THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Letter

pubs.acs.org/JPCL

Probing the Acid-Induced Packing Structure Changes of the Molten Globule Domains of a Protein near Equilibrium Unfolding

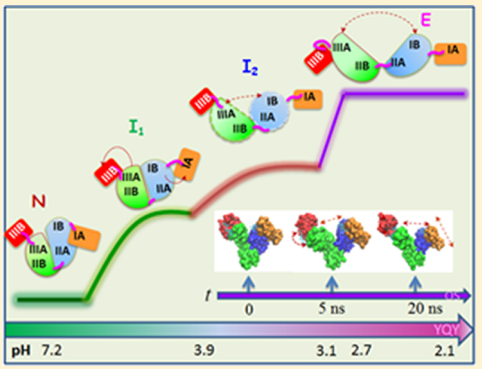
Yi-Qi Yeh,[†] Kuei-Fen Liao,[†] Orion Shih,[†] Ying-Jen Shiu,[†] Wei-Ru Wu,[†] Chun-Jen Su,[†] Po-Chang Lin,[†] and U-Ser Jeng*[,†,‡]

[†]National Synchrotron Radiation Research Center, Hsinchu 30076, Taiwan

[‡]Department of Chemical Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan

Supporting Information

ABSTRACT: Using simultaneously scanning small-angle X-ray scattering (SAXS) and UV−vis absorption with integrated online size exclusion chromatography, supplemental with molecular dynamics simulations, we unveil the long-postulated global structure evolution of a model multidomain protein bovine serum albumin (BSA) during acid-induced unfolding. Our results differentiate three global packing structures of the three molten globule domains of BSA, forming three intermediates $\mathbf{I_1}$, $\mathbf{I_2}$, and **E** along the unfolding pathway. The $\mathbf{I_1}$−$\mathbf{I_2}$ transition, overlooked in all previous studies, involves mainly coordinated reorientations across interconnected molten globule subdomains, and the transition activates a critical pivot domain opening of the protein for entering into the **E** form, with an unexpectedly large unfolding free energy change of −9.5 kcal mol^{-1}, extracted based on the observed packing structural changes. The revealed local packing flexibility and rigidity of the molten globule domains in the **E** form elucidate how collective motions of the molten globule domains profoundly influence the folding−unfolding pathway of a multidomain protein.



The residual tertiary structures of protein intermediates carry profound implications in understanding the protein folding−unfolding pathway or functioning.[1−8] Hydrophobic effects, resulting in a collapsed core of hydrophobic amino acids, are considered the major determinant of protein structural stability in aqueous solution.[3,8] Enhanced conformational entropy under environmental stimulations, however, competes with the van der Waals interactions for protein structural transitions toward functioning or folding−unfolding.[6−8] Recent progress has established a concept of molten globule domains for intrinsic disordered or partly unfolded proteins,[1−6] characterized by substantial secondary structure and a fluctuating tertiary structure with perturbed side-chain packing and hydrophobic cores (with or without core−water solvation).[3,8] The packing structure and melting sequence of the molten globule domains of a multidomain protein further influence the folding−unfolding pathway. Advances in molecular dynamics simulations have foreseen the dependence of folding−unfolding pathways of a multidomain protein on its topology of domain connectivity.[7,8] Nevertheless, experimental limitations have so far largely prohibited probing the global packing structures of the molten globule domains of multidomain proteins.

Probing the morphological evolution of protein intermediates has been challenging because of the generally less rigid structures that can blur or blind spectroscopic tools to different extents. These noncrystalline protein structures in solution exclude the use of powerful protein X-ray crystallography, whereas large and disordered structures often limit NMR sensitivity. Moreover, coexistence of protein intermediates further complicates the structural measurements and data interpretation. Nevertheless, optical spectroscopies including circular dichroism (CD), light absorption, fluorescence, and resonance Raman scattering are used to capture complementary local structural information on protein intermediates during folding−unfolding.[9,10] On the other hand, global structural conformations of native proteins in solution are increasingly better determined using small-angle X-ray or neutron scattering (SAXS and SANS).[11−16] Recent developments on combining SAXS with an online size exclusion high performance liquid chromatography (HPLC) column allow scanning along the elution path for SAXS data collection with online purified or separated protein species.[17] Rapid advance of model simulation with SAXS data further allows incorporation of local structural information (as fitting constraints) obtained from different structural probes in an integrated reconstruction of protein structure.[18] These developments have incubated a new approach that can break the conventional constraints on probing the packing structures and transitions of the molten globule domains in protein folding−unfolding pathways.

In this study, we perform scanning SAXS/UV−vis absorption measurements along the elution path of HPLC to probe the

Received: November 21, 2016
Accepted: January 3, 2017

DOI: 10.1021/acs.jpclett.6b02722
J. Phys. Chem. Lett. 2017, 8, 470−477

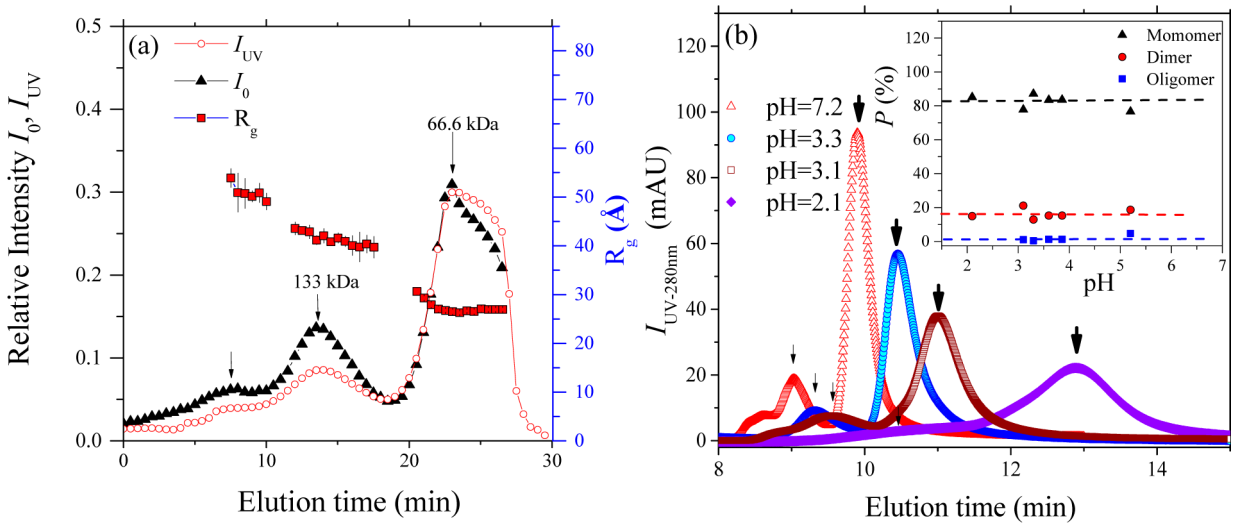


Figure 1. R_g and I_0 profiles extracted from the SAXS data measured along the chromatogram I_{UV} (280 nm wavelength) of a BSA solution at pH 7.2. The arrows mark the separated monomer, dimer, and oligomer peaks of BSA (molar mass selectively indicated). A slower HPLC elution rate was used for improved SAXS data collection time. (b) Representative chromatograms for BSA in the solutions at pH 7.2−2.1. The monomer and dimer elution peaks are marked with thick and thin arrows, respectively. The inset shows the molar fractions P extracted from the corresponding integrated peak areas of the species indicated.

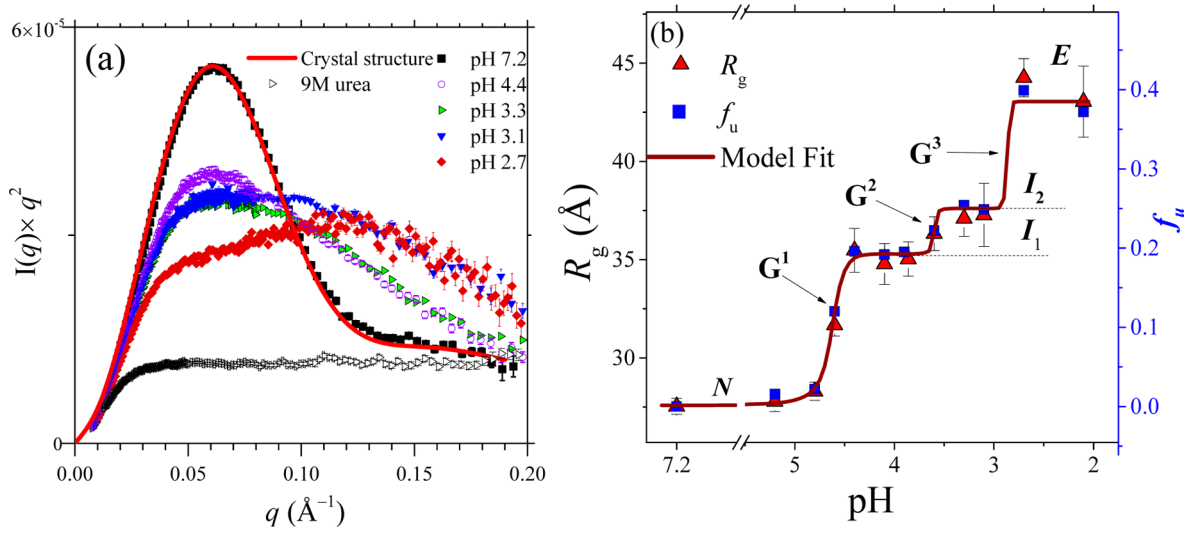


Figure 2. (a) pH-dependent SAXS data of BSA in the Kratky−Porod presentation. The SAXS data of the native form at pH 7.2 can be described well (solid profile) using the BSA crystal structure. SAXS data for an unfolded BSA with 9.0 M urea are shown for comparison. (b) Corresponding R_g values of BSA in the acid-induced unfolding. The unfolding fraction profile f_u (Y-axis to the right-hand-side) deduced from the protein unfolding dimension D_v is fitted (solid curve) using the modified Ising model with the three-group unfolding $\mathbf{G}^1$, $\mathbf{G}^2$, and $\mathbf{G}^3$ indicated for the transitions from the native state (**N**) to the intermediates $\mathbf{I}_1$, $\mathbf{I}_2$, and **E**, as marked.

intermediates of a model protein bovine serum albumin (BSA) partially unfolded in acid solutions between pH 2.1 and 7.2. The heart-shaped serum albumin, a transport protein comprised of three major domains, is one of the most abundant circulatory proteins in mammal blood.[19] Due to the distinguished multidomain structure, serum albumin can conform for noncovalent associations with fatty acids, hemin, bilirubin, and drugs, as a carrier.[19,20] Such nature renders BSA one of the most extensively utilized proteins in laboratory practice, serving as a human serum albumin substitute on many occasions. Despite extensive applications, the structural complexity has delayed BSA crystal structure determination until lately.[21] On the basis of local structural features, earlier studies[22−28] established two intermediates of serum albumin in acid-induced unfolding, termed as the fast (**F**) and expanded (**E**) forms.[9,23−28] The global packing structure of the three molten globule domains that constitute the intermediates, however, remains postulated to date. Here, we show that the scanning HPLC-SAXS approach can effectively exclude scattering contamination from coexisting oligomer species and collect SAXS of individual protein intermediates along the HPLC elution path. Integrating local structural information in the SAXS-based structure modeling[29,30] (rigid components mixed with flexible loops of a protein), we further elucidate subtle coordinated reorientations across interconnected molten globule subdomains, leading to a cleft in the domain−domain interfaces, prior to a critical pivot opening of the domains into an expanded intermediate of BSA. Such stepwise structural

DOI: 10.1021/acs.jpclett.6b02722
J. Phys. Chem. Lett. 2017, 8, 470−477

transitions cannot be explained by the conventional two-state model with an unfolding fraction determined from the populations of coexisting folded and unfolded proteins. A recently modified Ising model[14] is, therefore, adopted to interpret the observed multistep structural changes in acid-induced BSA unfolding, in terms of a topological collection of folding−unfolding units that constitute a protein. The modified Ising model provides a unique view to interpret the multistep unfolding process of BSA observed and allow extraction of the corresponding unfolding free energy changes of the grouped folding−unfolding units.

Figure 1a illustrates the zero-angle intensity I_0 and R_g values extracted using the Guinier approximation[11,12] from the SAXS data measured along the HPLC elution path of the native BSA sample solution at pH 7.2 (Figure S1 in the Supporting Information (SI)). The successively differentiated peaks indicate coexisting species in the sample solution. The last two peaks, corresponding to calibrated molecular masses of ∼132 and 66 k Da, are assigned to BSA dimer and monomer, respectively. The corresponding R_g values, 40.0 ± 0.5 and 28.1 ± 0.5 Å, are consistent with those calculated from the crystal structures of BSA dimer and monomer (R_g = 40.0 ± 0.2 and 27.5 ± 0.1 Å, as shown in Figure S2).[21] Shown in Figure 1b are the representative chromatograms measured for the BSA solutions at pH 7.2−2.1, each exhibiting three similar elution peaks as those in Figure 1a. Interestingly, the integrated peak areas of the three species remain of the same (molar) ratio of 82:16:2 for the monomer/dimer/oligomer in the solutions regardless of the pH values (inset of Figure 1b), implying no transformation among the three species in the pH range. The BSA monomer elution peak, however, systematically delayed and broadened with the decrease of the pH values. Such behavior is attributed to gradually enhanced nonspecific interactions between the increasingly positively charged BSA and the silica-based chromatography matrix (of negative surface charge) in the acidic environments below the protein isoelectric point (ca. pH 5). A previous report suggested an effective charge of +21 e (or a net charge of +65 e, which is consistent with our molecular dynamics simulation result) for BSA at pH 2.4; in contrast, the native protein has an effective charge of −8 e (or a net charge of −11 e at pH 6.8).[31] Most importantly, the monomer peaks for the BSA solutions at pH 3.3, 3.1, and 2.1 are separable and overlap little with that of the native state at pH 7.2 in the elution time, as shown in Figure 1b, suggesting that there is little coexistence of the BSA intermediates with the native form in the solutions.

Figure 2a illustrates the scanning HPLC-SAXS data corresponding to the chromatographs shown in Figure 1b, in the Kratky−Porod presentation.[32,33] The bell-shaped profile centered at $q_c \approx 0.06$ Å^{-1} for the BSA solution at pH 7.2 reveals a globular structure, with an estimated R_g of ∼28 Å using $q_cR_g \approx 3^{1/2}$.[32] When the solution pH value is lowered to 2.7−2.1, a relatively smeared and extended hump covering a wide q range between 0.05 and 0.2 Å^{-1} implies an extended structure.[32−35] For BSA partially unfolded in solution pH 4.8−3.1, SAXS profiles transit gradually from the bell-shape to the extended shape. For comparison, SAXS data for an unfolded BSA with 9 M urea (Figure 2a)[34] exhibit a characteristic plateau region of random coil behavior in the higher-q regime (>0.05 Å^{-1}).[32,33] Exhibited in Figure 2b are all of the pH-dependent R_g values of the acid-unfolded BSA extracted from the SAXS profiles simulated with the best-fitted models (as representatively shown in Figure 3), which are consistent with those obtained

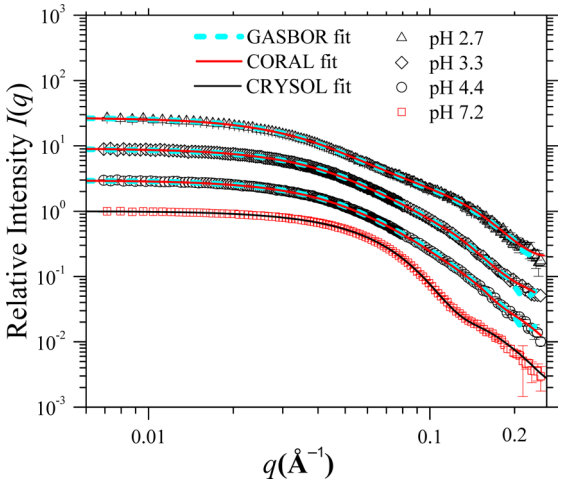


Figure 3. Representative SAXS data of BSA partially unfolded at the pH values indicated. The data are respectively fitted using the GASBOR (dashed curves) and CORAL (solid curves) models illustrated in Figure 4. The native BSA data (pH 7.2) are fitted (solid curve) using CRYSOL with PDB (4F5S).

using the Guinier approximation (Figure S3 and Table S1).[33] Concentration-independent SAXS profiles suggest little effect of interprotein interactions on these R_g values (Figure S4). These pH-dependent R_g values exhibit three-step transitions, beyond the one-step transition of the conventional two-state model, characterizing transitions from the native (N) to the $\mathbf{I}_1$, to the $\mathbf{I}_2$ intermediate, and then to the commonly recognized expanded **E** form. We note that our UV−vis absorption data simultaneously observed with the HPLC-SAXS at the same sample position reveal a smeared transition over $\mathbf{I}_1$ and $\mathbf{I}_2$ (Figure S5); previous CD and fluorescence results could not resolved these two intermediates $\mathbf{I}_1$ and $\mathbf{I}_2$ and smeared them as the **F** form.[36,37] The uniquely SAXS-revealed $\mathbf{I}_1$−$\mathbf{I}_2$ and $\mathbf{I}_2$−**E** transitions involve largely reorganization of the packing structure of the molten globule domains, which might be insensitive to local structural probes.

The multistep unfolding behavior of BSA shown in Figure 2b obviously cannot be interpreted using the convention two-state model, which assumes that the folded and unfolded proteins are populated according to a Boltzmann distribution determined by an unfolding free energy change and temperature of the system.[16] Moreover, the chromatographs in Figure 1b suggest little coexistence of **N** with the $\mathbf{I}_1$, $\mathbf{I}_2$, or **E** states in the acid-induced unfolding. Therefore, we adopt a modified Ising model recently developed for protein folding−unfolding that can describe the multistep unfolding behavior of BSA observed.[38−41] Briefly, the modified Ising model regards a protein as a topological collection (ensemble) of interacting units; each unit may take either the folded or unfolded state.[38−41] Both thermodynamic and near-equilibrium kinetic characteristics of protein folding−unfolding are assumed to be determined by the elementary process in folding−unfolding of the structural units inside the protein, expressed as the protein unfolded fraction f_u. When all units of a protein are in the folded (or unfolded) state, the protein behaves as fully folded with f_u = 0 (or unfolded with f_u = 1.0). The value of f_u can be between zero and unity for partially folded proteins, and all of the protein in the same environment can be in the same f_u of a common conformation, as revealed by our HPLC elution result. Furthermore, evolution of the protein morphology in a folding−unfolding process is

DOI: 10.1021/acs.jpclett.6b02722
J. Phys. Chem. Lett. 2017, 8, 470−477

Table 1. Thermodynamic Parameters Fitted with the Modified Ising Model for BSA in pH-Induced Unfolding[a]

group	γ	$2\Delta\varepsilon_m^I$ (kcal mol^{-1} M^{-1})	J^I (kcal mol^{-1})	ΔG_m^I (kcal mol^{-1})	ΔG_o^I (288 K) (kcal mol^{-1})	ΔG_J^I (kcal mol^{-1})
$\mathbf{G^1}$	0.19	$-(9.68 \pm 0.40) \times 10^4$	0.13 ± 0.01	-3.9 ± 0.2 (pH 4.4)	2.3 ± 0.1	-2.1 ± 0.2
$\mathbf{G^2}$	0.06	$-(1.96 \pm 0.44) \times 10^4$	0.28 ± 0.02	-9.8 ± 0.3 (pH 3.3)	4.9 ± 0.1	-6.0 ± 0.5
$\mathbf{G^3}$	0.14	$-(1.60 \pm 0.16) \times 10^4$	-0.13 ± 0.01	-32.0 ± 0.5 (pH 2.7)	21.9 ± 0.2	-9.5 ± 0.5

[a]The three folding−unfolding groups $\mathbf{G^1}$, $\mathbf{G^2}$, and $\mathbf{G^3}$ together contribute 39% (f_u) of the protein unfolding in D_v. The values for the free energy changes ΔG_m^I are deduced at 288 K. Note that the positive and negative J values of the three groups contribute respectively to the conformation instability and stability of the intermediates.

governed by the free energy changes of the folding−unfolding units of the protein in the modified Ising model and can have stepwise changes when the units are of distinct unfolding free energy changes. Thereby, the unfolding behavior of a protein containing multigroups (I) can be described by accumulated group unfolding $f_u = \sum_I \gamma_I f_u^I$, with the summation of the weighting coefficients γ_I (relative populations of the folding−unfolding unit) normalized to unity.

Under a mean-field approximation, the unfolded fraction of the Ith group of the modified Ising model is given by[38−41]

$$f_u^I = \left\{1 + \exp\left[\frac{\Delta G^I + 4J^I(1 - 2f_u^I)}{k_B T}\right]\right\}^{-1} \quad (1)$$

assuming that the canonical ensemble lies not far from thermodynamic equilibrium. In the above expression, only nearest-neighbor interactions (J) between folding−unfolding units in the same group are considered, and the unfolding free energy change of the Ith group is expressed as

$$\Delta G_J^I = \Delta G^I(C_m, T) + 4J^I(1 - 2f_u^I) \quad (2)$$

with $\Delta G_J^I = \Delta G_m^I + \Delta G_o^I(T)$ determined by the denaturant free energy change $\Delta G_m^I = 2\Delta\varepsilon_m^I C_m$ with the coefficient $2\Delta\varepsilon_m^I$ and the denaturant concentration C_m (in our case, the proton concentration), together with the thermal unfolding free energy change $\Delta G_o^I(T)$.[38−41]

We further propose an r^3-weighted protein dimension D_v as a measure of protein unfolding dimension, which is defined by the distance distribution function $p(r)$[11,42,43] Fourier transformed from the SAXS data as

$$D_v^3 = \frac{\int_0^{D_{max}} r^3 p(r)\, dr}{\int_0^{D_{max}} p(r)\, dr} \quad (3)$$

with $p(r)$ approaching zero smoothly at the maximum dimension $r = D_{max}$. We define the unfolding fraction profile as[40,41]

$$f_u = (D_v - D_{v,N})/(D_{v,N} - D_{v,U}) \quad (4)$$

with the minimum $D_{v,N}$ of the native state (at pH 7.2) and the maximum $D_{v,U}$ of the unfolded BSA (with the 9 M urea case). As shown in Figure 2b, the hence deduced unfolding fraction profile $f_u(D_v)$ of BSA can be satisfactorily described by a three-group unfolding of $\mathbf{G^1}$, $\mathbf{G^2}$, and $\mathbf{G^3}$, with the best fitted thermodynamics parameters shown in Table 1. These parameters suggest that a relatively larger group with γ^1 = 0.19 of $\mathbf{G^1}$ unfolds completely for the intermediate $\mathbf{I_1}$ when the pH value is ∼4.4, which is followed by the unfolding of a relatively smaller group $\mathbf{G^2}$ (γ^2 = 0.06) when the pH value reaches pH 3.3 for the intermediate $\mathbf{I_2}$. At pH 2.7, the third group $\mathbf{G^3}$ (γ^3 = 0.14) unfolds to the **E** form. With the three groups unfolded, the **E** form has an accumulated f_u = 0.39 (Figure 2b) and carries a substantial residual tertiary structure, as illustrated below. The corresponding group unfolding free energy changes of $\mathbf{G^1}$, $\mathbf{G^2}$, and $\mathbf{G^3}$ reinforce successively from ΔG_J^I = −2.1 to −6.0 to −9.5 kcal mol^{-1} (Table 1), in the stepwise multigroup unfolding. We note that the previously reported unfolding free energy change of ∼6.0 kcal mol^{-1} for BSA in a couple of occasions,[3,44] on the basis of local structural changes, likely corresponds to the unfolding free energy change ΔG_J^2 of the $\mathbf{G^2}$ group here. The significantly larger ΔG_J^3 = −9.5 kcal mol^{-1} of the $\mathbf{G^3}$ group unfolding (for the $\mathbf{I_2}$−**E** transition) revealed uniquely in our study can be attributed to large global packing structure changes associated with the pivot domain interactions of BSA, as detailed from the structural analysis below. We note that the proposed free energy changes for the respective transitions are based on the assumption of reversible transitions between the acid-induced intermediates and the native state, which is well supported by an observed refolding of BSA from the **E** form to the native state (Figure S6).

To reveal the corresponding packing structure changes of the molten globule domains in the three-step unfolding, we performed model simulations with the SAXS data. Our SAXS data analysis strategy is to obtain global protein envelopes of the intermediates using *ab initio* modeling of GASBOR for reference.[42] Guided by the best-fitted envelope morphology of the acid-induced intermediates of BSA, we further modeled the same set of SAXS data with the CORAL algorithm[43] that allows construction of structures from mixed molten globule domains (as rigid bodies) and flexible chains for an optimized conformation. In the CORAL structure modeling, we used an initial BSA structure model deduced from the original crystallographic BSA structure (PDB entry 4F5S)[21] and assigned the subdomains **A** and subdomains **B** (both are rich in antiparallel α-helices and with disulfide bonds between helices) in each of the three major domains **I**, **II**, and **III** of BSA as rigid domains in CORAL modeling. The subdomains are interconnected by long loops (12−13 dummy residues) to adjust the relative spacing and orientations.[21] After rigorous fitting attempts, we found that the fitting could be much improved with the CORAL rigid domains regrouped into Domain **IA** (residues 6−104), Domain **IB**−**IIA** (residues 118−291), Domain **IIB**−**IIIA** (residues 304−489), and Domain **IIIB** (residues 502−583), together with one free coil (5 residues in the N-terminal) and three flexible interdomain linkers for the four CORAL rigid domains. We further took into account the structural constraints that all of the local structures with the 17 disulfide bonds should be maintained in the acidic unfolding to restrict the freedom of nearby polypeptide chain movements.

Figure 3 shows the fitted SAXS data for **N**, $\mathbf{I_1}$, $\mathbf{I_2}$, and **E** states of BSA using the GASBOR and CORAL structural models

DOI: 10.1021/acs.jpclett.6b02722
J. Phys. Chem. Lett. 2017, 8, 470−477

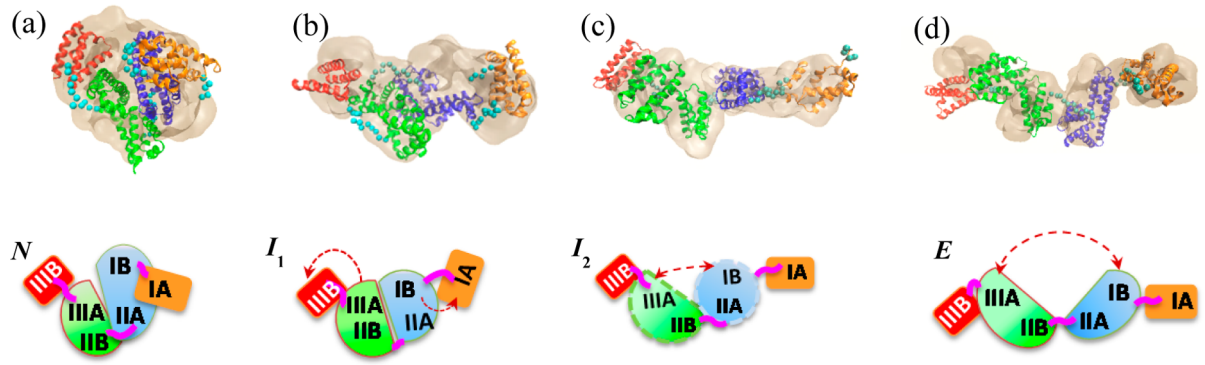


Figure 4. (a–d) CORAL models for the acid-induced unfolding of BSA at the states indicated. The small spheres represent the dummy residues of the flexible interdomain linkers between the four CORAL rigid domains (color-coded). Corresponding GABOR envelopes are shown in surface presentation for comparison. Corresponding cartoons illustrate the changes (dotted arrows) of the four assigned rigid domains interconnected by flexible chains (thick segments). The conventional assignments of the BSA major domains are marked with **I**, **II**, and **III**, each with **A** and **B** subdomains.

Table 2. SAXS Parameters for the BSA Solutions at the Selected pH Values[a]

state (pH value)	**N** (7.2)	$\mathbf{I}_1$ (4.4)		$\mathbf{I}_2$ (3.3)		**E** (2.7)	
$I(0)/c$ (cm^{-1} mg^{-1} mL)	2.42 ± 0.13	2.67 ± 0.13		2.72 ± 0.13		1.50 ± 0.13	
R_g (Å) of GA	27.5 ± 0.4	35.5 ± 1.1		37.1 ± 0.9		44.3 ± 1.0	
R_g (Å) of $p(r)$	27.6 ± 0.1	35.5 ± 0.1		38.1 ± 0.2		44.7 ± 0.3	
D_{max} (Å)	86.1	124.2		129.9		154.9	
fitting algorithm	CRYSOL	GASBOR	CORAL	GASBOR	CORAL	CORAL	EOM
fitting χ^2	1.21	1.28	1.01	1.00	1.00	1.60	1.53

[a]The zero-angle SAXS intensity $I(0)$ is scaled to the absolute intensity for 1.0 mg mL^{-1}, in units of cm^{-1} (cross section per unit volume).[50] Values for the maximum pair distance D_{max} are determined from the corresponding pair distance distribution functions $p(r)$ Fourier transformed from the SAXS data.[11,42,43] R_g values are extracted from the Guinier approximation (GA) and $p(r)$ functions.

shown in Figure 4. The best-fitted CORAL models overlap decently with the corresponding GASBOR envelopes (Figure 4 for the side views and Figure S7 for the top views). All extracted SAXS parameters are summarized in Table 2. These CORAL structural models reveal delicate transitions of the molten globule domains of BSA in the acid-induced unfolding. In the first unfolding step to the intermediate $\mathbf{I}_1$ from the native state, swing opening of the molten globule domains **IIIB** and rotation of **IA** (Figure 4b) complete the unfolding of the $\mathbf{G}^1$ group defined by the modified Ising model. Entering into $\mathbf{I}_2$ from the $\mathbf{I}_1$ state, however, involves a collective reorientation of the interconnected (**IIIA**–**IIB**) and (**IIA**–**IB**) domains (Figure 4c), forming a cleft in the pivot domain interfaces; this corresponds to the relative small fraction $\gamma \approx 6\%$ of the $\mathbf{G}^2$ group of the modified Ising model. Further transition to the **E** form at pH 2.7 or lower requires a follow-up pivot opening of the two interconnected domains (**IIIA**–**IIB**) and (**IIA**–**IB**), with the intersection of the **IIB** and **IIA** subdomains as the pivot (Figure 4d). The structural models in Figure 4 not only reveal the long-postulated packing structure evolution of the molten globule domains of the BSA intermediates but also characterize the structural transition sequence along the unfolding path. We also carefully checked local structural changes reported in many previous studies[9,22−28] and found that the available pieces of evidence are largely consistent with the three-step unfolding sequence proposed here (as detailed in the SI).

Previous NMR results suggested possible fluctuations of the two interconnected (**IIIA**–**IIB**) and (**IIA**–**IB**) domains in a similar acid-induced unfolding process.[45] In addition, hydrophobic solvation of the molten globule domains **I** and **II** for a softened structure of the **E** form was also reported lately.[3] These pieces of evidence suggest certain structure flexibility of the expanded **E** form. To further account for such structure flexibility in solution, we used the ensemble optimization method (EOM)[46] in search of an optimal structure ensemble that could best fit the same set of SAXS data (pH = 2.7). The EOM-generated structure pool for BSA contains 10 000 conformations, from the most compact ($R_{g\text{-}min}$ = 25 Å) to the most extended ($R_{g\text{-}max}$ = 55 Å) form. Moreover, the same constraints of the assignments of rigid domains and flexible chains used in the CORAL approach were also applied in the EOM model fitting with the same set of SAXS data. The best SAXS data fitting (χ^2 = 1.53; see Figure S8) was achieved with an ensemble comprising six equal-probable structural models representatively shown in Figure 5 (completely shown in Figure S9). We note that the resolution of the size exclusion column (ca. 6 Å in R_g) used in our SAXS data collection helped to reject certain ensembles containing conformations of unexpectedly large and small R_g values that would be excluded during the HPLC-SAXS measurement. The six conformations of the best ensemble of R_g values between 44 and 49 Å (Figure 5d) are overlapped respectively with the CORAL model structure (representing an average structure) in Figure 5 to reveal the conformational polydispersity of the **E** form of BSA. Similar SAXS analysis for protein or vesicle conformation polydispersity or flexibility was demonstrated previously.[46,47] The model overlapping in Figure 5 reveals that **IIIB** and **IA** subdomains are relatively more mobile than the other domains (Figure 5b-c), presumably resulted from partial domain melting (particularly in the interfaces of the subdomains as revealed from the local probes detailed in SI). In contrast, the two pivot

DOI: 10.1021/acs.jpclett.6b02722
J. Phys. Chem. Lett. 2017, 8, 470−477

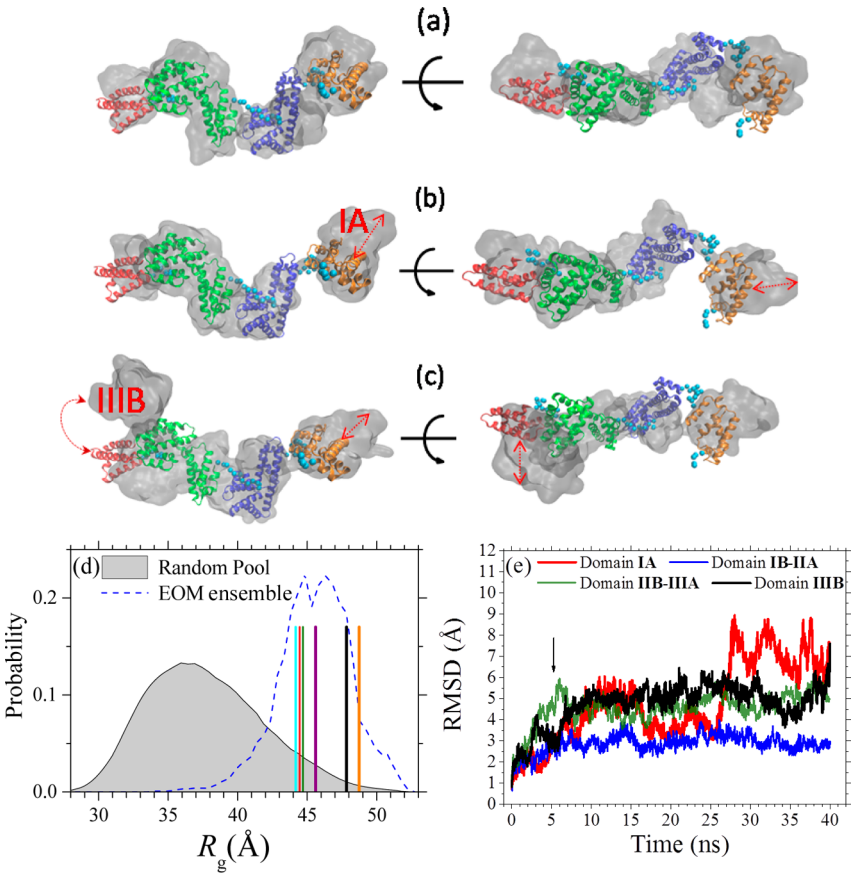


Figure 5. (a–c) Orthogonal views (surface presentations in gray) of the three representative (out of the six in Figure S9) models of the best-fit EOM ensemble for the **E** form of BSA. The models are overlapped with the CORAL model (of colored domains) to illustrate the regional conformation polydispersity, as indicated by the arrows. (d) R_g distribution of the EOM-SAXS selected ensemble (dotted blue profile) centered at ~46 Å in the higher R_g regime of the random pool (gray filled area). The six models that constitute the best-fit ensemble are marked by the vertical bars of equal probability. (e) Evolution of the RMSD for the four EOM-SAXS assigned rigid domains (as marked) during the constant-pH molecular dynamics simulation at pH 2.7. All RMSD values are calculated with respect to the initial coordinates. Larger fluctuations corresponding to the structural features of the **E** form are observed after ~5 ns of the simulation.

units of (**IIIA-IIB**) and (**IIA-IB**) are more steady in the six models as revealed in Figure 5 and Figure S9, suggesting a better core structure stability (or rigidity) of the **E** form. Because the linker (12 amino acids) between **IIA** and **IIB** is flexible, the regional rigidity presumably originates from the attractive domain–domain interactions.

To examine the SAXS-revealed local packing flexibility and rigidity of the molten globule domains in the **E** form, we further conduct constant-pH molecular dynamics simulations with the coupling between protonation and the conformation explicitly taken into account.[48,49] The hence captured dynamic behavior of the **E** form during the 40 ns simulation of the pH-unfolding process, starting from the native BSA conformation at pH 2.7 (300 K), confirms the features revealed by the SAXS EOM models (detailed in Figure S10). Specifically, swing open of subdomain **IIIB** is observed in the earlier stage of the molecular dynamics simulation; after t = 5 ns, pivot opening of the interconnected subdomains of **IB–IIA** and **IIB–IIIA** for the **E** form is observed. Subsequently, the pivot domains remain stable with relatively small root-mean-squared distance (RMSD), as shown in Figure 5e. In contrast, volatile trajectories for subdomains **IA** and **IIIB** continue along the rest of the 40 ns simulation path, resulting in constant mobile subdomains with oscillatory RMSD values (Figure 5e), as suggested by the EOM-SAXS analysis.

In summary, using scanning SAXS along the elution path of HPLC along with molecular dynamics simulation, we have elucidated the long-postulated packing structures of the three molten globule domains of BSA in acid-induced unfolding. The multistep changes of the packing structure observed are mapped onto a three-group unfolding process, on the basis of an unconventional view of protein folding–unfolding formulated with a modified Ising model. Collective changes in the packing structure of the three molten globule domains result in a cleft in the pivot domain interfaces followed by a pivot opening of the central interconnected rigid domains of BSA, characterizing the $\mathbf{I}_1$–$\mathbf{I}_2$ and $\mathbf{I}_2$–**E** transitions. These feature transitions together with the molecular dynamics simulation details provide hints in incorporating or releasing small molecules for biomedical or industrial applications. Most importantly, the observed evolution of the protein conformation strongly supports the previous simulation studies;[7,8] all suggest that collective motions of the molten globule domains, in addition to main-chain or side-chain interactions with the denaturant, profoundly influence the folding–unfolding pathway of a multidomain protein.

DOI: 10.1021/acs.jpclett.6b02722
J. Phys. Chem. Lett. 2017, 8, 470–477

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpclett.6b02722.

Materials and experimental methods; SAXS data analysis; UV−Vis data; and detailed SAXS parameters (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: usjeng@nsrrc.org.tw. Tel: +886-3-578-0281, ext-7108. Fax: +886-3-578-3813.

ORCID

Yi-Qi Yeh: 0000-0002-2387-3600

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was partially funded by the Minister of Science and Technology, Taiwan (MOST 104-2627-M-213).

■ REFERENCES

(1) Baldwin, R. L.; Frieden, C.; Rose, G. D. Dry Molten Globule Intermediates and the Mechanism of Protein Unfolding. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 2725−2737.

(2) Jha, S. K.; Marqusee, S. Kinetic evidence for a two-stage mechanism of protein denaturation by guanidinium chloride. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 4856−4861.

(3) Acharya, N.; Mishra, P.; Jha, S. K. Evidence for Dry Molten Globule-Like Domains in the pH-Induced Equilibrium Folding Intermediate of a Multidomain Protein. *J. Phys. Chem. Lett.* **2016**, *7*, 173−179.

(4) Bhattacharyya, S.; Varadarajan, R. Packing in molten globules and native states. *Curr. Opin. Struct. Biol.* **2013**, *23*, 11−21.

(5) Baldwin, R. L. Dynamic Hydration Shell Restores Kauzmann's 1959 Explanation of How the Hydrophobic Factor Drives Protein Folding. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 13052−13056.

(6) Baldwin, R. L.; Rose, G. D. Molten globules, entropy-driven conformational change and protein folding. *Curr. Opin. Struct. Biol.* **2013**, *23*, 4−10.

(7) Inanami, T.; Terada, T. P.; Sasai, M. Folding pathway of a multidomain protein depends on its topology of domain connectivity. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 15969−15974.

(8) Zhou, R.; Huang, X.; Margulis, C. J.; Berne, B. J. Hydrophobic Collapse in Multidomain Protein Folding. *Science* **2004**, *305*, 1605−1609.

(9) El Kadi, N.; Taulier, N.; Le Huerou, J. Y.; Gindre, M.; Urbach, W.; Nwigwe, I.; Kahn, P. C.; Waks, M. Unfolding and Refolding of Bovine Serum Albumin at Acid pH: Ultrasound and Structural Studies. *Biophys. J.* **2006**, *91*, 3397−3404.

(10) Serrano, A. L.; Waegele, M. M.; Gai, F. Spectroscopic studies of protein folding: Linear and nonlinear methods. *Protein Sci.* **2012**, *21*, 157−170.

(11) Jeffries, C. M.; Graewert, M. A.; Blanchet, C. E.; Langley, D. B.; Whitten, A. E.; Svergun, D. I. Preparing monodisperse macromolecular samples for successful biological small-angle X-ray and neutron-scattering experiments. *Nat. Protoc.* **2016**, *11*, 2122−2153.

(12) Trewhella, J. Small-angle scattering and 3D structure interpretation. *Curr. Opin. Struct. Biol.* **2016**, *40*, 1−7.

(13) Zheng, W.; Borgia, A.; Buholzer, K.; Grishaev, A.; Schuler, B.; Best, R. B. Probing the Action of Chemical Denaturant on an Intrinsically Disordered Protein by Simulation and Experiment. *J. Am. Chem. Soc.* **2016**, *138*, 11702−11713.

(14) Shiu, Y. J.; Hayashi, H.; Shih, O.; Su, C.; Tsai, M. Y.; Yeh, Y. Q.; Su, C. J.; Huang, Y. S.; Lin, S. H.; Jeng, U. Intrinsic coordination for revealing local structural changes in protein folding−unfolding. *Phys. Chem. Chem. Phys.* **2016**, *18*, 3179−3187.

(15) Bernadó, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I. Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering. *J. Am. Chem. Soc.* **2007**, *129*, 5656−5664.

(16) Segel, D. J.; Fink, A. L.; Hodgson, K. O.; Doniach, S. Protein Denaturation: A Small-Angle X-ray Scattering Study of the Ensemble of Unfolded States of Cytochrome c. *Biochemistry* **1998**, *37*, 12443−12451.

(17) Berthaud, A.; Manzi, J.; Pérez, J.; Mangenot, S. Modeling Detergent Organization around Aquaporin-0 Using Small-Angle X-ray Scattering. *J. Am. Chem. Soc.* **2012**, *134*, 10080−10088.

(18) Sung, T. C.; Li, C. Y.; Lai, Y. C.; Hung, C. L.; Shih, O.; Yeh, Y. Q.; Jeng, U. S.; Chiang, Y. W. Solution Structure of Apoptotic BAX Oligomer: Oligomerization Likely Precedes Membrane Insertion. *Structure* **2015**, *23*, 1878−1888.

(19) Quinlan, G. J.; Martin, G. S.; Evans, T. W. Albumin: Biochemical Properties and Therapeutic Potential. *Hepatology* **2005**, *41*, 1211−1219.

(20) Ghuman, J.; Zunszain, P. A.; Petitpas, I.; Bhattacharya, A. A.; Otagiri, M.; Curry, S. Structural Basis of the Drug-binding Specificity of Human Serum Albumin. *J. Mol. Biol.* **2005**, *353*, 38−52.

(21) Bujacz, A. Structures of bovine, equine and leporine serum albumin. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2012**, *D68*, 1278−1289.

(22) Ferrer, M. L.; Duchowicz, R.; Carrasco, B.; de la Torre, J. G.; Acuna, A. U. The Conformation of Serum Albumin in Solution: A Combined Phosphorescence Depolarization-Hydrodynamic Modeling Study. *Biophys. J.* **2001**, *80*, 2422−2430.

(23) Carter, D. C.; Ho, J. X. Structure of serum albumin. *Adv. Protein Chem.* **1994**, *45*, 153−204.

(24) Muzammil, S.; Kumar, Y.; Tayyab, S. Molten globule-like state of human serum albumin at low pH. *Eur. J. Biochem.* **1999**, *266*, 26−32.

(25) Dockal, M.; Chang, M.; Carter, D. C.; Rüker, F. Conformational Transitions of the Three Recombinant Domains of Human Serum Albumin Depending on pH. *J. Biol. Chem.* **2000**, *275*, 3042−3050.

(26) Khan, M. Y. Direct evidence for the involvement of domain III in the N-F transition of bovine serum albumin. *Biochem. J.* **1986**, *236*, 307−310.

(27) Curvale, R.; Masuelli, M.; Padilla, A. P. Intrinsic viscosity of bovine serum albumin conformers. *Int. J. Biol. Macromol.* **2008**, *42*, 133−137.

(28) Barbosa, L. R. S.; Ortore, M. G.; Spinozzi, F.; Mariani, P.; Bernstorff, S.; Itri, R. The Importance of Protein-Protein Interactions on the pH-Induced Conformational Changes of Bovine Serum Albumin: A Small-Angle X-Ray Scattering Study. *Biophys. J.* **2010**, *98*, 147−157.

(29) Bendedouch, D.; Chen, S. H. Structure and Interparticle Interactions of Bovine Serum Albumin in Solution Studied by Small-Angle Neutron Scattering. *J. Phys. Chem.* **1983**, *87*, 1473−1477.

(30) Skou, S.; Gillilan, R. E.; Ando, N. Synchrotron-based small-angle X-ray scattering (SAXS) of proteins in solution. *Nat. Protoc.* **2014**, *9*, 1727−1739.

(31) Böhme, U.; Scheler, U. Effective charge of bovine serum albumin determined by electrophoresis NMR. *Chem. Phys. Lett.* **2007**, *435*, 342−345.

(32) Receveur-Bréchot, V.; Durand, D. How Random are Intrinsically Disordered Proteins? A Small Angle Scattering Perspective. *Curr. Protein Pept. Sci.* **2012**, *13*, 55−75.

(33) Perez, J.; Vachette, P.; Russo, D.; Desmadril, M.; Durand, D. Heat-induced Unfolding of Neocarzinostatin, a Small All-b Protein Investigated by Small-angle X-ray Scattering. *J. Mol. Biol.* **2001**, *308*, 721−743.

(34) Durand, D.; Vivès, C.; Cannella, D.; Pérez, J.; Pebay-Peyroula, E.; Vachette, P.; Fieschi, F. NADPH oxidase activator $p67^{phox}$ behaves in solution as a multidomain protein with semi-flexible linkers. *J. Struct. Biol.* **2010**, *169*, 45−53.

DOI: 10.1021/acs.jpclett.6b02722
J. Phys. Chem. Lett. 2017, 8, 470−477

(35) Sinibaldi, R.; Ortore, M. G.; Spinozzi, F.; de Souza Funari, S.; Teixeira, J.; Mariani, P. SANS/SAXS study of the BSA solvation properties in aqueous urea solutions via a global fit approach. *Eur. Biophys. J.* **2008**, *37*, 673−681.
(36) Sogami, M.; Era, S.; Nagaoka, S.; Inouye, H. Circular dichroic and fluoropolarimetric studies on tryptophyl residues in acid-induced isomerization of bovine plasma albumin. *Int. J. Pept. Protein Res.* **1982**, *19*, 263−269.
(37) Era, S.; Nagaoka, S.; Sogami, M.; Watari, H.; Akasaka, K.; Kida, K.; Sogami, M.; Yoshida, A. Circular dichroic and fluorometric studies on the acid-induced isomerization of bovine plasma albumin−1 -anilino-8-naphthalenesulfonate complex. *Int. J. Pept. Protein Res.* **1985**, *26*, 575−583.
(38) Liang, K. K.; Hayashi, M.; Shiu, Y. J.; Mo, Y.; Shao, J.; Yan, Y.; Lin, S. H. Application of the generalized kinetic Ising model to the kinetics of protein folding. *J. Chin. Chem. Soc.* **2003**, *50*, 335−338.
(39) Liang, K. K.; Hayashi, M.; Shiu, Y. J.; Mo, Y.; Shao, J.; Yan, Y.; Lin, S. H. Thermodynamics and kinetics of protein folding: A mean field theory. *Phys. Chem. Chem. Phys.* **2003**, *5*, 5300−5308.
(40) Shiu, Y. J.; Jeng, U.; Su, C.; Huang, Y. S.; Hayashi, M.; Liang, K. K.; Yeh, Y. L.; Lin, S. H. A modified Ising model for the thermodynamic properties of local and global protein folding−unfolding observed by circular dichroism and small-angle X-ray scattering. *J. Appl. Crystallogr.* **2007**, *40*, s195−s199.
(41) Shiu, Y. J.; Jeng, U.; Huang, Y. S.; Lai, Y. H.; Lu, H. F.; Liang, C. T.; Hsu, I.-J.; Su, C. H.; Su, C.; Chao, I.; Su, A. C.; Lin, S. H. Global and Local Structural Changes of Cytochrome c and Lysozyme Characterized by a Multigroup Unfolding Process. *Biophys. J.* **2008**, *94*, 4828−4836.
(42) Svergun, D. I.; Petoukhov, M. V.; Koch, M. H. Determination of Domain Structure of Proteins from X-Ray Solution Scattering. *Biophys. J.* **2001**, *80*, 2946−2953.
(43) Petoukhov, M. V.; Franke, D.; Shkumatov, A. V.; Tria, G.; Kikhney, A. G.; Gajda, M.; Gorba, C.; Mertens, H. D. T.; Konarev, P. V.; Svergun, D. I. New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **2012**, *45*, 342−350.
(44) Togashi, D. M.; Ryder, A. G.; O'Shaughnessy, D. Monitoring local unfolding of Bovine Serum Albumin during denaturation using steady-state and time-resolved fluorescence spectroscopy. *J. Fluoresc.* **2010**, *20*, 441−452.
(45) Sadler, P. J.; Tucker, A. pH-induced structural transitions of bovine serum albumin Histidine pK_a, values and unfolding of the N-terminus during the N to F transition. *Eur. J. Biochem.* **1993**, *212*, 811−817.
(46) Tria, G.; Mertens, H. D. T.; Kachala, M.; Svergun, D. I. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* **2015**, *2*, 207−217.
(47) Castorph, S.; Riedel, D.; Arleth, L.; Sztucki, M.; Jahn, R.; Holt, M.; Salditt, T. Structure Parameters of Synaptic Vesicles Quantified by Small-Angle X-Ray Scattering. *Biophys. J.* **2010**, *98*, 1200−1208.
(48) Case, D. A.; Berryman, J. T.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E. I.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; Le Grand, S.; Li, P.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Monard, G.; Needham, P.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Salomon-Ferrer, R.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; York, D. M.; Kollman, P. A. *AMBER 2015*; University of California: San Francisco, 2015.
(49) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Struct., Funct., Genet.* **2004**, *55*, 383−394.
(50) Jeng, U.; Su, C. H.; Su, C. J.; Liao, K. F.; Chuang, W. T.; Lai, Y. H.; Chang, J. W.; Chen, Y. J.; Huang, Y. S.; Lee, M. T.; Yu, K. L.; Liu, D. G.; Chang, C. F.; Liu, C. Y.; Chang, C. H.; Liang, K. S.; Lin, J.-M. A small/wide-angle X-ray scattering instrument for structural characterization of air−liquid interfaces, thin films and bulk specimens. *J. Appl. Crystallogr.* **2010**, *43*, 110−121.

DOI: 10.1021/acs.jpclett.6b02722
J. Phys. Chem. Lett. 2017, 8, 470−477

Supporting Information

Probing the Acid-induced Packing Structure Changes of the Molten Globule Domains of a Protein near Equilibrium Unfolding

Yi-Qi Yeh,[a] Kuei-Fen Liao,[a] Orion Shih,[a] Ying-Jen Shiu,[a] Wei-Ru Wu,[a] Chun-Jen Su,[a] Po-Chang Lin,[a] and U-Ser Jeng,[a,b,*]

[a.] *National Synchrotron Radiation Research Center, Hsinchu 30076, Taiwan.*

[b.] *Department of Chemical Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan.*

S1. Experimental Methods

Sample Preparations.

Bovine serum albumin (BSA) was purchased from Sigma-Aldrich (M_w = 66.430 kDa; purity $\geqq$98%) without further purification. BSA aqueous solutions of concentration 10 mg mL^{-1} were prepared with 150 mM NaCl and 20 mM HEPES at pH values between 7.2-2.1, adjusted by $NaOH_{(aq)}$ and $HCl_{(aq)}$. All the buffers and samples were filtered (0.22 μm pore size) before use. All the samples were prepared and used within 24 hr. The sample concentrations after the size exclusion column of the HPLC for SAXS data collection, however, reduced below 3 mg mL^{-1} due to elution dilution (see Figure S4 below).

Scanning SAXS/HPLC and Data Analysis.

Scanning SAXS/HPLC was conducted at the 23A SWAXS endstation equipped with an on-line size-exclusion high performance liquid chromatographic (SE-HPLC) system (Agilent chromatographic system 1260 series, as detailed in Figure S1), at the National Synchrotron Radiation Research Center (NSRRC). 50 μL BSA sample solutions (10 mg mL^{-1}) were injected into the HPLC column with a flow rate 0.35 ml/min, and directed through a quartz capillary (2.0 mm dia.) thermostated at 288 K for simultaneous SAXS and UV-Vis absorption measurements at the same sample position (with orthogonal incidences). The sample exposure area to X-rays and UV-Vis was monitored using a remote controlled camera. Dry air was flowed through the sample exposure area during the measurement to prevent water condensation on the sample capillary tube. The buffer solutions used were at the same pH values as the corresponding sample solutions. SAXS data were collected successively during the HPLC sample elution with 1 data frame per 30 s using a Pilatus 1M-F area detector. Buffer solutions were measured under

identical conditions for background scattering subtraction. With 15 keV X-rays (wavelength λ= 0.8266 Å) and a sample-to-detector distance of 3.17 m, the scattering vector q, defined by $4\pi\lambda^{-1}\sin\theta$ with scattering angle 2θ, covered from 0.007 to 0.25 $Å^{-1}$. Data were corrected for electronic noise, sample transmission, and detector sensitivity, followed by a scaling to the absolute intensity $I(q)$ in units of cm^{-1} (scattering cross section per unit volume) via scattering from water at the protein sample conditions.[s1] Data were evaluated for radiation damage, background subtraction quality, and sample concentration effects, and well-overlapped SAXS profiles collected over the sample elution peak of HPLC were integrated for improved data statistics. Zero-angle scattering intensities (I_o) and the corresponding R_g values were extracted from SAXS data using the Guinier approximation.[s2] The R_g values used in the Ising model analysis were extracted from the SAXS profiles simulated with the best-fit models. To avoid effects of inter-particle interactions, all the R_g values discussed in the text are extracted from the concentration independent SAXS results for the native, $\boldsymbol{I_1}$, $\boldsymbol{I_2}$, and $\boldsymbol{E}$ states as shown in Figure S4, using samples solutions of BSA concentrations below 3.0 mg mL^{-1}.

SAXS data evaluation and model fittings were performed using the ATSAS package.[s2] The distance distribution function $p(r)$ was Fourier transformed from the SAXS data; the corresponding maximum dimension D_{max} and Porod volumes were evaluated using GNOM.[s2] Protein envelopes for different emphases were obtained using the programs of GASBOR, DAMMIF, and CORAL.[s2-s4] For each set of SAXS data, 7−10 best-fit models of similar χ^2 from DAMMIF or GASBOR were generated, with the average normalized spatial discrepancy (NSD) of the models smaller than unity; the model with the lowest NSD was used as the representative model.[s2] CRYSOL was used in the SAXS data fitting with the available crystal structure.[s5] Conformational flexibility of the $\boldsymbol{E}$ form of partly unfolded BSA was analyzed with the program EOM (Ensemble Optimization Method).[s6] EOM 2.0 uses a genetic algorithm to select a minimum set (six in our case) of models from the optimization ensemble for a linear combination of the corresponding theoretical scattering curves that could best fit the SAXS data. In the fitting, the individual weight of each conformer is optimized under a defined ensemble size. In the EOM-2.0 algorithm, a large pool of 10,000 independent structures with different conformations of protein extremities was generated by RANCH (RANdom CHain) first. Then GAJOE (Genetic Algorithm Judging Optimization of Ensembles) used a genetic algorithm to select a minimum set (six in our case) of models from the pool for a linear combination of the corresponding theoretical scattering curves that could best fit the SAXS data.

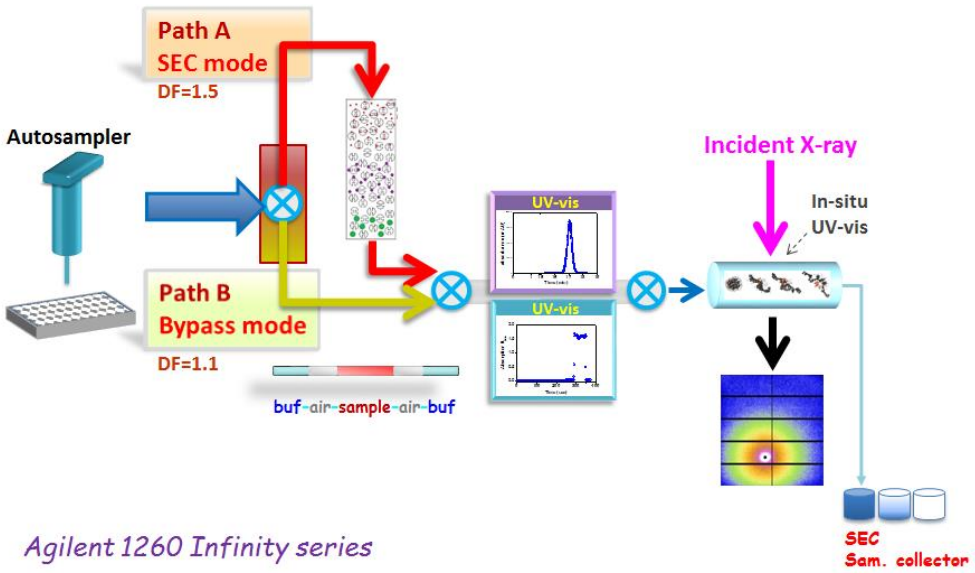


Figure S1. An on-line HPLC system (Agilent 1260 Infinity series) for simultaneous scanning SAXS and UV-Vis-absorption measurements at the same sample position. The system allows a column-bypass path option (Path B as indicated) for SAXS/UV-Vis measurements. At the end of the sample elution, fractioned sample can be collected for further analyses. The dilution factors (DF) with the SEC and bypass paths are 1.5 and 1.1, as indicated, with 100 μl sample injection.

S2. Determination of equivalent molecular mass (M_w) of the elution peaks.

A mass-calibration curve was constructed using standards of known M_w, including133 kDa BSA dimer, 66.5 kDa BSA monomer, 13.7 kDa Ribonuclease A, and 112 Da Uracil. A polynomial equation fitted from the elution peaks of the standards was used to extrapolate the equivalent molecular mass from the retention times of elution peaks observed from the solution of BSA.

S3. Native BSA dimer and monomer conformations in solution

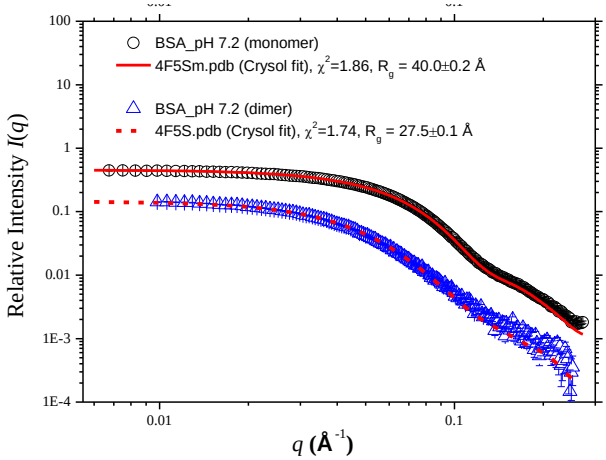


Figure S2. The SAXS profiles and CRYSOL fittings of BSA monomer (solid curves) and dimer (dashed curve) in a solution of pH 7.2. The fitting χ^2, R_g values, and the PDB crystal structures used are indicated.[S7]

S4. Complete pH-dependent SAXS data and fitting with the Guinier approximation

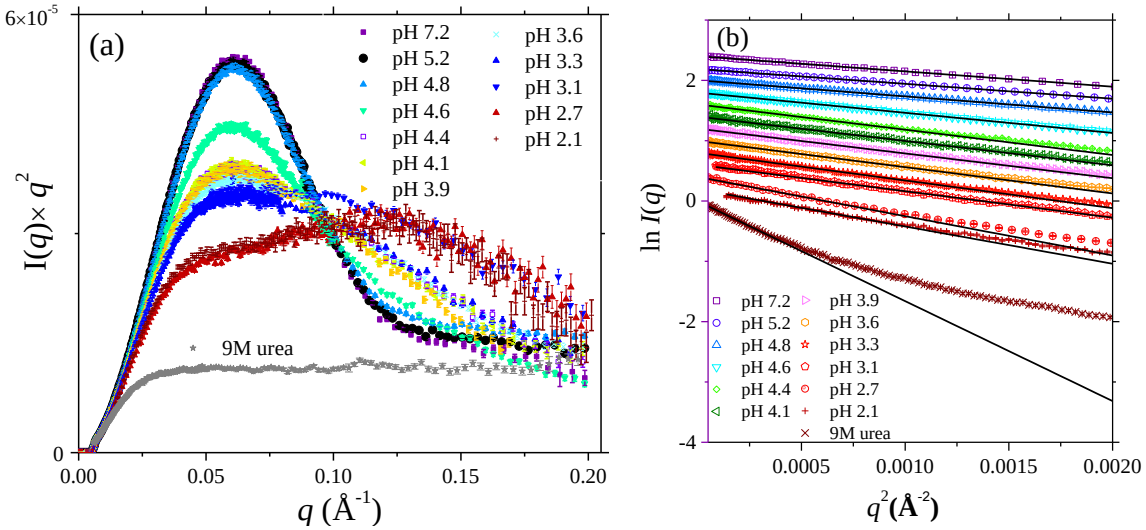


Figure S3. (a) pH-dependent SAXS data of BSA in the Kratky-Porod presentation. Also shown are the SAXS data for the unfolded BSA with 9.0 M urea. (b) Guinier plot for the data and fittings (solid lines) with the Guinier approximation. The corresponding R_g values fitted are selectively shown in Table S1.

S5. Concentration independent R_g values

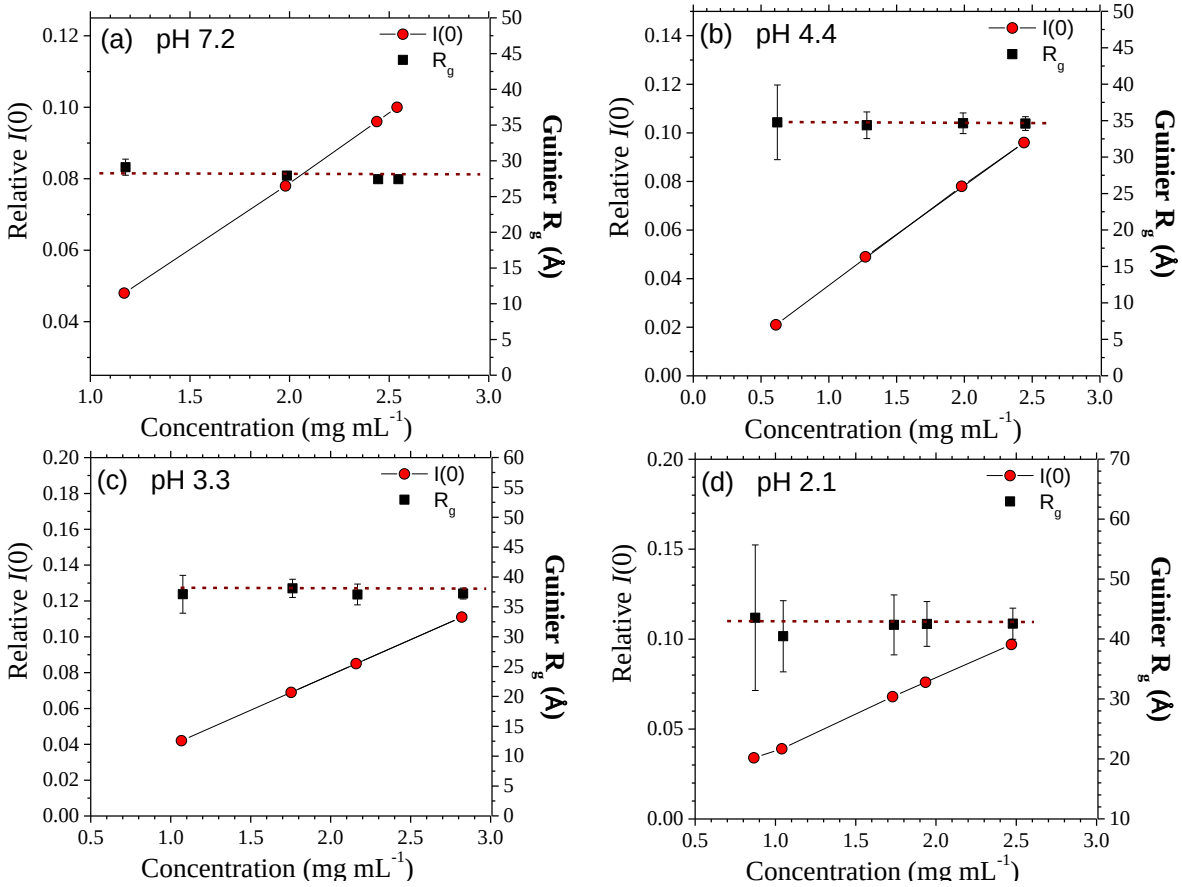


Figure S4. Concentration independent R_g values extracted from the SAXS profiles measured for BSA at (a) pH 7.2 (b) pH 4.4 (c) pH 3.3, and (d) pH 2.1, for the native, $\boldsymbol{I}_1$, $\boldsymbol{I}_2$, and $\boldsymbol{E}$ states, respectively. The results show little or no inter-particle interactions on the Rg values, with sample concentrations (after the HPLC elution dilution) below ca. 3 mg mL^{-1}.

S6. UV-Vis absorption

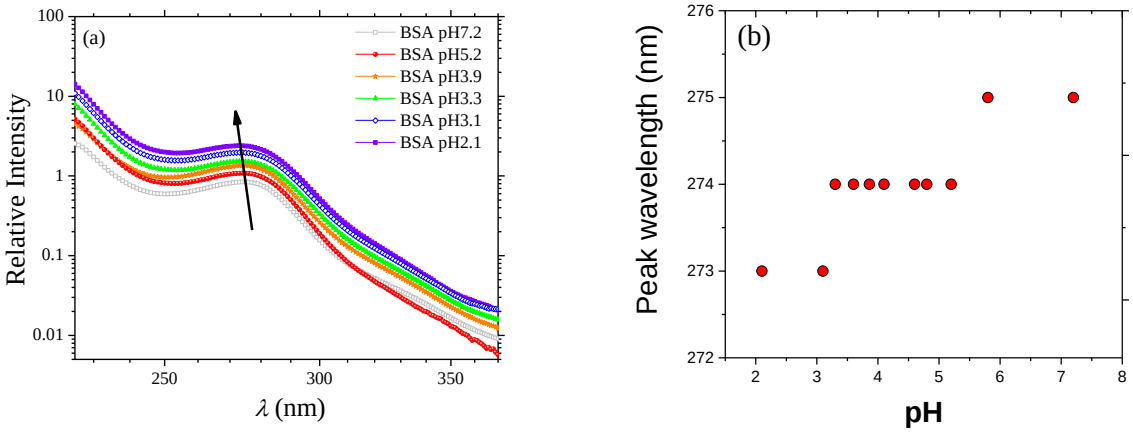


Figure S5. (a) UV-Vis absorption spectra and (b) the corresponding peak position near λ = 274 nm, simultaneously measured with the SAXS data for the acid-induced unfolding of BSA in solutions of the pH values indicated.

S7. Refolding of BSA

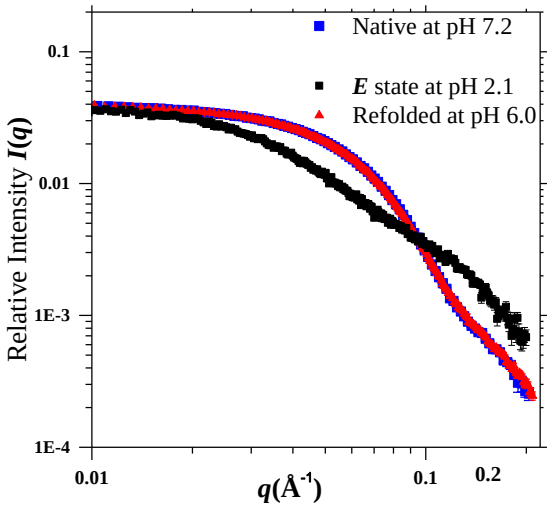


Figure S6. SAXS profiles measured for BSA at the *E*-state at pH 2.1 and the refolded BSA at pH 6.0. The refolded BSA sample solution was prepared by dialysis of the BSA sample solution at pH 2.1 in the environment of pH 7.2 for a final sample solution pH 6.0. The SAXS profile measured for the refolded BSA can overlap that for native BSA prepared at pH 7.2, demonstrating that the transitions between the acid-induced intermediates and the native state are reversible.

S8. Examining the proposed BSA unfolding process with the previously observed local structural information

- **Mobile IA domain in the higher-pH unfolding process followed by a flexible protein conformation in the lower-pH unfolding process, as revealed by electron paramagnetic resonance (EPR), near-CD**

Previous EPR result revealed opening of the Cys34 located crevice via a two-step transition over pH 6.4–2.2.[s8] The first transition occurred in the pH range of 4.5-3.7, suggesting the loosened sulfhydryl group. Below pH 3.7, the sulfhydryl group became largely exposed to solvent and Cys34 located crevice became largely opened for a flexible tertiary structure. Also, the disulfide bond was found to remain intact during these transitions. The near-CD spectra suggested perturbed Phe27 and Phe36 located in the same crevice at pH 4.0.[s9]

These results are consistent with the structural transition features described by the CORAL models in Figure 4, in the ***I*$_1$–*I*$_2$** with mobile **IA** domain, followed by ***I*$_2$–*E*** transitions with a flexible ***E***-form conformation revealed from the EOM-SAXS fitting.

- **NMR and viscosity results support the unfolding sequence proposed: partial opening in the native *N* to *I*$_1$ then *I*$_1$-*I*$_2$ transitions, followed by a wide structural opening in the transition to Expanded form (*E*).**

Previous NMR result of acid-induced HSA unfolding showed that His146, His336, and His533 (located at IB, IIB, and IIB of HSA) became solvent-accessible and unstructured over the pH range 4.5-4.0,[s10] corresponding to our proposed ***I*$_1$-*I*$_2$** transition. The solvent-exposed His residues were attributed to a partial opening of HSA. By homology, His145, His337, and His534 (located at **IB, IIB**, and **IIIA**) of BSA should have similar spectral changes. In corresponding to SA crystal structure, the change in His145 (**IB**/h8) would involve the detachment of subdomain **IA** from the other domains. The changes in His 337 (**IIB**/h8) would involve **IB-IIA** and **IIB-IIIA** interfaces, revealing a similar detachment of the subdomain **IIA-IIIA** interface. These were followed by a more expanded conformation in the proposed ***I*$_2$–*E*** transition in pH 3.5–2.5, as proposed in a previous intrinsic viscosity study.[S11]

- **Trp214 fluorescence of intact HSA (domain II) for single step transition between pH 4.0 to 2.5**

Trp214 of native HSA is located between the **IIA** and **IIIA** interface. A solvent-accessible environment of Trp214 was observed by fluorescence in pH 4.0– 2.5,[s12] suggesting a late change or structural rearrangement of the **IIA** and **IIIA** interface for the proposed cleft in the domain interface followed by a pivot opening in the respective ***I*$_1$-*I*$_2$** and ***I*$_2$–*E*** transitions.

- **Changes of the packing structure with rigid molten globule domains**

The recombinant fragments (Domain **I** and Domain **III**) retained molten globule domains in hydrodynamic characterization in the pH range above 4.0.[s13] The two CORAL rigid domains

IB-IIA and IIB-IIIA were found to deform but remain intact at pH lower than 3.0[s13-14], which is consistent with the packing structure changes with locally rigid molten globule domains proposed in this study. The result supports our rigid-domain assumption used in the CORAL (see Figure S5a, S5c, and S5d) and EOM structural modeling (see Figure S6e).

- **Domain-Domain interface detachment from the polar/non-polar features of Tyr residues for domain mobility**

BSA has one Trp-213 residue in subdomain **IIA** between **IIA** and **IIIA** interface and 20 tyrosine residues. By homology, all horse, human, and bovine serum albumins possibly would reveal similar polar/non-polar environment of the residues. According to Carter's prediction[s15], Tyr-147, 262, 331, 400, and 496 appeared to be exposed; whereas Tyr-84, 137, 139, 155, 156, 160, 318, 333, and 340 were partially buried, and Tyr-30, 352, 370, 410, and 451 were fully buried in a similar acid-induced unfolding.[s16] Based on the BSA structure (pdb: 4F5S) [s7], the partially buried Tyr-155,156, and 160 could be stable due to disulfide bond and anti-parallel helices confinement; the buried Tyr-30 and partially buried Tyr-84, 137, 139, 318, 333, and 340, however, were found to be exposed to solvent during a similar acid-induced unfolding.[s9]

These are consistent with the proposed cleft in the domain interface followed by the pivot opening of the **IIB-IIIA** and **IB-IIA** pivot domains for the detachment of the two interconnected domain-domain interface.

- **Tyr fluorescence and Helix loss for mobile domain-IA**

The tyrosyl fluorescence of the three recombinant HSA domains showed that the four out of seven buried Tyr residues in domain-**I** became more solvent accessible environment from neutral pH to pH 2.0.[s9] This leads to more protonation environment around the Tyr-30, 137, and 139. The far-UV CD spectra of the fragment domain-**I** showed 7% α-helix loss from pH 7.4 to 4.0 and 15% α-helix loss from pH 4.0 to 2.5. The α-helix loss in domain-**I** might result in a solvent-accessible Tyr-84.

These results are consistent with our SAXS-based CORAL and EOM analyses (Figure 5) for a mobile subdomain **IA** in the high-pH unfolding process with a relative rigid domain **IB**, due to weakened interface interactions of the subdomains.

- **Tyr fluorescence and Far-CD spectrum revealing weakened interface interactions between subdomains IIA and IIIB, leading to a more open structure of the *E* form with mobile Domain-IIIB**

There are four Tyrosine residues in BSA subdomain-**IIIA**, including buried Tyr-410 and 451 and exposed Tyr-400 and 496. Tyr-410 is located in the end of IIIA/h2 between the interface of subdomains **IIIA** and **IIB**. During acid-induced unfolding, Tyr-410 was expected to expose to solvent due to hydrogen-bond disruption, leading to a more opened Domain **III**. Tyr-451 is located in IIIA/h4 and confined by a disulfide bond (Cys436 and Cys447) and the anti-parallel helices (IIIA/h3-h4). Tyr-400 is located in IIIA/h2 near the turn between two successive helices,

solvent accessible surface. Tyr-496 is located in a solvent-accessible flexible coil between subdomains **IIIA** and **IIIB**. In Carson' prediction using RIBBON program,[s17] the helix 2, 3, and 4 could be preserved in the ***E*** form structure (pH 3.5-2.5). The far-UV CD spectra of the fragment Domain-**III** showed limited 15% α-helix loss from pH 7.4 to 4.0 and 8% α-helix loss from pH 4.0 to 2.5.[s9] In 1986, M. Y. Khan also reported that Tyr residues transferred from a buried to exposed environment in pH 4.5-3.5,[s14] suggesting detachment of Domain **III** from the other domains. These results are consistent with a more open structure of the ***E*** form, with mobile domain **IIIB**, as revealed in our CORAL structural model and the EOM models.

- **Sign of a cleft of the IIIA-IIB and IIA-IB interconnected domains and the rigid core structure of the *E* form**

Dockal's work[s9] showed that with Far-UV CD no helix loss occurred in Domain **II** in the pH range 7.4-2.5, and the **IIA-IIIA** subdomain interface was more open in the pH range 4.0-2.0. The results are consistent with our EOM analysis (Figure 5 in the text) for a rigid core structure (with the domain **II**) of the ***E*** form, and the cleft followed by the pivot opening of **IIIA-IIB** and **IIA-IB** interconnected domains in the $\boldsymbol{I_1}$**–**$\boldsymbol{I_2}$ followed by $\boldsymbol{I_2}$**–*E*** transitions (Figure 4 in the text).

S9. SAXS related parameters.

Table S1. SAXS related parameters for BSA acid-induced unfolding for the native *N*, intermediates $\boldsymbol{I_1}$, $\boldsymbol{I_2}$, and ***E*** form. $I(0)$ values are in the absolute intensity (cross-section per unit volume) for BSA solutions normalized to a sample concentration of 1.0 mg mL^{-1}. The SAXS profiles used for model fitting and the R_g values discussed in the text were obtained with samples solutions of BSA concentrations below 3.0 mg mL^{-1} (after HPLC elution dilution), to avoid protein-protein interaction effects (cf. Figure S4).

State (pH value)		***N*** (7.2)	$\boldsymbol{I_1}$ (4.4)	$\boldsymbol{I_2}$ (3.3)	***E*** (2.7)
$I(0)/c$ ($cm^{-1}mg^{-1}ml$) [from Guinier App.]		2.42	2.67	2.72	1.50*
R_g (Å) [from Guinier]		27.5±0.4	35.5 ± 1.1	37.1 ± 0.9	44.3 ± 1.0
$I(0)/c$ ($cm^{-1}mg^{-1}ml$) [from $P(r)$]		2.55	2.55	2.80	1.53
R_g (Å) [from $P(r)$]		27.6	35.5	38.1	44.7
D_{max} (Å)		86.1	124.2	129.9	154.9
V_{Porod} (Å^3)		87695	73966	75705	83231
Fitting quality /algorithm					
χ^2	CRYSOL	1.21			
	GASBOR		1.28	1.00	1.48
	CORAL		1.01	1.00	1.6
	EOM				1.53
Instrument[s1]		NSRRC 23A SWAXS beamline, Pilatus 1M-F area detector			
Primary data reduction kit		NSRRC 23A SWAXS package			
Data processing kit		ATSAS 2.5.1 / Scanning SAXS-UV_2_0.exe by NSRRC BL23A			

* $I(0)$ value is lower due to a reduced scattering contrast of BSA in the low-pH solution of high concentrations of $HCl_{(aq)}$.

S10. SAXS-based structural models built with GASBOR and COREL algorithms

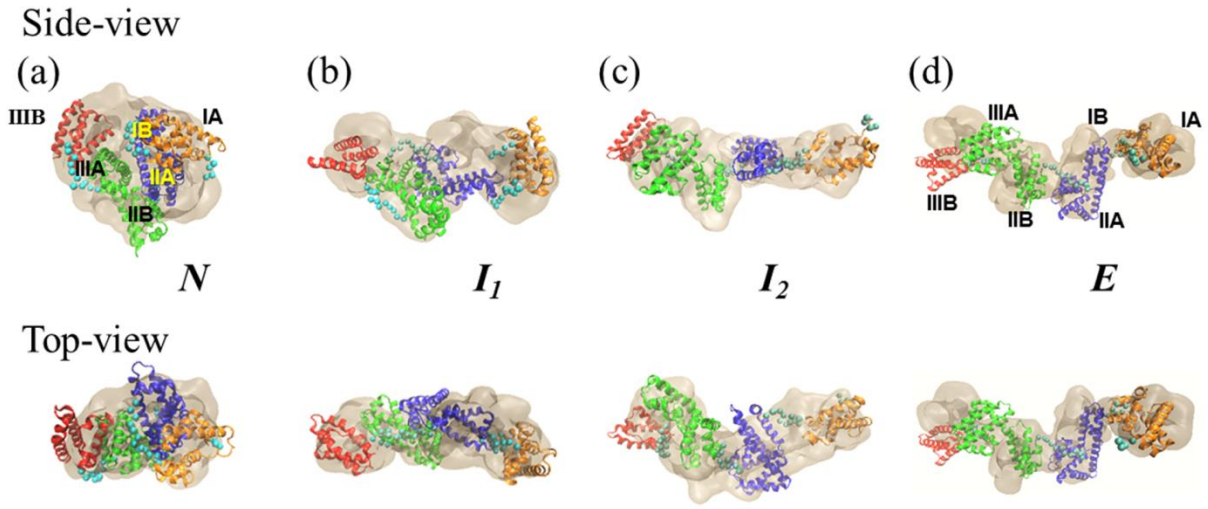


Figure S7. (a)-(d) are the orthogonal views (side and top views) of the CORAL structural models for the acid-induced unfolding of BSA at the native ***N*** and intermediate $\boldsymbol{I_1}$, $\boldsymbol{I_2}$, and ***E*** states indicated. The small beads in (b)–(d) represent the dummy residues of the flexible interdomain linkers between the four rigid domains (color coded) in CORAL fitting algorithm. The conventional assignments of the BSA major domains are marked with **I**, **II**, and **III**, each with **A** and **B** subdomains. Corresponding GABOR envelopes (surface presentation in light brown) are overlapped for comparison.

S11. SAXS data fitting with the ensemble optimization method (EOM) for the *E*-form of BSA

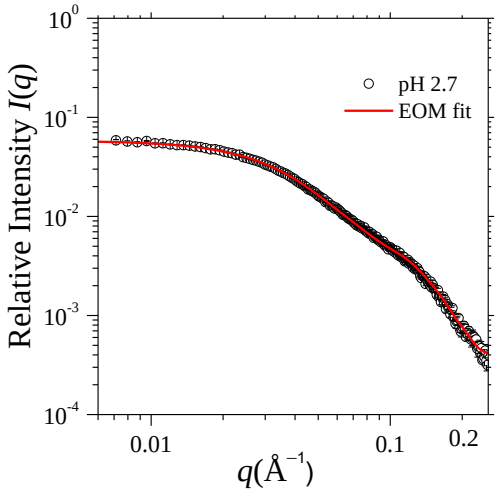


Figure S8. SAXS data for the ***E*** form of BSA partially unfolded at pH 2.7. Data are fitted with the best ensemble generated by EOM (χ^2 = 1.53), comprising the six structural models shown in Figure S9.

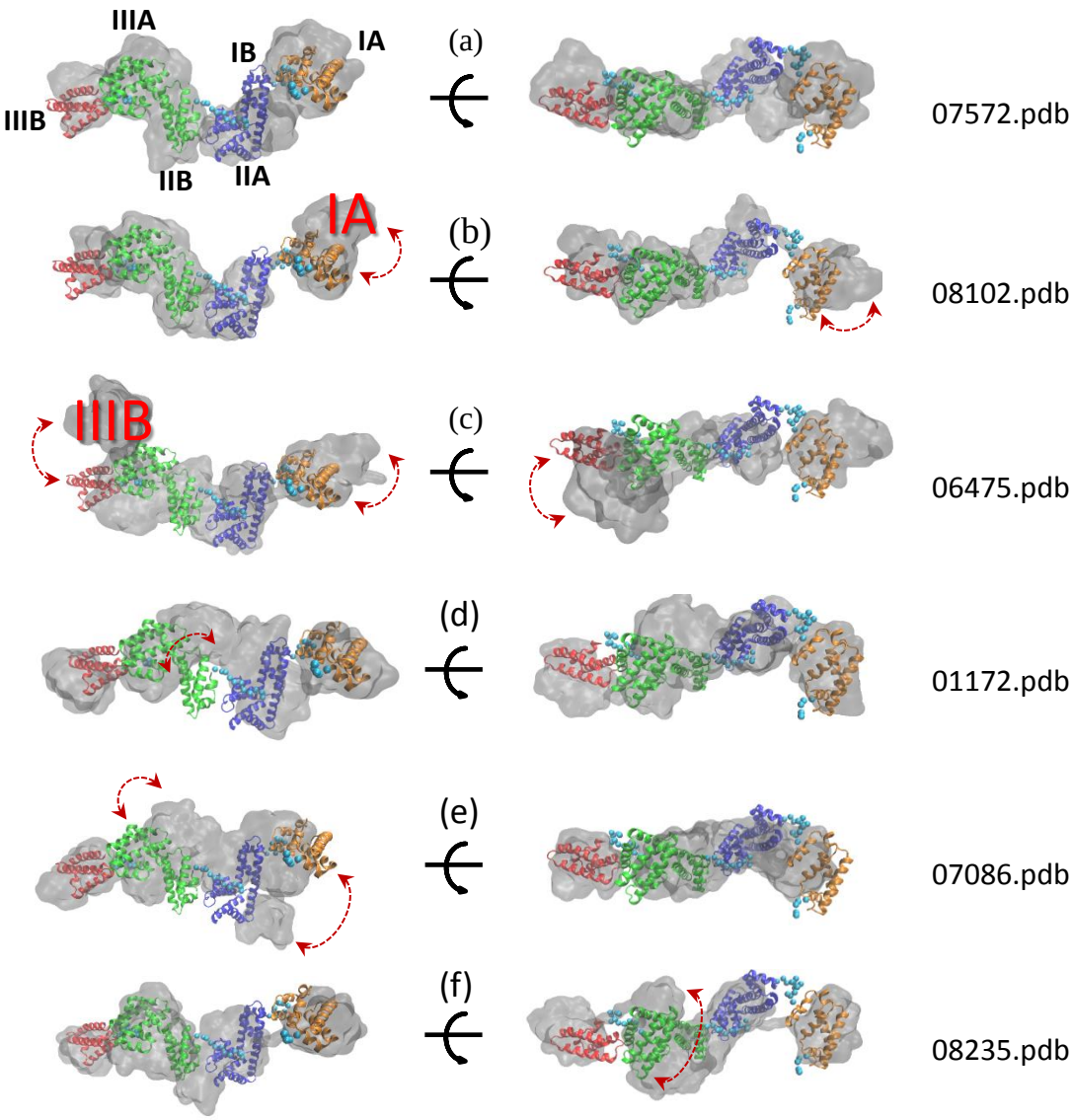


Figure S9. Orthogonal views (right and left) of the six model structures (surface presentations in gray) of the best-fit ensemble generated by EOM for the ***E***-form of BSA at pH 2.7. The models are overlapped with the most representative structure (color coded) obtained by CORAL, revealing the mobile domains through their small deviations from the CORAL model structure, as indicated by the arrows. Overall, **IA** and **IIIB** are relatively more mobile than the other subdomains.

S12. Constant pH Molecular Dynamics (CpHMD) methods

AMBER 14 with the ff14SB force field[s18] was used to perform CpHMD simulations of BSA solutions at pH 2.7. The initial configuration of BSA is taken from the coordinates available in the PDB (code: 4F5S). All the simulations employed the modified generalized Born model (igb = 2 implicit solvation model in AMBER).[s19,s20] Salt concentration was set at 0.1 M

and the cutoff for non-bonded interactions was set to infinity (1000Å). The first stage of the simulation is energy-minimizing the protein, restraining the backbone atoms to their starting positions with a harmonic potential of 10 kcal/mol/Å^2, then slowly heat the system up to 300K while keeping restraints on the backbone atoms. Next, a 2-ns MD simulation at 300 K with no position restraints was carried out to equilibrate the system. All acidic residues (Glu, Asp, and His, totaling 116 titratable residues) were allowed to change protonation states during the equilibrium and production run. The CpHMD method involves periodic Monte Carlo (MC) sampling of protonation states during the standard generalized Born simulation.[s21] At each MC step, a new protonation state is chosen randomly for one titratable residue and the free energy of the transition is computed. The obtained energy is used as the Metropolis criterion to determine whether the transition will be accepted. If the transition is accepted, the simulation continues with the new protonation state. Otherwise, the simulation continues without the change of protonation state. Since BSA has a large number of titratable residues, we use a shorter period for MC steps (protonation state change attempts were made every 2 fs). Finally, after the equilibrium step, the 40 ns production CpHMD simulation was carried out under the same condition as the equilibrium procedure with a time step of 2 fs using GPU-accelerated Amber code.[s22] Trajectories were collected every 2 ps and analyzed with cpptraj.[s18] Figure S10 illustrates the four representative snapshots (as indicated) taken during the 40-ns CpHMD simulation of the BSA pH-induced unfolding, starting from the native conformation at pH 2.7.

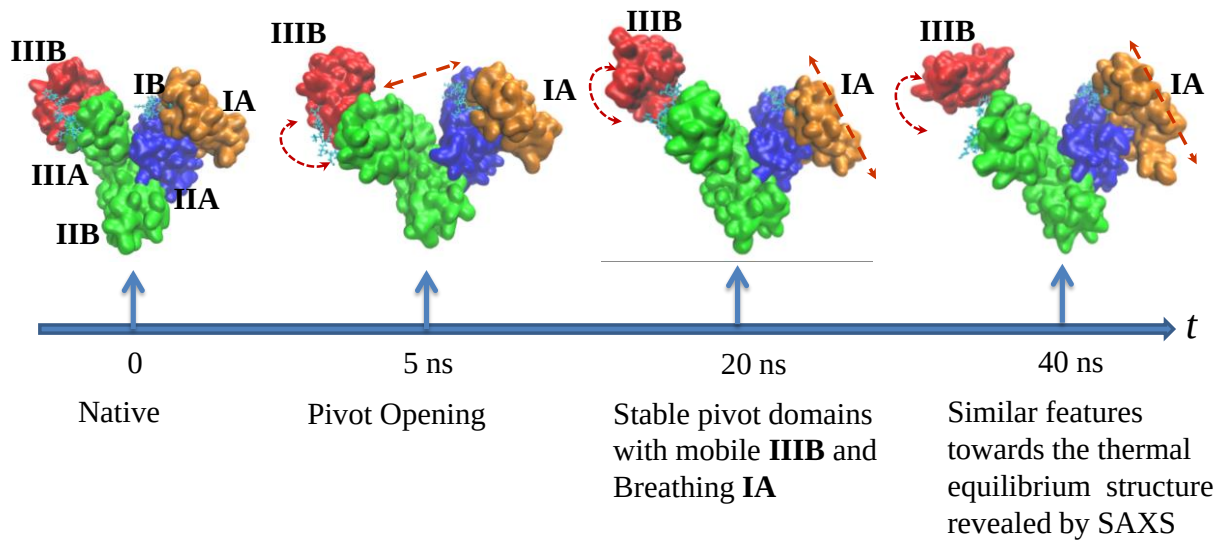


Figure S10. The four representative snapshots (as indicated) taken during the 40-ns CpHMD simulation of the BSA conformation evolution, starting from the native conformation at pH 2.7 (300 K). The structural features are marked by arrows and indicated below the simulation timeline *t*.

Supporting References

(s1) Jeng, U.; Su, C. H.; Su, C. J.; Liao, K. F.; Chuang, W. T.; Lai, Y. H.; Chang, J. W.; Chen, Y. J.; Huang, Y.-S.; Lee, M. T.; Yu, K. L.; Liu, D. G.; Chang, C. F.; Liu, C. Y.; Chang, C. H.; Liang, K. S. A small/wide-angle X-ray scattering instrument for structural characterization of air–liquid interfaces, thin films and bulk specimens. *J. Appl. Cryst.* **2010**, 43, 110.

(s2) Petoukhov, M. V.; Franke, D.; Shkumatov, A. V.; Tria, G.; Kikhney, A. G.; Gajda, M.; Gorba, C.; Mertens, H. D. T.; Konarev, P. V.; Svergun, D. I. New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Cryst.* **2012**, 45, 342-350.

(s3) Jeffries, C. M.; Graewert, M. A.; Blanchet, C. E.; Langley, D. B.; Whitten, A. E.; Svergun, D. I. Preparing monodisperse macromolecular samples for successful biological small-angle X-ray and neutron-scattering experiments. *Nat. Protoc.* **2016**, 11, 2122–2153.

(s4) Blanchet, C. E.; Svergun, D. I. Small-angle X-ray scattering on biological macromolecules and nanocomposites in solution. *Annu. Rev. Phys. Chem.* **2013**, 64, 37-54.

(s5) Svergun, D.; Barberato, C.; Koch, M. H. J. CRYSOL– a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Cryst.* **1995**, 28, 768-773.

(s6) Tria, G.; Mertens, H. D. T.; Kachala, M.; Svergun, D. I. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* **2015**, 2, 207-217.

(s7) Bujacz, A. Structures of bovine, equine and leporine serum albumin. *Acta Cryst.* **2012**, D68, 1278–1289.

(s8) Cornell, C. N.; Kaplan, L. J. Spin-label studies of the sulfhydryl environment in bovine plasma albumin. 1. The N-F transition and acid expansion. *Biochemistry* **1978**, 17, 1750-1754.

(s9) Dockal, M.; Chang, M.; Carter, D. C.; Rüker, F. Conformational Transitions of the Three Recombinant Domains of Human Serum Albumin Depending on pH. *J. Biol. Chem.* **2000**, 275, 3042-3050.

(s10) Sadler, P. J.; Tucker, A. pH-induced structural transitions of bovine serum albumin Histidine pKa, values and unfolding of the N-terminus during the N to F transition. *Eur. J. Biochem.* **1993**, 212, 811-817.

(s11) Curvale R.; Masuelli M.; Padilla A.P. Intrinsic viscosity of bovine serum albumin conformers. *Int. J. Biol. Macromol.* **2008**, 42,133–137.

(s12) Muzammil, S.; Kumar, Y.; Tayyab, S., Molten globule-like state of human serum albumin at low pH. *Eur. J. Biochem.* **1999**, 266, 26-32.

(s13) Khan, M. Y.; Salahuddin, A. Isolation, characterization and effect of acidic pH on the unfolding-refolding mechanism of serum albumin domains. *J. Biosci.* **1990**, 15, 361-376.

(s14) Khan, M. Y. Direct evidence for the involvement of domain III in the N-F transition of bovine serum albumin. *Biochem. J.* **1986**, 236, 307-310.

(s15) Cater, D. C.; Ho, J. X. Structure of serum albumin. *Adv. Protein Chem.* **1994**, 45, 153-204.

(s16) Peters, T. Jr. *All About Albumin: Biochemistry, Genetics, and Medical Application*; Academic Press: San Diego, CA, **1996**.

(s17) Carson, M. Ribbon models of macromolecules. *J. Mol. Graphics* **1987**, 5, 103-106.

(s18) Case, D. A.; Berryman, J. T.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E. I.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; Le Grand, S.; Li, P.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Monard, G.; Needham, P.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Salomon-Ferrer, R.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; York, D. M.; Kollman, P. A. *AMBER 2015*; University of California, San Francisco. **2015**.

(s19) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B* **2000**, 104, 3712-3720.

(s20) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, 55, 383-394.

(s21) Mongan, J.; Case, D. A.; McCammon, J. A. *J. Comput. Chem.* **2004**, 25, 2038-2048.

(s22) Salomon-Ferrer, R.; Gotz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. *J. Chem. Theory Comput.* **2013**, 9, 3878-3888.