NANO MICRO small

Supporting Information

for *Small*, DOI: 10.1002/smll.201603748

Exploring Protein–Nanoparticle Interactions with Coarse-Grained Protein Folding Models

*Shuai Wei, Logan S. Ahlstrom, and Charles L. BrooksIII**

Supporting Information

Exploring protein-nanoparticle interactions with coarse-grained protein folding models.

*Shuai Wei, Logan S. Ahlstrom, and Charles L. Brooks III**

S1. Model details

S1.1. Renormalization of intramolecular folding forces.

The simulated folding temperature (T_f^{sim}) of GB1 with the initial parameters for the G$\bar{o}$-like model was 307 K (Figure S1), while the corresponding experimental value (T_f^{exp}) is 353 K. Thus, we rescaled the strength of native contacts in the model to achieve T_f^{sim} = 350 K (**Figure S1**), which is close to the folding temperature of many proteins of a similar size as GB1. To obtain this desired folding temperature, the strength of each native contact pair is increased by the factor T_f^{exp}/T_f^{sim}. This rescaling, when done in conjunction with the surface potential parameterization should provide a model where both forces of adsorption and forces of folding are balanced, thus enabling us to make quantitative observations regarding the extent of unfolding of the protein on any particular adsorption isotherm or at a particular value of NP surface hydrophobicity.

S1.2. Calculating the adsorption free energy with umbrella sampling.

Specifically, we use a biasing potential of the form

$$V_{umb} = k_u(\xi - \xi_h)^2, \quad \text{(S1)}$$

where k_u = 10 kcal/mol/Å^2 is the force constant, ξ_h is the desired distance between the center of geometry of the protein and the NP center for a particular biasing window, and ξ is the instantaneous distance from the NP center. The values of ξ_h ranged from 401 to 500 Å from the center of the NP (which is 1 to 100 Å from the NP surface) in increments of 1 Å. At 100 Å the protein-NP surface interaction decays to zero. The canonical

ensemble is used and the temperature is maintained by the Nosé-Hoover-Chain integration method with three thermostats of mass $10^{-26}\ kg\ Å^2$.[1, 2, 3] Each simulation consisted of 10 million steps of equilibration and 30 million steps of production with a step size of 10 fs.

We constructed the potential of mean force (PMF) curve, $\omega(r)$, between the protein and the NP by calculating the radial distribution function, g(r), with the weighted histogram analysis method (WHAM).[4] The PMF is divided into two domains based on the distance (r_c) between the protein center of geometry and the center of the NP. The protein-NP adsorption equilibrium constant, K_{ad}, is derived as[5]

$$K_{ad} \approx \int_{r<r_c} 4\pi(r+400)^2 e^{-\beta\omega(r)} dr, \quad \text{(S2)}$$

where r is the distance between the protein and the NP center and $4\pi(r+400)^2 dr$ is the translational volume factor and K_{ad} has the units of volumn (e.g. $Å^3$). This allows for the adsorption free energy of the protein on the NP to be calculated as

$$\Delta G_{ad} = -\frac{1}{\beta}\ln(K_{ad} \cdot C°), \quad \text{(S3)}$$

where the $C°$ is a standard state concentration of 1 mole/L ($= \frac{1}{1661} Å^3$).[5]

S1.3. Analyzing protein stability on NP surfaces with replica exchange simulations.

To understand the effects of NP surface curvature and hydrophobicity on protein GB1 stability, we perform simulations and compare protein GB1 stability in bulk solution, on NPs with diameters of 6 nm, 20 nm, and 80 nm, and on a flat surface, while also on hydrophilic, moderately-hydrophilic, and hydrophobic NPs. The protein GB1 is initially randomly oriented and located at a distance of about 16 Å from the surface of the NP. The temperature range of 280 K to 490 K (for the protein in bulk solution) and of 240 K to 410 K (for the protein on NP/flat surfaces) are covered by 24 replicas (as shown in **Table S1** and **Table S2**), with a temperature spacing of 5 K for the replicas around the

transition point and with a spacing of up to 10 K for the replicas further away. The replicas exchange every 2000 steps.

To track protein folding/unfolding during these simulations, we determine the instantaneous folding fraction, f, or the ratio of the number of native contacts formed at a particular instance relative to the total number of native contacts possible. Over the time course of the simulation, the average of this progress variable is calculated as

$$f(T) = \langle f \rangle_T = \frac{\sum_U f(U)\Omega(U)e^{-\beta U}}{\sum_U \Omega(U)e^{-\beta U}}, \quad \text{(S4)}$$

where U is the potential energy. The key quantity needed to evaluate **Equation S4** is the density of states, $\Omega(U)$, which is calculated using WHAM[4] on the data obtained from replica exchange simulations.

The secondary structure of protein conformations obtained from the simulations were analyzed with PCCASO.[6] This method provides accurate secondary structural estimates based only on the location of the C_α atom of each residue, and is therefore nicely applicable to analyze configurations from the C_α-resolution G$\bar{o}$-like model[7] employed in this work.

S1.4. Hamiltonian Mapping.

To predict the dependence of protein folding on surface hydrophobicity, we employed the Hamiltonian Mapping formalism,[8, 9] which is rooted in WHAM. [4] This approach can efficiently extrapolate changes in the folding and binding behavior of coarse grained models as a function of environmental conditions.[10, 11] We first perform simulations on an original Hamiltonian (H_0, representing a reference hydrophobicity, χ_{ref}). We then reweight the probability distribution of f (the fraction of native contacts) obtained under H_0 to analyze changes in folding under a modified Hamiltonian (H_m, corresponding to a given target hydrophobicity, χ_{target}) using the following equation:

$$P_{reweighted}(f) = \frac{\sum_{k=1}^{R} N_k(f) e^{-\beta H_m^{\chi target}}}{\sum_{k=1}^{R} n_m e^{Fm - \beta m H_0^{\chi ref}}}, \quad \text{(S5)}$$

where $e^{-F_m} = \sum_r P_{reweighted,m}(f)$. R is the total number of simulations (e.g., the number of temperature windows from replica exchange) and n_m is the total number of snapshots in a given window. N_k is the histogram count of configurations with a particular value of f in the kth simulation. The free-energy shifts, F_m, are determined self-consistently[4] and the reweighed probabilities are computed at 298 K.

S2. Supplementary data

Both the original and renormalized G$\bar{o}$-like models of GB1 show a single folding transition point when analyzing the fraction of native contacts formed over a range of temperatures (Figure S1). Moreover, the heat capacity curves as a function of temperature for both cases exhibit one peak at their respective melting temperatures. This behavior indicates a two-state folding mechanism.

Figure S1. The fraction of native contacts formed (solid lines) and heat capacity (C_v, dashed lines) as a function of temperature for the original (black) and renormalized (red) G$\bar{o}$-like models of GB1.

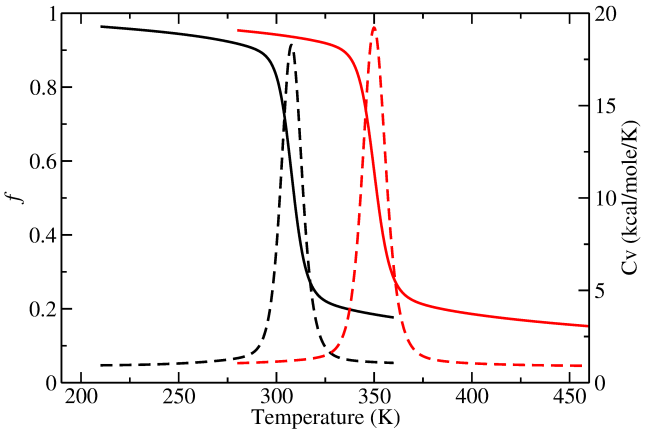
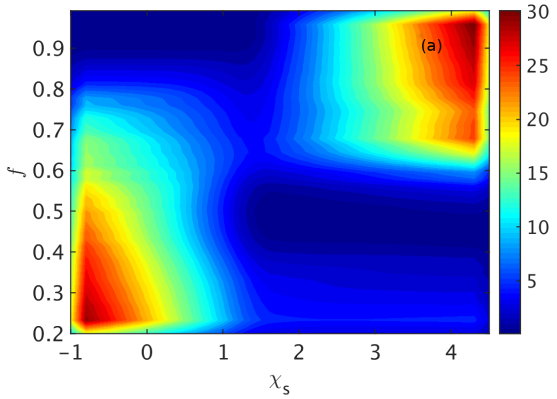
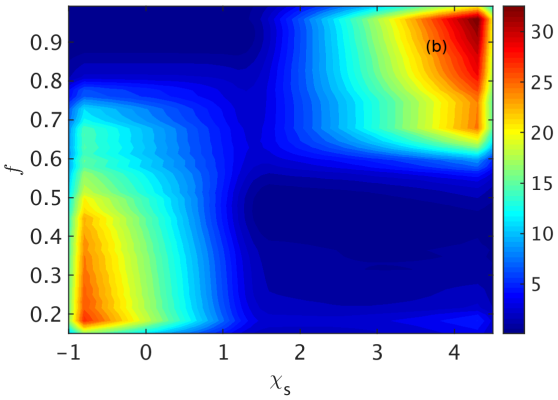


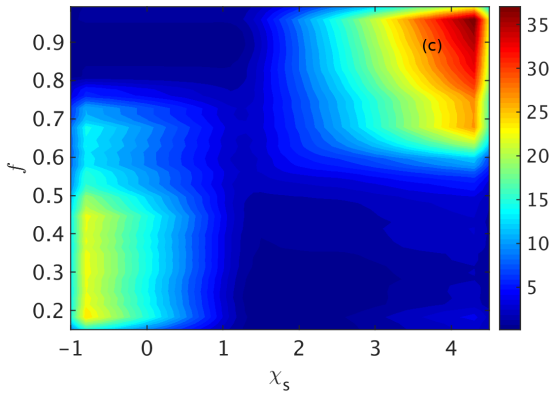
Figure S2. The dependency of the GB1 stability on NP surface hydrophobicity at different temperatures: (a) 290 K; (b) 300 K; (c) 310 K; (d) 320 K; (e) 330 K; (f) 340 K; (g) 350 K.



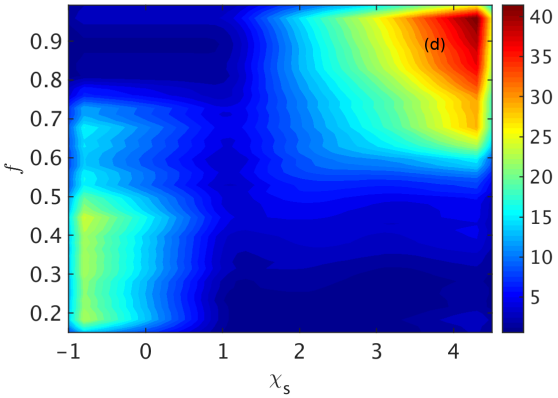
(Large positive value indicates high hydrophobicity)



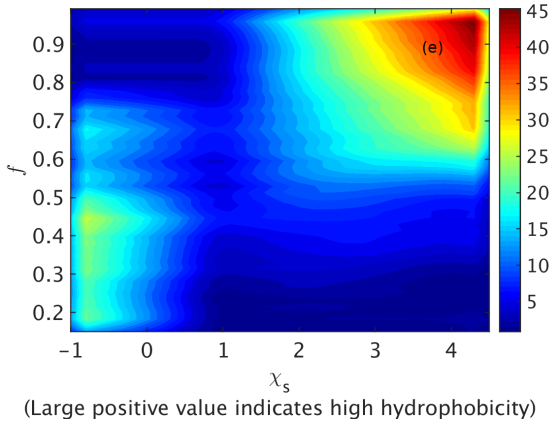
(Large positive value indicates high hydrophobicity)



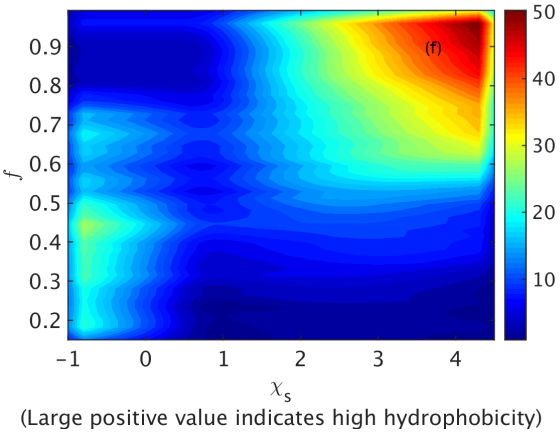
(Large positive value indicates high hydrophobicity)



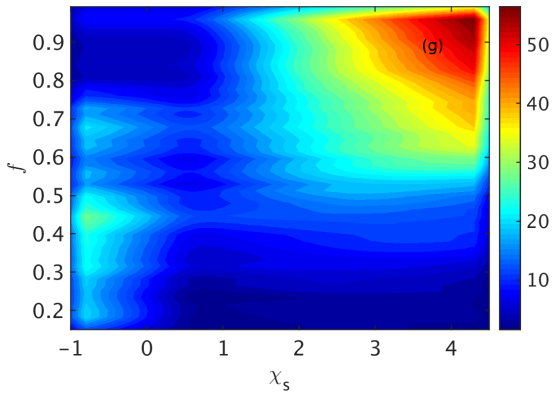
(Large positive value indicates high hydrophobicity)



(Large positive value indicates high hydrophobicity)



(Large positive value indicates high hydrophobicity)



(Large positive value indicates high hydrophobicity)

Table S1. Temperatures (K) used in the replica exchange simulations in the bulk.

280	290	300	310	320	330	340	350
360	365	370	375	380	390	400	410
420	430	440	450	460	470	480	490

Table S2. Temperatures (K) used in the replica exchange simulations on the NP or flat surfaces.

240	250	260	270	280	285	290	295
300	305	310	315	320	325	330	335
340	350	360	370	380	390	400	410

1. Hoover WG. Canonical dynamics: Equilibrium phase-space distributions. Physical Review A. 1985;31:1695-7.
2. Nosé S. A unified formulation of the constant temperature molecular dynamics methods. The Journal of chemical physics. 1984;81:511-9.
3. Nosé S. Constant-temperature molecular dynamics. Journal of Physics: Condensed Matter. 1990;2:SA115.
4. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. Journal of Computational Chemistry. 1992;13:1011-21.

5. Woo H-J, Roux B. Calculation of absolute protein–ligand binding free energy from computer simulations. Proceedings of the National Academy of Sciences of the United States of America. 2005;102:6825-30.
6. Law SM, Frank AT, Brooks CL. PCASSO: A fast and efficient Cα-based method for accurately assigning protein secondary structure elements. J Comput Chem. 2014;35:1757-61.
7. Karanicolas J, Brooks CL, III. The origins of asymmetry in the folding transition states of protein L and protein G. Protein Science. 2002;11:2351-61.
8. Law SM, Ahlstrom LS, Panahi A, Brooks CL. Hamiltonian Mapping Revisited: Calibrating Minimalist Models to Capture Molecular Recognition by Intrinsically Disordered Proteins. The Journal of Physical Chemistry Letters. 2014;5:3441-4.
9. Shea J-E, Nochomovitz YD, Guo Z, Brooks CL. Exploring the space of protein folding Hamiltonians: The balance of forces in a minimalist β-barrel model. The Journal of chemical physics. 1998;109:2895-903.
10. Ahlstrom LS, Law SM, Dickson A, Brooks III CL. Multiscale Modeling of a Conditionally Disordered pH-Sensing Chaperone. Journal of molecular biology. 2015;427:1670-80.
11. O'Brien EP, Brooks BR, Thirumalai D. Effects of pH on Proteins: Predictions for Ensemble and Single-Molecule Pulling Experiments. Journal of the American Chemical Society. 2012;134:979-87.