Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Structural and functional insights into an archaeal L-asparaginase obtained through the linker-less assembly of constituent domains

Rachana Tomar,[a]‡ Pankaj Sharma,[b]‡ Ankit Srivastava,[a] Saurabh Bansal,[a]§ Ashish[b] and Bishwajit Kundu[a]*

[a]Kusuma School of Biological Sciences, Indian Institute of Technology Delhi, New Delhi, India, and [b]CSIR – Institute of Microbial Technology, Chandigarh, India

‡ These authors contributed equally.
§ Present address: JUIT, Wakhnaghat, Solan 173 215, India.

Correspondence e-mail: bkundu@bioschool.iitd.ac.in

Received 11 July 2014
Accepted 23 October 2014

PDB references: *P. furiosus* L-asparaginase, wild type, 4q0m; conjoined, apo, 4ra6; complex with L-aspartic acid, 4nje; complex with citrate, 4ra9

Covalent linkers bridging the domains of multidomain proteins are considered to be crucial for assembly and function. In this report, an exception in which the linker of a two-domain dimeric L-asparaginase from *Pyrococcus furiosus* (PfA) was found to be dispensable is presented. Domains of this enzyme assembled without the linker into a conjoined tetrameric form that exhibited higher activity than the parent enzyme. The global shape and quaternary structure of the conjoined PfA were also similar to the wild-type PfA, as observed by their solution scattering profiles and X-ray crystallographic data. Comparison of the crystal structures of substrate-bound and unbound enzymes revealed an altogether new active-site composition and mechanism of action. Thus, conjoined PfA is presented as a unique enzyme obtained through noncovalent, linker-less assembly of constituent domains that is stable enough to function efficiently at elevated temperatures.

1. Introduction

L-Asparaginases are clinically important enzymes that catalyze the conversion of L-asparagine to L-aspartic acid and ammonia (Broome, 1961; Offman *et al.*, 2011). Structural and functional information on L-asparaginases from different sources suggest that they vary in composition and act as dimers, tetramers (dimers of dimers) or hexamers (trimers of dimers) (Bansal *et al.*, 2010; Cedar & Schwartz, 1967; Lubkowski *et al.*, 1994; Pritsa & Kyriakidis, 2001). Bacterial L-asparaginases are categorized as type I and type II based on their cellular localization (Schwartz *et al.*, 1966; Campbell *et al.*, 1967). Irrespective of variations, a dimeric assembly constitutes the basic functional enzyme in all cases. A dimer formed by a 'head-to-tail' arrangement of two monomeric subunits holds two axially opposite active sites at the interface (Swain *et al.*, 1993). In effect, a tetrameric enzyme is equipped with four active sites and a hexameric enzyme with six active sites. Each active site further consists of two catalytic triads (I and II) to carry out the acylation and deacylation steps, respectively, that are necessary for the conversion of the substrate to product (Sanches *et al.*, 2007; Ortlund *et al.*, 2000).

Previously, the superimposition of a modelled structure of the hyperthermophilic *Pyrococcus furiosus* L-asparaginase (PfA) with the crystal structure of the mesophilic *Escherichia coli* L-asparaginase (EcAII) led us to design active-site mutants of PfA with enhanced substrate affinity and activity (Bansal *et al.*, 2012). From these studies, the flexibility of an active-site loop has been proposed to be an important factor

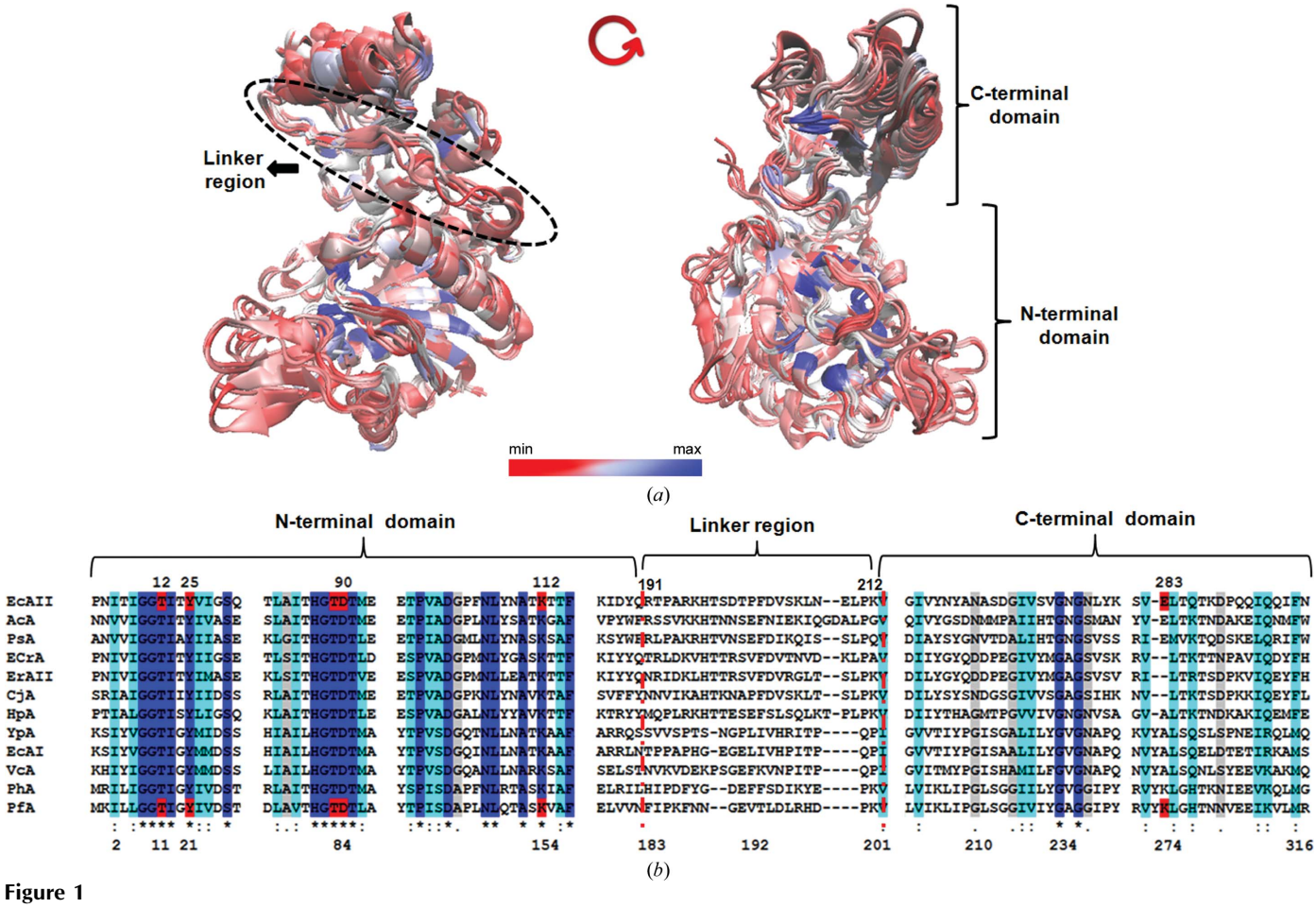

Figure 1
Sequence and structural comparison of bacterial and archaeal L-asparaginases. (*a*) The structural superimposition of the 12 available L-asparaginase structures is shown to highlight key conserved residues (blue) and nonconserved residues (red). The dashed ellipse shows the nonconserved linker region among all the structures. (*b*) In the sequence alignment, asterisks (sequences highlighted in blue) indicate fully conserved residues, colons (sequences highlighted in turquoise) indicate conservation between groups with strongly similar properties and points (sequences highlighted in grey) indicate conservation between groups with weakly similar properties. The active-site residues are highlighted in red for the PfA and EcAII sequences. The numbering corresponds to the EcAII (PDB entry 3eca) and PfA (model) structures. The linker boundary is shown as a dashed line (red). Only partial sequences are shown for brevity. Evidently, the most key conserved residues are found in the N-terminal domain and the least conserved in the C-terminal domain, supporting the structural alignment, while the linker region remains mostly nonconserved.

determining the catalytic efficiency of L-asparaginases (Yao *et al.*, 2005; Lubkowski *et al.*, 1996; Kozak *et al.*, 2000). PfA functions as a dimer, whereas EcAII and many other bacterial L-asparaginases function as tetramers, with each monomer composed of distinct N-and C-terminal domains connected by a linker (Bansal *et al.*, 2012; Swain *et al.*, 1993; Aghaiypour *et al.*, 2001*a*).

Linkers in multidomain proteins are considered to be indispensable as they help in tethering the domains and assist in communication between domains (Arviv & Levy, 2012; Bhaskara & Srinivasan, 2011; Gokhale & Khosla, 2000). Often, the linker length and sequence defines conformational flexibility for accommodating multiple domains and preventing non-native interactions between interfering domains (van Leeuwen *et al.*, 1997). In mesophiles and higher eukaryotes, the addition of linkers of variable lengths and sequence has been shown to improve protein stability, and in some cases provides added functionality (Robinson & Sauer, 1998). Multiple reports support the chaperone function of linkers (Chen *et al.*, 2012; Buske & Levin, 2013). Linker size in homologous proteins has also been correlated with evolution, with the shortest linkers being observed in archaeal proteins (Wang *et al.*, 2011).

We found that while the sequences of the N-terminal domains of L-asparaginases were highly conserved across species, their linkers were variable, with shorter linkers in the archaeal enzymes. In PfA, the linker was also found to be smaller and nonconserved compared with L-asparaginases of bacterial origin (Fig. 1, Supplementary Fig. S1 and Supplementary Table S1[1]). This, together with our earlier finding that the N-terminal domain of PfA is actually involved in the overall folding of the protein (Tomar *et al.*, 2013), led us to speculate that the linker of PfA may have a rather insignificant role. Thus, to evaluate whether the domains of PfA can function in isolation, or in a certain combination, without the

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: QH5016).

linker, we synthesized the N-terminal (NPfA) and C-terminal (CPfA) domains and studied their behaviour separately and in combination (Fig. 2). Here, we describe how these disjointed domains orient spatially to acquire enzymatic function.

2. Materials and methods

2.1. Alignment of sequence and structural data

The UniProt database (Bairoch *et al.*, 2005) was used to identify members of asparaginase family with structures deposited in the Protein Data Bank (PDB; Berman *et al.*, 2000). All structures retrieved from the PDB were aligned using *ClustalW*2 in the *MultiSeq* extension of *VMD* (Thompson *et al.*, 1994; Roberts *et al.*, 2006). The ECAII crystal structure (PDB entry 3eca) was chosen as a non-redundant structure for structural alignment using *STAMP* (Russell & Barton, 1992). Residues were coloured by similarity according to the BLOSUM60 matrix. The global similarities and differences of the PfA homology structure (Bansal *et al.*, 2012) compared with all other structures were quantified by calculating the r.m.s.d and homology ($Q_{\rm H}$) between the structures. Finally, only the 12 available bacterial and archaeal L-asparaginase structures with the highest homology ($Q_{\rm H}$ > 0.6) were considered for analysis (Supplementary Table S1). The following structures were aligned and analysed: PDB entries 2gvn, 3nxk, 3ntx, 2wt4, 1hfw, 3eca, 2p2d, 2ocd, 1wls, 1agx, 1djp and the PfA model. The partial and full sequence alignments were further visualized and represented using *Jalview* 2 (Waterhouse *et al.*, 2009). The sequence conservation was annotated using the Gonnet PAM 250 matrix defined in *ClustalW*2. The domain boundaries were ascertained based on this structural and sequence alignment along with Pfam multiple alignment corresponding to the asparaginase family (PF00710).

2.2. Cloning, expression and purification of proteins

For the cloning and expression of the N-terminal and C-terminal domains separately, a previously developed PfA clone was used as the template (Bansal *et al.*, 2012). Using the set of primer pairs (Supporting Information §1), PCR amplifications of the DNA sequences corresponding to NPfA and CPfA were performed. The PCR products were ligated separately in pET-28a vector (Novagen) using the *Nhe*I and *Bam*HI (New England Biolabs) sites, followed by transformation into *E. coli* DH5α cells and subsequently into the expression host *E. coli* Rosetta (DE3). Cultures grown in LB medium (HiMedia) containing 50 µg ml^{-1} kanamycin and 17 µg ml^{-1} chloramphenicol (Sigma) were induced with 1 m*M* IPTG (Sigma) at an A_{600} of 0.6 and were harvested 14 h post-induction. Cells were lysed by sonication followed by centrifugation. Expression was analyzed by 12% SDS–PAGE. A standard Ni–NTA affinity-based purification procedure under denaturing condition (Qiagen protocol) was followed to purify each domain. Purified fractions of each domain were pooled and subjected to refolding either independently or after mixing them in an equimolar mixture by dialysis at 4°C. The dialysis buffer used was 25 m*M* Tris, 50 m*M* NaCl at pH 8.0 for NPfA (pI 5.7), pH 9.0 for CPfA (pI 7.1) and pH 8.5 for the domain mixture. After dialysis, the protein samples were purified by passage through a Superdex 200 gel-filtration column attached to an ÄKTApurifier FPLC system (GE Healthcare). Purified proteins were stored in a freezer at −20°C until further use. The wild-type PfA was expressed and purified using a previously reported protocol (Bansal *et al.*, 2012).

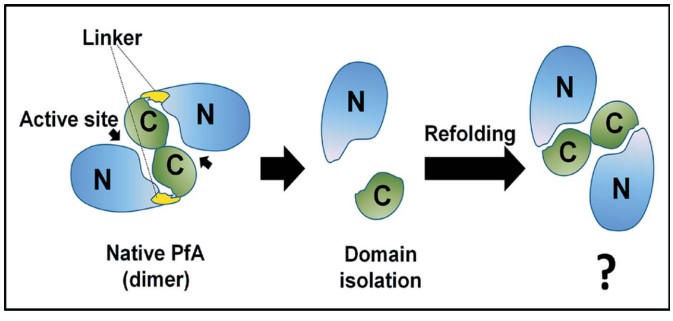

Figure 2
Schematic of possible domain assembly of PfA. Dissociated NPfA (blue) and CPfA (green) domains and their possible association without the linker (yellow) are shown.

2.3. Molecular mass and subunit association

To determine molecular association, a refolded mixture of NPfA, CPfA and wild-type proteins was analyzed on a Superdex 200 analytical gel-filtration column attached to an ÄKTApurifier FPLC system (GE Healthcare). Molecular mass and oligomeric nature were further confirmed by MALDI-TOF mass spectrometry (Bruker) and dynamic light scattering with a Zetasizer Nano ZS instrument (Malvern, UK), respectively.

2.4. Estimation of secondary structure

To determine secondary structure, each protein (0.2 mg ml^{-1}) in Tris–NaCl buffer was loaded into a 1 mm path-length quartz cuvette and the far-UV CD spectrum was measured from 250 to 200 nm in a spectropolarimeter (Jasco J-815) with a spectral bandwidth of 5 nm. An average of three scans was plotted against wavelength.

2.5. Activity assay

Activity was measured for the isolated domains, the wild-type and the conjoined PfA using a standard Nesslerization protocol (Mashburn & Wriston, 1963). The reaction was set up in buffer consisting of 200 µl 50 m*M* sodium phosphate pH 7.4 to which 200 µl 100 m*M* L-asparagine and 25 µl 4.2 µ*M* enzyme solution were added and the volume made up to 2 ml, followed by incubation for 10 min at 37°C. After incubation, the reaction was stopped by adding 100 µl 1.5 *M* trichloroacetic acid. The solution was centrifuged and the supernatant (500 µl) was diluted with water to 7 ml, to which 1 ml Nessler's reagent was added. The absorbance was measured at 480 nm to determine the enzymatic activity. One international unit

(IU) of L-asparaginase activity was defined as the amount of enzyme liberating 1 µmol ammonia in 1 min. Finally, the specific activity was defined in terms of units per milligram of protein.

2.6. Synchrotron SAXS data acquisition and processing

The SAXS data for the wild-type protein as well as for the conjoined PfA were collected using a charge-coupled detector (CCD) on the X9 beamline at the National Synchrotron Light Source, Brookhaven National Laboratory, Brookhaven, New York, USA. The beam wavelength and the ratio of the sample-to-detector distance to the diameter of the CCD were 0.873 Å and 20.8, respectively. To acquire the scattering data, 120 µl of wild-type and conjoined L-asparaginase at each of three different concentrations were used together with matched buffers. The exposure time for both the protein sample and the matched buffer was 120 s in a quartz flow cell at 15°C at a flow rate of 50 µl min^{-1}. The images recorded on the CCD were scaled, merged and circularly averaged using the Python script-based programs written by Dr Lin Yang (X9 beamline, National Synchrotron Light Source). The buffer contribution was subtracted to obtain the scattering intensity I as a function of the momentum-transfer vector q ($q = 4\pi\sin\theta/\lambda$, where λ and θ represent the wavelength of the X-rays and the scattering angle, respectively). All of the SAXS experiments described in this study were carried out in duplicate. Guinier and indirect Fourier transformation analysis were carried out using *PRIMUS* (Konarev *et al.*, 2003) and *GNOM* (Semenyuk & Svergun, 1991) as available in the *ATSAS*2.1 suite of programs (Konarev *et al.*, 2006). Structure reconstructions were carried out using *DAMMINIQ* (Svergun, 1999) and were averaged using the *DAMAVER* suite of programs (Volkov & Svergun, 2003).

2.7. Protein crystallization, data collection and structure refinement

To set up crystallization screens, freshly purified wild-type and conjoined L-asparaginase were used. The proteins were concentrated to 10 mg ml^{-1}. Initial crystallization trials were performed using the crystallization screening kits Crystal Screen, Index (Hampton Research), Structure Screen (Molecular Dimensions) and Wizard Screen (Emerald Bio). Diffraction-quality crystals of conjoined L-asparaginase without (apo form) and with substrate were obtained in three different conditions by the vapour-diffusion method at 20°C from hanging drops composed of 1 µl protein solution and 1 µl reservoir solution after approximately 3–7 d. Crystals of conjoined L-asparaginase (apo form) were obtained in two conditions: (i) 15%(*v*/*v*) reagent alcohol, 100 m*M* imidazole–HCl pH 7.5, 200 m*M* $MgCl_2$ and (ii) 0.2 *M* sodium citrate tribasic dihydrate, 0.1 *M* sodium cacodylate, 30%(*v*/*v*) 2-propanol at pH 6.0. Crystals of conjoined L-asparaginase with substrate were obtained in wells containing 0.2 *M* sodium citrate tribasic dihydrate, 0.1 *M* sodium cacodylate, 30%(*v*/*v*) 2-propanol pH 6.0, 150 m*M* L-asparagine. In each case, crystals were grown for a week and were then used to collect diffraction data. Attempts to obtain diffraction-quality crystals of the wild-type protein initially failed, but diffraction-quality crystals formed after two months in a hanging-drop crystallization setup. The reservoir solution in this case was 0.2 *M* ammonium dihydrogen phosphate, 0.1 *M* Tris pH 9.0, 50% MPD.

Diffraction data for conjoined as well as wild-type crystals were collected on an in-house MAR 345 dtb image-plate detector mounted on a Bruker AXS MICROSTAR-H or a Rigaku MicroMax-007 HF rotating-anode X-ray generator (λ = 1.5418 Å) operated at 40 kV and 30 mA. For cryoprotection, crystals were soaked in 20% ethylene glycol and 20% glycerol added to the corresponding mother liquor and were subsequently flash-cooled in liquid nitrogen at 100 K using an Oxford cryostream. Subsequently, diffraction data were collected from all of the crystals at the same temperature. The crystal-to-detector distance was kept at 200 mm for the apo, substrate-bound conjoined and wild-type crystals, while a crystal-to-detector distance of 175 mm was used for the crystals with citrate. Each frame was recorded for 10 min with 1° oscillation during the recording of each image for all crystals. Diffraction data processing including intensity integration and scaling was performed using *MOSFLM* and the *HKL*-2000 suite (Battye *et al.*, 2011; Otwinowski & Minor, 1997).

Initial structure determination of apo conjoined L-asparaginase was performed by the molecular-replacement method with *Phaser* (McCoy *et al.*, 2007) from the *CCP*4 suite (Winn *et al.*, 2011) using the *P. horikoshii* (PhA) structure (Yao *et al.*, 2005; PDB entry 1wls) as a search model. The solved structure of apo conjoined L-asparaginase was further used as a search model to solve the structures of the other conjoined as well as the wild-type PfA L-asparaginase. The number of chains in the asymmetric unit was determined using *MATTHEWS_COEF* (Kantardjieff & Rupp, 2003) from the *CCP*4 suite. Conjoined PfA was present as a dimer in space group $P4_1$, while conjoined PfA with substrate (L-aspartic acid) and with citrate ion were present as a monomer in the asymmetric unit in space group $P6_522$. Wild-type PfA was also present as a monomer in the asymmetric unit in space group *H*32. The initial models of all crystals were refined by rigid-body refinement using *REFMAC*5 (Murshudov *et al.*, 2011) followed by restrained refinement. Further refinement was performed by multiple rounds of manual inspection using *Coot* (Emsley & Cowtan, 2004) and *PHENIX* (Adams *et al.*, 2010) until the models were completely built. The addition of solvent molecules present in the solution began at the stage at which R_{work} reached around 0.25. Molecules were added to electron densities where the $F_o - F_c$ map was more than 3σ above the mean and the $2F_o - F_c$ map showed density at the 1σ level forming at least one hydrogen bond to a protein atom or another solvent atom. *PROCHECK* (Laskowski *et al.*, 1993) was used as a validation tool to assess the quality of all of the final refined models. Despite the different crystallization conditions and space groups, the biological assembly of both the wild-type and the reconstituted L-asparaginase appeared to be a dimeric form as validated using the EBI *PISA* server (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html).

3. Results and discussion

3.1. A functional conjoined molecule was reconstituted from co-refolding of PfA domains

The sequence alignment and structural superposition of different bacterial and archaeal L-asparaginases showed that while the N-terminal domains were highly conserved, the linkers were nonconserved and variable (Fig. 1 and Supplementary Fig. S1). Interestingly, in the archaeal enzymes (PfA and PhA) the linker was found to be relatively small, with a sequence differing significantly from many other bacterial asparaginases (Fig. 1, Supplementary Fig. S1 and Supplementary Table S1). One important role of the linker defined recently is to function as a chaperone and help the parent protein to fold (Chen *et al.*, 2012; Buske & Levin, 2013). Since in PfA this folding assistance is reportedly provided by the N-terminal domain, the specific role of a small and non-conserved linker in the parent protein was questioned (Tomar *et al.*, 2013). To investigate this, we purified the NPfA and CPfA domains separately devoid of linker by refolding them from inclusion bodies formed inside *E. coli* expression hosts. While the NPfA domain appeared as large soluble oligomers, the CPfA domain readily formed aggregates (Tomar *et al.*, 2013). Interestingly, when co-refolded, an equimolar mixture of NPfA and CPfA resulted in a soluble species. We investigated whether the soluble species was a conjoined entity of both the domains and whether it had enzymatic function (Fig. 2). We first characterized the soluble, linker-less species using size-exclusion chromatography, in which the co-refolded domains mainly eluted as a single peak around 13.8 ml (Fig. 3*a*). This confirmed the physical interaction between the domains. Furthermore, the resemblance of the elution profile to that of the wild-type protein indicated the acquisition of a size similar to that of the wild type by the domain-assembled species (Fig. 3*a*). Using the standard curve, the molecular weight of the reconstituted enzyme was found

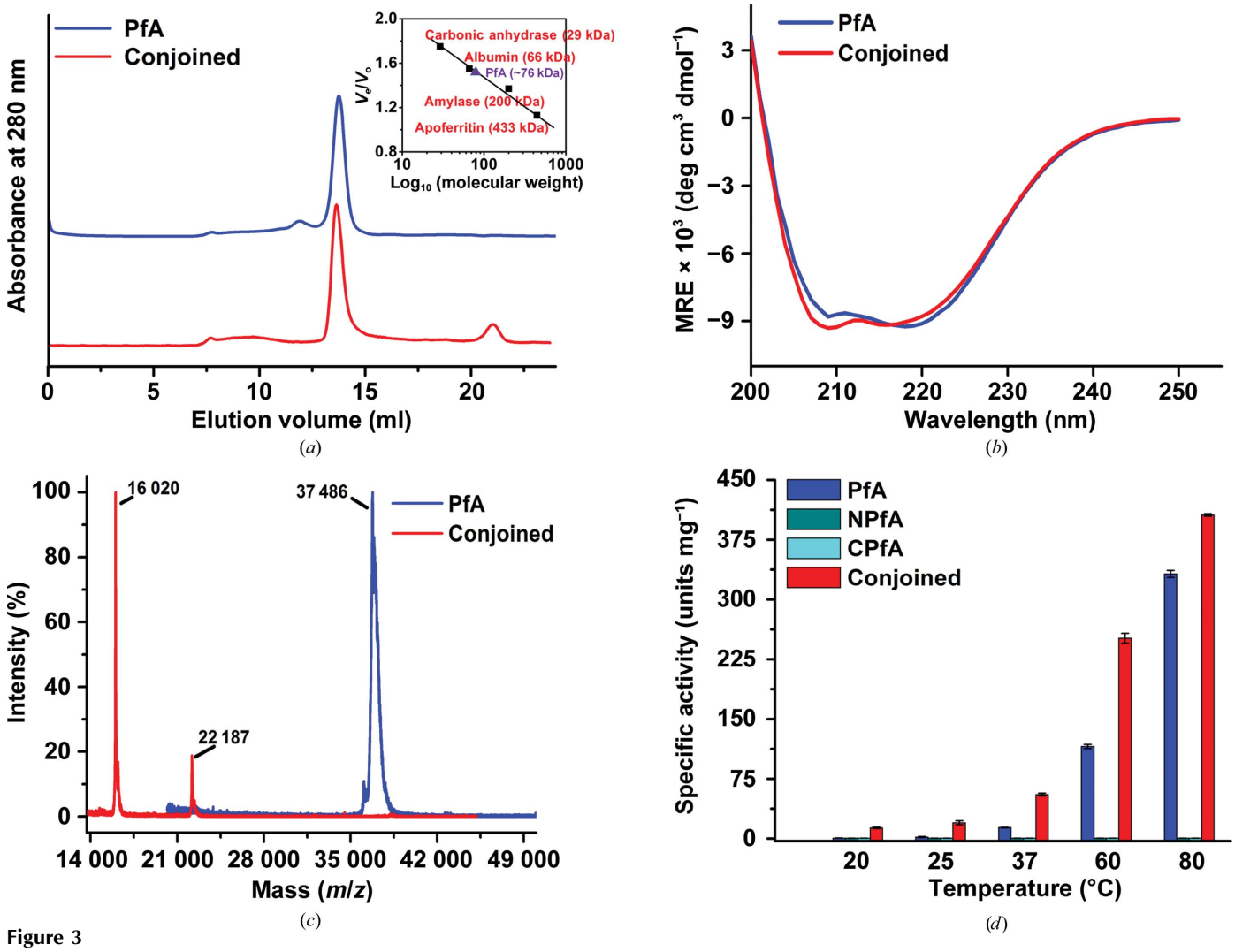

Figure 3
Comparison of molecular mass, secondary structure and activity of wild-type and conjoined PfA. (*a*) Elution profiles of proteins, showing overlapping chromatograms. The inset shows the standard curve for a Superdex 200 10/300 GL column. (*b*) The far-UV CD spectra showing similarity in secondary structure. (*c*) MALDI-MS data showing a mass of ∼37 kDa for wild-type PfA (blue) and conjoined PfA (red) displaying distinct peaks for the N-terminal (∼22 kDa) and C-terminal (∼16 kDa) domains, summing to ∼37 kDa. (*d*) Specific activity of the wild-type and conjoined enzymes compared with the isolated domains. The conjoined molecule displayed higher specific activity at all experimental temperatures.

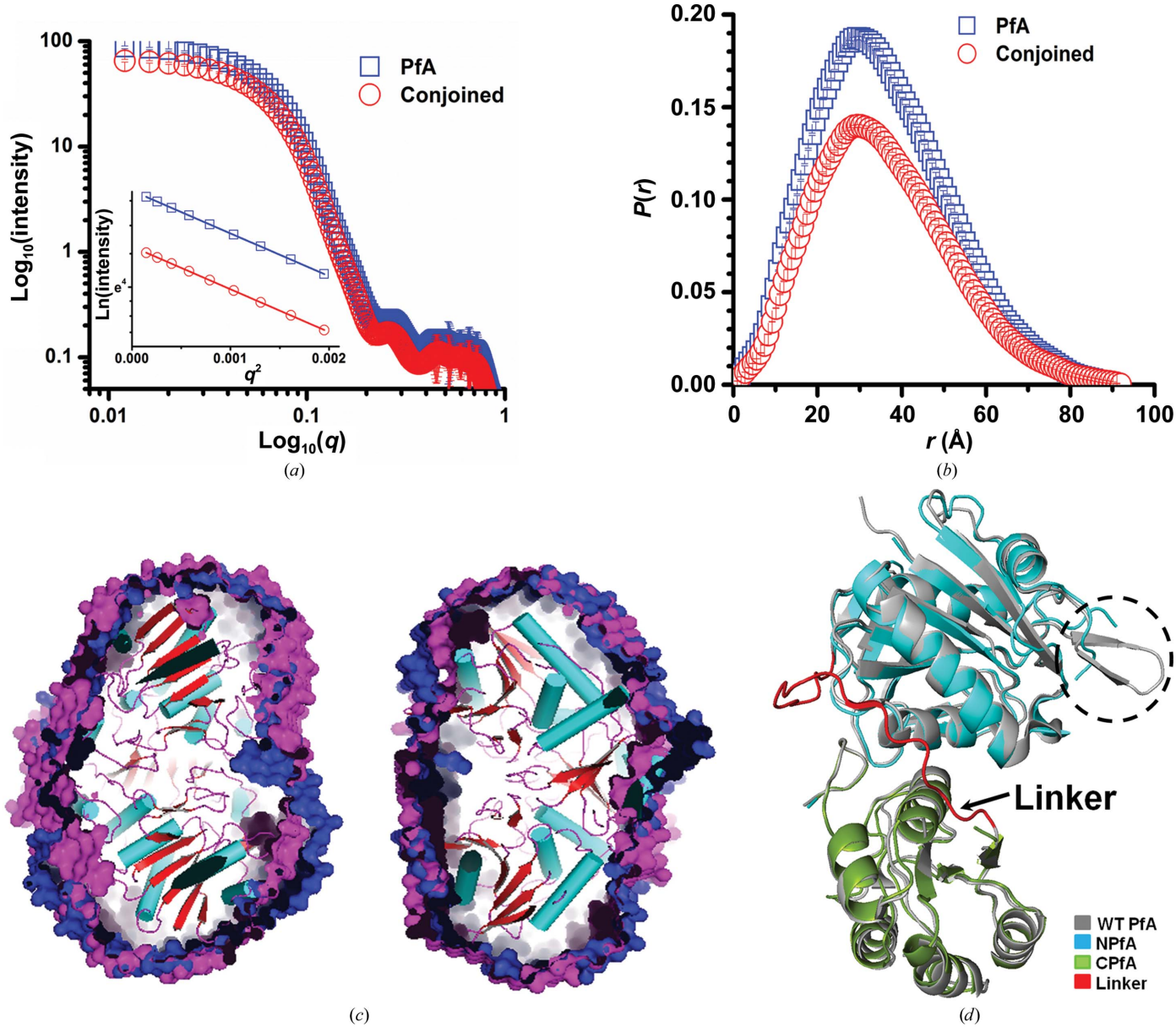

Figure 4
Structural comparison of the reconstituted conjoined molecule with the wild-type enzyme. (*a*) SAXS data profile shown as a log–log plot of intensity *versus* *q* for wild-type and conjoined proteins, showing a linear relationship. (*b*) Indirect Fourier transformation of the SAXS data sets for the wild-type (squares) and conjoined (circles) proteins. (*c*) SAXS data-based global shape of the wild-type (blue) and conjoined (magenta) enzymes, confirming similar profiles to the crystal structure of PhA (PDB entry 1wls). (*d*) Crystal structure of conjoined PfA (PDB entry 4ra6), in which NPfA (blue) and CPfA (green) superimposed with the corresponding domains of wild-type PfA (PDB entry 4q0m; grey) except for the inter-domain linker (red). The active-site loop is highlighted with a dashed circle. Note that the loop region is unstructured in conjoined PfA (apo form) as inferred from the absence of electron density in this region.

to be 76 kDa, which was almost equivalent to that of the wild-type protein (75.45 kDa; Fig. 3*a*, inset). Likewise, comparison of the dynamic light-scattering profiles verified that the two proteins have almost similar hydrodynamic radii, with the conjoined protein being slightly larger than the wild-type protein (Supplementary Fig. S2). Analysis of the CD data showed nearly matching secondary-structural contents of the two proteins, with a slight variation which may be owing to small structural changes in the loop region (Fig. 3*b*). This was also verified later from their respective crystal structures (Fig. 3*b* and Supplementary Table S2). The appearance of both NPfA and CPfA in the MALDI-TOF spectrum of the domain-associated molecule suggested that the entities are present in a conjoined tetrameric (2:2) assembly similar to the wild-type enzyme (Fig. 3*c*). Moreover, while each domain was inactive independently, the soluble conjoined PfA was endowed with catalytic activity (Fig. 3*d*). Interestingly, the specific activity of conjoined PfA was higher than that of the wild-type enzyme at all experimental temperatures (Fig. 3*d*). On an increase in temperature, an increment in specific activity was observed.

Table 1
Data-collection statistics for the crystals of wild-type PfA and linker-less (conjoined) PfA in the absence and presence of ligands.

Values in parentheses are for the highest resolution shell. All crystals were grown at 293 K and the wavelength of the X-ray beam used for data collection was 1.5418 Å.

		Conjoined PfA		
	Wild-type PfA	Apo	With L-aspartic acid	With citrate
Data collection				
Resolution (Å)	50–2.23	50–2.50	50–2.50	50–2.05
Space group	$H32$	$P4_1$	$P6_522$	$P6_522$
Unique reflections	19525	19946	16068	29876
Unit-cell parameters				
a (Å)	116.38	61.43	91.59	90.7
b (Å)	116.38	61.43	91.59	90.7
c (Å)	153.94	156.13	188.77	190.0
α (°)	90	90	90	90
β (°)	90	90	90	90
γ (°)	120	90	120	120
Completeness (%)	98.1 (82.2)	95.6 (76.6)	94.7 (99.7)	99.8 (97.5)
R_{merge}	0.11 (0.43)	0.12 (0.76)	0.13 (0.71)	0.09 (0.80)
Multiplicity	7.3 (6.6)	4.6 (3.7)	13.4 (13.7)	14.5 (9.2)
Average $I/\sigma(I)$	10.7 (4.01)	13.3 (1.5)	13.5 (3.9)	39.0 (2.0)
Refinement				
R_{work} (%)	19.0	18.9	22.84	19.7
R_{free} (%)	24.4	25.9	27.01	22.7
Solvent content (%)	55.9	35.9	59.7	58.0
Chains	A, B	A, B, P, Q	A, B	A, B
R.m.s.d. from ideality				
Bonds (Å)	0.008	0.008	0.009	0.008
Angles (°)	1.054	1.171	1.234	1.067
Wilson B factor (Å^2)	23.58	44.5	42.78	37.3
Ramachandran plot statistics (%)				
Most favoured	90.4	87.8	89.0	93.2
Allowed region	9.2	11.6	10.2	6.4
Generously allowed	0.4	0.6	0.8	0.4
Disallowed	0	0	0	0
PDB code	4q0m	4ra6	4nje	4ra9

This hinted towards sufficiently tight association of the PfA domains even in the absence of linker to retain activity and resist thermal denaturation.

3.2. Solution scattering studies showed similar molecular alignment of the conjoined and wild-type enzymes

To gain further insights into the global shape and quaternary structure of the conjoined protein, SAXS data profiles from solutions of both of the proteins were analysed. A log–log plot of intensity *versus* q showed the monodisperse nature of the scattering species in the samples (Fig. 4*a*). Further support was obtained from indirect Fourier transformation of the SAXS data (q range of 0.009–0.5 Å^{-1}), which showed that the predominant solution shapes of the wild-type and conjoined enzymes are very similar and can be characterized by a D_{max} of 89 and 92 Å and an R_g of 27.3 and 27.5 Å, respectively (Fig. 4*b*). Inertial axes alignment of the dummy-residue models of wild-type and conjoined PfA on the crystal structure of the homologous protein from *P. horikoshii* (PhA; PDB entry 1wls; Yao *et al.*, 2005) confirmed that both of the PfA proteins adopted a dimeric association, as seen in the case of PhA (Fig. 4*c*). *CRYSOL* calculated a χ^2 value of 1.7 and 2.1 between the calculated SAXS profile of the crystal structure (PDB entry 1wls) and the experimental SAXS data acquired for the wild-type and conjoined proteins, respectively. Similarity in the SAXS data-based parameters of the two systems upheld that the global shapes of the intact wild-type and conjoined species are almost identical.

3.3. The structural integrity of the conjoined enzyme is maintained through domain–domain interactions rather than the covalent linker

To determine the forces responsible for maintaining the structural integrity, we obtained crystals of the conjoined PfA. After refinement (parameters are listed in Table 1), we concluded that the conjoined protein (PDB entry 4ra6) acquired a similar structure to that of wild-type PfA (PDB entry 4q0m) (Fig. 4*d*). The two isolated domains, *i.e.* NPfA and CPfA, in the conjoined protein were oriented in a tetrameric conformation similar to that of the parent protein. The structure of linker-containing PfA aligned with the linker-less conjoined PfA with a low r.m.s.d (0.49 Å), suggesting that the linker does not play a role in the assembly of the domains (Fig. 4*d*). For a typical multi-domain multimeric protein, *in vitro* folding is an onerous task, given that each domain has to independently fold and then undergo recognition to assume the correct molecular assembly. In the latter stage, the finding of a cognate partner by the individual domains is possibly assisted by the presence of covalent linker(s) which keep the right partners tethered together. Therefore, the conjoined PfA presents a unique example of folding and assembly in which the domains not only folded but also recognized their cognate partners without being guided or assisted by a covalent linker.

Interaction analysis using the EBI *PDBSUM* server (http://www.ebi.ac.uk/pdbsum; Laskowski, 2009) revealed that a large number of hydrophobic interactions are mainly responsible for stabilizing and maintaining the spatial orientation of the domains, even in the absence of a linker (Supplementary Fig. S3). Additionally, a few interfacial hydrogen bonds are also present, contributing further to the stability. Additionally, in the wild-type protein a linker appeared to hinder the mobility of the domains with respect to each other, while in the case of the conjoined protein we assume that the domains are free to reorient spatially in a conformation which is more accessible to the substrate. Thus, the reconstituted enzyme has a lower energy barrier to properly orient the subunits and active-site residues at the subunit interface to be more active. This also explains the temperature-dependent difference at lower temperatures, where we observed a much higher activity ratio

of the reconstituted (conjoined) enzyme compared with the wild-type enzyme (Fig. 3*d*). From this study, we speculate that in the pro-archaeal stages the association of domains evolved independently of linkers. Evolutionarily, the hostile archaeal habitats possibly enforced the domains to acquire linkers so that they remain tethered together and worked in close association without drifting away. However, the linker came with a penalty in the form of a mild inhibitory effect on activity owing to the extra polypeptide length, probably posing steric hindrance for the substrate to enter the enzyme active site. In mesophiles, the linkers progressively increased in size and flexibility to accommodate more domains and subunits, thereby increasing the number of active sites per molecule. This hypothesis can be tested on other multidomain archaeal proteins in order to understand the significance of the linkers.

3.4. Conversion of an unstructured loop into a rigid β-hairpin upon substrate binding indicates a 'post-substrate entry' gatekeeping mechanism

To gain insights into the active site, crystals were obtained both in the absence and the presence of the substrate L-asparagine. Since we obtained crystals in the product (L-aspartate)-bound state, we further obtained crystals in the presence of citrate, a nonhydrolyzable substrate mimic. In the crystal structure of the apo L-asparaginase (PDB entry 4ra6), mainly electron density of the active-site loop residues was missing, in addition to a few residues from the other subunit (Supplementary Fig. S4*a*). This involved residues Val15–Tyr21 from the N-terminal domain of one subunit and residues Arg272′–Val275′ from the C-terminal domain of the other subunit, encompassing the aforementioned regions. This disordered loop region is analogous those in EcAII and other type II asparaginases (PDB entry 3eca; Fig. 4*d*, Supplementary Figs. S4*a* and S5), clearly showing its highly flexible nature (Swain *et al.*, 1993; Aung *et al.*, 2000).

This could also be the conformation of the loop in the wild-type enzyme in a substrate-free form. However, the crystal structure of wild-type PfA (PDB entry 4q0m) obtained with phosphate ion displayed definite electron density at the loop region, indicating a rigid loop conformation analogous to that reported in a homologous L-asparaginase from *P. horikoshii* (PhA; PDB entry 1wls) in the unliganded form (Fig. 4*d* and Supplementary Fig. S6*c*; Yao *et al.*, 2005). The structures of the L-aspartate-bound (PDB entry 4nje) and the citrate-bound (PDB entry 4ra9) conjoined forms also exhibited a rigid β-hairpin loop (Fig. 5 and Supplementary Figs. S4*b*, S4*c*, S5, S6*a* and S6*b*). It can thus be concluded that the disordered loop acts as a gatekeeper which becomes rigid when the active site is occupied. The substrate-induced active-site rearrangement is similar to that observed for EcA II (Swain *et al.*, 1993; Aung *et al.*, 2000; Michalska & Jaskolski, 2006), but the conversion of an unstructured loop into a β-hairpin is a novel finding. Interestingly, an interaction between Glu22 and

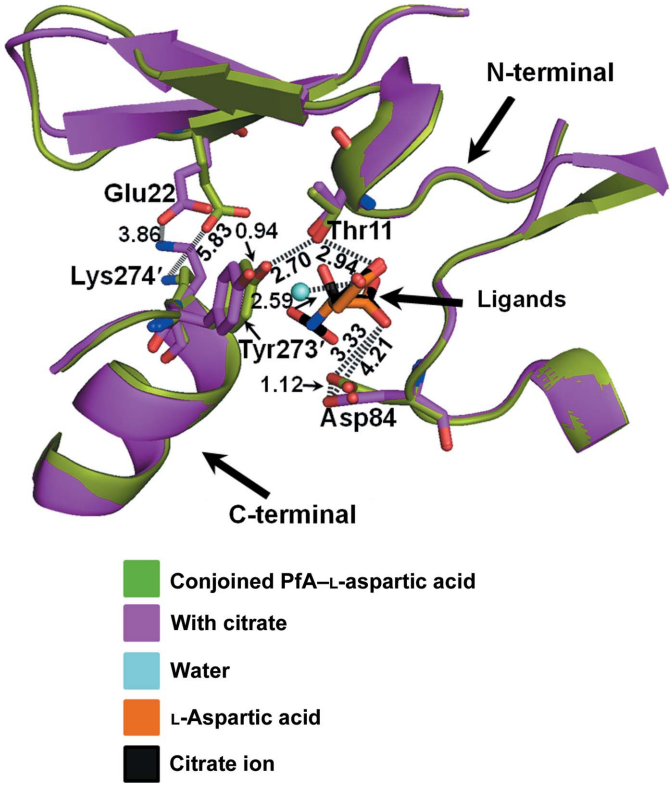

Figure 5
Active-site composition. Superimposition of the active-site residues of conjoined PfA with L-aspartic acid (green; PDB entry 4nje) and with citrate (magenta; PDB entry 4ra9) in which L-aspartic acid is shown in orange and citrate in black. A conserved water molecule was also observed in the crystal structure (cyan).

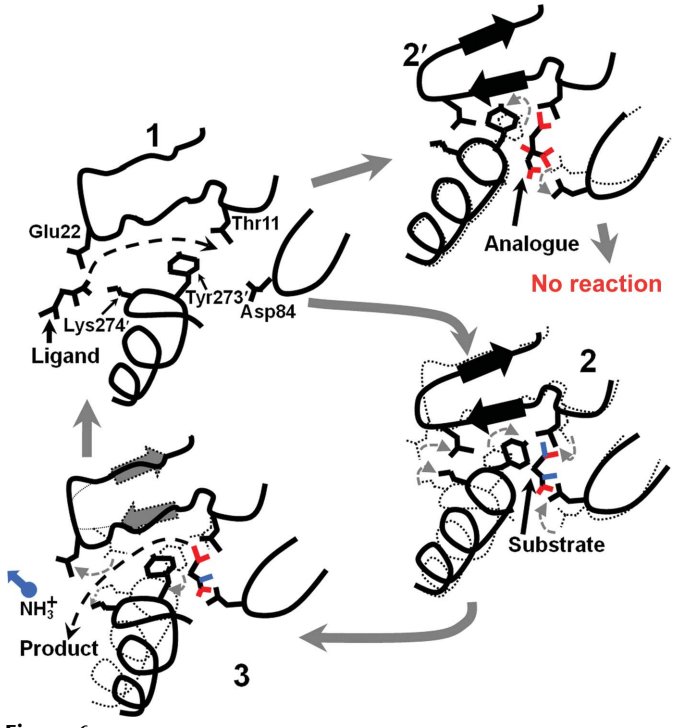

Figure 6
Schematic representation of the reaction mechanism. Conformational changes at the active site during (1) the introduction of ligand, (2) and (2′) the entry of substrate and analogue, respectively, and (3) product formation. Note that the loop acquires a stable β-hairpin conformation upon ligand binding.

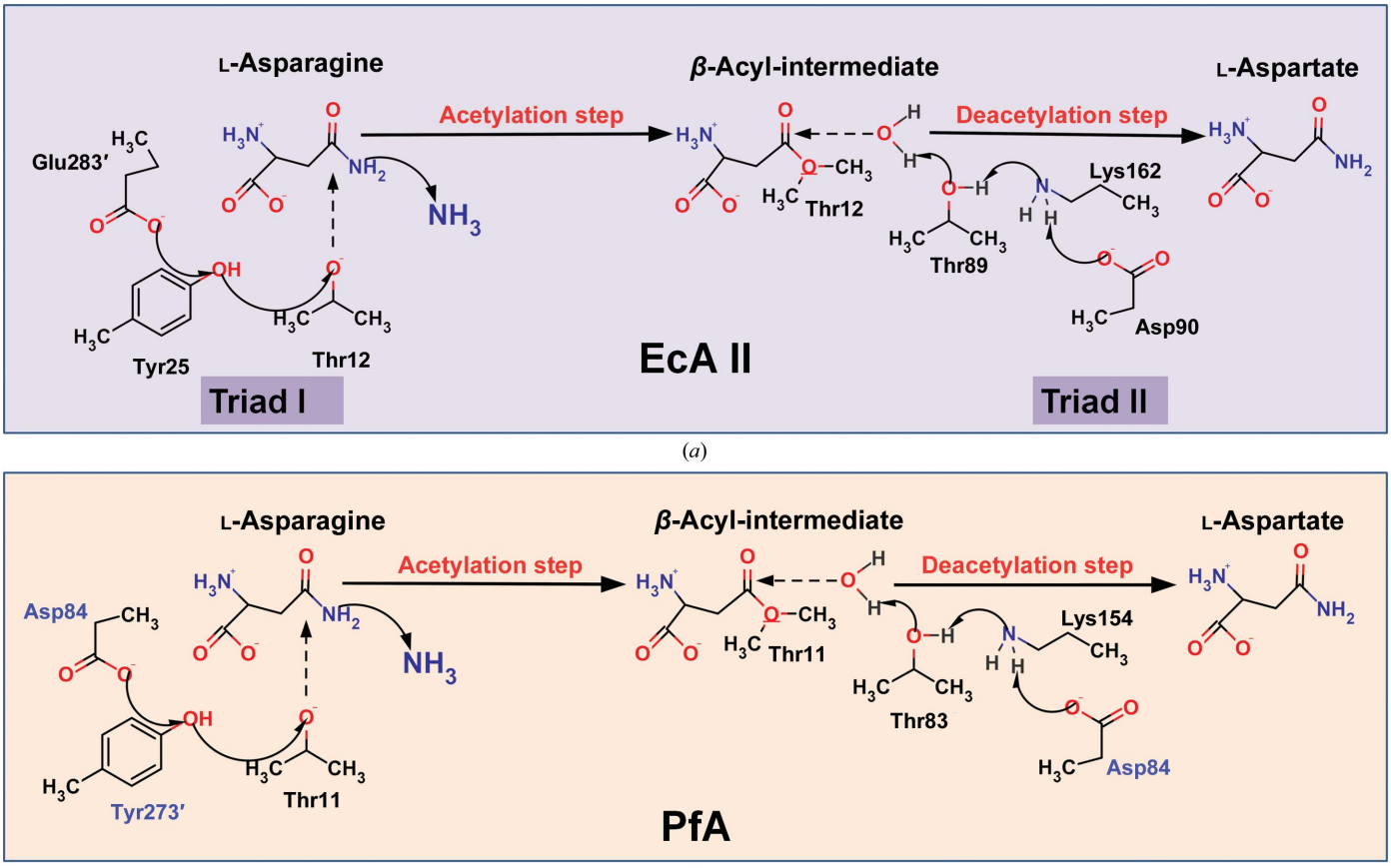

Figure 7
Reaction mechanism at the EcAII (*a*) and PfA (*b*) catalytic triads. A common acidic residue Asp84 is present in both triads I and II of PfA, with Tyr273′ in triad I as the basic residue.

Lys274′, which is known to constitute an important 'pre-substrate entry' gatekeeping mechanism in PhA (Yao *et al.*, 2005), did not appear in our structures. In contrast, our structures support a 'post-substrate entry' gatekeeping in which the loop reorients and acquires a rigid conformation. The loop lid closes down after substrate entry, and in the case of the nonhydrolyzable ligand citrate remained permanently blocked to further substrate entry, halting enzyme turnover. Asparagine (the natural hydrolyzable substrate) brings about a similar closing of the loop lid after its entry, recruiting all of the active-site residues for catalysis. At the same time, the closed loop helps to prevent the entry of any other ligand before completion of the reaction in a relatively closed environment, thus acting as a gatekeeper. However, once the substrate has been hydrolyzed, a local variation in charge ensues, resulting in ejection of the product by reopening of the lid. The subsequent relaxed, disordered state attained by the active-site loop becomes ready for another cycle of accepting substrate. This mechanism is shown schematically in Fig. 6.

3.5. The rearranged active-site residues in conjoined PfA suggest a novel reaction mechanism

Our structures also provide an insight into the catalytic mechanism of PfA, which differs from the accepted and experimentally proven mechanism of bacterial L-asparaginases (Figs. 7*a* and 7*b*; Sanches *et al.*, 2007; Kozak *et al.*, 2000; Swain *et al.*, 1993; Palm *et al.*, 1996; Miller *et al.*, 1993; Ehrman *et al.*, 1971). Irrespective of their quaternary structures, in bacterial L-asparaginases the C-terminal and N-terminal domains of each monomeric subunit are engaged with the other monomer, forming axially opposite active sites in a head-to-tail arrangement (Swain *et al.*, 1993). Each active site consists of two catalytic triads (I and II). While triad I acylates the substrate (L-asparagine) to form a *β*-aspartyl enzymatic intermediate, triad II deacylates the intermediate in the presence of a water molecule to release L-aspartic acid and ammonia as products. As per the current concept, each of these catalytic triads functions through an independent acid–base–nucleophile machinery in a concerted manner. For example, triad I in EcAII consists of Glu283′ (acid), Tyr25 (base) and Thr12 (nucleophile), while triad II consists of Asp90 (acid), Lys162 (base) and Thr89 (nucleophile) (Sanches *et al.*, 2007). The Thr89 residue in the second catalytic triad activates a water molecule that acts as the second nucleophile in the deacylation step (Fig. 7*a*). The catalytic triad II residues were found to be conserved amongst bacterial L-asparaginases. In the case of catalytic triad I in PfA, while two of the residues found to interact with substrate were Thr11 and Asp84 from NPfA, the third residue Tyr273′ was from CPfA

(Figs. 5 and 7*b* and Supplementary Fig. S6A). This was an unusual finding, as in other bacterial enzymes this third residue (acting as a base) is contributed by the N-terminal domain (Tyr21 in PhA and Tyr25 in EcAII; Sanches *et al.*, 2007; Yao *et al.*, 2005; Kozak *et al.*, 2000; Aghaiypour *et al.*, 2001*b*). Here, the notable thing is that this basic residue in PfA is not located on a flexible loop as was observed in the cases of PhA and EcAII. Thus, in the presence of substrate the spatial orientation of Tyr273′ and Asp84 in PfA ensured positional conjugation to activate the Thr11 nucleophile, initiating acylation to produce the β-aspartyl intermediate. In the citrate-bound structure, however, only Thr11 showed interaction with citrate, whereas Asp84 and Tyr273′ moved about 1 Å away compared with the product-bound state, thus halting any chance event of catalytic conversion (Figs. 5 and 6 and Supplementary Fig. 6*b*). This suggests that the accurate positioning of active-site residues is a specific substrate-induced phenomenon which cannot be reproduced faithfully by other ligands such as citrate (see the schematic in Fig. 6). The deacylation step in the presence of an activated water molecule at triad II consisting of Asp84, Lys154 and Thr83, however, is similar to bacterial L-asparaginases (Figs. 7*a* and 7*b*).

3.6. Comparison of conjoined PfA with other type I and type II L-asparaginases

Although PfA shares 22, 33 and 66% sequence similarity with EcAII, EcAI and PhA, respectively, their overall folded topologies were similar. Structural comparison using the *PyMOL* visualization tool (DeLano, 2002) demonstrated that the r.m.s.d.s of the corresponding C^{α} atoms between conjoined PfA and the other asparaginases were in the range 0.48–1.65 Å. In spite of this, some crucial differences amongst these enzymes were notable. When the sequences were compared, Tyr21 of PfA was found to align with Tyr21 of PhA, Tyr25 of EcAII and Tyr24 of EcAI. While in the EcAII crystal structure (PDB entry 3eca) Tyr25 is oriented towards the core of the active site, constituting a key basic residue for catalysis, Tyr21 in both of the archaeal enzymes (PhA and PfA) was found to be oriented away from the active site (PDB entries 1wls and 4nje) (Supplementary Figs. S7*a* and 7*b*). Interestingly, the equivalent residue Tyr24 of EcAI (PDB entry 2p2d) was also found to be oriented away like those of the archaeal enzymes (Supplementary Fig. 7*c*). This indicates that the basic function of abstracting protons from the acidic residue of the catalytic triad is performed by an altogether different residue in those enzymes in which this Tyr residue is oppositely oriented. This was indeed the case, as amongst the residues of triad I of PfA the basic residue Tyr273′ came from another subunit, possibly acting as a suitable substitute for Tyr21. Tyr273′ was also found to align similarly in the catalytic plane of PhA (Supplementary Fig. S7*b*). However, in the absence of supporting biochemical data and the crystal structure of substrate-bound PhA, no function has previously been assigned to this residue (Yao *et al.*, 2005). In the EcAI sequence, Tyr273′, although present, neither appeared in the electron-density map of the catalytic site nor was any apparent function assigned (PDB entry 2p2d; Yun *et al.*, 2007; Supplementary Fig. S7*c*). In contrast to this, in the EcAII structure the Tyr273′ residue was absent (Supplementary Fig. S7*a*).

The monomeric PfA structure resembled that of *E. coli* AnsA (type I), which shares 33% sequence identity. The major difference is that AnsA functions as a tetramer rather than a dimer (Yun *et al.*, 2007). With the dimeric interface of PfA showing very little similarity to the tetrameric interface of AnsA, the key interfacial residues of AnsA responsible for allosteric reorganization of the enzyme were not found in PfA (Yun *et al.*, 2007). In fact, superposition of the two enzyme structures revealed that this region deviates greatly, in spite of secondary-structural similarities. The dimeric PfA structure also suggests that this enzyme would not be allosteric because of three reasons. Firstly, unlike in AnsA, which involves movement of the entire tetramer, in PfA no gross quaternary-structural changes were observed between the substrate-bound and unbound forms except for loop rigidification (Yun *et al.*, 2007). Secondly, in spite of structural similarity between PhA and PfA in a putative allosteric region near helix α8, we did not find additional electron density for substrate at this site in our substrate-bound PfA structure (PDB entry 4nje). Thus, we believe that since the available PhA structure is of the apo form, the proposed allosteric binding region may be purely speculative based on structural alignment with EcAI. Thirdly, and most importantly, a typical sigmoidal curve characteristic of allosteric enzyme kinetics was not displayed by PfA (Bansal *et al.*, 2012).

We conclude that upon substrate entry the loop closes with concomitant positioning of Tyr273′ in close proximity to Thr11 and Asp84, resulting in a functional catalytic triad I. Hence, the reaction proceeds through an 'in-and-out' dynamic flipping of the Tyr273′ and Asp84 residues, reciprocating between closed and open gate conformations depending on the presence or absence of substrate, respectively (see the schematic in Fig. 6). This confirmed our previous prediction of a dynamic flipping mechanism of the Tyr273′ residue based on modelling and MD simulations (Bansal *et al.*, 2012). Conclusive evidence of the importance of Tyr273′ came from experiments performed on a Tyr273′-to-Ala mutant, which showed a drastic reduction in enzyme activity (Bansal *et al.*, 2012). Together, the fusion of catalytic triads I and II involving a common acid moiety (Asp84) represents a novel mechanism of action in PfA involving an amalgamated pentad orchestrating the conversion of L-asparagine into L-aspartic acid (Figs. 7*a* and 7*b*). This might be evidence of a reductive evolution of the archaeal machinery.

The conjoined PfA is an example in which isolated domains of a multidomain protein undergo molecular recognition to form a structure with matching physical properties to those of its parent protein. Furthermore, the perfectly oriented active-site residues in the linker-less PfA raise questions about the evolutionary importance of linkers in enzymes of this class.

RT and PS acknowledge CSIR and AS acknowledges ICMR, Government of India for research fellowships. Ashish

and BK acknowledge IMTECH Chandigarh and IIT Delhi, respectively, for infrastructural and financial support. Use of the National Synchrotron Light Source, Brookhaven National Laboratory was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-98CH10886. Dr S. Singh and Mr S. P. Yadav are acknowledged for their help at various stages of the work. The authors acknowledge Dr S. Karthikeyan for his help in data analysis. Professor B. Jayaram and Professor J. Gomes are acknowledged for reading the manuscript critically and giving valuable comments.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.

Aghaiypour, K., Wlodawer, A. & Lubkowski, J. (2001*a*). *Biochemistry*, **40**, 5655–5664.

Aghaiypour, K., Wlodawer, A. & Lubkowski, J. (2001*b*). *Biochim. Biophys. Acta*, **1550**, 117–128.

Arviv, O. & Levy, Y. (2012). *Proteins*, **80**, 2780–2798.

Aung, H.-P., Bocola, M., Schleper, S. & Röhm, K.-H. (2000). *Biochim. Biophys. Acta*, **1481**, 349–359.

Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N. & Yeh, L.-S. L. (2005). *Nucleic Acids Res.* **33**, D154–D159.

Bansal, S., Gnaneswari, D., Mishra, P. & Kundu, B. (2010). *Biochemistry*, **75**, 375–381.

Bansal, S., Srivastava, A., Mukherjee, G., Pandey, R., Verma, A. K., Mishra, P. & Kundu, B. (2012). *FASEB J.* **26**, 1161–1171.

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

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Bhaskara, R. M. & Srinivasan, N. (2011). *Sci. Rep.* **1**, 40.

Broome, J. D. (1961). *Nature (London)*, **191**, 1114–1115.

Buske, P. J. & Levin, P. A. (2013). *Mol. Microbiol.* **89**, 249–263.

Campbell, H. A., Mashburn, L. T., Boyse, E. A. & Old, L. J. (1967). *Biochemistry*, **6**, 721–730.

Cedar, H. & Schwartz, J. H. (1967). *J. Biol. Chem.* **242**, 3753–3755.

Chen, Z., Chen, X.-J., Xia, M., He, H.-W., Wang, S., Liu, H., Gong, H. & Yan, Y.-B. (2012). *Biophys. J.* **103**, 558–566.

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

Ehrman, M., Cedar, H. & Schwartz, J. H. (1971). *J. Biol. Chem.* **246**, 88–94.

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

Gokhale, R. S. & Khosla, C. (2000). *Curr. Opin. Chem. Biol.* **4**, 22–27.

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

Konarev, P. V., Petoukhov, M. V., Volkov, V. V. & Svergun, D. I. (2006). *J. Appl. Cryst.* **39**, 277–286.

Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 1277–1282.

Kozak, M., Jaskólski, M. & Röhm, K.-H. (2000). *Acta Biochim. Pol.* **47**, 807–814.

Laskowski, R. A. (2009). *Nucleic Acids Res.* **37**, D355–D359.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Leeuwen, H. C. van, Strating, M. J., Rensen, M., de Laat, W. & van der Vliet, P. C. (1997). *EMBO J.* **16**, 2043–2053.

Lubkowski, J., Palm, G. J., Gilliland, G. L., Derst, C., Röhm, K.-H. & Wlodawer, A. (1996). *Eur. J. Biochem.* **241**, 201–207.

Lubkowski, J., Wlodawer, A., Ammon, H. L., Copeland, T. D. & Swain, A. L. (1994). *Biochemistry*, **33**, 10257–10265.

Mashburn, L. T. & Wriston, J. C. Jr (1963). *Biochem. Biophys. Res. Commun.* **12**, 50–55.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Michalska, K. & Jaskolski, M. (2006). *Acta Biochim. Pol.* **53**, 627–640.

Miller, M., Rao, J. K., Wlodawer, A. & Gribskov, M. R. (1993). *FEBS Lett.* **328**, 275–279.

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.

Offman, M. N., Krol, M., Patel, N., Krishnan, S., Liu, J., Saha, V. & Bates, P. A. (2011). *Blood*, **117**, 1614–1621.

Ortlund, E., Lacount, M. W., Lewinski, K. & Lebioda, L. (2000). *Biochemistry*, **39**, 1199–1204.

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

Palm, G. J., Lubkowski, J., Derst, C., Schleper, S., Röhm, K.-H. & Wlodawer, A. (1996). *FEBS Lett.* **390**, 211–216.

Pritsa, A. A. & Kyriakidis, D. A. (2001). *Mol. Cell. Biochem.* **216**, 93–101.

Roberts, E., Eargle, J., Wright, D. & Luthey-Schulten, Z. (2006). *BMC Bioinformatics*, **7**, 382.

Robinson, C. R. & Sauer, R. T. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 5929–5934.

Russell, R. B. & Barton, G. J. (1992). *Proteins*, **14**, 309–323.

Sanches, M., Krauchenco, S. & Polikarpov, I. (2007). *Curr. Chem. Biol.* **1**, 75–86.

Schwartz, J. H., Reeves, J. Y. & Broome, J. D. (1966). *Proc. Natl Acad. Sci. USA*, **56**, 1516–1519.

Semenyuk, A. V. & Svergun, D. I. (1991). *J. Appl. Cryst.* **24**, 537–540.

Svergun, D. I. (1999). *Biophys. J.* **76**, 2879–2886.

Swain, A. L., Jaskólski, M., Housset, D., Rao, J. K. & Wlodawer, A. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 1474–1478.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.

Tomar, R., Garg, D. K., Mishra, R., Thakur, A. K. & Kundu, B. (2013). *FEBS J.* **280**, 2688–2699.

Volkov, V. V. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 860–864.

Wang, M., Kurland, C. G. & Caetano-Anollés, G. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 11954–11958.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. (2009). *Bioinformatics*, **25**, 1189–1191.

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

Yao, M., Yasutake, Y., Morita, H. & Tanaka, I. (2005). *Acta Cryst.* D**61**, 294–301.

Yun, M. K., Nourse, A., White, S. W., Rock, C. O. & Heath, R. J. (2007). *J. Mol. Biol.* **369**, 794–811.