Supporting information:

Relative Interfacial Cleavage Energetics of Protein Complexes Revealed by Surface Collisions

Sophie R. Harvey, Justin T. Seffernick, Royston S. Quintyn, Yang Song, Yue Ju, Jing Yan, Aniruddha N. Sahasrabuddhe, Andrew Norris, Mowei Zhou, Edward J. Behrman, Steffen Lindert, and Vicki H. Wysocki.

Materials and methods

Synthesis of the Perfluorothiol (FC_{12})

2-Perfluorodecyl ethane-1-thioacetate [$CF_3(CF_2)_9CH_2CH_2SCOCH_3$]

The procedure of Naud *et al*.(1) was modified as follows. A 50 mL three-neck round-bottom flask was charged with 2.52 g (4.02 mmol) of 2- perfluorodecyl ethyl bromide (Oakwood Chemical Co.), 0.92 g (10.7 mmol) potassium thioacetate, 20 mL ethanol, and a stirring bar. A reflux condenser and two septa were attached, the apparatus was flushed with nitrogen, and a balloon fitted to the top of the condenser. The mixture was refluxed for 4 hr with stirring. The reaction mixture was then cooled on ice, acidified with HCl, and poured with stirring into cold water. The ester was filtered on a Büchner funnel, washed with water, and air-dried. Yield: 2.2-2.4 g, 91%. The IR spectrum of the thioester is similar to that of the bromide except for a strong absorption due to the carbonyl group. Melting point (Mp): 63-64°C. IR(Nujol): 1695, 1441, 1239, 1216, 1150, 967, 884 cm^{-1}. The bromide: 1452, 1202, 1147, 1076, 951, 882 cm^{-1}. ^{1}H-NMR(600 MHz, $CDCl_3$, TMS):δ 2.37(s, $-CH_3$), 2.38(tt, J= 8.2 Hz, $-CF_2\mathbf{CH_2}-$), 3.09(t, J=8.2 Hz, $-\mathbf{CH_2}S-$). The three-bond proton-proton and proton-fluorine couplings are coincidentally the same.

1-mercapto-2-perfluorodecyl ethane[$CF_3(CF_2)_9CH_2CH_2SH$]

A 50 mL three-neck flask was charged with 1.24 g (2 mmol) of the ester, 20 mL methanol, and a stirring bar. A reflux condenser and one septum was then installed and the air displaced with nitrogen. A solution of NaOH in water (0.52 ml, 0.15 g, 3.75 mmol) was then added under nitrogen. The second septum was added, remaining air displaced, and a balloon used to cap the condenser. The mixture was refluxed for 4 hr with stirring. The thiol was isolated in the same way as for the ester. Yield: 0.97- 1.04 g, 89%. The precipitate may have such fine particles that it is difficult to filter. In this case, it should be extracted with chloroform. The product contained about 7% disulfide. ^{1}H-NMR (600 MHz,$CDCl_3$, TMS): δ 1.61(t, J=8.2 Hz, -SH), 2.43(tt, J=8.2 Hz, $-CF_2\mathbf{CH_2}-$), 2.76(dt, J=8.2 Hz, $-\mathbf{CH_2}SH$). IR(Nujol):1456, 1446, 1344, 1306, 1241, 1219, 1150, 1111, 1078, 995, 883, 665, 643, 556, 528 cm^{-1}. The IR spectra of the thiol and the disulfide are almost identical but an absorption at 950 cm^{-1} is characteristic of the disulfide and absent in the thiol whereas the band at 995 cm^{-1} is absent in the disulfide and present in the thiol. The –SH stretch was not observed in the IR, but the Raman spectrum shows the –SH stretch at 2576 cm^{-1}. Mp 67-68°(literature 68°, ref 1)). The thiol may be separated from the disulfide cleanly by dissolution in dichloromethane in which, remarkably, the disulfide is hardly soluble. Thin layer chromatography on silica plates with hexane as solvent and iodine as the detection agent separates the thiol (Retardation factor (Rf) 0.5) and the disulfide (Rf 0.25).

Carrying out the hydrolysis in ethanol rather than methanol (78° vs 65°) led to destruction of the thiol. Naud *et al*.(1) recommend differential solubility in ethanol to separate the thiol and the disulfide; this was unsatisfactory in our hands as the yield of thiol was only about 10% and still contained about 10% disulfide.

www.pnas.org/cgi/doi/10.1073/pnas.1817632116

Hydrolysis using HCl according to Graupe *et al.*(2) gave a product after precipitation with water which contained about 10% disulfide. The yield, however, was poorer than the alkaline method.

The products formed from hydrolysis of the ester under these conditions were quantitated by proton NMR in $CDCl_3$. The thiol and the corresponding disulfide have spectra that differ in three ways: The triplet due to the –SH group is absent in the disulfide. The methylene resonances of the disulfide are downfield of those due to the thiol by about 0.1 ppm. The splitting of the $–CH_2S-$ group is a doublet of triplets in the thiol and appears as a pseudo-quartet whereas in the disulfide the resonance is almost a simple triplet except for the four-bond H-F coupling ($^4J_{HF}$= 1.5 Hz).

Attempts to reduce the disulfide using dithiothreitol or sodium borohydride were unsuccessful. However, monolayers of the same structure are formed by both thiols and disulfides(3-5) so that a small quantity of disulfide in these preparations is tolerable.

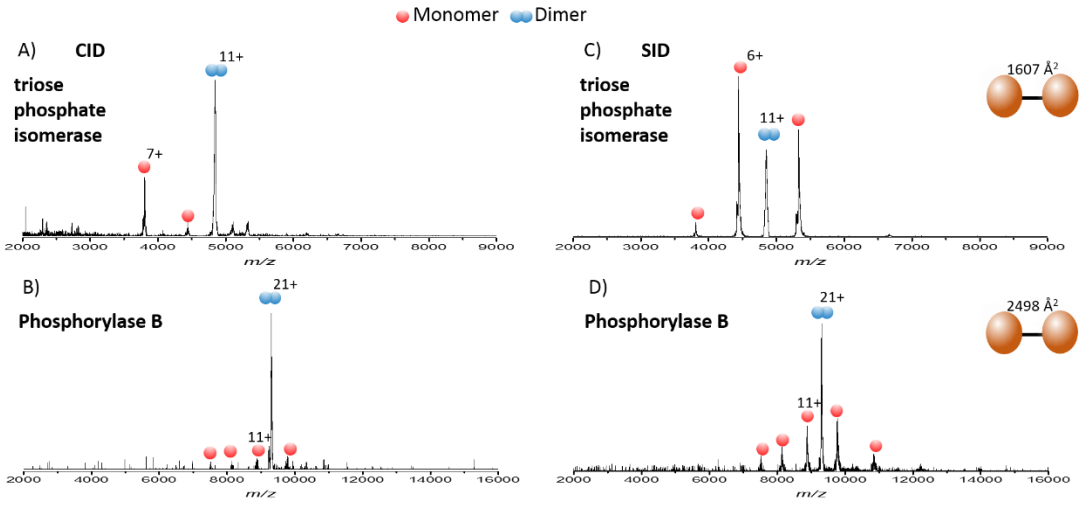

Figure S1: CID spectra (left) obtained for dimeric protein complexes under charge reducing conditions for A) 11+ triose phosphate isomerase at 1980 eV and B) 21+ Phosphorylase B at 4200 eV. SID spectra (right) obtained for dimeric protein complexes under charge reducing conditions for C) 11+ triose phosphate isomerase at 440eV and D) 21+ Phosphorylase B at 1680 eV. Insets are cartoon representations of the solved structure (PDBs 8TIM and 1ABB respectively) with interfacial area labeled, as determined from PISA.

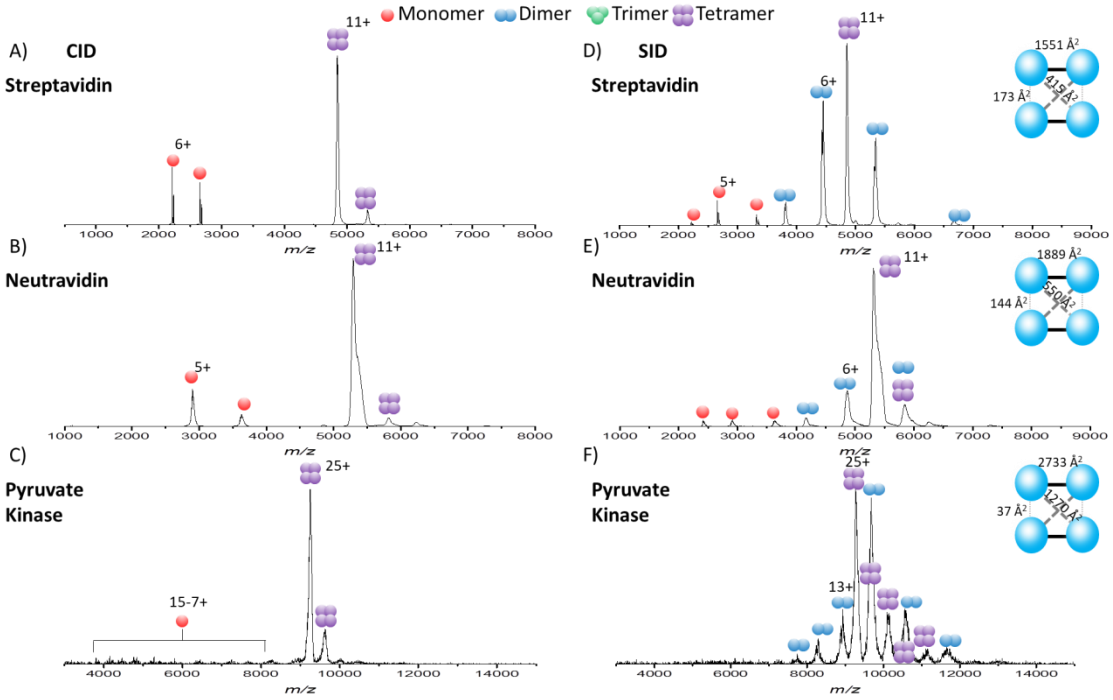

Figure S2: CID spectra (left) obtained for tetrameric protein complexes under charge reducing conditions for A) streptavidin 11+ at 1210 eV B) neutravidin 11+ at 1540 eV, and C) pyruvate kinase 25+ at 4750 eV. Monomer is the dominant product in all CID spectra as expected. When trimer is observed it is not at equal intensity to the monomer, which is attributed to the known higher transmission efficiency of the highly charged monomer in this instrument type. SID spectra (right) obtained for D) streptavidin 11+ at 440eV E) neutravidin 11+ at 440 eV, and F) pyruvate kinase 25+ at 1500 eV. In contrast to CID, dimers are the dominant product here, which is consistent with the dimer-of-dimer structures of these complexes. Insets are cartoon representations of the solved structure (PDBs 1SWB, 1VYO and 1AQF respectively) with interfacial area labeled, as determined from PISA.

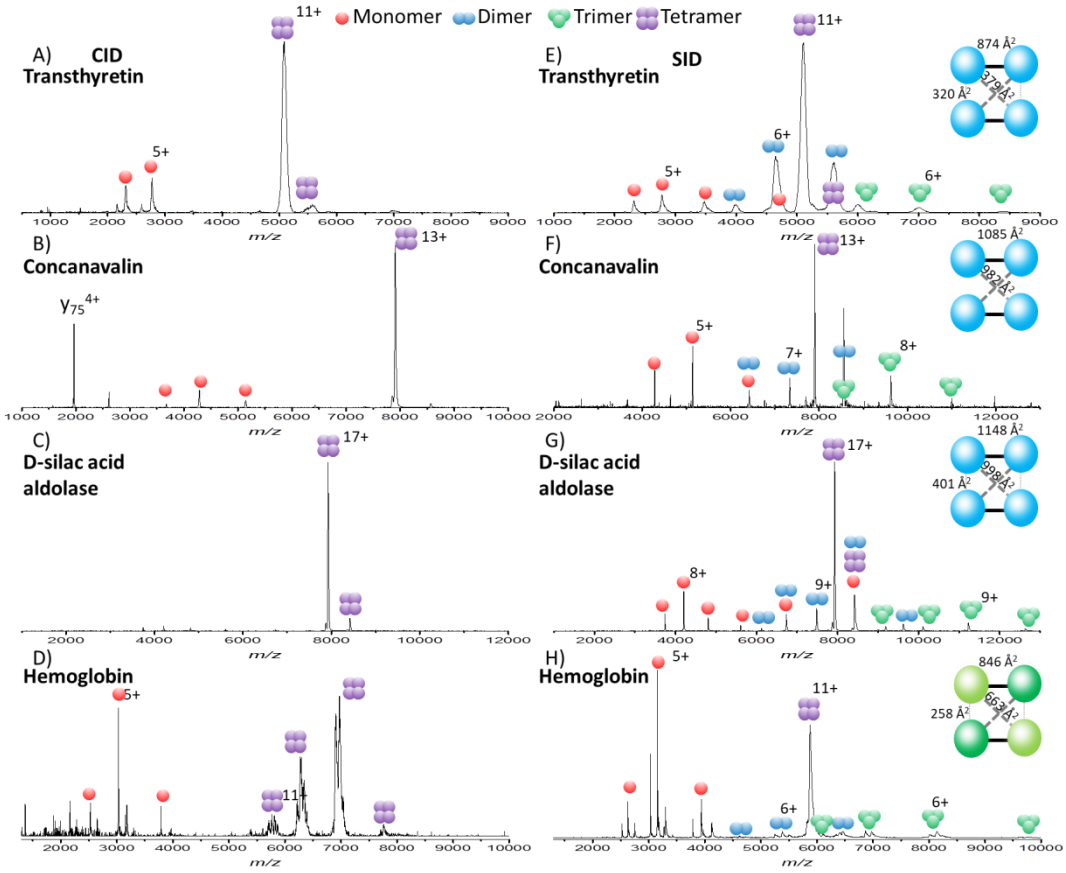

Figure S3: CID spectra (left) obtained for tetrameric protein complexes under charge reducing conditions for A) transthyretin 11+ at 1540 eV B) concanavalin A 13+ at 2080 eV, and C) D-sialic acid aldolase 17+ at 2720 eV, and, D) hemoglobin 11+ at 1650 eV. Monomer is the dominant product in all CID spectra as expected, when trimer is observed it is not at equal intensity to the monomer, which is attributed to the known higher transmission efficiency of the highly charged monomer. Note for CID of ConA the major product is a 7.8 kDa peptide (Y_{75}), as has been previously reported under charge reduced conditions. (6) SID spectra (right) obtained for E) transthyretin 11+ at 440eV F) concanavalin A 13+ at 910eV, G) d- sialic acid aldolase 17+ at 1190eV and H) hemoglobin 11+ at 660 eV, which produces both alpha and beta monomers and can also lose the heme group. Insets are cartoon representations of the solved structure (PDBs 1F41, 3CNA, 3LBM and 1GZX respectively) with interfacial area labeled, as determined from PISA.

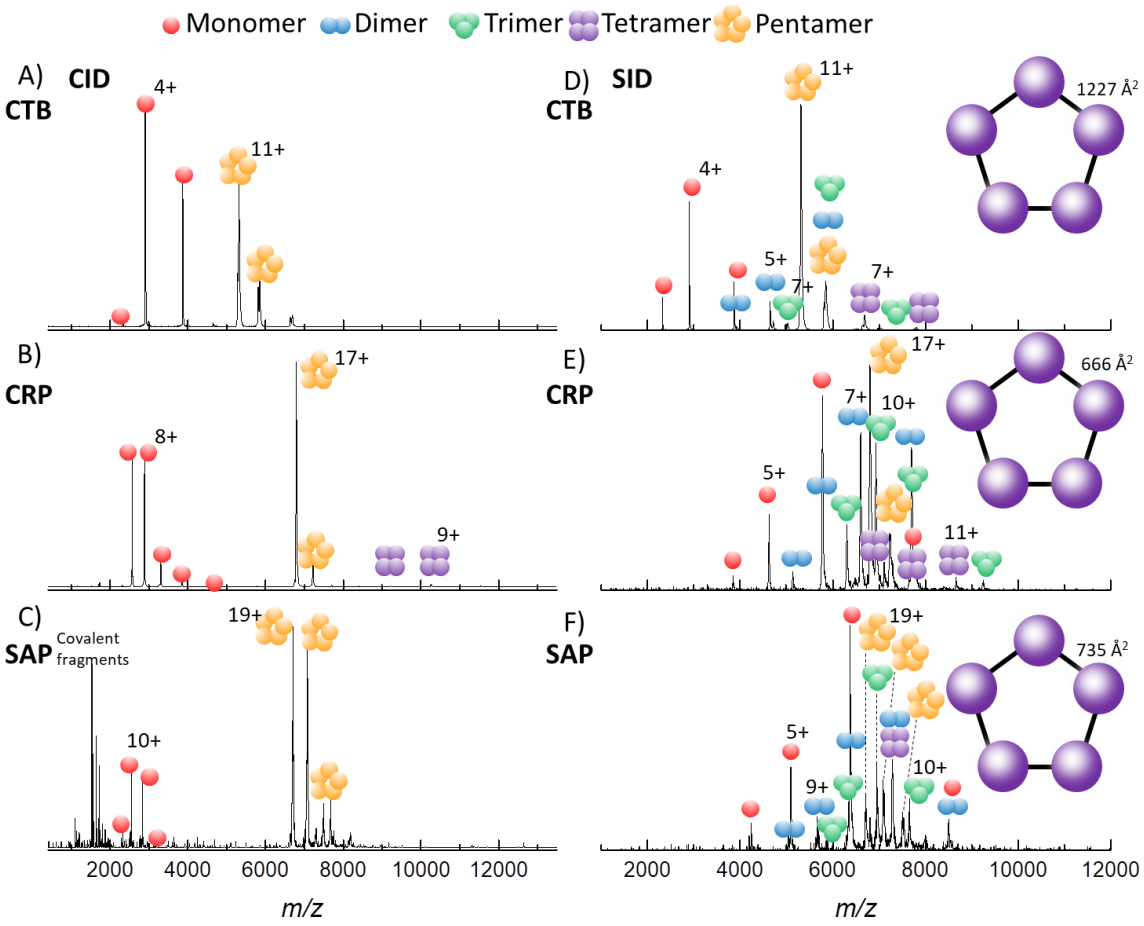

Figure S4: CID spectra (left) acquired for the pentameric protein complexes under charge reducing conditions A) CTB 11+ at 1320 eV B) CRP 17+ at 2040 eV and C) SAP 19+ at 2850 eV. Low energy SID spectra (right) acquired for the pentameric protein complexes under charge reduced conditions D) CTB 11+ at 330 eV, E) CRP 17+ at 680 eV, and F) serum amyloid P 19+ at 760 eV. Insets are cartoon representations of the solved structure (PDBs 1FGB, 1GNH and 1SAC respectively) with interfacial area labeled, as determined from PISA.

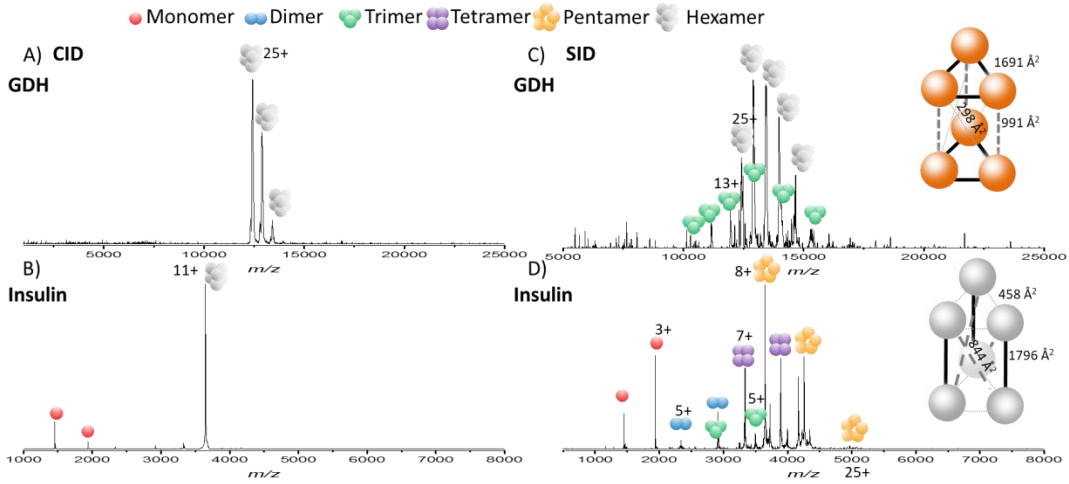

Figure S5: CID spectra (left) acquired for the hexameric protein complexes under charge reducing conditions A) bovine GDH 25+ at 5000 eV, and B) insulin 11+ at 440 eV. Low energy SID spectra (right) acquired for the hexameric protein under charge reduced conditions C) Bovine glutamate dehydrogenase 27+ at 2700 eV, and D) insulin 11+ at 440 eV. Insets are cartoon representations of the solved structure (PDBs 3MVO and 2AIY respectively) with interfacial area labeled, as determined from PISA.

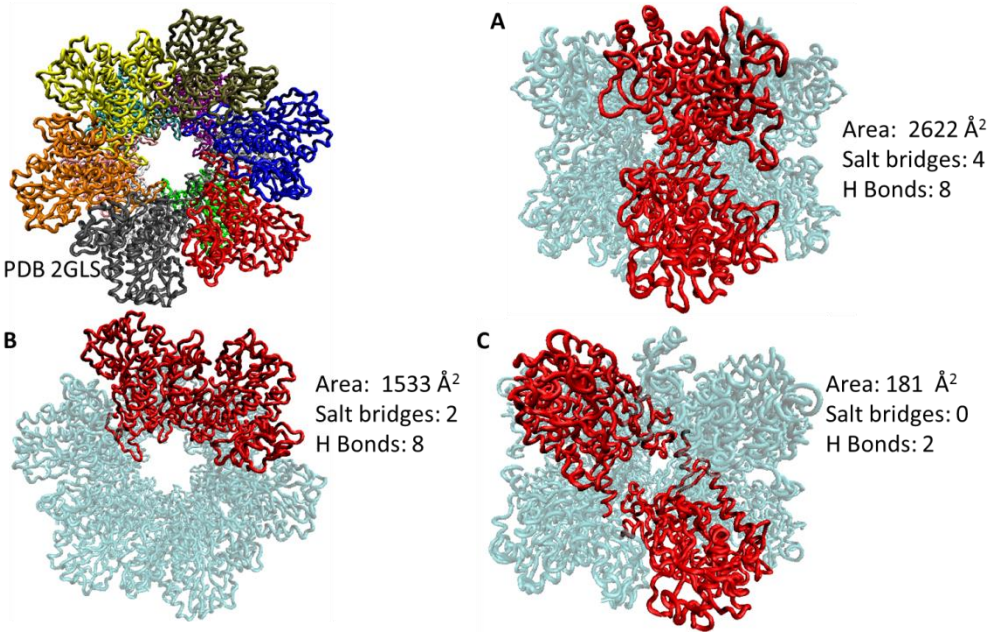

Figure S6: Representation of the interfaces of glutamine synthetase. A, C: side view of stacked rings; B: top view of hexameric ring.

Interfacial Analysis: interface area (IA)/ Å², salt bridges (SB), hydrogen bonds (HB)			Subunits produced Corresponding interfacial area broken on dissociation						Expected product	Observed Product
Glutamine synthetase, 624 kDa (PDB 2GLS)										
A: IA:2622 SB:4 HB:4	B: IA:1533 SB:2 HB:8	C: IA 181 SB:0 HB:2	**M+U:** **6958**	**D+De** **6494**	T+N: 8935	**Q+O:** **6494**	P+He :8935	**2H:** **6494**	M+He, D+De Q+O, 2H	M+He, D+De Q+O, 2H

Table S1: Interfacial information for glutamine synthetase studied by SID in addition to the predicted and observed SID products. Interfacial Analysis was performed with PISA analysis and is reported in the form of interface area (IA)/ Å², salt bridges (SB), hydrogen bonds (HB).For products: M represents monomer, D dimer, T trimer, Q tetramer, P pentamer, H hexamer, He heptamer, De decamer, U undecamer, and Do dodecamer. For glutamine synthetase interfaces are shown in Figure S6, where A and C are inter-ring interfaces and B is the intra-ring interface.

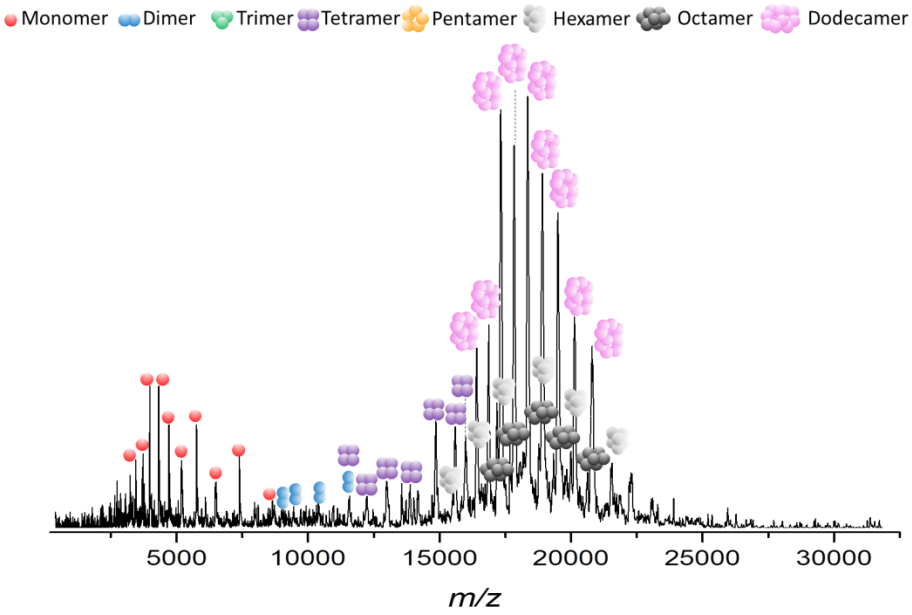

Figure S7: SID spectrum acquired for charge states 36-42+ of glutamine synthetase protein complex acquired at 180 V.

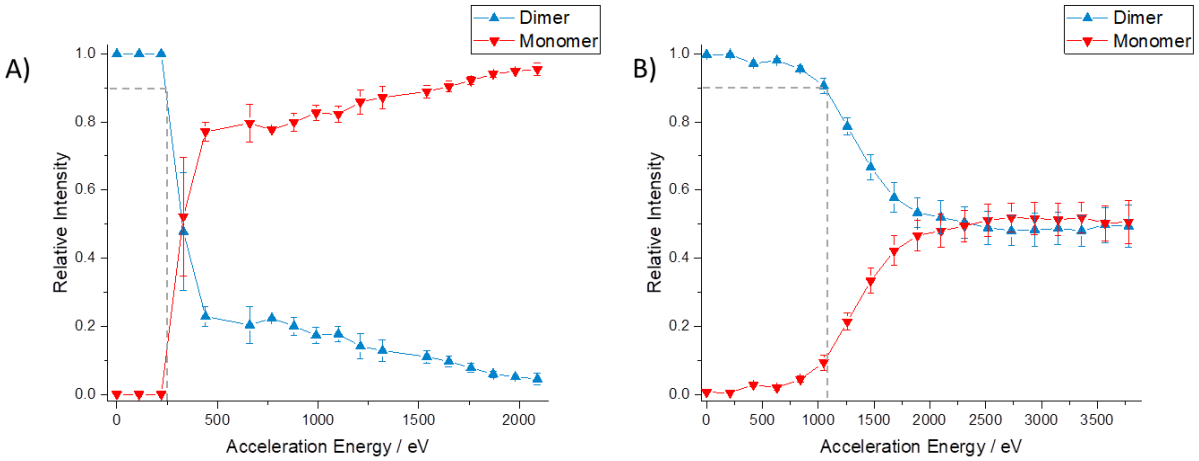

Figure S8: SID energy resolved mass spectrometry (ERMS) plots for A) 11+ triose phosphate isomerase and B) 21+ Phosphorylase B. Relative intensity was determined from the summed intensity of every charge state observed, and error bars represent the standard deviation between three repeats. Dashed grey lines highlight 10% fragmentation, the appearance energy used in computational models.

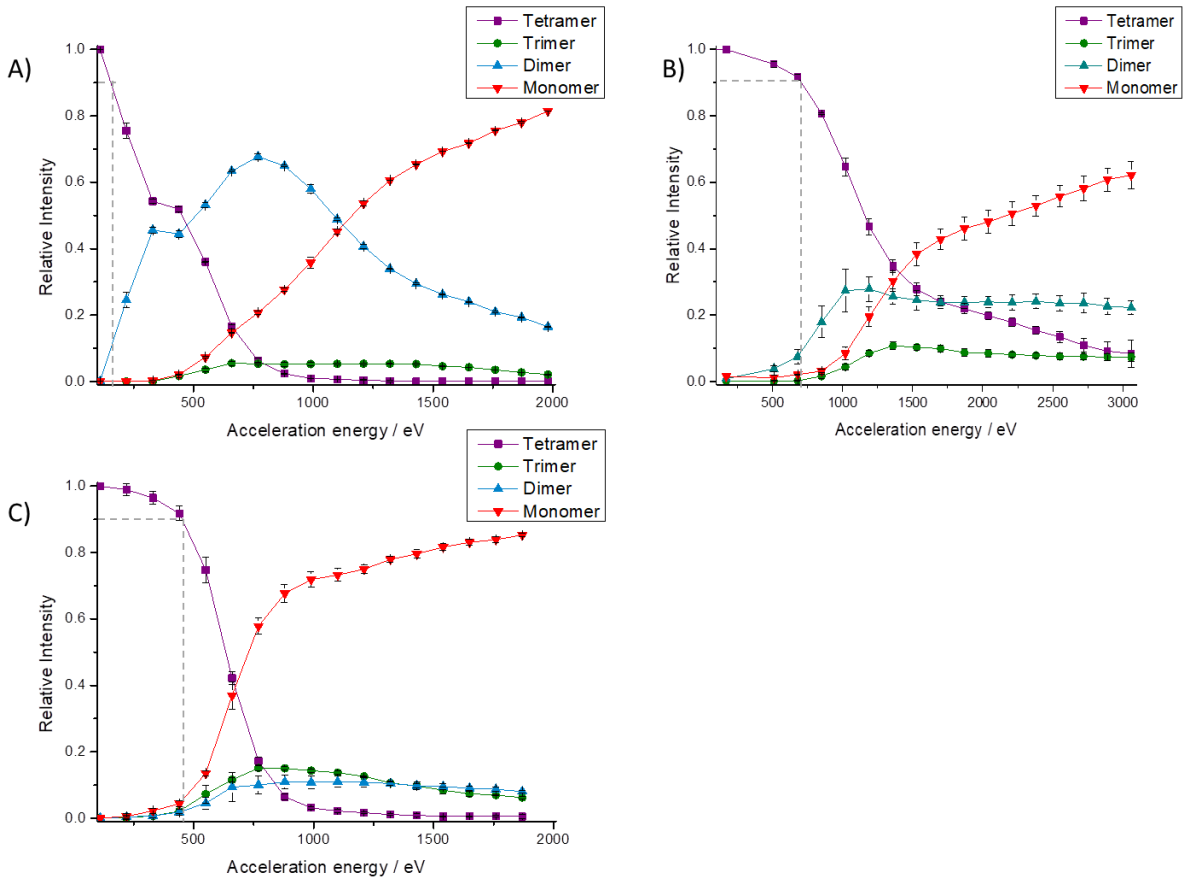

Figure S9: SID ERMS plots for A) 11+ Streptavidin, B) 17+ D-silac acid aldolase, and C) 11+ Hemoglobin. Relative intensity was determined from the summed intensity of every charge state observed, and error bars represent the standard deviation between three repeats. Dashed grey lines highlight 10% fragmentation, the appearance energy used in computational models.

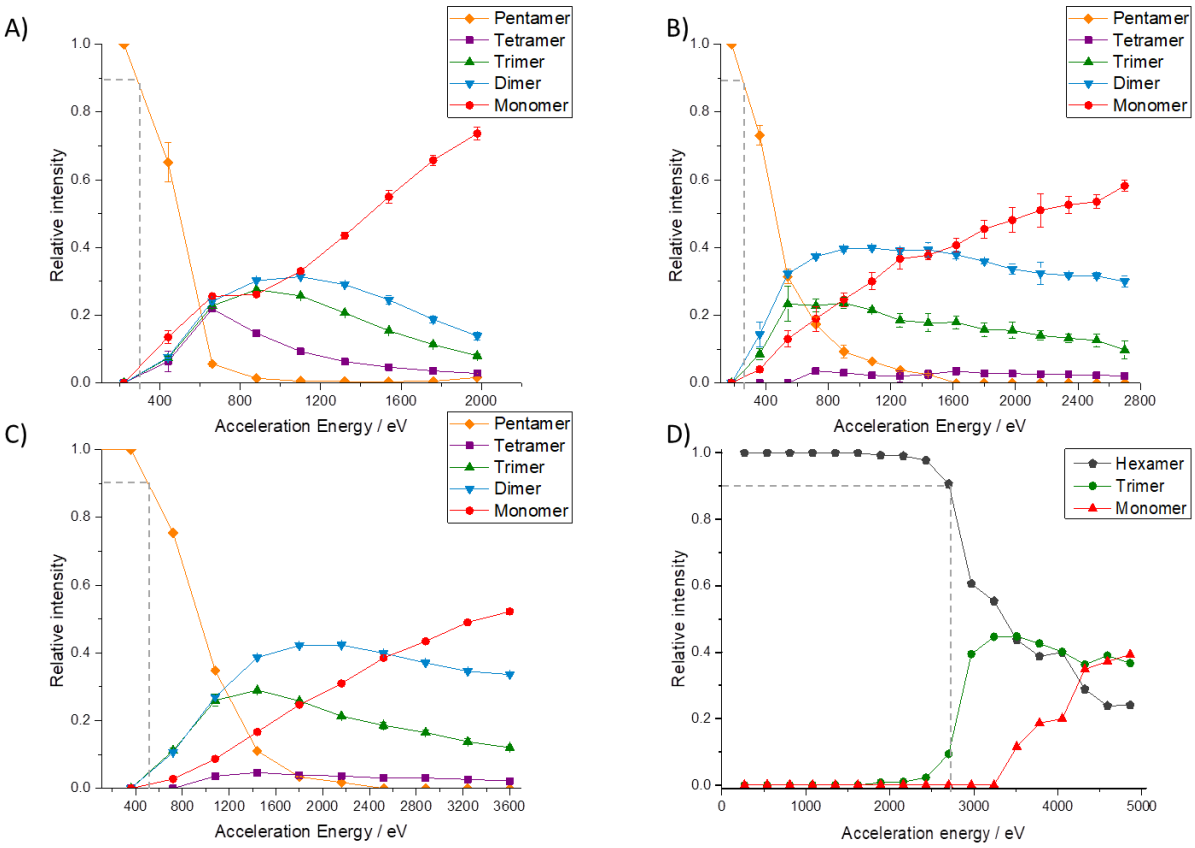

Figure S10: SID ERMS plots for A) 11+ cholera toxin B, B) 18+ C-reactive protein, C) 18+ serum amyloid P, D) Bovine glutamate dehydrogenase. Relative intensity was determined from the summed intensity of every charge state observed, and error bars represent the standard deviation between three repeats. Dashed grey lines highlight 10% fragmentation, the appearance energy used in computational models.

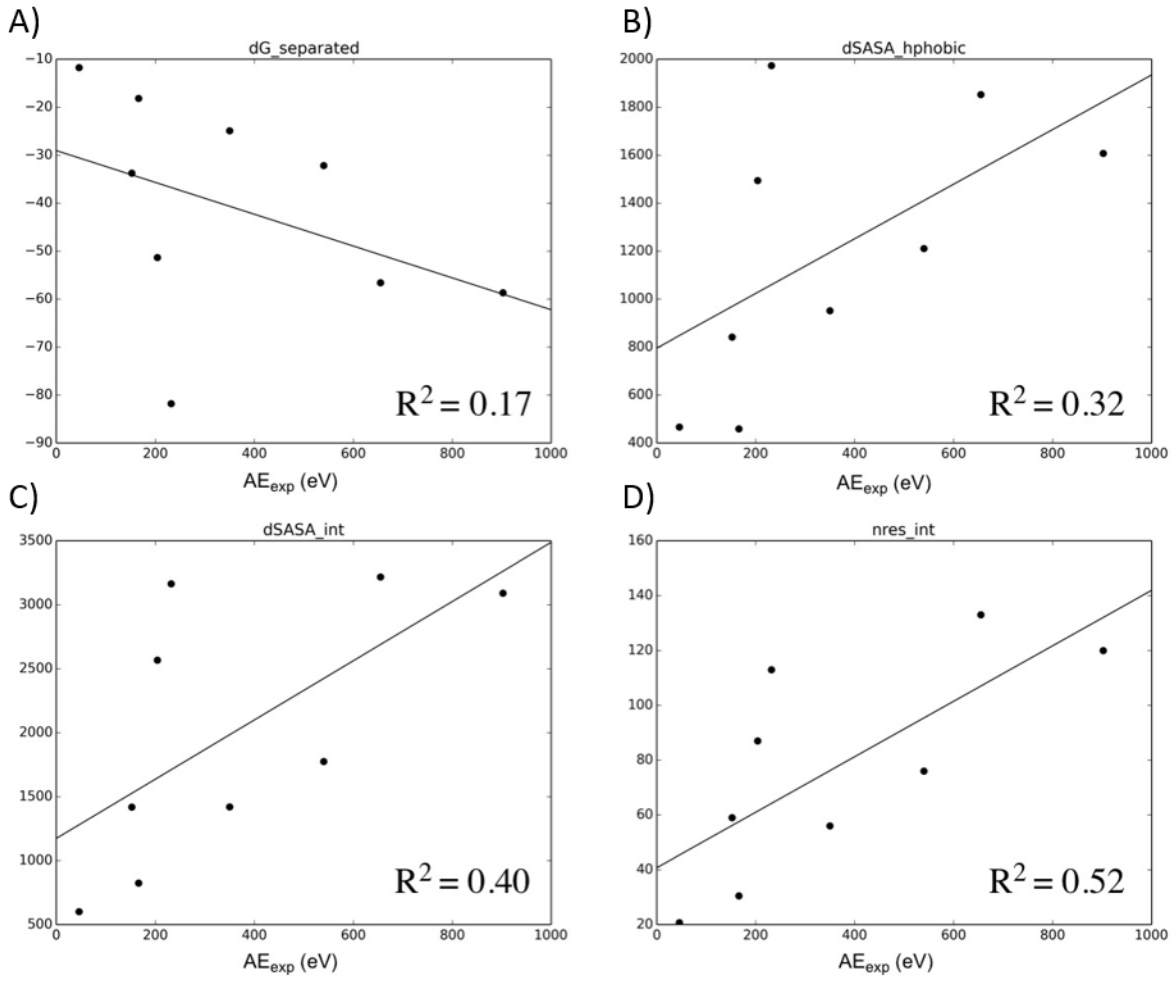

Figure S11: Individual correlations for four of the best calculated interface parameters from Rosetta with experimental DE. (A) dG_separated represents the change in Rosetta energy when the interface subunits are separated (lower corresponds to stronger interface). (B) dSASA_hphobic and (C) dSASA_int represent the hydrophobic and total surface interface areas respectively. (D) nres_int represents the number of interacting residues at the interface. While all shown parameters show some correlation, the correlations are not strong enough individually to use as a predictive model.

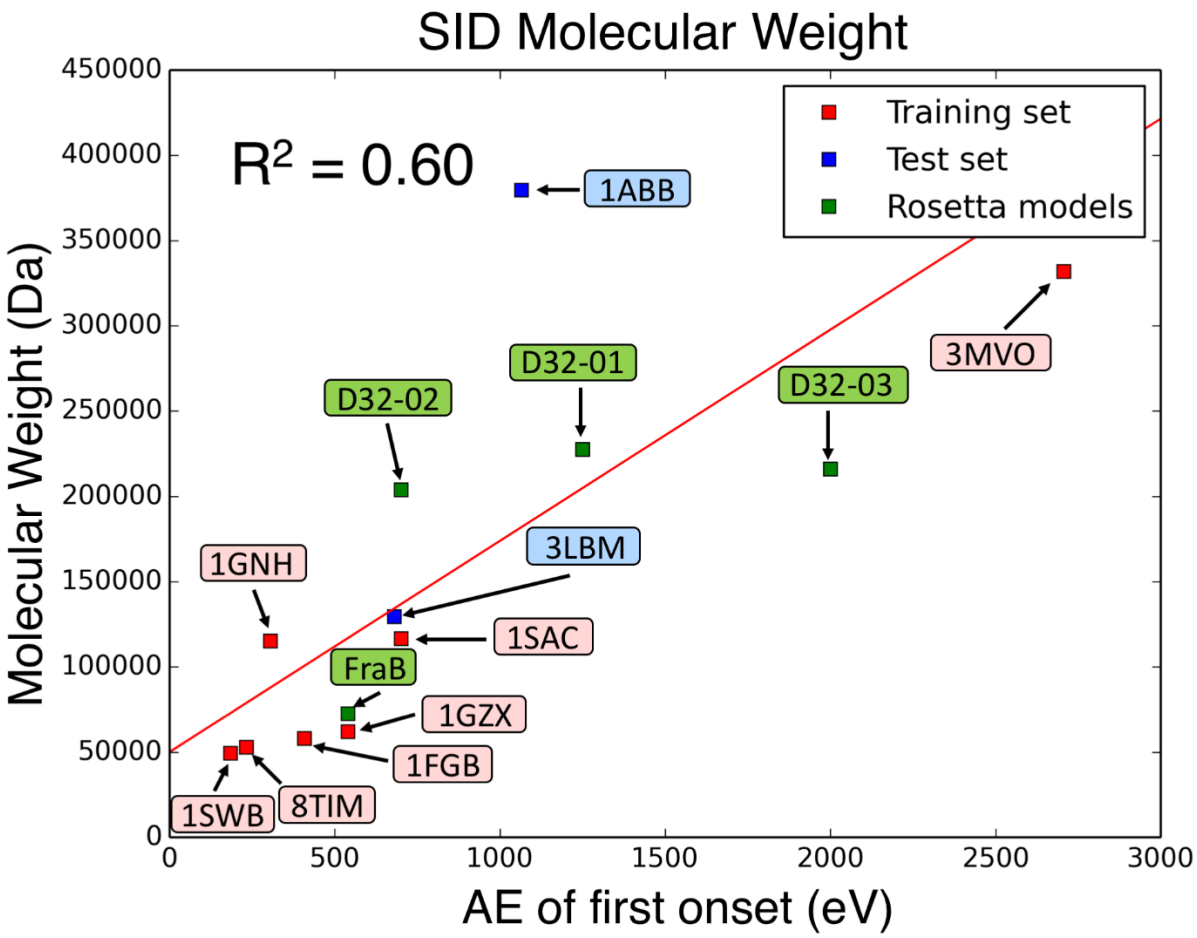

Figure S12: Correlation between molecular weight and un-normalized appearance energy of the first onset. Correlation is lower than the model which combines interfacial properties.

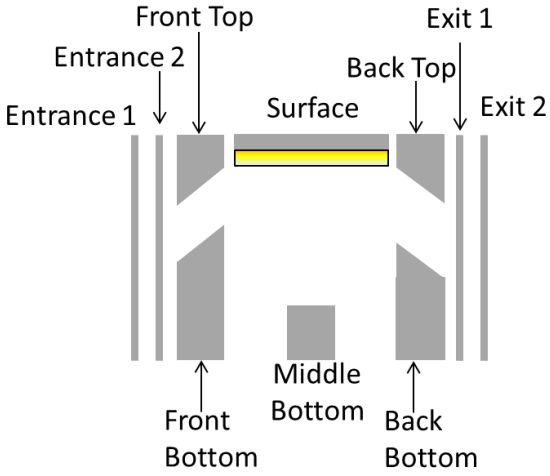

Figure S13: Schematic of the SID device used in the Synapt G2 and G2S platforms.

Lens	Applied potential / V
Entrance 1	10
Entrance 2	-30
Front Top	-90
Front Bottom	5
Middle Bottom	-60
Surface	-30
Back Top	-155
Back Bottom	-58
Exit 1	-55
Exit 2	-62

Table S2: Representative SID lens voltage settings for SID 50 V on a Waters Synapt G2 mass spectrometer, operating in IM mode. Kinetic energy is defined by the SID voltage times the charge state.

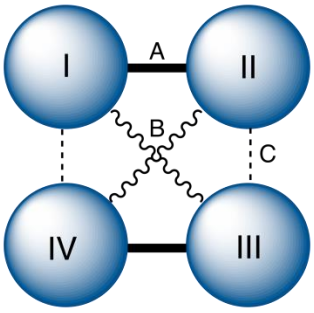

Figure S14: Cartoon representation of a tetrameric protein complex, highlighting the different subunits and interfaces that could be present.

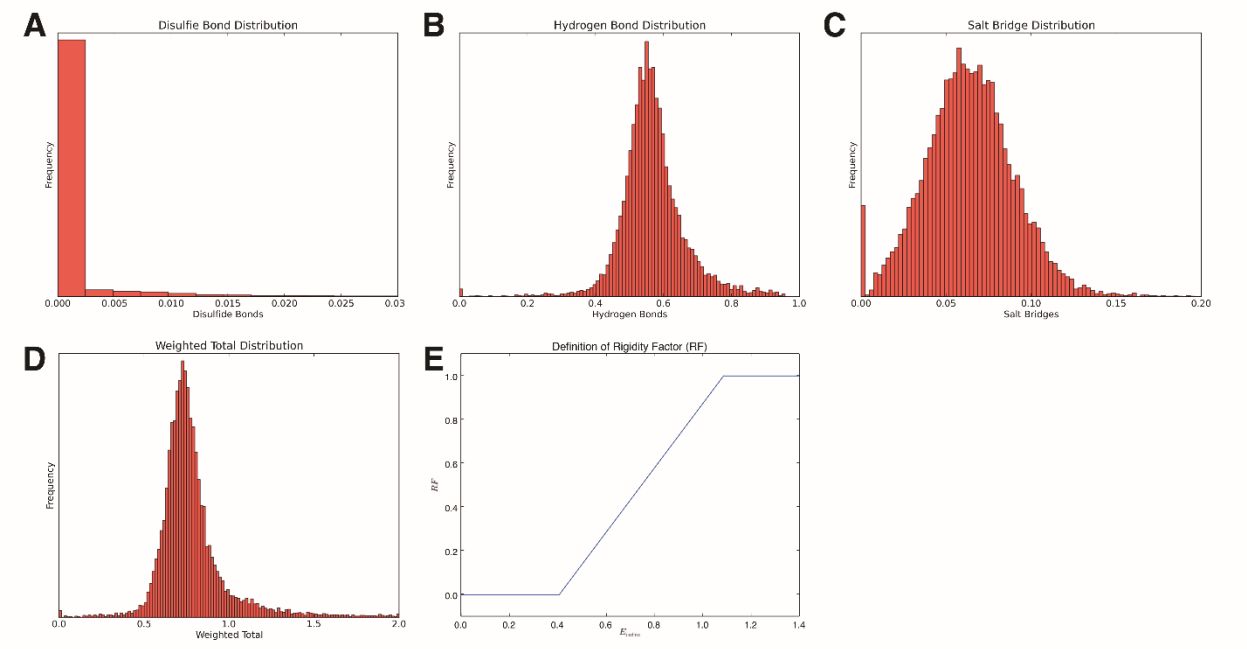

Figure S15: Individual distributions for disulfide bonds (A), hydrogen bonds (B), and salt bridges (C) for the 12,591 unique, non-redundant subunits culled from PISCES. (D) shows the E_{intra} distribution, which is a relative approximation of the total intramolecular energy of each subunit, as shown in Equation 2. (E) shows the definition of RF, calculated using E_{intra}. RF was defined to be 0 if E_{intra} was less than 2 standard deviations from the mean, 1 if E_{intra} was greater 2 standard deviations from the mean, with a linear extrapolation for subunits lying between the extrema.

References

1. Naud C, Calas P, Blancou H, & Commeyras A (2000) Synthesis of terminally perfluorinated long-chain alkanethiols, sulfides and disulfides from the corresponding halides. *J. Fluorine Chem.* 104(2):173-183.
2. Graupe M*, et al.* (1999) Terminally perfluorinated long-chain alkanethiols. *J. Fluorine Chem.* 93(2):107-115.
3. Schönherr H & Ringsdorf H (1996) Self-assembled monolayers of symmetrical and mixed alkyl fluoroalkyl disulfides on gold. 1. Synthesis of disulfides and investigation of monolayer properties. *Langmuir* 12(16):3891-3897.
4. Grönbeck H, Curioni A, & Andreoni W (2000) Thiols and disulfides on the Au (111) surface: the headgroup− gold interaction. *J. Am. Chem. Soc.* 122(16):3839-3842.
5. Love JC, Estroff LA, Kriebel JK, Nuzzo RG, & Whitesides GM (2005) Self-assembled monolayers of thiolates on metals as a form of nanotechnology. *Chem. Rev.* 105(4):1103-1170.
6. Zhou M, Dagan S, Wysocki VH, (2013) Impact of Charge State on Gas-Phase Behaviors of Noncovalent Protein Complexes in Collision Induced Dissociation and Surface Induced Dissociation. Analyst 138:1353-1362.