Green Fluorescent Proteins: Empirical Force Field for the Neutral and Deprotonated Forms of the Chromophore. Molecular Dynamics Simulations of the Wild Type and S65T Mutant

Nathalie Reuter,*,†,‡ Hai Lin,† and Walter Thiel*,†

Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr, Germany, and Neuroendocrinologie et biologie cellulaire digestives, Institut National de la Santé et de la Recherche Médicale, U410, Faculté de médecine Xavier Bichat, 75018 Paris, France

Received: December 10, 2001; In Final Form: March 29, 2002

CHARMM force field parameters are reported for the neutral and anionic form of the *p*-hydroxybenzylidene-imidazolinone chromophore of green fluorescent protein (GFP). They were derived by fitting against ab initio (RHF/6-31G*) and density functional (B3LYP/6-31G*) reference data for the isolated chromophore model compounds and their complexes with water. More specifically, the parameters for the bonded interactions were calibrated against B3LYP geometries and rotational barriers of the model compounds, the atomic charges for the nonbonded Coulomb interactions were fitted against RHF geometries and binding energies of the complexes, and the van der Waals parameters were taken from related CHARMM residues. The optimized parameters reproduce the reference data well. For further validation, molecular dynamics (MD) simulations were carried out for wild-type GFP with a neutral chromophore and for the S65T mutant with an anionic chromophore. The average MD structures from 400 ps production runs are generally in reasonable agreement with the available X-ray structures, e.g., concerning the geometry of the chromophore in the protein and the hydrogen bonding network around the chromophore. In some aspects, however, the present MD simulations differ from the published X-ray results: in particular, they raise the issue whether the residue Glu222 might be deprotonated in the S65T mutant, contrary to the conclusions from the X-ray work.

1. Introduction

In the bioluminescent jellyfish *Aequorea victoria*, the blue chemoluminescence of photoprotein Aequorin is converted into green fluorescence by energy transfer to green fluorescent protein (GFP; for reviews, see refs 1 and 2). GFP is a spontaneously fluorescent protein with 238 amino acids. The cloning of the wild-type GFP gene[3,4] and its subsequent expression in heterologous systems[4−6] have established GFP as a novel genetic reporter system. When expressed in bacteria or eukariotic cells and illuminated by blue or UV light, GFP yields a bright green fluorescence without the need for any cofactor, substrate, or additional reactant. In addition, several mutants of GFP are now available with somewhat different absorption and emission spectra. GFP has become an extremely powerful noninvasive marker in living cells and an invaluable tool in molecular and cell biology.

From the numerous X-ray studies available,[7−13] it is known that the protein consists of an 11-stranded β barrel which is nearly a perfect cylinder, 42 Å long and 24 Å in diameter (cf. Figure 1). Its chromophore is generated by an auto-catalytic mechanism: cyclization of three adjacent residues (serine 65, tyrosine 66, and glycine 67) followed by 1,2-dehydrogenation of the tyrosine. The resulting product is a *p*-hydroxybenzylidene-imidazolinone, containing a phenolic ring (from Tyr 66) linked to an imidazolinone group by a bridging carbon. The chromophore lies in the center of the β barrel, which may help to protect it from quenching by oxygen and from attack by hydronium ions. The wild-type GFP has two absorption maxima at 395 and 475 nm that are believed to be due to the neutral and anionic form of the chromophore, respectively. The two different forms occur in a ratio of 6 to 1,[14] and the equilibrium appears to be governed by a hydrogen-bonding network that permits proton transfer between neighboring residues.[14] There exist several variants of GFP that exhibit different emission and absorption spectra due to mutation of one or more amino acids in the surrounding. This may perturb the chromophore either

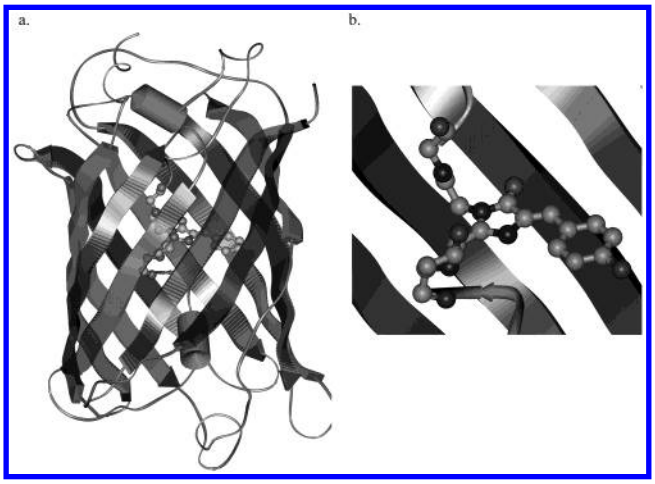

Figure 1. Position of the chromophore in the 3D structure of the green fluorescent protein. a. Secondary structure: the β barrel. b. Zoom on the chromophore.

* To whom correspondence should be addressed. N. Reuter (E-mail: reuter@bichat.inserm.fr) and W. Thiel (E-mail: thiel@mpi-muelheim-.mpg.de).
† Max-Planck-Institut für Kohlenforschung.
‡ Institut National de la Santé et de la Recherche Médicale.

10.1021/jp014476w CCC: $22.00
Published on Web 05/29/2002

directly when mutations affect residues 65, 66, or 67 or indirectly by changing the interactions between the chromophore and its neighboring residues. The mutations may modify the H-bond network and, as a consequence, may affect the proton transfer and/or the stabilization of the different protonation states of the chromophore. For instance, one of the better characterized mutants, the S65T mutant (replacement of serine 65 by a threonine), lacks the peak at 395 nm in the absorption spectra, which is evidence of a permanently deprotonated chromophore. Despite the huge amount of literature concerning the GFPs and the knowledge of the structural details brought by the X-ray studies,[7−13] there are still many open questions that can be addressed by theoretical calculations. The protonation states of the chromophore and its surrounding residues as well as their change upon photoconversion are not clear yet, and there is a controversy on the destination of the proton leaving the chromophore. Some groups[9,11,14] report it to be transferred to Glu 222, but a Fourier transform infrared (FTIR) study[15] does not support this hypothesis without identifying another proton acceptor. Recent ab initio calculations on the isolated chromophore[16] reconcile the available FTIR data[15] with a probable Glu222 terminus of the proton transfer in the protein, whereas the Raman spectra of wild-type GFP and the S65T mutant[17] provide no conclusive evidence on the protonation state of Glu222. A better understanding of the proton transfer and the role of the surrounding residues in the photochemical process would be of a great help to tune the efficiency of the existing mutants and even to design new mutants with specific photochemical properties.

Several theoretical studies on GFP have been reported so far. The formation of the chromophore has been studied in the gas phase[18] and in the protein.[19,20] Quantum chemical model computations have addressed the ground-state vibrational spectra[16] as well as the fluorescence phenomena and the excited states[21−26] of the isolated chromophore. Electrostatic calculations[27] have focused on the protonation and conformational substates in wild-type GFP, whereas a recent classical molecular dynamics (MD) simulation[28] has employed the AMBER force field[29] to investigate the overall stability of the β barrel and the network of strong hydrogen-bonding interactions surrounding the chromophore in the wild-type GFP.

We plan to study GFP with the use of the CHARMM force field[30] employing both classical simulations and combined quantum mechanical (QM) and molecular mechanical (MM) approaches. The chromophore in GFP does not resemble any natural amino acids, so that the usual force field parameters cannot be applied. To ensure a consistent description of GFP at the CHARMM level, we have decided to derive parameters for the chromophore (both in the neutral and anionic form) following the standard procedure described by MacKerell and co-workers.[30−32] Section 2 describes the development of these parameters. In section 3, we assess the quality of the new force field by comparison with ab initio and density functional results for the neutral and anionic chromophore covering geometries, rotational barriers, and chromophore−water intermolecular interaction energies. Section 4 reports 400 ps molecular dynamics (MD) simulations of both the wild-type protein and the S65T mutant, which are known to contain the neutral chromophore and its anionic form, respectively. Comparing the structural data obtained from the MD trajectories to X-ray data allows us to check how the new parameters perform in the protein environment. The MD simulations also provide new evidence on the protonation states in the S65T mutant.

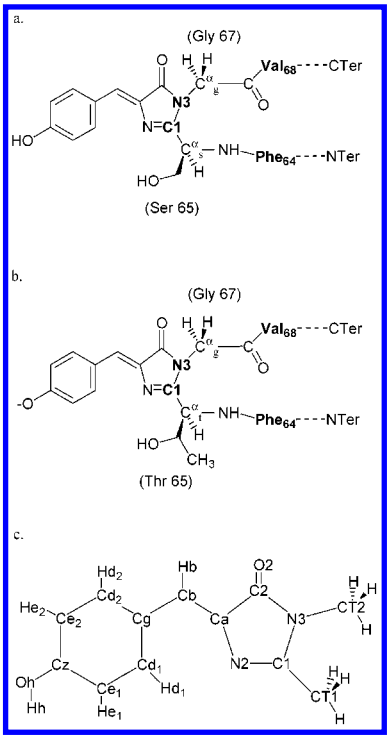

Figure 2. Chromophore and its linkage to the protein: (a) in the wild-type protein (neutral chromophore) and (b) in the S65T mutant (anionic chromophore) where a threonine replaces serine 65. c. Model compound for the neutral chromophore and atomic labels used in the text and tables. Labels are analogous for the anionic model (Hh being removed).

2. Parameter Development

A. General Strategy. The chromophore and its linkage to the protein are shown in Figures 1 and 2. It is a *p*-hydroxybenzylideneimidazolinone, linked to the backbone of the protein by two bonds involving the atoms labeled C1 and N3 (cf. Figure 2). The chromophore of GFP shows strong electron delocalization, especially in the anion, and is therefore quite different from any compound already parametrized in CHARMM22. This makes it necessary to develop new parameters for the chromophore. These need to reproduce the geometrical features of the chromophore and its environment that are crucial for the characteristics of the absorption spectra like the twist of the six-membered ring relative to the imidazolinone ring and the hydrogen bonding network between the chomophore and its neighboring residues.

To keep the new parameters consistent with those already available in the CHARMM22 force field, we employed the same parametrization strategy as described for CHARMM22[30] and CHARMM27.[32] The overall procedure is aimed at reproducing the internal geometries and intermolecular interaction energies of small model compounds. Target data were obtained by quantum mechanical studies against which parameters were then fitted via an iterative approach. Details of the procedure are given in the following paragraphs and the Supporting Information.

B. Model Compounds. Figure 2c shows the model chosen for the neutral chromophore, as well as the labels used for each atom. The anionic model is analogous except that Hh is removed. The model compounds include the entire chromophore (tyrosyl ring, bridging bonds, and imidazolinone ring) to account for the effects of electron delocalization. They contain two methyl groups (bound to C1 and N3) to represent the linkage of the chromophore to the protein. The next atoms in the sequence, both in the direction to the C- and N-terminal part,

belong to usual amino acids and can be described by standard parameters; therefore, they do not appear in our model compounds.

C. Potential Energy Function in CHARMM. The potential energy function of CHARMM is given by eq 1. The stretching, bending and out-of-plane deformations are represented by harmonic potentials. The dihedral term is a cosine function. The van der Waals interactions are represented by a 12−6 Lennard-Jones potential and the electrostatic interactions by a Coulomb potential:

$$E = \sum_{\text{bonds}} K_r(r - r^0)^2 + \sum_{\substack{\text{valence}\\ \text{angles}}} K_\theta(\theta - \theta^0)^2 + \sum_{\text{dihedrals}} K_\phi(1 + \cos(n\phi - \phi^0)) + \sum_{\text{impropers}} K_\varphi(\varphi - \varphi^0)^2 + \sum_{\substack{\text{nonbonded}\\ i-k\ \text{pairs}}} \frac{q_i q_k}{Dr_{ik}} + \sum_{\substack{\text{nonbonded}\\ i-k\ \text{pairs}}} \epsilon_{ik}\left(\left(\frac{R_{\min,ik}}{r_{ik}}\right)^{12} - 2\left(\frac{R_{\min,ik}}{r_{ik}}\right)^{6}\right) \quad (1)$$

For all bonds, valence angles, and improper terms, force constants (K_r, K_θ, K_φ) and equilibrium values (r^0, θ^0, φ^0) must be defined. For each dihedral term, three parameters are needed: the force constant K_ϕ, the periodicity n, and the phase ϕ^0. For a correct description of the electrostatic nonbonded interactions of the chromophore, each atom needs to be assigned a proper partial charge q_i, different in the neutral and the anionic form. The van der Waals parameters are the usual atom-type dependent parameters: well depth, ϵ_{ik}, and minimum interaction radius, $R_{\min,ik}$.

D. Optimization of the Parameters. The development of new parameters for CHARMM starts with a reasonably chosen initial set of parameters which is then optimized against various target data obtained mainly, in our case, from ab initio restricted Hartree−Fock (RHF) and density functional DFT calculations (optimized geometries, adiabatic rotational barriers, and intermolecular interaction energies). The DFT calculations employed Becke's three-parameter functional (B3LYP)[33] as implemented in the Gaussian suite of programs.[34] Following the standard iterative CHARMM parametrization procedure,[30−32,35] we have derived parameters for two new residues in the CHARMM force field, i.e., for the neutral GFP chromophore and for its anionic form (see Supporting Information for details).

Nonbonded Interactions. The van der Waals parameters for the chromophore are set equal to already existing parameters in CHARMM22: we adopt the values of the tyrosine residue (for the tyrosyl group), the imidazole ring of the histidine residue (for the imidazolinone ring), and the propene fragment (for the CH bridge). These parameters are similar in the anionic and the neutral species, except for the oxygen atom of the tyrosine ring. They are summarized in Table 1 which lists all atoms of the chromophore and the associated CHARMM22 atom types for the van der Waals parameters. The charges obtained for the model compounds are also summarized in Table 1. They were obtained by fitting ab initio (RHF/6-31G*) interaction energies and geometries of complexes between model compounds and water molecules (see Figure 3 and section 3.A). Details are given in the Supporting Information. The charges of the atoms connecting to the rest of the proteins (Ser/Thr65 and Gly67) are the ones defined in CHARMM for those types of residues, except for the C^α atom of the glycine whose charge is changed to −0.18e to keep the neutrality of the newly defined residue for the chromophore. In the case of the neutral species, the carbon and the hydrogen atoms of the six-membered ring carry charges of ±0.115e which are identical to the charges of the six-membered ring of the tyrosine residue in CHARMM. The charges on C_z, O_h, and H_h are also taken directly from CHARMM.

TABLE 1: Atom Labels (cf. Figure 2) in the Anionic and Neutral Model Compounds, CHARMM22 Atom Types for the van der Waals Parameters, and Point Charges for the Neutral and Anionic Model Compounds

atom label	van der Waals types in CHARMM22	point charges (e) neutral	anion
	Imidazolinone		
C1	CPH2	0.76	0.50
N2	NR2	−0.55	−0.60
N3	NR1	−0.64	−0.57
C2	CPH1	0.80	0.57
O2	O	−0.61	−0.57
Ca	CPH1	0.24	0.10
	Bridging Bonds		
Cb	CE1	−0.10	−0.14
Hb	HA1	0.10	0.21
	Six-Membered Ring		
Cg	CA	0.00	−0.09
Cd1	CA	−0.115	−0.08
Hd1	HP	0.115	0.14
Cd2	CA	−0.115	−0.08
Hd2	HP	0.115	0.14
Ce1	CA	−0.115	−0.28
He1	HP	0.115	0.10
Ce2	CA	−0.115	−0.28
He2	HP	0.115	0.10
Cz	CA	0.11	0.45
Oh	OH1[a]/O[b]	−0.54	−0.62
Hh	H	0.43	

[a] Neutral. [b] Anionic.

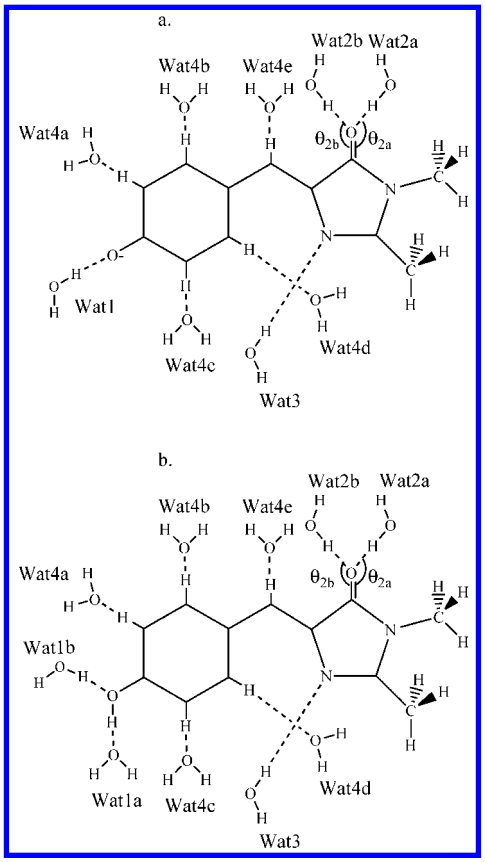

Figure 3. Complexes between the model compounds and water. a. Anionic chromophore. b. Neutral chromophore.

TABLE 2: Optimized Stretching Parameters for the Model Chromophore: Force Constants K_r (kcal mol^{-1} Å^{-2}) and Equilibrium Distance r^0 (Å)[a]

	anion		neutral	
bond	K_r	r^0	K_r	r^0
C1−C$^{\alpha}$t/s	**354.0**	1.490	**320.0**	1.490
N3−C$^{\alpha}$g	**396.0**	**1.440**	**352.0**	**1.450**
N3−C1	400.0[b]	1.390	400.0[b]	1.390
N3−C2	400.0[b]	1.410	400.0[b]	1.410
C2−O2	**854.0**	**1.240**	**807.0**	**1.220**
C1−N2	400.0[b]	1.300	400.0[b]	1.300
C2−Ca	410.0[b]	**1.460**	410.0[b]	**1.490**
N2−Ca	400.0[b]	1.400	400.0[b]	1.400
Ca−Cb	**500.0**	**1.390**	**560.0**	**1.360**
Cb−Hb	360.5[c]	1.100[c]	360.5[c]	1.100[c]
Cb−Cg	**437.0**	**1.410**	**370.0**	**1.450**
Cg−Cd1	305.0[d]	**1.430**	305.0[d]	**1.375**[d]
Cg−Cd2	305.0[d]	**1.430**	305.0[d]	**1.375**[d]
Cd1−Hd1	340.0[d]	1.080[d]	340.0[d]	1.080[d]
Cd2−Hd2	340.0[d]	1.080[d]	340.0[d]	1.080[d]
Cd1−Ce1	305.0[d]	**1.350**	305.0[d]	**1.375**[d]
Cd2−Ce2	305.0[d]	**1.350**	305.0[d]	**1.375**[d]
Ce1−He1	340.0[d]	1.080[d]	340.0[d]	1.080[d]
Ce2−He2	340.0[d]	1.080[d]	340.0[d]	1.080[d]
Ce1−Cz	305.0[d]	**1.455**	305.0[d]	**1.375**[d]
Ce2−Cz	305.0[d]	**1.455**	305.0[d]	**1.375**[d]
Cz−Oh	**842.0**	**1.250**	**334.3**[d]	**1.411**[d]
Oh−Hh			545.0[d]	0.960[d]

[a] Parameters that differ in the neutral and anionic species are given in bold. [b] From CHARMM22, imidazole ring of the histidine residue, bonds (bond types): Nδ1-Cϵ1(NR1−CPH2), Nδ1-Cγ(NR1−CPH1), Nϵ2-Cϵ1(NR2−CPH2), Cδ2-Cγ(CPH1−CPH1) and Nϵ2-Cδ2(NR2−CPH1). [c] From CHARMM22, propene. [d] From CHARMM22, tyrosine residue parameters, CA-CA, CA-HP, CA-OH or OH−HH bond types.

Bond Stretching. Whenever possible, harmonic potentials already existing in the force field were utilized. For bonds that exhibit geometrical and chemical similarities with a tyrosyl ring, or an imidazole ring or propene, we used the parameters defined in CHARMM22 for those compounds. The details of these comparisons and procedures are given as Supporting Information. Force constants and equilibrium values for all bond types are listed in Table 2, which also specifies the origin of the parameters. Parameters without superscript are new. The parameters that are different in the neutral and anionic chromophore are printed in bold characters. A major difference between these two species is the geometry of the tyrosyl ring. In the case of the neutral chromophore, the DFT(B3LYP/6-31G*) bond lengths are similar for all C−C bonds in this ring (about 1.40 ± 0.01 Å), whereas in the anionic form, those for Cd1−Ce1 or Cd2−Ce2 are significantly shorter (1.37 Å) than the others (1.43−1.46 Å). This is reflected in different stretching potentials for these bonds (see r^0 values in Table 2). Another important difference concerns the bonds linking the two rings: Ca−Cb and Cb−Cg. In the anionic species, they have partial double-bond character (1.39 and 1.41 Å, respectively), whereas in the neutral form, they have the characteristics of a double bond (Ca−Cb, 1.36 Å) and of a single bond (Cb−Cg, 1.45 Å). These two bonds have thus been assigned different force constants and equilibrium values.

Valence Angles. For the bending potentials, we followed exactly the same procedure as for the stretching potentials; we used existing parameters whenever it made sense regarding the similarities between the model chromophore on one hand and the histidine, tyrosine, and propene residues defined in CHARMM on the other hand. The valence angles and the corresponding optimized potential parameters are listed in Table 3, which also gives the origin of the parameters, like in the case of the stretching potentials. The anion and the neutral species differ mainly in the equilibrium values, θ^0. The force constants of the angles centered on C2, N3, and C1 in the imidazolinone ring also differ, being slightly higher in the case of the anion. These angles are less rigid in the neutral chromophore, where the electronic delocalization is less pronounced, as found from ab initio RHF and DFT calculations. Furthermore, in the tyrosyl ring, the C−C−C angles are essentially equivalent in the neutral form but slightly different in the anion which is reflected in the θ^0 values of the angles centered on atoms Ce or Cd and Cg or Cz.

TABLE 3: Optimized Bending Parameters for the Model Chromophore: Force Constants K_θ (kcal mol^{-1} Rad^{-2}) and the Equilibrium Angles θ^0 (degree)[a]

			anion		neutral	
valence angle			K_θ	θ^0	K_θ	θ^0
Ha	C$^{\alpha}$t/s	C1	50.0[b]	109.5[b]	50.0[b]	109.5[b]
N	C$^{\alpha}$t/s	C1	50.0[b]	107.0[b]	50.0[b]	107.0[b]
C$^{\beta}$t/s	C$^{\alpha}$t/s	C1	52.0[b]	108.0[b]	52.0[b]	108.0[b]
N2	C1	C$^{\alpha}$t/s	40.0	125.0	40.0	125.0
N2	C1	N3	130.0[c]	**113.3**	130.0[c]	**114.0**
C1	N2	Ca	130.0[c]	**106.6**	130.0[c]	**106.0**
N3	C1	C$^{\alpha}$t/s	**40.0**	**121.7**	**35.0**	**121.4**
C1	N3	C2	130.0[c]	107.9	130.0[c]	107.9
C1	N3	C$^{\alpha}$g	**40.0**	**128.3**	**36.0**	**129.0**
N2	Ca	C2	130.0[c]	108.3	130.0[c]	108.3
N2	Ca	Cb	45.8[c]	129.5	45.8[c]	129.5
C2	N3	C$^{\alpha}$g	**40.0**	**123.8**	**32.0**	**123.4**
N3	C2	O2	**50.0**	**124.0**	**42.0**	**126.0**
N3	C2	Ca	130.0[c]	103.0	130.0[c]	103.0
N3	C$^{\alpha}$g	H$^{\alpha}$g	48.0[c]	108.0[c]	48.0[c]	108.0[c]
N3	C$^{\alpha}$g	Cg	50.0[c]	107.0[c]	50.0[c]	107.0[c]
O2	C2	Ca	**44.0**	**133.0**	**38.0**	**132.0**
C2	Ca	Cb	45.8[c]	**122.7**	45.8[c]	**122.0**
Ca	Cb	Cg	130.0	**133.2**	130.0	**130.0**
Ca	Cb	Hb	42.0[d]	**112.0**	42.0[d]	**114.0**
Cb	Cg	Cd1(2)	45.8[e]	**120.0**	45.8[e]	**121.0**
Hb	Cb	Cg	42.0	**115.0**	42.0	**116.0**
Cg	Cd1(2)	Ce1(2)	40.0[e]	**122.0**	40.0[e]	**120.0**[e]
Cd1(2)	Cg	Cd2(1)	40.0[e]	**116.0**	40.0[e]	**120.0**[e]
Cd1(2)	Ce1(2)	Cz	40.0[e]	**122.0**	40.0[e]	**120.0**[e]
Ce1(2)	Cz	Ce2(1)	40.0[e]	**115.0**	40.0[e]	**120.0**[e]
He1(2)	Ce1(2)	Cz	30.0[e]	120.0[e]	30.0[e]	120.0[e]
He1(2)	Ce1(2)	Cd1(2)	30.0[e]	120.0[e]	30.0[e]	120.0[e]
Hd1(2)	Cd1(2)	Ce1(2)	30.0[e]	120.0[e]	30.0[e]	120.0[e]
Hd1(2)	Cd1(2)	Cg	30.0[e]	120.0[e]	30.0[e]	120.0[e]
Oh	Cz	Ce1(2)	45.2	120.0[e]	45.2[e]	120.0[e]
Hh	Oh	Cz			65.0[e]	108.0[e]

[a] Parameters that differ in the neutral and anionic species are given in bold. [b] From CHARMM22, serine/threonine, angle type (atom names): NH1−CT1−HB (N−CA−HA), NH1−CT1−C (N−CA−C), HB−CT1−C (HA−CA−C), CT2/CT1−CT1−C (CB−CA−C). [c] From CHARMM22, histidine; imidazole ring and Cβ−Cγ−Nδ, Cβ−Cγ−Cδ, angle types: CT2−CPH1−NR2, CT2−CPH1−CPH1. [d] From CHARMM22, propene: HEL1−CEL1−CEL2. [e] From CHARMM22, tyrosine, angle types: CT2−CA−CA, CA−CA−CA, HP−CA−CA, OH1−CA−CA, H−OH1−CA.

Dihedral Angles. We define only four new dihedral potentials: for the rotation around bonds linking the two rings (Ca−Cb and Cb-Cg) and around bonds linking the chromophore to the backbone of the protein (C1−Ct1 and N3−Ct2). The other torsion parameters were taken from CHARMM22 and are identical for the anionic and neutral form of the chromophore. The parameters for our newly defined potentials are summarized in Table 4. The barriers to rotation around the Ca−Cb and Cb-Cg bonds, respectively, are quite different in the neutral chromophore and the anion because of a more pronounced electronic delocalization in the latter (see above). This is reflected in the corresponding torsional force constant (see Table 4).

TABLE 4: Optimized Torsional Parameters for the Model Chromophores: Force Constant K_ϕ (kcal mol^{-1}), Multiplicity *n* and Phase ϕ^0 (Degrees)

dihedral angle				anion			neutral		
				K_ϕ	*n*	ϕ^0	K_ϕ	*n*	ϕ^0
X_1 (X_1=C2, N2)	Ca	Cb	X_2 (X_2=Cg, Hb)	3.900	2	180	6.840	2	180
X_1 (X_1=Ca, Hb)	Cb	Cg	X_2 (X_2=Cd1, Cd2)	2.700	2	180	1.400	2	180
C1	N3	C^αg	X (X= H^α_g, C)	0.032	3	0	0.067	3	0
C2	N3	C^αg	X (X= H^α_g, C)	0.032	3	180	0.067	3	180
N3	C1	C^αt/s	X (X=N, $H^\alpha_{t/s}$, $C^\beta_{t/s}$)	0.105	3	0	0.100	3	0
N2	C1	C^αt/s	X (X=N, $H^\alpha_{t/s}$, $C^\beta_{t/s}$)	0.105	3	180	0.100	3	180

TABLE 5: Complexes between Anionic Chromophore and Water (Interaction Energies E in kcal/mol, Distances R in Å, Angles θ in Degree)

complexes[b]	ab initio[a]			empirical		
	R	θ	*E*	*R*	θ	*E*
C_zO_h−HW^1	1.87		−10.67	1.70		−10.81
C_2O_2−HW^{2a}	1.89	145.0	−9.81	1.75	148.0	−9.15
C_2O_2−HW^{2b}	1.91	139.0	−8.35	1.70	134.5	−9.42

[a] RHF/6-31G*. [b] See Figure 3a. W denotes Wat.

Impropers. Both in the neutral and anionic chromophore, improper potentials are included at the bridging atoms (Ca and Cb) and at some atoms of the imidazolinone group (C1, N3, and N2) to reproduce the DFT torsion potentials. These harmonic potentials depend on the force constant K_φ and the equilibrium value φ_0, with the latter being 0 in all cases presently. The optimized force constants K_φ (in kcal mol^{-1} rad^{-2}) for (C1, N3, C2, Ca, and Cb) are (50, 50, 50, 50, and 30) for the anion and (0.5, 0.45, 0.50, 220, and 30) for the neutral chromophore. In the latter case, the small values for C1, N3, and N2 indicate that the imidazolinone ring is quite flexible, whereas the high value for Ca reflects the rigidity of the double bond.

3. Comparison between Force-Field and ab Initio Calculations on the Model Compounds

A. Chromophore−Water Interactions. The complexes between the model compounds and one water molecule that were used during parametrization and validation are indicated in Figure 3. Interaction energies and selected geometrical variables are given in Tables 5 and 6 for those complexes which are bound at the RHF/6-31G* level. Their geometries were determined by a partial optimization: the monomers were held fixed at the optimized geometry of the isolated model compound and the TIP3P[36,37] water geometry, and normally only the hydrogen bond distance was optimized assuming a linear hydrogen bond in the chosen orientation (Figure 3). For the complexes Wat2 (a and b), the angles θ_{2a} and θ_{2b} were also optimized. The interaction energy was calculated as the difference between the total energy of the complex and the sum of the monomer energies.

Anionic Chromophore. From Table 5, one can see that the interaction energies and geometries calculated with our final set of parameters are in reasonable agreement with the ab initio results. The hydrogen bond distances are underestimated by approximately 0.2 Å, consistent with previous studies,[31,32] and the angles are close to the RHF values. During the parametrization, it was impossible to accurately reproduce both interaction energies and geometries. We chose to tune the charges such that the pattern of the resulting deviations was similar as elsewhere in CHARMM,[31,32] and suitable for the intended applications.

TABLE 6: Complexes between Neutral Chromophore and Water (Interaction Energies *E* in kcal/mol, Distances *R* in Å, and Angles θ in Degrees)

complexes[b]	scaled ab initio[a]			empirical		
	R	θ	*E*	*R*	θ	*E*
O_hH_h−OW^{1a}	1.99		−7.53	1.86		−7.02
C_zO_h−HW^{1b}	2.13		−4.01	1.90		−4.55
C_2O_2−HW^{2a}	2.00	139.5	−7.25	1.75	144.0	−7.13
C_2O_2−HW^{2b}	1.99	129.4	−7.69	1.75	134.0	−8.00
N_2−HW^3	3.97		−1.61	3.60		−1.26
$C_{e2}H_{e2}$−OW^{4a}	2.58		−1.19	2.70		−0.36
$C_{d2}H_{d2}$−OW^{4b}	2.50		−2.57	2.62		−2.02
$C_{e1}H_{e1}$−OW^{4c}	2.63		−2.31	2.60		−2.30
$C_{d1}H_{d1}$−OW^{4d}	6.60		−0.94	6.10		−1.44
C_bH_b−OW^{4e}	2.36		−1.50	2.46		−1.92

[a] RHF/6-31G*, energies are scaled by a factor of 1.16.[31] [b] See Figure 3b. W denotes Wat.

Neutral Chromophore. According to Table 6, the empirical interaction energies reproduce the scaled ab initio RHF values quite well when using the recommended scale factor of 1.16.[31] An attractive interaction potential is found for all 10 complexes which are also bound at the ab initio level. In general, the trends in the hydrogen bond lengths and angles are well reproduced. Weak complexes arise in the case of the N2 and Cd1−Hd1 groups, because of repulsive interactions caused by their spatial proximity. From the computed ab initio distances, it is clear that N2−HW3 (R = 3.97 Å) and Cd1−OW4d (R = 6.60 Å) are not hydrogen-bonded but only held together by long-range electrostatic interactions. These features are reproduced by our empirical potential in a qualitative sense.

B. Geometries and Rotational Barriers of the Chromophore. The average absolute difference between the DFT target bond lengths and the empirical bond lengths in the optimized structures is 0.0094 Å for the anionic chromophore and 0.0118 Å for the neutral form. The bond lengths on the isolated model compounds are thus well reproduced by our parameters. The average absolute difference between bond angles of the empirically optimized and the DFT optimized geometry is 1.27° for the anion and 1.11° for the neutral molecule which is again satisfactory.

The empirical rotational barriers around the bridging bonds compare well with the DFT and OM2[24] values (cf. Table 7), both for the anionic or the neutral form. Compared with DFT, the barriers around Ca−Cb and Cb-Cg tend to be slightly underestimated, whereas the relative energy of the trans conformation is realistic for the anion and too high for the neutral chromophore. These slight discrepancies are unimportant for our purposes because MD simulations starting from the crystallographic structure at 300 K (with a cis-conformation) will remain in the "cis region" and will not cross rotational barriers of the order of 30 kcal/mol.

4. Molecular Dynamics Simulations

Having established that the new parameters describe the isolated chromophores realistically, we now turn to MD simulations of different forms of GFP to check their performance in the protein environment.

In the wild-type GFP, the majority species contains the chromophore in its neutral state.[14] There exist different mutants of the GFP (see refs 1 and 2 for reviews), among them S65T[38]

TABLE 7: Energies (kcal/mol) Relative to the cis-Conformation Shown in Figure 2c

a. Anion

	trans[a]	Ca−Cb[b]	Cb-Cg[c]	C1−Ct1[d]	N3−Ct2[e]
DFT	2.50	32.74	24.44	1.63	1.12
OM2[24]	1.0	26.2	20.9		
emp.	3.03	31.49	21.28	1.60	1.11

b. Neutral Chromophore

	trans[a]	Ca−Cb[b]	Cb-Cg[c]	C1−Ct1[d]	N3−Ct2[e]
DFT	−2.70	55.00	9.57	1.63	0.90
OM2[24]	−0.3	54.7	4.1		
emp.	2.33	51.85	9.50	1.73	0.71

[a] Trans-conformation with a dihedral angle ϕ(N2−Ca−Cb−Cg) = 180°. The MP2/6-31G* value for the anion is 3.87 kcal/mol. [b] Barrier for rotation around the Ca−Cb bond with a dihedral angle ϕ(N2−Ca−Cb−Cg) = 90°. For the anion, the DFT value refers to a transition state (TS) calculation, whereas for the neutral chromophore, it refers to an energy minimization with the dihedral angle fixed to 90°. The MP2/6-31G* value (TS) for the anion is 31.95 kcal/mol. [c] Barrier for rotation around the Cb-Cg bond with a dihedral angle ϕ(Ca−Cb−Cg−Cd1) = 90°. Both for the anion and the neutral chromophore, the DFT values refer to TS calculations. [d] Barrier for rotation around the C1−Ct1 bond. [e] Barrier for rotation around the N3−Ct2 bond.

where the chromophore is deprotonated (see Figure 2b). X-ray structures of each of the two forms of the chromophore (neutral or anionic) in its natural environment are available.[7−9] We report molecular dynamics simulations of the two forms of the protein and compare the structural data derived from the MD trajectories with the X-ray structures of the S65T mutant[8] and the wild-type GFP.[9] All of the steps described below (setup, solvation, geometry optimization, MD simulation, and analysis of the trajectories) were performed with the CHARMM program,[39] version c27b2, using the CHARMM22 all atom force field,[35] with the new parameters for the chromophores.

A. Computational Details. The initial set of coordinates used for the MD simulations were prepared from X-ray structures obtained from the Protein Data Bank[40,41] under the reference 1ema (1.9 Å resolution) for the mutant[8] and 1emb (2.1 Å resolution) for the wild type.[9] In the latter structure, the chromophore is in its natural protonation state (neutral). We chose this structure as our starting point for the wild type because it is a monomer (which we want to simulate), whereas an earlier structure (pdb code 1gfl[7]) with a slightly better resolution (1.9 Å) refers to a dimer. For the mutant, missing side chains of several residues were added (for instance, lysines 26, 52, 101, 107, 131, 156, 158, 162, and 214; glutamic acids 6, 124, and 132; arginine 122; glutamin 157; asparagin 212). Then the coordinates of the missing hydrogen atoms were determined using the HBUILD[42] facility in CHARMM, at pH 7.

It has been proposed[9,11] from a detailed analysis of the two crystal structures that the photoconversion of GFP involves an intramolecular proton transfer, initiated by deprotonation of the phenolic moiety which in turn leads to rearrangements of the hydrogen-bonding network in the protein and eventually to protonation of anionic Glu 222. However, later FTIR studies[15] that clearly show the deprotonation of the chromophore upon photoconversion do not provide any indications for a change of the protonation state of Glu222. Hence, it is presently not clear where the proton of the chromophore is transferred to or on which amino acid it may reside transiently. In our simulations of the S65T mutant with the anionic chromophore, we decided to protonate the carboxylic group of Glu222 to be consistent with the chosen X-ray structures and the hypothesis of the groups who have refined them.[8,9] For the simulation of the wild type, the chromophore is protonated, and the glutamic acid is negative, as it should be at pH 7. As a result, the following types of residues were charged (total charge different from 0): aspartic acid (−1), glutamic acids (−1) except Glu 222 for 1ema, lysine (+1), arginine (+1). Considering the surrounding of their imidazole group, the histidine residues numbered 25, 148, 181, 199, and 217 were protonated at their δ nitrogen atom, and the others (77, 81, 139, and 169) were protonated at their ϵ nitrogen atom. The total charge of the protein is −6 e in both cases.

For the mutant protein, we used only the 95 crystallographic water molecules given in the X-ray structure (1ema). For the wild-type species, the 74 water molecules listed in the PDB file (1emb) were included and supplemented, as suggested previously,[28] by two water molecules corresponding to W320 and W383 in the mutant (see W75 and W76 in Figure 6b, both in the vicinity of W19). All subsequent calculations (minimizations and MD) employed a constant dielectric function with $D = 1$. We used a 14 Å cutoff with a switch truncation function for the van der Waals interactions and a shift function for the electrostatic interactions.

The initially built structures were first progressively minimized in vacuum: at the beginning, tight harmonic constraints were applied on the backbone, and looser constraints were applied on the side chains, and then the force constants of the constraints were slowly decreased to finally vanish. In the next step, the system (protein + crystal waters) was solvated in a sphere (37 Å radius) of preequilibrated water molecules, centered on the center of the protein.[43] It was then minimized for 100 steps of steepest descent and equilibrated around the fixed protein for 2 ps, using stochastic boundary conditions,[44−48] with a buffer region of 5 Å. For all simulations, the time step was 1 fs, and stochastic boundary conditions were used as well as SHAKE[49] to constrain the bond lengths between hydrogen and heavy atoms. The whole system was then heated from 100 to 300 K, increasing the temperature by 1 K each 100 steps. Once the system had reached 300 K, it was equilibrated for 200 ps under NVT conditions. After the equilibration phase, production runs were performed for 400 ps under NVE conditions. The average temperatures during these production runs were 296.6 K (standard deviation of 2.1 K) for the S65T mutant and 296.7 K (standard deviation of 2.0 K) for the wild-type form of GFP.

B. RMS Deviations from the X-ray Structures vs Simulation Time. Figure 4 shows the root-mean-square (rms) deviations between the MD and X-ray structures, as a function of the simulation time. For each system, two sets of rms values are given: (i) for the ensemble of the heavy atoms of the protein and the crystallographic water molecules (solid lines) and (ii) for only the backbone of the protein (dashed lines). Conformations were stored every 10 ps during heating and equilibration and every 1 ps during the production run. The rms values of the heavy atoms (i) increase continuously, reflecting mostly the mobility of some of the crystallographic water molecules which join the bulk of surrounding water. The evolution of the rms values of the backbone (ii) during the production run is stable and allows us to use those 400 ps to calculate different structural properties (see below). To be more specific, from the beginning to the end of the 400 ps production runs, the rms deviations (ii) change from 0.68 to 0.75 Å for the S65T mutant and from 0.87 to 0.91 Å for the wild-type GFP. The ranges of the rms deviation values are thus comparable to each other, and they are of the same order of magnitude as the ones published for the C^{α} atoms

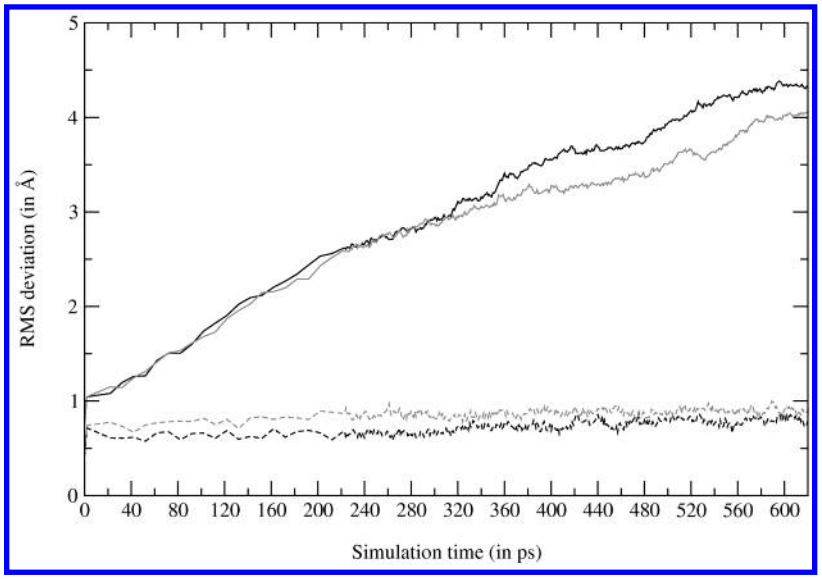

Figure 4. RMS deviations of the simulated S65T mutant (black lines) and the wild-type GFP (grey lines) from 1ema and 1emb X-ray structures, respectively, during the different steps of the MD: equilibration of the water bulk (0−2 ps), heating (2−22 ps), NVT equilibration (22−222 ps), and production run (222−622 ps). The rms deviations of the heavy atoms (proteins + X-ray water molecules) are shown with continuous lines, and dashed lines are used for the rms deviations of only the backbone atoms.

over a 1 ns MD simulation by Helms et al.[28] In fact, our systems remain slightly closer to the X-ray structures because the average rms values for the C^{α} atoms are 0.72 Å for the mutant and 0.83 Å for the wild type, whereas Helms et al. report rms deviations between 0.8 and 1.1 Å for the C^{α} atoms of the wild type during their longer 1 ns simulation.

C. Atomic Mean Square Fluctuations. The experimental *B* factors from X-ray diffraction can be converted to atomic rms fluctuations:[50]

$$\langle \Delta r_j^2 \rangle^{1/2} = \sqrt{\frac{3}{8\pi^2} B_j} \qquad (2)$$

where $\langle \Delta r_j^2 \rangle^{1/2}$ and B_j denote the root-mean-square fluctuations and the temperature factor of atom *j*, respectively. The rms atomic fluctuations can also be calculated over an MD simulation. In general, comparison between rms fluctuations obtained from eq 2 and from MD can only be qualitative.[47,51] It may serve as a check whether the MD simulation correctly reproduces the difference in mobility of the different parts of the molecule as given by the crystallographic data.

For both systems, the value of $\langle \Delta r_j^2 \rangle$ of each heavy atom of the backbone was calculated with respect to its average position during the MD simulations. Figure 5 shows the weighted average (atomic weight) per residue. The values obtained for both the mutant and the wild-type species are plotted as well as those calculated according to eq 2 from the experimental *B* values given in the PDB files. The theoretical atomic rms fluctuations are somewhat smaller than the experimental ones, as also found previously.[28] Even so, each MD simulation reproduces the global trends well; that is, the residues that show the highest mobility during the simulations are also the ones with the highest experimental temperature factors. The residues that fluctuate most during the MD simulations are those at the two extremities of the β barrel; they belong to loops or random coils. The chromophore and the residues constituting the barrel are much less mobile. Overall, our MD simulations reproduce the rigidity of the β barrel very well.

Interestingly, not a single water molecule from the bulk enters the β barrel during the MD simulations showing how well the inside of the barrel is protected from the solvent. In the barrel, the water molecules are not very mobile. In both forms of the protein, seven water molecules (numbered 316, 344, 345, 346, 370, 374, and 395 in 1ema and 22, 14, 3, 7, 19, 27, and 12 in 1emb) surrounding the chromophore are of particular interest because they are involved in a very stable hydrogen-bonding network. If we disregard number 374, the oxygen atoms of the six others exhibit fluctuations which remain below 0.92 Å. In addition, for both systems, the average position of the oxygen atom of each of those six molecules during the simulations deviates by less than 1.0 Å from its position in the X-ray structure. The water numbered 374 in the 1ema convention is a little bit more mobile than the others in the mutant species (fluctuations of 1.63 Å and deviation of 1.26 Å from the X-ray position) but rather fixed in the wild-type protein (fluctuation of 0.60 Å and deviation of 0.54 Å from the X-ray position). This difference between the two structures is a consequence of the difference between the two hydrogen bond networks around the chromophore (see below). The other water molecules inside of the barrel are also rather immobile, in both simulations. These results on the mobility of the water molecules and the shielding of the β barrel from the outside solvent environment are consistent with previous findings.[1,28]

D. Chromophore Geometry: Bond Lengths and Bond Angles. Table 8 compares bond lengths averaged over the MD production runs to the corresponding DFT and X-ray values. The results are given for both the mutant and wild-type species. The distances between hydrogen and heavy atoms are not discussed here because they are kept constant and equal to their default value in the force field due to the use of the SHAKE restraints. The rms deviation between X-ray and averaged MD bond lengths is equal to 0.07 Å for the mutant (anionic chromophore) and 0.05 Å for the wild-type protein (neutral chromophore). Most of the bond lengths are well reproduced (error $\leq$ 0.03 Å), but for some of them, the difference between MD and X-ray values is quite high and significantly exceeds the standard deviation of the MD value (bold numbers in Table 8). One extreme case is the distance between Cz and Oh for both forms of the chromophore; it is 1.25 Å (MD) instead of 1.48 Å (X-ray) for the mutant and 1.40 Å (MD) instead of 1.52 Å (X-ray) for the wild type. The MD values are in line with the DFT bond lengths calculated on the model compound, i.e.,

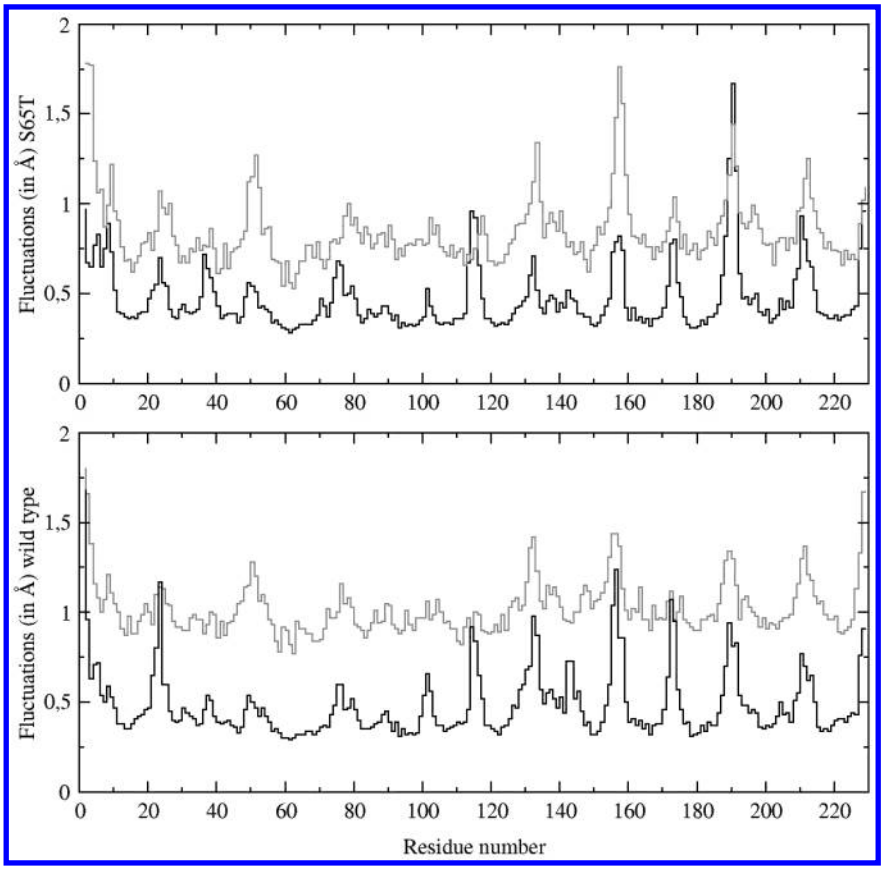

Figure 5. Atomic isotropic fluctuations averaged by residue (black lines) during the MD simulation (400 ps) and compared to the experimental values (grey lines) derived from the X-ray *B* factors.

TABLE 8: Bond Lengths (Å) of the Anionic and Neutral Chromophore

	anion			neutral		
bond	MD[a]	DFT[b]	X-ray[c]	MD[a]	DFT[b]	X-ray[d]
C1−C$^{\alpha}$t/s	1.51(0.03)	1.49[e]	1.51	1.52(0.03)	1.49[e]	1.49
N3−C$^{\alpha}$g	1.47(0.03)	1.43[f]	1.45	1.49(0.03)	1.45[f]	1.44
N3−C1	1.40(0.02)	1.39	**1.34**	1.40(0.03)	1.39	1.38
N3−C2	1.41(0.03)	1.42	1.38	1.41(0.03)	1.41	1.42
C2−O2	1.24(0.02)	1.24	1.20	1.22(0.02)	1.22	1.22
C1−N2	1.31(0.02)	1.30	**1.39**	1.30(0.03)	1.30	**1.46**
C2−Ca	1.47(0.03)	1.46	1.48	1.50(0.02)	1.49	1.49
N2−Ca	1.40(0.02)	1.40	1.41	1.40(0.02)	1.40	**1.44**
Ca−Cb	1.40(0.02)	1.39	**1.35**	1.37(0.02)	1.36	1.35
Cb-Cg	1.42(0.02)	1.41	**1.49**	1.47(0.03)	1.45	1.49
Cg-Cd1	1.45(0.03)	1.44	1.41	1.41(0.03)	1.41	1.40
Cg-Cd2	1.45(0.03)	1.43	1.45	1.41(0.03)	1.41	1.42
Cd1−Ce1	1.37(0.03)	1.37	**1.43**	1.40(0.03)	1.39	**1.44**
Cd2−Ce2	1.37(0.03)	1.37	**1.41**	1.40(0.03)	1.39	1.40
Ce1−Cz	1.45(0.03)	1.46	1.44	1.40(0.03)	1.40	**1.43**
Ce2−Cz	1.46(0.03)	1.46	**1.39**	1.40(0.03)	1.40	1.40
Cz-Oh	1.25(0.02)	1.25	**1.48**	1.40(0.03)	1.36	1.52

[a] Average values from 400 ps MD production runs, standard deviations are given in parentheses. [b] DFT(B3LYP/6-31G*) optimized geometry of the isolated molecule. [c] 1ema. [d] 1emb. [e] Bond C1−Ct1 in the model compounds. [f] Bond N3−Ct2 in the model compounds.

1.25 and 1.36 Å, respectively. In this case, it is clear that the bond lengths calculated from the X-ray structures are much longer than usual C═O or C−O bonds. Indeed, a survey of nine X-ray structures of other GFP variants found in the Protein Data Bank (1emc,[11] 1eme,[11] 1emk,[11] 1eml,[11] 1emm,[11] 1emg,[13] 1gfl,[7] and 2yfp[12]) gives values between 1.26 Å (2yfp) and 1.40 Å (1eml) for the Cz−O bond length. A screening of the Cambridge Data Base for molecules of the type para−X−Phe−OH (X being an aliphatic chain) yields 504 hits with Cz−O bond lengths between 1.32 and 1.42 Å, the average being 1.37 Å. Hence, there are no documented examples of a C−O bond as long as the one found in the X-ray files (1ema and 1emb). Likewise, a recently published CPMD study of the chromophore[26] reports values of 1.284 and 1.382 Å for optimized geometries of the anionic and neutral forms, respectively, which are close to our values. We conclude from these comparisons that the Cz−Oh distances from the X-ray structures 1ema[8] and 1emb[9] are not realistic and should be reexamined.

Such large deviations in the bond lengths of the chromophore are not found between the current DFT and MD calculations, and the corresponding rms deviations are therefore only 0.014 Å for the anionic form and 0.017 Å for the neutral form. Hence, the overall agreement between the DFT and MD structures is much better than that between the X-ray and MD structures. Similar results are found for the bond angles in the chromophore: the rms deviations for the S65T mutant (anion) and the wild type (neutral) respectively are 4.34° and 3.72° for MD vs X-ray, compared with 1.46° and 1.59° for MD vs DFT.

E. Chromophore Geometry: Dihedral Angles of the Bridging Bonds. The MD simulation of the wild-type protein (neutral chromophore) yields an average dihedral angle between the atoms N2−Ca−Cb−Cg of 2.3°, with a standard deviation of 4.8°. For Ca−Cb−Cg−Cd1, the average value is −2.2° (std dev. = 7.7°). These values are in good agreement with the X-ray structure where the corresponding dihedrals are found to be +7.1° and −5.7°, respectively. In the anionic species, the agreement is not satisfactory: from the 400 ps simulations, we find average values (standard deviations) for these two dihedral angles of 3.5°(6.1°) and 0.7°(7.1°), whereas the X-ray structure (1ema) gives +17.6° and −13.0°. These X-ray values are quite large considering the heights of the rotational barriers around

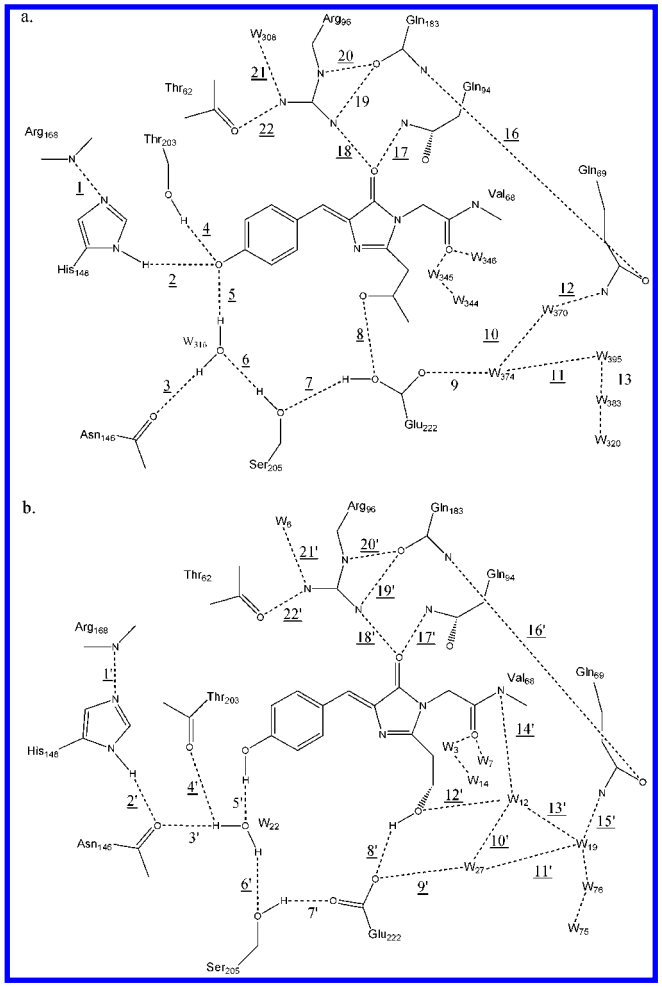

Figure 6. Hydrogen bonding network around (a) the anionic chromophore in the S65T mutant and (b) the neutral chromophore in the wild-type protein. For clarity, hydrogen atoms are not always shown.

these bonds which have partial double bond character. It seems unlikely that the protein lowers these barriers enough to favor such a large dihedral angle. It should also be noticed that none of the other published X-ray structures of GFP exhibit such high dihedral angles. Their average absolute values obtained from eight pdb files (1emc, 1eme, 1emk, 1eml, 1emm, 1gfl, 1emg, and 2yfp) are 1.85° for N2−Ca−Cb−Cg (minimum 0.20° in 1emc, maximum 4.50° in 1gfl) and 3.15° for Ca−Cb−Cg−Cd1 (min. 0.14° in 1emc, max. 6.98° in 1gfl). Hence, the deviation from planarity is always much smaller than in 1ema. Under these circumstances, we prefer not to draw any firm conclusion from this discrepancy between simulations and X-ray data (1ema).

F. Hydrogen Network between the Chromophore and Its Surrounding. From a visual inspection of the trajectories of the MD simulations and the measurement of distances between hydrogen bond donors and acceptors in the surrounding of the chromophore (the analysis was restricted to this area), we are able to identify hydrogen bonds that persisted throughout simulations. These interactions are summarized in Figure 6 which shows the hydrogen bond network in the S65T mutant and the wild-type protein. Table 9 compares characteristic distances from the X-ray structure to average values obtained from the MD simulations; we report the average distance between H-bond acceptor and H-bond donor (heavy atoms partners in the hydrogen bond) as well as the standard deviation in parentheses. The indices used in the tables refer to the numbers in Figure 6. Bold characters are used for distances characterizing the hydrogen bonds that directly involve one atom of the chromophore, more precisely one of the 19 (anionic chromophore) or 20 (neutral form) atoms forming the rings and the bridging bonds. Focusing first on the direct interactions of the chromophore, we find the same hydrogen bonds as described in the literature.[8,9,13,28] There are five hydrogen bonds directly involving the anionic chromophore in the simulation of S65T, and they are numbered 2 (Oh and Nδ from His 148), 4 (Oh and Oγ of Thr 203), 5 (Oh and water 316), 17 (O2 and Nϵ2 from Gln94), and 18 (O2 and Nh2 from Arg96) in Figure 6a. The agreement between the distances measured from the X-ray structure and the average MD values is quite good. For interactions 2, 4, 17, and 18, the deviation Δ (0.18, 0.10, 0.05, and 0.03 Å) is below the standard deviation of the MD (0.24, 0.14, 0.22, and 0.09 Å). In the case of interaction 5, the average distance between Oh and W316 (3.20 Å) is larger than the X-ray value (2.70 Å), but the standard deviation is quite high (0.43 Å), showing a high mobility of the water molecule in this pocket, and indeed the X-ray value is often reached during the simulation (17.5% of the conformations stored exhibit distances shorter or equal to 2.70 Å). We also believe that the Cz−Oh distance is overestimated in the X-ray structure by at least 0.2 Å (see section 4.D), leading to a misplacement of the oxygen atom that will contribute to the observed deviation.

In the wild-type form, only three hydrogen bonds involve Oh and O2 directly, and they are numbered 5′ (Oh and water W22), 17′ (O2 and Nϵ2 from Gln94), and 18′ (O2 and Nh2 from Arg96). Interaction 17′ is very well reproduced by the MD (Δ = 0.03 Å), whereas interaction 18′ appears to be underestimated by the simulation (Δ = 0.32 Å, std dev. = 0.09 Å). However, the structure published of the wild-type dimer (pdb code 1gfl[7]) shows a distance of 2.71 (3.08) or 2.67 Å (2.96 Å) for interaction 18′ (17′) depending on which monomer (chain A or B) is considered (see Table 9b). Those values are much closer to what we obtain from the MD simulations. Interaction 5′ between Oh and the water W22 is again problematic, even though the deviation (Δ = 0.2 Å) between the X-ray and MD distances is smaller than in the anionic case.

We now consider all of the hydrogen bonds shown in Figure 6. We generally find the same hydrogen bonds as reported in the X-ray studies.[8] In an overall assessment, most of the average distances obtained from MD are in good agreement with the X-ray values (see Table 9). Concerning the wild-type protein, the average deviation Δ from the X-ray data (1emb) is 0.22 Å. Some individual discrepancies are significantly above the standard deviation of the MD values (see Table 9b). As in the case of interaction 18′ (see above), the differences Δ between X-ray and MD results for the interactions 2′, 5′, 8′, 9′, and 15′ are much smaller when referring to pdb structure 1gfl (0.06, 0.12, 0.06, 0.01, and 0.05 Å) rather than to 1emb (0.50, 0.24, 0.39, 0.60, and 0.32 Å). In fact, concerning the overall structure of the chromophore and its surrounding, the MD simulation is clearly closer to structure 1gfl (average error 0.14 Å with chain A and 0.15 Å with chain B) than 1emb.

In the case of the S65T mutant, there are large discrepancies between the MD and X-ray results for the hydrogen bonds numbered 7, 8, 9, and 11, in the vicinity of Glu222 (see Table 9a). It should be pointed out first that the MD distances for interactions 9 and 11 are not meaningful because there is actually an exchange at about t = 180 ps between water molecules W395 and W374 (see Figure 6a). Comparing the situation before and after this exchange, the average distance between the oxygen atoms of Glu222 and W374 increases from 2.71 (0.16) to 5.16 (0.54) Å, whereas that between the oxygen atoms of Glu222

TABLE 9: Distances (in Å) between H-Bond Acceptors and H-Bond Donors Involved in the Interactions Shown in Figure 6

a. Simulation of Mutant S65T vs X-ray(1ema)

H-bond #	X-ray[a]	MD[b]	$\Delta = \lvert(\text{X-ray})-(\text{MD})\rvert$[c]
1	3.18	3.08(0.14)	0.10
2	**2.85**	**3.03(0.24)**	**0.18**
3	2.85	2.85(0.22)	0.00
4	**2.67**	**2.77(0.14)**	**0.10**
5	**2.70**	**3.20(0.43)**	**0.50**
6	2.70	2.85(0.16)	0.15
7	3.72	2.81(0.13)	0.91
8	2.81	3.94(0.35)	1.13
9	2.50	4.05(1.30)	1.55
10	2.80	2.79(0.14)	0.01
11	2.80	4.22(0.55)	1.42
12	2.80	3.03(0.23)	0.23
13	3.20	3.20(0.48)	0.00
16	3.09	2.92(0.17)	0.17
17	**3.03**	**3.08(0.22)**	**0.05**
18	**2.73**	**2.70(0.09)**	**0.03**
19	2.84	2.84(0.13)	0.00
20	2.94	2.77(0.11)	0.17
21	2.82	2.84(0.12)	0.02
22	2.74	2.71(0.09)	0.03

b. Simulation of Wild-Type vs X-ray
(1emb, 1gfl from chain A and 1gfl from chain B)

	X-ray[a]				$\Delta = \lvert(\text{X-ray})-(\text{MD})\rvert$[c]		
H-bond#	1emb	1gfl[d]	1gfl[e]	MD[b]	1emb	1gfl[d]	1gfl[e]
1′	3.19	2.94	2.98	3.09(0.16)	0.10	0.15	0.11
2′	2.69	3.20	3.25	3.19(0.33)	0.50	0.01	0.06
3′	3.16	2.86	2.95	3.18(0.41)	0.02	0.32	0.23
4′	2.98	3.44	3.43	3.03(0.34)	0.05	0.41	0.40
5′	**2.52**	**2.63**	**2.64**	**2.76(0.12)**	**0.24**	**0.13**	**0.12**
6′	2.59	2.70	2.60	2.87(0.25)	0.28	0.17	0.27
7′	2.60	2.74	2.81	2.71(0.15)	0.11	0.03	0.10
8′	3.05	2.73	2.60	2.66(0.12)	0.39	0.07	0.06
9′	3.35	2.66	2.76	2.75(0.18)	0.60	0.09	0.01
10′	3.02	3.02	2.78	3.18(0.42)	0.16	0.16	0.40
11′	2.60	2.79	2.80	2.87(0.22)	0.27	0.08	0.07
12′	2.66	2.71	2.75	2.90(0.26)	0.24	0.19	0.15
13′	5.06	5.21	5.14	4.73(0.60)	0.33	0.48	0.41
14′	2.86	2.86	2.93	3.11(0.26)	0.25	0.25	0.18
15′	2.70	3.01	3.07	3.02(0.25)	0.32	0.01	0.05
16′	3.17	3.12	3.14	3.22(0.29)	0.05	0.10	0.08
17′	**3.01**	**2.96**	**3.08**	**3.04(0.20)**	**0.03**	**0.08**	**0.04**
18′	**3.03**	**2.67**	**2.71**	**2.71(0.09)**	**0.32**	**0.04**	**0.00**
19′	2.72	3.00	2.89	2.80(0.12)	0.08	0.20	0.09
20′	2.95	2.72	2.82	2.80(0.12)	0.15	0.08	0.02
21′	2.96	2.93	2.93	2.83(0.11)	0.13	0.10	0.10
22′	2.60	2.95	2.91	2.80(0.13)	0.20	0.15	0.11

[a] The values reported for the X-ray distances are slightly different from those in the publications because they were calculated directly from the Cartesian coordinates given in the PDB file. [b] Average distances are calculated over 400 ps production runs, standard deviations are given in parentheses. [c] Absolute error between X-ray and MD average values. [d] Chain A. [e] Chain B.

and W395 decreases from 4.19 to 2.87 Å. So there always exists, during the simulation, one hydrogen bond between Glu222 and one water molecule. However, even after accounting for this exchange process the hydrogen bond pattern around Glu222 is different in the MD simulation and in the X-ray structure (1ema[8]).

This led us to question the protonation state of the glutamic acid (see section 4.A). Therefore, we ran under exactly the same conditions as before, an MD simulation of 1ema with Glu222 being anionic, with other things being equal to the previous system simulated. The analysis of the trajectory shows that the overall structure of the protein is not affected significantly by this deprotonation, apart from changes in two regions (cf. Figure 6a). First, around Glu222 and W374, the average distances decrease strongly for interactions 8 (2.98 Å), 9 (2.81 Å), and 11 (3.19 Å) which are now much closer to the corresponding X-ray values from 1ema (see Table 9a; the standard deviations of the MD distances are 0.30−0.37 Å). Second, around Ser205 and W316, the interactions 3 (3.36 Å) and 6 (3.25 Å) become weaker, and there are clear discrepancies between the average MD and the X-ray structure in this region because of partial rearrangement of hydrogen bonds. It is not surprising that such deviations must arise in an MD run for an artificial system where one proton is missing at Glu222 (without putting it elsewhere). In our opinion, the results from this MD raise the possibility that Glu222 might actually be deprotonated in the S65T mutant and that the proton might be located elsewhere. We plan to address this issue through additional MD studies.

5. Conclusions

Following established procedures, CHARMM force field parameters have been derived for the neutral and anionic forms

of the chromophore of the GFP. Comparisons with ab initio and density functional calculations show that these parameters reproduce with sufficient accuracy the geometry and rotational barriers of the isolated chromophore as well as the geometries and interaction energies in complexes with water molecules. MD simulations of the wild-type protein and of the S65T mutant have been carried out to assess the quality of the parameters in the protein environment. The analysis of the trajectories highlights particular structural properties of the protein as a whole that are in line with previously published studies: the rigidity of the β barrel, the protection of its inside against solvent molecules, and the stable hydrogen bond network around the chromophore involving water molecules in a frozen-like state. Comparison between the average MD and the X-ray structures shows that the internal geometry of the chromophore is generally well reproduced, although there is a significant discrepancy for the Cz−Oh bond length (where we believe that the 1ema X-ray value is unrealistic). The MD and X-ray structures contain the same number and type of hydrogen bonds involving the chromophore. Concerning the hydrogen-bonding network in the environment of the chromophore, there is general qualitative agreement between the MD and X-ray results for the wild-type GFP with a neutral chromophore (quantitatively better for 1gfl than for 1emb). In the case of the S65T mutant with an anionic chromophore, there are qualitative differences between the MD and X-ray (1ema) networks around Glu222, and we therefore tend to believe, contrary to interpretation of the X-ray data,[9,11] and in view of conflicting assessments of the FTIR data,[15,16] that Glu222 might not be protonated in the S65T mutant. Further work is needed to resolve this issue.

In an overall assessment of our validation studies, the newly derived CHARMM parameters seem to offer a sufficiently accurate description of the chromophore of GFP and its interactions with water molecules and the protein environment. They should therefore be suitable for MM and QM/MM studies of GFP.

Acknowledgment. We thank Dr. Roland Stote and Dr. Alexander MacKerell for their help with CHARMM. Thanks are also due to Dr. Klaus Angermund, Dr. Ulrich Fleischer and Dr. Frank Terstegen for helpful discussions.

Supporting Information Available: Detailed description of the optimization of the new force field parameters for the chromophore. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509−544.

(2) Sacchetti, A.; Ciccocioppo, R.; Alberti, S. *Histol. Histopathol.* **2000**, *15*, 101−107.

(3) Prasher, D. C.; Eckenrode, V. K.; Ward, W. W.; Prendergast, F. G.; Cormier, M. J. *Gene* **1992**, *111*, 229−233.

(4) Inouye, S.; Tsuji, F. I. *FEBS Lett.* **1994**, *341*, 277−280.

(5) Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C. *Science* **1994**, *263*, 802−805.

(6) Wang, S. H. T. *Nature* **1994**, *369*, 400−403.

(7) Yang, F.; Moss, L. G.; Phillips, G. N., Jr. *Nat. Biotechnol.* **1996**, *14*, 1246−1251.

(8) Ormo, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J. *Science* **1996**, *273*, 1392−1395.

(9) Brejc, K.; Sixma, T. K.; Kitts, P. A.; Kain, S. R.; Tsien, R. Y.; Ormo, M.; Remington, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2306−2311.

(10) Wachter, R. M.; King, B. A.; Heim, R.; Kallio, K.; Tsien, R. Y.; Boxer, S. G.; Remington, S. J. *Biochemistry* **1997**, *36*, 9759−9765.

(11) Palm, G. J.; Zdanov, A.; Gaitanaris, G. A.; Stauber, R.; Pavlakis, G. N.; Wlodawer, A. *Nat. Struct. Biol.* **1997**, *4*, 361−365.

(12) Wachter, R. M.; Elsliger, M. A.; Kallio, K.; Hanson, G. T.; Remington, S. J. *Structure* **1998**, *6*, 1267−1277.

(13) Elsliger, M. A.; Wachter, R. M.; Hanson, G. T.; Kallio, K.; Remington, S. J. *Biochemistry* **1999**, *38*, 5296−5301.

(14) Chattoraj, M.; King, B. A.; Bublitz, G. U.; Boxer, S. G. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8362−8367.

(15) van Thor, J. J.; Pierik, A. J.; Nugteren-Roodzant, I.; Xie, A.; Hellingwerf, K. J. *Biochemistry* **1998**, *37*, 16915−16921.

(16) Yoo, H.-Y.; Boatz, J. A.; Helms, V.; McCammon, J. A.; Langhoff, P. W. *J. Phys. Chem. B* **2001**, *105*, 2850−2857.

(17) Bell, A. F.; He, X.; Wachter, R. M.; Tonge, P. J. *Biochemistry* **2000**, *39*, 4423−4431.

(18) Siegbahn, P. E. M.; Wirstam, M.; Zimmer, M. *Int. J. Quantum Chem.* **2001**, *81*, 169−186.

(19) Branchini, B. R.; Lusins, J. O.; Zimmer, M. *J. Biomol. Struct. Dyn.* **1997**, *14*, 441−448.

(20) Branchini, B. R.; Nemser, A. R.; Zimmer, M. *J. Am. Chem. Soc.* **1998**, *120*, 1−6.

(21) Voityuk, A. A.; Michel-Beyerle, M. E.; Rösch, N. *Chem. Phys. Lett.* **1997**, *272*, 162−167.

(22) Voityuk, A. A.; Michel-Beyerle, M. E.; Rösch, N. *Chem. Phys. Lett.* **1998**, *296*, 269−276.

(23) Voityuk, A. A.; Michel-Beyerle, M. E.; Rösch, N. *Chem. Phys.* **1998**, *231*, 13−25.

(24) Weber, W.; Helms, V.; McCammon, J. A.; Langhoff, P. W. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 6177−6782.

(25) Helms, V.; Winstead, C. L.; Langhoff, P. W. *J. Mol. Struct. (THEOCHEM)* **2000**, *506*, 179−189.

(26) Tozzini, V.; Nifosi, R. *J. Phys. Chem. B* **2001**, *105*, 5797−5803.

(27) Scharnagl, C.; Raupp-Kossmann, R.; Fischer, S. F. *Biophys. J.* **1999**, *77*, 1839−1857.

(28) Helms, V.; Straatsma, T. P.; McCammon, J. A. *J. Phys. Chem. B* **1999**, *103*, 3263−3269.

(29) Cornell, W.; Cieplak, P.; Bayly, C.; Gould, I.; Merz, K. M.; Ferguson, D.; Spellmeyer, D.; Fox, T.; Caldwell, J.; Kollman, P. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(30) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(31) Pavelites, J. J.; Gao, J. L.; Bash, P. A.; Mackerell, A. D. *J. Comput. Chem.* **1997**, *18*, 221−239.

(32) Foloppe, N.; MacKerell, A. D. *J. Comput. Chem.* **2000**, *21*, 86−104.

(33) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(34) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.7; Gaussian, Inc.: Pittsburgh, PA, 1998.

(35) MacKerell, A. D. J.; Bashford, D.; Bellott, M., Jr.; Dunbrack, R. L.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Wiorkiewicz-Kuczera, J.; Karplus, M. *FASEB J.* **1992**, *6A*, 143.

(36) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

(37) Jorgensen, W. L. *J. Phys. Chem.* **1986**, *90*, 1276−1284.

(38) Heim, R.; Cubitt, A. B.; Tsien, R. Y. *Nature* **1995**, *373*, 663−664.

(39) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.

(40) Abola, E.; Bernstein, F. C.; Bryant, S. H.; Koetzle, T. F.; Weng, J. *Crystallographic Databases-Information Content, Software Systems, Scientific Applications*; Data Commission of the International Union of Crystallography: Bonn, 1987; pp 107−132.

(41) Bernstein, F. C.; Koetzle, T. F.; Willams, G. J. B.; Meyer, E. F. J.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535−542.

(42) Brünger, A. T.; Karplus, M. *Proteins* **1988**, *4*, 148−156.

(43) Stote, R.; Dejaegere, A.; Kuznetsov, D.; Falquet, L. *http://www.ch.embnet.org/MD_tutorial/*, 1999.

(44) Brooks, C. L.; Karplus, M. *J. Chem. Phys.* **1983**, *79*, 6312−6323.

(45) Brünger, A. T.; Brooks, C. L.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 8458−8462.
(46) Brooks, C. L.; Brünger, A. T.; Karplus, M. *Biopolymers* **1985**, *24*, 843−865.
(47) Brooks, C. L.; Karplus, M.; Pettitt, B. M. *Adv. Chem. Phys.* **1988**, *71*, 1−259.
(48) Brooks, C. L.; Karplus, M. *J. Mol. Biol.* **1989**, *208*, 159−181.
(49) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(50) Willis, B. T. M.; Prior, A. W. *Thermal Vibrations in Crystallography*; Cambridge University Press: New York, 1975; pp 1−280.
(51) Stote, R.; Dejaegere, A.; Karplus, M. In *Computational Approaches to Biochemical Reactivity*; Náray-Szabó, G., Warshel, A., Eds.; Kluwer Academic Publisher: Dordrecht, The Netherlands, 1997; pp 153−198.