw)	Gln104C···Ile116C'	Thr113C···Thr113C'
Glu16C'(blw)···W407···Lys118C(ref)	Tyr20A'(ref)···Glu16A(blw)	Arg41C(ref)···Val161B'(blw)	Ile116C···Gln104C'	Ala115C
Lys19C'(blw)···Glu16C(ref)	Lys118A'(ref)···Glu16A(blw)		Lys91A···Gln89C'	
Tyr20C'(blw)···W407···Glu16C(ref)	Asp23A'(ref)···W255···Lys19A(blw)			
	Asp23A'(ref)···W279···Lys19A(blw)			

(*b*) $P6_522$.

Intermolecular, cross-helical H1	Intermolecular, sixfold screw, parallel helical H1	C-terminal	Intramolecular	Intramolecular, sixfold screw
Glu16B'(blw)···Lys19B(ref)	Glu16C'(blw)···Tyr20C(ref)	Val161D(ref)···Arg41B(abv)	Gln89B···Lys91C	Val111C
Lys19B'(blw)···Glu16B(ref)	Glu16C'(blw)···Lys118C(ref)	Arg41B'(ref)···Val161D(abv)	Leu99	Thr113C···Thr113C'
Tyr20B'(blw)···W408···Glu16B(ref)	Tyr20C'(blw)···Glu16C(ref)	Val161D(ref)···Ser11B(abv)	Leu100	Ala115C···Val103C'
Glu16B'(blw)···W408···Lys118B(ref)	Lys118C'(blw)···Gln16C(ref)	Val161D(ref)···Thr12B(abv)	Val103	Trp13
	Glu16C'(blw)···W225···Glu16C(ref)		Gln104B···Ile116A	Trp15
	Trp15C'(blw)···W291···Tyr20C(ref)		Val114	
			Ala115	

(*c*) $C2$.

Intermolecular	C-terminal	Intramolecular Parallel	Cross	C-terminal
Lys19H(ref)···Glu16H(abv)	Val161C(ref)···Ser11I(abv)	Lys19L···Lys19D	**Lys19I···Glu16A**	**Val161G···Arg41A**
	Val161C(ref)···Thr12I(abv)	Glu16L···Glu16D		Val161G···Thr12A
	Val161C(ref)···Arg41I(abv)			Val161G···Ser11A
	Val161E(ref)···Ser11H(blw)			
	Val161E(ref)···Thr12H(blw)			
	Val161E(ref)···Arg41H(blw)			

another contact between the carbonyl O atom of Val103 from subunit C and the amine N atom of Ala115 from subunit C'.

The contacts along the two high-symmetry axes (6_5 and 3_1) are mediated by the two mobile subdomains (Met1–Leu58 and Phe135–Val161) that protrude from the basic disc-like shape of the octamer. These contacts can be described as parallel helical (involving H1 and H5), cross-helical (H1 and H5) and the contacts provided by the ordered C-terminus (H6 and its extension) (Figs. 4*c*, 4*d* and 4*e*).

The intermolecular interactions along the 6_5 screw axis are established by helices H1 (Glu16, Tyr20 and Asp23) and H5 (Lys118) of subunit C of the reference octamer and the same residues of the octamer below (subunit C'). The O atom (O^ε) of Glu16 forms hydrogen bonds to the hydroxyl O atom of Tyr20 and the N atom (N^ζ) of Lys118. The contacts are repeated by a crystallographic dyad perpendicular to the 6_5 axis. There are two water molecules (W225 and W291) that mediate the contact between the octamers. On one side, water W291 mediates between the hydroxyl O atom of Tyr20 and the N atom (N^ε) of Trp15. On the other side, water W225 forms a bridge between the O atoms (O^ε) of the two Glu16 residues (Fig. 4*c*).

The other contact along the *c* axis is approximately 12.2 Å from the origin of the octamer along the diagonal direction; this coincides with the position of the 3_1 axis in the trigonal form and we refer to it as the cross-helical contact formed by residues in helices H1 (Glu16, Lys19 and Tyr20) and H5 (Lys118). The O atom (O^ε) of Glu16 in subunit B in the reference octamer forms a hydrogen bond to N^ζ of Lys19 in subunit B' of the subunit below. There are four hydrogen bonds mediated by water W408. Three of them are within the same subunit (Lys19, Tyr20 and Lys118) and the partner is the O atom (O^ε) of Glu16 in subunit B' (below; Fig. 4*d*). The side-chain conformations of the individual residues in these two contacts are not the same.

The carboxyl O atom of the C-terminal residue Val161 in subunit D of the reference octamer forms a hydrogen bond (2.91 Å) to the NH1 atom of Arg41 in subunit B of the octamer above. It should be noted that residue Arg41 is part of the active site of PurE. The same carboxyl group also forms a hydrogen bond to the O atom (O^γ) of Thr12 and the O atom (O^γ) of Ser11 in subunit B of the top octamer (Fig. 4*e*).

3.4.2. The trigonal $P3_121$ form. The packing in the *ab* plane is similar to that in the hexagonal form, except for the fact that

the crystallographic twofold is now along the *b* axis and results in a different handedness in the arrangement of the octamers along the axis (Fig. 5*a*).

Along the 3_1 screw axis, the quadrangular particles stack around a centre located between subunits *A* and *C* (Fig. 5*b*) that is approximately 12.2 Å from the origin of the octamer along the diagonal direction. In this space group, the octameric particles pivot around the position of the 3_1 screw axis in the neighbourhood of residues Gln89, Gln104 and Ile116 in the *C* chain and Lys91 in the *A* chain and make contacts with the corresponding residues across the perpendicular dyad forming the octamer (Fig. 5*b*). A summary of the contacts is presented in Table 3.

The intermolecular interactions along the 3_1 screw axis are established by cross-helical contact of H1 (Glu16, Lys19 and Tyr20) and H5 (Lys118) from subunit *C* of the reference octamer with the same residues in subunit *C′* of the octamer below. The carboxyl O atom (O$^\varepsilon$) of Glu16 forms a hydrogen bond to the amine N atom (N$^\zeta$) of Lys19. Water W407 mediates the other hydrogen bonds. The O atom (O$^\varepsilon$) of Glu16 in subunit *C* (reference) interacts through water 407 with the hydroxyl O atom of Tyr20 in subunit *C′* (below). The N atom (N$^\zeta$) of Lys118 of subunit *C* (reference) interacts through water 407 (symmetric) with the O atom (O$^\varepsilon$) of Glu16 of subunit *C′* (below). There is only one water molecule W407 (and a related symmetry molecule) that mediates
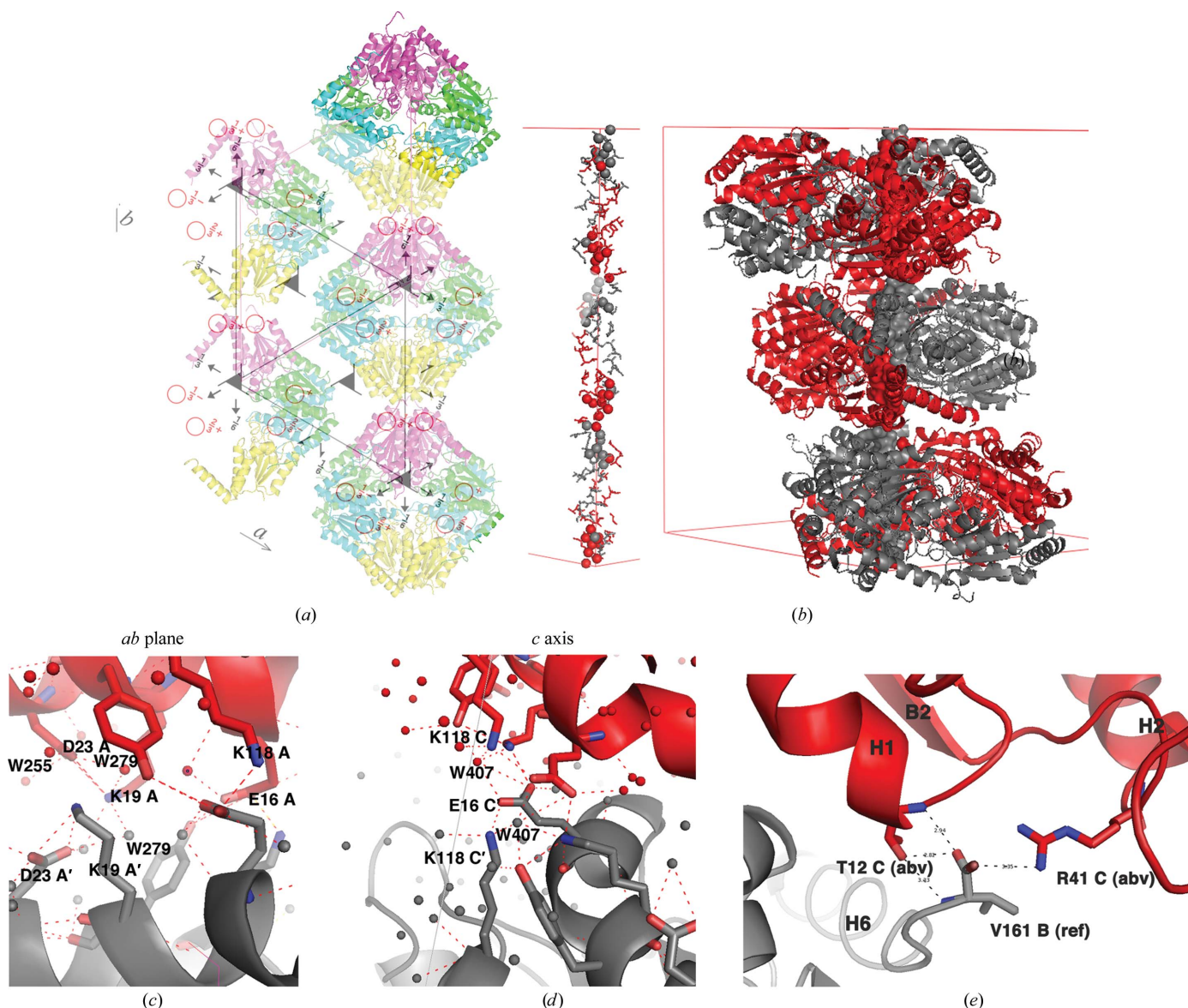

Figure 5
Crystal packing for $P3_121$. Comparative views of the trigonal form. (*a*) Packing view of the *ab* plane. (*b*) View of the unit cell perpendicular to the 3_1 screw axis with the contact residues that are within a maximum distance of 5 Å from the *c* axis, showing the same protruding area making the contacts along the trigonal axis. The red and black colours correspond to the two halves of the PurE octamer formed by the dyad perpendicular to the high-symmetry axis. (*c*) Closer view of the 3_1 screw-axis contacts (subunits *A–A′*) and (*d*) *C–C′* 60° away. As before, helix H1 mediates most of the contacts but the specific residue–residue contacts are different, including the side-chain conformations. (*e*) Critical crystal contact of the ordered C-terminus of subunit *B* with residues of H1 and H2 of subunit *C* of the particle above. Dashed lines correspond to hydrogen bonds.

interactions between residues from two octamers (Fig. 5c).

We have observed other kinds of interactions established by parallel helical contacts formed by H1 (Glu16, Tyr20 and Asp23) and H5 (Lys118) interactions between subunit *A* of the octamer below and subunit *A'* of the reference octamer. The O atom (O$^\varepsilon$) of Glu16 forms a hydrogen bond to the hydroxyl O atom of Tyr20 and the N atom (N$^\zeta$) of Lys118. Two water molecules, W255 and W279, act as linkers between the carboxyl O atom of Asp23 and the N atom (N$^\zeta$) of Lys19 (Fig. 5d).

The C-terminus is ordered in the two internal chains *A* and *B*. The carboxyl O atom of the C-terminal residue Val161 in subunit *B* of the reference octamer forms a hydrogen bond (2.91 Å) to the amine N atom of Arg41 in subunit *C'* (octamer below). The same carboxyl O atom forms hydrogen bonds to the amino N atom and the O$^\gamma$ atom of Thr12 in subunit *C'* (octamer below). These two interactions are part of an inter-octamer contact (Fig. 5e). In contrast, the C-terminus of subunit *A* interacts with the indole N atom of Trp15 as part of an intra-octamer contact.

3.4.3. The monoclinic C2 form. The relationship between the *C*2 unit cell and the corresponding hexagonal and trigonal unit cells (Otte & Crocker, 1965) is shown in Fig. 6(b). In the

ab plane the contacts between different layers of the crystal are mediated by three intermolecular contacts along the vertices of the octamer (*F* and *E*, *A* and *B*), including residues in the N-terminus, with the symmetry-related subunits (*I'* and *J'*, *A'* and *B'*) (Fig. 6a, Table 3).

Similar contacts to those described above for other space groups can be observed in the *bc* or *ac* planes. These include parallel, cross-helical and C-terminal contacts. The only difference is that in this space group (*C*2) the contacts are within the two molecules in the asymmetric unit and the symmetry observed is noncrystallographic. The intramolecular contacts between the octamer and tetramer in the asymmetric unit are summarized in Table 3.

In this crystal form, there are three fully ordered C-termini: two in the full octamer (chains *E* and *G*) and a third in subunit *C* of the tetramer. The carboxyl O atom of the C-terminal residue Val161 in subunit *E* of the reference octamer forms a salt bridge to the N atom NH1 of Arg41 in subunit *H'* (below). The same carboxyl forms a hydrogen bond to the amino N atom and the O atom (O$^\gamma$) of Thr12 in subunit *H'* (Fig. 6e). Only the ordered C-terminus of subunit *G* of the full octamer in the asymmetric unit makes contact with the neighbouring half-octamer in the asymmetric unit (intra-octamer contact).
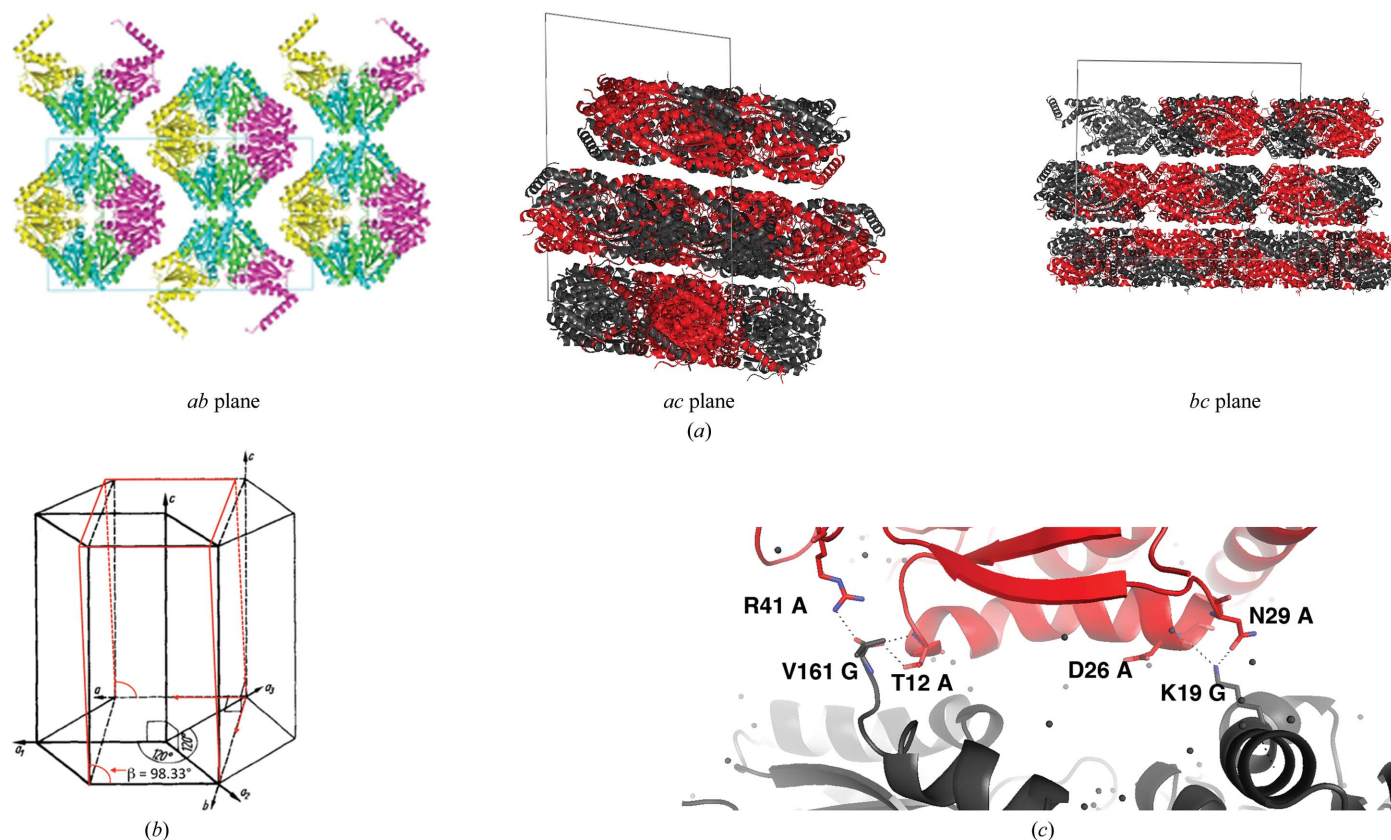

ab plane   *ac* plane   *bc* plane

(a)

(b)   (c)

Figure 6
Crystal packing for *C*2. Comparative views of the monoclinic form. (*a*) Packing views of the *ab*, *ac* and *bc* planes. Although somewhat distorted in the monoclinic cell, the contacts are very similar to those described above. (*b*) The geometrical relationship between the different unit-cell parameters: hexagonal, trigonal and the distorted *C*222 cell that results in monoclinic *C*2 upon distortion of the interaxial angle β. (*c*) A closer view of the ordered C-terminal intramolecular interaction between subunits *A* and *G* of the two different octamers in the unit cell: Arg41 of subunit *A* interacts with the C-terminal Val161 of chain *G* in the adjacent octamer below in the asymmetric unit. Chain labelling is as indicated in the text (§4).

Table 4
Acetate molecules found in the different crystal forms of PurE.

Intramolecular acetate molecules are numbered ACT1E–ACT4E in the hexagonal form; ACT9E is intermolecular and is located at a crystal contact. Only two acetate molecules were found in the trigonal form and none were found in the C2 form. The amino acids involved in the different interactions are represented by their single-letter codes. The subscripts (H) and (I) indicate hydrogen bonds and interface residues, respectively.

	M9–W15	V36–Y45	G66–G73	V88	S92–N94	P106	K118	R147	V154	S158
$P6_522$										
ACT1E	$A_{(H)}$	$A_{(H)}$	$A_{(I)}$			$C_{(I)}$				
ACT2E	$B_{(I)}$	$B_{(H)}$	$B_{(I)}$			$D_{(I)}$				
ACT3E	$C_{(I)}$	$C_{(H)}$	$C_{(I)}$			$A_{(I)}$				
ACT4E	$D_{(I)}$	$D_{(H)}$	$D_{(I)}$			$B_{(I)}$			$A_{(I)}$	
ACT9E					$B_{(H)}$(SX)†					
$P3_121$										
ACT6E	$C_{(I)}$	$C_{(H)}$	$C_{(I)}$			$A_{(I)}$				
ACT9E	$D_{(I)}$	$D_{(H)}$	$D_{(I)}$			$B_{(I)}$				

† Crystallographic symmetry $x - y$, $-y$, $-z$ ($P6_522$).
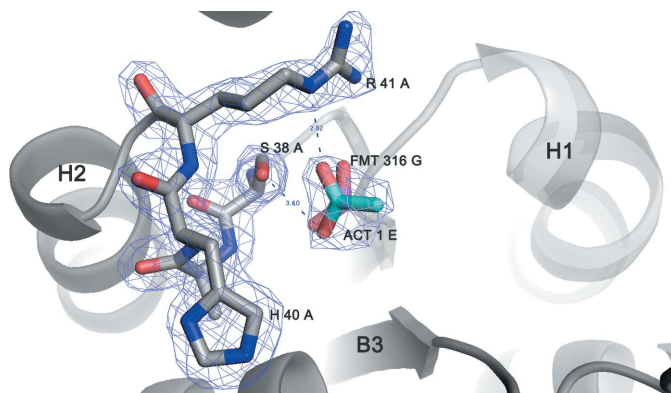
Figure 7
Active-site formate and acetate. View of the binding mode of the acetate molecule bound at the active site of chain A in the hexagonal form superposed on the revised coordinate set of PDB entry 1xmp (courtesy of the authors). The $2F_o - F_c$ density map was contoured at the 1σ level. This revised set included a formate near Ser43G (E. coli PurE). The acetate molecule corresponds to the formate found in subunit G of the monoclinic C2 form of B. anthracis PurE published previously (Boyle et al., 2005). A formate molecule was also found at a similar position in the active site of M. tuberculosis PurE (PDB entry 3lp6).

3.5. The solvent structure

3.5.1. Acetate sites. The vast majority of the crystal forms obtained for PurE grew under high-salt conditions (≥ 0.2 M ammonium dihydrogen phosphate and ≤ 2 M ammonium sulfate; Table 1). Among them, small organic monovalent anions such as sodium formate (0.8 M for PDB entry 1xmp and 1.8 M for PDB entry 3lp6) and sodium acetate are common. In particular, our three crystal forms were grown using sodium acetate at concentrations of 0.3 M ($P6_522$), 0.75 M ($P3_121$) and 0.3–0.4 M (C2). Thus, it is not surprising that acetate molecules are found at certain positions.

The most consistent and repeated acetate site in the structure of PurE is found within the active site near Ser38 and its binding pocket, as illustrated in Fig. 7. The position and orientation of this acetate ion coincide with those of the formate ion found bound near Ser38 of chain G in the struc-

ture of B. anthracis PurE (revised data set for PDB entry 1xmp courtesy of the authors) and with the sulfate ion in Thermotoga maritima PurE (PDB entry 1o4v). It is also coincident with one of the formate molecules found at the active site (near Ser43) in the tetragonal form of M. tuberculosis PurE (PDB entry 3lp6). However, this anionic site does not correspond to the nitrate moiety of the N^5-CAIR product as suggested by the three-dimensional structure of the complex of E. coli PurE with N^5-CAIR (PDB entry 2ate; Hoskins et al., 2007) or to the carboxyl of CAIR (PDB entry 2nsj; Hoskins et al., 2007).

In the hexagonal form ($P6_522$), the crystallization conditions for the crystal contained 0.3 M sodium acetate and 15% PEG 4000. Four acetate molecules could be assigned with confidence from the refined electron-density maps. Consistently, they are located in the active site of the PurE enzyme within the Ser38–Thr42 loop and in particular near Ser38 and Arg41 (Fig. 7). One of the acetate O atoms forms a hydrogen bond to the O atom (O^γ) of Ser38 and the other one interacts with the N atom (N^ε) of Arg41. In addition, there is a well defined acetate near the centre of the octamer that makes contact with Asn94 of chains A and D in the asymmetric unit and the corresponding chains (B' and C') by crystal symmetry (Table 4).

The crystals of the trigonal ($P3_121$) form grow in 0.75 M sodium acetate. Two acetate molecules were unambiguously found in difference density maps. The binding mode of the two acetate molecules in the active sites of subunits C and D is similar to that described above in the hexagonal form. There are no well defined acetates in the C2 crystal form (Table 4).

3.5.2. Bulk solvent. A critical question in understanding the different behaviour of the two high-symmetry forms in relation to the changes in RH is the role that the bulk solvent could play in the transition. Based on the unit-cell parameters at each RH, we analysed the bulk-solvent content of the different crystal forms upon changes in the RH and the results have briefly been discussed previously (§3.2, Fig. 3). The number of water molecules per residue ranges from eight to ten in all of the crystal forms. For the hexagonal and monoclinic forms this remains essentially constant. In contrast, the trigonal form shows a distinct upward trend beginning at approximately 90% humidity. Interestingly, the transition between the trigonal and hexagonal lattices takes place between 90 and 95% humidity (Fig. 2c). It is worth noticing that the line corresponding to the changes in the hexagonal lattice crosses the curve describing the changes in the trigonal lattice at approximately 95% humidity. As hypothesized earlier, this might suggest that the lattice transition is also related to changes in the bulk solvent.

3.5.3. Ordered solvent. In the refined hexagonal PurE structure there are 0.55 water molecules per ordered residue

Table 5
Statistics of polygon counts for the different crystal forms.

Results of the distribution of water polygons as provided by the *Water Polygon Finder* program (Lee & Kim, 2009).

	$P6_522$	$P3_121$	$C2$
No. of waters in asymmetric unit	352	497	483
Layer 1	312	407	453
Layer 2	40	90	30
No. of polygons in unit cell	84	157	20
3 vertices	60 (71.4%)	36 (23.0%)	
4 vertices	24 (28.6%)	15 (9.5%)	4 (20%)
5 vertices		76 (48.4%)	16 (80%)
6 vertices		18 (11.5%)	
7 vertices		12 (7.6%)	
No. of ordered protein residues	632	637	1890
Waters per residue	0.55	0.78	0.26

in the asymmetric unit, in contrast to the trigonal structure, in which there are 0.78 water molecules per ordered residue. In the monoclinic form refined at 2.5 Å resolution there are 0.26 water molecules per ordered residue, as shown in Table 5. The discrepancy between the number of water molecules per residue in the hexagonal and trigonal forms is larger than expected based on the resolutions of the two structures.

We compared the locations of the ordered waters in the two structures determined at high resolution. The $P6_522$ crystal structure has 352 water molecules and the $P3_121$ form has 497. Upon superposition, water positions within less than 1.5 Å were considered to be conserved. The number of shared waters is around 277 and these could be considered to be structural waters. The number of waters unique to each crystal form were 72 for the hexagonal form and 214 for the trigonal form. We proposed that the large number of waters in the trigonal form (three times the number) is greater than would be expected based on the resolutions of the structures. It is of note that these waters are predominantly located in the intermolecular contacts of the crystal (Fig. 8) and along the high-symmetry axis contacts.

The systematic comparison of the effect of the changes in the RH of the crystal on the hydration of the corresponding protein for the three different crystal forms presented earlier (§3.2, Fig. 3) provided insight into the different responses of the three different crystal forms to the humidity of the environment, which could have an effect on the amount of ordered solvent. The responses of the hexagonal and monoclinic forms are rather similar and are represented by an essentially horizontal line. In contrast, the trigonal form showed a more pronounced response to changes in the RH, particularly above 90% RH. It should be noted that it is precisely near this RH value that the trigonal form undergoes a lattice transition from the trigonal form to a hexagonal form (probably $P622$; Figs. 2c and 3a).

An additional factor affecting the susceptibility of the $P3_121$ lattice to changes in the RH of the environment could be the association of the water in higher order aggregates. In particular, by applying a suitable pattern-search routine to molecular configurations of crystalline water, it is possible to analyse how it is organized in polygons (Stillinger, 1980). We

have analyzed the distribution of water polygons (see §2.2.6) and the results suggest that this is an additional factor that could contribute to the different response of the three crystalline forms upon changes in the RH of the environment.

In their initial analysis of 1500 structures, Lee & Kim (2009) observed that the most common water polygons in protein structures were triangles (43%) followed by quadrilaterals (12%) and pentagons (24%). This aggregation pattern is similar to that observed in the more elastic $P6_522$ form of PurE, in which 71% of the clusters are triangles and the
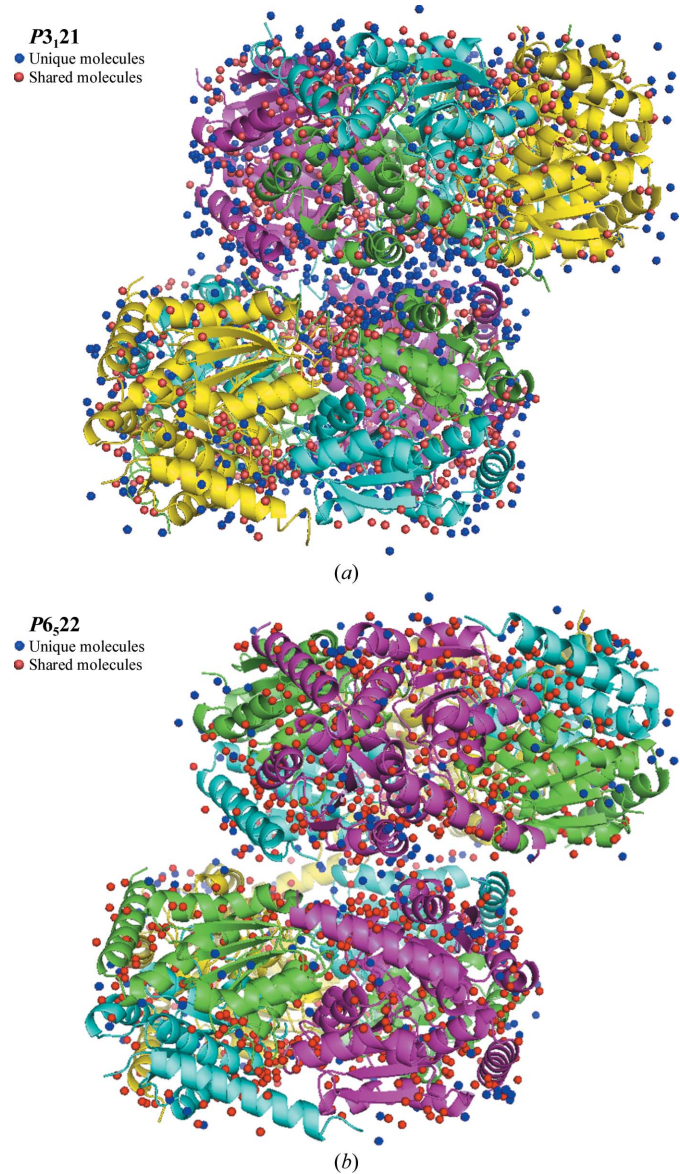
Figure 8
Comparison of the number and distribution of waters along the *c*-axis contacts in the hexagonal and trigonal forms of PurE. The figures show the hydration shells of two stacked octamers along the high-symmetry axes running upwards in the images. Subunits in the asymmetric unit are coloured from *A* to *D* and repeated by the crystallographic dyad. In (*a*) (trigonal) the upper octamer is related to the lower one by a 120° rotation and is shifted by *c*/3. In (*b*) (hexagonal) the upper octamer is related to the lower one by 60° and is shifted by *c*/6. Red spheres represent shared water molecules and blue spheres correspond to those that are unique to the space group.

remaining clusters are exclusively quadrilaterals (29%). The other two structures presented in this work, the $P3_121$ and $C2$ forms, have a distribution of clusters dominated by pentagons, as shown in Fig. 9. A possible explanation of the different clustering of water molecules in the crystals is the composition of the crystallization medium. The $P6_522$ crystals were obtained in the presence of PEG and salt, whereas the $P3_121$ and $C2$ crystals grew in the presence of sodium acetate and cacodylate (Table 1).

4. Summary of results

The systematic variation of the RH in three different crystal forms of PurE revealed a distinctly different behaviour in two different high-symmetry forms: $P6_522$ and $P3_121$. The crystals of the $P6_522$ form responded elastically to changes in the RH of the crystal environment in that they retained the crystallographic parameters (unit-cell parameters and space group) over a wide range of RH values (95–70%) and were able to sustain these properties during cyclic RH changes (Fig. 2a). In addition, the crystals retained the quality and extent of the diffraction pattern upon returning to the optimal RH (~85%). In contrast, a similar geometrical arrangement of PurE octamers in a $P3_121$ crystallographic lattice responded to similar changes in RH in an inelastic manner. The diffraction pattern of the trigonal form deteriorated irreversibly when subjected to similar RH changes (95–70%) and their quality of diffraction did not recover to the optimal value. In addition, under certain conditions the $P3_121$ lattice underwent a lattice transition to a hexagonal lattice with a doubled c-axis length that could be the $P6_522$ form. The $C2$ lattice also discussed is geometrically related to the other two by a distortion of the hexagonal lattice, which can also be indexed as an orthorhombic $C222$ cell, upon an expansion of the β angle from 90° to 98.3° (Fig. 6d). Overall, the $C2$ form was considered to be less amenable to systematic study of the effects of RH changes owing to the lower resolution diffraction and the large number of subunits in the asymmetric unit (12, equivalent to a full octamer and a tetramer).

In an attempt to find the factors that could contribute to the dramatically different responses to changes in the RH, a detailed analysis of the refined structures of the three different crystal forms was performed at the optimal RH and at cryogenic temperatures, including determination of (i) the three-dimensional structures of PurE at the tertiary and quaternary levels, (ii) the packing in the three different crystal forms and (iii) the solvent structure for both ordered and bulk solvent.

The static structure of the corresponding 422 octamers does not differ in any dramatic way from the octamers described in the quaternary structure of PurE from other species, including the previously determined structure of PurE from *B. anthracis*. However, analysis of the temperature factors (B factors; Fig. 1d) of the amino acids in the different subunits (A–D for the hexagonal and trigonal forms and A–D and E–L for the $C2$ form) and the distribution of large r.m.s.d. deviations among the different subunits gives credence to the suggestion that the structure of PurE is better described in terms of two flexible

domains on either side of a more rigid core (Fig. 10). The flexible domain at the N-terminus comprises residues Lys2–Leu58 ($\beta 1$–H1–$\beta 2$–H2) and is followed by a more rigid central core containing residues Lys59–Ser134 ($\beta 3$–H3–$\beta 4$–H4–$\beta 5$–H5). The smaller final flexible subdomain mainly consists of H6 and the final extension (Gly157-Ser158-Glu159-Leu160-Val161), which adopts different conformations in the various subunits.

The packing of the disc-like PurE octamer in the *ab* plane in the three crystal forms can approximately be described as hexagonal close packing of a square particle, with minor variations in the contacts along the high-symmetry axes in the hexagonal and trigonal cells. These contacts are always provided (although in slightly different side-chain conformations) by the flexible subdomains introduced above that turn out to be the most salient features of the PurE octameric discs. These protruding features provide three types of contacts referred to as parallel helical (involving H1), cross-helical (also involving H1 but in a different orientation) and C-terminal (involving the ordered helix H6). Small variations in the interfaces of these three contact areas in the trigonal and hexagonal lattices, including the conformation of the external side chains of H1 (Glu16, Tyr20 and Asp23) and H5 (Lys118) and the contacts with a few (1–3) water molecules, result in the differences between the $P3_121$ and $P6_522$ packing, including their different handedness.

Another structural element that is important in establishing the handedness of the packing arrangement in the two different space groups is the extent of ordering of the C-terminal flexible subdomain in the different crystal forms, particularly in the external subunits that make the contacts between the reference molecule and those above and below it. In the $P3_121$ packing, the internal subunits B and C are fully ordered and the external subunits are only ordered to Ser158 (subunit C) and Gly157 (subunit D). In contrast, in the $P6_522$ tetramer subunits B and D are fully ordered and subunits A and C are only ordered to Gly157. In the more complex
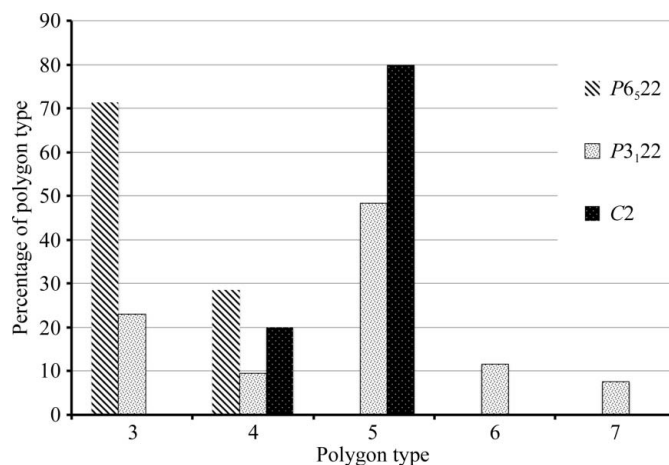

Figure 9
Histogram of polygon distribution. Histogram of the percentages of different types of polygons (3–7 fused) found in the three different crystals forms of PurE. The statistics were obtained using the program *Water Polygon Finder* (Lee & Kim, 2009) as described in §2.2.6.

packing arrangement in the *C*2 crystal, the first tetramer (subunits *A–D*) only has a fully ordered subunit *C*; in the other full octamer in the asymmetric unit only subunits *E* and *G* are fully ordered. The other subunits are only ordered to the end of helix H6 (Gly157 or Glu156). Incidentally, when the C-terminus is ordered the carboxy-terminus interacts with Arg41 in the corresponding subunit, affecting the conformation of this active-site residue and hindering the entrance of possible ligands or inhibitors. This could explain our failed attempts to soak ligands into the trigonal and hexagonal forms despite the 'dilation' of the two crystal lattices using changes in the RH of the crystal (Abad-Zapatero *et al.*, 2011).

The refined structures also provide important information regarding the solvation of the PurE octamers and the amount of solvent (both bulk and ordered) involved in the crystal contacts. For the purpose of the discussion of the different diffraction properties, we focused on the trigonal and hexagonal forms.
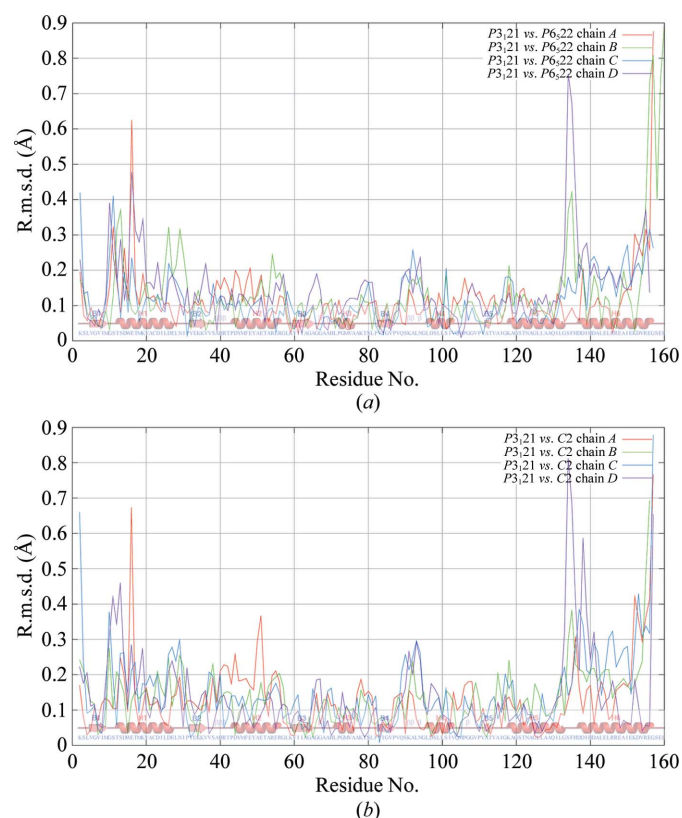

Figure 10
Tertiary-structure variations among different PurE crystal forms. Root-mean-square deviations (r.m.s.d.s) between C$^\alpha$ atoms for different subunits of PurE. (*a*) R.m.s.d. *versus* residue number between subunit *A* of the trigonal form and the corresponding subunits of the tetramer (*A–D*) in the asymmetric unit of the hexagonal form. (*b*) R.m.s.d. *versus* residue number between subunit *A* of the trigonal form and the corresponding subunits (*A–D*) of the half-octamer in the asymmetric unit of the *C*2 form. Note the clustered high r.m.s.d. values in the N-terminal flexible subdomain (residues Lys2–Leu58) and at the C-terminal end. Local large differences in the more rigid structural core occur at the β-turn (Lys91–Gly95) near the centre of the particle (fourfold axis) and prior to the C-terminal mobile subdomain at Phe135. A similar profile of large r.m.s.d. deviations occurs when comparing the structures of the different subunits in the *C*2 form.

Our detailed analysis of the solvent distribution in the trigonal and hexagonal packings revealed an important difference that cannot be related to the different resolutions of the two refined structures (1.76 *versus* 2.0 Å). The detailed comparison of the solvent structures, presented previously, revealed that there were many more unique waters that were involved in intermolecular contacts along the 3$_1$ screw axis in the *P*3$_1$21 form compared with the hexagonal form (80 *versus* 26; Figs. 8*a* and 8*b*).

5. Discussion and conclusions

The summary of the results outlined above suggests a possible explanation for the different behaviour of the two main lattices (hexagonal *versus* trigonal) upon changes in the RH of the crystal environment and could lead to the identification of possible causes of and factors participating in this effect.

We surmise that in the *P*6$_5$22 form the intermolecular (crystal) contacts are better defined (more rigid) owing to the smaller amount of water in the interface (Fig. 8*b*). The hexagonal lattice only expands by a certain amount owing to the limited number of water molecules, and when it is expanded by the additional hydration and contracted by the changes in RH the subunit contacts return to the same well defined state. Owing to the small number of water molecules in the interface, the disc-like aggregates can 'readily' find the previously existing contacts without any ambiguity and thus the lattice is restored to one of very few possibilities, guided by the parallel, cross-helical and C-terminal contacts (Fig. 4). All of these factors could have been facilitated by the presence of PEG in the crystallization medium, as has been suggested in other studies (Russi *et al.*, 2011). The resilience of the hexagonal lattice to the changes in RH could be indicated by the relatively robust response to the increase in the number of waters per amino-acid residue (Fig. 3*a*).

In the *P*3$_1$21 form, a detailed analysis of the unit-cell expansion, the solvent structure and the distribution of the solvent in terms of higher polygon aggregates points to a different behaviour of the crystal upon changes in the RH. The majority (80 *versus* 26) of the unique waters observed in this crystal packing are located in the inter-octamer space relating the reference molecule to those above and below it (Fig. 8*a*). The pattern of ordering of the C-terminal residues within the various subunits is different (Fig. 5) and it is conceivable that the C-terminus of subunit *B* could be disordered upon changes in the RH of the crystal and that the three residues of the C-terminus of subunit *D* could become ordered to make it more like the *P*6$_5$22 form. However, the transition from the *c*-axis dimension observed in the *P*3$_1$21 form (~132 Å) to one more like that in the hexagonal *P*622 form would not require a strict replication of the interactions. Minor alterations in the number of water molecules, the redistribution of some water molecules and/or the rearrangement of the C-terminal residues would suffice to induce the ~270 Å axis.

A full experimental strategy to prove or disprove the suggestions proposed above and to unambiguously identify the most likely factors and the order of occurrence in a

sequential cascade will require the data collection, structure determination and full refinement of structures of PurE crystals collected at the critical RH points identified above. Of particular interest, although it might prove to be particularly challenging, would be the collection of complete data sets at high resolution at the transition point between the trigonal and hexagonal lattices observed in the experiments described above.

References

Abad-Zapatero, C., Oliete, R., Rodriguez-Puente, S., Pous, J., Martinelli, L., Johnson, M. E. & Guasch, A. (2011). *Acta Cryst.* F**67**, 1300–1308.

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.

Bernal, J. D. & Crowfoot, D. (1934). *Nature (London)*, **133**, 794–795.

Berthou, J., Cesbron, F. & Laurent, A. (1972). *J. Mol. Biol.* **71**, 809–813.

Blevins, R. A. & Tulinsky, A. (1985). *J. Biol. Chem.* **260**, 8865–8872.

Böttcher, J., Jestel, A., Kiefersauer, R., Krapp, S., Nagel, S., Steinbacher, S. & Steuber, H. (2011). *Methods Enzymol.* **493**, 61–89.

Boyle, M. P., Kalliomaa, A. K., Levdikov, V., Blagova, E., Fogg, M. J., Brannigan, J. A., Wilson, K. S. & Wilkinson, A. J. (2005). *Proteins*, **61**, 674–676.

Carugo, O. & Bordo, D. (1999). *Acta Cryst.* D**55**, 479–483.

Chen, X., Weber, I. & Harrison, R. W. (2008). *J. Phys. Chem. B*, **112**, 12073–12080.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Greenspan, L. (1977). *J. Res. Natl Bur. Stand. A*, **81**, 89–96.

Hoskins, A. A., Morar, M., Kappock, T. J., Mathews, I. I., Zaugg, J. B., Barder, T. E., Peng, P., Okamoto, A., Ealick, S. E. & Stubbe, J. (2007). *Biochemistry*, **46**, 2842–2855.

Huxley, H. E. & Kendrew, J. C. (1953). *Acta Cryst.* **6**, 76–80.

Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.

Kiefersauer, R., Than, M. E., Dobbek, H., Gremer, L., Melero, M., Strobl, S., Dias, J. M., Soulimane, T. & Huber, R. (2000). *J. Appl. Cryst.* **33**, 1223–1230.

Kleywegt, G. J. (1996). *Acta Cryst.* D**52**, 842–857.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

Lee, J. & Kim, S.-H. (2009). *Protein Sci.* **18**, 1370–1376.

Madhusudan, Kodandapani, R. & Vijayan, M. (1993). *Acta Cryst.* D**49**, 234–245.

Mathews, I. I., Kappock, T. J., Stubbe, J. & Ealick, S. E. (1999). *Structure*, **7**, 1395–1406.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Nakasako, M. (2004). *Philos. Trans. R. Soc. Lond. B*, **359**, 1191–1204.

Otte, H. M. & Crocker, A. G. (1965). *Phys. Status Solidi*, **9**, 441–451.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Perutz, M. (1985). *Methods Enzymol.* **114**, 3–18.

Russi, S., Juers, D. H., Sanchez-Weatherby, J., Pellegrini, E., Mossou, E., Forsyth, V. T., Huet, J., Gobbo, A., Felisaz, F., Moya, R., McSweeney, S. M., Cusack, S., Cipriani, F. & Bowler, M. W. (2011). *J. Struct. Biol.* **175**, 236–243.

Samant, S., Lee, H., Ghassemi, M., Chen, J., Cook, J. L., Mankin, A. S. & Neyfakh, A. A. (2008). *PLoS Pathog.* **4**, e37.

Sanchez-Weatherby, J., Bowler, M. W., Huet, J., Gobbo, A., Felisaz, F., Lavault, B., Moya, R., Kadlec, J., Ravelli, R. B. G. & Cipriani, F. (2009). *Acta Cryst.* D**65**, 1237–1246.

Schwarzenbacher, R. et al. (2004). *Proteins*, **55**, 474–478.

Settembre, E. C., Chittuluru, J. R., Mill, C. P., Kappock, T. J. & Ealick, S. E. (2004). *Acta Cryst.* D**60**, 1753–1760.

Stillinger, F. H. (1980). *Science*, **209**, 451–457.

Thoden, J. B., Holden, H. M., Paritala, H. & Firestine, S. M. (2010). *Biochemistry*, **49**, 752–760.

Tranchimand, S., Starks, C. M., Mathews, I. I., Hockings, S. C. & Kappock, T. J. (2011). *Biochemistry*, **50**, 4623–4637.

Winn, M. D. et al. (2011). *Acta Cryst.* D**67**, 235–242.

Zhang, Y., Morar, M. & Ealick, S. E. (2008). *Cell. Mol. Life Sci.* **65**, 3699–3724.