Electronic Supplementary Material (ESI) for Chemical Science.

Supplementary Information

Determine Protein Conformation and Orientation at Buried Solid/Liquid Interfaces

Wen Guo[1], Tieyi Lu[1], Ralph Crisci,[1] Satoshi Nagao[2], Tao Wei[3], Zhan Chen[1*]

[1]Department of Chemistry, University of Michigan, 930 North University Avenue, Ann Arbor, Michigan 48109, USA.

[2]Graduate School of Science, University of Hyogo, 3-2-1 Koto, Ako-gun Kamigouri-cho, Hyogo, Japan, 678-1297

[3]Department of Chemical Engineering, Howard University, 2366 Sixth Street, NW Washington, DC 20059, USA.

Table of Contents

S1. Preparation of WT GB1 and MT GB1 with and without Isotopic Labels

A cold-shock expression system was employed for high efficiency isotopic labeling of the proteins. The protocol used for isotopic labeling was similar to the previously reported protocol.[1] The pCold IV plasmid (Takara Bio, Japan) and *E. coli* BL21(DE3) strain (New England Biolabs, MA) were used for expression of the WT GB1 and MT GB1 genes. A single colony of *E. coli* was precultured in 10 mL LB medium at 37°C overnight, and the whole cell suspension was placed in 1 L LB medium and continued cultivation at 37°C until the OD_{600} value reached 0.6−0.8. The cells were collected by centrifugation and resuspended in 1L M9 minimal medium containing non-isotope-labeled ammonium chloride and D-glucose with or without 0.1 g ^{13}C-labeled amino acid at the carbonyl position (L-leucine (1-13C, 99%), L-valine (1-13C, 99%), L-phenylalanine (1-13C, 99%), L-lysine:2HCl (1-13C, 99%), and L-isoleucine (1-13C, 99%), Cambridge Isotope Laboratories, Inc., MA). After about one hour cultivation at 37°C, the medium was chilled to 15°C and continued to cultivate until the OD_{600} value reached 0.8−1.0. Then, isopropyl β-D-1-thiogalactopyranoside (IPTG) (final concentration: 0.5 mM) was added to the medium for the induction and cultivated at 15°C for 24 hours. The cells were collected by centrifugation at 8,500g for 5 min. The harvested cells were resuspended with lysis buffer (50 mM potassium phosphate, pH 7.0), and then disrupted by sonication. The cell debris was removed by centrifugation at 34,000g for 30 min, and the supernatant was collected. The WT GB1 and MT GB1 proteins were purified from the supernatant by anion exchanging (HiPrep DEAE FF 16/10, GE Healthcare, Buckinghamshire) and size-exclusion (HiLoad 26/600 Superdex 75 pg, GE Healthcare, Buckinghamshire) chromatography using a FPLC system (BioLogic DuoFlow 10, Bio-Rad, CA). Molecular weights of the purified WT GB1 and MT GB1 proteins were analyzed by MALDI-TOF MS spectrometry (ultrafleXtreme, Bruker Daltonics, USA) using sinapinic acid as a matrix in

linear mode (Figure S1). The mass peaks of WT NL and MT NL were observed at m/z = 6067 and 5967, respectively, which are corresponding to the average molecular weights of WT NL ([M+H^+] = 6065.62) and MT NL ([M+H^+] = 5965.54) without N-terminal Met. On the other hand, the mass peaks of WT Leu, MT Leu, MT Val, MT Phe, MT Lys, and MT Ile were observed at 1~5 higher m/z compared to those of WT NL and MT NL. The results suggest that WT GB1 and MT GB1 proteins with and without isotope labels were successfully obtained.

To confirm specific amino acids were isotope-labeled in the GB1 proteins, we analyzed the mass of trypsin-digested fragments by MALDI-TOF MS spectrometry using α-cyano-4-hydroxy-cinnamic acid in reflector mode. The GB1 proteins were incubated for digestion using trypsin (final concentration: 2 μg/mL) at 37°C for 3 hours, and then at 30 °C for 18 hours. The trypsin digestion of the GB1 proteins was quenched by addition of trifluoroacetate (final concentration: 0.2 %). The amino acid sequence, sequence number, and monoisotopic mass number of the trypsin-digested peptide fragments of WT GB1 and MT GB1 proteins were shown in Figure S2(a). As shown in the mass spectra of non-digested GB1 proteins (Figure S1), the [1] fragment (N-terminal Met) was naturally cleaved during expression in *E. coli* cells or purification from the cells. There are seven fragments if the GB1 proteins are digested by trypsin that can cleave at the position after each Lys residue (the GB1 proteins have no Arg residues). If the trypsin-digested peptide fragment contains isotope-labeled amino acids, its mass peak should be shifted +1 Da for each labeled residue. This means that isotope labeling efficiency can be determined by the peak shift in the mass number of the labeled sample from the monoisotopic sample. In the mass spectra, the mass peaks of the [51-56] fragment (C-terminal region of the GB1 proteins) could not be separately observed due to overlapping with those of modified fragments, but all other fragments were identified (Figure S2(b)-(g)). For example, the [5-10] fragment contains two Leu,

one Lys, and one Ile. Therefore this fragment exhibited +2 Da shift for WT Leu and MT Leu, and a +1 Da shift for MT Lys and MT Ile (Figure S2(c)). On the other hand, the mass peak attributed to the [5-10] fragment from the other GB1 proteins (not labled with Leu, Lys, or Ile) were observed at its monoisotopic mass number ([M+H+] = 657.43). The shifts in the mass number by isotope labeling of the GB1 proteins are shown in Table S1. WT GB1 and MT GB1 proteins contain three Leu, four Val, two Phe, six Lys, and one Ile. This amino acid composition corresponds well to the total shifts in the mass number in Table S1, although the information of the [51-56] fragment is missing. Additionally, the increase in the mass number of the whole GB1 proteins (without trypsin digestion) also supports that the isotope labeling of the amino acids in the GB1 proteins was highly specific and efficient. The present isotope labeling method of specific amino acids in a protein and the characterization method using mass spectrometry are general, which can be applied to other proteins.

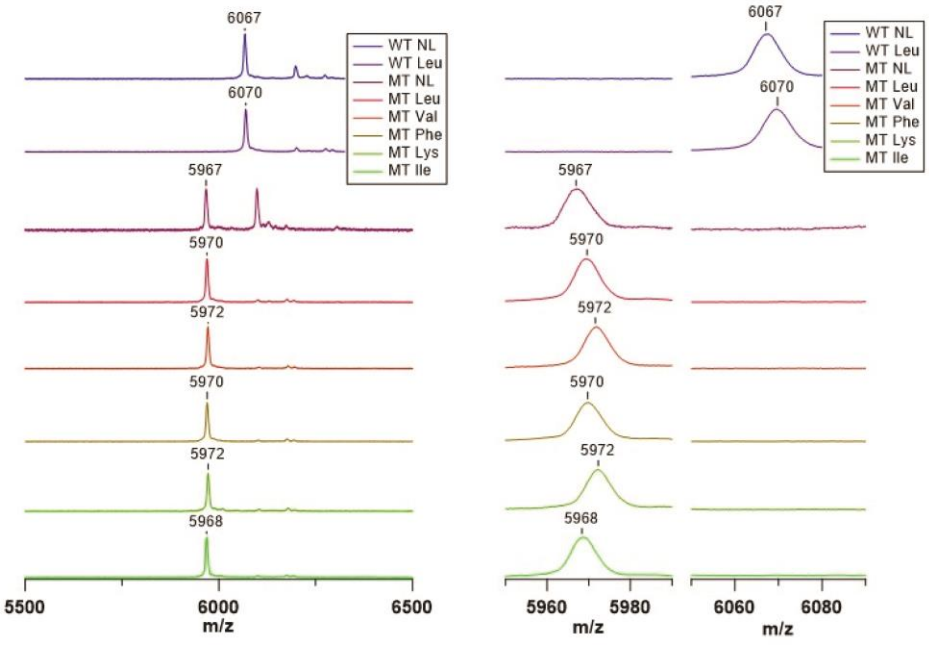

Figure S1. MALDI-TOF-MS spectra of WT GB1 and MT GB1 proteins with and without isotope labels. Right and left figures show the same data while the right figures have a smaller m/z range.

(a)

WT GB1

M	TYK	LILNGK	TLK	GETTTEAVDAATAEK	VFK	QYANDNGVDGEWTYDDATK	TFTVTE
[1]	[2-4]	[5-10]	[11-13]	[14-28]	[29-31]	[32-50]	[51-56]
	411.22	657.43	361.24	1493.70	393.25	2161.89	697.34

MT GB1

M	TYK	LILNGK	TLK	GETTTEAVDAATAEK	VFK	AYAADNGVDGEWTYDDATK	TFTVTE
[1]	[2-4]	[5-10]	[11-13]	[14-28]	[29-31]	[32-50]	[51-56]
	411.22	657.43	361.24	1493.70	393.25	2061.87	697.34

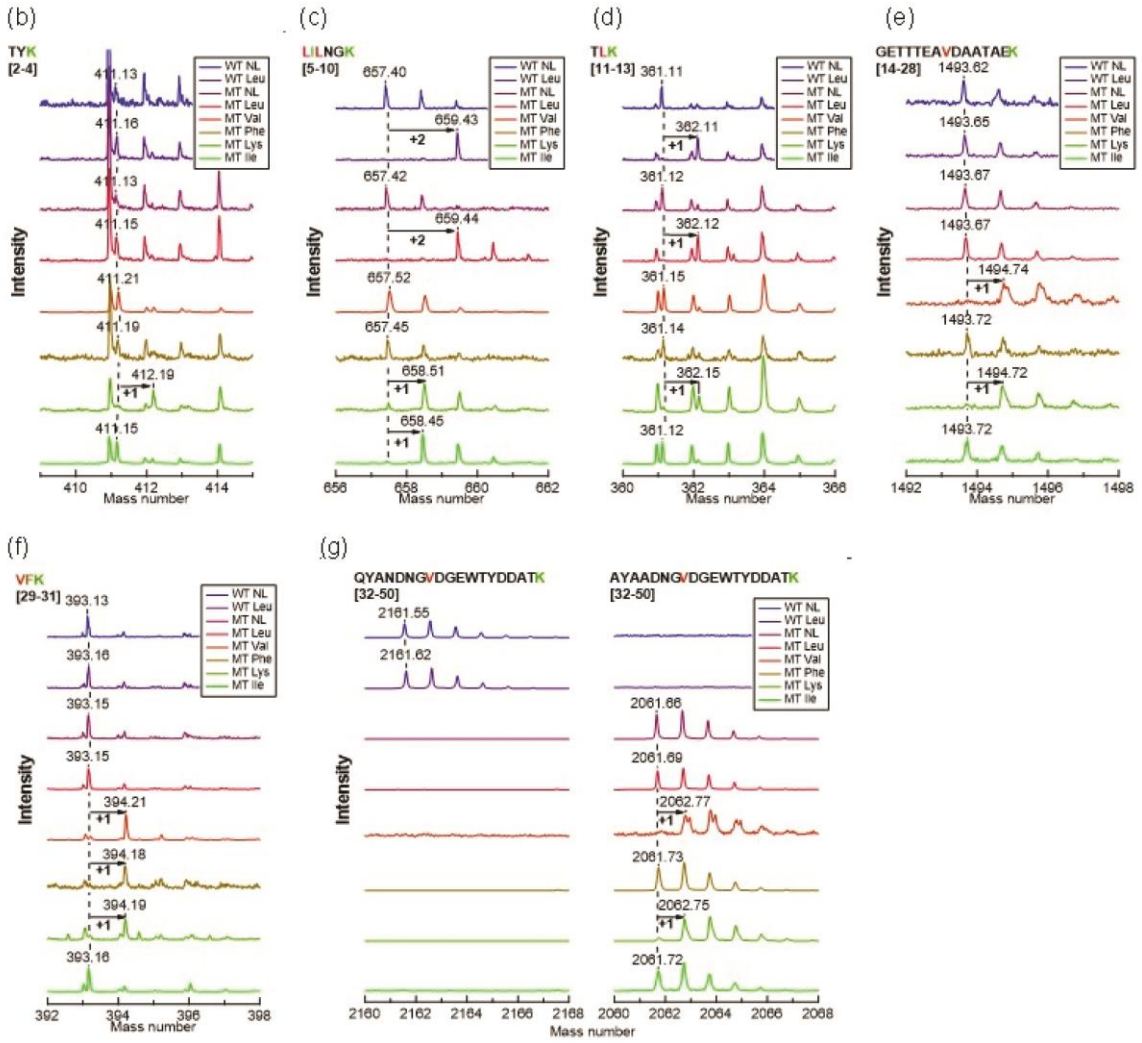

Figure S2. Digested peptide fragments and MALDI-TOF-MS spectra of trypsin-digested WT and MT GB1 proteins. (a) Expected peptide fragments obtained by trypsin digestion of WT and MT GB1. The number inside the bracket represents the amino acid sequence number of the peptide, and the number under the bracket represents the monoisotopic mass of its protonated molecular ion. (b-h) Offset mass spectra of the peptide fragments: (b) [2-4], (c) [5-10], (d) [11-13], (e) [14-28], (f) [29-31], and (g) [32-50] (left: WT GB1, right: MT GB1, (g) only).

Table S1. The mass peak shifts of protein fragments from the monoisotopic mass peaks of WT GB1 and MT GB1 protein fragments.

	Trypsin-digested peptide fragment						
GB1 samples	[2-4]	[5-10]	[11-13]	[14-28]	[29-31]	[32-50]	[51-56]
WT NL	0	0	0	0	0	0	N/D
WT Leu	0	+2	+1	0	0	0	N/D
MT NL	0	0	0	0	0	0	N/D
MT Leu	0	+2	+1	0	0	0	N/D
MT Val	0	0	0	+1	+1	+1	N/D*
MT Phe	0	0	0	0	+1	0	N/D*
MT Lys	+1	+1	+1	+1	+1	+1	N/D
MT Ile	0	+1	0	0	0	0	N/D

*The [51-56] fragment contains one Val and one Phe, but those peak shifts could not be identified due to the peak overlapping.

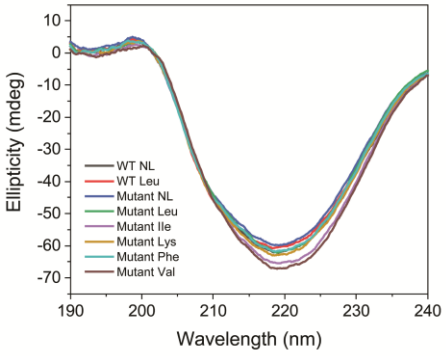

Figure S3. CD spectra of WT GB1 and MT GB1 samples. The spectra are similar for all the samples, indicating minimal impact on the GB1 secondary structures by mutation and isotope

labeling. The CD spectra were collected using J-1500 instrument (Jasco Inc, Japan) with protein solution concentrations of 0.5 mg/mL.

S2. SFG Spectra of the Polystyrene Surface

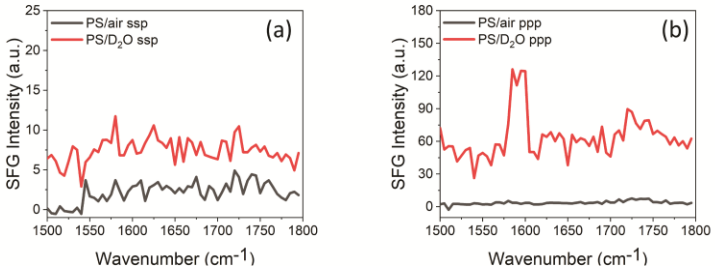

Figure S4. SFG ssp (a) and ppp (b) spectra of polystyrene/air interface (black) and polystyrene/D_2O interface (red). No noticeable SFG signal was detected from the PS/air interface. No SFG signal was detected from the ssp spectrum from the PS/D_2O interface. A peak was observed in the ppp spectrum from the PS/D_2O interface, contributed by the C=C stretching mode. The different intensities of the ssp and ppp signals at 1600 cm^{-1} is related to the PS phenyl orientation. The thickness of PS film is around 150 nm, thus the surface of the film will not be influenced by the substrate which supports the film.[2]

S3. SFG Fitting Parameters Used for Protein Spectral Reconstruction

Experimental SFG spectra were fitted to obtain the parameters using the following equation (SE1), same as the equation (1) in the main text:

$$I_{SFG} \propto \left|\chi_{eff}^{(2)}\right|^2 = \left|\chi_{NR}^{(2)} + \sum_q \frac{A_q}{\omega_{IR} - \omega_q + i\Gamma_q}\right|^2 \quad \text{(SE1)}$$

where $\chi_{eff}^{(2)}$ is the effective second-order susceptibility, and $\chi_{NR}^{(2)}$ is the non-resonant contribution. A_q, ω_q, and Γ_q are the SFG signal amplitude, the vibrational frequency, and the damping coefficient (or peak width) of the vibrational mode q, respectively.

To compare the experimental measurements with the calculated spectra for a protein, it is necessary to only consider the observed SFG signal contributions from the protein while excluding the non-resonant background contributions and the contributions from other molecules. The reconstructed experimental protein SFG spectra were calculated using the following equation (SE2). The protein contributions used to reconstruct the experimental SFG spectra obtained from spectral fitting are listed in Table S2.

$$I_{recon,SFG} \propto \left| \sum_{peak\ i\ from\ a\ protein} \frac{A_i}{\omega_{IR} - \omega_i + i\Gamma_i} \right|^2 \quad \text{(SE2)}$$

Table S2. SFG Fitting Parameters Used for Protein Spectral Reconstruction

	amp1	x1 (cm^{-1})	w1 (cm^{-1})	amp2	x2 (cm^{-1})	w2 (cm^{-1})	amp3	x3 (cm^{-1})	w3 (cm^{-1})
WT NL ssp	51.8	1624.6	21.2	-21.0	1655.0	10.0			
WT NL ppp	60.8	1623.9	13.5	-32.0	1649.0	14.5			
WT Leu ssp	60.0	1633.3	27.1	-38.0	1655.3	15.3			
WT Leu ppp	32.0	1624.6	11.6	-12.0	1656.0	17.3			
MT NL ssp	47.2	1621.4	23.0	-20.0	1651.3	12.1	32.4	1685.1	22.1
MT NL ppp	62.5	1625.7	13.5	-40.0	1675.3	15.2			
MT Leu ssp	68.9	1628.7	22.7	-18.0	1647.8	22.6	31.5	1687.3	29.4
MT Leu ppp	87.9	1627.0	15.1	-30.0	1656.7	17.5			
MT Val ssp	61.4	1626.4	19.5	-30.0	1670.3	13.0			
MT Val ppp	95.2	1627.9	16.4	-45.0	1670.0	16.0			
MT Phe ssp	50.7	1623.3	18.4	-18.0	1655.8	19.0	34.6	1684.9	22.3
MT Phe ppp	110.4	1629.3	17.7	-50.0	1664.3	13.0			
MT Lys ssp	20.1	1619.4	14.6	-15.0	1642.9	16.1	29.8	1676.2	22.8
MT Lys ppp	42.4	1627.7	11.0	-25.0	1664.0	15.6			
MT Ile ssp	50.4	1618.0	21.0	-20.0	1649.0	17.1	35.6	1684.0	24.1

MT Ile ppp	65.2	1627.9	15.1	-25.0	1665.0	17.7			

S4. Snapshots of the Initial Configurations of WT GB1 and MT GB1 on the PS Surface

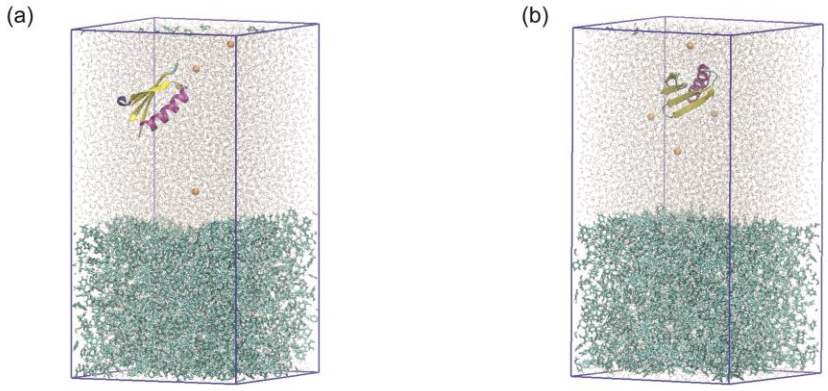

Figure S5. Snapshots of initial configurations of (a) a WT GB1 and (b) a MT GB1 on a PS surface in an aqueous environment with counterions (Na^+). The PS molecules are shown in cyan; counterions are shown in orange and water molecules are shown in gray. The thickness of the simulated PS film is around 5.5 nm. According to the simulation results of PS films[2], the mobile free surface is 2 nm (less than 5.5 nm used in this study), showing that our MD simulation of the PS film can be justified.

S5. Additional Information for SFG Spectral Calculation and Scoring System Used for Spectral Comparison

The calculated SFG response in the laboratory frame can be converted to the Jones frame to compare to the experimentally collected reconstructed data. Formulas (SE3) – (SE9) shown below can convert the SFG response in the laboratory frame (x, y, z) to the Jones frame (s, p) by considering the Fresnel coefficients.

$$\chi^{(2)}_{eff,ssp} = F_{ssp/yyz}\chi^{(2)}_{yyz} \quad \text{(SE3)}$$

$$\chi^{(2)}_{eff,ppp} = F_{ppp/zzz}\chi^{(2)}_{zzz} + F_{ppp/xxz}\chi^{(2)}_{xxz} + F_{ppp/xzx}\chi^{(2)}_{xzx} + F_{ppp/zxx}\chi^{(2)}_{zxx} \quad \text{(SE4)}$$

where $F_{Jones/Lab}$ represents the following:

$$F_{ssp/yyz} = L_{yy}(\omega_{SUM})L_{yy}(\omega_{VIS})L_{zz}(\omega_{IR})\sin\delta_2 \quad \text{(SE5)}$$

$$F_{ppp/zzz} = L_{zz}(\omega_{SUM})L_{zz}(\omega_{VIS})L_{zz}(\omega_{IR})\sin\delta\sin\delta_1\sin\delta_2 \quad \text{(SE6)}$$

$$F_{ppp/xxz} = -L_{xx}(\omega_{SUM})L_{xx}(\omega_{VIS})L_{zz}(\omega_{IR})\cos\delta\cos\delta_1\sin\delta_2 \quad \text{(SE7)}$$

$$F_{ppp/xzx} = -L_{xx}(\omega_{SUM})L_{zz}(\omega_{VIS})L_{xx}(\omega_{IR})\cos\delta\sin\delta_1\cos\delta_2 \quad \text{(SE8)}$$

$$F_{ppp/zxx} = L_{zz}(\omega_{SUM})L_{xx}(\omega_{VIS})L_{xx}(\omega_{IR})\sin\delta\cos\delta_1\cos\delta_2 \quad \text{(SE9)}$$

where L_{ii} (i = x, y, or z) are the Fresnel factors for the local field correction, and δ, δ_1 and δ_2 are the incident angle between the surface normal and the light of the output sum frequency signal beam, the input visible beam and the input IR beam, respectively.

The reconstructed experimental ssp SFG spectra and the calculated ssp SFG spectra were normalized to [0, 1] and meanwhile the reconstructed experimental ppp SFG spectra and the calculated SFG ppp spectra were scaled to maintain the original ppp/ssp intensity ratios (the ratio was determined by the maximum of I_{ppp} divided by the maximum of I_{ssp}).

The comparison (difference or diff.) between the reconstructed experimental ssp (or ppp) SFG spectra and the calculated ssp (or ppp) SFG spectra was done by using a point-to-point least square method, as shown in equation (SE10):

$$diff. = \sum_{i=1}^{i=301} \sqrt[2]{(I_{exp,i} - I_{cal,i})^2} \quad \text{(SE10)}$$

where i means the i-th data point in the range from 1500 cm^{-1} to 1800 cm^{-1} (increment of 1 cm^{-1}, 301 data points in one spectrum). $I_{exp,i}$ and $I_{cal,i}$ are the experimental intensity and calculated SFG intensity at the i-th frequency, respectively.

Then the score of the spectral comparison of a protein with a specific orientation is define with equation (SE11):

$$Score_{protein\ x} = \sqrt{\frac{1}{ssp\ diff.} \times \frac{1}{ppp\ diff.}} \quad \text{(SE11)}$$

The final score of WT GB1 or MT GB1 with a specific orientation can be obtained by the following equation (SE12):

$$Final\ Score\ = \sqrt[n]{\prod_1^n Score_{protein\ x}} \quad \text{(SE12)}$$

Here for WT GB1, n = 2 (for two protein samples). For MT GB1, n = 6 (for 6 protein samples). The final score can be plotted as a function of protein orientation (θ, ψ) for each protein if rotation grid search is applied to show the matching score heat map. The final score can also be plotted as a function of atomistic MD simulation time for proteins without rotations (at their (0°, 0°) orientation position).

S6. Distance between the Protein and the PS Surface as a Function of Time in Simulation

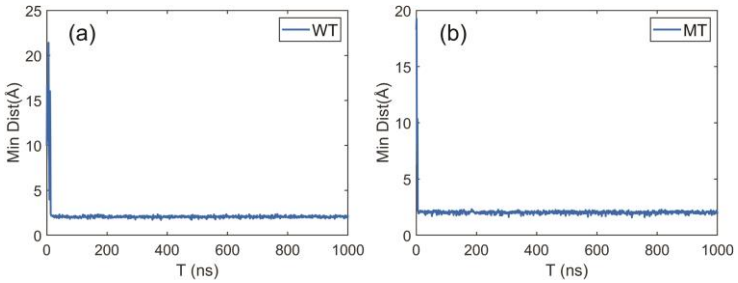

Figure S6. Distance between the protein (a) WT GB1 or (b) MT GB1 and the PS surface as a function of time in simulation. WT GB1 and MT GB1 were released at ~ 2.0 nm from the PS

surface. The adsorption processes happened quickly for both cases (~20 ns for WT GB1 and ~ 10 ns for MT GB1). After landing on the PS surface, the protein-surface distance has almost no change for each case along the entire simulation period.

S7. Radius of Gyration of the Proteins as a Function of Time in Simulation

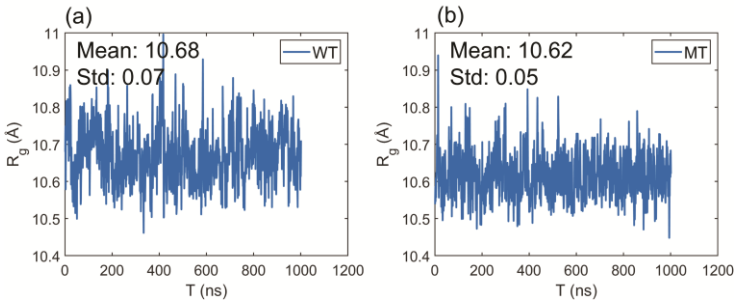

Figure S7. Radius of gyration (R_g) fluctuations of (a) WT GB1 and (b) MT GB1 along the simulation time. The mean R_g value of WT GB1 is similar to the mean R_g value of MT GB1, with slightly larger variations. This indicates that the atom distribution along the principal axis of inertia of WT GB1 is similar to that of MT GB1 along the entire simulation period. This result infers that the mutation of Q32A and N35A has little to no effect on the folding of GB1.

S8. RMSD of the Proteins (compared to the 0 ns) as a Function of Time in Simulation

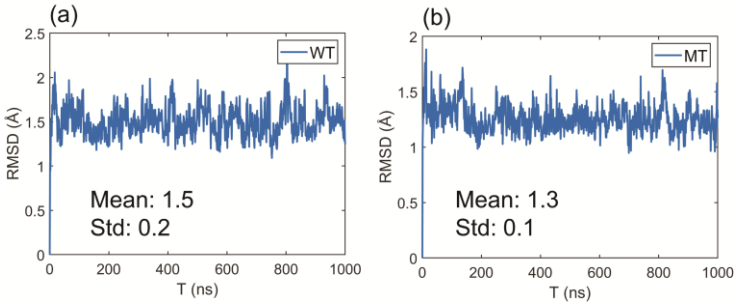

Figure S8. Root mean square deviation (RMSD) between the protein structure at 0 ns and the protein structure from 0 ns to 1000 ns. (a) is the RMSD plot of WT GB1 and (b) is the RMSD plot of MT GB1. WT GB1 and MT GB1 possess similar mean value and variation of RMSDs, indicating that the mutation of Q32A and N35A causes little to no conformation changes due to the GB1 – PS interactions.

S9. Secondary Structure Revolution of Simulated WT GB1 and MT GB1

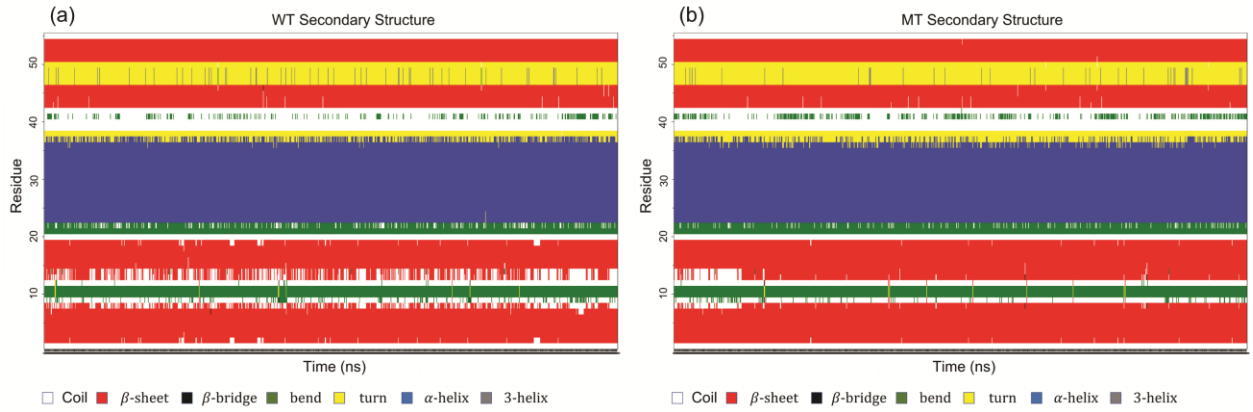

Figure S9. Secondary structure map of (a) WT GB1 and (b) MT GB1 as a function of the simulation time. WT GB1 and MT GB1 have similar conformations in the entire simulation time. The major conformational difference is that for MT GB1, N8, L11 and K13 formed β-sheet structure when stabilized, while for WT GB1 these three residues could sometimes be coil structure.

The only conformational difference is that for MT GB1, N8, L11 and K13 formed β-sheet structure when stabilized while for WT GB1, these three residues could sometimes have coil structure. Detailed analysis (shown in Section S18) revealed that N8, L11 and K13 are located at the contact site between MT GB1 and the PS surface. The interaction between the PS surface and the protein causes the contact site to form a turn structure in the MT GB1. While in the WT GB1, N8, L11 and K13 are not near the surface contact site and thus these three residues could sometimes be random.

S10. θ_a as a Function of Time

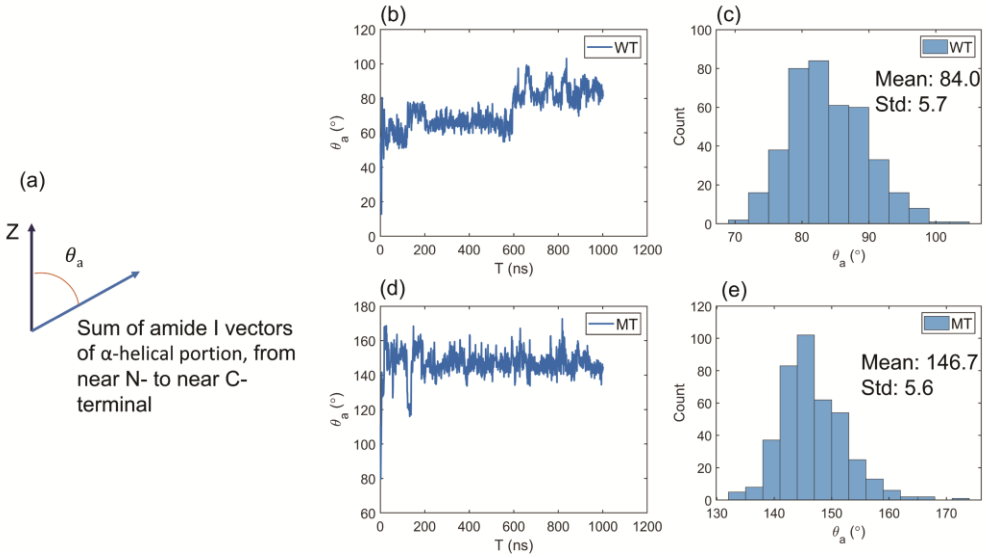

Figure S10. (a) θ_a is defined as the angle between the z axis and the sum of the amide I vectors of α-helix (from residue 21 to residue 36, pointing from near N-terminal to near C-terminal) of protein GB1. (b) and (d) are the changes of θ_a of the simulated WT GB1 and MT GB1, respectively, from 0 ns to 1000 ns with the increment of 1 ns. (c) and (e) are the θ_a distributions of last 400 ns of the simulations of WT GB1 and MT GB1, respectively. WT GB1 781 ns at (30°, 50°) has a θ_a of 90.2°, and at (150°, 230°) has a θ_a of 89.8°. MT GB1 972 ns at (30°, 100°) has a θ_a of 148.2°, and at (150°, 280°) has a θ_a of 31.8°. After 600 ns, θ_a reached equilibrium, and WT GB1 and MT GB1 possess different mean θ_a values with similar standard deviations for the last 400 ns. θ_a could be used to differentiate the best matching pairs with opposite absolute orientations, which could not be separated by homodyne SFG measurements.

S11. Orientation Visualization of WT GB1 781 ns (150°, 230°) and MT GB1 972 ns (150°, 280°)

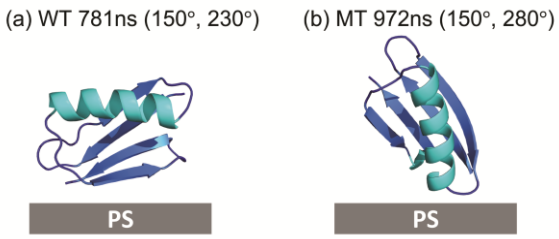

Figure S11. Orientation visualization of (a) WT GB1 781 ns at the orientation of (150°, 230°), having the highest matching score between Hamiltonian calculated WT GB1 SFG spectra and the reconstructed experimental WT GB1 SFG spectra, and (b) MT GB1 972 ns at the orientation of (150°, 280°), having the highest matching score between Hamiltonian calculated MT SFG spectra and the reconstructed experimental MT SFG spectra. These two orientations could not match the atomistic MD simulation results and adopt opposite absolute orientations compared to the simulated orientations. We therefore exclude the possibility that these two orientations are the most likely protein orientations at interfaces.

S12. Structure Overlaps between the Crystal Structure and the Simulated Structures

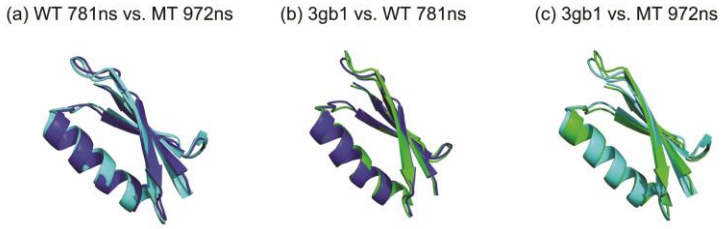

Figure S12. (a) For WT 781 ns (blue) vs. MT 972 ns (cyan), RMSD = 2.0 Å, larger than the conformational fluctuations of either WT GB1 or MT GB1 along the entire simulation time (RMSD mean of WT is 1.5 Å and RMSD mean of MT is 1.3 Å). Therefore, we believe that the conformation difference between WT GB1 and MT GB1 is distinct, not a random fluctuation. (b)

For 3gb1 (green) vs. WT 781 ns (blue), RMSD = 1.7 Å, and (c) for 3gb1 (green) vs. MT 972 ns (cyan), RMSD = 1.6 Å. (b) and (c) show that it is reasonable to use the crystal structure to approximate the interfacial structures for both WT GB1 and MT GB1 because the conformational changes between the crystal structure and the MD simulated structures are small.

S13. Crystal Structure for SFG Spectral Calculation

We have used crystal structure for SFG data analysis previously, which is a valid method if the protein structure is rigid.[3–6] Here again we assumed that the crystal structure could approximate the structures of both WT GB1 and MT GB1 at the PS/protein solution interfaces, and the calculated spectra were compared with the reconstructed experimentally collected spectra using the same strategy as described above. We believe that this assumption is reasonable because the crystal structure is similar to the simulated WT GB1 structure at 781 ns (RMSD = 1.7 Å) and the simulated MT GB1 structure at 972 ns (RMSD = 1.6 Å), shown in the Section S12.

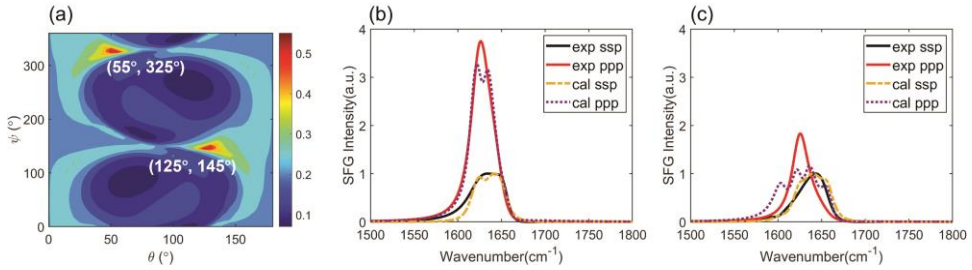

Figure S13. (a) Final score heat map of spectral matching between the reconstructed WT SFG spectra and the calculated WT SFG spectra by rotating the crystal structure 3gb1. The orientations at (55°, 325°) and (125°, 145°) possess highest matching scores (score = 0.56). The spectral comparisons between the reconstructed experimental spectra and the calculated spectra using 3gb1 at orientation of (55°, 325°) (or (125°, 145°)) of (b) WT NL and (c) WT Leu. By using the crystal structure, the best matched calculated ppp spectra could not match the reconstructed experimental ppp spectra well, leading to lower matching scores.

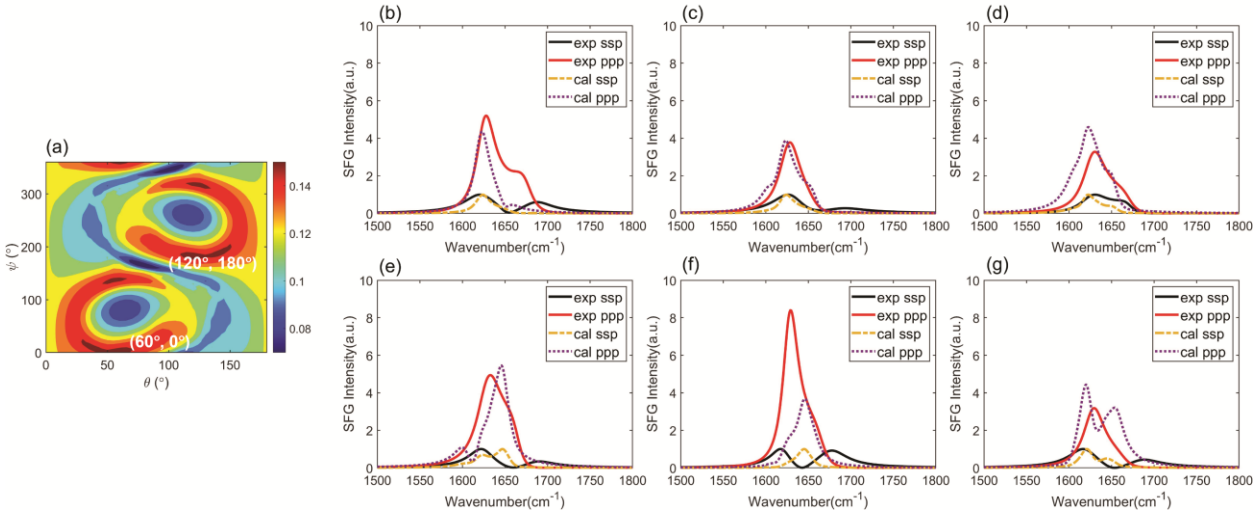

Figure S14. (a) Final score heat map between the reconstructed experimental MT SFG spectra and the calculated MT SFG spectra using crystal structure 3gb1. The orientations at (60°, 0°) and (120°, 180°) possess highest matching scores (score = 0.15). The spectral comparisons between reconstructed experimental spectra and calculated spectra by rotating 3gb1 at orientation of (60°, 0°) (or (120°, 180°)) of the case of (b) MT NL, (c) MT Leu, (d) MT Val, (e) MT Phe, (f) MT Lys and (g) MT Ile. By using the crystal structure, the calculated spectra and the experimental reconstructed spectra could not match well. For example, almost all the ppp spectra except MT Leu could not match well in peak intensity, peak center, etc., leading to lower matching scores.

The final score heat maps obtained based on the crystal structures of WT GB1 and MT GB1, along with the spectral comparisons of the top-ranking orientations are shown in Figure S13 and Figure S14 respectively. For WT GB1, the best matched orientations are (55°, 325°) (visualized in Figure 6(a)) and (125°, 145°) (visualized in Figure 6(b)) with a score of 0.56. This score (obtained via the crystal structure) is lower than the best matched score (0.66) obtained by using the rotated simulated WT GB1 structures. This indicates that the atomistic MD simulation could effectively modify the protein structures at the interface/surface by taking the protein-surface

interaction into consideration. This results in better matching of the experimental data and interpretation of the experimental measurements. The visualized crystal structure 3gb1 at the orientation of (55°, 325°), although not very different from that of the simulated WT GB1 structure at 781 ns of (30°, 50°, Figure 6(c)), had the α-helical chain slightly tilting up, making the planar residues within this chain not closely in contact with the surface. This could impair the protein-PS interactions and thus is not reasonable. Clearly, the results obtained from the crystal structure input for WT GB1 are less rational compared to those from the MD simulated structure inputs.

For MT GB1, the best matched orientations using the crystal structure input are (60°, 0°) (visualized in Figure 6(d)) and (120°, 180°) (visualized in Figure 6(e)) with a score of 0.15. This score is lower than that obtained by using the rotated simulated MT GB1 structures (with the highest score of 0.19). The visualized most likely MT GB1 orientation of (120°, 180°) deduced by using the crystal structure 3gb1 is similar to the most likely orientation of the simulated MT GB1 structure (Figure 6(f)). However, the α-helical chain is totally above the bottom of the β-sheet portion, hindering it from contacting the PS surface. The other best matched orientation based on the crystal structure, (60°, 0°), has the planar residues on the helix far from the PS surface, which is highly unlikely. Thus, it can be concluded that here although using the crystal structure for spectral calculations could differentiate the orientations between WT GB1 and MT GB1 on the PS surface, the accuracy may not be as high as that obtained from the results based on the use of the simulated GB1 structures. As mentioned above, we believe that the atomistic MD simulation could capture the protein structure deviation from the crystal structure when considering the protein – surface interaction, improving the accuracy of the SFG data analysis. In addition, atomistic MD simulation results can be used to further validate SFG data analysis.

S14. Spectral Comparisons of WT GB1 at 821 ns

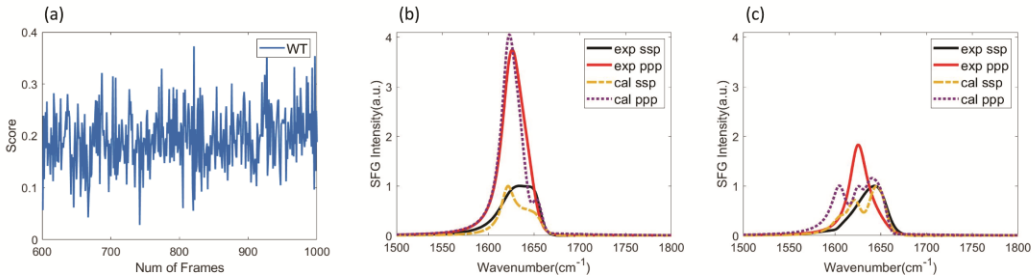

Figure S15. (a) Score plot of WT GB1 configurations (last 400 ns). Spectra of the configuration with the highest score (WT GB1 821 ns, score = 0.37) for the case of (b) WT NL and (c) WT Leu. The matching quality using MD configurations without rotation is lower than using rotated MD configurations (score = 0.66).

S15. Spectral Comparisons of MT GB1 at 887 ns

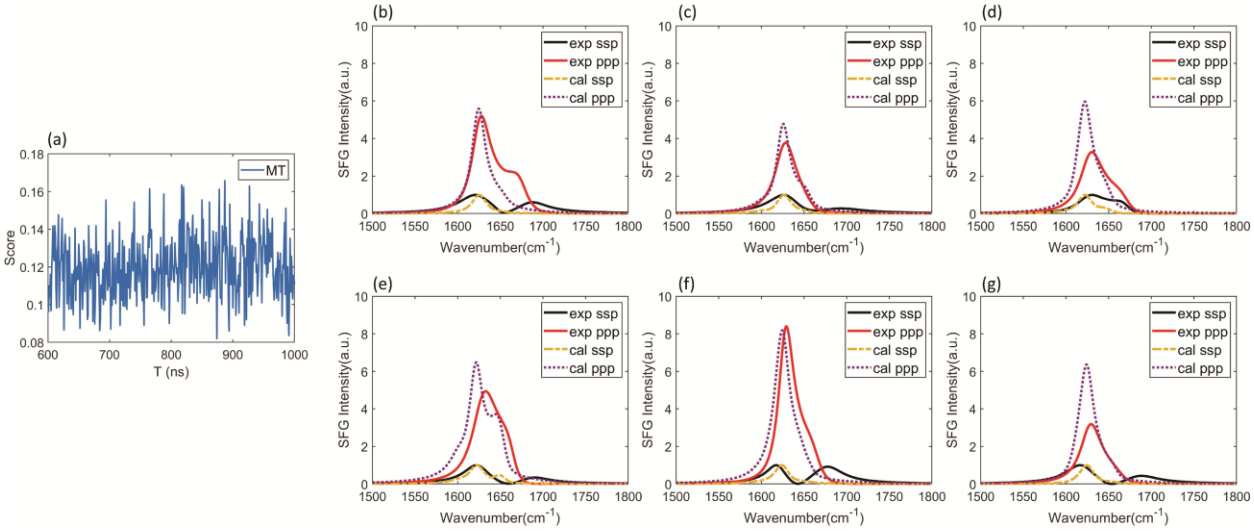

Figure S16. (a) Score plot of MT GB1 configurations (last 400 ns). Spectra of the configuration with the highest score (MT GB1 887 ns, score = 0.17) for the case of (b) MT NL, (c) MT Leu, (d) MT Val, (e) MT Phe, (f) MT Lys and (g) MT Ile. The matching quality using the MD simulated structures without rotation is lower than using rotated MD simulated structures (score = 0.19).

S16. Top Ranked Configurations of Non-Rotated WT GB1 and Their Orientation Visualizations

(a)

Configuration (ns)	Score
927	0.37
997	0.35
967	0.33
993	0.33

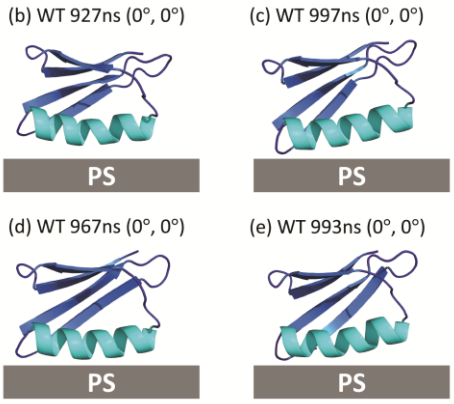

Figure S17. (a) Table of top scores of non-rotated WT GB1. Orientation visualizations of (b) WT GB1 927 ns, (c) WT GB1 997 ns, (d) WT GB1 967 ns and (e) WT GB1 993 ns. These four orientations all adopt a helix lying-down orientation and are similar to WT GB1 821 ns (highest ranked non-rotated WT GB1) and WT GB1 781 ns at the orientation of (30°, 50°), which is the best matched orientation among all rotated simulated structures. Thus, the conclusion that WT GB1 adopts a lying-down orientation on the PS surface is reliable.

S17. Top Ranked Configurations of Non-Rotated MT GB1 and Their Orientation Visualizations

(a)

Configuration (ns)	Score
817	0.16
927	0.16
820	0.16
765	0.16

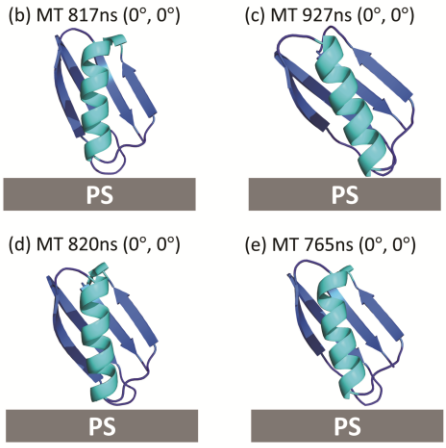

Figure S18. (a) Table of top scores of non-rotated MT GB1. Orientation visualizations of (b) MT GB1 817 ns, (c) MT GB1 927 ns, (d) MT GB1 820 ns and (e) MT GB1 765 ns. These four orientations all adopt a stand-up orientation and are similar to MT GB1 887 ns (highest ranked non-rotated MT GB1) and MT GB1 972 ns at the orientation of (30°, 100°), which is the best matched orientation among all rotated MD simulated structures. Thus, the conclusion that MT GB1 adopts a stand-up orientation on the PS surface is reliable.

S18. Residual RMSD

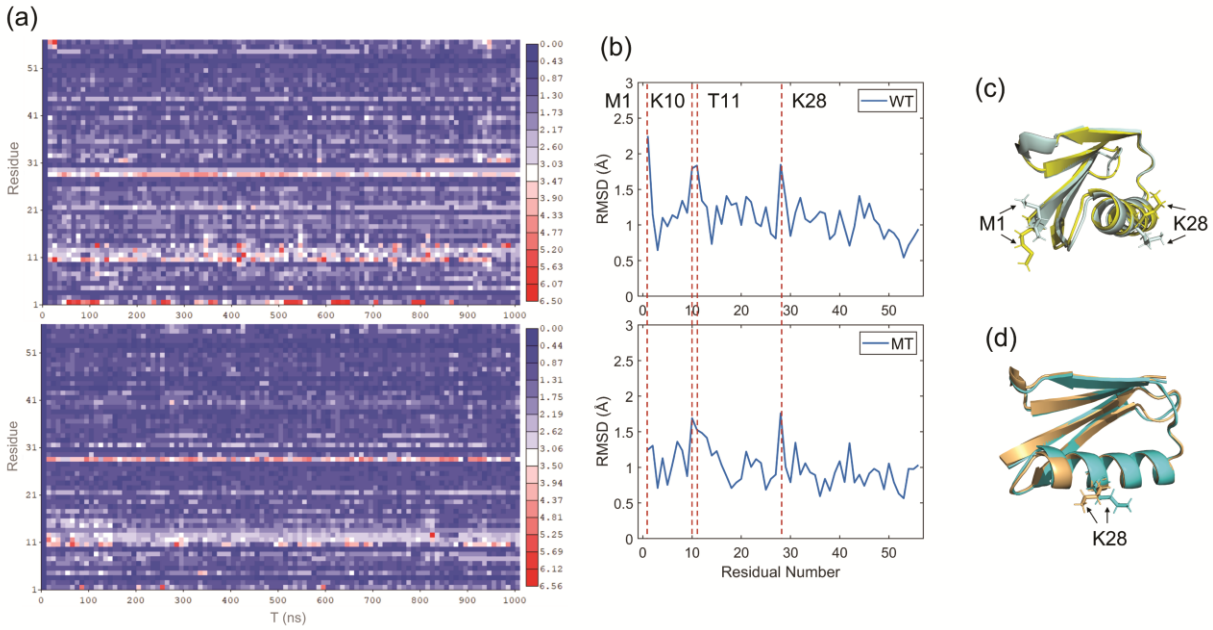

Figure S19. (a) Residual RMSD heat maps of WT GB1 (top) and MT GB1 (bottom) between the simulated protein structure at 0 ns and the simulated protein structure from 0 ns to 1000 ns with a step of 10 ns. (b) Residual RMSD plots of WT GB1 (top) and MT GB1 (bottom) between the simulated protein structure at 0 ns and the best matched simulated protein structure (781 ns for WT GB1 and 972 ns for MT GB1). (c) Superimposed structure of the WT GB1 0 ns (palecyan) and 781 ns (yellow). Residue M1 and K28 were marked with sticks. (d) Superimposed structure of the MT GB1 0 ns (cyan) and 972 ns (orange). Residue K28 was marked with sticks. It is

worth noting that the orientations of the simulated structures at 781 ns and 972 ns shown in (c) and (d) are not the matched orientations from the spectral comparisons. The heat maps show that obvious fluctuations occurred at M1, K10, T11 and K28 for WT GB1 and at K10, T11 and K28 for MT GB1 along the simulation.

To further understand the structural difference between the stabilized GB1 structure in the bulk environment (0 ns in MD simulation) and the simulated GB1 structures on the PS surface which have the best matching scores, residual RMSDs were calculated by considering the protein's heavy backbone excluding the hydrogen atoms. The Residual RMSD heat maps (Figure S19(a)) show that compared to the stabilized structure in solution (0 ns in MD simulation), the bend region near the N-terminus (K10 and T11) for both WT GB1 and MT GB1 displays above-average fluctuations because of the flexibility of this region. For the WT GB1, additional fluctuations also occur at M1 and K28. When visualized, it is clear that with a lying-down orientation (shown in Figure 7(a)) of WT GB1 on the PS surface, the hydrophobic side chain of M1 tilts towards the hydrophobic PS surface while the charged K28 side chain tilts away from the PS surface and interacts with the solvent molecules (water). The favorable interactions between M1 and the PS surface and between K28 and water lead to more distinct structural differences at these two sites compared to the stabilized WT GB1 in solution, exhibiting large RMSD values. For MT GB1, K28 also displays a high RMSD value. However, different from WT GB1, K28 in MT GB1 is not located at the near-surface region for the best matched orientation (shown in Figure 7(b)). If the MT GB1 lies down on the surface, similar to WT GB1 (Figure S19(c)), the charged residue K28 would point towards the surface, which is less reasonable because the charged residues interact with the solvent more favorably. Thus, it is more reasonable for MT GB1 to adopt a stand-up pose.

S19. Visualization of Contact Areas of WT GB1 and MT GB1 on PS

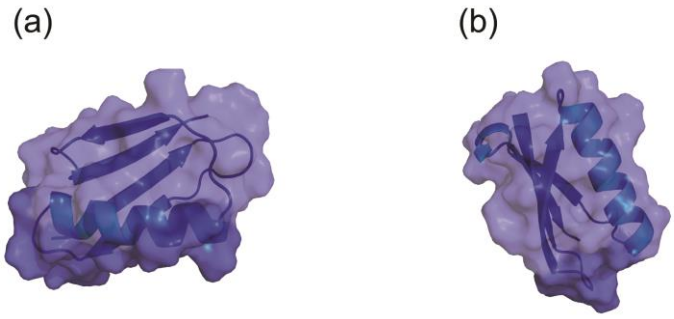

Figure S20. Visualization of Contact Areas (in dark blue) of (a) WT GB1 and (b) MT GB1 on PS. The contact area of WT GB1 on PS is obviously larger than the contact area of MT GB1 on PS.

S20. Movie Showing the Deduced Structures of Wild-type Protein GB1 with High Matching Scores between Experimental and Calculated Data

This movie is uploaded as a separate file, which shows the deduced structures of wild-type protein GB1 with high matching scores (with scores between 0.59 and 0.66) between experimental and calculated data.

S21. Movie Showing the Deduced Structures of Mutant GB1 with High Matching Scores between Experimental and Calculated Data

This movie is uploaded as a separate file, which shows the deduced structures of mutant protein GB1 with high matching scores (with scores between 0.17 and 0.19) between experimental and calculated data.

S22. References

1 T. Sugiki, K. Furuita, T. Fujiwara and C. Kojima, *Biochemistry*, 2018, **57**, 3576–3589.
2 B. Li, S. Zhang, J. S. Andre and Z. Chen, *Progress in Polymer Science*, 2021, **120**, 101431.
3 T. Lu, W. Guo, P. M. Datar, Y. Xin, E. N. G. Marsh and Z. Chen, *Chem. Sci.*, 2022, **13**, 975–984.
4 A. P. Boughton, P. Yang, V. M. Tesmer, B. Ding, J. J. G. Tesmer and Z. Chen, *Proc. Natl. Acad. Sci.*, **13**, 975–984.

5 M. Xiao, S. Wei, J. Chen, J. Tian, C. L. Brooks III, E. N. G. Marsh and Z. Chen, *J. Am. Chem. Soc.*, 2019, **141**, 9980–9988.
6 X. Zou, S. Wei, S. Badieyan, M. Schroeder, J. Jasensky, C. L. Brooks, E. N. G. Marsh and Z. Chen, *J. Am. Chem. Soc.*, 2018, **140**, 16560–16569.