

# Exploring the Molecular Mechanism of Stabilization of the Adhesion Domains of Human CD2 by N-Glycosylation

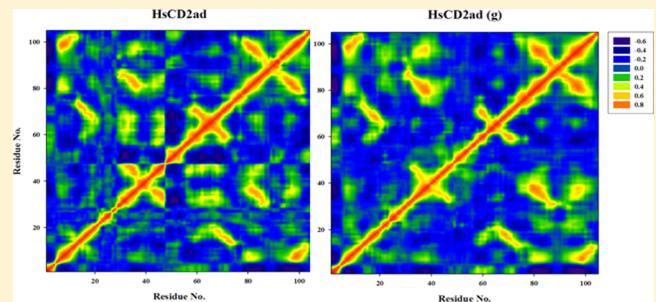
Xing Y. Wang<sup>†,‡</sup> Chang G. Ji<sup>\*,†</sup> and John Z. H. Zhang<sup>\*,†,‡</sup>

<sup>†</sup>State Key Laboratory of Precision Spectroscopy and Department of Physics, Institute of Theoretical and Computational Science, East China Normal University, Shanghai 200062, China

<sup>‡</sup>Department of Chemistry, New York University, New York, New York 10012, United States

## Supporting Information

**ABSTRACT:** N-Glycosylation is one of the most common cotranslational and post-translational modifications occurring in protein biosynthesis and plays a critical role in protein folding and structural diversification. Molecular dynamics studies of two benchmark systems, the NH<sub>2</sub>-terminal human CD2 adhesion domain (HsCD2ad), and the NH<sub>2</sub>-terminal rat CD2 adhesion domain (RnCD2ad) were carried out to investigate the energetic and dynamic effect of N-glycosylation on protein's stability. Our study revealed that N-glycosylation of HsCD2ad at the type I  $\beta$ -bulge turn strengthens the relevant hydrogen bonds, in particular, the hydrogen bond between Asn<sup>65</sup>OD1-Thr<sup>67</sup>HG1. Dynamic cross correlation map analysis showed that nonglycosylated HsCD2ad has strong anticorrelated motions, whereas glycosylated HsCD2ad largely destroyed this anticorrelated motion. As a result, N-glycosylation energetically and dynamically stabilizes HsCD2ad. In contrast, N-glycosylation of RnCD2ad does not display observable effect on protein's stabilization. The current theoretical result is in excellent agreement with the recent thermodynamic experiment of Culyba et al. and indicates that enthalpy and entropy may both contribute to the stabilization of human CD2 by N-glycosylation.



## INTRODUCTION

Glycosylation is a form of cotranslational and post-translational modification that occurs in protein biosynthesis. In eukaryotic cells, the majority of proteins synthesized in the rough endoplasmic reticulum (ER) undergo glycosylation. Glycosylation is linked to some important processes in proteins including structural alteration and modulation of molecular interaction.<sup>1</sup> N-glycosylation is reported to have many structural and functional roles, such as stabilization of protein structure, acceleration of protein folding, promotion of secondary structure formation, reduction of aggregation, shielding of hydrophobic surfaces, facilitation of disulfide pairing, and increase in folding cooperativity.<sup>2–7</sup> The N-glycan addition is noted to occur on asparagine in the sequence context of Asn-X-Ser/Thr, where X denotes any amino acid except proline. Previous studies established that glycosylation could extrinsically affect protein-folding efficiency<sup>2,8</sup> or intrinsically stabilize proteins.<sup>9–12</sup> However, quantitative details and specific molecular mechanisms of glycan-induced stabilization are not well understood.

Human CD2 is a cell surface glycoprotein present on T lymphocytes and natural killer cells and is important in mediating both cellular adhesion and signal transduction by means of interactions with its counter-receptor CD58.<sup>13</sup> The NH<sub>2</sub>-terminal human CD2 adhesion domain (HsCD2ad) is small (105 amino acids), has no prolines or cysteins to complicate its folding, and contains only a single consensus N-

glycosylation site at Asn-65 (Figure 1). The NH<sub>2</sub>-terminal rat CD2 adhesion domain (RnCD2ad) is the ortholog of HsCD2ad and contains a very similar glycosylated region, a type I  $\beta$ -bulge turn geometry<sup>14</sup> (Figure 2B). However, the crystal structure of RnCD2ad is folded without glycosylation, which is due to the lack of Asn-Gly-Thr sequon. Thus, an Asn-Gly-Thr sequon is introduced to RnCD2ad by mutating Asp<sup>61</sup> to Thr<sup>61</sup>. Furthermore, to ensure that there is only one glycosylated site, three mutations Asn<sup>66</sup> to Gln<sup>66</sup>, Asn<sup>76</sup> to Gln<sup>76</sup>, and Asn<sup>83</sup> to Asp<sup>83</sup>, are performed, and the mutated structure is referred to as RnCD2\*. Wild-type RnCD2\* has Glu and Leu at positions 55 and 57, instead of Lys and Phe, as in HsCD2ad<sup>12,15</sup> (Figure 2).

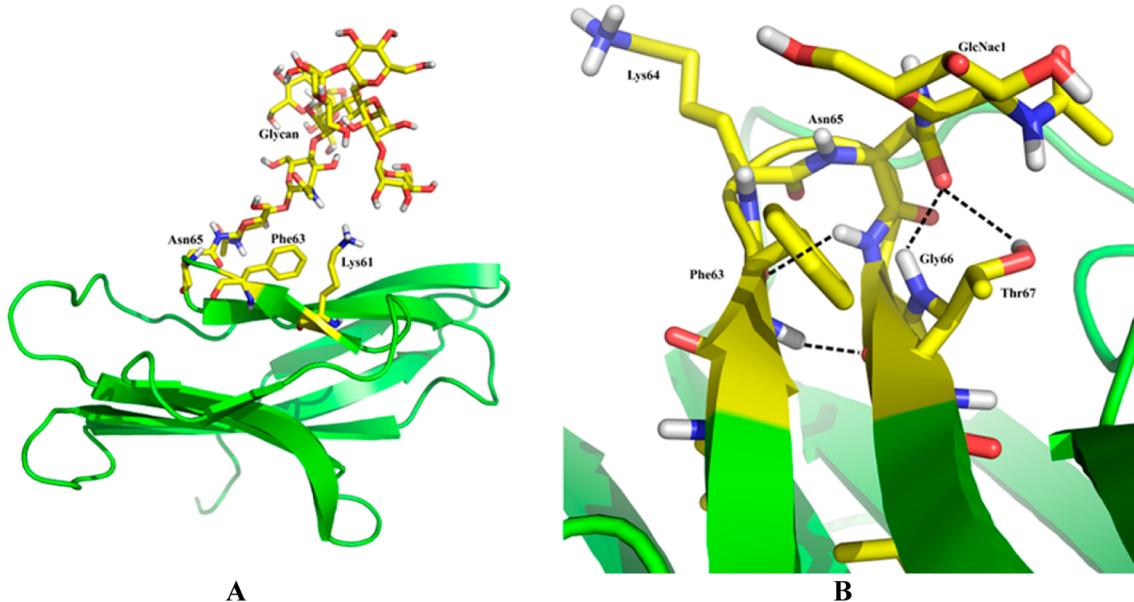
Recently, Kelly and coworkers performed a series of thermodynamics studies for a number of proteins including HsCD2ad and RnCD2\*.<sup>12,15</sup> Their study aimed to provide quantitative energetic analysis on the effect of N-glycosylation on thermostability and folding kinetics. Experimental results showed that glycosylated HsCD2ad is 3.1 kcal/mol more stable and unfolds 50 times more slowly than its nonglycosylated counterpart. Although there was no direct assessment, three hydrogen-bonding interactions within the type I  $\beta$ -bulge turn were implicated for their energetic contribution to the stability

Received: April 29, 2012

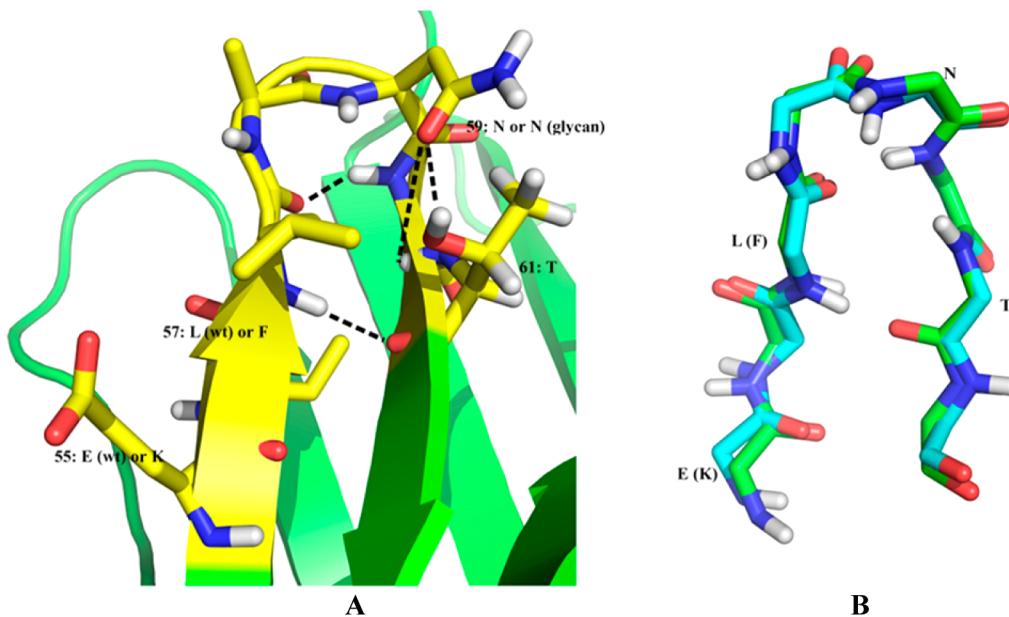
Revised: September 2, 2012

Published: September 4, 2012





**Figure 1.** (A) Crystal structures of HsCD2ad (PDB code 1GYA). Protein is shown in the Ribbon diagram. The N-glycan, Asn<sup>65</sup>, Phe<sup>63</sup>, and Lys<sup>61</sup> are represented with a stick model. (B) Stick representations of the type I  $\beta$ -bulge turn in HsCD2ad. Dashed lines indicate hydrogen bonds. All structures rendered in Pymol ([www.pymol.org](http://www.pymol.org)).



**Figure 2.** (A) Ribbon diagram of the published NMR structure of RnCD2ad (PDB code 1HNG). The type I  $\beta$ -bulge turn (residue 55–61) is highlighted in yellow. Dashed lines indicate hydrogen bonds. (B) Comparison of the type I  $\beta$ -bulge turn of RnCD2\* with the corresponding structure of HsCD2ad, shown with stick models. Corresponding key amino acids in HsCD2ad are marked within parentheses. All structures rendered in Pymol ([www.pymol.org](http://www.pymol.org)).

of the protein upon glycosylation. However, the stabilization effect is not obvious for glycosylated wild-type RnCD2\* compared with its nonglycosylated counterpart. However, stabilization becomes more evident when Lys and Phe are introduced to the glycosylated type I  $\beta$ -bulge turn in mutated RnCD2\*.

To help understand the molecular mechanism of glycosylation and the importance of Asn-Gly-Thr sequon as well as other possible amino acids from a theoretical aspect, we carried out molecular dynamics (MD) simulation for HsCD2ad and RnCD2\* in this work. Specifically, we investigated energetic impact of four hydrogen bonds within the glycosylated type I  $\beta$ -

bulge turn, which were implicated in the experimental study of Culyba et al.,<sup>12,15</sup> in particular, the hydrogen bond between Asn<sup>65</sup>OD1-Thr<sup>67</sup>HG1 in HsCD2ad. In addition, dynamical aspect related to the possible entropic effect of N-glycosylation was also performed to provide a better understanding of the stabilization effect.

## THEORETICAL METHODS

**Initial Structures.** Crystal structures of glycosylated HsCD2ad (PDB code: 1GYA) and nonglycosylated RnCD2ad (PDB code: 1HNG) were obtained from Protein Data Bank

(www.pdb.org). RnCD2\* were built by mutating Asp<sup>61</sup> to Thr<sup>61</sup>, Asn<sup>66</sup> to Gln<sup>66</sup>, Asn<sup>76</sup> to Gln<sup>76</sup>, and Asn<sup>83</sup> to Asp<sup>83</sup>. These mutations were introduced to ensure that the Asn-Gly-Thr amino acid sequence was constructed and only one glycosylation site was kept for RnCD2ad. The nonglycosylated structure was obtained by simply removing the glycan from glycosylated HsCD2ad. The same glycan were also added to Asn<sup>59</sup> of RnCD2\* to build the glycosylated RnCD2\*. Single mutations Glu55Lys and Leu57Phe as well as double mutation Glu55Lys/Leu57Phe were built for both glycosylated and nonglycosylated RnCD2\* to study the impact of Lys and Phe on N-glycosylation. All variants of HsCD2ad and RnCD2\* are listed in Table 1.

**Table 1. Variants Used for Molecular Dynamics Simulations<sup>a</sup>**

HsCD2ad variant	
HsCD2ad	no glycan; wild type sequence
HsCD2ad (g)	glycan; wild type sequence
RnCD2* variant	
RnCD2*	no glycan; wild type sequence
RnCD2* (g)	glycan; wild type sequence
RnCD2*-K	no glycan; Glu55Lys
RnCD2*-K (g)	glycan; Glu55Lys
RnCD2*-F	no glycan; Leu57Phe
RnCD2*-F (g)	glycan; Leu57Phe
RnCD2*-K, F	no glycan; Glu55Lys/Leu57Phe
RnCD2*-K, F (g)	glycan; Glu55Lys/Leu57Phe

<sup>a</sup>There are two variants for HsCD2ad, eight variants for RnCD2\*.

**MD Simulation.** MD simulations were performed using GROMACS package (www.gromacs.org)<sup>16</sup> for all variants. Amber99SB force field<sup>17</sup> and GLYCAM06 force field<sup>18</sup> of AMBER10 were used for proteins and carbohydrates, respectively. GLYCAM was developed to be compatible with AMBER force field<sup>19</sup> and their combination was successfully applied in several carbohydrate–protein complexes and glycoproteins.<sup>20–24</sup> Proteins were placed in rhombic dodecahedron solvation boxes, and periodic boundary conditions were used. Prior to MD simulations, systems were energy minimized to avoid any steric conflicts generated during the initial setup. NVT and NPT equilibration of 100 ps each were performed to help the system reach the desired temperature and pressure. After releasing the position restraints on heavy atoms, a 200 ns production run was performed for each system for data collection and analysis.

## RESULTS AND ANALYSIS

**HsCD2ad. Effect of Glycosylation on Hydrogen Bond Energy.** Hydrogen bonding is one of the dominant interactions in maintaining an intact secondary structure, which plays an important role in protein folding and stabilization. It is defined as a type of attractive interaction between an electronegative atom and a hydrogen atom bonded to another electronegative atom. The experimental studies on HsCD2ad by Kelly and coworkers showed that glycosylated HsCD2ad is 3.1 kcal/mol more stable and unfolds 50 times more slowly than the nonglycan structure, confirming the stabilizing effect of N-glycosylation on HsCD2ad. The crystal structure revealed a compact structure formed among the Phe<sup>63</sup> side chain, the hydrophobic  $\alpha$  face of GlcNAc1 on Asn<sup>65</sup>, and the Thr<sup>67</sup> side

chain. Kelly and coworkers hypothesized that their interactions contributed to the glycan-induced stabilization. Although no direct measurement was made, three hydrogen bonds located within the type I  $\beta$ -bulge turn were implicated to contribute directly to glycan-induced stabilization. These three hydrogen bonds are formed among Thr<sup>67</sup>O-Phe<sup>63</sup>H, Asn<sup>65</sup>OD1-Thr<sup>67</sup>H, and Asn<sup>65</sup>OD1-Thr<sup>67</sup>HG1.

Because direct experimental measurement of energy of these individual hydrogen bonds is not straightforward, theoretical study could provide an attractive means to compute the energetic effect of glycosylation on these individual hydrogen bonds. In the present study, a total of four hydrogen bonds were found within the glycosylated region. A fourth one, neglected in the experimental study by Culyba et al.,<sup>15</sup> is formed between Phe<sup>63</sup>O-Gly<sup>66</sup>H. These four hydrogen bonds are indicated by dashed lines in Figure 1B, and Table 2 lists the

**Table 2. Hydrogen-Bonding Properties (Percentage of Occupancy, Average Distance, and Average Angle) of Glycosylated HsCD2ad and Nonglycosylated HsCD2ad<sup>a</sup>**

	hscd2ad (g)			hscd2ad		
	% occupied	distance	angle	% occupied	distance	angle
Phe <sup>63</sup> O-Gly <sup>66</sup> H	99.54	2.91	152	98.46	2.96	154
Thr <sup>67</sup> O-Phe <sup>63</sup> H	97.34	3.03	159	90.15	3.03	156
Asn <sup>65</sup> OD1-Thr <sup>67</sup> H	97.34	3.13	156	89.96	3.09	159
Asn <sup>65</sup> OD1-Thr <sup>67</sup> HG1	86.37	2.74	163	38.17	2.85	156

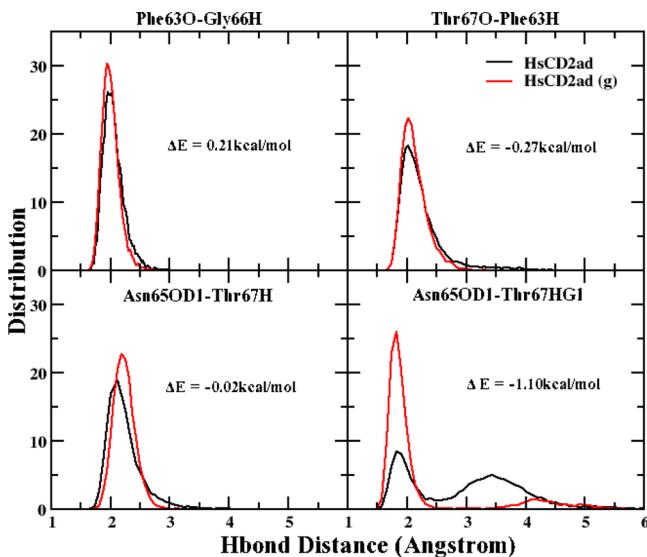
<sup>a</sup>Distances were calculated between two heavy atoms (oxygen and nitrogen).

basic properties of these four hydrogen bonds. The hydrogen bond energy is defined as the sum of electrostatic and van der Waals (vdW) energy. The vdW energy is given by a 6–12 potential, and the electrostatic interaction is modeled by a Coulomb interaction of atom-centered point charges. The hydrogen bond energy is thus given by the following formula

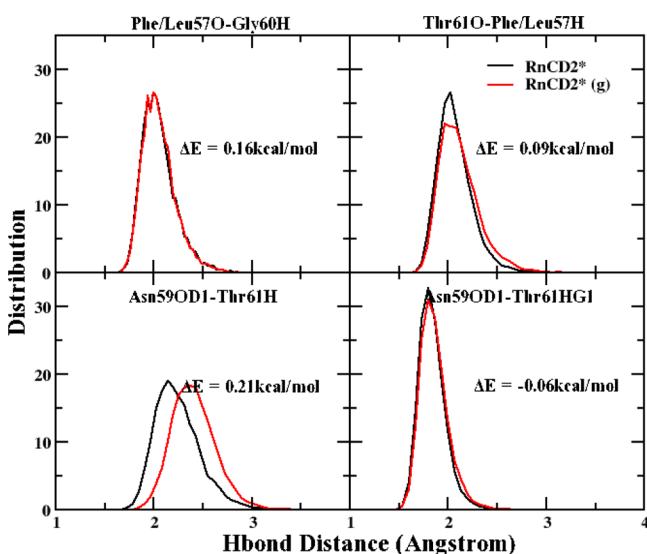
$$E_{\text{bonding}} = \sum_{i < j} \left[ \frac{A_{ij}}{R_{ij}^{12}} \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

where vdW and charge parameters are obtained from Amber99SB force field.<sup>25</sup> To investigate the influence of glycosylation on the structure and energy of these four hydrogen bonds, we plotted distribution of bond lengths for both glycosylated and nonglycosylated hydrogen bonds in Figure 3. The distance in Figure 3 is defined between the oxygen and hydrogen atoms, which is different from the distance defined between two heavy atoms in Table 2.

The two hydrogen bonds (Phe<sup>63</sup>O-Gly<sup>66</sup>H and Thr<sup>67</sup>O-Phe<sup>63</sup>H) are backbone hydrogen bonds within the  $\beta$  turn. They are very stable with more than 90% occupancies in both glycan-form and non-glycan-form. For both of them, percentage occupancies are marginally larger, and distributions of their hydrogen bond distances are only slightly more concentrated in the glycan-form. As a result, the energies of these two hydrogen bonds are not materially affected by the glycosylation. For example, the energy difference for Thr<sup>67</sup>O-Phe<sup>63</sup>H is ~0.27 kcal/mol (lower in glycan-form), and that of Phe<sup>63</sup>O-Gly<sup>66</sup>H is 0.21 kcal/mol (higher in glycan-form). These differences are quite small, and their net effect almost cancels out. The OD1

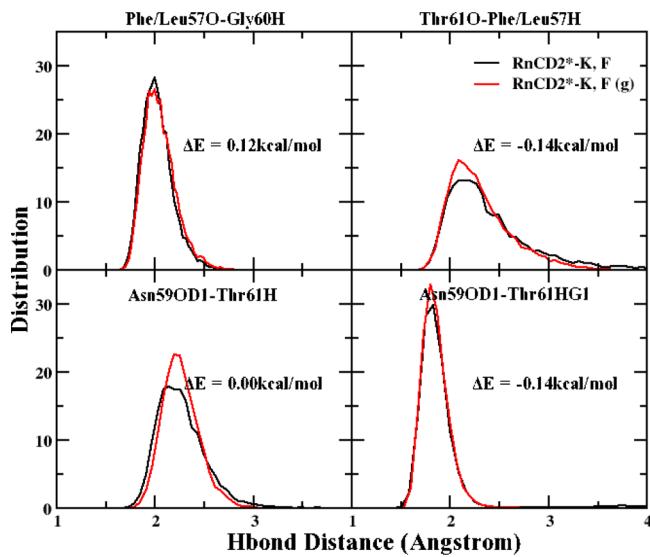


**Figure 3.** Distances (unit: Å) of hydrogen bonds as a function of distribution. Comparison of HsCD2ad and HsCD2ad (g). Distances were calculated between hydrogen and oxygen.  $\Delta E$  indicates the hydrogen-bonding energy difference between glycan-form and non-glycan-form. Upper left: Phe<sup>63</sup>O-Gly<sup>66</sup>H,  $\Delta E = 0.21$  kcal/mol. Upper right: Thr<sup>67</sup>O-Phe<sup>63</sup>H,  $\Delta E = -0.27$  kcal/mol. Lower left: Asn<sup>65</sup>OD1-Thr<sup>67</sup>H (main chain),  $\Delta E = -0.02$  kcal/mol. Lower right: Asn<sup>65</sup>OD1-Thr<sup>67</sup>HG1 (side chain),  $\Delta E = -1.10$  kcal/mol.



**Figure 4.** Distances (unit: Å) of hydrogen bonds as a function of distribution. Comparison of RnCD2\* and RnCD2\*(g). Distances were calculated between hydrogen and oxygen.  $\Delta E$  indicates the hydrogen-bonding energy difference between glycan-form and non-glycan-form. Upper left: Phe/Leu<sup>57</sup>O-Gly<sup>60</sup>H,  $\Delta E = 0.16$  kcal/mol. Upper right: Thr<sup>61</sup>O-Phe<sup>57</sup>H,  $\Delta E = 0.09$  kcal/mol. Lower left: Asn<sup>59</sup>OD1-Thr<sup>61</sup>H (main chain),  $\Delta E = 0.21$  kcal/mol. Lower right: Asn<sup>59</sup>OD1-Thr<sup>61</sup>HG1 (side chain),  $\Delta E = -0.06$  kcal/mol.

on the Asn<sup>65</sup> side chain forms two hydrogen bonds, with both backbone hydrogen and side-chain hydrogen of Thr<sup>67</sup>. Similarly, the calculated energy difference between glycan-form and non-glycan-form of the hydrogen bond formed between Asn<sup>65</sup>OD1-Thr<sup>67</sup>H is negligible (only 0.02 kcal/mol) in this study.



**Figure 5.** Distances (unit: Å) of hydrogen bonds as a function of distribution. Comparison of RnCD2\*-K,F and RnCD2\*-K,F (g). Distances were calculated between hydrogen and oxygen.  $\Delta E$  indicates the hydrogen-bonding energy difference between glycan-form and non-glycan-form. Upper left: Phe/Leu<sup>57</sup>O-Gly<sup>60</sup>H,  $\Delta E = 0.12$  kcal/mol. Upper right: Thr<sup>61</sup>O-Phe<sup>57</sup>H,  $\Delta E = -0.14$  kcal/mol. Lower left: Asn<sup>59</sup>OD1-Thr<sup>61</sup>H (main chain),  $\Delta E = 0.00$  kcal/mol. Lower right: Asn<sup>59</sup>OD1-Thr<sup>61</sup>HG1 (side chain),  $\Delta E = -0.14$  kcal/mol.

**Table 3.** Hydrogen-Bonding Properties (Percentage of Occupancy, Average Distance, And Average Angle) of Glycosylated RnCD2\* and Nonglycosylated RnCD2\*

	RnCD2* (g)			RnCD2*		
	% occupied	distance	angle	% occupied	distance	angle
Leu <sup>57</sup> O-Gly <sup>60</sup> H	97.86	2.92	146	98.82	2.94	149
Thr <sup>61</sup> O-Phe/ Leu <sup>57</sup> H	97.56	3.04	162	98.95	3.01	164
Asn <sup>59</sup> OD1- Thr <sup>61</sup> H	84.06	3.24	154	94.30	3.12	155
Asn <sup>59</sup> OD1- Thr <sup>61</sup> HG1	99.82	2.75	164	99.74	2.73	164

<sup>a</sup>Distances were calculated between two heavy atoms (oxygen and nitrogen).

**Table 4.** Hydrogen-Bonding Properties (Percentage of Occupancy, Average Distance, and Average Angle) of Glycosylated RnCD2\*-K,F and Nonglycosylated RnCD2\*-K,F<sup>a</sup>

	RnCD2*-K,F (g)			RnCD2*-K,F		
	% occupied	distance	angle	% occupied	distance	angle
Phe <sup>57</sup> O-Gly <sup>60</sup> H	98.32	2.92	147	98.79	2.91	149
Thr <sup>61</sup> O-Phe/ Leu <sup>57</sup> H	85.47	3.12	158	77.74	3.12	157
Asn <sup>59</sup> OD1- Thr <sup>61</sup> H	96.70	3.15	155	92.25	3.13	155
Asn <sup>59</sup> OD1- Thr <sup>61</sup> HG1	99.79	2.74	164	96.49	2.73	164

<sup>a</sup>Distances were calculated between two heavy atoms (oxygen and nitrogen).

However, the energy of the hydrogen bond formed between Asn<sup>65</sup>OD1-Thr<sup>67</sup>HG1 exhibits the largest effect on N-

**Table 5. Hydrogen Bonding Energy Difference ( $\Delta E = E(\text{HsCD2ad (g)}) - E(\text{HsCD2ad})$ ), unit: kcal/mol) between Glycan-Form and Non-Glycan-Form for HsCD2ad Variants**

	$\Delta E$
Phe <sup>63</sup> O-Gly <sup>66</sup> H	0.21
Thr <sup>67</sup> O-Phe <sup>63</sup> H	-0.27
Asn <sup>65</sup> OD1-Thr <sup>67</sup> H	-0.02
Asn <sup>65</sup> OD1-Thr <sup>67</sup> HG1	-1.10
total	-1.2
experiment	$\Delta G = -3.1 \text{ kcal/mol}$

glycosylation of HsCD2ad. When the glycan is attached, the percentage of its hydrogen-bonding occupancy increased significantly from 38.2 to 86.4%, indicating a strong stabilizing effect of glycosylation. The distribution of the hydrogen bond length is also significantly altered after glycosylation. As shown in Figure 3, the hydrogen bond length in non-glycan-form has a wide distribution with an additional outer peak near 3.5 Å, which is consistent with its observed low occupancy (38.2%). Upon glycosylation, however, the distribution of the bond length shifted toward a short distance with a value centered around 1.8 Å, indicating a much stable hydrogen bond, as shown in Figure 3. Also, the average hydrogen bond angle is increased from about 156 to 163 °. The result clearly indicates that a much stronger hydrogen bond is formed after glycosylation. Our calculation showed that the energy of this hydrogen bond was lowered by ~1.1 kcal/mol after glycosylation (see Table 5). We believe that this is the major structural effect leading to the more stabilized HsCD2ad after glycosylation.

The above analysis on hydrogen bond energy is highly consistent with the experiment of Culyba et al.,<sup>12,15</sup> which found that the total free energy of the glycan-form HsCD2ad (g) was lowered by 3.1 kcal/mol. Because the present calculation provides just the specific energies of these four hydrogen bonds, it is not appropriate to compare our calculated energy difference directly with the total free energy difference measured by the experiment for the entire protein system. In addition, polarization effect is not included in the present study because the standard nonpolarizable force field is used in our calculation. As a result, the calculated hydrogen bond energy is likely underestimated in the theoretical calculation. In view of these factors, the agreement between theoretical calculation and experimental measurement is quite remarkable.

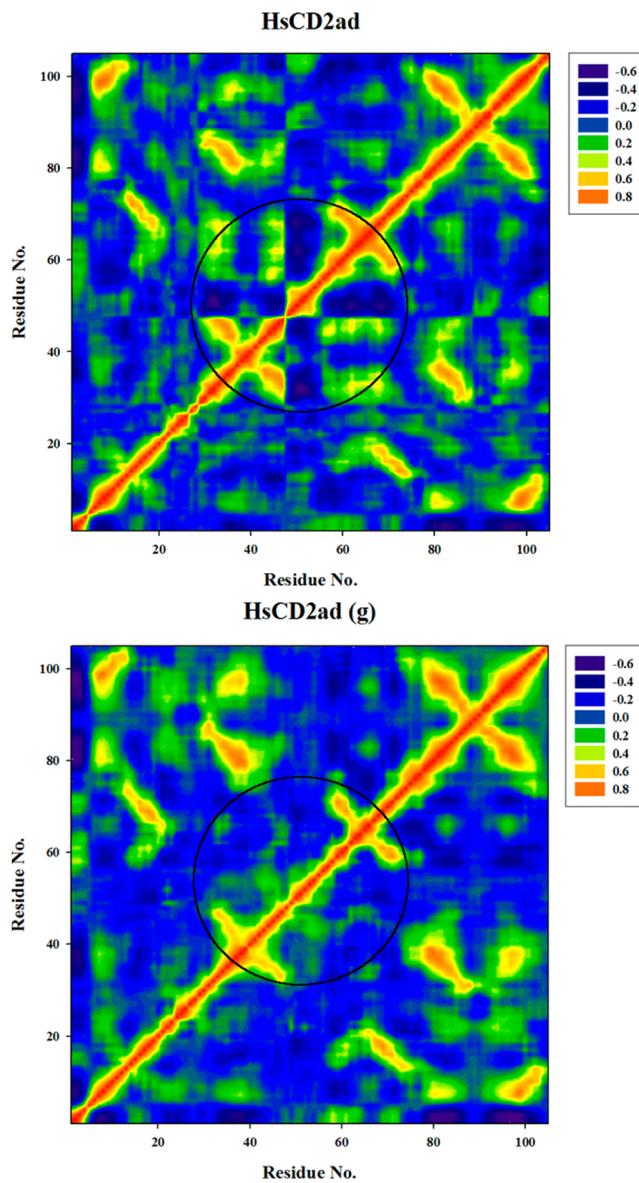
**Internal Dynamics.** In folded proteins, the movements of many amino acids tend to be correlated. To investigate the effect of glycosylation on correlated motions of amino acids in HsCD2ad, we performed longer time MD simulation up to 200 ns to examine dynamical cross-correlation matrices (DCCMs) for both glycan-form (HsCD2ad (g)) and non-glycan-form (HsCD2ad). DCCM is a 3D matrix representation that graphically displays time-correlated information among the residues of the protein. For each pair of C-alpha atoms  $i$  and  $j$ , the cross-correlation coefficient is given by

$$C_{ij} = \frac{\langle \Delta r_i \cdot \Delta r_j \rangle}{\sqrt{\langle (\Delta r_i)^2 \Delta r_j^2 \rangle}}$$

where  $\Delta r_i$  is the displacement from the mean position of the  $i$ th atom and the symbol  $\langle \rangle$  represents time average over the MD trajectory. The magnitude of all pair-wise cross-correlation coefficients is plotted to assess the atomic fluctuations/

displacements of the systems.<sup>26,27</sup> On the DCCM map, each point represents a correlation  $C_{ij}$  of atoms  $i$  and  $j$ . If  $C_{ij} = 1$ , then the fluctuations of atoms  $i$  and  $j$  are completely correlated (same period and same phase). If  $C_{ij} = -1$ , then the fluctuations of atoms  $i$  and  $j$  are completely anticorrelated (same period and opposite phase). If  $C_{ij} = 0$ , then the fluctuations of atoms  $i$  and  $j$  are not correlated.

The DCCM in Figure 6 shows that global dynamics of the non-glycan-form and glycan-form are very similar to each other



**Figure 6.** DCCM of HsCD2ad (Upper Figure: HsCD2ad; Lower Figure: HsCD2ad (g)).

because glycosylation does not change protein's original secondary structures. However, major differences exist in some regions. (See the correlations between residues 49–53 and residues 30–70 in the circled areas in Figure 6.) These differences may reflect how glycosylation changes protein's local dynamics and conformational subspace. For nonglycosylated HsCD2ad, residues 49–53 and residues 30–70 have strong anticorrelated movements, which are indicated by dark blue/purple (upper figure in Figure 6). These anticorrelated

movements disappeared in glycosylated HsCD2ad (lower figure in Figure 6), which indicates a change of the protein's interaction network. For example, when glycan is attached, the original interaction network formed between protein residues and water molecules breaks, and a new interaction network is formed between protein residues and glycan and nearby solvent molecules. The allowed conformational subspace of these residues involved in this interaction network also changes as well as the joint conformational subspace of these residues. The former corresponds to local fluctuation of individual residues, and the latter corresponds to large-scale fluctuations between protein residues. Changes in both local and large-scale fluctuations in protein dynamics contribute to entropy of the system, and DCCM reflects large-scale correlated motions of the protein system. The observed difference in DCCM between HsCD2ad and HsCD2ad (g) indicates that their entropic contribution to thermodynamic stability is different.

**Entropy Analysis.** To estimate the quantitative effect of glycosylation on protein's entropy, configurational entropies of HsCD2ad and its glycan-form were calculated using the quasiharmonic analysis method.<sup>28</sup> In this approach, effective vibrational modes were extracted from trajectories generated in explicit MD simulation, and configurational entropies were calculated using those effective vibrational modes. Consequently, large-scale correlated motions of protein were included in entropy analysis. Entropy calculated from quasiharmonic analysis covers both local and large-scale correlated fluctuations. To preserve the same number of atoms for entropy comparison, glycan and glycan-attached Asn were not included in entropy calculation. The calculated values of entropy are somewhat dependent on the simulation time used.<sup>29,30</sup> Figure 7

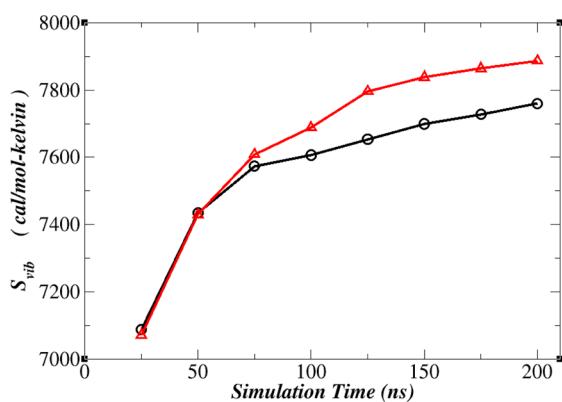


Figure 7. Vibrational entropy computed by quasiharmonic analyses for different lengths of MD trajectories (Black: HsCD2ad; Red: HsCD2ad (g)).

plots the entropies calculated from various time segments of MD trajectories with 25, 50, 75, 100, 125, 150, 175, and 200 ns, respectively. In our calculation, trajectory was saved every 0.1 ps.

The result shows that protein's internal configurational entropy increases after glycosylation. This is consistent with the above analysis from DCCM map. (See Figure 6.) Considering the thermodynamic equation:  $\Delta G = \Delta H - T\Delta S$ , the increase in vibrational entropy could decrease the free energy and contribute to the glycan-induced stabilization. The result suggests that glycosylation could also affect the stability of folded structure by perturbing vibrational motion of protein.

Such effect may be part of the functionalization process of glycosylation.

**RnCD2\*.** The energies of the same four key hydrogen bonds, Phe/Leu<sup>57</sup>O-Gly<sup>60</sup>H, Thr<sup>61</sup>O-Phe/Leu<sup>57</sup>H, Asn<sup>59</sup>OD1-Thr<sup>61</sup>H, and Asn<sup>59</sup>OD1-Thr<sup>61</sup>HG1 (Figure 2A), are investigated for RnCD2\*. To study the impact of Phe<sup>57</sup> and Lys<sup>55</sup> at type I  $\beta$ -bulge turn for glycosylation-induced stabilization, we studied variants built by mutating Glu<sup>55</sup> to Lys<sup>55</sup> and Leu<sup>57</sup> to Phe<sup>57</sup> as well.

**Wild Type and Single Mutation.** Our result shows that glycosylation has no significant stabilization impact on wild-type RnCD2\*. In the glycan-form, percentage occupancies of three hydrogen bonds (Phe/Leu<sup>57</sup>O-Gly<sup>60</sup>H, Thr<sup>61</sup>O-Phe/Leu<sup>57</sup>H, and Asn<sup>59</sup>OD1-Thr<sup>61</sup>H) are all slightly lower than its nonglycosylated counterpart. The energies of these three hydrogen bonds are 0.16, 0.09, and 0.21 kcal/mol higher in glycosylated structure. The Asn<sup>59</sup>OD1-Thr<sup>61</sup>HG1 hydrogen has slightly stronger bonding energy in glycan-form. The total energy of these four hydrogen bonds is 0.40 kcal/mol higher than that in nonglycosylated form. Our study also shows that there is no significant difference in energy between glycan- and non-glycan-forms in singly mutated RnCD2\* (see Table 6),

Table 6. Hydrogen-Bonding Energy Difference between Glycan-Form and Non-Glycan-Form for RnCD2\* Variants<sup>a</sup>

	$\Delta E(W)$	$\Delta E(K)$	$\Delta E(F)$	$\Delta E(K,F)$
Phe/Leu <sup>57</sup> O-Gly <sup>60</sup> H	0.16	0.22	0.22	0.12
Thr <sup>61</sup> O-Phe/Leu <sup>57</sup> H	0.09	-0.29	0.19	-0.14
Asn <sup>59</sup> OD1-Thr <sup>61</sup> H	0.21	0.26	-0.04	0.00
Asn <sup>59</sup> OD1-Thr <sup>61</sup> HG1	-0.06	0.04	-0.17	-0.14
total	0.40	0.23	0.20	-0.16
experiment $\Delta G$ (kcal/mol)	$-0.6 \pm 0.6$	$-1.5 \pm 0.6$	$-1.8 \pm 0.4$	$-3.7 \pm 0.6$

<sup>a</sup> $\Delta E(X) = E(\text{RnCD2* (g)}) - E(\text{RnCD2*})$  (in kcal/mol), X = W, for wild type; X = K, for mutation Glu55Lys; X = F, for mutation Leu57Phe; X = K, F, for mutations Glu55Lys and Leu57Phe.

with the mutated RnCD2\* being slightly more stable in the glycan-form than the wild type. The theoretical result is quite consistent with the experimental measurement of Culyba et al.<sup>15</sup>

**Double Mutation.** The stabilization effect becomes relatively evident, however, when both mutations Glu55Lys and Leu57Phe occur at the glycosylation loop. Among the four hydrogen bonds, the total energy is lowered by  $\sim 0.16$  kcal/mol after glycosylation. Despite the fact that our calculated energy difference is quite small, this is qualitatively consistent with the experimental result of Culyba et al. showing enhanced stabilizing effect of glycosylation in doubly mutated RnCD2\*.<sup>15</sup> This might help confirm the experimental hypothesis that Lys and Phe play important roles in stabilizing the glycan-form. The basic properties of the four hydrogen bonds for wild and double mutants of RnCD2\* are listed in Tables 3 and 4 and distributions of hydrogen bond lengths are plotted in Figures 4 and 5.

## CONCLUSIONS AND DISCUSSION

In this work, long-time MD simulations (200 ns each) in explicit solvent were performed for both glycan-form and non-

glycan-form of HsCD2ad and RnCD2\*. Detailed hydrogen-bonding energetics and dynamical analysis were performed, and the molecular mechanism of N-glycosylation-induced stabilization was analyzed.

In the HsCD2ad system, four hydrogen bonds Phe<sup>63</sup>O-Gly<sup>66</sup>H, Thr<sup>67</sup>O-Phe<sup>63</sup>H, Asn<sup>65</sup>OD1-Thr<sup>67</sup>H, and Asn<sup>65</sup>OD1-Thr<sup>67</sup>HG1 are found within the glycosylated type I  $\beta$ -bulge turn. Except Phe<sup>63</sup>O-Gly<sup>66</sup>H, the N-glycosylation decreases bonding energies of the other three hydrogen bonds by 0.27, 0.02, and 1.1 kcal/mol, respectively. In particular, the hydrogen bond between Asn<sup>65</sup>OD1-Thr<sup>67</sup>HG1 largely increases its percentage occupancy during the MD simulation from 38.2 to 86.4%, indicating its greatest effect on N-glycosylation of HsCD2ad. As a result, total hydrogen-bonding energies of these four hydrogen bonds are 1.2 kcal/mol lower in glycan-form, which is consistent with the observed total free energy difference of 3.1 kcal/mol (lower in glycan-form) from thermo experiment.<sup>15</sup>

The inner dynamic motion in HsCD2ad is also affected by N-glycosylation. When the glycan is attached, the original strong anticorrelated motion between residues 49–53 and residues 30–70 becomes less correlated, indicating the extension of the system's conformational subspace. The larger conformational subspace means that the glycosylated protein is entropically more stable than nonglycosylated form. This result indicates that stabilization of glycosylated HsCD2ad is also dynamically driven.

In the RnCD2\* system, no significant stabilization impact of N-glycosylation is observed on wild-type RnCD2\*. Although the stabilization effect is still not significant with single mutation of either Glu55Lys or Leu57Phe, its glycan-form is relatively more stable than wild type. When both mutations Glu55Lys and Leu57Phe are present, the stabilization effect is somewhat enhanced. The theoretical result is in good agreement with recent experiment and helped confirm the importance of Phe and Lys at the glycosylated region on N-glycosylation.

Glycosylation dramatically impacts protein stability, protein solubility, protein folding dynamics, and function. Previous experimental studies indicate that N-linked glycans have different roles in small peptides<sup>31</sup> and large glycoproteins.<sup>13,32</sup> For small peptides, attached glycans usually help stabilize the compact conformation.<sup>33</sup> For large proteins, glycans may either stabilize or destabilize folded proteins.<sup>34–36</sup> Previous computational studies on model peptide systems suggested that conformations are uniquely influenced by the attached saccharide.<sup>33,36,37</sup> However, few computational efforts have been made to study large glycoproteins. The current simulation study shows that although no substantial conformational change occurred in HsCD2ad (g), the strength of certain hydrogen bonds and some large-scale correlated motions of protein was affected upon glycosylation of HsCD2ad.

In the current study, two possible microscopic mechanisms of glycosylation on protein's stability were proposed from theoretical calculation and analysis. One is that glycan may shield intraprotein hydrogen bond from solvent and thus energetically stabilizes specific hydrogen bonds. Energetic contribution of hydrogen bond to protein folding can be affected through change of hydrophobicity of the glycosylation site, as has been demonstrated in double-mutation experiment<sup>38</sup> and computational study.<sup>39</sup> The other mechanism of glycosylation is that glycan may affect local and large-scale correlated motions of protein and thus impact the entropy of

the folded protein. Previous study of DeKoster et al. showed that glycosylation-induced entropic stabilization is the main contributor in Ovomucoid first domain.<sup>40</sup> The current theoretical study and previous experimental investigation suggest that glycosylation may play complex roles in protein's stability depending on protein's interaction network around the glycosylation site. This is the origin of "context-dependent" glycosylation effect. Further work along this line to uncover molecular mechanism may help us understand how glycosylation shapes protein's function under different environment.

Our current study was limited in scope because only folded structure and dynamics were investigated. To obtain more complete information on thermodynamic contribution of the above-mentioned effect, we needed to analyze the unfolded state.<sup>34,41–43</sup> Although such study may be computationally possible for small model peptides, it is computationally prohibitive for large realistic proteins in all atom MD simulation. For example, HsCD2ad has over 200 residues, and its conformational space is too large to sample in unfolded state. However, recent work using coarse-grained model to study the influence of glycosylation on unfolded state of protein has been reported,<sup>34</sup> in which the authors suggested that the stabilization effect may result from the raised energy of unfolded state rather than the lowered energy of folded state as a result of glycosylation. Thus further detailed theoretical study is needed to provide a more complete picture for the mechanism of stabilization of protein by N-glycosylation, in which both folded and unfolded states of protein should be investigated in a consistent fashion.

## ASSOCIATED CONTENT

### S Supporting Information

Plots of potential energy and rmsd as a function of simulation time for HsCD2ad and RnCD2\* in both glycan and non-glycan-forms. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: chicago.ji@gmail.com (C.J.); john.zhang@nyu.edu (J.Z.).

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank the National Natural Science Foundation of China (Grant Nos. 21003048, 10974054, and 20933002) and Shanghai PuJiang program (09PJ1404000) for financial support. C.G.J. is also supported by "the Fundamental Research Funds for the Central Universities" and Open Research Fund of the State Key Laboratory of Precision Spectroscopy, East China Normal University. We also thank the Computational Center of ECNU for providing computational time.

## REFERENCES

- (1) Varki, A. *Glycobiology* 1993, 3, 97–130.
- (2) Helenius, A.; Aeby, M. *Science* 2001, 291, 2364–2369.
- (3) Wormald, M. R.; Dwek, R. A. *Structure* 1999, 7, 155–160.
- (4) Imperiali, B.; O'Connor, S. E. *Curr. Opin. Struct. Biol.* 1999, 3, 643–649.
- (5) Mitra, N.; Sinha, S.; Ramya, T. N.; et al. *Trends Biochem. Sci.* 2006, 31, 156–163.

- (6) Petrescu, A. J.; Milac, A. L.; Petrescu, S. M.; et al. *Glycobiology* **2004**, *14*, 103–114.
- (7) O'Connor, S. E.; Pohlmann, J.; Imperiali, B.; et al. *J. Am. Chem. Soc.* **2001**, *123*, 6187–6188.
- (8) Molinari, M. *Nat. Chem. Biol.* **2007**, *3*, 313–320.
- (9) Glzman, R.; Okiyoneda, T.; Mulvihill, C. M.; et al. *J. Cell Biol.* **2009**, *184*, 847–862.
- (10) Wormald, M. R.; Wooten, E. W.; Bazzo, R.; et al. *Eur. J. Biochem.* **1991**, *198*, 131–139.
- (11) Andreotti, A. H.; Kahne, D. *J. Am. Chem. Soc.* **1993**, *115*, 3352–3353.
- (12) Hanson, S. R.; Culyba, E. K.; Hsu, T. L.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 3131–3136.
- (13) Wyss, D. F.; Choi, J. S.; Li, J.; et al. *Science* **1995**, *269*, 1273–1278.
- (14) Jones, E. Y.; Davis, S. J.; Williams, A. F.; et al. *Nature* **1992**, *360*, 232–239.
- (15) Culyba, E. K.; Price, J. L.; Hanson, S. R.; et al. *Science* **2011**, *331*, 571–575.
- (16) Hess, B. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (17) Hornak, V.; Abel, R.; Okur, A.; et al. *Proteins* **2006**, *65*, 712–725.
- (18) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; et al. *J. Comput. Chem.* **2008**, *29*, 622–655.
- (19) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; et al. *AMBER 11*; University of California: San Francisco, 2010.
- (20) DeMarco, M. L.; Woods, R. J. *Glycobiology* **2008**, *18*, 426–440.
- (21) Guardiani, C.; Signorini, G. F.; Livi, R.; et al. *J. Phys. Chem. B* **2012**, *116*, 5458–5467.
- (22) Islam, S. M.; Richards, M. R.; Taha, H. A.; et al. *J. Chem. Theory Comput.* **2011**, *7*, 2989–3000.
- (23) Cezard, C.; Trivelli, X.; Aubry, F.; et al. *Phys. Chem. Chem. Phys.* **2011**, *13*, 15103–15121.
- (24) Yao, J.; Nellas, R. B.; Glover, M. M.; et al. *Biochemistry* **2011**, *50*, 4097–4104.
- (25) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; et al. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (26) Swaminathan, S., Jr.; Harte, W. E.; Beveridge, D. L. *J. Am. Chem. Soc.* **1991**, *113*, 2717–2721.
- (27) Arnold, G. E.; Ornstein, R. L. *Biophys. J.* **1997**, *73*, 1147–1159.
- (28) Karplus, M.; Kushick, J. N. *Macromolecules* **1981**, *14*, 325–332.
- (29) Harris, S. A.; Gavathiotis, E.; Searle, M. S.; et al. *J. Am. Chem. Soc.* **2001**, *123*, 12658–12663.
- (30) Gohlke, H.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 238–250.
- (31) Live, D. H.; Kumar, R. A.; Beebe, X.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 12759–12761.
- (32) Narhi, L. O.; Arakawa, T.; Aoki, K. H.; et al. *J. Biol. Chem.* **1991**, *266*, 23022–23026.
- (33) Ellis, C. R.; Maiti, B.; Noid, W. G. *J. Am. Chem. Soc.* **2012**, *134*, 8184–8193.
- (34) Cho, S. S.; Levy, Y.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 434–439.
- (35) Choi, S.; Reixach, N.; Connelly, S.; et al. *J. Am. Chem. Soc.* **2010**, *132*, 1359–1370.
- (36) Lu, D.; Yang, C.; Liu, Z. *J. Phys. Chem. B* **2012**, *116*, 390–400.
- (37) Bosques, C. J.; Tschampel, S. M.; Woods, R. J.; et al. *J. Am. Chem. Soc.* **2004**, *126*, 8421–8425.
- (38) Gao, J.; Bosco, D. A.; Powers, E. T.; et al. *Nat. Struct. Mol. Biol.* **2009**, *16*, 684–690.
- (39) Ji, C. G.; Xiao, X. D.; Zhang, J. Z. *H. J. Chem. Theory Comput.* **2012**, *8*, 2157–2164.
- (40) DeKoster, G. T.; Robertson, A. D. *Biochemistry* **1997**, *36*, 2323–2331.
- (41) Shental-Bechor, D.; Levy, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 8256–8261.
- (42) Shental-Bechor, D.; Levy, Y. *Curr. Opin. Struct. Biol.* **2009**, *19*, 524–533.
- (43) Price, J. L.; Shental-Bechor, D.; Dhar, A.; et al. *J. Am. Chem. Soc.* **2010**, *132*, 15359–15367.