Amino Acids (2012) 43:2457–2468
DOI 10.1007/s00726-012-1324-9

ORIGINAL ARTICLE

Wards in the keyway: amino acids with anomalous pK_as in calycins

Ivano Eberini · Cristina Sensi · Michele Bovi · Henriette Molinari · Monica Galliano · Franco Bonomi · Stefania Iametti · Elisabetta Gianazza

Received: 16 January 2012 / Accepted: 15 May 2012 / Published online: 30 May 2012

Abstract As a follow-up to our recent analysis of the electrostatics of bovine β-lactoglobulin (Eberini et al. in Amino Acids 42:2019–2030, 2011), we investigated whether the occurrence in the native structure of calycins—the superfamily to which β-lactoglobulin belongs—of amino acids with anomalous pK_as is an infrequent or, on the contrary, a common occurrence, and whether or not a general pattern may be recognized. To this aim, we randomly selected four calycins we had either purified from natural sources or prepared with recombinant DNA technologies during our previous and current structural and functional studies on this family. Their p*I*s vary over several pH units and their known functions are as diverse as carriers, enzymes, immunomodulators and/or extracellular chaperones. In our survey, we used both in silico prediction methods and in vitro procedures, such as isoelectric focusing, electrophoretic titration curves and spectroscopic techniques. By comparing the results under native conditions (no exposure of the proteins to chaotropic agents) to those after protein unfolding (in the presence of 8 M urea), a shift is observed in the pK_a of at least one amino acid per protein, which results in a measurable change in p*I*. Three types of amino acids are involved: Cys, Glu, and His, their position varies along the calycin sequence. Although no common mechanism may thus be recognized, we hypothesize that the 'normalization' of anomalous pK_as may be the phenomenon that accompanies, and favors, structural rearrangements such as those involved in ligand binding by these proteins. An interesting, if anecdotal, validation to this view comes from the behavior of human retinol binding protein, for which the p*I* of the folded and liganded protein is intermediate between those of the folded and unliganded and of the unfolded protein forms. Likewise, both solid (from crystallography) and solution state (from CD spectroscopy) data confirm that the protein undergoes structural rearrangement upon retinol binding.

Keywords Calycins · Electrostatics · Unfolding · Urea · Electrophoresis · Molecular modeling

Abbreviations

cl-BABP	Chicken liver bile acid-binding protein or chicken liver basic fatty acid binding protein
ETC	pH-mobility, or electrophoretic titration curves
IEF	Isoelectric focusing
IPG	Immobilized pH gradient
PGDS	Glutathione-independent, lipocalin-type prostaglandin D synthase or prostaglandin-H2 D-isomerase (EC = 5.3.99.2) or β-trace protein

Electronic supplementary material The online version of this article (doi:10.1007/s00726-012-1324-9) contains supplementary material, which is available to authorized users.

I. Eberini · C. Sensi · E. Gianazza (✉)
Gruppo di Studio per la Proteomica e la Struttura delle Proteine, Dipartimento di Scienze Farmacologiche, Università degli Studi di Milano, via Giuseppe Balzaretti 9, 20133 Milan, Italy
e-mail: elisabetta.gianazza@unimi.it

M. Bovi · H. Molinari
Dipartimento Scientifico e Tecnologico, Università degli Studi di Verona, Strada Le Grazie 15, 37134 Verona, Italy

M. Galliano
Dipartimento di Biochimica "Alessandro Castellani", Università degli Studi, via Taramelli 3B, 27100 Pavia, Italy

F. Bonomi · S. Iametti
Sezione di Biochimica, Dipartimento di Scienze Molecolari Agroalimentari, Università degli Studi di Milano, via Celoria 2, 20133 Milan, Italy

PMF	Potential of mean force
PTM	Post-translational modifications
SDS-PAGE	Polyacrylamide gel electrophoresis in presence of sodium dodecyl sulfate

Introduction

Overall amount and spatial distribution of electrostatic charge are critical parameters in protein function (Goldenberg and Steinberg 2010), yet these features are being experimentally assessed in very few cases. When isoelectric focusing (IEF) was first introduced, measuring the native p*I* was a somewhat standard step in the characterization of any newly purified protein (Righetti 1983; Righetti and Drysdale 1974). Restraints due to practical considerations (ion interference, protein solubility, staining sensitivity) have caused a progressive decrease of the number of focusing experiments on individual native proteins. Meanwhile, the analytical paradigm has shifted towards the fractionation of whole proteomes under denaturing and dissociating conditions; current protocols allow qualitative and quantitative assessment of a sample components through MS techniques after 2DE (IPG-SDS-PAGE) and even after just 1DE (SDS-PAGE) procedures (Ahrens et al. 2010).

As a result of these opposite trends, only in few cases the p*I*s under native and under denaturing conditions have been experimentally compared for any protein. Early literature on IEF was especially careful in dismissing possible artifacts, so the generalized finding of identity in p*I* between folded and unfolded proteins was stressed, and the only example of a large shift in p*I* quoted time after time was the one of serum albumin (Salaman and Williamson 1971; Ui 1971). More recently the concern has moved instead to the agreement between predicted and found for p*I* values under denaturing conditions. Coincidence between the two data has been proved for a large number of proteins (Bjellqvist et al. 1993) and the notion generalized to all proteins without post-translational modifications (PTM); conversely, divergence between predicted and found has been assumed to imply the presence of PTM (Bjellqvist et al. 1994; Gianazza 1995). Although not formalized in any way, the current assumption tends to stretch this conclusion—and the suggestions from the few old experimental data—by considering the p*I* of all proteins, under both native and denaturing conditions, to be identical to the values computed from compositional or sequence data.

We have reported recently a clearly measurable p*I* shift between native and denatured forms for the calycin β-lactoglobulin (Eberini et al. 2011). In the present report, a combination of in silico and in vitro techniques is used to analyze the electrostatics of four more calycins (only the former for a fifth calycin). p*I* shifts are observed upon protein unfolding and the reasons for the diversity from one protein to another in both direction and extent of these changes are clarified. Electrostatics evaluations carried out with reference to the known crystallographic structure of these proteins (or of a reliable model built in silico on the basis of homologous structures) are used to recognize which amino acid side chains are affected in each case. All calycins have an internal cavity designed to accommodate ligands (although the natural molecules actually interacting with some of them have not been identified). The presence in the native structure of all test proteins of amino acids with anomalous pK_a may suggest that—as already known for some calycins (Eberini et al. 2004, 2008a, b)—these residues may be involved in rearrangements that accompany and modulate ligand binding, with a combination of steric and electrostatic effects.

Materials and methods

List of the test calycins

We studied the behavior of four lipocalins of human origin—lipocalin 15, retinol binding protein, α_1-microglobulin, and glutathione-independent, lipocalin-type prostaglandin D synthase (or prostaglandin-H2 D-isomerase, EC = 5.3.99.2, or β-trace protein; h-PGDS)—and of one from *Gallus gallus*, chicken liver bile acid binding protein (or basic fatty acid binding protein; cl-BABP). In some cases, we compared our current findings with data previously obtained for bovine β-lactoglobulin (Eberini et al. 2011). Figure 1 shows the sequence and structure alignment of the six proteins.

Protein preparation and/or purification

Two calycins were purified from biological sources: human retinol binding protein from urine, human α_1-microglobulin both from urine and from amniotic fluid; two were produced by recombinant DNA technology: human lipocalin 15, and chicken liver bile acid binding protein.

For α_1-microglobulin, the purification procedure was similar to what already reported in Amoresano et al. (2000) and Sala et al. (2004). Starting from human amniotic fluid collected at weeks 16–18 of pregnancy, dialyzed samples were loaded on a high-performance Q-Sepharose column equilibrated with 6.25 mM 1,3-bis(tris(hydroxymethyl)methylamino)propane (BISTRIS-propane), pH 7.5, and eluted with a linear gradient between the starting condition and a 6.25 mM BISTRIS-propane, pH 9.5,

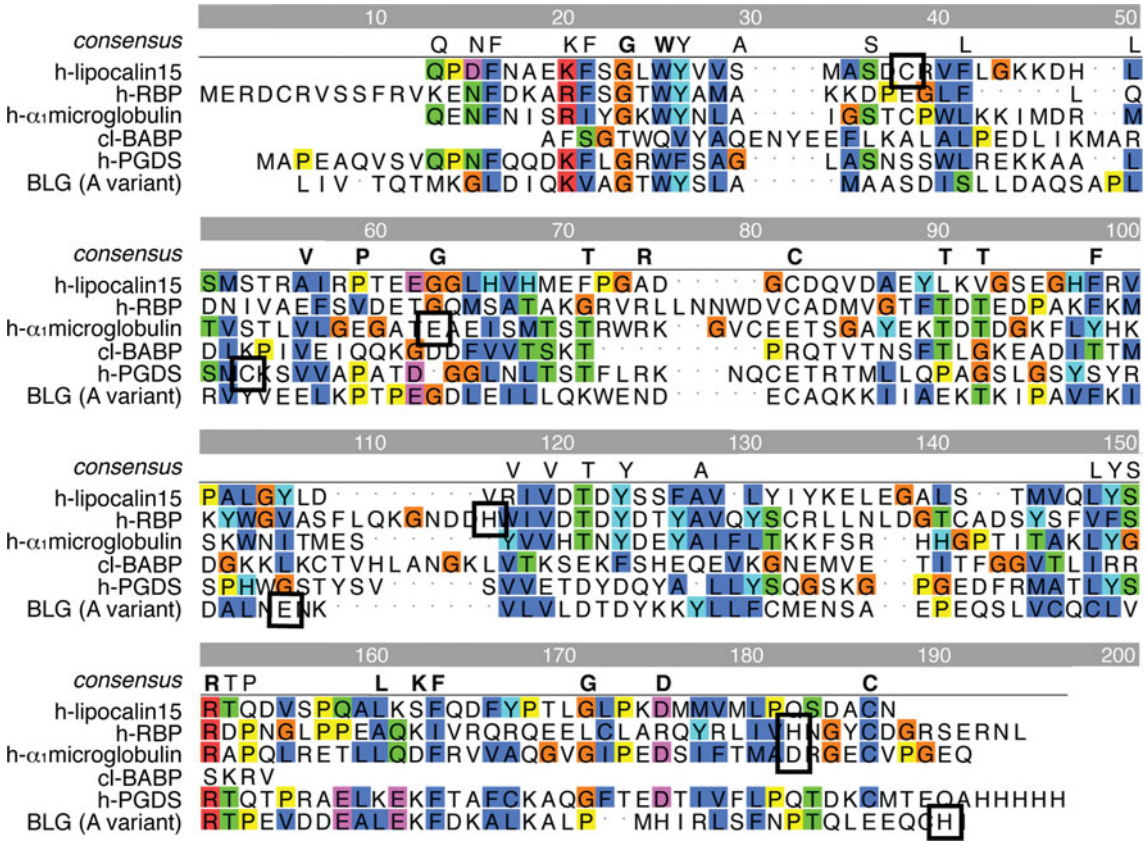

Fig. 1 Sequence and structure alignment of currently and previously investigated calycins. The amino acids whose pK_as are maximally affected by the transition between folded and unfolded form are *boxed*. *Color coding* (according to Clustal-X consensus calculation) highlights amino acid similarity, *consensus* identifies the amino acid most often found at the specified position

350 mM NaCl buffer. The α_1-microglobulin-containing fractions were fractionated by gel filtration on a Superdex 75 column with a 20 mM Tris/HCl, pH 8.0, 150 mM NaCl buffer. Starting from the urine of hemodialized patients, the protein was precipitated at 80 % of ammonium sulfate saturation; the precipitate was applied on a DEAE-cellulose column equilibrated with 50 mM sodium acetate, pH 5.8, and eluted with 300 mM NaCl. The eluted proteins were then gel filtered on a Sephadex G 100 column with a 20 mM Tris/HCl, pH 8.0, 150 mM NaCl buffer.

The procedure for the purification of retinol binding protein from urine was similar to the above, except that the DEAE-cellulose column was equilibrated with 50 mM sodium acetate, pH 4.8. Elution was with 100 mM NaCl until the absorbance was negligible, followed by the same buffer containing 300 mM NaCl (unpublished results). Two fractions containing retinol-binding protein eluted at different NaCl concentration from the ionic exchange column (100 and 300 mM, respectively) were processed in parallel, and for short are referred to in the following as 'low' and 'high salt'.

Recombinant cl-BABP was expressed as soluble protein in *E. coli* BL21 (DE3) bearing the recombinant plasmid pET24d, as reported in (Ragona et al. 2006). Protein expression was induced with 0.7 mM isopropyl β-D-1-thiogalactopyranoside. The cells were harvested and lysed in 50 mM Tris, 10 % sucrose, 1 mM EDTA, 10 mM 2-mercaptoethanol, pH 8.0. The supernatant was loaded on a DEAE-cellulose column equilibrated with 50 mM Tris/acetate, pH 7.8. Fractions containing cl-BABP were gel filtered on a Sephacryl S-100 HR column equilibrated with 50 mM Tris/HCl, pH 7.2, 200 M NaCl.

The plasmid containing the clone for human lipocalin 15 was obtained from the IMAGE consortium (RZPD, Deutsches Ressourcenzentrum fuer Genomforschung GmbH, ID number 7939738). The cDNA was amplified by PCR using primers designed to introduce restriction sites for *Bam*HI and *NdeI*; the fragment was then ligated into the expression vector pET15 respecting the reading frame. Protein synthesis was induced with isopropyl β-D-1-thiogalactopyranoside in *E. coli* BL21 cells transformed with the resulting vector. Under these conditions of subcloning, the protein was fused to a six histidine tag through its N terminus and could be affinity purified by passing the bacterial extracts through a nickel-Sepharose IMAC column. The tag was then removed due to the presence of a digestion site for thrombin (unpublished results).

Before use in electrophoresis and CD experiments, all protein samples were dialyzed against 20 mM Tris/HCl pH 7.0.

Isoelectric focusing

Immobilized pH gradients (IPG) (Bjellqvist et al. 1982), covering with a non-linear course at a range of 4–10 (Gianazza et al. 1985), were polymerized, washed and dried according to the standard protocols (Gianazza 2009; Righetti 1990). Strips of the gel were then re-swollen in urea solutions of different concentration. The gels were run at 15 °C for 15,000 V × h. Coomassie staining was carried out according to Righetti and Drysdale (1974).

pH-mobility curves

The pH-mobility (or electrophoretic titration) curves (ETC) (Gianazza et al. 1999; Righetti and Gianazza 1980; Rosengren et al. 1977) of the test calycins under varying conditions were evaluated by running the proteins across a wide, or narrow, pH range set by focused carrier ampholytes. Empty gels, 120 × 120 × 0.5 mm^3, with a T5 C4 matrix, were cast in a mold (Eberini et al. 2011), washed, dried and reswollen in solutions containing either 2 % w/v Ampholine 3.5–10, or 1.6 % w/v Ampholine 3.5–10 and 0.4 % w/v Pharmalyte 8–10.5, together with appropriate concentrations of urea. The first dimension run lasted 1, or 1.5, h at 15, or 4 °C, and the pH gradient was established in the gel by delivering a constant power of 3 W over an anode–cathode distance of 10 cm. The run of the test proteins, orthogonal to the direction of the pH gradient, was at 30 V/cm, with a duration increasing from 30 to 60 min depending on the protein and on absence/presence of additives. After extensive washing with 30:10:60 v:v:v ethanol:acetic acid:water to remove most carrier ampholytes, the protein pattern was stained with Coomassie Blue R250 according to (Righetti and Drysdale 1974).

Spectroscopy

Circular dichroism (CD) spectra for proteins in 20 mM Tris/HCl pH 7.0 were recorded on a Jasco J810 spectropolarimeter at 20 °C in 0.1-cm pathlength cells, and analyzed by means of the Jasco J800 software. Emission fluorescence spectra were recorded at 20 °C in a Perkin-Elmer LS-50 spectrofluorometer, with excitation at 335 nm and both excitation and emission slits set at 5 nm. Absorbance spectra were recorded at 20 °C in a Perkin-Elmer Lambda 2 instrument. In all cases, the temperature was controlled through a computer-driven, Peltier-equipped thermostatted cell holder.

Selection of reference crystal structures, model building

3D crystallographic structures of all the investigated proteins, but α_1-microglobulin and lipocalin 15, were downloaded from the RCSB PDB (http://www.rcsb.org) with the following identifiers: 1JYD for human retinol binding protein 4, 1TVQ for cl-BABP, and 2WWP for h-PGDS. Although a crystallographic structure for human lipocalin 15 is available from RCSB PDB (2XST), we decided to disregard it, as it contains 9 unsolved residues; we used instead a complete crystallographic structure recently solved in one of our labs (unpublished results). In h-PGDS, we mutated in silico into a Met residue Cys67, which is involved in disulfide bonds not solved in the crystallographic structures. Since no crystallographic structure is available for human α_1-microglobulin, a 3D structure was predicted through comparative modeling; we used the program Homology Model of the Homology module in the Molecular Operating Environment (MOE), version 2011.10, and made reference to lipocalin q83 (RCSB PDB ID: 1JZU.A) as template. The target sequence was aligned with the Align program of the Homology module with respect to a freezed pre-aligned protein set containing lipocalin q83 from *Coturnix coturnix*, prostaglandin D synthase homolog from *Danio rerio*, uncharacterized protein from *PTGDS* gene from *Takifugu rubripes*, CALII from *Gallus gallus*, LCN15 from *Homo sapiens*. This alignment, manually optimized, is reported in Supplementary Fig. 1. Since side chain orientation affects the pK_a values, we modeled the backbone, then sampled 100 different side chain orientations, at 300 K, with the Boltzmann-weighted randomized modeling procedure (Levitt 1992). The average model was used for the prediction of aminoacid pK_as.

Prediction of aminoacid pK_a

pK_a values for all the titratable residues were computed through the residue pK_a program of the Biopolymer module in the MOE suite. This program is based on the PROPKA method (Li et al. 2005), which evaluates the empirical shifts in pK_a values associated with the different types of interactions that each group may establish. Values can be computed very rapidly as a simple pairwise scoring function.

$$pK_a = pK_{aq} + \Delta pK_{des} + \Delta pK_{HB} + \Delta pK_{chgchg} \quad (1)$$

where pK_{aq} is the pK_a value of the free residue in aqueous solution, ΔpK_{des} is the desolvation component (both global and local desolvations are measured as a function of burial), ΔpK_{HB} is the hydrogen bonding contribution, and ΔpK_{chgchg} is the buried charge-charge interaction component.

The pK_a prediction algorithm is iterative. An initial set of pK_a values is computed then the charge-charge interactions and the hydrogen bond terms between groups with similar pK_a values are recomputed. As a rule, the group with the higher pK_a will lower that of the interacting partner, while its own pK_a will increase by a corresponding amount. Once this is done for all relevant pairwise interactions, the process is repeated until it converges.

Results and discussion

The calycins we selected for this survey are the ones that were most easily available in our laboratories, where they have been thoroughly investigated in the recent past also in collaborative studies (Beringhelli et al. 2002; Cairoli et al. 1994; Eberini et al. 2008a, b; Eliseo et al. 2007; Fogolari et al. 2005, 2000; Monaco et al. 1995, 1984; Nichesola et al. 2004; Ragona et al. 2003; Rasmussen et al. 2007). Ongoing research is still devoted to the novel features of those proteins as well as to newly identified members of this superfamily.

The proteins investigated may be taken as largely representative of the whole group. They are of interest for the approaches presented here, since, as computed from sequence data (http://www.uniprot.org/), their p*I*s vary over several pH units (Table 1), and their known functions are as diverse as carriers (Flower 1995, 1996), enzymes (Urade and Hayaishi 2000), immunomodulators (Logdberg and Wester 2000) and/or extracellular chaperones (Kanekiyo et al. 2007).

Figure 2a compares the focusing positions of four of the proteins listed in Table 1 on IPG strips swollen either in 15 % v/v glycerol or in 8 M urea. Two samples have been run for α_1-microglobulin: one purified from human urine and largely homogeneous, another from human amniotic fluid and heterogeneous as a result of N- and O-glycosylation. Recombinant cl-BABP failed to focus under equilibrium conditions (not shown), and therefore ETC was used for assessing its p*I* (Fig. 2b). ETC is a kinetic procedure that identifies a protein p*I* as the pH at which the migration curve crosses the sample application trench. The advantage of ETCs over equilibrium focusing is the fast sample separation in the already established pH gradient. This prevents artifactual findings due to slow migration with shallow $\delta u/\delta pH$ around p*I*. Moreover, the short migration time minimizes association/dissociation and/or rearrangement in susceptible proteins. Thus, values from ETCs may be more reliable than p*I*s determined under prolonged exposure to the IEF medium (Gianazza et al. 1999).

For all test calycins, different p*I*s are observed for the native and for the denatured protein, but the extent of the shift varies. In all test cases, the unfolded have a higher p*I* than the folded forms, whereas for β-lactoglobulin a lower p*I* was apparent upon unfolding [Table 1, entry with dark gray background; data from Eberini et al. (2011)].

To investigate why in their native state these calycins have more acidic p*I*s than expected from their amino acid composition, we carried out an in silico evaluation of the individual amino acid pK_as in their 3D structure, as in Eberini et al. (2011). Except for α_1-microglobulin, all of the selected proteins have been crystallized, some of them in one of our laboratories, and the coordinates of their solid state are either published in the RCBS PDB (http://www.rcsb.org/pdb/) or available to us as unpublished results.

For human α_1-microglobulin, we built a 3D model through comparative modeling, using lipocalin q83 as template. The lipocalins display unusually low levels of overall sequence conservation, with pairwise sequence identity often falling below 20 %, generally regarded as the threshold for reliable alignment (Flower 1996). Despite the low sequence identity between α_1-microglobulin and lipocalin q83 (20.4 %), we are confident about the accuracy of our model in the frame of the present study. Indeed, the direction of the observed p*I* shift is correctly predicted and its magnitude fits computation as closely as for the proteins whose 3D structure is experimentally known.

We computed the pK_a values for all the titrable residues of the test proteins. Detailed data are reported in

Table 1 p*I*s computed for currently and previously investigated calycins in their folded and unfolded states

Protein (UniProt entry)	Computed[a] from p*K*model[b] values (theoretically unfolded protein)	Computed[a] from MOE pK_a program values (folded protein)
h-Lipocalin 15 (Q6UWW0)	4.77	4.50
h-Retinol binding protein 4 (P02753)	5.49	4.82
h-α_1-Microglobulin (P02760)	6.95	6.73
cl-BABP (P80226)	9.00	8.68
h-PGDS (P41222)	8.79	8.44
Bovine β-lactoglobulin	4.76 (isoform A)	4.59 (isoform A)
	4.84 (isoform B)	4.68 (isoform B)

[a] Computed through a C++ program written in our lab

[b] As listed in Supplementary Tables S1–S5

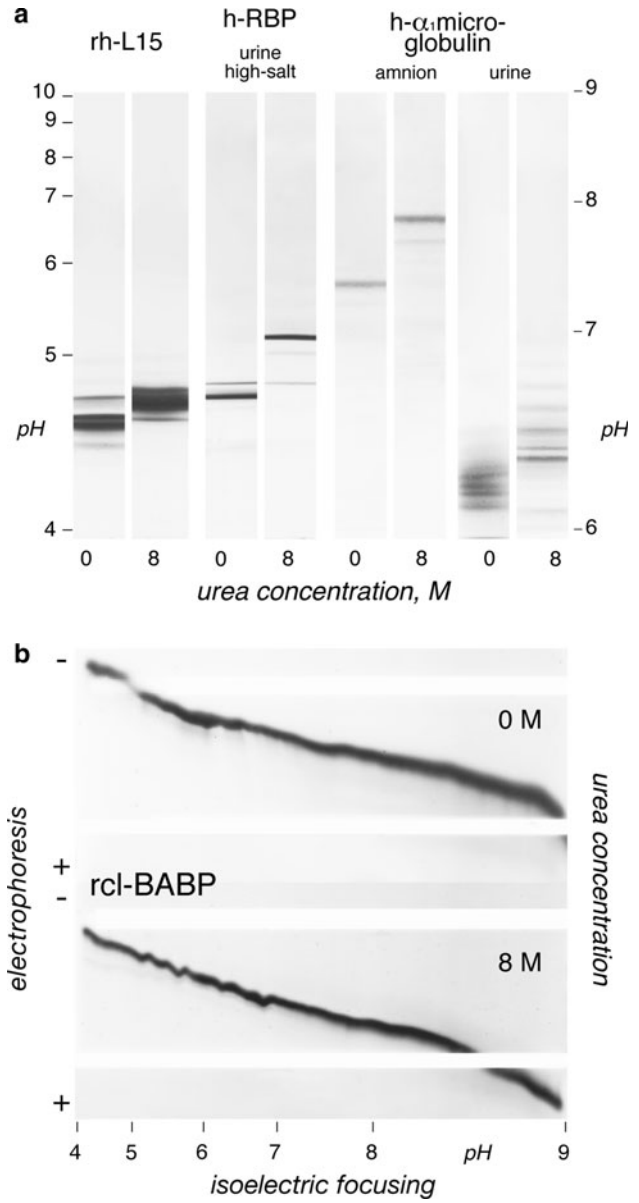

Fig. 2 **a** Isoelectric focusing under native (0 M urea) and denaturing conditions (8 M urea) of the test proteins on 4–10 NL IPG. L15 = lipocalin 15; RBP = human retinol binding protein; h = human; r = recombinant. pH gradient course is marked on the *left*. **b** Electrophoretic titration curve under native (0 M urea, *top*) and denaturing conditions (8 M urea, *bottom*) of chicken liver fatty acid binding protein

Supplementary Tables S1–S5. For each protein, we highlighted the amino acids that satisfy the following relationships:

$$(\mathrm{p}K_{\mathrm{a}_i\mathrm{folded}} < \text{theoretical p}I < \mathrm{p}K_{\mathrm{a}_i\mathrm{unfolded}}) \quad \mathrm{U} \quad (\mathrm{p}K_{\mathrm{a}_i\mathrm{unfolded}} < \text{theoretical p}I < \mathrm{p}K_{\mathrm{a}_i\mathrm{folded}}) \tag{2}$$

$$|\Delta \mathrm{p}K_{\mathrm{a}_i}| \geq 1 \tag{3}$$

Specifically, for lipocalin 15, the most relevant residue involved in the observed p*I* shift is Cys26, whose pK_a value varies from 9.00 (unfolded form) to 4.72 (native form). Very low pK_a values for cysteines have been described in enzyme catalytic sites, where their sulfide atom acts as nucleophile (Karala et al. 2010). Other residues involved, with a less extensive effect, are Glu47, Glu48, and Glu76; all of them are borderline for both relationships (2) and (3).

The p*I* shift in retinol binding protein can be explained by the change in pK_a values between folded and unfolded form for two histidine residues. Both His104 and His170 satisfy relationships (2) and (3), changing their pK_as from 6.50 (unfolded form) to, respectively, 1.11 and 3.99 (native form). However, we hypothesize that His170 has a lower influence on protein electrostatics, since—as already demonstrated for β-lactoglobulin (Eberini et al. 2011)—its pK_a value can be quite unstable because of end effects. The very low pK_a value for His104 is due to a buried conformation of its side chain, which entails relevant contributions from global (−3.22 p*K* units) and local (−2.17 p*K* units) desolvation of the group. Neither hydrogen bonds nor charge interactions are directly involved in this shift.

α_1-Microglobulin has a glutamic acid (Glu59), whose pK_a value shifts from 4.50 (unfolded form) to 6.70 (native form). Also, two borderline histidines (His 91 and His 105) could contribute to the p*I* shift, as observed for lipocalin 15.

cl-BABP has only one Cys residue (Cys80) satisfying both the relationships; its pK_a value varies from 9.00 (unfolded form) to 7.85 (folded form). The availability for this protein of NMR data produced in our labs (Pedo et al. 2009) allows the comparison between the experimental and computational pK_a value of His98 under native conditions. Both experimental and in silico procedures detect an especially low value, corresponding to 4.5–4.8 with the former approach and to 3.0 with the latter (Supplementary Table S4).

Supplementary Table S5 contains data about a fifth calycin, h-PGDS. The most affected residue is Cys43, whose pK_a changes from 9.00 (unfolded form) to 6.67 (native form), which is expected to greatly contribute to the observed p*I* shift from 8.44 to 8.79 (Table 1, entry with light gray background). No experimental validation could be obtained for these computed values, since the only soluble form that could be obtained for PGDS expressed as a recombinant protein required the mutation of Cys43 to Ala (unpublished data). Although only derived from computation, the evidence on h-PGDS electrostatics fits into a pattern common to the other test calycins.

Table 2 summarizes the above results, suggesting that three types of residues (Cys, Glu, and His) play the main role in all the p*I* shift effects.

Due to its large shift in p*I*, we decided to study in greater detail the properties of retinol binding protein. A first issue relates to the occurrence of two bands (Fig. 2) with a different behavior when the protein is run in 8 M urea. The

Table 2 Summary of the amino acids whose pK_as are mostly affected in the folded structures of currently and previously investigated calycins

Protein	Glu	His	Cys
h-Lipocalin 15			26
h-Retinol binding protein		104 (170)[a]	
h-α_1-Microglobulin	59		
cl-BABP			80
h-PGDS			43
Bovine β-lactoglobulin	44, 89	(161)[a]	

[a] The pK_a of the amino acids in the C-terminal part of the protein is strongly influenced by end effects, as shown in Eberini et al. (2011) for β-lactoglobulin

staining intensity of the one with higher p*I* hardly changes, whereas the one with lower p*I* drastically decreases. MS analysis confirmed that both bands correspond to retinol binding protein and did not provide evidence for (differential) PTM (not shown).

The two fractions obtained from ion-exchange chromatography during the purification of this protein were therefore compared. Spectrophotometric and spectrofluorimetric data (Supplementary Fig. 3a and 3b) hint to the presence in both cases of a high-conjugation unsaturated ligand, i.e. retinol or a related compound, bound at a much higher concentration in the low-salt fraction. Far-UV CD spectra of the two protein fractions have similar shape (Supplementary Fig. 3c), but, in the near-UV CD spectra, the low-salt fraction has higher molar ellipticity and much more prominent features between 250 and 300 nm than the high-salt fraction (Supplementary Fig. 3d). Also the proportion of the two bands in IEF is different, as the more alkaline prevails in the low-salt fraction and the more acidic in the high-salt one (Fig. 3). When the two samples were run in the presence of increasing urea concentrations, between 3 and 5 M, the intensity of the higher p*I* band hardly changed, whereas that of the lower p*I* band steadily decreased to become negligible at the highest urea concentration. At the same time a much higher p*I* band became evident, as already observed in Fig. 2.

Together, this evidence allows the identification of the bands, from anode to cathode, as folded apo-, folded holo- and unfolded retinol binding protein. In agreement with several findings, also from our laboratories (Beringhelli et al. 2002; Eberini et al. 2006; Evans and Williams 1980; Ewbank and Creighton 1993; Santagati et al. 1997), these data prove the higher resistance to unfolding of the liganded than of the unliganded form as demonstrated also with the naturally occurring ligands in β-lactoglobulin (Barbiroli et al. 2011).

Supplementary Fig. 3 shows the ETCs of the two fractions of retinol binding protein we have analyzed, in presence of urea concentrations increasing stepwise from 0 to 8 M. The tracing associated with the holo-form hardly changes in intensity at any pH or chaotrope concentration. At acidic pH the tracing associated with the folded apo-form decreases in intensity, in favor of the unfolded protein, starting from 4 M urea (arrow); conversely, at alkaline pH, the unfolding is still incomplete at the end of the duration of the run (1 h).

Figure 4 shows in greater detail the ETCs in 8 M urea. Due to the incomplete unfolding of retinol binding protein under the selected running conditions, it is possible to compare side by side the tracings of the folded and unfolded apo-protein. The maximal divergence between the slopes of the two curves is observed in the acidic-to-neutral region, i.e. at a pH corresponding to the pK_a of 'model' histidine residues. This finding is in close agreement with the predictions of in silico computations.

A change in p*I* upon binding of a neutral ligand as seen for retinol binding protein is uncommon but not unique: apo-Gc globulin has a p*I* of 5.10, its complex with vitamin D a p*I* of 4.95 (Constans et al. 1980), and the difference in behavior between the two forms is large enough to allow for a clear-cut separation by pseudo-affinity chromatography (Chapuis-Cellier et al. 1982; Gianazza and Arnaud 1982). Different mobility between apo- and holo-form in electrophoresis had already been reported for retinol

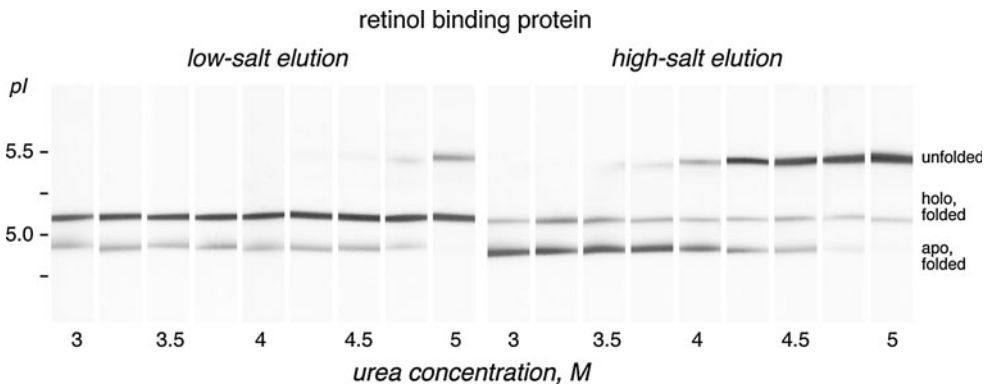

Fig. 3 Isoelectric focusing in presence of increasing urea concentrations (from 3 to 5 M, through 0.25 M steps) of two fractions of retinol binding protein purified from urine. Crop from 4 to 10 NL IPGs; pH gradient course on the *left*. Identification of the main protein bands on the *right*

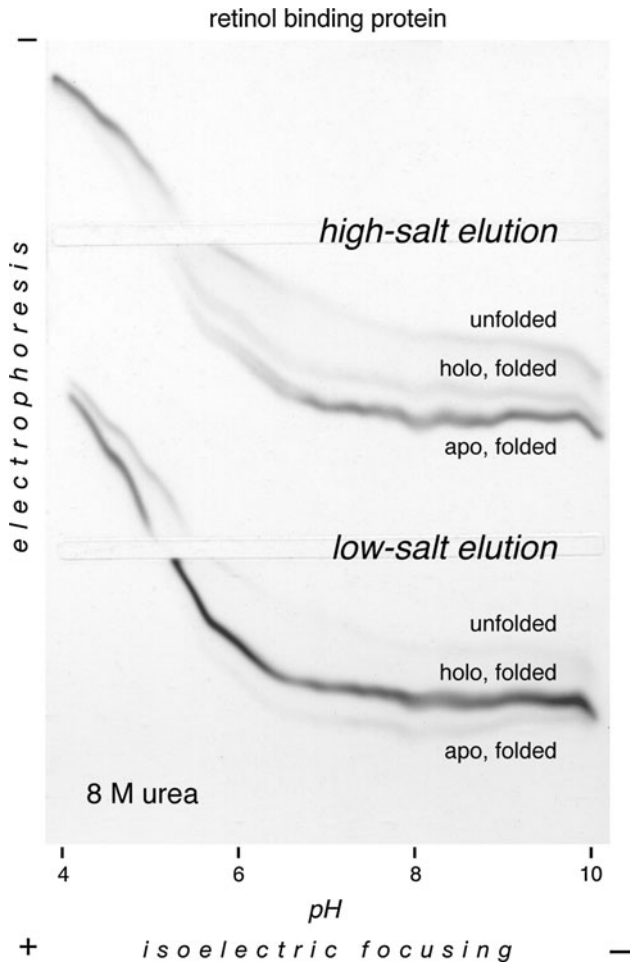

Fig. 4 ETC of retinol binding protein across a transverse pH gradient in presence of 8 M urea. Polarity as marked. Identification of the main protein bands on the *right*

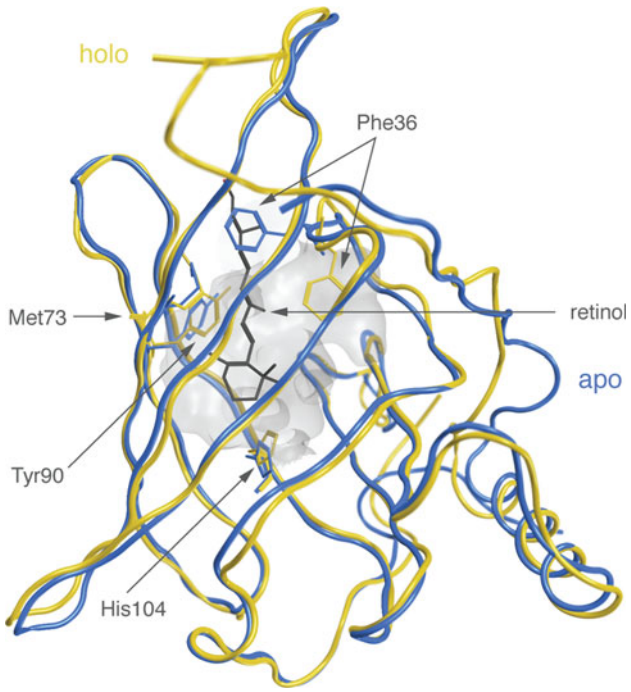

Fig. 5 Orientation of the amino acids whose pK_as most extensively shift between folded and unfolded form in apo- (*blue*) and holo- (*yellow*) h-retinol binding protein. The apo-form corresponds to the crystal with PDB ID 1BRQ, the holo-form to 1QAB and contains bound retinol. The protein backbone is rendered as *ribbon*, the amino acid side chains and the ligand retinol as *balls* and *sticks*

binding protein; the holo-form had been identified through its fluorescence and the apo-form from its increase after solvent extraction (Fex and Hansson 1979).

Among the possible causes leading to a p*I* shift upon ligand binding, two seem more likely: (1) a direct steric interference or dielectric perturbation on amino acids in/near the binding site, (2) an indirect effect mediated by conformational rearrangement (similar to the induced fit first described for enzymes bound to their substrates) and affecting amino acids in any position on the protein.

In the case of retinol binding protein, the hypothesis of a structural rearrangement is supported by the difference between the CD spectra of low- and high-salt fractions (Supplementary Fig. 2c and 2d) that correspond to holo- and apo-protein, respectively. Crystallographic data on retinol binding protein demonstrate that the main difference between unliganded and liganded forms is at the level of the entrance to the calyx (specifically, between amino acid residues 34 and 37) (Zanotti et al. 1993). The large concerted motions of the backbone are influenced by the presence of the ligand as shown by molecular dynamics simulation and essential dynamics analysis (Elenewski and Hackett 2010), although these features cannot be observed in short simulations (Chau et al. 1999). Steered molecular dynamics simulations and free energy methods demonstrate conformational rearrangement during retinol unbinding that involves the Phe36-Met73-Tyr90 ring (see Fig. 5). Computation of the potential of mean force (PMF) for unbinding of retinol from the complex and PMF decomposition through adaptive biasing force calculations into contributions from constituent interactions allows to evaluate a substantial involvement of electrostatic, as well as van der Waals, interactions to drive the intermediate interaction into a stable configuration (Elenewski and Hackett 2010).

To test whether the changes in conformation upon ligand binding documented by crystallographic structures do influence the pK_a of any of the amino acids, we repeated the computation on all the apo- and holo-forms of h-retinol binding protein available in RCBS database (http://www.rcsb.org/pdb/). The list of such structures and the results from computation are listed in Table 3. The change in pK_a (ΔpK_a) between folded and unfolded forms does not significantly differ between apo- and holo-h-retinol binding

Table 3 Differences between the pK_as of two affected amino acids, His104 and His 170, in the shift from the folded to the unfolded structure of h-retinol binding protein as a function of the liganded/unliganded status in various crystal structures

PDB ID	Ligand	$\Delta pK_{a\ His104}$	$\Delta pK_{a\ His170}$
1BRQ	None (apo form)	5.17	2.42
1JYD	None (apo form)	5.39	2.51
1JYJ	None (apo form)	5.68	3.03
1QAB	Retinol	5.18	2.90
1BRP	Retinol	5.47	4.46
1RBP	Retinol	5.54	1.49
1RLB	Retinoic acid	5.34	1.64
2WQA	Oleic acid	5.61	2.04
2WQ9	Oleic acid	5.49	2.40
2WR6	Linoleic acid	5.64	2.29
3BSZ	Retinol	5.53	1.74
3FMZ	$C_{20}H_{19}F_3N_2O_3$	5.81	1.99

protein, either for His104 or for His170. However, when the detailed structure of the binding site is displayed (Fig. 5), the relationship of His104 to the β-iononic ring of retinol in the holo-form becomes evident. So, while the structural rearrangement that accompanies ligand binding only marginally involves this amino acid, a change in its microenvironment does occur. This modification is not taken into account by our computational approach, mainly because it is tightly related to the ligand and not to the protein properties.

Overview: a unifying hypothesis?

Conversion between different functional states of a protein (unliganded/liganded, inactive/active) often occurs through conserved mechanisms. Some of such cases involve broadly defined processes such as for instance the activation of many proteases through proteolytic mechanisms. However, identical mechanisms have been detected in some protein families that are conserved down to the molecular details for most/all family members. This is for instance the case of the proteolytic cleavage and conformational rearrangement occurring in serpins during their interaction with the target proteases (Gettins 2002; Huntington 2006).

The peculiarities in the electrostatics of the native states of some calycins, as we have described in the present report, do not feature an identical behavior of specific amino acids. Both the chemical nature (Cys, Glu, and His) and the position (Fig. 6 and boxed items in Fig. 1) of the amino acids with anomalous pK_a vary as do the amplitude of the pK_a shifts and their compound effect on pI. We like to stress once more that, although the number of proteins analyzed is a small percentage of over 150 members of the calycin superfamily, the inclusion criteria were not biased by preference for specific chemico-physical or biological properties. Due to their chance extraction from the whole group, the repeated finding of anomalous pK_as seems to imply a widespread, if not a constant, occurrence.

One possible reason for such a phenomenon is the equilibrium between the local effects aimed at stabilizing side chains with anomalous pK_a values and the reorganization energy of the same side chains during their rearrangement towards more favorable energetic conformations. One of such occurrences may be ligand binding, which is known to involve concerted motions in all the calycins characterized so far (see for instance Eberini et al. (2008a, b)). Conformational transitions finely tune the protein binding abilities according to the prevailing milieu, as already described for bovine β-lactoglobulin (Beringhelli et al. 2002; Eberini et al. 2004). This feature seems to be linked to the anomalous pK_a values of some key titrable residues.

Whether this is the case also for other proteins of the same family (and/or other ligands) would require a series of planned experiments. However, an experimental proof could be obtained only under those conditions (especially, pH) that allow the complex to be stable throughout the duration of the run with equilibrium techniques such as IEF or at least for a long enough time as to influence the overall migration with kinetic-based techniques such as ETC.

Yet another point of interest is the possibility that the observed changes in pI, either related to ligand binding or to loss/gain of structural features, may alter the interaction pattern(s) of the involved protein(s) with physiological or pathological partners, and thus affect more deeply the functional features of systems which are much more complex than the protein itself. This could especially be true of those calycins that are thought to have a function as molecular chaperones (Kanekiyo et al. 2007).

Together, these considerations leave us with question marks in our writing and with questions still to be addressed with our experimental work.

In this specific application, computational tools have demonstrated their value in the treatment of protein properties at atomic resolution, with a very satisfactory agreement between experimental evidence and computational data. However, it must be appreciated that computations make reference to model patterns and imply a number of approximations so that the reliability of the results, which is excellent for the description of general trends, somewhat decreases for more detailed accounts. Table 3 provides the evidence of the extent of result variability when reference is made to different crystal structures even of the same apoprotein or of the same protein–ligand complex. A further complication comes from the relative motions of different parts of the protein when solvated, which could be

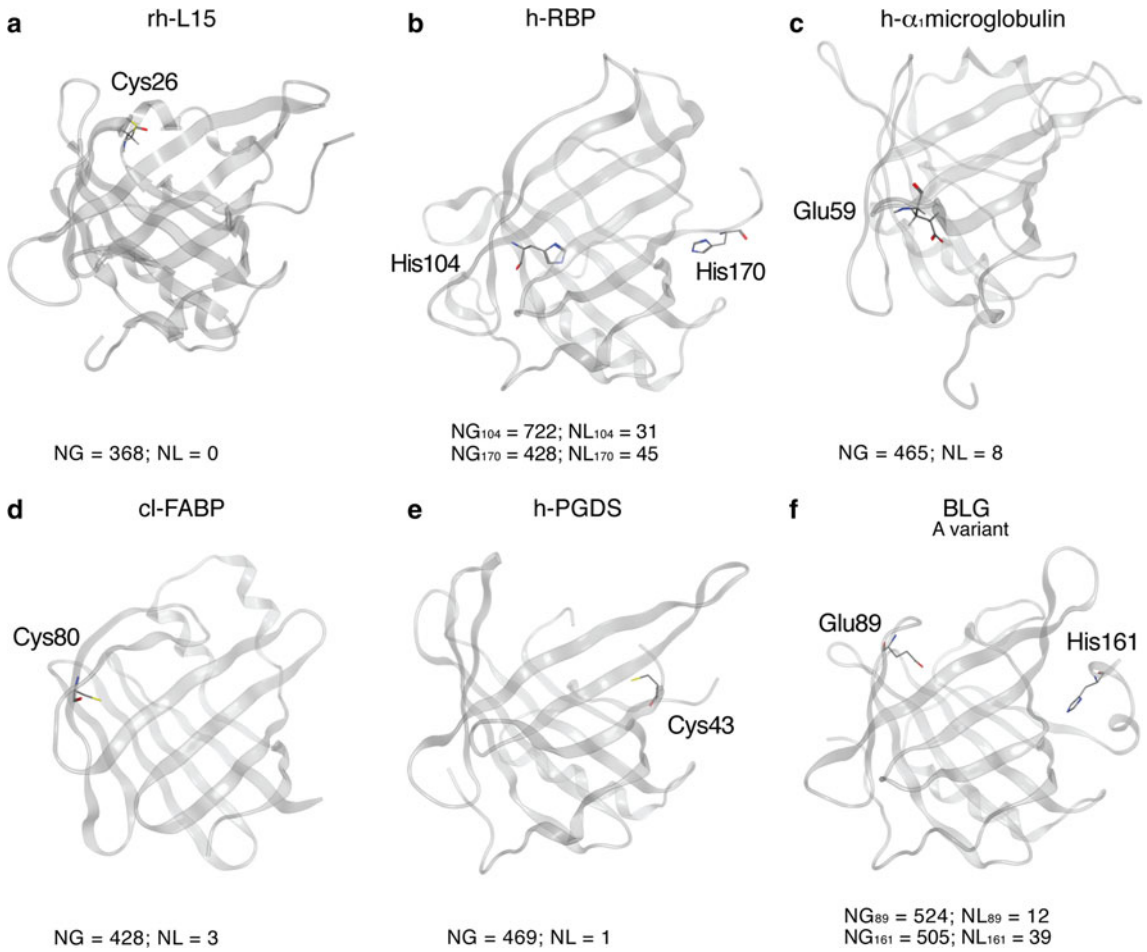

Fig. 6 Position of the amino acids whose pK_as most extensively shift between folded and unfolded form of the test calycins. Two parameters are marked for each amino acid: NGlob = number of atoms within a 15.5 Å radius (it is a measure of bulk burial; atoms with NGlob >400 are considered to be buried), and NLoc = number of atoms within a local radius (it varies from 3.5 to 6.0 Å and is used to estimate the solvent-accessible surface of the group). Further detail in Supplementary Tables S1–S5

underestimated in crystals, and may allow substantial changes in pK_a for individual amino acids as demonstrated for the highly flexible C-terminal region of β-lactoglobulin (Eberini et al. 2011). Finally, as confirmed by our data on apo- and holo-h-retinol binding protein (Table 3), the computation of pK_as focuses on protein sequence and thus misses the contribution from ligands, when bound in a complex. For these reasons, results from simulations should never be accepted uncritically, but rather taken as guides either to interpret experimental results or to plan experimental validation to computational results.

Conflict of interest None.

References

Ahrens CH, Brunner E, Qeli E, Basler K, Aebersold R (2010) Generating and navigating proteome maps using mass spectrometry. Nat Rev Mol Cell Biol 11:789–801

Amoresano A, Minchiotti L, Cosulich ME, Campagnoli M, Pucci P, Andolfo A, Gianazza E, Galliano M (2000) Structural characterization of the oligosaccharide chains of human alpha1-microglobulin from urine and amniotic fluid. Eur J Biochem 267:2105–2112

Barbiroli A, Bonomi F, Ferranti P, Fessas D, Nasi A, Rasmussen P, Iametti S (2011) Bound fatty acids modulate the sensitivity of bovine beta-lactoglobulin to chemical and physical denaturation. J Agric Food Chem 59:5729–5737

Beringhelli T, Eberini I, Galliano M, Pedoto A, Perduca M, Sportiello A, Fontana E, Monaco HL, Gianazza E (2002) pH and ionic strength dependence of protein (un)folding and ligand binding to bovine β-lactoglobulins A and B. Biochemistry 41:15415–15422

Bjellqvist B, Ek K, Righetti PG, Gianazza E, Görg A, Postel W, Westermeier R (1982) Isoelectric focusing in immobilized pH gradients: principles, methodology and some applications. J Biochem Biophys Methods 6:317–339

Bjellqvist B, Hughes GJ, Pasquali C, Paquet N, Ravier F, Sanchez JC, Frutiger S, Hochstrasser D (1993) The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. Electrophoresis 14:1023–1031

Bjellqvist B, Basse B, Olsen E, Celis JE (1994) Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points

correlate with polypeptide compositions. Electrophoresis 15:529–539
Cairoli S, Iametti S, Bonomi F (1994) Reversible and irreversible modifications of beta-lactoglobulin upon exposure to heat. J Protein Chem 13:347–354
Chapuis-Cellier C, Gianazza E, Arnaud P (1982) Interaction of group-specific component (vitamin D-binding protein) with immobilized Cibacron Blue F3-GA. Biochim Biophys Acta 709:353–357
Chau PL, van DM Aalten, Bywater RP, Findlay JB (1999) Functional concerted motions in the bovine serum retinol-binding protein. J Comput Aided Mol Des 13:11–20
Constans J, Viau M, Gouaillard C, Bouissou C, Clerc A (1980) In: Radola BJ (ed) Electrophoresis '79. W. deGruyter, Berlin, pp 701–709
Eberini I, Baptista AM, Gianazza E, Fraternali F, Beringhelli T (2004) Reorganization in apo- and holo-β-lactoglobulin upon protonation of Glu89: molecular dynamics and pKa calculations. Proteins 54:744–758
Eberini I, Fantucci P, Guerini Rocco A, Gianazza E, Galluccio L, Maggioni D, Ben ID, Galliano M, Mazzitello R, Gaiji N, Beringhelli T (2006) Computational and experimental approaches for assessing the interactions between the model calycin β-lactoglobulin and two antibacterial fluoroquinolones. Proteins 65:555–567
Eberini I, Guerini Rocco A, Ientile AR, Baptista AM, Gianazza E, Tomaselli S, Molinari H, Ragona L (2008a) Conformational and dynamics changes induced by bile acids binding to chicken liver bile acid binding protein. Proteins 71:1889–1898
Eberini I, Rocco AG, Mantegazza M, Gianazza E, Baroni A, Vilardo MC, Donghi D, Galliano M, Beringhelli T (2008b) Computational and experimental approaches assess the interactions between bovine beta-lactoglobulin and synthetic compounds of pharmacological interest. J Mol Graph Model 26:1004–1013
Eberini I, Sensi C, Barbiroli A, Bonomi F, Iametti S, Galliano M, Gianazza E (2011) Electrostatics of folded and unfolded bovine beta-lactoglobulin. Amino Acids 42:2019–2030
Elenewski JE, Hackett JC (2010) Free energy landscape of the retinol/serum retinol binding protein complex: a biological host-guest system. J Phys Chem B 114:11315–11322
Eliseo T, Ragona L, Catalano M, Assfalg M, Paci M, Zetta L, Molinari H, Cicero DO (2007) Structural and dynamic determinants of ligand binding in the ternary complex of chicken liver bile acid binding protein with two bile salts revealed by NMR. Biochemistry 46:12557–12567
Evans RW, Williams J (1980) The electrophoresis of transferrins in urea/polyacrylamide gels. Biochem J 189:541–546
Ewbank JJ, Creighton TE (1993) Structural characterization of the disulfide folding intermediates of bovine a-lactalbumin. Biochemistry 32:3694–3707
Fex G, Hansson B (1979) Retinol-binding protein from human urine and its interaction with retinol and prealbumin. Eur J Biochem FEBS 94:307–313
Flower DR (1995) Multiple molecular recognition properties of the lipocalin protein family. J Mol Recognit 8:185–195
Flower DR (1996) The lipocalin protein family: structure and function. Biochem J 318:1–14
Fogolari F, Ragona L, Licciardi S, Romagnoli S, Michelutti R, Ugolini F, Molinari H (2000) Electrostatic properties of bovine β-lactoglobulin. Proteins 39:317–330
Fogolari F, Moroni E, Wojciechowski M, Baginski M, Ragona L, Molinari H (2005) MM/PBSA analysis of molecular dynamics simulations of bovine beta-lactoglobulin: free energy gradients in conformational transitions? Proteins 59:91–103
Gettins PG (2002) Serpin structure, mechanism, and function. Chem Rev 102:4751–4804
Gianazza E (1995) Isoelectric focusing as a tool for the evaluation of post-translational processing and chemical modifications of proteins. J Chromatogr 705:67–87
Gianazza E (2009) Casting immobilized pH gradients. In: Walker JM (ed) The protein protocol handbook. Humana Press, Totowa, pp 305–322
Gianazza E, Arnaud P (1982) A general method for the fractionation of plasma proteins: pseudo-ligand affinity chromatography on immobilized Cibacron Blue F3-GA. Biochem J 201:129–136
Gianazza E, Giacon P, Sahlin B, Righetti PG (1985) Non-linear pH courses with immobilized pH gradients. Electrophoresis 6:53–56
Gianazza E, Castiglioni S, Eberini I (1999) Low-tech electrophoresis, small but beautiful, and effective: electrophoretic titration curves of proteins. Electrophoresis 20:1325–1338
Goldenberg NM, Steinberg BE (2010) Surface charge: a key determinant of protein localization and function. Cancer Res 70:1277–1280
Huntington JA (2006) Shape-shifting serpins—advantages of a mobile mechanism. Trends Biochem Sci 31:427–435
Kanekiyo T, Ban T, Aritake K, Huang ZL, Qu WM, Okazaki I, Mohri I, Murayama S, Ozono K, Taniike M, Goto Y, Urade Y (2007) Lipocalin-type prostaglandin D synthase/beta-trace is a major amyloid beta-chaperone in human cerebrospinal fluid. Proc Natl Acad Sci USA 104:6412–6417
Karala AR, Lappi AK, Ruddock LW (2010) Modulation of an active-site cysteine pKa allows PDI to act as a catalyst of both disulfide bond formation and isomerization. J Mol Biol 396:883–892
Levitt M (1992) Accurate modeling of protein conformation by automatic segment matching. J Mol Biol 226:507–533
Li H, Robertson AD, Jensen JH, Vacca JP (2005) Very fast empirical prediction and rationalization of protein pKa values. Proteins 61:704–721
Logdberg L, Wester L (2000) Immunocalins: a lipocalin subfamily that modulates immune and inflammatory responses. Biochim Biophys Acta 1482:284–297
Monaco HL, Zanotti G, Ottonello S, Berni R (1984) Crystallization of human plasma apo-retinol-binding protein. J Mol Biol 178:477–479
Monaco HL, Rizzi M, Coda A (1995) Structure of a complex of two plasma proteins: transthyretin and retinol-binding protein. Science 268:1039–1041
Nichesola D, Perduca M, Capaldi S, Carrizo ME, Righetti PG, Monaco HL (2004) Crystal structure of chicken liver basic fatty acid-binding protein complexed with cholic acid. Biochemistry 43:14072–14079
Pedo M, D'Onofrio M, Ferranti P, Molinari H, Assfalg M (2009) Towards the elucidation of molecular determinants of cooperativity in the liver bile acid binding protein. Proteins 77:718–731
Ragona L, Fogolari F, Catalano M, Ugolini R, Zetta L, Molinari H (2003) EF loop conformational change triggers ligand binding in β-lactoglobulins. J Biol Chem 278:38840–38846
Ragona L, Catalano M, Luppi M, Cicero D, Eliseo T, Foote J, Fogolari F, Zetta L, Molinari H (2006) NMR dynamic studies suggest that allosteric activation regulates ligand binding in chicken liver bile acid-binding protein. J Biol Chem 281:9697–9709
Rasmussen P, Barbiroli A, Bonomi F, Faoro F, Ferranti P, Iriti M, Picariello G, Iametti S (2007) Formation of structured polymers upon controlled denaturation of beta-lactoglobulin with different chaotropes. Biopolymers 86:57–72
Righetti PG (1983) Isoelectric focusing: theory, methodology and applications. Elsevier, Amsterdam
Righetti PG (1990) Immobilized pH gradients: theory and methodology. Elsevier, Amsterdam
Righetti PG, Drysdale JW (1974) Isoelectric focusing in gels. J Chromatogr 98:271–321

Righetti PG, Gianazza E (1980) pH-mobility curves of proteins by isoelectric focusing combined with electrophoresis at right angle. In: Radola BJ (ed) Electrophoresis '79. W. de Gruyter, Berlin, pp 23–38

Rosengren Å, Bjellqvist B, Gasparic V (1977) A simple method of choosing optimum pH conditions for electrophoresis. In: Radola BJ, Graesslin D (eds) Electrofocusing and isotachophoresis. W. deGruyter, Berlin, pp 165–171

Sala A, Campagnoli M, Perani E, Romano A, Labo S, Monzani E, Minchiotti L, Galliano M (2004) Human alpha-1-microglobulin is covalently bound to kynurenine-derived chromophores. J Biol chem 279:51033–51041

Salaman MR, Williamson AR (1971) Isoelectric focusing of proteins in the native and denatured states. Anomalous behaviour of plasma albumin. Biochem J 122:93–99

Santagati S, Gianazza E, Agrati P, Vegeto E, Patrone C, Pollio G, Maggi A (1997) Oligonucleotide squelching reveals the mechanism of estrogen receptor autologous down-regulation. Mol Endocrinol 11:938–949

Ui N (1971) Isoelectric points and conformation of proteins. I Effect of urea on the behavior of some proteins in isoelectric focusing. Biochim Biophys Acta 229:567–581

Urade Y, Hayaishi O (2000) Biochemical, structural, genetic, physiological, and pathophysiological features of lipocalin-type prostaglandin D synthase. Biochim Biophys Acta 1482:259–271

Zanotti G, Ottonello S, Berni R, Monaco HL (1993) Crystal structure of the trigonal form of human plasma retinol-binding protein at 2.5 A resolution. J Mol Biol 230:613–624