Detection of the protein dimers, multiple monomeric states and hydrated forms of *Plasmodium falciparum* triosephosphate isomerase in the gas phase

Suman S. Thakur[1], P.D. Deepalakshmi[2], P. Gayathri[1],
Mousumi Banerjee[1], M.R.N. Murthy[1] and P. Balaram[1,3]

[1]Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012
and [2]Mass Spectrometry Facility, National Centre for Biological Sciences-
TIFR, GKVK PO, Bellary Road, Bangalore 560 065, India

[3]To whom correspondence should be addressed.
E-mail: pb@mbu.iisc.ernet.in

Dimeric and monomeric forms of the enzyme triosepho-sphate isomerase (TIM) from *Plasmodium falciparum* (*Pf*) have been detected under conditions of nanoflow by electrospray mass spectrometry. The dimer ($M = 55\ 663$ Da) exhibits a narrow charge state distribution with intense peaks limited to values of 18^+ to 21^+, maximal intensity being observed for charge states 19^+ and 20^+. A monomeric species with a charge state distribution ranging from 11^+ to 16^+ is also observed, which may be assigned to folded dissociated subunits. Complete dimer dissociation results under normal electrospray condition. The effects of solution pH and source temperature have been investigated. The observation of four distinct charge state distributions which may be assigned to a dimer, folded monomer, partially folded monomer and unfolded monomer is reported. Circular dichromism and fluorescence studies of *Pf* TIM at low pH support the retention of substantial secondary and tertiary structures. Satellite peaks in mass spectra corresponding to hydrated species are also observed and isotope shift upon deuteration is demonstrated. The analysis of all available independent crystal structures of *Pf* TIM and TIMs from other organisms permits identification of structurally conserved water molecules. Hydration observed in the dimer and folded monomeric forms in the gas phase may correspond to these conserved sites.
Keywords: gas phase conformation/gas phase non-covalent interactions/mass spectrometry/protein hydration/triosephosphate isomerase

Introduction

Electrospray ionisation mass spectrometry (ESI-MS) is rapidly emerging as a powerful method for direct characterisation of the constituents of non-covalent complexes in the gas phase (Benesch and Robinson, 2006). Multiply charged protein molecules are inserted into the mass spectrometer under highly energetic conditions in a normal electrospray experiment, resulting in both subunit dissociation and dehydration. The resultant mass measurement provides a highly accurate determination of protein molecular weight. If ionisation and injection into the mass spectrometer are achieved under much milder conditions, the determination of the masses of protein–protein complexes and protein–ligand adducts is achievable. Rapid

developments in the technology for nanospray methods and for the control of the internal energies imparted to molecules upon injection have dramatically expanded the range of non-covalent interactions that can be characterised directly by using mass spectrometry (Daniel *et al.*, 2002, 2003; Wendt *et al.*, 2003; Takats *et al.*, 2004; Kitova *et al.*, 2005; Huang *et al.*, 2005; Sharon and Robinsonm, 2007). The observation of the 70S subunit of ribosomes is a *tour de force* in this area (McKay *et al.*, 2006). It has become possible even with commercially available mass spectrometers to vary the conditions of electrospray to detect non-covalent complexes.
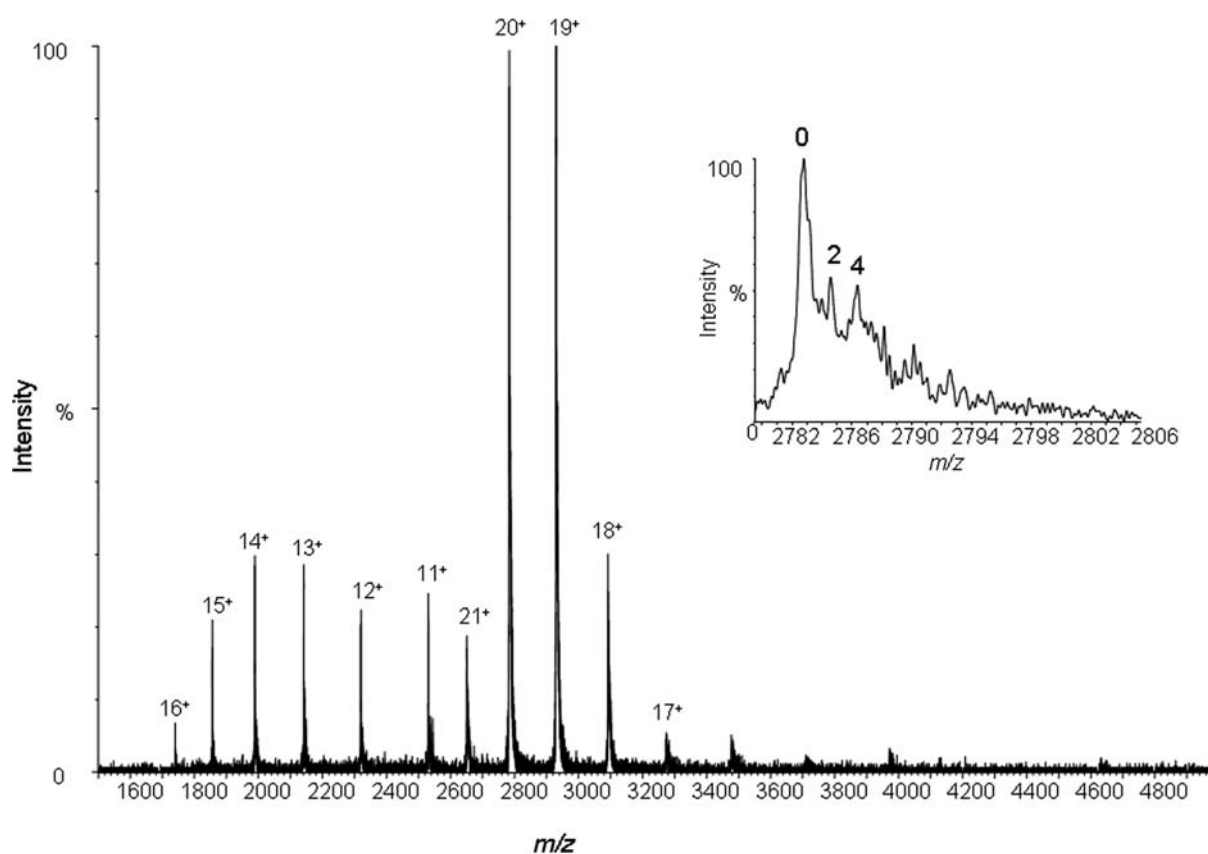
We have attempted to examine *Plasmodium falciparum* triosephosphate isomerase (*Pf* TIM) under very soft ionisation conditions in order to characterise dimeric and folded monomeric states and to estimate the number of water molecules that remain bound to both the monomeric and the dimeric forms of the protein. Nanoflow procedures are particularly useful for preservation of non-covalent interactions. Experimental parameters are optimised to preserve protein–protein contacts, which are usually destroyed under normal conditions of electrospray. The results of ESI-MS studies are often interpreted in terms of the structures that exist in the solutions that are sprayed, suggesting retention of memory of the solution state conformation in the gas phase (Loo, 1997; Jurchen and Williams, 2003).

Obligatory homo-oligomeric proteins are generally characterised by very high values of inter-subunit association constants (Darnall and Klotz, 1975), with the result that monomeric subunits are undetectable in solution, except under strongly denaturing conditions. The analysis of the unfolding reaction of multimeric proteins involves a dissection of the competing processes of subunit dissociation and chain unfolding. In situations where subunit dissociation precedes chain unfolding, folded monomeric units should, in principle, be detected. Characterisation of such isolated, folded subunits is generally difficult in solution using standard biophysical techniques. Protein engineering procedures have been used to design mutant sequences which can exist as stable monomeric forms in solutions. For example, monomeric forms of the well-studied enzyme TIM have been reported using a deletion mutant in which seven residues from a protruding loop at the dimer interface have been excised (Borchert *et al.*, 1994). TIM is the prototype $\alpha_8\beta_8$ barrel, an extremely robust structure, which retains appreciable structure even in 8 M urea (Gokhale *et al.*, 1999).

In this report, we describe gas phase detection of four forms of the dimeric, glycolytic enzyme TIM from *Pf* (*Pf* TIM), which are assigned to the dimer, a folded monomer, a partially unfolded state and a largely unfolded monomeric species. In addition, we demonstrate the observation of water molecules tightly bound to all the four forms.

Table I. TIMs structure used for identification of bound water molecule

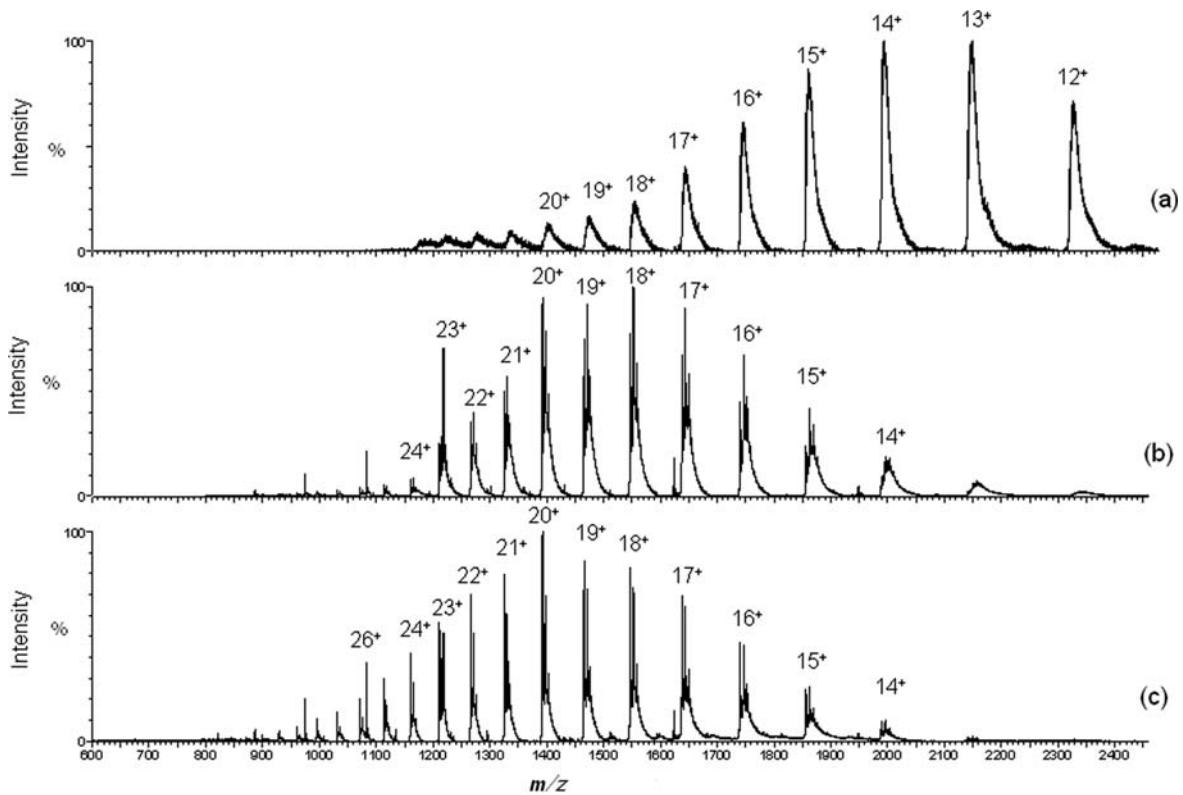| Sl. no	PDB code	Species	References
1	1LYX	*Plasmodium falciparum* (Malaria parasite)	Parthasarathy *et al.* (2002a)
2	1LZO	*Plasmodium falciparum* (Malaria parasite)	Parthasarathy *et al.* (2002a)
3	1M7O	*Plasmodium falciparum* (Malaria parasite)	Parthasarathy *et al.* (2002b)
4	1M7P	*Plasmodium falciparum* (Malaria parasite)	Parthasarathy *et al.* (2002b)
5	1O5X	*Plasmodium falciparum* (Malaria parasite)	Parthasarathy *et al.* (2003)
6	1VGA	*Plasmodium falciparum* (Malaria parasite)	Eaazhisai *et al.* (2004)
7	1WOA	*Plasmodium falciparum* (Malaria parasite)	Eaazhisai *et al.* (2004)
8	1WOB	*Plasmodium falciparum* (Malaria parasite)	Eaazhisai *et al.* (2004)
9	1YDV	*Plasmodium falciparum* (Malaria parasite)	Velanker *et al.* (1997)
10	2VFD	*Plasmodium falciparum* (Malaria parasite)	Unpublished (doi:10.2210/pdb2vfd/pdb)
11	2VFE	*Plasmodium falciparum* (Malaria parasite)	Unpublished (doi:10.2210/pdb2vfe/pdb)
12	2VFF	*Plasmodium falciparum* (Malaria parasite)	Unpublished (doi:10.2210/pdb2vff/pdb)
13	2VFG	*Plasmodium falciparum* (Malaria parasite)	Unpublished (doi:10.2210/pdb2vfg/pdb)
14	1AMK	*Leishmania mexicana*	Williams *et al.*(1999)
15	1TCD	*Trypanosoma cruzi*	Maldonado *et al.* (1998)
16	5TIM	*Trypanosoma brucei brucei*	Wierenga *et al.* (1991)
17	1WYI	*Homo sapiens* (Human)	Kinoshita *et al.* (2005)
18	1YPI	*Saccharomyces cerevisiae* (Yeast)	Lolis *et al.* (1990)
19	1R2R	*Oryctolagus cuniculus* (Rabbit)	Aparicio *et al.* (2003)
20	1TIM	*Gallus gallus* (Chicken)	Banner *et al.* (1976)
21	1MO0	*Caenorhabditis elegans* (Worm)	Symersky *et al.* (2003)
22	1M6J	*Entamoeba histolytica* (Amoeba)	Rodriguez-Romero *et al.* (2002)
23	1TRE	*Escherichia coli*	Noble *et al.* (1993)
24	1AWI	*Homo sapiens* (Human)	Mahoney *et al.* (1997)
25	1YYA	*Thermus thermophilus* (Bacteria)	Yamamoto *et al.* unpublished (doi:10.2210/pdb1yya/pdb)
26	1BTM	*Bacillus stearothermophilus*	Delboni *et al.* (1995)
27	1B9B	*Thermotoga maritima*	Maes *et al.* (1999)
28	1WOM	*Bacillus subtilis*	Kaneko *et al.* (2005)
29	1HG3	*Pyrococcus woesei*	Walden *et al.* (2001)
30	2H6R	*Methanocaldococcus jannaschii*	Gayathri *et al.* (2007)

Fig. 1. Nano-ESI-MS spectrum of *Pf* TIM. The charge states are indicated. Inset: Expansion of 20^+ charge state.

An analysis of bound water molecules in *Pf* TIM crystal structures is presented. The number of structurally conserved water molecules observed in multiple crystal structures of the enzyme correlates well with the number of protein-bound water molecules detectable in the gas phase.
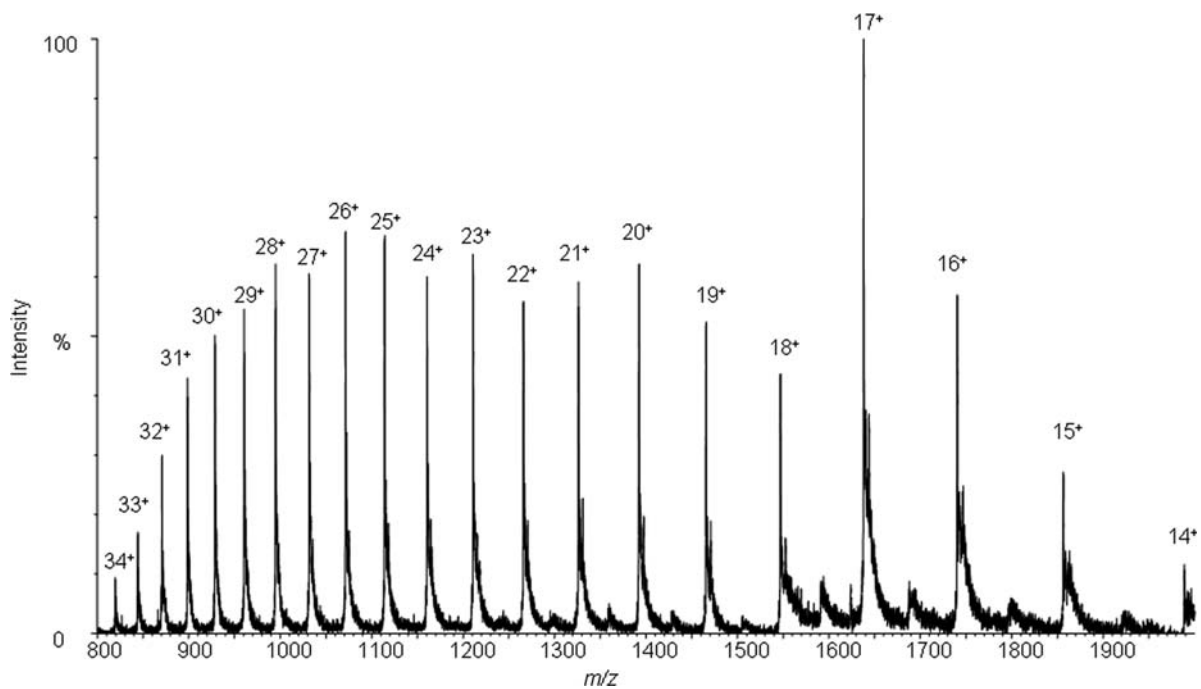
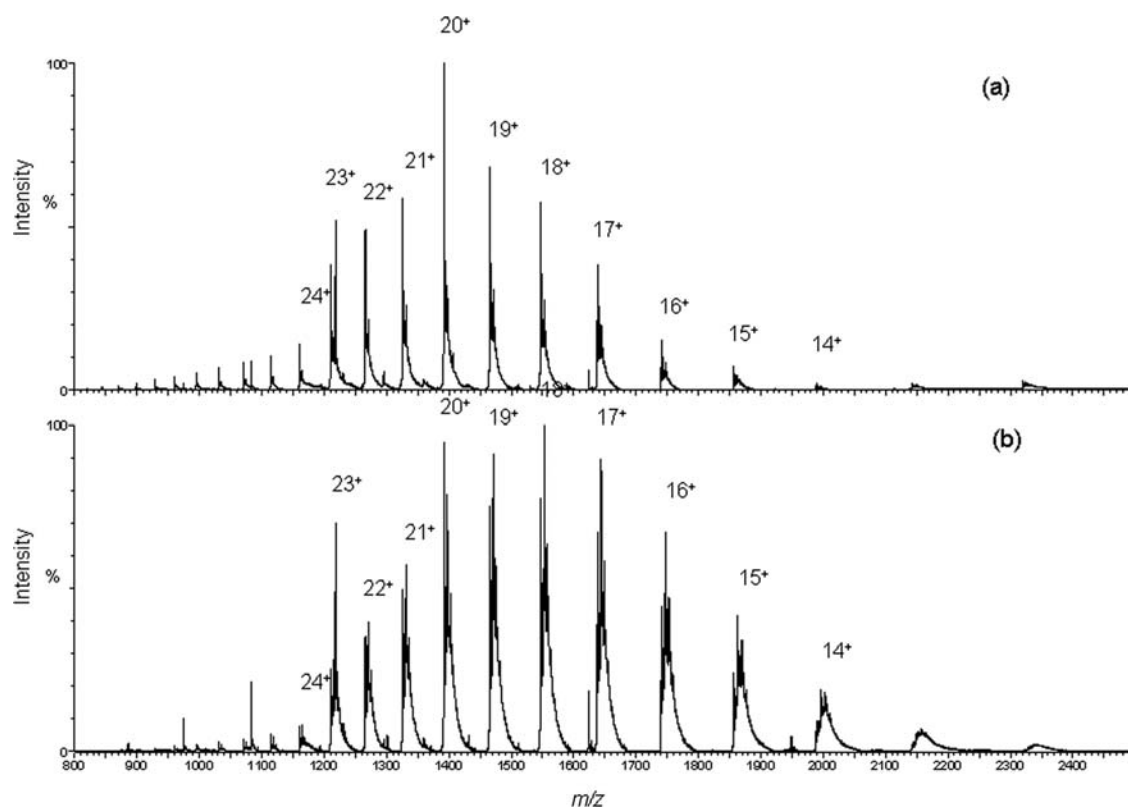## Experimental methods

Protein purification

The procedure for the purification of *Pf* TIM has been described previously (Velanker *et al.*, 1997). Purity was checked by SDS–PAGE and LC-ESI-MS which yielded a mass of 27 831.2 Da, in full agreement with the sequence.
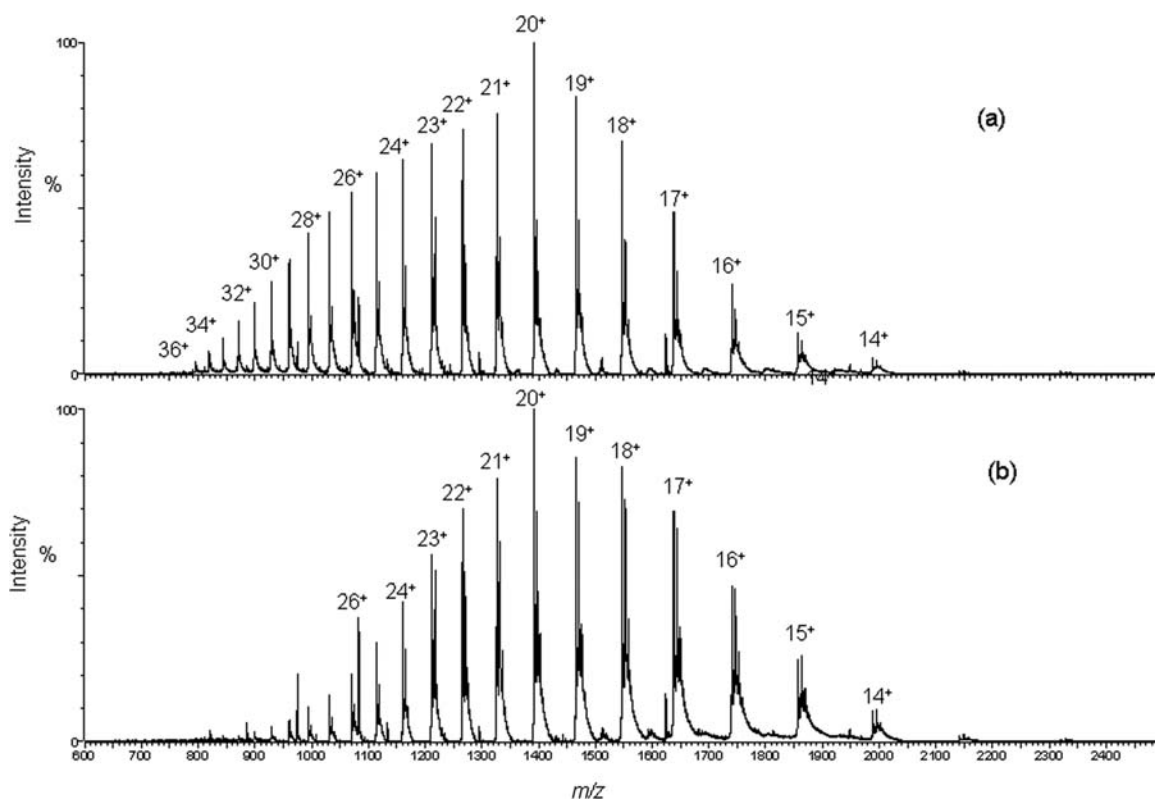
Fig. 2. ESI-MS spectra of *Pf* TIM at pH 6.8 and source temperatures (**a**) 20°C, (**b**) 60°C and (**c**) 80°C.

Fig. 3. ESI-MS spectrum of *Pf* TIM at pH 2.8 and source temperature 60°C.

Fig. 4. ESI-MS spectra of *Pf* TIM as a function of solution pH and source temperature. Temperature 60°C (**a**) pH 1.8 and (**b**) pH 6.8. Note that differences in charge state distribution are observed at pH 2.8 (Fig. 3).

Fig. 5. ESI-MS spectra of *Pf* TIM as a function of solution pH and source temperature. Temperature 80°C (**a**) pH 1.8 and (**b**) pH 6.8.

292

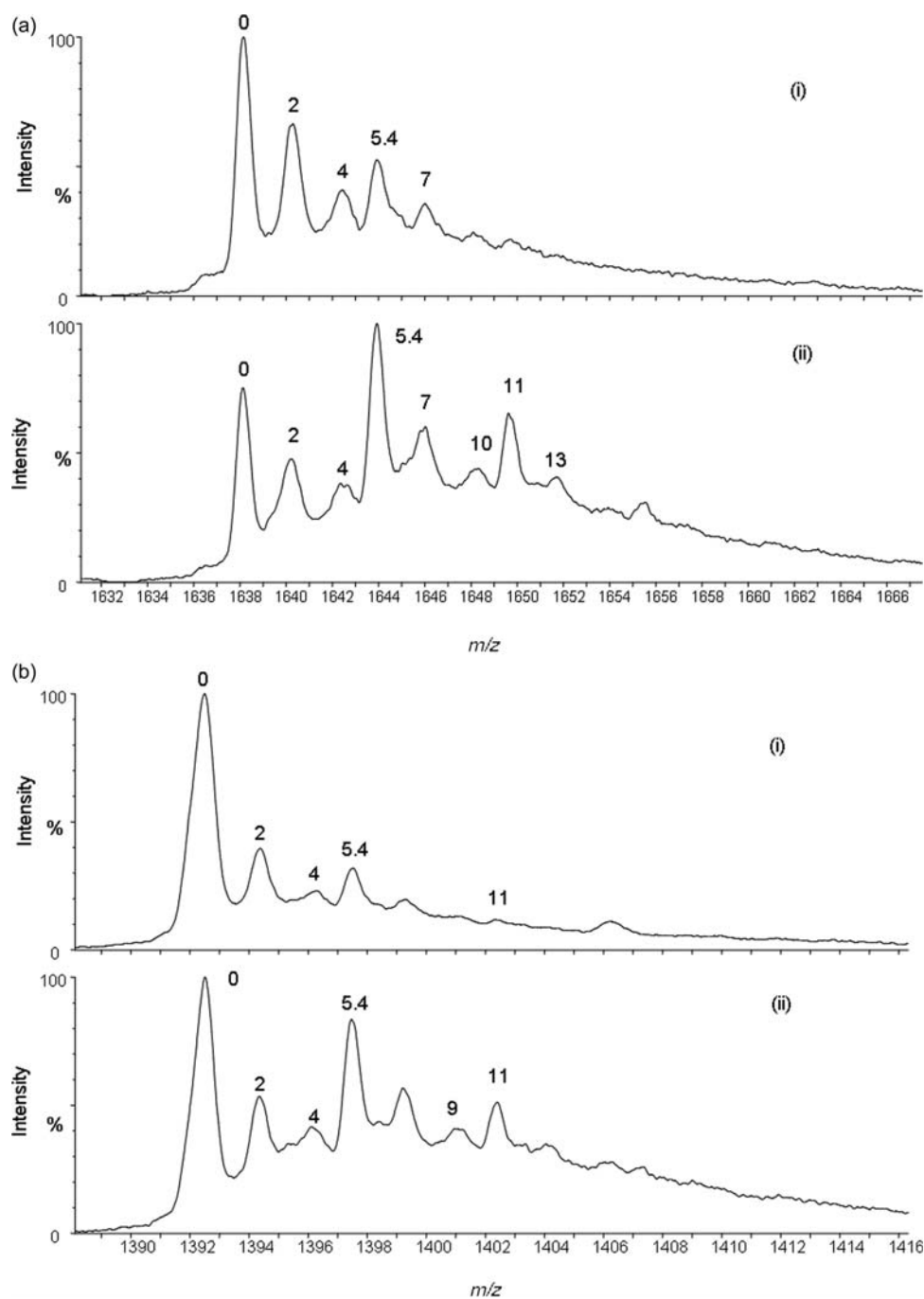TIM concentrations were determined using a molar extinction coefficient of $\epsilon_{280} = 25\,710$/M/cm.

Mass spectrometry

Samples for mass spectrometry were prepared by dissolving lyophilised protein in deionised water (Milli Q, France), yielding a solution with pH 6.8. Low pH solutions were prepared by acidification with formic acid to yield a final pH of ~1.8 or 2.8. The reported pH values in the electrospray mass spectra presented correspond to that measured in bulk solution, prior to infusion into the mass spectrometer. Samples were injected through a glass capillary needle (nanoflow probe tips, type E: Waters Corporation, Milford, MA, USA)

and also through a normal electrospray capillary. Electrospray flow rate was ~5 µl/min and the nanospray needle flow rate was ~200 nl/min.

QTOF-nano-ESI

All spectra were acquired on an ESI QTOF Ultima API (Waters Corporation, Milford, MA, USA) equipped with a Z-spray nanoflow source. External calibration was done using caesium iodide in the range of $100-4000$ *m/z*. A resolving power ($m/\Delta m$) of ~10 000 and a mass accuracy of <10 ppm were obtained using the standards. The mass spectrometer was operated in the positive ion mode. Caesium iodide was used as a calibrant (Sigma, St Louis, MO, USA).

Fig. 6. Expansion of charge states indicating satellite peaks due to hydration at 60°C. (**a**) $z = 17^+$, (i) pH 1.8 and (ii) pH 6.8, (**b**) $z = 20^+$, (i) pH 1.8 and (ii) pH 6.8 and (**c**) $z = 23^+$, (i) pH 1.8 and (ii) pH 6.8.
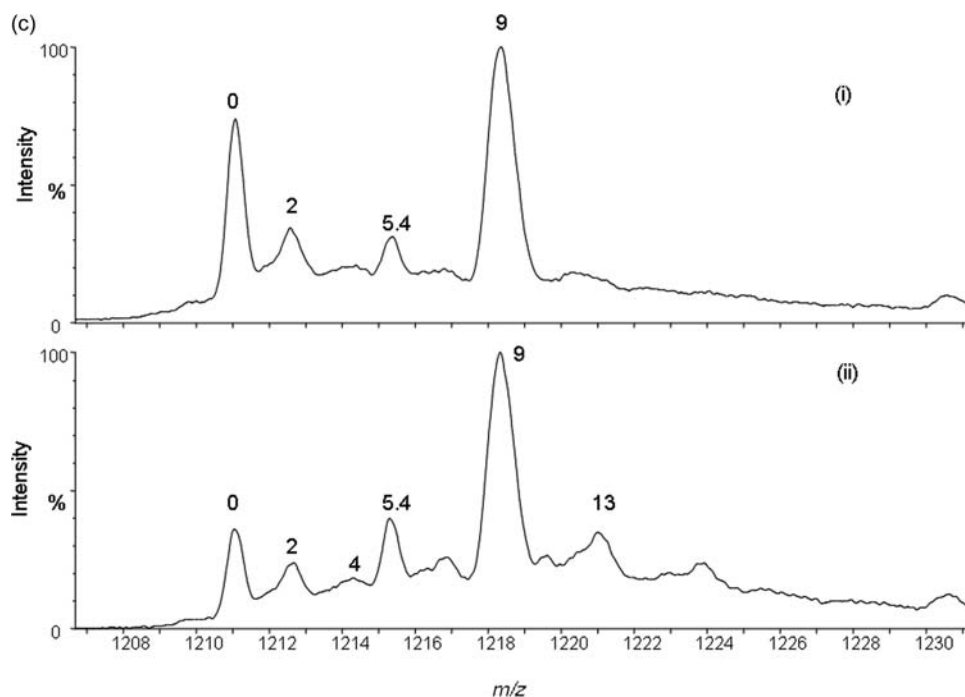
Fig. 6. Continued.

For the detection of the TIM dimer in nano mode, using glass capillaries, a capillary voltage of 1.8–2.4 kV, cone voltage of 70 V, collision energy of 5 eV and source block temperature of 20°C were used. Ion transfer stage pressure 2.36e0 mBar, quadrupole analyser pressure of 3.88e−5 mBar and TOF analyser pressure of 6.18e−7 mBar were maintained to preserve non-covalent interactions.

The following experimental parameters were used in normal electrospray mode: capillary voltage 3–3.5 kV, cone voltage 70–100 V, source block temperature 20°C, 60°C and 80°C, desolvation temperature 200°C and a collision energy of 10 eV. The tubing was flushed with H_2O before injection of the sample.

Data acquisition was done using the MassLynx software (version 4.0). The mass spectral data are presented as raw data, on an *m/z* scale, with minimal smoothening by the Savitzky–Golay method and were analysed using MaxEnt1 application software within the MassLynx platform and also checked manually.

Deuterium exchange

Deuterium exchange of *Pf* TIM was carried out by dissolving a lyophilised protein sample in deuterated water (D_2O, 2.5 mg protein/ml). Positive ion mode spectra were recorded in D_2O by direct injection.

Circular dichromism

Far UV-circular dichromism (CD) measurements were carried out on a Jasco J-715 spectropolarimeter. A path length of 1 mm was used and the spectra were averaged over four scans at a scan speed of 10 nm/min.

Fluorescence spectroscopy

Fluorescence emission spectra were recorded on a Hitachi F-2500 fluorescence spectrometer. The protein samples were excited at 280 nm and the emission spectra were recorded from 300 to 400 nm. Excitation and emission band pass were kept as 5 nm.

Analysis of crystal structures

In order to identify conserved water molecules in the *Pf* TIM structure, the 35 monomeric units from all *Pf* TIM structures available till date were superposed using the SSM Superpose option (Krissinel and Henrick, 2004) in COOT version 0.1.2 (Emsley and Cowtan, 2004). Thirteen *Pf* TIM structures available in the PDB and 17 structures of TIM from other organisms were used for superposition. These are listed in Table I. The water numbers refer to the water number of the PDB entry 1O5X. Initially, the water molecules in the superposed A subunits of all the crystal structures were visually examined and those present in seven or more structures within a radius of 1.8 Å were retained. The presence of these water molecules was checked in the rest of the 35 superposed subunits of *Pf* TIM. In order to confirm the presence of a water molecule, the corresponding electron density maps were examined. If a density corresponding to 2Fo-Fc at 1σ level or Fo-Fc at 3σ level was observed at the expected position of the water, it was considered to be present, even if the corresponding water molecule was absent in the deposited PDB file. The B-factors of the water molecules were examined and atoms in contact with it were identified by using a cutoff radius of 4 Å using the CONTACT program of the CCP4 program suite (Collaborative Computational Project Number 4, 1994). Polar atoms lying within the radius of 3.8 Å from a water molecule were considered to be probable hydrogen bonding partners. The accessible surface areas
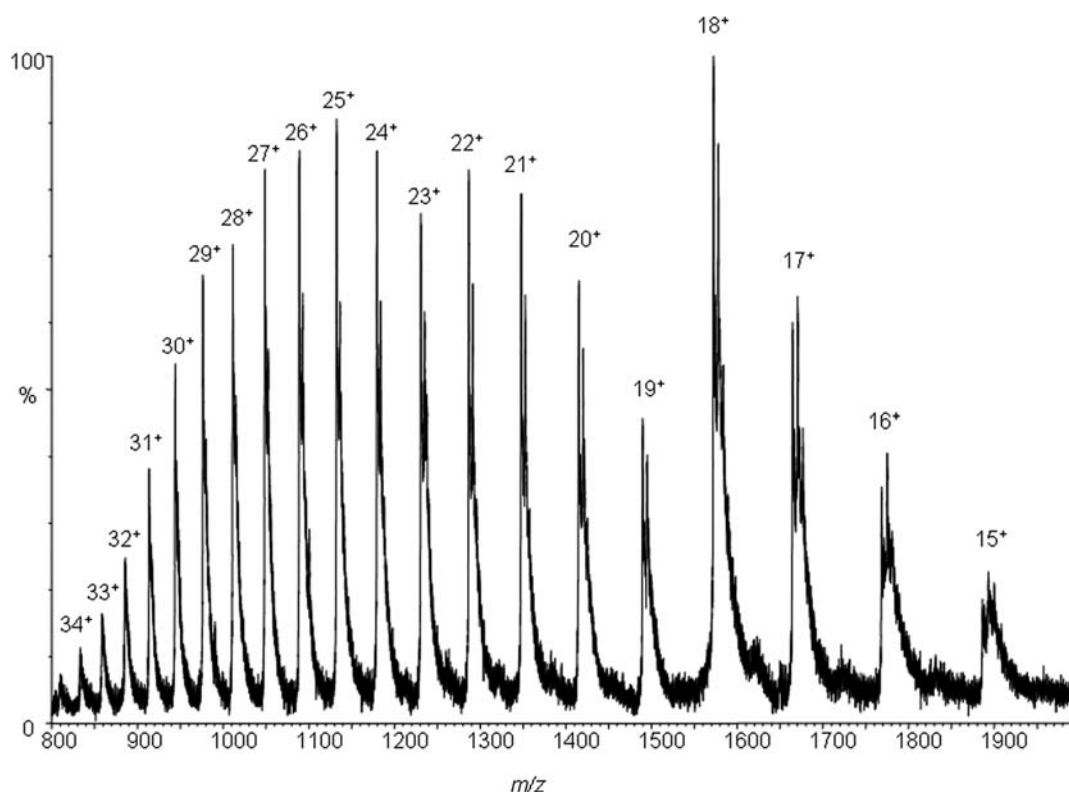
294

Fig. 7. ESI-MS spectra of deuterated *Pf* TIM at 60°C.

Fig. 8. Expansion of charge state $z = 17^+$ of deuterated *Pf* TIM at 60°C. Note the shift of satellite peaks upon deutration.

(ASA) of water molecules were calculated using NACCESS (Hubbard and Thornton, 1996).

Results

Mass spectrometry of Pf TIM

Figure 1 shows the charge state distribution obtained upon injection of a solution of *Pf* TIM at pH 6.8, with a source temperature of 20°C and collision energy of 5 eV under condition of nanospray. It is evident that two clear distributions are observable in the *m/z* range of 1600–3500, corresponding to the 17^+ to 22^+ charge states of a dimer and the 11^+ to 16^+ charge states of a monomer. The calculated masses of both the monomer and the dimer agree well with the expected values (monomer, M_{calc}: 27 831.5 Da; M_{obs}: 27 831.2 Da; dimer, M_{calc}: 55 663 Da; M_{obs}: 55 659.9 Da.). An interesting feature of the charge state distribution shown

in Fig. 1 is that the most intense peaks in the monomeric species correspond to the charge state 13^+ and 14^+, whereas in the dimer, the most intense charge states are 19^+ and 20^+. This may correspond to species in which only 9 or 10 sites per monomer are indeed protonated in the dimeric species. Notably, several acidic residues E-77, D-85 and E-97 lie proximal to the interface. Figure 1 (inset) shows an expansion of the 20^+ charge state of the TIM dimer. Two satellite peaks correspond to a mass addition of 36 and 72 Da, suggesting that they arise from hydrated species containing two and four bound water molecules, respectively.

Figure 2 compares the charge state distribution of *Pf* TIM using a solution of pH 6.8, as a function of the source temperature, obtained under normal electrospray conditions. In all three cases, source temperatures 20°C, 60°C and 80°C (Fig. 2), we observe peaks corresponding to the monomer, with no trace of a dimeric species. This may be contrasted

with the results observed under nanoflow conditions (Fig. 1). It is evident that upon increasing the source temperature, higher charge states are more populated, consistent with the idea that there is enhanced unfolding of the monomeric species. At low source temperature, 20°C, the peaks are considerably broadened suggestive of inefficient desolvation. At higher source temperatures (60° and 80°C), there is an appreciable sharpening of the mass spectral peaks, with clear observation of satellite peaks, which arise due to hydrated species.

Figure 3 shows the charge state distribution observed at pH 1.8 at a source temperature of 60°C, under conditions of normal electrospray. No dimeric species is observed, but two clear distributions are seen, which presumably correspond to partially folded and unfolded monomeric states. The partially folded species is characterised by charge states 14^+ to 17^+, whereas the charge states 18^+ to 34^+ represent populations of unfolded species. This interpretation is of course subject to the caveat that significant differences in gas phase conformations may be observed between neighbouring charge states. For example, the 5^+ and 7^+ charge states of cytochrome c show dramatically different conformations in the gas phase (Woenckhaus *et al.*, 1997a). It must also be noted that protein charge state distributions can be influenced by electrospray conditions, especially the nature of excess ions that are present in the small droplet before ion evaporation (Verkerk *et al.*, 2003).
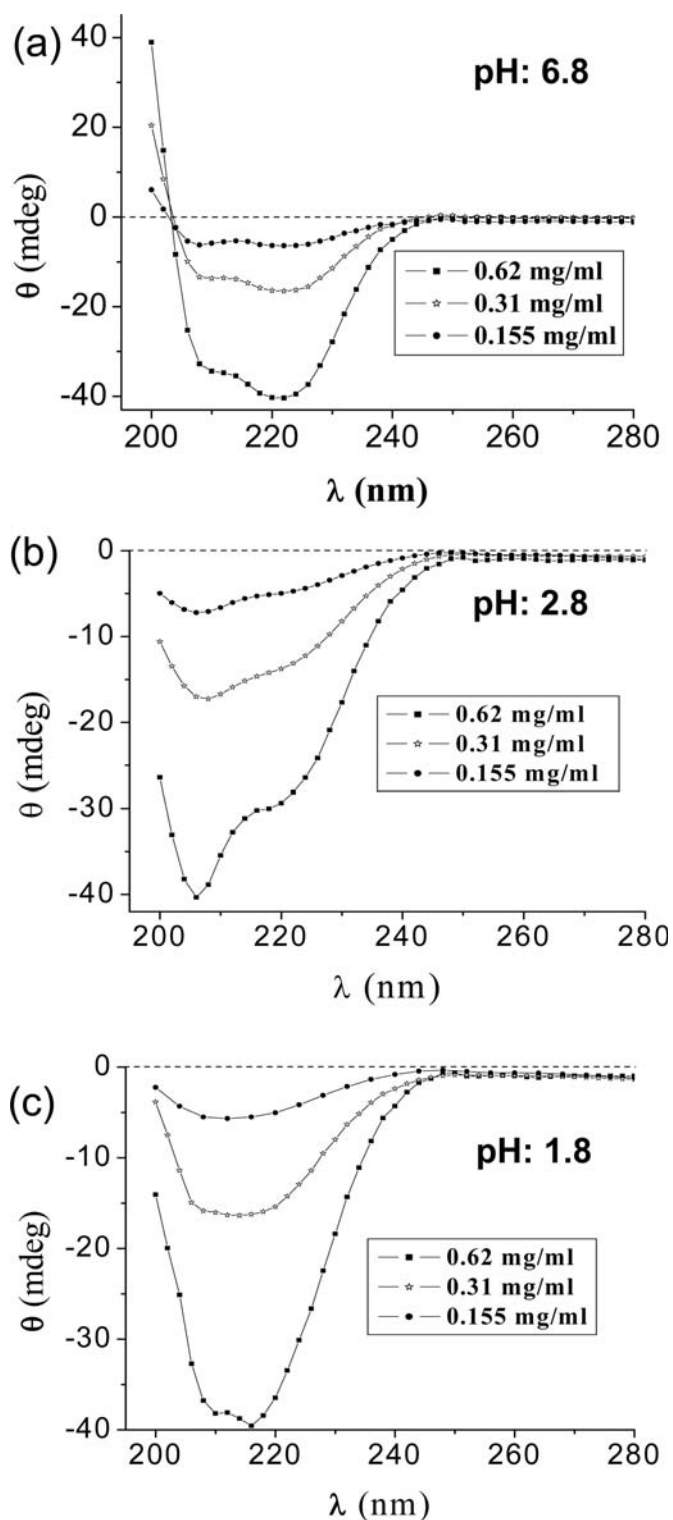
A notable feature of the spectra in Figs 4 and 5 is the observation of satellite peaks which correspond to hydrated species at source temperature 60° and 80°C, respectively. Figures 6a–c shows an expanded view of charge states 17^+, 20^+ and 23^+ at source temperature 60°C (pH 1.8 and 6.8).

A comparison of the charge state distributions at pH 1.8 (Figs 4a and 5a) and pH 6.8 (Figs 4b and 5b) with source temperatures of 60°C and 80°C (Figs 4 and 5) immediately reveals a much wider distribution of observed charge states at higher temperature. Charge states ranging from 14^+ to 36^+ can be detected at pH 1.8, whereas at pH 6.8, the charge state distribution is narrower, from 14^+ to 28^+ at 80°C (Fig. 5). In all cases, deconvolution yields the exact mass of the *Pf* TIM monomer (27 831.2 Da). The differences at the two pH values are more dramatic, when the source temperature is raised to 80°C. The observation of satellite peaks, separated from the parent peak by *m/z* values of $18/n$, clearly suggests that these correspond to species from which water molecules have not been stripped off. The *n* value yields the number of adducted waters.

Deuteration exchange studies

Interpretation of the satellite peaks in terms of hydrated species can be difficult due to the appearance of peaks corresponding to alkali ion adducts. An attempt was therefore made to examine the effect of deuteration on *m/z* values of the satellite peaks.

Pf TIM can be exhaustively deuterated by dissolution in D_2O permitting exchange to occur over a period of 1 week at 25°C. Under these conditions, almost complete exchange of all labile NH, OH and SH groups are achieved. The results of ESI MS spectra (Fig. 7) yield a mass of 28 286.2 Da. *Pf* TIM has 459 exchangeable hydrogens, so the expected mass of fully deuterated protein is 28 290 Da. Figure 8 shows an expansion of the satellite peak observed for charge state 17^+. Peaks are separated by *m/z* values of $20/z$, suggesting that the

Fig. 9. Far UV-CD spectrum of *Pf* TIM at concentrations of range 0.15, 0.31 and 0.62 mg/ml at pH (**a**) 6.8, (**b**) 2.8 and (**c**) 1.8.

anticipating shift upon deuteration is indeed observed, thereby confirming the assignment of the satellite peaks to hydrated species.

Biophysical characterisation of Pf TIM at low pH

The effect of pH on the solution state conformation of the protein was also probed using CD. Figure 9 shows the far

Table II. Conserved water molecules in *Pf* TIM crystal structures

No.	Water[a]	Conservation in *Pf* TIM structures	B-factor	ASA[b] (Å^2)	No. of H-bonds with		Conservation across organisms[c]
					Protein	Water	
1	HOH-30	35	8.4	0	4	2	17
2	HOH-8	35	8.7	0	3	1	7
3	HOH-13	34	8.1	0	3	1	16
4	HOH-44	33	10.6	0	4	2	11
5	HOH-6	33	9.7	0	3	1	4
6	HOH-69	32	17.9	1.1	3	2	3
7	HOH-63	32	9.7	0	5	0	13
8	HOH-145	32	20.3	3.7	3	2	16
9	HOH-1	31	7.4	0	4	2	3
10	HOH-119	30	21.8	1.6	4	5	7
11	HOH-17	29	9.1	0	6	1	12
12	HOH-52	28	9.5	0	4	2	15
13	HOH-62	27	11.6	0	5	1	10
14	HOH-5	25	9.1	0	3	2	10

[a]The water numbers refer to the water number of the PDB entry 1O5X.
[b]ASA, accessible surface areas of water molecules calculated using NACCESS. A fully exposed water molecule has an ASA of 98.5 Å^2.
[c]A total of 18 TIM structures from distinct organisms were examined. The numbers represent the number of examples in which the water molecules are found.

UV-CD spectrum of *Pf* TIM over a concentration range 0.15–0.62 mg/ml at pH 6.8, 2.8 and 1.8. The nanospray ESI-MS studies were carried out using relatively high concentrations (up to 5 mg/ml) to maximise the detection of dimeric species. CD studies, however, have to be carried out at much lower concentrations due to excessive absorption artefacts.
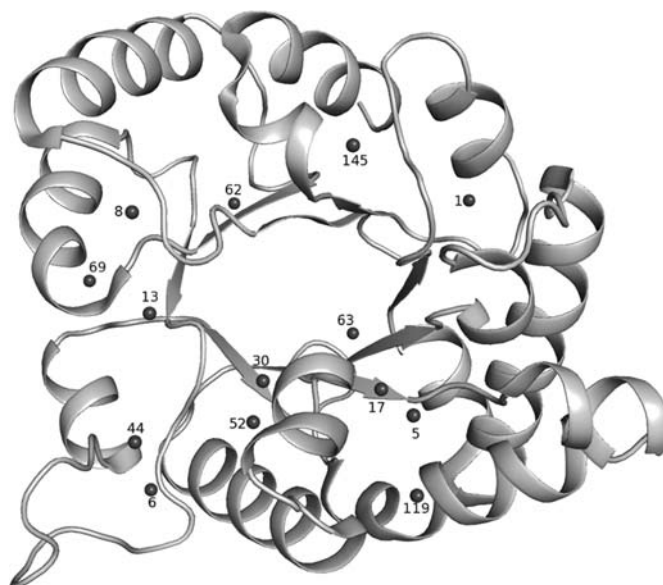
Inspection of the CD spectrum in Fig. 9 reveals that there is a significant change in secondary structure at low pH. The native folded protein exhibits minima at 220 and 208 nm, characteristic of the $\alpha_8\beta_8$ barrel (Gokhale *et al.*, 1999). In contrast, at pH 2.8 and 1.8, there is a significant change in both band positions and relative intensities of the two bands, suggestive of a significant change in secondary structure. Notably, *Pf* TIM retains appreciable structure even at pH 1.8.

Pf TIM contains two tryptophan residues W11 and W168, whose fluorescence emission properties serve as reporters of the local environment (Pattanaik *et al.*, 2003). At pH 6.8, *Pf* TIM shows an emission maximum of 330 nm (excitation 280 nm). At low pH, a red shift is observed to 338 nm (pH 2.8 and 1.8), suggestive of greater degree of solvent exposure of tryptophan residues upon acid unfolding. The observed emission maximum in 6 M guanidium chloride is 346 nm, which may be compared with a value of 355 nm for free tryptophan in water. Taken together with the CD results, these observations suggest that *Pf* TIM retains considerable secondary and tertiary structure even at low pH, corresponding to the conditions under which the protein has been subjected to ESI-MS.
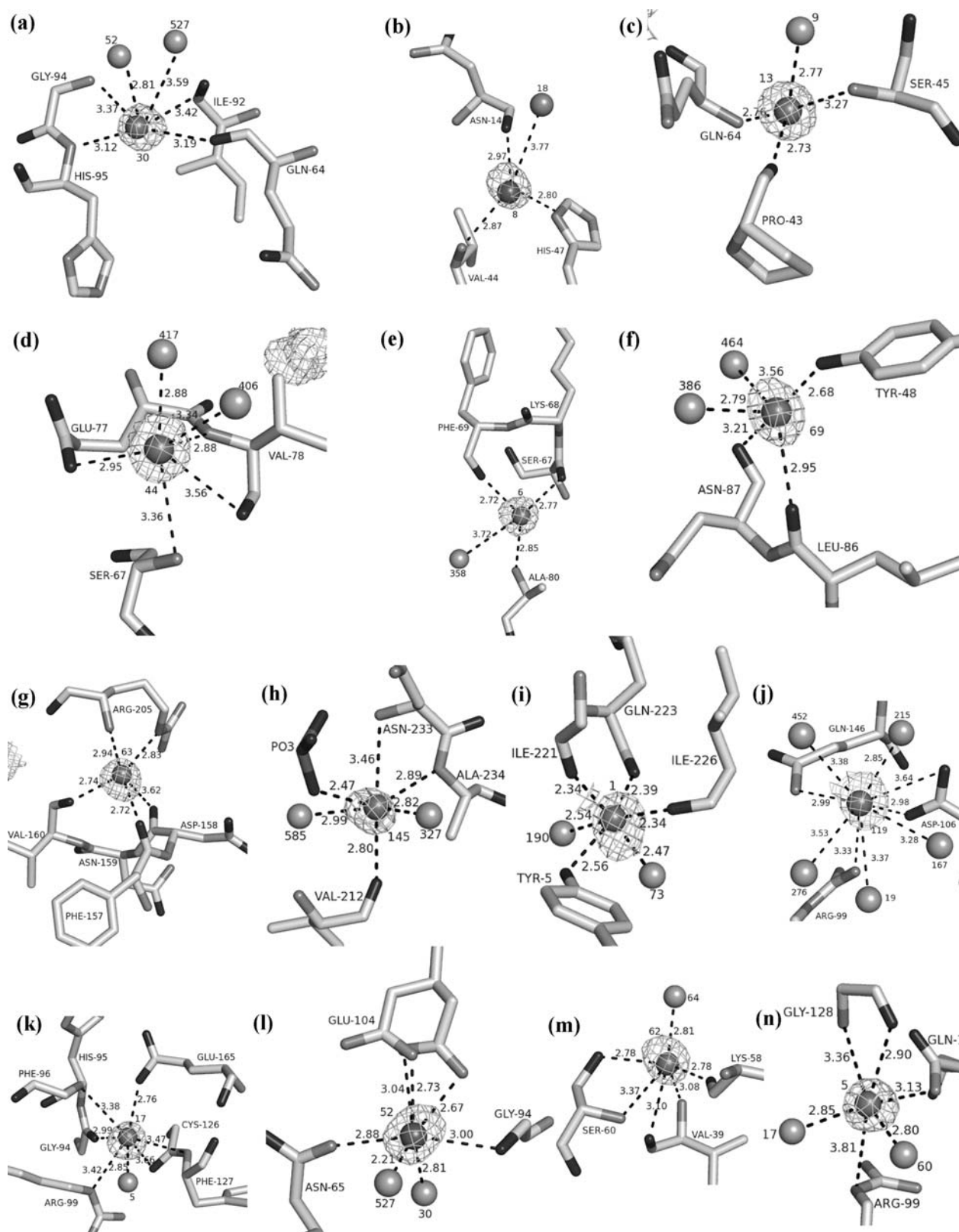
Water molecules in the *Pf* TIM crystal structure

Hydration of the monomer

Thirteen independent crystal structures of *Pf* TIM, which include both site-specific mutants and protein ligand complexes, have been previously determined in this laboratory. These 13 structures were examined for the presence of conserved water molecules. In all, 35 monomers which

Fig. 10. Ribbon diagram of the *Pf* TIM monomer with conserved water molecules observed in the crystal structure (PDB code: 1O5X).

constitute the crystallographic asymmetric units of *Pf* TIM were examined. Table II lists 14 water molecules which are observed in at least 25 out of 35 monomeric units. The relevant B factors and solvent ASA are also listed. Of these, only two water molecules (HOH 30 and HOH 8) are observed in all the 35 cases. We have also examined the conservation of these 14 water molecules in crystal structures of TIM from various other organisms. One representative structure was chosen when several entries were found in the Protein Data Bank (PDB). Interestingly, 9 of the 14 water molecules listed in Table II are conserved in the structure of TIMs from at least 10 organisms. The locations of the 14 water molecules bound to a single subunit of TIM are illustrated in Fig. 10. The environment of specific water molecules and the number of hydrogen bonds formed to protein atoms and other water molecules are summarised in Fig. 11.

Fig. 11. Environment of the conserved water molecules in *Pf* TIM (PDB code: 1O5X). The hydrogen bonds are shown as dashed lines and the distances labelled. The electron density shown around the water molecules corresponds to the 2Fo-Fc map contoured at 0.8σ. (**a**) HOH-30, (**b**) HOH-8, (**c**) HOH-13, (**d**) HOH-44, (**e**) HOH-6, (**f**) HOH-69, (**g**) HOH-63, (**h**) HOH-145, (**i**) HOH-1, (**j**) HOH-119, (**k**) HOH-17, (**l**) HOH-52, (**m**) HOH-62 and (**n**) HOH-5.

HOH-30, HOH-13, HOH-145 and HOH-52 are conserved in 17, 16, 16 and 15, respectively, of the 18 TIM structures available from diverse species. Out of these, HOH-30, HOH-13 and HOH-52 can be considered as water molecules

with a structural role. HOH-30 is held by hydrogen bonds with the main chain O of Gln64 (end of strand 3) and main chain N of Gly94 (end of strand 4) (Fig. 11a). The conserved HOH-52 is also hydrogen bonded to HOH-30, and is further

bound to N of Gly94, Asn65 side chain and the side chain carboxyl group of Glu104 which occurs in alternate conformations in many *Pf* TIM structures. In spite of the alternate side chain positions of Glu104, the location of HOH-52 is maintained (Fig. 11l). Notably, Glu104 is a completely conserved residue in all TIM sequences determined so far. Mutation of this residue to Asp104 leads to a human genetic disorder, probably because the mutant protein becomes thermally labile (Daar *et al*., 1986; Repiso *et al*., 2002). A recent report also suggests that this mutation affects dimerisation (Ralser *et al*., 2006). HOH-13 also occupies a similar position connecting the ends of strands 2 and 3, by forming hydrogen bonds with the main chain O of Ser45 and the main chain N of Gln64 (Fig. 11c). Interestingly, both HOH-30 and HOH-13 are attached to the main chain atoms of Gln64—a highly conserved residue across TIMs from various organisms. Gln64 appears to have a role in stabilizing an unusual backbone conformation of the active site lysine residue, which has positive Ramachandran ϕ, ψ angles in all known TIM structures. HOH-145 is present in the substrate binding pocket, between loops 7 and 8, and our analysis shows that it is present irrespective of occupancy of the active site by ligands. This water molecule may play an important role in the activity of TIM. The remaining water molecules which are found only in *Pf* TIM structures are bound to the highly variable loop regions, which exhibit significant differences in the side chain and the main chain conformations.

Water molecules at the dimer interface

Subunit interfaces in multimeric proteins are often more polar and more hydrated than protein interiors (Tsai *et al*., 1996; Xu *et al*., 1997; Valdar and Thornton, 2001). Interface residues are also generally highly conserved in proteins with association constants for multimerisation $\geq 10^9$ M (Klotz *et al*., 1970; Darnall and Klotz, 1975; Jaenicke and Lilie, 2000), with the result that subunit dissociation is not observed even at nanomolar concentrations, which is often the lowest detectable limit in most biophysical studies in solution. Complementary surfaces of interacting proteins have presumably been shaped by evolutionary selection, resulting in high specificity and high strength of interactions. Hydration of interfaces is common and water molecules appear to promote interactions between the protein surfaces (Xu *et al*., 1997).

Table III lists the number of water molecules observed at the *Pf* TIM dimer interface, which show a high degree of conservation across the 13 independent crystal structures determined. Water molecules which interact with atoms from both A and B subunits have been identified. Six water molecules are conserved in at least 25 cases. Interestingly, four water molecules, HOH-16, HOH-65, HOH-72 and HOH-17, are conserved in as many as 17, 15, 13 and 11 structures of TIM from other species. Figure 12 shows a view of the water molecules at the subunit interface and Fig. 13 illustrates the environment of the six most conserved water molecules.

Discussion

The observed mass spectra of *Pf* TIM established that a dimer of 55 659.9 Da can indeed be detected under nanospray conditions at pH 6.8, with a low source temperature of 20°C. *Pf* TIM contains 8 Arg, 22 Lys, 5 His and a free amino terminus resulting in 36 possible sites of protonation. In acidic media (pH 1.8, Fig. 5), charge states all the way up to 36^+ can be detected, suggesting that under conditions which promote chain unfolding in the gas phase, complete charging at all sites is indeed possible. In contrast, major charge states detected for the dimeric species (Fig. 1) are 19^+ and 20^+, suggesting that between 9 and 10 protons are added to each protein subunit.

The charge states observed between *m/z* values 2531.2 and 1740.4 are assigned to monomeric species carrying charges between 11 and 16 protons. It may be noted that the peak at *m/z* 2531 corresponds to the overlap of charge state 11^+ of a monomer and 22^+ of a dimer (Fig. 1). Dimer dissociation during the nano-electrospray process should result in largely folded monomeric species, which are formed from dimeric species carrying 22 or more charges. It is likely that the monomeric states carrying 11^+ to 16^+ charges correspond to largely folded species. Normal electrospray conditions result in dimer dissociation (Fig. 2). The low pH (1.8), high source temperature (60°C) spectra shown in Fig. 3 illustrate the two distinct charge state distributions from the monomeric species, which differ appreciably from those observed in high pH and low source temperature spectra (Fig. 2a). The charge states which give rise to the distribution between 14^+ and 18^+ (Fig. 3) are clearly more unfolded than the states exhibiting

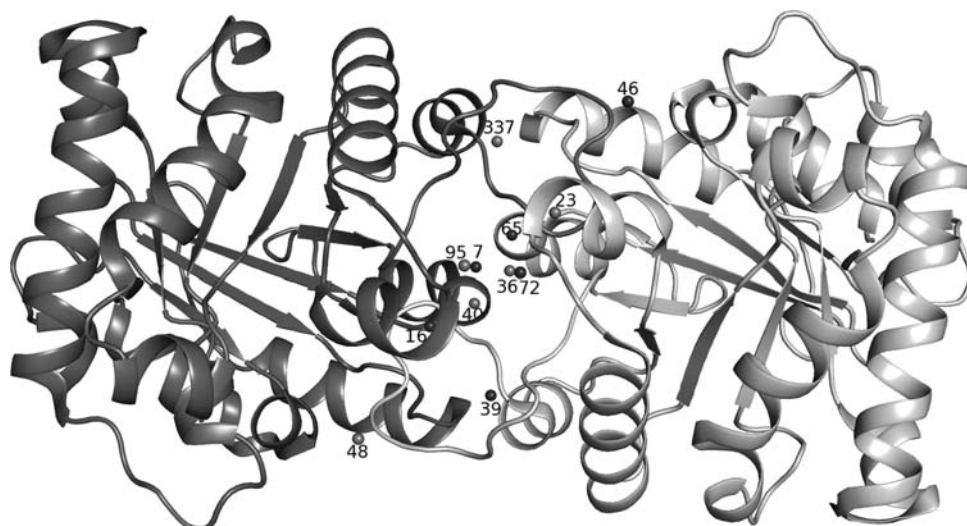**Table III.** Conserved interface water molecules in *Pf* TIM crystal structures

No.	Water[a]	Conservation in *Pf* TIM structures	B-factor	ASA[b] (Å^2)	No. of H-bonds with			Conservation across organisms[c]
					Protein		Water	
					A	B		
1	HOH-39	33	11.1	15.2	2	1	0	4
2	HOH-7	32	9.6	25.5	0	4	3	11
3	HOH-72	31	13.1	3.6	1	1	4	13
4	HOH-16	31	9.3	25.5	3	1	1	17
5	HOH-65	28	9.3	1.2	2	3	1	15
6	HOH-46	26	12.7	9.3	2	2	2	9

[a]The water numbers refer to the water number of the PDB entry 1O5X.
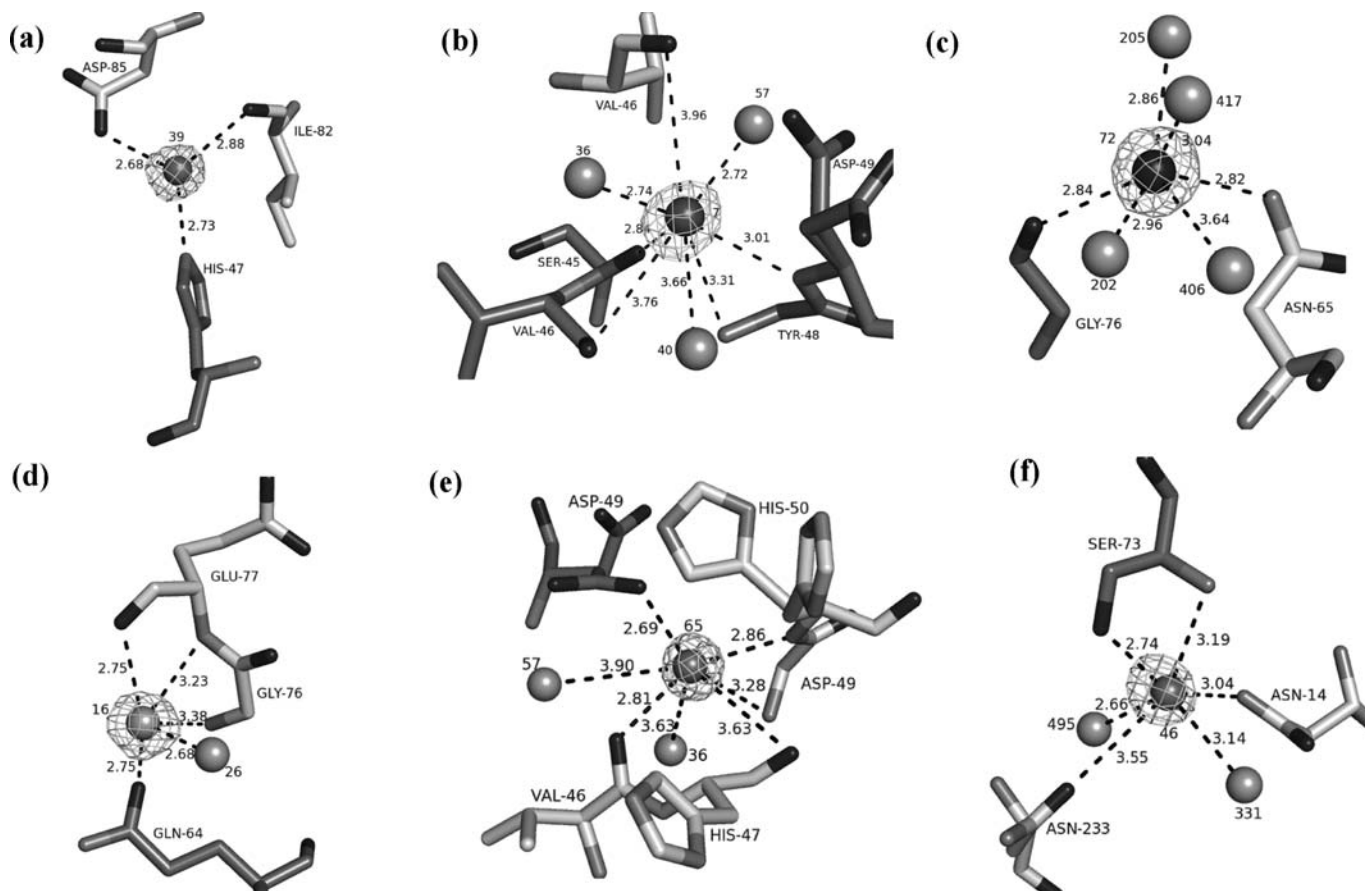[b]ASA, accessible surface areas of water molecules calculated using NACCESS. A fully exposed water molecule has an ASA of 98.5 Å^2.
[c]A total of 18 TIM structures from distinct organisms were examined. The numbers represent the number of examples in which the water molecules are found.

Fig. 12. Ribbon diagram of the *Pf* TIM dimer indicating conserved interface water molecules (PDB code: 1O5X).
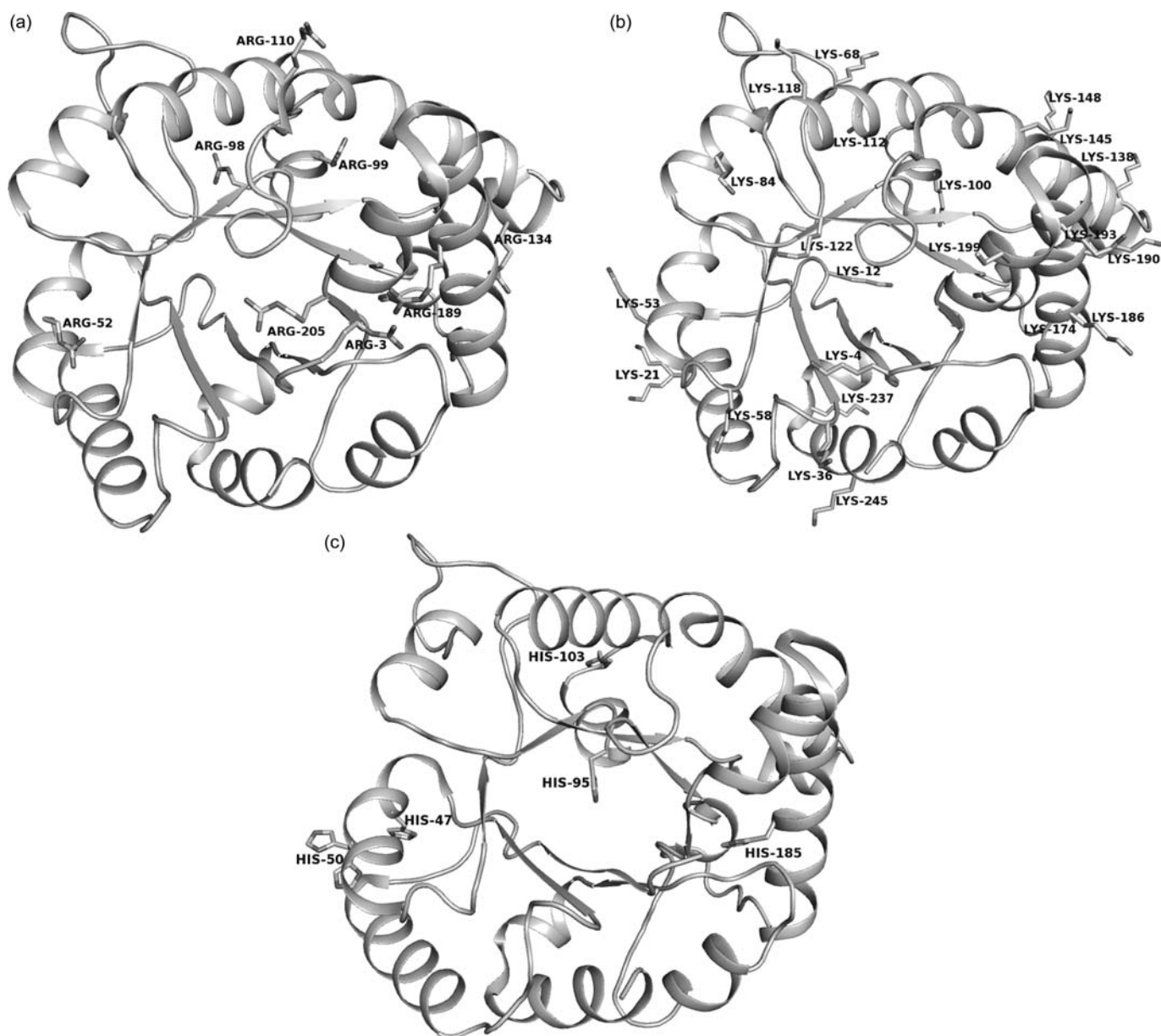

Fig. 13. Environment of the conserved dimer interface water molecules in *Pf* TIM (PDB code: 1O5X): (**a**) HOH-39, (**b**) HOH-7, (**c**) HOH-72, (**d**) HOH-16, (**e**) HOH-65 and (**f**) HOH-46. A and B subunits of the dimer are shown in light grey and dark grey, respectively, and the water molecules at the interface related by the 2-fold symmetry are shown in light grey.

a distribution of charges between 11^+ and 16^+ (Fig. 1). The second distribution observed may be assigned to a largely unfolded monomeric species having observed charge states 18^+ to 34^+ (Fig. 3). Assignment of the distinct charge state distributions to differentiate unfolded species is supported by CD and fluorescence experiments,

which suggest that non-native, partially unfolded states are predominantly populated at low pH.

Although an identification of the specific protonation sites is not possible from mass spectrometric data, a comparison may be made with the environments of the sites in the crystal structures. Figure 14a–c shows the distribution of the

Fig. 14. Environment of basic residues in *Pf* TIM: (**a**) Arg, (**b**) Lys and (**c**) His.
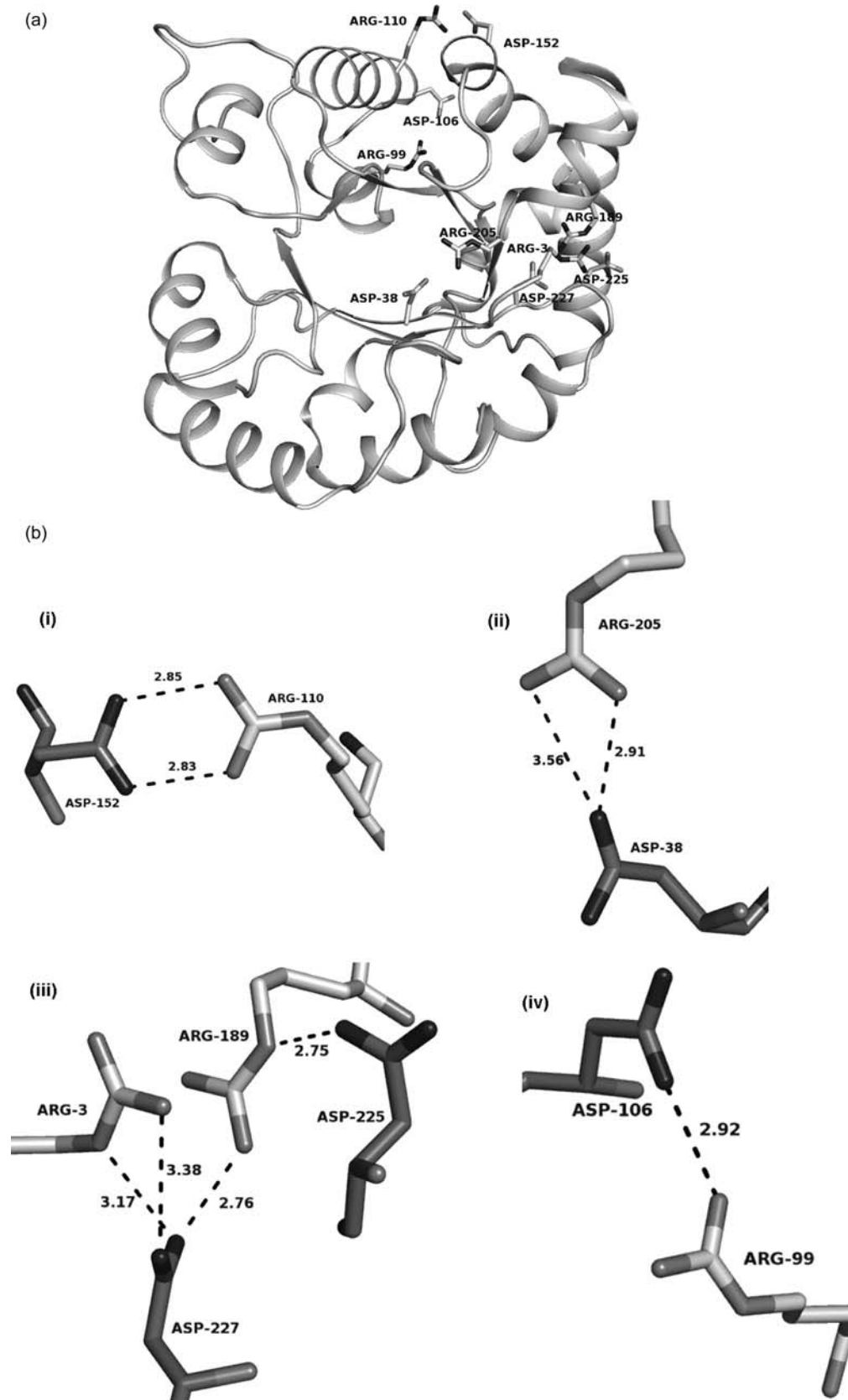
Arg, Lys and His residues which are potential sites for protonation. Of the eight Arg residues which have high pKa values, as many as seven are involved in interactions with neighbouring, negatively charged residues in the crystal structures (Fig. 15). Similarly, 6 out of 22 lysine residues are involved in interactions with neighbouring negatively charged residues (not shown). In some cases, ion pair interactions are mediated through bridging water molecules (Fig. 13). Interestingly, only one of the five His residues participates in electrostatic interactions (Fig. 16g). Assuming that the memory of solution phase structures is maintained in the gas phase, under conditions where a dimeric state is populated, it is possible that these water molecules are mass spectrometrically detectable.

At pH 6.8, the carboxylic acid groups are largely negatively charged in *Pf* TIM. There are 19 Glu, 14 Asp and a C-terminal carboxylic acid, which are potentially ionisable. Thus, maintaining an overall charge of 9^+ to 10^+ on each

subunit necessarily requires the protonation of an almost equal number of carboxylic acid side chains. The detection of clearly asymmetric charge state distributions for the monomeric species suggests that largely folded and partially folded states of isolated monomeric subunits are populated in the gas phase.

Hydration is an inevitable consequence of the milieu in which proteins exist in cells and may have an important functional role (Meyer, 1992). The crystal structures of proteins have provided a great deal of detailed structural information on water molecules which are bound to protein surfaces (Bottoms *et al.*, 2006). The presence of a very large number of water molecules in the crystal lattice permits observation of not only strongly bound solvent molecules, but also allows visualisation of water networks which hold proteins together in the solid state. In the case of multimeric proteins and protein–protein complexes, water molecules often act as bridges between
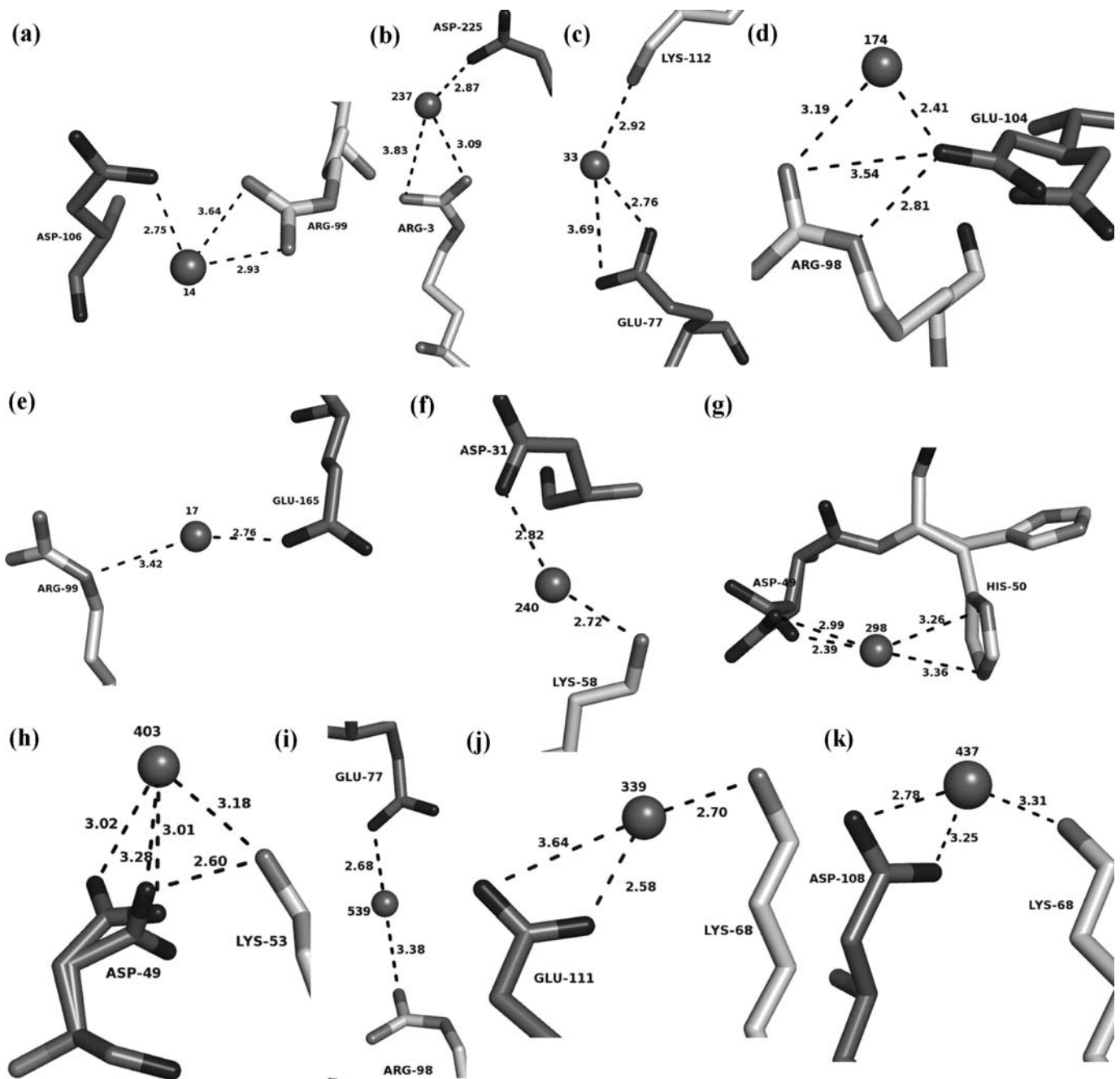
Fig. 15. (**a**) Position of intrasubunit salt bridges in *Pf* TIM: Asp-Arg, (**b**) Asp-Arg salt bridges in *Pf* TIM. (i) Asp152-Arg110, (ii) Asp38-Arg205, (iii) Asp227-Arg3 and Asp225-Arg189 and (iv) Asp106-Arg199.

protruding side chains of amino acid residues on the interacting surfaces (Paliwal *et al.*, 2005). Although the strength of interaction between the individual water

molecules and protein residues is hard to estimate, crystallographic observations may be used to identify tightly bound and invariant water molecules in protein structures,

Fig. 16. Water-mediated ionic interactions in *Pf* TIM. (**a**) Arg99-HOH-14-Asp106, (**b**) Arg3-HOH-237-Asp225, (**c**) Lys112-HOH-33-Glu77, (**d**) Arg98-HOH-174-Glu104, (**e**) Arg99-HOH-17-Glu165, (**f**) Lys58-HOH-240-Asp31, (**g**) His50-HOH-298-Asp49, (**h**) Lys53-HOH-403-Asp49, (**i**) Arg99-HOH-539-Glu77, (**j**) Lys68-HOH-339-Glu111 and (**k**) Lys68-HOH-437-Asp108.

determined from differentially hydrated crystals (Nagendra *et al.*, 1998; Ramirez and Freymann, 2006). Simplistically, the water molecules which show multiple interactions to proteins are likely to be bound more strongly to the macromolecular surface.

Interestingly, water bound to the proteins can also be detected in the gas phase by ESI-MS. If ionisation can be accomplished under extremely 'soft' conditions, water molecules which are tightly attached can remain bound to the proteins, thus permitting mass spectral detection (Woenckhaus *et al.*, 1997a; Fye *et al.*, 1998; Rodriguez-Cruz *et al.*, 1999). Notably in the case of *Pf* TIM, the number of water molecules detected which are bound to the diverse charge states under

different conditions corresponds closely to the number of highly conserved water molecules, which are identified in the crystal structures (see preceding section). In the dimer structure, two invariant water molecules are found in almost all structures. Interestingly, species corresponding to two bound water molecules are observed in all the charge states of the monomeric protein. Curiously, the 23^+ charge state alone shows a significant intensity for species containing nine bound water molecules per subunit. It is relevant to note that preferential hydration of specific charge states has been observed in BPTI (Woenckhaus *et al.*, 1997b).

HOH-17 and HOH-39 are conserved in most of the crystal structures examined and mediate interaction between

positively and negatively charged sites (Fig. 15). It is likely that such waters may indeed be strongly bound to the protein even in the gas phase. Hydrogen bonds which are formed between ionic and neutral species are estimated to be exceptionally strong (Meot-Ner, 2005) and the stability will presumably be enhanced in a solvent-free, low dielectric environment. Further information on gas phase hydration may be obtained by comparisons between wild-type and mutant protein structures, where specific hydration sites are eliminated by suitable mutation.

Acknowledgements

Funding

References

Aparicio,R., Ferreira,S.T. and Polikarpov,I. (2003) *J. Mol. Biol.*, **334**, 1023–1041.

Banner,D.W., Bloomer,A., Petsko,G.A., Phillips,D.C. and Wilson,I.A. (1976) *Biochem. Biophys. Res. Commun.*, **72**, 146–155.

Benesch,J.L. and Robinson,C.V. (2006) *Curr. Opin. Struct. Biol.*, **16**, 245–251.

Borchert,T.V., Abagyan,R., Jaenicke,R. and Wierenga,R.K. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 1515–1518.

Bottoms,C.A., White,T.A. and Tanner,J.J. (2006) *Proteins*, **64**, 404–421.

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

Daar,I.O., Artymiuk,P.J., Philips,D.C. and Maquat,L.E. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 7903–7907.

Daniel,J.M., Friess,S.D., Rajagopalan,S., Wendt,S. and Zenobi,R. (2002) *Int. J. Mass Spectrom.*, **216**, 1–27.

Daniel,J.M., McCombie,G., Wendt,S. and Zenobi,R. (2003) *J. Am. Soc. Mass Spectrom.*, **14**, 442–448.

Darnall,D.W. and Klotz,I.M. (1975) *Arch. Biochem. Biophys.*, **166**, 651–682.

Delboni,L.F., Mande,S.C., Rentier-Delrue,F., Mainfroid,V., Turley,S., Vellieux,F.M., Martial,J.A. and Hol,W.G. (1995) *Protein Sci.*, **4**, 2594–2604.

Eaazhisai,K., Balaram,H., Balaram,P. and Murthy,M.R.N. (2004) *J. Mol. Biol.*, **343**, 671–684.

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

Fye,J.L., Woenckhaus,J. and Jarrold,M.F. (1998) *J. Am. Chem. Soc.*, **120**, 1327–1328.

Gayathri,P., Banerjee,M., Vijayalakshmi,A., Azeez,S., Balaram,H., Balaram,P. and Murthy,M.R.N. (2007) *Acta Crystallogr., Sect. D*, **63**, 206–220.

Gokhale,R.S., Ray,S.S., Balaram,H. and Balaram,P. (1999) *Biochemistry*, **38**, 423–431.

Huang,H.H., Liao,H.K., Chen,Y.J., Hwang,T.S., Lin,Y.H. and Lin,C.H. (2005) *J. Am. Soc. Mass Spectrom.*, **16**, 324–332.

Hubbard,S.J. and Thornton,J.M. (1996) NACCESS, version 2.1.1. Department of Biochemistry and Molecular Biology, University College London.

Jaenicke,R. and Lilie,H. (2000) *Adv. Protein Chem.*, **53**, 329–401.

Jurchen,J.C. and Williams,E.R. (2003) *J. Am. Chem. Soc.*, **125**, 2817–2826.

Kaneko,T., Tanaka,N. and Kumasaka,T. (2005) *Protein Sci.*, **14**, 558–565.

Kinoshita,T., Maruki,R., Warizaya,M., Nakajima,H. and Nishimura,S. (2005) *Acta Crystallogr., Sect. F*, **61**, 346–349.

Kitova,E.N., Daneshfar,R., Marcato,P., Mulvey,G.L., Armstrong,G. and Klassen,J.S. (2005) *J. Am. Soc. Mass Spectrom.*, **16**, 1957–1968.

Klotz,I.M., Langerman,N.R. and Darnall,D.W. (1970) *Annu. Rev. Biochem.*, **39**, 25–62.

Krissinel,E. and Henrick,K. (2004) *Acta. Crystallogr. D*, **60**, 2256–2268.

Loo,J.A. (1997) *Mass Spectrom. Rev.*, **16**, 1–23.

Lolis,E., Alber,T., Davenport,R.C., Rose,D., Hartman,F.C. and Petsko,G.A. (1990) *Biochemistry*, **29**, 6609–6618.

Maes,D., *et al.* (1999) *Proteins*, **37**, 441–453.

Mahoney,N.M., Janmey,P.A. and Almo,S.C. (1997) *Nat. Struct. Biol.*, **4**, 953–960.

Maldonado,E., Soriano-Garcia,M., Moreno,A., Cabrera,N., Garza-Ramos,G., de Gomez-Puyou,M., Gomez-Puyou,A. and Perez-Montfort,R. (1998) *J. Mol. Biol.*, **283**, 193–203.

McKay,A.R., Ruotolo,B.T., Ilag,L.L. and Robinson,C.V. (2006) *J. Am. Chem. Soc.*, **128**, 11433–11442.

Meot-Ner,M. (2005) *Chem. Rev.*, **105**, 213–284.

Meyer,E. (1992) *Protein Sci.*, **1**, 1543–1562.

Nagendra,H.G., Sukumar,N. and Vijayan,M. (1998) *Proteins*, **32**, 934–937.

Noble,M.E., Zeelen,J.P., Wierenga,R.K., Mainfroid,V., Goraj,K., Gohimont,A.C. and Martial,J.A. (1993) *Acta Crystallogr. D*, **49**, 403–417.

Paliwal,A., Asthagiri,D., Abras,D., Lenhoff,A.M. and Paulaitis,M.E. (2005) *Biophys. J.*, **89**, 1564–1573.

Parthasarathy,S., Ravindra,G., Balaram,H., Balaram,P. and Murthy,M.R. (2002a) *Biochemistry*, **41**, 13178–13188.

Parthasarathy,S., Balaram,H., Balaram,P. and Murthy,M.R. (2002b) *Acta Crystallogr. D*, **58**, 1992–2000.

Parthasarathy,S., Eaazhisai,K., Balaram,H., Balaram,P. and Murthy,M.R. (2003) *J. Biol. Chem.*, **278**, 52461–52470.

Pattanaik,P., Ravindra,G., Sengupta,C., Maithal,K., Balaram,P. and Balaram,H. (2003) *Eur. J. Biochem.*, **270**, 745–756.

Ralser,M., Heeren,G., Breitenbach,M., Lehrach,H. and Krobitsch,S. (2006) *PLoS ONE*, **1**, e30, 1–12.

Ramirez,U.D. and Freymann,D.M. (2006) *Acta Crystallogr. D*, **62**, 1520–1534.

Repiso,A., Boren,J., Ortega,F., Pujades,A., Centelles,J., Vives-Corrons,J.L., Climent,F., Cascante,M. and Carreras,J. (2002) *Hematologica*, **87**, 4, ECR 12.

Rodriguez-Cruz,S.E., Klassen,J.S. and Williams,E.R. (1999) *J. Am. Soc. Mass Spectrom.*, **10**, 958–968.

Rodriguez-Romero,A., Hernandez-Santoyo,A., Del Pozo-Yauner,L., Kornhauser,A. and Fernandez-Velasco,D.A. (2002) *J. Mol. Biol.*, **322**, 669–675.

Sharon,M. and Robinsonm,C.V. (2007) *Annu. Rev. Biochem.*, **76**, 167–193.

Symersky,J., Li,S., Carson,M. and Luo,M. (2003) *Proteins Struct. Funct. Genet.*, **51**, 484–486.

Takats,Z., Wiseman,J.M., Gologan,B. and Cooks,R.G. (2004) *Anal. Chem.*, **76**, 4050–4058.

Tsai,C.J., Lin,S.L., Wolfson,H.J. and Nussinov,R. (1996) *Crit. Rev. Biochem. Mol. Biol.*, **31**, 127–152.

Valdar,W.S. and Thornton,J.M. (2001) *Proteins*, **42**, 108–124.

Velanker,S.S., Ray,S.S., Gokhale,R.S., Suma,S., Balaram,H., Balaram,P. and Murthy,M.R.N. (1997) *Structure*, **5**, 751–761.

Verkerk,U.H., Peschke,M. and Kebarle,P. (2003) *J. Mass Spectrom.*, **38**, 618–631.

Walden,H., Bell,G.S., Russell,R.J., Siebers,B., Hensel,R. and Taylor,G.L. (2001) *J. Mol. Biol.*, **306**, 745–757.

Wendt,S., McCombie,G., Daniel,J., Kienhofer,A., Hilvert,D. and Zenobi,R. (2003) *J. Am. Soc. Mass Spectrom.*, **14**, 1470–1476.

Wierenga,R.K., Noble,M.E., Vriend,G., Nauche,S. and Hol,W.G. (1991) *J. Mol. Biol.*, **220**, 995–1015.

Williams,J.C., Zeelen,J.P., Neubauer,G., Vriend,G., Backmann,J., Michels,P.A., Lambeir,A.M. and Wierenga,R.K. (1999) *Protein Eng.*, **12**, 243–250.

Woenckhaus,J., Mao,Y. and Jarrold,M.F. (1997a) *J. Phys. Chem. B*, **101**, 847–851.

Woenckhaus,J., Hudgins,R.R. and Jarrold,M.F. (1997b) *J. Am. Chem. Soc.*, **119**, 9586–9587.

Xu,D., Tsai,C.J. and Nussinov,R. (1997) *Protein Eng.*, **10**, 999–1012.