BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Experimental Identification and Theoretical Analysis of a Thermally Stabilized Green Fluorescent Protein Variant

Saori Akiyama,[†] Atsushi Suenaga,[‡,§] Takayuki Kobayashi,[†] Tetsuya Kamioka,[†] Makoto Taiji,[‡] and Yutaka Kuroda*[,†]

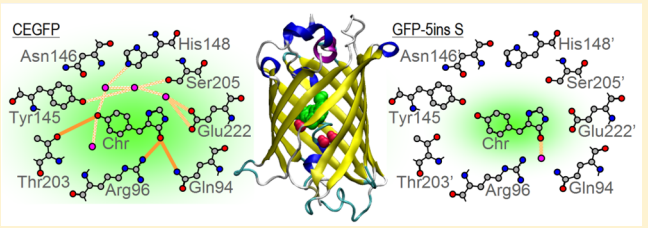
[†]Department of Biotechnology and Life Sciences, Graduate School of Engineering, Tokyo University of Agriculture and Technology, Koganei-shi, Nakamachi, Tokyo 184-8588, Japan

[‡]Computational Biology Research Core, Quantitative Biology Center, RIKEN.R303, 7-1-26 minatojima-Minami-machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan

[§]Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology (AIST), 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan

Supporting Information

ABSTRACT: In this study, we aim to relate experimentally measured macroscopic properties to dynamic and structural changes as calculated by molecular dynamics (MD) simulations. We performed the analysis on four GFP (green fluorescent protein) variants, which have amino acid replacements or insertion in a flexible region on the protein surface and which resulted from a previous protein splicing reaction optimization experiment. The variants are a reference GFP (CEGFP), GFP-N144C, GFP-N144C/Y145F, and a GFP with five residues inserted between Y145 and N146 (GFP-5ins). As a result, we identified a single Y145F mutation that increased the thermal stability of GFP-N144C/Y145F by 3−4 °C. Because circular dichroism measurements indicated that the overall GFP β-barrel fold was maintained in all variants, we presumed that the fluorescence activity and thermal stability related to local changes that could be detected by standard MD simulations. The 60 ns MD simulations indicated that the Y145's hydroxyl group, which is straight and buried in the crystal structure, was bent avoiding the hydrophobic core during the simulation in both CEGFP and GFP-N144C. This local strain was relieved in GFP-N144C/Y145F, where the tyrosine's hydroxyl group was replaced with the F145 hydrophobic aliphatic carbon. F145 remained indeed buried during the simulation maintaining local compactness, which presumably reflected the improved thermal stability of GFP-N144C/Y145F. Furthermore, the analysis of internal water molecules localized within the GFP's β-barrel suggested that a change in the local hydrogen bonding pattern around the chromophore correlated with a strong fluorescence activity decrease in GFP-5ins. Although relating experimental observation with calculated molecular features proved to be delicate, this study suggested that some microscopic features could be useful reporters for redesigning GFPs and other proteins. The newly identified GFP-N144C/Y145F was among the most stable GFP variant and demonstrates the potential of such computer-aided design.



Green fluorescent protein (GFP) from jellyfish *Aequorea victoria* is a fluorescent protein first identified in 1962[1] whose gene was isolated in 1992.[2] The self-catalytic maturation of its chromophore and its ability to emit light without an additive represent a great advantage for use as a biomarker in biomedical research, and various mutants with altered fluorescent properties have been developed. Examples of such mutants are a blue fluorescent protein (BFP),[3,4] a yellow fluorescent protein (YFP),[5,6] a cyan fluorescent protein (CFP),[3] cycle-3, which has enhanced stability and can mature even at 37 °C,[7] EGFP, which is excited by cyan light,[8] Sapphire, which is excited by violet light,[9] and Venus, which has efficient maturation efficiency and is tolerant to an acidic environment.[10]

Several studies have examined the behaviors of GFP variants using molecular dynamics (MD) simulation, and structural insights about excitation and emission wavelength shifts have been reported. To date, the free energy differences between two chromophore's structural states calculated by molecular simulation for both EGFP and EGFP-T203Y agreed well with experimental values measured by absorption spectroscopy.[11] In addition, they demonstrated that the population of the chromophore's state was reversed with the S65T mutation and rationalized the population reversal by a change in the hydrogen bond network around the chromophore. Furthermore, the relationship between the molecular structure and the red shifts of the absorption maximum of S65G and S65T GFP was examined by MD simulation and quantum mechanics/molecular mechanics calculations.[12] Their analysis identified minute structural changes that may cause the red shift of the

Received: May 3, 2012
Revised: September 1, 2012
Published: September 10, 2012

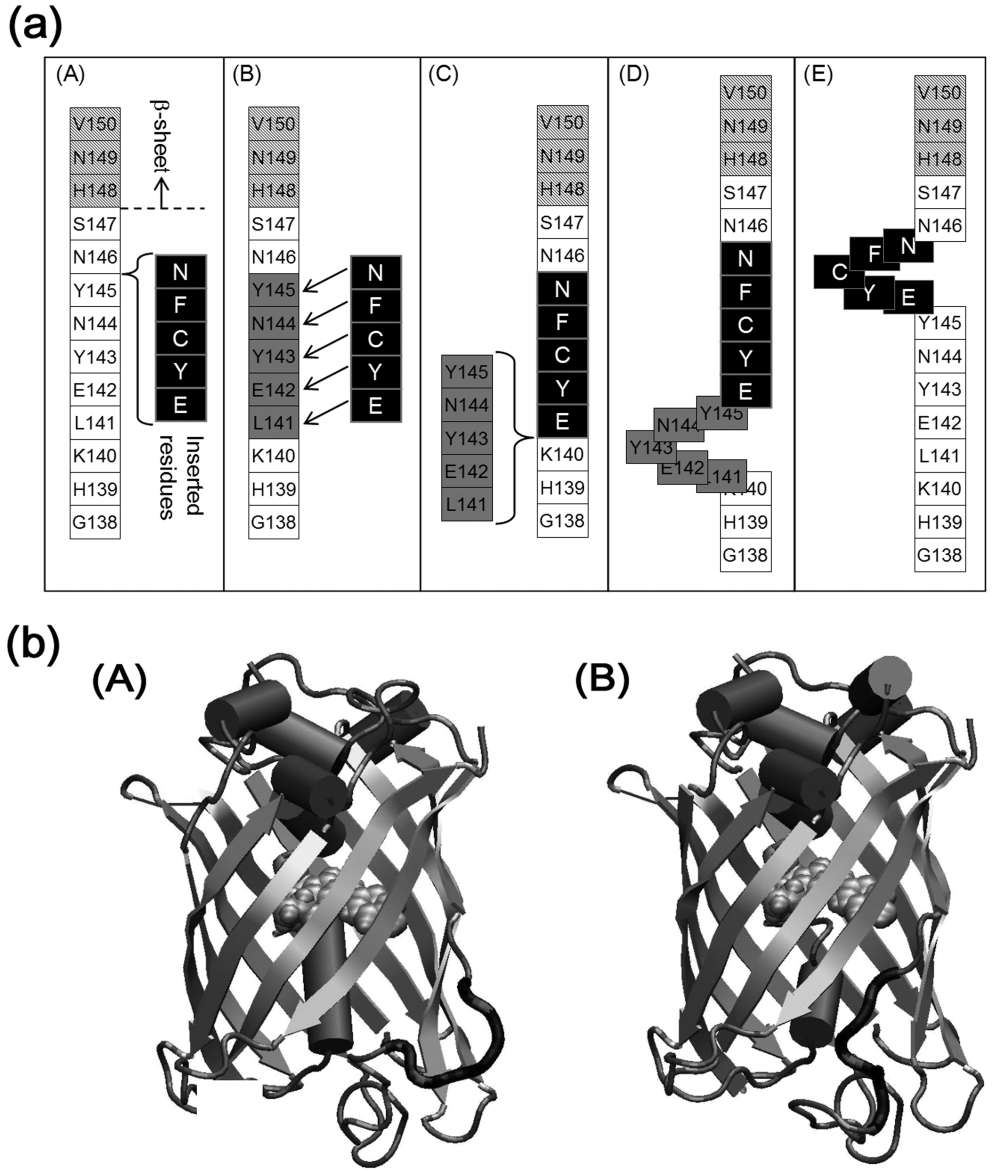


Figure 1. (a) Schematic representation of GFP-5ins building. Numbers in each box represent the residue numbers. The five inserted residues (EYCFN) are shown by black boxes. (b) Ribbon models of GFP-5ins S (A) and GFP-5ins K (B). The chromophore is represented with a van der Waals model. The five inserted residues are shown with a thick black line.

absorption maximum and revealed that the hydrogen bonding around the chromophore differed between S65G and S65T.

Relating a protein's macroscopic properties with changes in molecular structural and dynamic features remains challenging. For example, we recently analyzed how single mutations in bovine pancreatic trypsin inhibitor could result in >10 °C thermal stabilizations. Although the stabilized variants' crystal structures were determined at very high resolutions and we found that an alanine to glycine mutation in a surface-exposed loop region removed potential steric clashes, structural information alone could not explain how an alanine to valine mutation improved stabilization, which occurred probably through an increase in entropy.[13]

Sophisticated methods are effective for analyzing thermodynamic properties and large conformational changes caused by a single amino acid mutation, but they require extensive computational resources.[14] Here, we aim to relate changes in macroscopic biochemical properties produced by a single mutation or a few mutations to dynamic and structural changes as calculated by standard molecular dynamics simulations. Molecular dynamics simulation of the GFPs suggested that a change in the local hydrogen bonding pattern around the chromophore correlated with the decreased fluorescence activity in a GFP with five residues inserted between Y145 and N146 (GFP-5ins). Further, a Y145F mutation that increased the GFP thermal stability by >3 °C correlated with the replacement of the tyrosine hydroxyl group with a hydrophobic aliphatic carbon that improved local compactness by keeping residue 145 buried. This study indicates that some local microscopic molecular features seem to correlate well with experimentally observed changes, and monitoring these features could prove to be efficient for computer-aided screening of large sets of mutants (e.g., ref 15).

dx.doi.org/10.1021/bi300580j | *Biochemistry* 2012, 51, 7974−7982

MATERIALS AND METHODS

Protein Expression and Purification. The gene encoding GFPs was amplified by polymerase chain reaction and inserted into pET21c using the NdeI/BamHI site, as previously reported.[16] The plasmids were transformed in *Escherichia coli* strain JM109(DE3) pLysS, and the cells were grown in LB medium. The expression was induced by adding IPTG at a final concentration of 1 mM when the OD_{590} reached 0.6. EGFP, GFP-N144C, and GFP-N144C/Y145F were expressed for 4 h at 37 °C, whereas GFP-5ins was expressed for 13 h at 25 °C. *E. coli* cells were harvested by centrifugation and lysed by sonication in 50 mM Tris-HCl buffer with 150 mM NaCl (pH 8.7). The lysate was centrifuged at 8000*g* for 20 min, and the precipitate was again sonicated in the same buffer. The supernatants were applied to Ni-NTA resin (QIAGEN), and the proteins were eluted with elution buffer [50 mM NaH_2PO_4, 300 mM NaCl, and 250 mM imidazole (pH 8.0)], which was then removed by dialysis. The protein's concentration was calculated on the basis of its molar extinction coefficient at 280 nm.

Fluorescence Measurements. Fluorescence was measured at 20 °C in 5 mM phosphate buffer (pH 7.0). The emission spectra were recorded from 440 to 640 nm (3 nm slit) with an excitation wavelength of 493 nm (1 nm slit) using a JASCO FP-6500 fluorescence spectrometer. The baselines of the buffer alone were subtracted, and the fluorescence intensities were adjusted to a EGFP protein concentration of 500 nM.

Circular Dichroism Measurements. Circular dichroism (CD) spectra were recorded at 20 °C in 5 mM phosphate buffers (pH 7.0) using a JASCO J-720 circular dichroism spectropolarimeter with a 1 mm path length quartz cell. The spectra were averaged over three scans and smoothed. The baselines of the buffer alone were subtracted, and the CD values were converted to residual ellipticity using the protein concentrations determined above. The secondary structure contents were estimated by using a k2d algorithm.[17]

Thermal denaturation curves were monitored in the temperature range of 25−95 °C by the CD signal at 205 nm (2 nm slit) using 1 cm path length quartz cells at pH 7.0 (5 mM phosphate buffer) and at pH 4.6 (5 mM acetate buffer). The melting temperatures (T_m) were calculated by least-squares fitting the experimental data with a two-state model using Origin version 6.1J.

Construction of Initial Structures. We used the coordinates of the cycle-3 EGFP [Protein Data Bank (PDB) entry 2awk][18] as a template for building the structure of the four GFP variants. The model structures of CEGPF, GFP-N144C, and GFP-N144C/Y145F were built by erasing the side chain coordinates and manually editing the residue name in the PDB files and growing replaced side chains using the leap module in AMBER. Replaced residues versus the template were as follows: CEGFP (Q79R/M95R), GFP-N144C (Q79R/M95R/N140C), and GFP-N144C/Y145F (Q79R/M95R/N140C/Y141F). For GFP-5ins, two model structures were built. In the first one (S model), five residues were inserted by employing the segment match method,[19] and energy minimization was performed in vacuum with all atom positions fixed except for the inserted segment (five residues) by using the modeling software MOE (Chemical Computing Group Inc., Montreal, QC) [GFP-5ins S (Figure 1a, E)]. In the second one (K model), because the *B* factor of the region where the five residues were inserted was not specially low (Figure S3 of the Supporting Information), we kept the coordinates of the original residues and replaced the side chains with those of the five inserted amino acids (Figure 1a, A and B). The five displaced residues were inserted at other places where there was a high *B* factor [GFP-5ins K (Figure 1a, C and D)]. After that, the energy of the five inserted residues was minimized in vacuum.

MD Simulation Protocol. Prior to MD simulation, seven sodium ions were added to the systems for charge neutralization. The system was immersed in a water box containing TIP3P water molecules.[20] The box dimensions (80 Å × 90 Å × 90 Å) were chosen such that the minimal distance of any solute atom from the wall of the water box was 15 Å. The total number of atoms for all systems was approximately 52000. The time step was set to 2.0 fs.

A series of MD simulations were conducted using Amber 8.0[21] on a personal computer (Xeon 3.2 GHz) equipped with special-purpose computer boards for MD simulations, MDGRAPE-3.[14,22,23] The all-atom point-charge force field, ff03, was chosen to represent the protein,[24] and we used the parameters developed by Reuter et al.[25] for the chromophore. All bond lengths were constrained to equilibrium lengths using the SHAKE method,[26] and the long-range Coulomb interactions were treated with the particle-mesh Ewald (PME) method.[27] The real-space component of the PME method was calculated using MDGRAPE-3, while the wave-number-space component and the bonded interactions were calculated by the host computer. To optimize the balance between the calculation times for these components, a cutoff distance of 14 Å was used for the real-space component. After a 5000-step energy minimization had been performed, all systems were gradually heated to 300 K during the first 50 ps, at a heating rate of 6 K/ps. Subsequently, the temperature and pressure were maintained at constant values of 300 K and 1 atm, respectively, with a coupling constant of 1.0 ps.[28] Simulations for each model were repeated three times, starting with the same initial coordinates but with different initial random velocities. Each simulation was performed for 20 ns, and the coordinates were saved every 10 ps. In total, 6000 snapshots were saved, and their main chain structures were categorized by principal component analysis for selecting average structures; a sampling of 1000 snapshots was used in the following analysis.

Analyses of MD Trajectories. The root-mean-square deviation (rmsd), β factor, distance, angle, and positional displacement between main chains of average structures were calculated using the ptraj module of AMBER. The solvent accessibilities of residues were calculated with DSSP,[29] and the secondary structures were assigned using the SECSTR module of PROCHECK.[30] For hydrogen bond analysis, we classified atom pairs as making a hydrogen bond, when the distance between the acceptor and the donor atoms was <3.5 Å and the angle between a hydrogen atom and the donor or acceptor mediated by the acceptor or donor was <40°.

RESULTS AND DISCUSSION

We analyzed relatively short 60 ns molecular dynamic simulations of GFP's variants to identify microscopic features that would correlate with experimentally measured macroscopic properties. We performed the analysis on four GFP variants that resulted from a previous experiment aiming at improving a protein splicing reaction,[16] a post-translational modification.

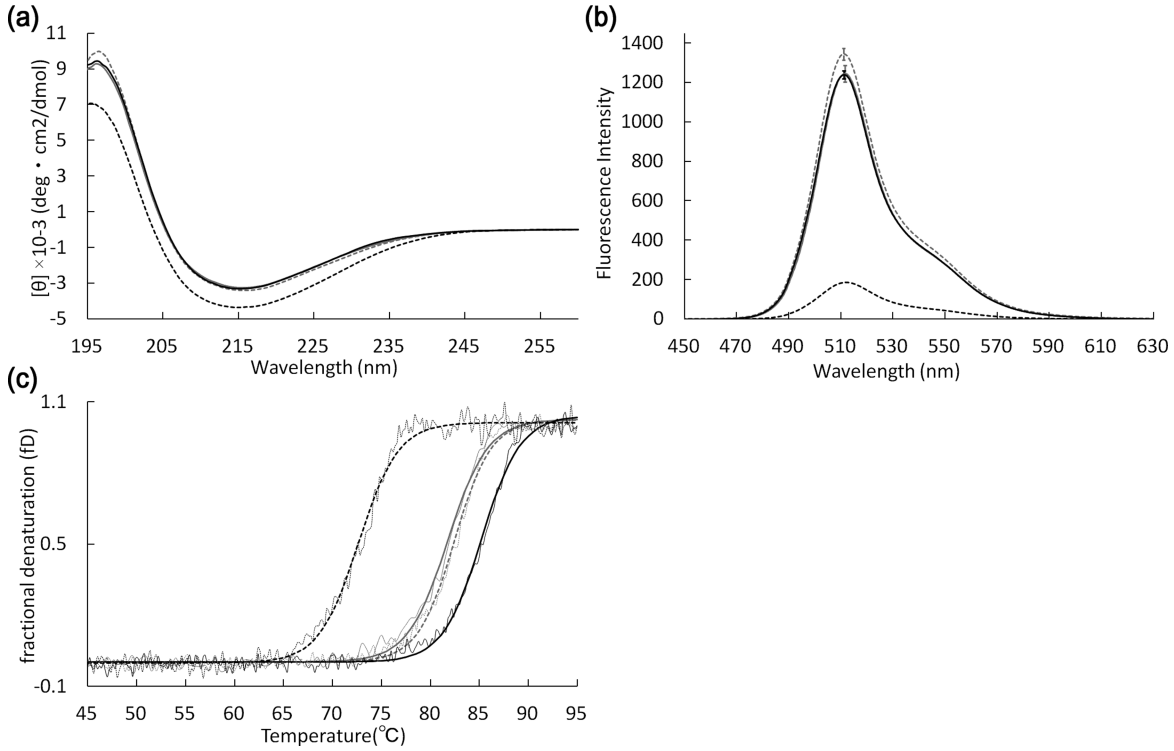


Figure 2. Experimental characterization of GFP variants. Data for CEGFP are shown with a dashed gray line, data for GFP-N144C with a solid gray line, and for GFP-N144C/Y145F and GFP-5ins with dashed and solid black lines, respectively. (a) CD spectra of GFP variants. The spectra were measured in 5 mM phosphate buffer (pH 7.0) at 20 °C, with 1 mm path length quartz cells. (b) Fluorescence spectra of GFP variants. The measurements were taken in 5 mM phosphate buffer (pH 7.0) at a protein concentration of 500 nM at 20 °C, and with an excitation wavelength of 493 nm. (c) Thermal denaturation curves of GFP variants monitored by the CD signal at 205 nm. The thin lines show the experimental raw data, and the fine line shows the fitted curve. The measurements were performed with a 1 cm path length quartz cell at pH 7.0 in 5 mM phosphate buffer.

Table 1. Experimental Results[a]

	CEGFP	GFP-N144C	GFP-N144C/Y145F	GFP-5ins
fluorescence activity (excitation at 493 nm)				
maximal intensity	1343 ± 30	1245 ± 42	1239 ± 20	186 ± 1
maximal wavelength (nm)	511.1	511.7	511.2	512.2
thermal stability T_m (°C)				
pH 7.0	82.4	81.7	85.1	72.6
pH 4.6	61.9	61.9	66.4	55.9
secondary structure content (%)				
α/β/coil	6/52/43	6/52/43	6/52/43	10/47/43
secondary structure content (residues)				
α/β/coil	12/117/98	12/117/98	12/117/98	23/109/100
residual ratio of fluorescence intensity at 80 °C and pH 7.0 (%)				
0 min	100.0	100.0	100.0	100.0
10 min	20.1	14.0	26.3	0.0
20 min	10.0	5.6	19.8	0.0
30 min	5.0	2.6	15.2	0.0

[a]The fluorescence activity was measured in 5 mM phosphate buffer (pH 7.0) at a protein concentration of 500 nM with an excitation wavelength of 493 nm. The melting temperature was calculated by fitting the thermal denaturation curves monitored by the CD signal at 205 nm. The measurements were taken with 1 cm path length quartz cells at pH 7.0 (5 mM phosphate buffer) and pH 4.6 (5 mM acetate buffer). Secondary structure contents were estimated by CD spectra, which were measured at 20 °C in 5 mM phosphate buffer (pH 7.0), with 1 mm path length quartz cells. Residual ratios of fluorescence intensity were measured at 80 °C in 5 mM phosphate buffer (pH 7.0) in 1 cm path length quartz cells.

Our reference protein, called CEGFP (PDB entry 2awk),[18] is a cycle-3[7] enhanced GFP[8] in which six residues are mutated from the original GFP sequence (F64L/S65T/Q80R/F99S/M153T/V163A). The three variants used in our study were GFP-N144C, GFP-N144C/Y145F, and a GFP with five residues (EYCFN) inserted between Y145 and N146 (GFP-5ins). The mutated region was in a flexible loop at the protein surface with no main chain atom forming H-bonds. Initial structures for the MD simulations were modeled from the CEGFP crystal structure. Two structures, with slightly different structures of the mutated loop, were modeled for GFP-5ins (Figure 1; see also Materials and Methods).

Thermal Stability and Fluorescence Activity. Circular dichroism (CD) indicated that all of the variants were folded at

dx.doi.org/10.1021/bi300580j | *Biochemistry* 2012, 51, 7974−7982

room temperature and that the CD spectra, at 20 °C, of GFP-N144C and GFP-N144C/Y145F fully overlapped with that of CEGFP. The spectrum of GFP-5ins differed slightly from that of CEGFP. Its estimated β-sheet content marginally decreased by 5%, while its α-helix content increased by 4% (Figure 2a and Table 1). The midpoint temperature (T_m) at pH 7.0 of GFP-5ins, calculated from the CD denaturation curves, decreased by ~9 °C, whereas that of GFP-N144C/Y145F increased by 3 °C relative to that of CEGFP (Figure 2c and Table 1).

The fluorescence activity at 20 °C of GFP-5ins was markedly reduced, whereas that of the GFP-N144C and GFP-N144C/Y145F was identical to that of CEGFP (Figure 1b and Table 1). Additionally, the fluorescence intensity of GFP incubated at 80 °C (pH7.0), where all variants were partially denatured (Table 1), indicated that GFP-5ins became completely inactive after incubation for <10 min, whereas the other variants had higher residual fluorescence activities, with GFP-N144C/Y145F having the highest. These results corroborated the thermal stability as assessed by the thermal denaturation curves (Figure 2c) and showed that a relatively modest increase in thermal stability can result in a significant improvement in GFP's thermoresistance as measured by residual fluorescence activity. The residual activity of GFP-N144C/Y145F was twice that of CEGFP, after a 20 min incubation, though the difference in melting temperature was a mere 3 °C. Finally, because CD measurements indicated that the overall GFP β-barrel fold was maintained in all variants at room temperature (Figure 2 and Table 1), we presumed that the alteration of fluorescence activity and thermal stability relates to local structural and dynamic changes that might be detectable by standard molecular dynamics simulations as performed in our study.

Average Tertiary and Secondary Structures Determined by MD Calculations. Simulations (20 ns) were repeated three times, and 6000 snapshots were saved for each model. Three independent simulations were conducted because that number would, at least partially, overcome the sampling problem,[31,32] especially because no large structural change took place upon mutation as demonstrated by circular dichroism experiments.

The main chain structures were analyzed by principal component analysis, and we selected the most densely populated area for further analysis (Figure S1 of the Supporting Information). We selected this area because it corresponds to the lowest (or one of the lowest)-energy region in the PCA space and because structures in this region can thus be assumed to be free of artifacts introduced by the manual modeling of the initial structures. We concentrated on analyzing 1000 structures in this lowest-energy region and defined the central structure as the average structure for representation in Figure S1 of the Supporting Information. The rmsd of the main chain atoms against the X-ray crystal structure converged to 1.0−1.5 Å, and the root-mean-square fluctuation (rmsf) per residue was essentially in line with the experimental b factor of the X-ray structure in all models (Figures S2 and S3 of the Supporting Information).

There were no large changes in the main chain structure of GFP around the mutation site in terms of rmsd, rmsf, or secondary structure (Figure 3a and Table 2). The mutated residues [144 (N144C) and 145 (Y145F)] formed a loop both in the crystal structure and during the simulation. Secondary structure prediction[33] (http://cib.cf.ocha.ac.jp/bitool/CF/) indicated a high loop-forming propensity for the five inserted residues (EYCFN), and they indeed formed a loop when GFP-5ins K was used; however, a β-bridge was formed between the Y and N residues when the S model was used as an initial structure, presumably because the two residues were close to each other in this model.

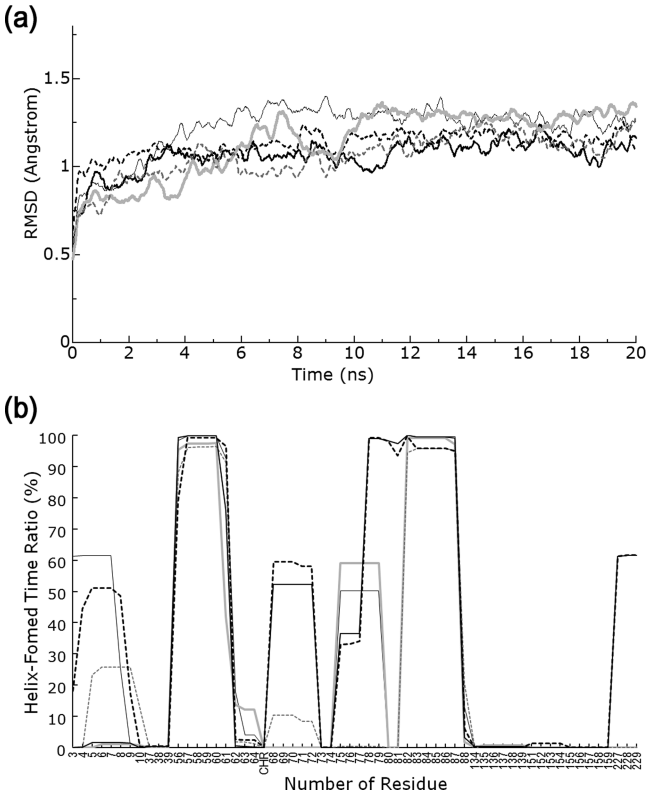


Figure 3. Results of MD simulations. Data for the CEGFP model are shown with a thin solid black line, data for the GFP-N144C model with a solid gray line, data for the GFP-N144C/Y145F model with a dashed gray line, data for the GFP-5ins S model with a thick black line, and data for the GFP-5ins K model with a dashed black line. (a) rmsd of the main chain against the X-ray crystal structure. The rmsd value was averaged per 100 ps. In CEGFP, GFP-N144C, and GFP-N144C/Y145F models, the deviation was calculated without the 10 starting residues and the 10 terminal residues. In GFP-5ins K and S, the value was calculated without the 10 starting residues and the 10 terminal residues, five inserted residues, and the five neighboring residues of the inserted region. (b) Time ratio of helix structure at each residue. The X-axis shows the residue number (residues that never formed a helix were removed).

In contrast, the secondary structure of residues 68−81, which is located far from the mutated region, differed between GFP-5ins and the other models. Residues 68−81 formed a helix in the reference crystal structure and in GFP-5ins, both S and K models, during the MD simulation. However, residues 68−81 formed loops in EGFP, GFP-N144C, and GFP-N144C/Y145F during the MD simulation (Figure 3b). This result is in line with the CD measurements indicating a slight increase in the helical content of GFP-5ins (Table 1). Furthermore, the region of residues 68−81 is located near the chromophore (which is formed by residues 65−67), and we speculate that the formation of helix in these regions influences the hydrogen bond network around the chromophore, which in turn could decrease the fluorescence intensity (see Hydrogen Bond Network Surrounding the Chromophore and Water Molecules in the β-Barrel's Interior).

dx.doi.org/10.1021/bi300580j | Biochemistry 2012, 51, 7974−7982

Table 2. Secondary Structure Content (percent) during the Simulations[a]

secondary structure	CEGFP	GFP-N144C	GFP-N144C/Y145F	GFP-5ins S	GFP-5ins K
3-10-helix	2.92	3.64	3.34	5.82	6.91
α-helix	5.90	3.18	2.99	2.69	2.97
β-bridge	0.76	0.65	0.88	1.48	0.47
β-sheet	50.11	48.90	50.46	49.47	49.43
strand	5.25	5.85	5.39	5.37	4.89
turn	17.74	16.04	16.15	19.18	14.36

[a]Secondary structure contents calculated from the 10 ns simulations. The secondary structures were assigned with PROCHECK,[30] and the contents were calculated from residue numbers and time ratios.

Table 3. Time Ratios of Hydrogen Bond Formation during Simulations[a]

H-bond pair		time ratio of H-bond formation (%)				
residue type, residue atom [main (m) or side (s)]		CEGFP	GFP-N144C	GFP-N144C/Y145F	GFP-5ins S	GFP-5ins K
Y39, O (m)	R73, NE (s)	0.1	5.5	14.2	4.3	4.6
Y39, O (m)	R73, NH1 (s)	9.2	3.3	19.3	3.4	9.6
G134, O (m)	G138, N (m)	20.4	22.8	46.2	37.5	0.7
K140, O (m)	E172, N (m)	38.3	0	95.7	95.5	0
N/C144, O (m)	N170, ND2 (s)	0.1	0	12.4	0	0
N146, OD1 (s)	R168, NE (s)	0	0	17.0	0	0
H148, NE2 (s)	K166, O (m)	0	21.1	45.4	0	0

[a]Time ratio of hydrogen bond formation during simulations. The time ratio of H-bond formation columns give the time ratio out of the 10 ns simulations, during which they formed a hydrogen bond.

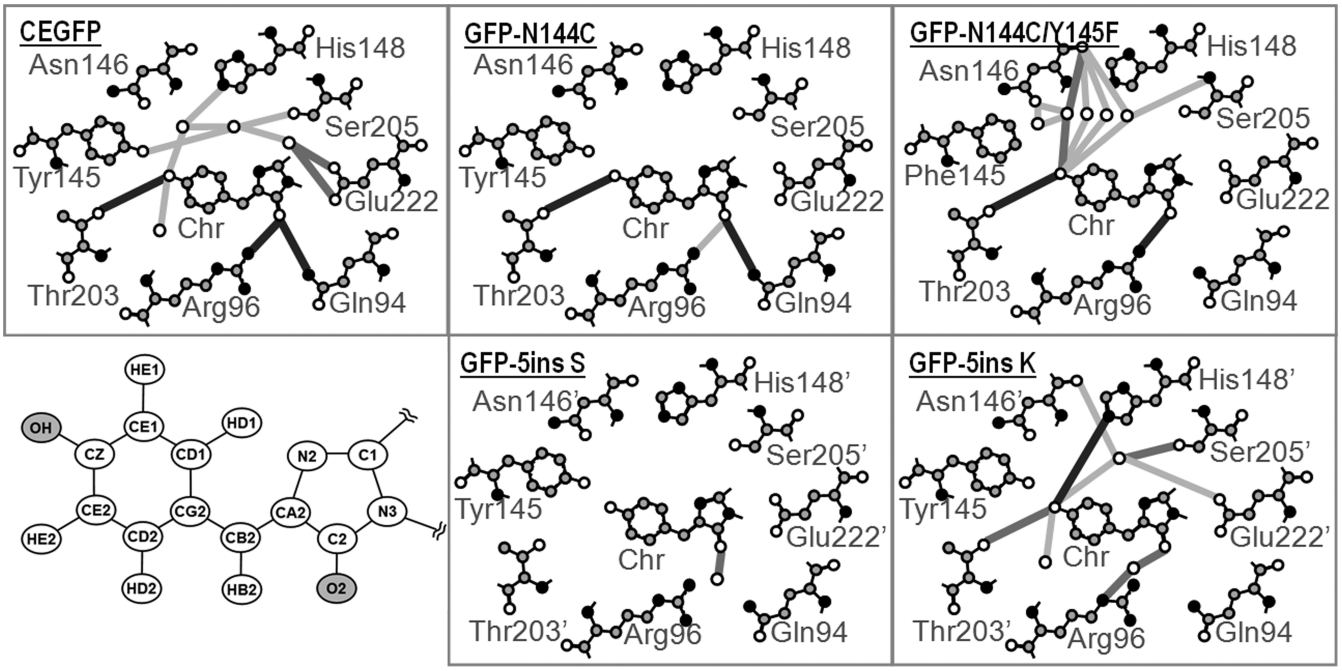


Figure 4. Cartoon representation of the hydrogen bond network surrounding the chromophore. The white, gray, and black circles represent oxygen, carbon, and nitrogen atoms, respectively. Atoms are shown using ball-and-stick representation, with thin sticks representing covalent bonds within amino acids. Thick lines represent hydrogen bonds, and the colors show the time ratio of hydrogen bond formation during the simulation: black for strong (80−100%), dark gray for intermediate (50−80%), and light gray for weak (20−50%). The bottom left panel shows our model GFP's chromophore. The gray-colored OH represents the phenolic group of 66Y, and the gray-colored O2 represents the carbonyl group of 66Y. The full list of hydrogen bonds is available in Table S1 of the Supporting Information.

Thermal Stabilization and Local Structural Changes. We examined the local dynamics and structural changes around the mutated region that could correlate with the thermal stability of the variants. As mentioned previously, we found few differences between the overall main chain structures of the variants, but local main chain structure and side chain configurations varied, sometimes significantly. The most significant change was observed at the Y145F side chain. In CEGFP and GFP-N144C, the side chain of Y145 bent and pointed away from the chromophore avoiding the hydrophobic core. In contrast, in the GFP-N144C/Y145F model, the aromatic ring of the replaced F145 turned straight to the hydrophobic core and notably approached the chromophore. (The averaged distance over 10 ns between the CZ atoms, a

dx.doi.org/10.1021/bi300580j | Biochemistry 2012, 51, 7974−7982

Table 4. Numbers of Internal Water Molecules (IWMs) and Their Closest Atoms in the Chromophore[a]

	CEGFP	GFP-N144C	GFP-N144C/Y145F	GFP-5ins S	GFP-5ins K
	Total Number of IWMs				
	7	7	6	12	6
	Number of IWMs				
internal water's closest atom in the chromophore[1]					
OH	6	6	5	11	3
N2	1	0	0	0	0
C1	0	1	0	0	1
O2	0	0	1	1	2

[a]Numbers of the internal water molecules and their closest atoms in the chromophore during simulations. The water molecules that stayed more than 0.5 ns within 3.0 Å of the chromophore were defined as internal waters. The atoms are shown in the bottom left panel of Figure 4.

part of the aromatic rings, of the chromophore and Y/F145 was 7.69 ± 0.47 Å in CEGFP, 7.61 ± 0.64 Å in GFP-N144C, and 4.48 ± 0.45 Å in GFP-N144C/Y145F.) Furthermore, the neighboring N146 side chain poked out and was completely exposed to the solvent in CEGFP and GFP-N144C, whereas in GFP-N144C/Y145F, it pointed toward the neighboring sheet (β8) to fill a gap between the sheet and the mutated loop. [The accessible solvent area (ASA) averaged over 10 ns was 137 Å^2 in CEGFP, 112 Å^2 in GFP-N144C, and 46 Å^2 in GFP-N144C/Y145F.] Additionally, the aromatic ring of H148 was completely bent into the β-barrel in CEGFP, whereas in GFP-N144C and GFP-N144C/Y145F, it was solvent-exposed and extended its side chain toward the neighboring sheet (β8). The averaged ASA of H148 was 42 Å^2 in CEGFP, 64 Å^2 in GFP-N144C, and 60 Å^2 in GFP-N144C/Y145F. Finally, the χ_1 side chain rotational angle of C144 was less constrained in GFP-N144C than in CEGFP. This is probably because the Asn side chain is smaller than that of the Cys, presumably allowing the Asn side chain to rotate rather freely in the folded state (Figure S4 of the Supporting Information). Overall, these observations suggest that the mutations increased both the local compactness and the number of hydrogen bonds in GFP-N144C/Y145F (Table 3), which could provide some rationals for its thermal stabilization.

Hydrogen Bond Network Surrounding the Chromophore and Water Molecules in the β-Barrel's Interior. We examined the trajectories of the MD simulations by focusing on the local dynamics and structural changes around the chromophore and the mutation or insertion site at residue 145 or 146. Among the many features examined, the hydrogen bond network, as defined by the atom's pairs distance and angle (see Materials and Methods), around the chromophore correlated with the loss of activity (Figure 4 and Table S1 of the Supporting Information). In all of the three active GFP variants (CEGFP, GFP-N144C, and GFP-N144C/Y145F), the chromophore formed hydrogen bonds with the side chains of Thr203, Arg96, and/or Gln94 most of the time. These hydrogen bonds were observed in previous simulation studies of GFPs and were shown to contribute to the rigidity of the chromophore.[11,12] These hydrogen bonds were either fully or partly absent from the inactive GFP-5ins, in a manner independent from the initial model, suggesting that the loss of hydrogen bonds between the chromophore and neighboring side chains relates to the loss of fluorescence activity.

Additionally, we analyzed the dynamic behavior of water molecules that remained in the interior of the GFP's barrel to assess their role in the hydrogen bond networks around the chromophore. To this end, we defined water molecules remaining within 3.0 Å of the chromophore for more than 0.5 ns as "internal water molecules" and analyzed their behaviors (Table 4). The numbers of internal water molecules were almost the same in all four models (seven water molecules in EGFP and GFP-N144C models and six in GFP-N144C/Y145F and GFP-5ins K models) except the GFP-5ins S model, which contained 12 internal water molecules (double that of other models). The chromophore's atom that is closest to most of the internal water molecules was the hydroxyl group of Y66 in EGFP, GFP-N144C, GFP-N144C/Y145F, and GFP-5ins S models (Figure 4; Y66 belongs to the chromophore, colored gray). On the other hand, in the GFP-5ins K model, the hydroxyl group was closest to only half (three of six) of the internal water molecules, and the remainder (two of six) were closer to the carbonyl group of the chromophore (Figure 4, colored gray). The hydrogen bond between the hydroxyl group and T203 that is present in EGFP, GFP-N144C, and GFP-N144C/Y145F is weakened or absent in the K and S models. We speculate that in the S model, this is caused by the increased motility of inner water molecules surrounding the hydroxyl group [11 water molecules (Table 4)], whereas internal waters provide alternative hydrogen bond partners in the K model (Figure 4 and Table S1 of the Supporting Information), which in turn affect the fluorescence activity. Similarly, water molecules around the chromophore's carbonyl group in both the K and S model weaken the hydrogen bonds with the surrounding E94 or R96, to which it is strongly bonded in EGFP, GFP-N144C, and GFP-N144C/Y145F (Figure 4 and Table S1 of the Supporting Information).

Reliability of the Correlations: Local versus Global Changes. Our study shows that the reliability and insightfulness of the calculations depended much on the initial GFP models and on the type of macromolecular properties examined. Dynamic and structural features of GFP-N144C and GFP-N144C/Y145F, where the structures of a mere single or two side chains were modeled, correlated relatively well with their macroscopic properties. For example, the stabilization of GFP-N144C/Y145F correlated with an increase in local compactness because of the replacement of the Y145's hydroxyl group with a hydrophobic aliphatic carbon in F145. The calculation indeed helped identify the thermostabilized GFP-N144C/Y145F variant. In contrast, in the GFP-5ins models, where the initial main chain and the side chain structures were extensively modeled, many features depended on the initial model and predictions seemed less reliable. For example, no microscopic feature related unambiguously to the thermal stability of GFP-5ins. On the other hand, the loss of activity of GFP-5ins was most likely related to a loss of hydrogen bonds around the chromophore and was predicted in a manner that was independent of the initial model.

Finally, even for GFP-N144C and GFP-N144C/Y145F, which differ by a mere single residue (Y145F), it was difficult to relate macroscopic properties to global dynamic or large structural changes. For example, we were essentially unable to relate the thermal stability or the activity changes to global molecular features such as a breakage of the β-barrel structure [in terms of β-sheet percentage, *B* factor, number of hydrogen bonds of the strands, etc. (Figures S5−S7 of the Supporting Information)]. Similarly, with regard to GFP-5ins, there is no rational way to distinguish the reliability of the S and K initial models, and we thus suspect that explaining the loss of thermal stability in GFP-5ins would need much longer calculation or more sophisticated techniques[34,35] versus those used in this study.

CONCLUSIONS

Computational molecular designs are subject to force field artifacts, sampling problems, and time scale limitations and hence cannot provide definitive insights. However, the significance of this study is to show that the combination of experimental and canonical simulation methods seems to identify local features that correlate with macroscopic protein properties. Specifically, it suggests that one could potentially design thermostabilized GFPs by monitoring local structural and dynamic changes around the mutated region. Likewise, but more speculatively, GFP variants with novel fluorescent properties might be identified just by focusing on the hydrogen bond network around the chromophore. Such computer-aided strategy should be effective when screening large sets of mutants before experimental assessment, as demonstrated by the identification of the thermostabilized GFP-N144C/Y145F variant.

ASSOCIATED CONTENT

Supporting Information

Additional results, including the PCA plots, rmsd, and rmsf calculated from molecular dynamic simulations. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: ykuroda@cc.tuat.ac.jp. Phone: +81-422-388-7794.

Funding

This work was partly supported by a JSPS grant in aid for scientific research (21300110, Y.K.), a special priority research area grant (21107505, Y.K.), and a contracted research "Protein 3000 Project" by the Ministry of Education, Culture, Sports, Science and Technology of Japan (A.S. and M.T.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Dr. Teppei Ebina, Hiromi Shimada, and members of the Kuroda group for discussion and technical expertise. We thank Prof. Masafumi Yohda (Tokyo University of Agriculture and Technology) for use of the fluorescence spectrometer in his laboratory and Patricia McGahan for English proofreading.

REFERENCES

(1) Shimomura, O., Johnson, F. H., and Saiga, Y. (1962) Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*. *J. Cell. Comp. Physiol. 59*, 223−239.

(2) Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G., and Cormier, M. J. (1992) Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene 111*, 229−233.

(3) Heim, R., and Tsien, R. Y. (1996) Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr. Biol. 6*, 178−182.

(4) Heim, R., Prasher, D. C., and Tsien, R. Y. (1994) Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. U.S.A. 91*, 12501−12504.

(5) Wachter, R. M., Elsliger, M. A., Kallio, K., Hanson, G. T., and Remington, S. J. (1998) Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure 6*, 1267−1277.

(6) Ormo, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y., and Remington, S. J. (1996) Crystal structure of the *Aequorea victoria* green fluorescent proteim. *Science 273*, 1392−1395.

(7) Crameri, A., Whitehorn, E. A., Tate, E., and Stemmer, W. P. (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol. 14*, 315−319.

(8) Cormack, B. P., Valdivia, R. H., and Falkow, S. (1996) FACS-optimized mutants of the green fluorescent protein (GFP). *Gene 173* ((Spec. Issue 1)), 33−38.

(9) Zapata-Hommer, O., and Griesbeck, O. (2003) Efficiently folding and circularly permuted variants of the Sapphire mutant of GFP. *BMC Biotechnol. 3*, 5.

(10) Nagai, T., Ibata, K., Park, E. S., Kubota, M., Mikoshiba, K., and Miyawaki, A. (2002) A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol. 20*, 87−90.

(11) Nifosi, R., and Tozzini, V. (2003) Molecular dynamics simulations of enhanced green fluorescent proteins: Effects of F64L, S65T and T203Y mutations on the ground-state proton equilibria. *Proteins 51*, 378−389.

(12) Patnaik, S. S., Trohalaki, S., and Pachter, R. (2004) Molecular modeling of green fluorescent protein: Structural effects of chromophore deprotonation. *Biopolymers 75*, 441−452.

(13) Islam, M. M., Sohya, S., Noguchi, K., Kidokoro, S., Yohda, M., and Kuroda, Y. (2009) Thermodynamic and structural analysis of highly stabilized BPTIs by single and double mutations. *Proteins 77*, 962−970.

(14) Higo, J., Ikebe, J., Kamiya, N., and Nakamura, H. (2012) Enhanced and effective conformational sampling of protein molecular systems for their free energy landscapes. *Biophys. Rev. 1*, 27−44.

(15) Hondoh, T., Kato, A., Yokoyama, S., and Kuroda, Y. (2006) Computer-aided NMR assay for detecting natively folded structural domains. *Protein Sci. 15*, 871−883.

(16) Kamioka, T., Tawa, M., Sohya, S., Yamazaki, T., and Kuroda, Y. (2009) Improved protein splicing reaction for low solubility protein fragments without insertion of native extein residues. *Biopolymers 92*, 465−470.

(17) Andrade, M. A., Chacon, P., Merelo, J. J., and Moran, F. (1993) Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng. 6*, 383−390.

(18) Wood, T. I., Barondeau, D. P., Hitomi, C., Kassmann, C. J., Tainer, J. A., and Getzoff, E. D. (2005) Defining the role of arginine 96 in green fluorescent protein fluorophore biosynthesis. *Biochemistry 44*, 16211−16220.

(19) Levitt, M. (1992) Accurate modeling of protein conformation by automatic segment matching. *J. Mol. Biol. 226*, 507−533.

(20) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Comput. Phys. 79*, 926−935.

(21) Case, D. A., Darden, T., Cheatham, T. E., III, Simmerling, C., Wang, J., Duke, R. E., Luo, R., Merz, K. M., Wang, B., Pearlman, D. A., Crowley, M., Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak, V., Cui, G., Beroza, P., Schafmeister, C., Caldwell, J. W., Ross, W. S., and

Kollman, P. A. (2004) *Amber*, version 8.0, University of California, San Francisco.
(22) Taiji, M., Narumi, T., Ohno, Y., and Konagaya, A. (2004) *MDGRAPE-3 chip: A 165 Gflops application specific LSI for molecular dynamics simulations*, IEEE Computer Society (CD-ROM).
(23) Narumi, T., Ohno, Y., Okimoto, N., Koishi, T., Suenaga, A., Futatsugi, N., Yanai, R., Himeno, R., Fujikawa, S., and Taiji, M. (2006) A 185 Tflops simulation of amyloid-forming peptides from Yeast prion Sup35 with the special-purpose computer system MDGRAPE-3. *Proceedings of Supercomputing 2006*, No. 49.
(24) Duan, Y., Wu, C., Chowdhury, S., Lee, M., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J., and Kollman, P. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem. 24*, 1999−2012.
(25) Reuter, N., Hai, L., and Thiel, W. (2002) Green fluorescent proteins: Empirical force field for the neutral and deprotonated forms of the chromophore. Molecular dynamics simulations of the GFP and S65T mutant. *J. Phys. Chem. B 106*, 6310−6321.
(26) Ryckaert, J., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys. 23*, 327−341.
(27) Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Comput. Phys. 98*, 10089−10092.
(28) Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Comput. Phys. 81*, 3684−3690.
(29) Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*, 2577−2637.
(30) Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr. 26*, 283−291.
(31) Jayachandran, G., Vishal, V., and Pande, V. S. (2006) Using massively parallel simulation and Markovian models to study protein folding: Examining the dynamics of the villin headpiece. *J. Chem. Phys. 124*, 164902.
(32) Suenaga, A., Hatakeyama, M., Kiyatkin, A., Radhakrishnan, R., Taiji, M., and Khlodenko, B. (2009) Molecular dynamics simulations reveal that Tyr-317 phosphorylation reduces Shc binding affinity for phosphotyrosyl resideus of epidermal growth factor receptor. *Biophys. J. 96*, 2278−2288.
(33) Chou, P. Y., and Fasman, G. D. (1974) Prediction of protein conformation. *Biochemistry 13*, 222−245.
(34) Hansmann, U. H. E., Okamoto, Y., and Eisenmenger, F. (1996) Molecular dynamics, Langevin, and hybrid Monte Carlo simulations in multicanonical ensemble. *Chem. Phys. Lett. 259*, 321−330.
(35) Shirai, H., Nakajima, N., Higo, J., Kidera, A., and Nakamura, H. (1998) Conformational sampling of CDR-H3 in antibodies by multicanonical molecular dynamics simulation. *J. Mol. Biol. 278*, 481−496.