Supporting information

Protein-associated cation-clusters in aqueous arginine solutions

Vincent Vagenende[1]*, Alvin X. Han[1], Monika Mueller[1] and Bernhardt L. Trout[2]

(1) Bioprocessing Technology Institute, A*STAR (Agency for Science, Technology and Research), 20 Biopolis Way #06-01 Centros, Singapore 138668

(2) Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue Cambridge, E19-502b, Massachusetts 02139, USA

Supporting Methods

Materials

Hen egg-white lysozyme (L6876), α-chymotrypsinogen A (C4879), carbonic anhydrase (C2522), ovalbumin (A7641), bovine serum albumin (BSA; A7906), L-arginine hydrochloride (A5131) and MES (M3671) were obtained from Sigma-Aldrich. Ammonium sulfate (101217), sodium chloride (106404), sodium azide (10668), sodium dihydrogen phosphate monohydrate (106346), di-sodium hydrogen phosphate dihydrate (106576) and EDTA (324503) were obtained from Merck-Chemicals. Blue dextran 2000 (17-0360-01) was obtained from GE-Healthcare.

Size-exclusion chromatography

The effects of arginine and NaCl on protein size were investigated by size-exclusion chromatography (SEC) experiments in buffer A (50 mM MES, pH 6.5, 5mM EDTA, 0.05% sodium azide (w/v)) at increasing concentrations of arginine or NaCl (0 – 2 M). SEC experiments were carried out on a Shimadzu Class-VP HPLC system with a Superdex 75 PC 3.2/30 column (GE Healthcare). Prior to protein loading, the system was pre-equilibrated with 2 column volumes of buffer, and a flow rate of 0.04 mL min^{-1} was maintained throughout the experiment. Protein samples were prepared by dissolving them in the respective buffers to a concentration of 10 mg mL^{-1}. Protein sample (10 μL) was loaded onto the column, and protein elution was monitored by UV absorbance at 280nm for the next 85 minutes. Each experiment was run in duplicate.

Elution times were determined for five standard protein samples (BSA – 66.3 kDa, ovalbumin – 45 kDa, carbonic anhydrase – 31 kDa, α-chymotrypsinogen A – 25.6 kDa and lysozyme – 14.3 kDa) in Buffer A, and plotted against the stated protein's molecular weight (Mw) on a double logarithmic scale graph (Figure S1). Linear regression resulted in a good fit ($R^2 > 0.99$) and the corresponding equation was used to calculate apparent protein radii from

the respective elution times by assuming proteins as spheres with average density of 1.37 kg m^{-3} (*1, 2*) (Table S1, Table S2, Table S3, Table S4 and Table S5).

The possibility that arginine and NaCl effects on protein elution times result from solvent-induced changes of the chromatographic matrix was investigated by characterizing the excluded and total volume of the chromatographic matrix at various arginine and NaCl concentrations (Table S6). We found no significant effects of arginine and NaCl on the excluded volume of the chromatographic matrix over the investigated concentration range (0 – 2 M). Elution times of the total volume marker (acetone) increased moderately in the presence of arginine and NaCl. NaCl-induced increments of elution times of the total volume marker are much smaller than NaCl-induced increments of protein elution times, and arginine-induced increments of elution times of the total volume markers are opposite to arginine-induced changes of protein elution times (Table S1, Table S2, Table S3, Table S4 and Table S5). The observed changes in protein elution times can therefore not be attributed to solvent-induced changes of the chromatographic matrix.

The increase of the elution time of acetone at 2 M ArgHCl by about 2 min (Table S6) is substantial with respect to the arginine-induced decrease of the elution time of lysozyme at 2 M ArgHCl by about 5 min (Table S1). This suggests that the arginine-induced increment of total column volume could partially offset the arginine-induced decrease of protein elution times because of increasing protein hydrodynamic radii. The reported increases of apparent protein hydrodynamic radius, which do not account for solvent-induced changes of the total volume, should therefore be considered as lower estimates of the actual increases of protein hydrodynamic radius at higher arginine concentrations.
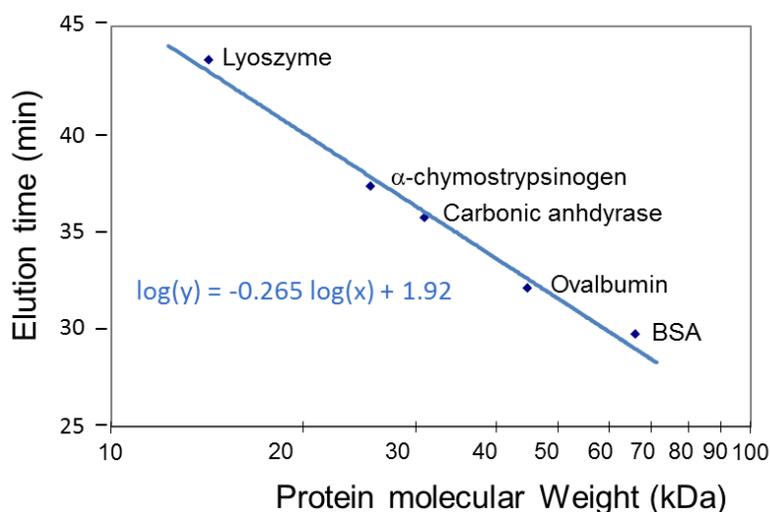

Figure S1: Elution times of five standard proteins of increasing molecular weight and the corresponding fit.

Table S1: Size-exclusion chromatography elution times and corresponding apparent protein radii of lysozyme.

ArgHCl [M]	Elution Time [min.]	Protein radius [Å]	NaCl [M]	Elution Time [min.]	Protein radius [Å]
0	42.8 ± 0.2	15.7 ± 0.1	0	42.8 ± 0.2	15.7 ± 0.1
0.25	42.6 ± 0.0	15.7 ± 0.0	0.25	44.1 ± 0.0	15.1 ± 0.0
0.5	41.5 ± 0.0	16.3 ± 0.0	0.5	44.4 ± 0.0	15.0 ± 0.0
0.75	40.1 ± 0.0	17.0 ± 0.0	0.75	44.6 ± 0.0	14.9 ± 0.0
1	39.7 ± 0.0	17.2 ± 0.0	1	45.6 ± 0.0	14.5 ± 0.0
1.5	39.0 ± 0.0	17.6 ± 0.0	1.5	47.8 ± 0.0	13.6 ± 0.0
2	38.2 ± 0.0	18.1 ± 0.0	2	50.3 ± 0.0	12.8 ± 0.0

Table S2: Size-exclusion chromatography elution times and corresponding apparent protein radii of α-chymotrypsinogen.

ArgHCl [M]	Elution Time [min.]	Protein radius [Å]	NaCl [M]	Elution Time [min.]	Protein radius [Å]
0	35.5 ± 0.0	19.8 ± 0.0	0	35.5 ± 0.0	19.8 ± 0.0
0.25	35.4 ± 0.0	19.9 ± 0.0	0.25	35.9 ± 0.0	19.5 ± 0.0
0.5	35.1 ± 0.0	20.1 ± 0.0	0.5	36.1 ± 0.0	19.4 ± 0.0
0.75	34.4 ± 0.0	20.7 ± 0.0	0.75	36.0 ± 0.0	19.5 ± 0.0
1	34.3 ± 0.0	20.7 ± 0.0	1	36.6 ± 0.0	19.1 ± 0.0
1.5	34.2 ± 0.0	20.8 ± 0.0	1.5	37.4 ± 0.0	18.6 ± 0.0
2	33.7 ± 0.0	21.1 ± 0.0	2	37.7 ± 0.2	18.4 ± 0.1

Table S3: Size-exclusion chromatography elution times and corresponding apparent protein radii of carbonic anhydrase.

ArgHCl [M]	Elution Time [min.]	Protein radius [Å]	NaCl [M]	Elution Time [min.]	Protein radius [Å]
0	33.9 ± 0.0	21.0 ± 0.0	0	33.9 ± 0.0	21.0 ± 0.0
0.25	33.9 ± 0.0	21.0 ± 0.0	0.25	34.3 ± 0.0	20.7 ± 0.0
0.5	33.5 ± 0.0	21.3 ± 0.0	0.5	34.4 ± 0.0	20.6 ± 0.0
0.75	32.8 ± 0.0	21.9 ± 0.0	0.75	34.2 ± 0.0	20.8 ± 0.0
1	32.8 ± 0.0	21.9 ± 0.0	1	34.4 ± 0.0	20.6 ± 0.0
1.5	32.6 ± 0.0	22.1 ± 0.0	1.5	34.4 ± 0.0	20.6 ± 0.0
2	32.0 ± 0.1	22.6 ± 0.0	2	34.1 ± 0.0	20.9 ± 0.0

Table S4: Size-exclusion chromatography elution times and corresponding apparent protein radii of ovalbumin.

ArgHCl [M]	Elution Time [min.]	Protein radius [Å]	NaCl [M]	Elution Time [min.]	Protein radius [Å]
0	30.6 ± 0.0	23.9 ± 0.0	0	30.6 ± 0.0	23.9 ± 0.0
0.25	30.8 ± 0.0	23.7 ± 0.0	0.25	31.2 ± 0.0	23.3 ± 0.0
0.5	30.6 ± 0.0	23.9 ± 0.0	0.5	31.4 ± 0.0	23.2 ± 0.0
0.75	30.0 ± 0.0	24.5 ± 0.0	0.75	31.1 ± 0.0	23.4 ± 0.0
1	29.9 ± 0.0	24.6 ± 0.0	1	31.4 ± 0.0	23.2 ± 0.0
1.5	29.6 ± 0.0	24.9 ± 0.0	1.5	31.5 ± 0.0	23.1 ± 0.0
2	29.1 ± 0.0	25.5 ± 0.0	2	31.1 ± 0.0	23.4 ± 0.0

Table S5: Size-exclusion chromatography elution times and corresponding apparent protein radii of BSA.

ArgHCl [M]	Elution Time [min.]	Protein radius [Å]	NaCl [M]	Elution Time [min.]	Protein radius [Å]
0	28.6 ± 0.2	26.0 ± 0.2	0	28.6 ± 0.2	26.0 ± 0.2
0.25	28.8 ± 0.0	25.8 ± 0.0	0.25	29.1 ± 0.0	25.4 ± 0.0
0.5	28.6 ± 0.0	26.1 ± 0.0	0.5	29.2 ± 0.0	25.4 ± 0.0
0.75	28.1 ± 0.0	26.7 ± 0.0	0.75	28.9 ± 0.0	25.7 ± 0.0
1	28.0 ± 0.0	26.7 ± 0.0	1	29.1 ± 0.0	25.4 ± 0.0
1.5	27.9 ± 0.0	26.8 ± 0.0	1.5	29.2 ± 0.0	25.4 ± 0.0
2	27.5 ± 0.0	27.3 ± 0.0	2	28.9 ± 0.0	25.7 ± 0.0

Table S6: Size-exclusion chromatography elution times of excluded volume marker (dextran blue 2000) and total volume marker (acetone).

ArgHCl [M]	Blue dextran 2000 [min.]	Acetone [min.]	NaCl [M]	Blue dextran 2000 [min.]	Acetone [min.]
0	23.6 ± 0.2	53.5 ± 1.3	0	23.6 ± 0.2	53.5 ± 1.3
0.25	24.6 ± 0.4	52.9 ± 0.0	0.25	23.7 ± 0.2	52.7 ± 0.1
0.5	23.9 ± 0.1	53.3 ± 0.0	0.5	23.9 ± 0.2	53.2 ± 0.0
0.75	23.6 ± 0.0	53.3 ± 0.0	0.75	23.5 ± 0.1	53.0 ± 0.0
1	23.7 ± 0.2	54.2 ± 0.0	1	23.6 ± 0.0	53.8 ± 0.0
1.5	23.7 ± 0.0	55.5 ± 0.0	1.5	23.5 ± 0.1	54.6 ± 0.0
2	23.9 ± 0.3	55.9 ± 0.0	2	23.4 ± 0.1	54.6 ± 0.0

Hydrophobic interaction chromatography

Hydrophobic Interaction Chromatography (HIC) experiments were performed for lysozyme, ovalbumin and BSA on two different hydrophobic resins (Toyopearl Butyl 600M and Phenyl 650M, Tosoh Bioscience). 1 ml of resin (settled volume) was packed in a 5/50 Tricorn column (GE Healthcare), and experiments were carried out on an ÄKTA Explorer system (GE Healthcare) at a flow rate of 1 mL min^{-1}. The system was equilibrated with 20 ml of buffer B (50 mM phosphate, pH 7.2) containing 2.3 M ammonium sulfate and the specified concentration of ArgHCl. Proteins dissolved in the same equilibration buffer were loaded onto the column and the column was washed for 10 ml with equilibration buffer. Protein elution was carried out by a linear gradient from 2.3 to 0 M ammonium sulfate over 30 ml. Lysozyme and BSA eluted in a single peak. The elution time was determined from the peak center. Elution of ovalbumin occurred over multiple broad peaks and determination of the effect of ArgHCl on HIC retention times was difficult. Each HIC experiment was run in duplicate.

The possibility that arginine effects on protein elution times result from arginine-induced changes of the chromatographic matrix was investigated by characterizing the total volume of the HIC resins in the respective buffers. The total volume of the packed columns was characterized by monitoring elution times of a total volume marker (buffer with 5 % water) based on the conductivity signal. Arginine has nearly no effect on the marker elution times (Table S7), indicating that arginine does not change the chromatographic matrix of the HIC resins. Arginine effects on protein elution times can therefore not be attributed to changes of the chromatographic matrix.

Table S7: Elution times corresponding with the total volume of the HIC columns in the respective buffers.

HIC resin	buffer B	buffer B + 1 M arginine	buffer B + 2.3M am. sulfate	buffer B + 2.3M am. sulfate + 1M arginine
	[min.]	[min.]	[min.]	[min.]
Butyl	1.67 ± 0.02	1.71 ± 0.02	1.61 ± 0.02	1.59 ± 0.02
Phenyl	1.81 ± 0.00	1.84 ± 0.01	1.66 ± 0.01	1.64 ± 0.01

Differential scanning calorimetry

Differential Scanning Calorimetry (DSC) measurements were performed using VP-Capillary DSC system (Microcal Inc.). Protein samples were prepared at a concentration of 1 mg ml^{-1} in buffer A (50 mM MES, pH 6.5, 5 mM EDTA, 0.05% NaN$_3$ (w/v)) with the specified concentrations of ArgHCl. Proteins were scanned from 30 to 100°C at 1 °C min^{-1} and each experiment was performed in triplicate. Thermograms were corrected by deduction of buffer blank scans and normalized to the concentration. The melting temperature, T_m, was determined from the peak of the transition curve using the Origin 7.0 software (OriginLab Corporation).

To determine the reversibility of the thermal unfolding, protein samples at the respective arginine concentrations were scanned twice from 55 to 80 °C at 1°C min^{-1}. For lysozyme, thermograms of repeated scans were nearly identical. This indicates the reversibility of the thermal unfolding of lysozyme, which is consistent with previous studies that report reversible unfolding of lysozyme around 74 ˚C (*3, 4*). For ovalbumin, no thermal transition was observed for the second scan. This observation is in agreement with previous studies that have pointed out the difficulty of refolding ovalbumin (*5*). Nevertheless, several research groups have shown that the thermal transition of ovalbumin at 76 ˚C corresponds with reversible protein unfolding (*6, 7*). We conclude therefore that the reported T_m-values correspond with reversible unfolding of the proteins.

Since T_m-values may significantly depend on buffer conditions, the effects of arginine on T_m-values were also measured in 40 mM TrisHCl pH 7.5 (buffer B). Arginine effects in buffer A and buffer B were similar (Table S8 and Figure S9), suggesting that the observed arginine effects are independent of buffer conditions near neutral pH. This is consistent with the fact that carboxyl- and Gdm$^+$-groups, which enable the formation of protein-associated Arg$^+$-clusters, are oppositely charged over a relatively wide pH range around neutral pH. Interestingly, Arakawa and

Tsumoto (*3*) observed no change T_m-values of lysozyme in buffer B at arginine concentrations ranging from 0 to 2 M. The apparent discrepancy with our results could be attributed to the higher precision of our measurements.

Table S8: Melting temperature T_m for lysozyme and ovalbumin at different arginine concentrations in buffer A (50 mM MES, pH 6.5, 5 mM EDTA, 0.05% NaN$_3$ (w/v)) and for lysozyme in buffer B (40 mM TrisHCl pH 7.5).

ArgHCl [M]	lysozyme buffer A first scan [˚C]	ovalbumin buffer A first scan [˚C]	lysozyme buffer B first scan [˚C]
0	74.49 ± 0.21	76.58 ± 0.14	73.64 ± 0.04
0.2	73.69 ± 0.13	75.41 ± 0.02	72.66 ± 0.15
0.5	73.35 ± 0.13	75.04 ± 0.04	72.30 ± 0.23
1	73.32 ± 0.06	75.32 ± 0.16	72.47 ± 0.07
1.5	74.10 ± 0.07	76.47 ± 0.10	73.19 ± 0.27
2	75.41 ± 0.25	78.01 ± 0.07	74.42 ± 0.12

Characterization of local protein solvation

Molecular dynamics simulations

Molecular dynamics simulations were performed for lysozyme in aqueous solutions of ArgHCl, GdmCl and glycine (Table S9). A high resolution crystallographic structure of lysozyme was extracted from the protein data bank (PDB Id: 1VFB (*8*)), and the setup of each simulation was carried out with CHARMM (*9*) version c32b2. Charges of Arg, Lys and the N-terminus were set positive, charges for Asp, Glu and the C-terminus were set negative, and all other amino acids were neutral. The protein was solvated in mixtures of water and cosolvent (ArgHCl, GdmCl and glycine) and the protein charge (+8) was neutralized by Cl-atoms. The total number of water (n_w) and cosolvent (n_x) molecules in the solvent boxes for the respective simulations are listed in Table S9. For all simulations, a minimum of 10 Å between the protein and the boundary of the solvent box was kept. The CHARMM22 parameter set (*10*) was used to model protein atoms, and water was modeled by the TIP3-model (*11*). The force field parameters for arginine and glycine were taken from the CHARMM22 force field with the N-terminal protonated and the C-terminal deprotonated. The parameters for the N- and C-terminal were taken from the CTER and NTER parameters available in CHARMM. Force field parameters of the Gdm$^+$-ion were based on the forcefield parameters of arginine with the atomic partial charges assigned symmetrically (*12*).

The solvated protein system is minimized for 1000 steps in NAMD 2.7 (*13*) and simulations were run in the NpT ensemble (1 atm, 298 K) using a 2 fs step size while keeping the bond lengths of hydrogen constant with the SHAKE algorithm. The system temperature was controlled using Langevin dynamics with a coupling coefficient of 5 ps^{-1}. The system pressure was controlled using Langevin piston pressure control with a period of 100 fs and a barostat damping timescale of 50 fs. For all simulations, periodic boundary conditions were used, and images and non-bonded lists were updated every 10 steps using a 12 Å cutoff distance. The van der Waals potential energy was

smoothly switched off between 8 and 10 Å, and the particle-mesh Ewald method was used to calculate long-range electrostatic interactions.

To facilitate comparison of solvation characteristics for different additives (ArgHCl, GdmCl and glycine), all protein coordinates were constrained with respect to the crystal structure with a force constant of 2 kcal mol^{-1} Å^{-2}. To assess the effects of local conformational changes on protein solvation by arginine, an additional simulation was performed with unconstrained side-chains for lysozyme in 1 M arginine. Cluster-like Arg$^+$-regions appeared regardless of protein side-chains motions and the overall preferential interaction coefficient Γ_{XP} did not change significantly (Table S9). All systems were simulated for at least 130 ns, which is considerably longer than the longest characteristic residence time of arginine at the protein surface ($\sim$ 50 ns) (Table S9). For two simulations the simulation time was extended to 310 ns. The analysis of cluster formation and characteristic residence times based on 130 ns and on 310 ns gave similar results. This suggests that molecular dynamics simulations of 130 ns are sufficiently long to characterize cluster formation on a protein in an aqueous arginine solution.

Table S9: Simulation systems and global protein solvation properties.

Protein	Cosolvent	Concentration	Time[a]	n_w	n_x	R'[b]	$\Gamma_{XP}(R')$
		[M]	[ns]	[-]	[-]	[Å]	[-]
Lysozyme	ArgHCl	0.2	130	6616	24	6	3.2 ± 1.6
Lysozyme	ArgHCl	0.5	130	6293	60	6	-1.1 ± 1.6
Lysozyme	ArgHCl	1	310	5791	120	6	-5.4 ± 1.3
Lysozyme[c]	ArgHCl	1	130	5791	120	6	-4.0 ± 1.1
Lysozyme	ArgHCl	2	130	4961	241	8	-25.7 ± 3.3
Lysozyme	GdmCl	1	310	5791	120	6	2.2 ± 0.6
Lysozyme	Glycine	1	130	4961	241	6	3.6 ± 0.8

[a] Total simulation time; the first 10 ns of the simulation time was excluded from subsequent analysis.

[b] R' is the radial distance from the protein surface beyond which $\Gamma_{XP}(R)$ reaches a plateau (Eq. S6).

[c] For this simulation, only backbone coordinates were constrained, whereas of all other simulations all protein backbones were constrained.

Characteristic residence times

Residence times of solvent molecules near the protein surface ($r < 6$ Å) and survival functions for Arg$^+$ and water, i.e. $N_{Arg^+}(t)$ and $N_{wat}(t)$, are calculated as described previously (*14*). Characteristic residence times are obtained by fitting the following survival functions for Arg$^+$ and water, respectively:

$$N_{Arg^+}(t) \cong n_1 e^{-t/\tau_1} + n_2 e^{-t/\tau_2} \qquad (S1)$$

$$N_{wat}(t) \cong n_{W,1}e^{-t/\tau_{W,1}} + n_{W,2}e^{-t/\tau_{W,2}} + n_{W,3}e^{-t/\tau_{W,3}} + c_W \qquad \text{(S2)}$$

Equation S1 and S2 allows for good fitting of $N_{Arg^+}(t)$ and $N_{wat}(t)$ at all concentrations (Figure S2):
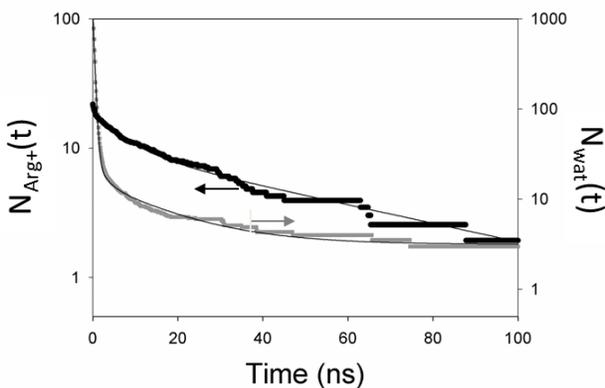

Figure S2: Survival functions $N_{Arg^+}(t)$ (black) and $N_{wat}(t)$ (gray) for lysozyme in 1 M ArgHCl, and their respective fits (Eq. S1 and S2).

Fitted parameters of $N_{Arg^+}(t)$ and $N_{wat}(t)$ are listed in Table 1 of the main text and in Table S10, respectively.

Table S10: Characteristic residence times of water at the protein surface in 1 M ArgHCl[a].

ArgHCl	n_1	τ_1	n_2	τ_2	n_3	τ_3	c
[M]	[-]	[ns]	[-]	[ns]	[-]	[ns]	[-]
0.2	1427.5	0.3	0.0	2.0	6.2	13.6	3.1
0.5	1370.3	0.3	2.6	2.0	10.9	12.9	2.8
1	1283.2	0.3	16.0	2.0	12.9	16.0	3.1
2	1225.7	0.5	125.4	2.0	8.3	16.9	3.4

[a] Parameters obtained by fitting to $N_{wat}(t)$ to Eq. S2.

Preferential interaction coefficient

The preferential interaction coefficient Γ_{XP} quantifies the excess of cosolvent at the protein surface and is determined by Kirkwood-Buff (KB) integrals $G_{\alpha P}$ for water (W) and cosolvent (X) (15-19):

$$\Gamma_{XP} = c_{X,bulk}(G_{XP} - G_{WP}) \qquad \text{(S3)}$$

$$G_{\alpha P} \equiv \int_V \left[\frac{c_\alpha(\vec{r})}{c_{\alpha,bulk}} - 1 \right] dV(\vec{r}) \text{ with } \alpha = X,W \qquad \text{(S4)}$$

With $c_\alpha(\vec{r})$ the local concentration of α at position $\vec{r}$ with respect to the protein P, and $c_{\alpha,bulk}$ the bulk concentration of α. Notably, only solvent regions where local solvent concentrations differ from the respective bulk concentrations contribute to Γ_{XP} (Eq. S4). Γ_{XP} can be determined from molecular dynamics simulations by truncating the above integral at a radial distance R and counting water and cosolvent molecules within R from the protein van der Waals surface(14):

$$\Gamma_{XP}(R) = \left\langle n_{XP}(r < R) - \frac{n_X - n_{XP}(r < R)}{n_W - n_{WP}(r < R)} n_{WP}(r < R) \right\rangle_{\tau_{run}} \tag{S5}$$

Brackets $\langle\ \rangle_{\tau_{run}}$ refer to the time average over the entire simulation time τ_{run}; n_X and n_W are the total number of cosolvent and water molecules in the simulation box; and $n_{XP}(r < R)$ and $n_{WP}(r < R)$ are the number of cosolvent and water molecules for which the center of mass is within a distance R from the protein van der Waals surface. Values of $\Gamma_{XP}(R)$ are calculated for increasing radial distances and R is fixed at the radial distance R' beyond which $\Gamma_{XP}(R)$ reaches a plateau (within the respective standard errors). The distance R' is the boundary between the local domain, where average solvent concentrations differ from the bulk solvent, and the bulk solvent, and we get:

$$\Gamma_{XP} \cong \Gamma_{XP}(R') \tag{S6}$$

For a 1:1 electrolyte the preferential interaction coefficient Γ_{XP} is determined by contributions of cations and anions (20):

$$\Gamma_{XP} = 0.5\left[\Gamma_{+,XP} + \Gamma_{-,XP} - |Z_P|\right] \tag{S7}$$

With $\Gamma_{+,XP}$ and $\Gamma_{-,XP}$ the preferential interaction coefficients of cations and anions, respectively, and Z_P the protein charge. Since $\Gamma_{+,XP}$ equals $\Gamma_{-,XP} - Z_P$ (20), Eq. S7 can be simplified into the following equations for positively and negatively charged proteins, respectively:

$$\Gamma_{XP} = \Gamma_{+,XP} \qquad Z_P \geq 0 \tag{S8a}$$

$$\Gamma_{XP} = \Gamma_{-,XP} \qquad Z_P \leq 0 \tag{S8b}$$

From Eq. S5, S6 and S8, we derive the following equations for calculating Γ_{XP} of proteins in aqueous solutions of 1:1 electrolytes:

$$\Gamma_{XP} \cong n_{+,XP}(r < R') - \frac{n_{+,X} - n_{+,XP}(r < R')}{n_W - n_{WP}(r < R')} n_{WP}(r < R') \qquad Z_P \geq 0 \tag{S9a}$$

$$\Gamma_{XP} \cong n_{-,XP}(r < R') - \frac{n_{-,X} - n_{-,XP}(r < R')}{n_W - n_{WP}(r < R')} n_{WP}(r < R') \qquad Z_P \leq 0 \tag{S9b}$$

With $n_{+,XP}(r < R')$ and $n_{-,XP}(r < R')$ the number of cations and anions within the distance R' from the protein surface, and $n_{+,X}$ and $n_{-,X}$ the total number of cations and anions in the simulation box. The distance R' is determined from Figure S3, and Equation S9a is used for calculating Γ_{XP} of ArgHCl and GdmCl for lysozyme (Table S9) and standard errors are calculated as described previously (*14, 21*).
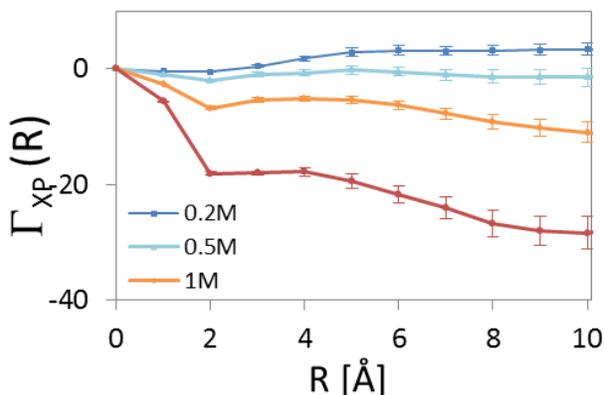

Figure S3: $\Gamma_{XP}(R)$ at increasing radial distances R from the protein surface at various ArgHCl concentrations. The radial distance R' equals the radial distance R where $\Gamma_{XP}(R)$ reaches a plateau (within the standard errors).

Estimation of increase of apparent protein radius by protein-associated Arg$^+$-ions

Taking into account the partial molar volume of Arg$^+$ of 118 cm^3 mol^{-1} (*22*), Arg$^+$-ions can be modeled as spheres with diameter d_{Arg^+} = 7.2 Å. A monolayer of Arg$^+$-spheres associated with the protein surface would therefore increase the apparent protein radius by ~ 7.2 Å. The increase of apparent protein radius by protein-associated class II Arg$^+$-ions can then be estimated as:

$$\Delta r_{protein}^{class\ II} = \frac{n_2^+}{n_{mono}^+} d_{Arg^+} \tag{S10}$$

With n_2^+ the average number of protein-associated class II Arg$^+$-ions (Table 1 in the main text), d_{Arg^+} the radius of Arg$^+$ (7.2 Å), and n_{mono}^+ the number of Arg$^+$-ions to form a monolayer at the protein surface.

Taking into account the surface area occupied by an Arg$^+$-sphere ($\sim d_{Arg^+}{}^2$) and the solvent accessible surface area of lysozyme (~ 7000 Å^2), the number of Arg$^+$-ions to form a monolayer at the surface of lysozyme is estimated to be 135. Equation S10 can then be used to calculate the increase of the apparent protein radius of lysozyme at various ArgHCl concentrations (Table S11).

Table S11: Estimated increase of the apparent radius of lysozyme by protein-associated Arg[+]-ions

ArgHCl	n_2^+	$\Delta r_{protein}^{class\ II}$ [a]	$\Delta r_{lysozyme}^{exp}$ [b]
[M]	[-]	[Å]	[Å]
0.2	3.3	0.18	0.20[c]
0.5	5.5	0.29	0.51
1	9.8	0.52	1.49
2	20.9	1.11	2.05

[a] Increase of lysozyme radius by protein-associated Arg[+] ions estimated from Eq. S10.

[b] Increase of lysozyme radius measured by Size-Exclusion Chromatography.

[c] Data measured at 0.25 M ArgHCl.

Supporting Figures
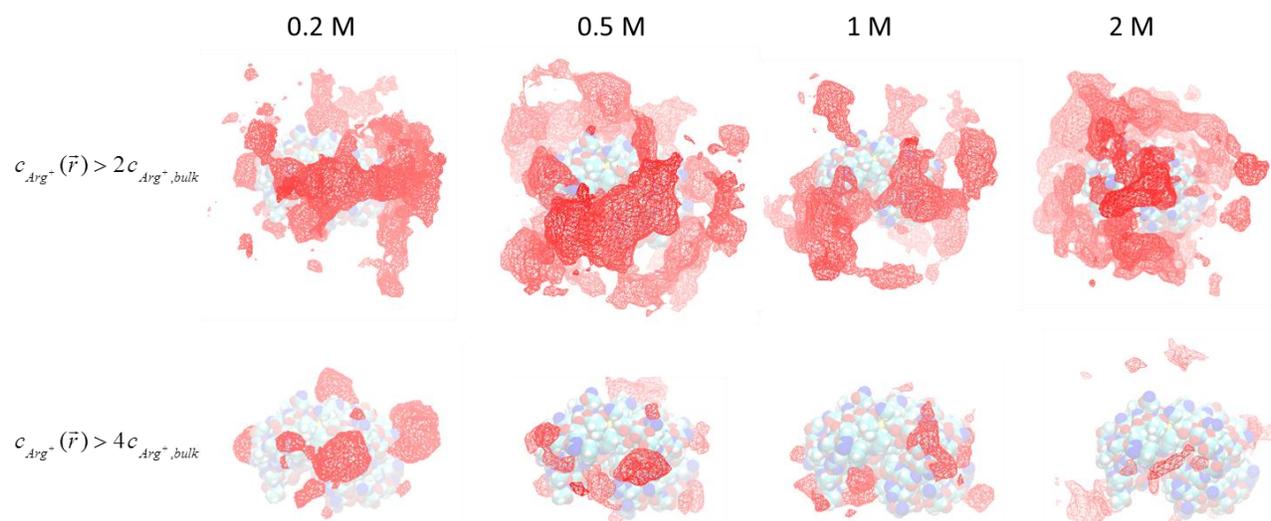
Figure S4: Local concentration maps of Arg^+ for lysozyme at various ArgHCl concentrations (0.2 – 2 M) with cutoffs of $2c_{Arg^+,bulk}$ (top) and $4c_{Arg^+,bulk}$ (bottom), respectively.
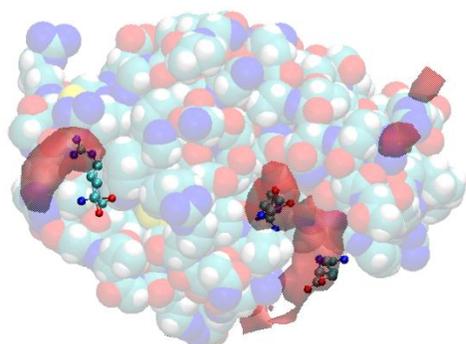
Figure S5: Superimposition of crystal-resolved Arg^+-ions (PDB: 3AGI (*23*)) with local concentration maps of Arg^+ for lysozyme in 1 M ArgHCl ($c_{Arg^+}(\vec{r}) > 4c_{Arg^+,bulk}$). Note that each of the 3 crystal-resolved Arg^+-ions is located near a high Arg^+ concentration region.
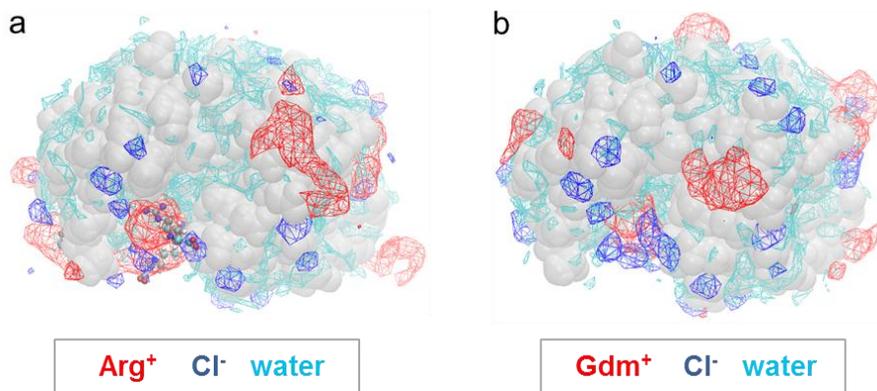
Figure S6: Local concentration maps of Cl^-, water, and (a) Arg^+ or (b) Gdm^+. Most Cl^--regions do not overlap with Arg^+-regions, and the location of Cl^--regions is nearly identical in ArgHCl and GdmCl solutions. Data for lysozyme in 1M ArgHCl and GdmCl, respectively.
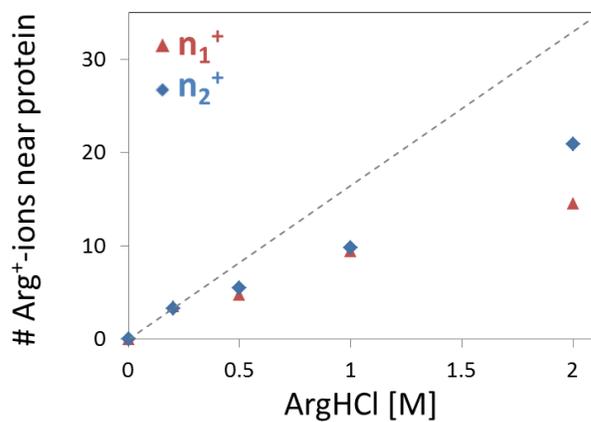

Figure S7: The number of class I Arg^+-ions (n_1^+) and class II Arg^+-ions (n_2^+) in function of ArgHCl concentration. The dashed line represents a linear increase extrapolated from 0.2 M ArgHCl.
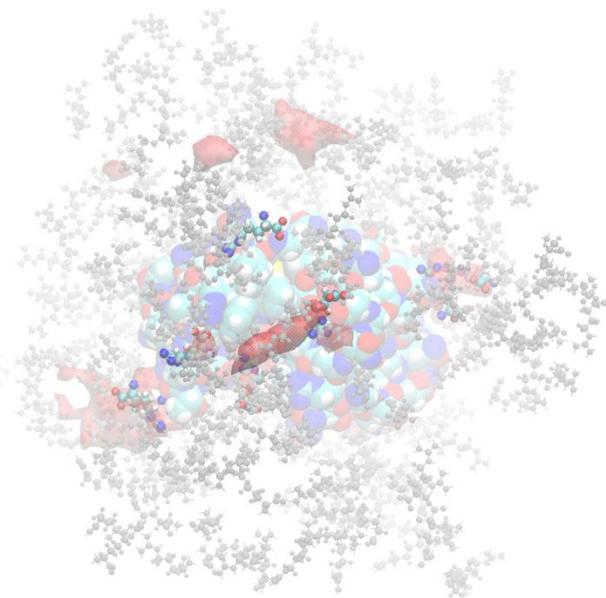
Figure S8: Snapshot of lysozyme in 2 M ArgHCl. Class II Arg$^+$-ions are represented in colored ball-and-stick models, and all other Arg$^+$-ions are represented as gray ball-and-stick models. Local concentration maps ($c_{Arg^+}(\vec{r}) > 4 c_{Arg^+,bulk}$) are colored in transparent red.
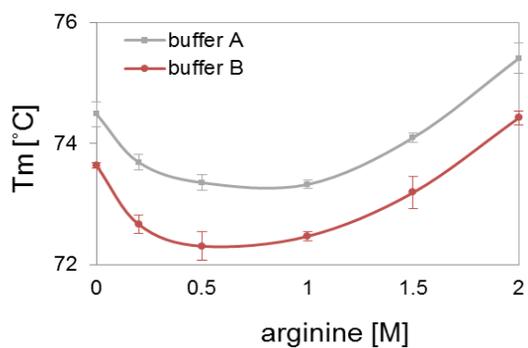

Figure S9: Effects of arginine on the conformational stability of lysozyme in two different buffers. Protein melting temperatures T_m were measured by differential scanning calorimetry in buffer A (50 mM MES pH 6.5, 5 mM EDTA, 0.05% NaN$_3$) and buffer B (40mM TrisHCl pH 7.5) at various arginine concentrations.

Supporting movies

Movie 1: Trajectory of Arg^+- and Cl^--ions (red and blue ball-and-stick models, respectively) for lysozyme in 0.2 M ArgHCl. Red transparent regions represent high concentration regions of Arg^+. Snapshots are rendered with VMD 1.9 (*24*) for every 1 ns.

Movie 2: Trajectory of Arg^+-ions associated with lysozyme residues Arg14 and Glu7 in 0.2 M ArgHCl from 0 to 130 ns. Class II Arg^+-ions and adjacent Arg^+-ions are colored in bright and shaded ball-and-stick models, respectively. Snapshots are rendered with VMD 1.9 (*24*) for every 1 ns. Note the repeated orientational change of each class II Arg^+-ion between 'double-associated' orientations whereby both its Gdm^+ and carboxyl moieties contact the protein surface, and 'single-associated' orientations whereby only one of these moieties contact the protein surface and the other moiety is free to interact with the other Arg^+-ions. Furthermore, the protein surface locus near Arg14 is occupied by a class II Arg^+-ion for almost the entire simulation time, while the protein surface locus near Glu7 is occupied by a class II Arg^+-ion for about half of the simulation time. Such high occupancy rates reflect the high affinity for Arg^+.

Supporting references

1. Gekko, K., and Noguchi, H. (1979) COMPRESSIBILITY OF GLOBULAR-PROTEINS IN WATER AT 25-DEGREES-C, *J. Phys. Chem. 83*, 2706-2714.
2. Squire, P. G., and Himmel, M. E. (1979) HYDRODYNAMICS AND PROTEIN HYDRATION, *Arch. Biochem. Biophys. 196*, 165-177.
3. Arakawa, T., and Tsumoto, K. (2003) The effects of arginine on refolding of aggregated proteins: not facilitate refolding, but suppress aggregation, *Biochem. Biophys. Res. Commun. 304*, 148-152.
4. Reddy, R. C., Lilie, H., Rudolph, R., and Lange, C. (2005) L-Arginine increases the solubility of unfolded species of hen egg white lysozyme, *Protein Sci. 14*, 929-935.
5. Takahashi, N., and Hirose, M. (1992) REVERSIBLE DENATURATION OF DISULFIDE-REDUCED OVALBUMIN AND ITS REOXIDATION GENERATING THE NATIVE CYSTINE CROSS-LINK, *J. Biol. Chem. 267*, 11565-11572.
6. Tani, F., Shirai, N., Onishi, T., Venelle, F., Yasumoto, K., and Doi, E. (1997) Temperature control for kinetic refolding of heat-denatured ovalbumin, *Protein Sci. 6*, 1491-1502.
7. Takahashi, N., Onda, M., Hayashi, K., Yamasaki, M., Mita, T., and Hirose, M. (2005) Thermostability of refolded ovalbumin and S-ovalbumin, *Biosci. Biotechnol. Biochem. 69*, 922-931.
8. Bhat, T. N., Bentley, G. A., Boulot, G., Greene, M. I., Tello, D., Dallacqua, W., Souchon, H., Schwarz, F. P., Mariuzza, R. A., and Poljak, R. J. (1994) Bound Water-Molecules and Conformational Stabilization Help Mediate an Antigen-Antibody Association, *Proc. Natl. Acad. Sci. U. S. A. 91*, 1089-1093.
9. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations, *J. Comput. Chem. 4*, 187-217.
10. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J. Phys. Chem. B 102*, 3586-3616.
11. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water, *J. Chem. Phys. 79*, 926-935.
12. Mason, P. E., Neilson, G. W., Enderby, J. E., Saboungi, M. L., Dempsey, C. E., MacKerell, A. D., and Brady, J. W. (2004) The structure of aqueous guanidinium chloride solutions, *J. Am. Chem. Soc. 126*, 11462-11470.
13. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD, *J. Comput. Chem. 26*, 1781-1802.
14. Vagenende, V., Yap, M. G. S., and Trout, B. L. (2009) Molecular Anatomy of Preferential Interaction Coefficients by Elucidating Protein Solvation in Mixed Solvents: Methodology and Application for Lysozyme in Aqueous Glycerol, *J. Phys. Chem. B 113*, 11743-11753.
15. Shimizu, S., and Boon, C. L. (2004) The Kirkwood-Buff theory and the effect of cosolvents on biochemical reactions, *J. Chem. Phys. 121*, 9147-9155.
16. Schurr, J. M., Rangel, D. P., and Aragon, S. R. (2005) A contribution to the theory of preferential interaction coefficients, *Biophys. J. 89*, 2258-2276.
17. Shulgin, I. L., and Ruckenstein, E. (2005) A protein molecule in an aqueous mixed solvent: Fluctuation theory outlook, *J. Chem. Phys. 123*, 054909.
18. Shimizu, S., and Matubayasi, N. (2006) Preferential hydration of proteins: A Kirkwood-Buff approach, *Chem. Phys. Lett. 420*, 518-522.
19. Jiao, Y., and Smith, P. E. (2011) Fluctuation theory of molecular association and conformational equilibria, *J. Chem. Phys. 135*, 014502.
20. Record, M. T., and Anderson, C. F. (1995) INTERPRETATION OF PREFERENTIAL INTERACTION COEFFICIENTS OF NONELECTROLYTES AND OF ELECTROLYTE IONS IN TERMS OF A 2-DOMAIN MODEL, *Biophys. J. 68*, 786-794.
21. Allen, M. P., and Tildesley, D. J. (1987) *Computer simulation of liquids*, Clarendon Press, Oxford, UK
22. Rao, M. V. R., Atreyi, M., and Rajeswari, M. R. (1984) PARTIAL MOLAR VOLUMES OF ALPHA-AMINO-ACIDS WITH IONOGENIC SIDE-CHAINS IN WATER, *J. Phys. Chem. 88*, 3129-3131.
23. Ito, L., Shiraki, K., Matsuura, T., Okumura, M., Hasegawa, K., Baba, S., Yamaguchi, H., and Kumasaka, T. (2011) High-resolution X-ray analysis reveals binding of arginine to aromatic residues of lysozyme surface: implication of suppression of protein aggregation by arginine, *Protein Eng. Des. Sel. 24*, 269-274.
24. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics, *J. Mol. Graphics 14*, 33-38.