Charge-state dependent variation of signal intensity ratio between unbound protein and protein-ligand complex in electrospray ionization mass spectrometry: the role of solvent-accessible surface area

Konstantin Chingin[1] and Konstantin Barylyuk[2]

[1]*Jiangxi Key Laboratory for Mass Spectrometry and Instrumentation, East China University of Technology, Guanglan road 418, Nanchang, Jiangxi, China, 330013*
[2]*Department of Biochemistry, University of Cambridge, Hopkins Building, Tennis Court Road, Cambridge, CB2 1QW, United Kingdom*

Correspondence to: chingin.k@gmail.com; kb601@cam.ac.uk.

Keywords: native mass spectrometry, protein-ligand complexes; binding constant; charge state distribution; charged residue model; electrospray ionization.

SUPPORTING INFORMATION

Methods

SASA calculation

The data on the high-resolution 3D structure of the protein complexes with their respective ligands were retrieved from the RCSB Protein Data Bank. The atomic coordinates were used to generate molecular surfaces of the proteins and the ligands in UCSF Chimera[1] using embedded software from the MSMS package.[2] *SASA* was calculated using the 'rolling ball' algorithm with a probe radius of 1.4 Å.[3] All the data sets passed D'Agostino – Pearson normality test ($\alpha = 0.05$). A paired t-test was performed to evaluate the statistical significance of $\Delta SASA = SASA_{PL} - SASA_{P}$.

Only a few ligand-free structures of the proteins considered in this study are available in PDB, stymying statistical analysis of the ligand-induced differences in *SASA*. To overcome this limitation, we generated apo-protein structures used for $SASA_P$ calculation by deleting the ligand atoms from the PL structure. A comparison of apo-protein structures generated this way for lysC and hCA I with true unbound protein structures (PDB IDs 2HEW vs. 1LYZ and 1AZM vs. 2CAB, respectively) revealed negligible differences in atomic coordinates: *RMSD* between pairs of backbone atoms was 0.320 Å and 0.326 Å, respectively (**Figure S1a-d**). In the case of Bcl-xL, a greater value of *RMSD* = 2.370 Å found for backbone atom pairs in the free protein structure (PDB ID 2M03) and the apo-protein structure generated from the complex with Bak-peptide (PDB ID 1BXL) reflects a conformational change induced by the ligand. More specifically, there is a conformational rearrangement around the BC groove (the binding site) upon Bak-peptide binding that includes winding of additional turns in α2 and concomitant unwinding in α3 (the two alpha helices flanking the BC groove), along with major side-chain rearrangements on both sides of the BC groove[4] (**Figure S1e-f**).

We also additionally tested on a subset of lysC, hCA I, and Bcl-xL structures that ligand removal does not result in an unfavorable conformation of the generated apo-proteins by performing energy minimization of protein structures with and without the ligand using the Molecular Modelling Toolkit (MMTK)[5] routines included in UCSF Chimera (see Chimera User's Guide for details: https://www.cgl.ucsf.edu/chimera/docs/). Typically, 100-1000 steepest descent steps followed by 10-100 conjugate gradient steps using a step size of 0.02 Å for both were used, with all other parameters set to defaults. Hydrogen atoms were added to protein structures when necessary and partial charges were assigned to atoms using DockPrep routine in UCSF Chimera prior to minimization. Only minor changes in atom coordinates occurred upon minimization that did not change the outcome of the statistical analysis of Δ*SASA* (**Figure S2, Table S4**).

The buried surface area was calculated as $bSASA = (SASA_P + SASA_L - SASA_{PL})/2$.

CSD analysis

The P and PL peak intensities were extracted from the spectra to calculate Z_{AV}:

$$Z_{AV} = \sum_{Z=Z_{min}}^{Z_{max}} Z \frac{I_Z}{I_{total}} \quad (\mathbf{S1})$$

where Z_{min} and Z_{max} are the observed minimum and maximum charge states, I_Z is the peak intensity of the ion of the Z-th charge state, and I_{total} is the sum of intensities of either all PL or all P peaks observed in the spectrum. All the data sets passed D'Agostino – Pearson normality test ($\alpha = 0.05$). Statistical significance of the observed $\Delta Z_{AV} = Z_{AV(PL)} - Z_{AV(P)}$ was evaluated by a paired t-test. The data on Z_{AV} of P and PL were paired to account for spectrum-to-spectrum variation of CSDs due to technical reasons, such as, for example, differences in the outlet diameter and the position of nano-ESI emitter tip relative to the inlet of the mass spectrometer.

CSD shift simulation

Experimental CSD of hCA I (**Figure 1**) was calculated according to **Equation S1** using data from an ESI mass spectrum measured under near-native conditions (protein concentration 10 µM, 50 mM ammonium acetate solution, pH 7.3). A Gaussian distribution was fitted to the extracted data to approximate the CSD. The fitted Gaussian was then shifted by $\Delta Z_{AV} = 0.07$ (≈ 1 % of $SASA_P$) to simulate the CSD of a hypothetical PL complex. The following assumptions were made: 1) the ligand is present at an equimolar amount with the protein (c_L = 10 µM); 2) 50 % of the available binding sites are occupied by the ligand (K_a = 0.2 µM^{-1}). The resulting peak intensity ratios of P and PL and the corresponding K_a values showed a pronounced variation with the charge state. A calculator to estimate the effect of ΔZ_{AV} on CSD and K_a is available at the link: https://www.desmos.com/calculator/6ub01ptsai.

Analysis of the dependence of ΔZ_{AV} on the concentration of ligand

The PL systems in this study were chosen partly based on the availability of ligand titration data. This allowed the opportunity to perform a linear regression analysis of the dependence of both Z_{AV} of P and PL, and ΔZ_{AV} on ligand molar excess in the reaction mixture (**Table S5**, **Figure S4**). An F-test was performed to test whether the slope of the fitted lines is significantly non-zero ($\alpha = 0.05$). In most cases, the slope of the regressed line did not significantly differ from 0, indicating the absence of any trend. We found significant trends in Z_{AV} of both P and PL in the case of hCA I (**Figure S4a**). However, there was no significant difference between the slopes of these trends, only between the intercepts. The fitted lines are parallel, and, hence is the absence of a significant trend in ΔZ_{AV}. A weak concentration dependence was found for ΔZ_{AV} of lysozyme-NAG_3 (**Figure S4d**), which is most likely an artefact due to the small number of available data points. Finally, we found significant trends in $Z_{AV(P)}$ (negative slope) and ΔZ_{AV} (positive slope) for Bcl-xL (**Figure S4e**), which we attribute to an ion suppression effect caused by a relatively large peptide ligand.

Supplementary Tables and Figures

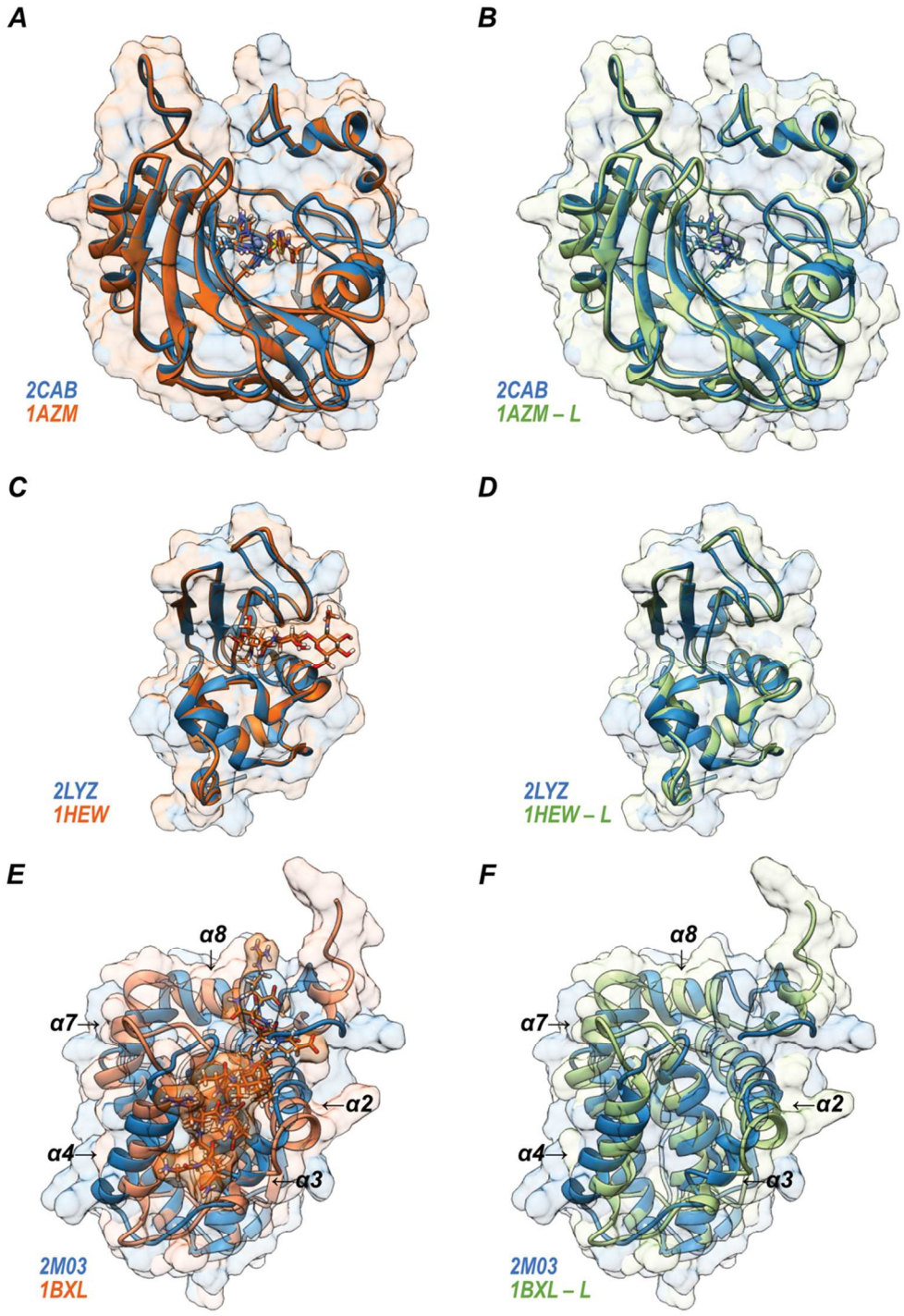

Figure S1. Superposition of free protein and protein-ligand complex structures of hCA I (***A-B***), lysC (***C-D***), and Bcl-xL (***E-F***). Protein chains are shown as ribbons, ligands are shown as sticks, and molecular surfaces are overlaid. Free proteins are in blue, complexes are in orange (***A, C, E***), and apo-protein structures generated by removing the ligand are in green (***B, D, F***). All structures were energy-minimized. PDB IDs

are indicated next to the diagrams and color-coded accordingly. For Bcl-xL, alpha-helices with the most pronounced conformational change induced by the ligand are indicated (***E-F***).

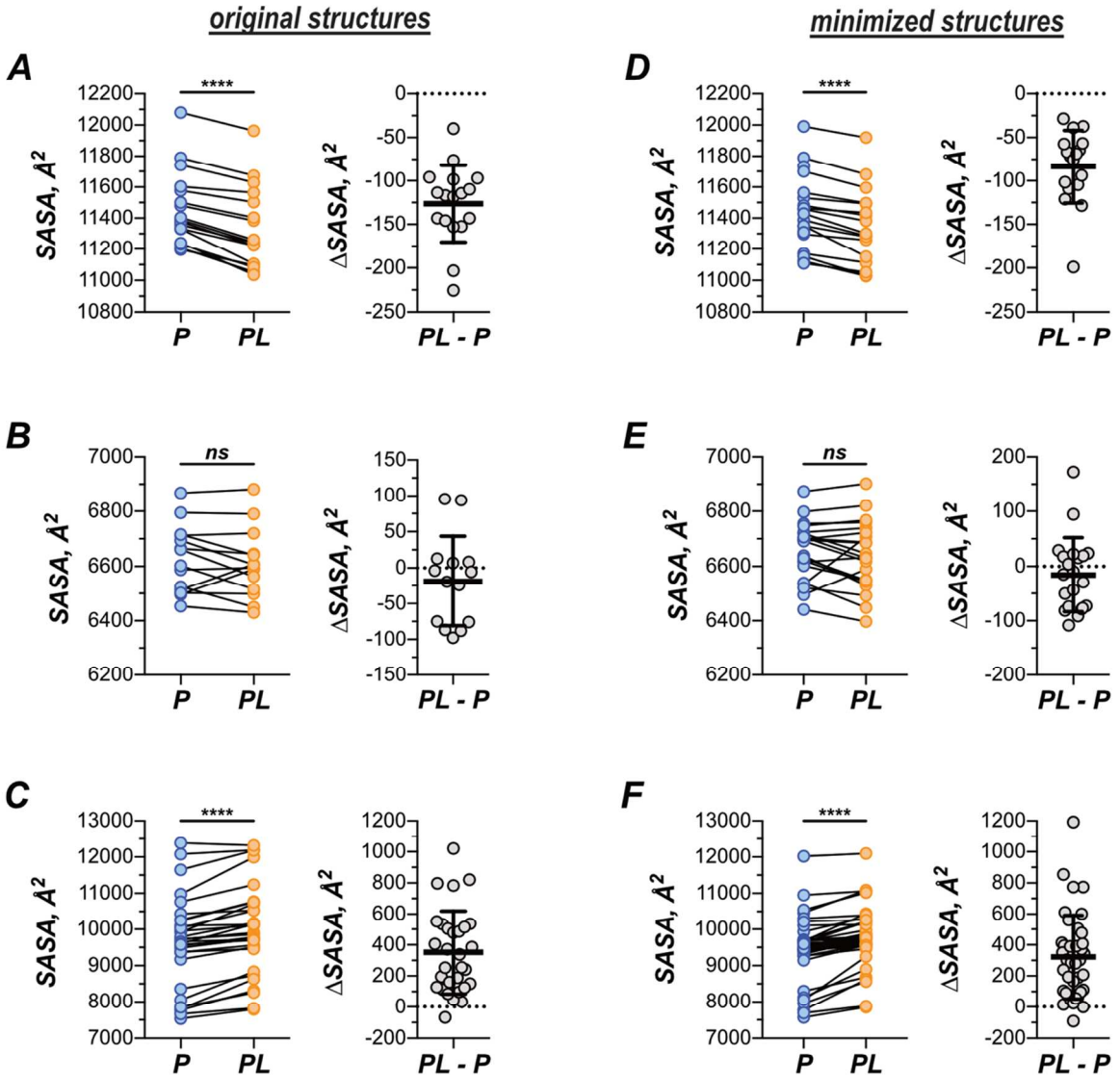

Figure S2. Comparison of $SASA_P$ (blue circles), $SASA_{PL}$ (orange circles), and Δ*SASA* (grey circles) values before (***A-C***) and after (***D-F***) energy minimization of the structures of hCA I (***A, D***), lysC (***B, E***), and Bcl-xL (***C, F***). Paired t-test was performed to evaluate the statistical significance of $\Delta SASA = SASA_{PL} - SASA_P$. Sample sizes were 16 and 17 for hCA I (***A, D***), 14 and 21 for lysC (***B, E***), and 30 and 39 for Bcl-xL (***C, F***). The summary of *p*-value is as follows: **** - $p < 0.0001$, *ns* – not significant ($p \geq 0.05$). Horizontal bars and whiskers in ΔSASA plots show means with the standard deviations.

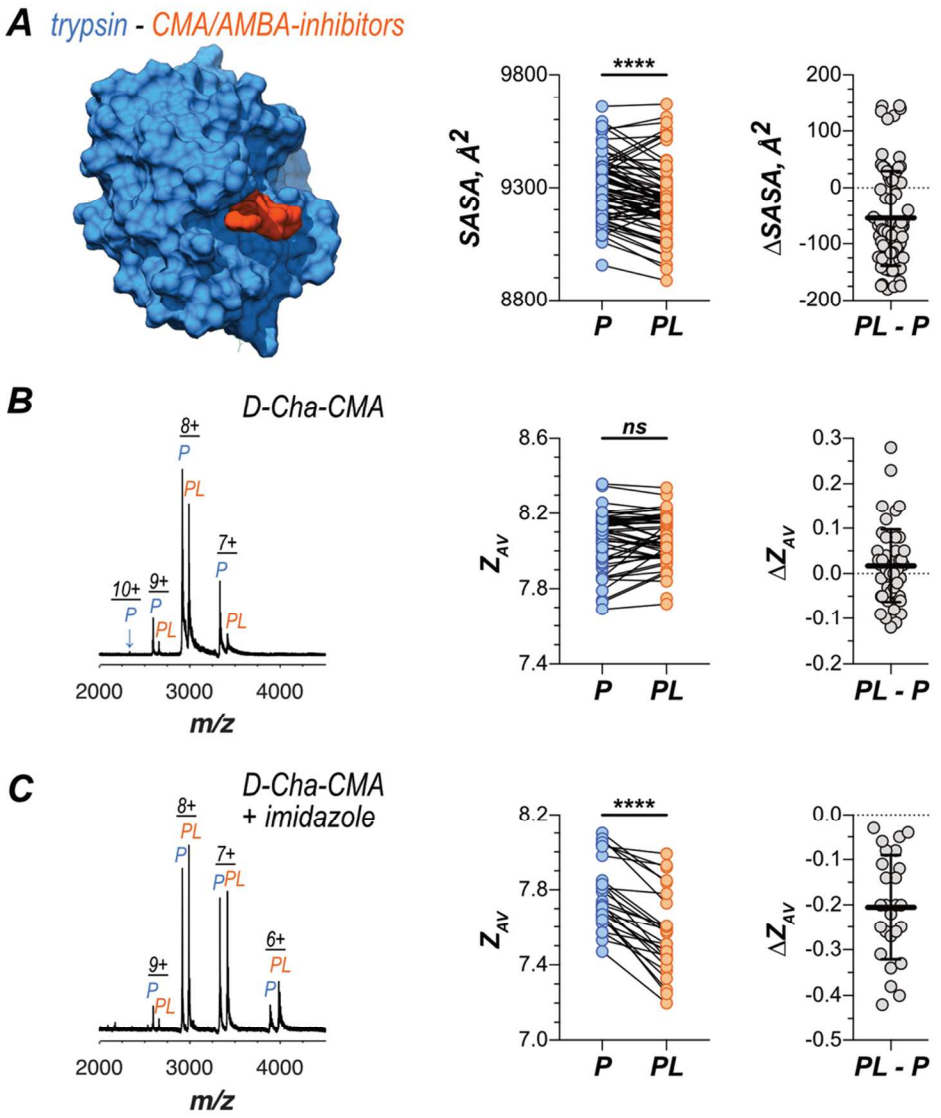

Figure S3. Analysis of Δ*SASA* (***A***) and ΔZ_{AV} (***B***, ***C***) of bovine cationic trypsin complexes with synthetic inhibitors containing 2-aminomethyl-5-chlor-benzylamide (CMA) or 4-amidinobenzylamide (AMBA) moiety. Molecular surface of a complex of trypsin (blue) with benzylsulfonyl-*D*-Arg-Gly-AMBA (orange) is shown in ***A*** (PDB ID 3PMJ[6]). A comparison of $SASA_P$ with $SASA_{PL}$ revealed a significant $\Delta SASA < 0$ (paired t-test, **** - $p < 0.0001$, n = 75). Example native ESI mass spectra of 5 µM trypsin mixed with 50 µM benzylsulfonyl-*D*-Cha-Pro-CMA alone (***B***) or in the presence of 10 mM imidazole (***C***) demonstrate that imidazole stabilizes weak noncovalent interactions and induces charge reduction in native ESI-MS.[7] No significant difference was found between Z_{AV} of electrosprayed ions of trypsin alone and bound with various CMA-inhibitors (***B***; paired t-test, *ns* – no significant difference, n = 51). Paired t-test performed on the native ESI-MS data acquired for the same protein-ligand mixtures in the presence of 10 mM imidazole revealed significant $\Delta Z_{AV} < 0$ (***C***; paired t-test, **** - $p < 0.0001$, n = 26). Horizontal bars and whiskers indicate mean values and standard deviations.

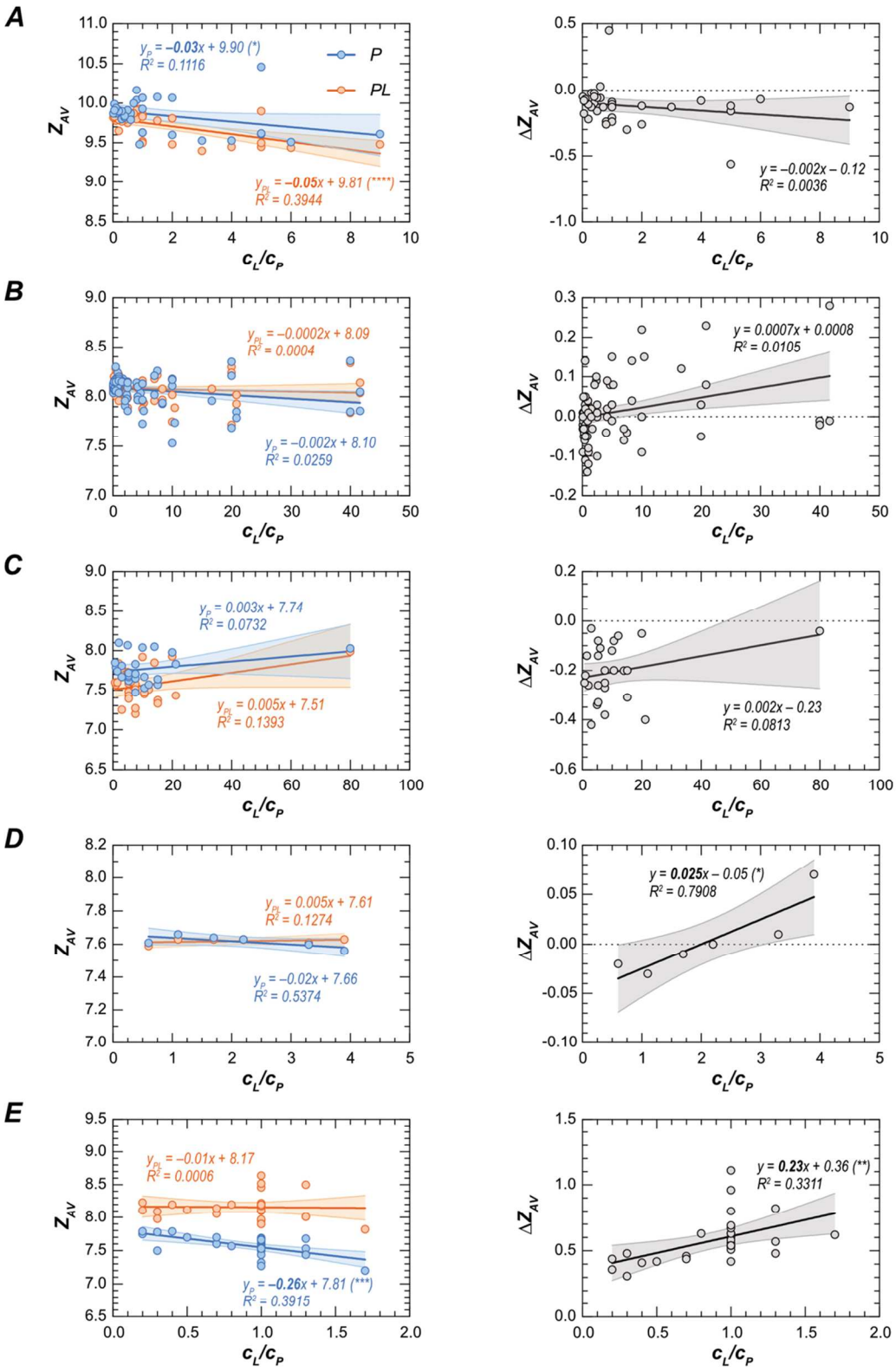

Figure S4. Linear regression analysis of the dependence of Z_{AV} (left column; blue for P, orange for PL) and ΔZ_{AV} (right column) on the ligand molar excess for ESI-MS titration data on hCA I (***A***)[8], trypsin with and

without 10 mM imidazole (***B*** and ***C***, respectively)[7], lysozyme (***D***)[9], and Bcl-xL (***E***)[10]. The lines of best fit are shown with 95-% prediction bands (shaded area). The fitted linear equations are indicated in the plots. The statistical significance of the slope deviation from zero was evaluated by F-test, and where a significant deviation was found the slope parameter value is indicated in bold and the respective summary of *p*-value is given in the brackets (* - $p < 0.05$, ** - $p < 0.01$, *** - $p < 0.001$, **** - $p < 0.0001$). See **Table S5** for details.

Table S1. Summary of the data on the solvent-accessible surface area (*SASA*) and average charge (Z_{AV}) in ESI-MS of selected protein-ligand complexes analyzed in this study.

Protein		Ligand		*SASA*[b], Å^2		Z_{AV}[c]		Shift				Ref.
Name	*MW*, Da	Name	*MW*[a], Da	P	PL	P	PL	Δ*SASA*, Å^2	*p*[d]	Δz_{AV}	*p*[d]	
lysozyme	14299	NAG-containing oligosaccharides	627.59	6628 ± 123	6610 ± 124.8	7.62 ± 0.04	7.62 ± 0.01	-18.4 ± 62.42	*ns*	0.002 ± 0.036	*ns*	[9]
human carbonic anhydrase I	28739	sulfonamides	157.19-258.32	11461 ± 237.8	11335 ± 258.5	9.85 ± 0.21	9.73 ± 0.16	-126.2 ± 44.26	****	-0.12 ± 0.14	**	[8]
human Bcl-xL (1-209, Δ45-84)	20650	BH3 peptides	1724.9	9228 ± 1129	9585 ± 1106	7.58 ± 0.15	8.16 ± 0.19	357.2 ± 238.9	****	0.58 ± 0.18	****	[10]
bovine cationic trypsin	23299	CMA-inhibitors	478.99-575.17	9299 ± 138.9	9245 ± 156	8.05 ± 0.17	8.07 ± 0.13	-54.6 ± 84.14	****	0.02 ± 0.08	*ns*	[7]
bovine cationic trypsin	23299	AMBA-inhibitors	457.55-553.72	9299 ± 138.9	9245 ± 156	8.10 ± 0.06	8.10 ± 0.04	-54.6 ± 84.14	****	-0.002 ± 0.054	*ns*	[7]
Se155-4 single-chain variable fragment	26539	oligosaccharides	120.17-1819.7	10717 ± 525.1	10721 ± 550.7	9.56 ± 0.04	9.61 ± 0.99	4.56 ± 210.8	*ns*	0.04 ± 0.99[e]	*ns*	[11,12]
human galectin 3, C-terminal domain	16330	oligosaccharides	488.44-1438.3	7331 ± 107.5	7484 ± 159.7	7.84 ± 0.33	8.17 ± 0.29	152.3 ± 105.3	****	0.33 ± 0.44[e]	*	[13]
bovine cationic trypsin	23299	CMA-inhibitors (with imidazole)	478.99-575.17	9299 ± 138.9	9245 ± 156	7.77 ± 0.18	7.57 ± 0.22	-54.6 ± 84.14	****	-0.21 ± 0.12	****	[7]

a – the range of molecular weight is provided only for those ligands that have associated ESI-MS data;

b – *SASA* – solvent-accessible surface area;

c – Z_{AV} – average charge in ESI mass spectrum;

d – summary of *p*-value computed in paired t-test: *ns* = no significant difference; * = $p < 0.05$; ** = $p < 0.01$; **** = $p < 0.0001$;

e – shift in the average charge was evaluated by unpaired t-test.

Supplementary References

(1) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004,** *25,* 1605-1612.
(2) Sanner, M. F.; Olson, A. J.; Spehner, J. C. *Biopolymers* **1996,** *38,* 305-20.
(3) Shrake, A.; Rupley, J. A. *J. Mol. Biol.* **1973,** *79,* 351-71.
(4) Chipuk, J. E.; Moldoveanu, T.; Llambi, F.; Parsons, M. J.; Green, D. R. *Molecular cell* **2010,** *37,* 299-310.
(5) Hinsen, K. *J. Comput. Chem.* **2000,** *21,* 79-85.
(6) Tziridis, A.; Rauh, D.; Neumann, P.; Kolenko, P.; Menzel, A.; Brauer, U.; Ursel, C.; Steinmetzer, P.; Sturzebecher, J.; Schweinitz, A.; Steinmetzer, T.; Stubbs, M. T. *Biol Chem* **2014,** *395,* 891-903.
(7) Cubrilovic, D.; Biela, A.; Sielaff, F.; Steinmetzer, T.; Klebe, G.; Zenobi, R. *J. Am. Soc. Mass Spectrom.* **2012,** *23,* 1768-1777.
(8) Jecklin, M. C.; Schauer, S.; Dumelin, C. E.; Zenobi, R. *Journal of Molecular Recognition* **2009,** *22,* 319-329.
(9) Cubrilovic, D.; Zenobi, R. *Anal. Chem.* **2013,** *85,* 2724-2730.
(10) Cubrilovic, D.; Barylyuk, K.; Hofmann, D.; Walczak, M. J.; Graber, M.; Berg, T.; Wider, G.; Zenobi, R. *Chemical Science* **2014,** *5,* 2794-2803.
(11) El-Hawiet, A.; Kitova, E. N.; Arutyunov, D.; Simpson, D. J.; Szymanski, C. M.; Klassen, J. S. *Anal. Chem.* **2012,** *84,* 3867-3870.
(12) Yao, Y.; Shams-Ud-Doha, K.; Daneshfar, R.; Kitova, E.; Klassen, J. *J. Am. Soc. Mass Spectrom.* **2015,** *26,* 98-106.
(13) Han, L.; Shamsuddoha, K.; Kitova, E. N.; Klassen, J. S. *Anal. Chem.* **2016,** *88,* 8224.