

Theoretical Studies of Chromophore Maturation in the Wild-Type Green Fluorescent Protein: ONIOM(DFT:MM) Investigation of the Mechanism of Cyclization

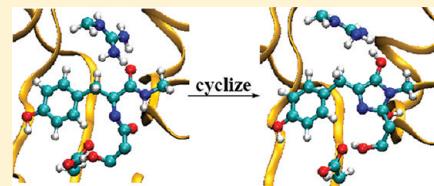
Yingying Ma,^{†,‡} Qiao Sun,^{*,‡} Zhen Li,[‡] Jian-Guo Yu,^{*,†} and Sean C. Smith^{*,‡,§}

[†]College of Chemistry, Beijing Normal University, Beijing 100875, P. R. China

[‡]Centre for Computational Molecular Science, Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, Qld 4072, Brisbane, Australia

[§]Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, United States

ABSTRACT: The availability of a gene encoding green fluorescence immediately stimulates interest in the puzzle of autocatalytic formation of the green fluorescent protein (GFP) chromophore. Numerous experimental and theoretical studies have indicated that cyclization is the first and most important step in the maturation process of the GFP. In our previous paper based on cluster models [*J. Phys. Chem. B* 2010, 114, 9698–9705], two possible mechanisms have been investigated with the conclusion that the backbone condensation initiated by deprotonation of the Gly67 amide nitrogen is easier than deprotonation of the Tyr66 α -carbon. However, the impact of the protein environment on the reaction mechanism remains to be explored. In this paper, we investigated the two possible mechanisms with inclusion of protein environmental effects by using molecular dynamics (MD) and combined quantum mechanics/molecular mechanics (QM/MM) calculations. Our calculations reveal no hydrogen bonding network that would facilitate deprotonation of the amide nitrogen of Gly67, although it is the lower energy pathway in the cluster model system. Contrastingly, there is a hydrogen bonding network between Tyr66 α -carbon and Glu222, which is in good agreement with X-ray data. The ONIOM studies show that proton transfer from Tyr66 α -carbon to Glu222 is a long-distance charge transfer process. The charge distribution of the MM region has a significant perturbation to the wave function for the QM region, with the QM energy for the proton transfer product being increased under the influence of the electrostatic protein environment. The barrier for the rate-limiting step in cyclization is quite high, about 40.0 kcal/mol in the case of ONIOM-EE.



1. INTRODUCTION

The green fluorescent protein (GFP) has become a ubiquitous tool for tagging and observing gene expression, protein localization, and cell developments.^{1–5} The GFP possesses a barrel-like structure, the wall of the barrel consisting of 11 β -sheets. The chromophore of the GFP is in the center of the barrel and linked by the α -helix which runs through the center of the barrel. It was reported that wild-type GFP absorbs light mainly at 395 nm, with a smaller peak at 475 nm, due to the neutral and anionic chromophores, respectively.^{1,6} Upon irradiation, the neutral chromophore undergoes excited state proton transfer, generating the anionic chromophore. The mechanisms of proton transfer in the GFP, as well as mechanistic and structural aspects of other fluorescent proteins, have been studied by our group among a number of others.^{7–25} The particular usefulness of GFPs and their mutants is because of their relative photostability and high quantum yield,²⁶ and all these are owing to the chromophores which are formed in an autocatalytic cyclization of a peptide from their own backbone structure that does not require a cofactor (only molecular oxygen).²⁷ It is well-known that the autocatalytic cyclization of wild-type GFP is completed by its own three amino acids Ser65, Tyr66, and Gly67. The GFP fluorescence is observed at about 90 min to 4 h after protein synthesis.^{1,28} Recently, Iizuka

et al.²⁹ studied the kinetics of de novo chromophore maturation of various GFP variants and pointed out some mutations that may promote rapid chromophore maturation.

Despite the investigations summarized above, the detailed mechanism for the formation of the chromophore in GFP is not yet clear. In 1994, Tsien's group^{1,2} proposed that the first step in chromophore formation is Gly67 amide nitrogen attack on the Ser65 carbonyl carbon, followed by dehydration and then a final oxidation step, based on expression of GFP in a wide variety of organisms and mutation experiments. Three years later, Reid and Flynn³⁰ investigated the S65T-GFP chromophore maturation process in vitro, proposing three distinct kinetic steps: first, protein folding precedes any chromophore modification and occurs fairly slowly ($k_f = 2.44 \times 10^{-3} \text{ s}^{-1}$); second, cyclization of the tripeptide with a rate constant of $3.8 \times 10^{-3} \text{ s}^{-1}$; and third, oxidation of the cyclized chromophore ($k_{ox} = 1.51 \times 10^{-4} \text{ s}^{-1}$). On the basis of the density functional calculations, Siegbahn et al.³¹ subsequently proposed that dehydrogenation of residue Tyr66 prior to cyclization (oxidized mechanism) is more favorable in energy

Received: September 10, 2011

Revised: December 14, 2011

Published: January 2, 2012

than the commonly accepted reduced mechanism (wherein cyclization precedes dehydrogenation). However, it has been suggested that their calculations are problematic because parts of their cluster models do not depend on the reliable experiment data, for example, the position and charge of Arg96.²⁷

Recently, the mechanism of GFP chromophore maturation has continued to attract attention. Two mechanisms have been proposed, respectively, by Getzoff's group^{32–37} and Wachter's group^{38–45} based on their experiments. The first mechanism, proposed by Getzoff and his co-workers,^{32–37} was called cyclization–dehydration–oxidation. The first step involves cyclization, followed by dehydration of the Ser65 carbonyl group, and finally oxidation of the C_α–C_β bond in Tyr66. Their explanation is based on crystal data,³⁶ which identifies an enolate moiety, cyclized and dehydrated but not oxidized. The experiment was carried out under anaerobic conditions and with dithionite reducing GFPsol (F64L S65T F99S M153T V163A);³⁵ however, this might not translate to the real case because the whole process of chromophore formation *in vivo* is under aerobic conditions. The second mechanism proposed by Wachter et al.^{38–45} is cyclization–oxidation–dehydration, in which the cyclized intermediate is trapped by oxidation, followed by the dehydration involving the loss of hydroxyl on the five-membered heterocycle and a proton from C_β in Tyr66. Both the crystal structure (PDB entry 1S6Z)³⁸ and mass spectroscopy data⁴⁵ support Wachter et al.'s mechanism. Maybe the strongest evidence for this mechanism is the kinetic isotope effect studies⁴⁴ which indicate that the proton on C_β in Tyr66 is not removed in the oxidation step. Meanwhile, Wachter et al. suggest two mechanisms⁴⁰ by which cyclization is mediated, namely, backbone condensation initiated by either deprotonation of the Tyr66 α-carbon or deprotonation of the Gly67 amide nitrogen.

In the formation of the chromophore, whether the second step is dehydration or oxidation may depend on the oxygen concentration and the efficiency of ring dehydration for the particular fluorescent protein variant.⁴⁶ However, the great majority of current evidence suggests that the first step is cyclization; i.e., the Gly67 amide nitrogen (N) attacks on the Ser65 carbonyl carbon (C), forming an imidazolone ring. There are two possible mechanisms for the cyclization⁴⁰ in which backbone condensation is initiated either by deprotonation of the Tyr66 α-carbon or by deprotonation of the Gly67 amide nitrogen. These two proposed mechanisms for backbone condensation in the cyclization process of GFP chromophore maturation have been investigated and compared by us⁴⁷ based on the cluster model system, which suggested that backbone condensation initiated by deprotonation of the Gly67 amide nitrogen is easier than deprotonation of the Tyr66 α-carbon. However, the cluster model does not include any protein environment; hence how inclusion of the protein environment would affect these mechanistic conclusions was not clear. It is therefore a significant and also nontrivial agenda to investigate the mechanism of chromophore maturation of the GFP with explicit inclusion of the solvated protein environment in the calculations.

2. COMPUTATIONAL METHODS

2.1. Protein Setup. The system is built based on the crystal structure (PDB entry 2AWJ,³² resolution 1.6 Å), in which six mutations are included: F64L, S65T, R96M, F99S, M153T, and V163A. These six mutations especially R96M have their

functions in maintaining the GFP in the precyclized state. We study the chromophore maturation process of the wild-type GFP, being the brightest fluorescent protein available and archetypal of the wider class of fluorescent proteins. The wild-type GFP has a mutation Q80R. It is necessary to make sure that the amino acid sequence in the precyclized state is the same as the wild-type. The details are: the 64th, 65th, 96th, 99th, 153rd, 163rd, and 80th residues in crystal structure (PDB entry 2AWJ) are changed from Leucine to Phenylalanine, from Threonine to Serine, from Methionine to Arginine, from Serine to Phenylalanine, from Threonine to Methionine, from Alanine to Valine, and from Glutamine to Arginine, respectively, using DeepView⁴⁸ software.

There are 227 amino acids in the system. All Lysine and Arginine residues are protonated, and all the Glutamic Acid and Aspartic Acid residues are unprotonated. The Histidines are neutral. The total charge of the protein is –6. To neutralize the system, 6 Na⁺ are added to the system. The simulated system is solved in a truncated octahedral box formed by 5843 water molecules of TIP3P type.⁴⁹ The solvated GFP consists of a total of 22 117 atoms, which contain 3571 solute atoms, 337 crystallographic water molecules, 5843 solvent molecules, and 6 sodium ions.

The equilibration protocol prior to the production dynamics was as follows. First, the system in vacuo was minimized for 500 steps where all the heavy atoms in the crystal structure were restrained by a force constant of 500 kcal mol^{–1} Å^{–2} and all the hydrogen atoms were allowed to relax. Such minimization was intended to release the bad contacts generated during adding hydrogen atoms, followed by 500 steps of full minimization of the atoms in vacuo without any restraint. On the third minimization of 500 steps, protein backbone atoms and crystal water were constrained by a force constant of 500 kcal mol^{–1} Å^{–2}, and such minimization was intended to release the bad contacts generated during the salvation of the system. A fourth minimization was then performed with no restrictions on the whole system for another 5000 steps.

Initially, we ran 60 ps of molecular dynamics (MD) at constant volume, and in this process our system was heated from 0 to 300 K with a weak restraint on backbone atoms and crystal water (a force constant of 10.0 kcal mol^{–1} Å^{–2}). Once our system had equilibrated over approximately 60 ps, we then switched off the restraints and changed to constant pressure before running a further 10.0 ns of equilibration at 300 K. The time step was 2 fs in both the heating process and production equilibration. A 12.0 Å cutoff was used in generating the nonbonded pair list. The Amber 9⁵⁰ was used for this kind of calculations.

2.2. QM/MM Calculations. A two-layer ONIOM (QM/MM) scheme,^{51–58} in which the interface between the QM and MM regions is treated by hydrogen link atoms,^{59,60} was then used for the calculations. The total energy of the system was calculated by the following equation

$$E^{\text{ONIOM}} = E^{\text{real,MM}} + E^{\text{model,QM}} - E^{\text{model,MM}} \quad (1)$$

Herein, $E^{\text{real,MM}}$ is the MM energy of the entire system, called the real system. $E^{\text{model,QM}}$, denoted by E_{QM} later in this study, is the QM energy of the model system, a chemically important part of the real system. $E^{\text{model,MM}}$ is the MM energy of the model system. The protein effect can be evaluated as follows: $\Delta E_{\text{MM}} = \Delta E_{\text{MM,real}} - \Delta E_{\text{MM,model}}$. If any atom is within 30 Å of the Tyr66 α-carbon, the whole residue will be included in the QM/MM calculations, and if any atom is within 20 Å of the

Tyr66 α -carbon, the whole residue is allowed to move when optimized. The QM part has 83 atoms in total, and link atoms were attached to C_A of Arg96, C_A of Glu222, C_A of Phe64, C_B of Val68, and C of Val68 (see Figure 1). In this study, the QM

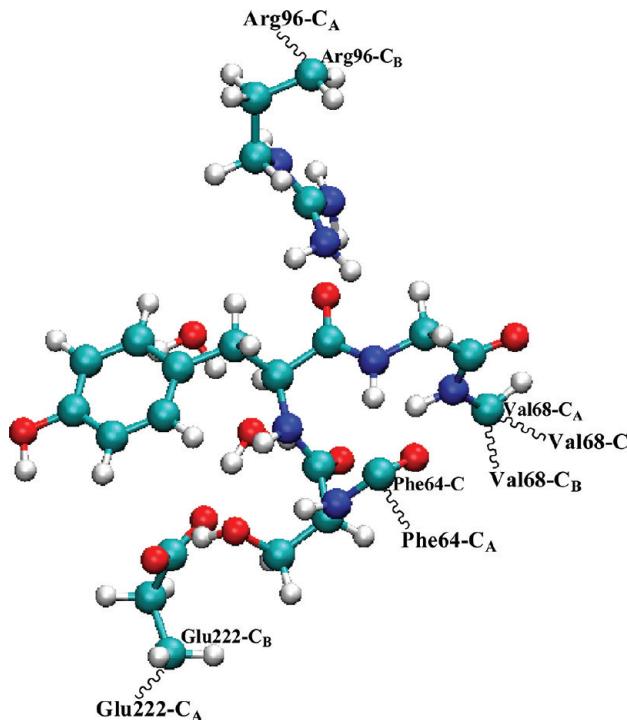


Figure 1. QM region for QM/MM calculations. Arg96- C_A , Val68- C_A , Val68- C_B , Phe64- C_A , and Glu222- C_A are treated as link atoms in DFT/MM calculations.

method was the density functional B3LYP^{61,62} with a 6-31G(d,p) basis set, and the MM method was the Amber96 force field.⁶³ Geometries were optimized in the ONIOM-ME (mechanical embedding) scheme and the ONIOM-EE (electronic embedding) scheme, respectively. Mechanical embedding treats the electrostatic interactions between the QM part and MM part at the molecular mechanics level, while electronic embedding incorporates the electrostatic interaction in the QM Hamiltonian and allows the wave function to be polarized by the charge distribution of the MM region.⁵⁸ A new quadratic coupled algorithm was used for transition state optimization.⁶⁴ The Gaussian 09 program⁶⁵ was used for these calculations.

3. RESULTS AND DISCUSSIONS

3.1. MD Results. In this part, we investigated the dynamic motions of the GFP at an atomic level by 10 ns MD simulation. Figure 2 shows the root-mean-square deviation (rmsd) between the structures in the process of MD and the starting structure of MD, as a function of simulation time. The rmsd values are for the backbone of the protein. It can be seen that the rmsd of the GFP backbone atoms remains low for the first 0.06 ns, and this is due to the restraints. Upon removing the restraints, the rmsd shot up as the GFP relaxed within the solvent. After that, our rmsd is fairly stable with no wild oscillations, which shows that our system is fairly stable in the MD simulation process. From Figure 3, it can be seen that there are some important hydrogen bonding networks around the precyclized chromophore, which may contribute to chromophore maturation. The average value

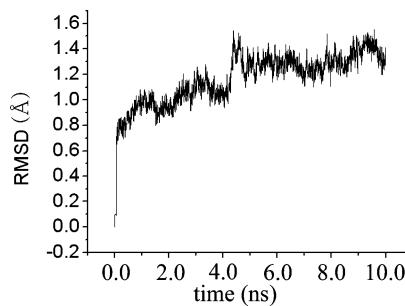


Figure 2. Root-mean-square deviations (rmsd) between the structures in the process of MD and the starting structure of MD as a function of simulation time.

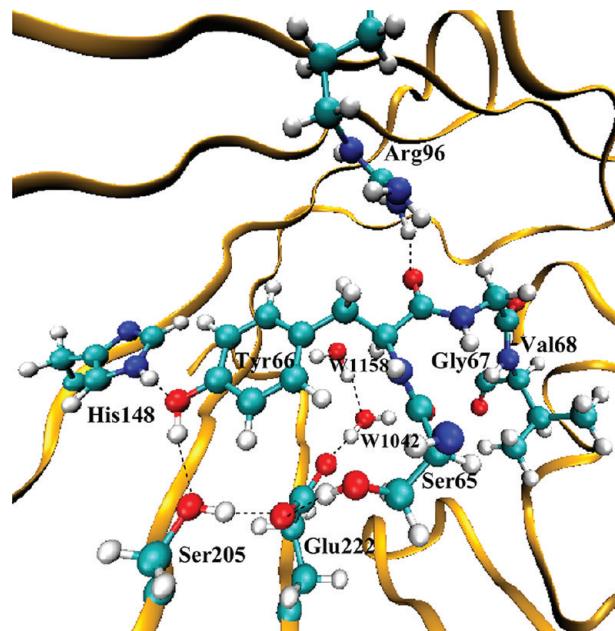


Figure 3. Hydrogen bonding network around the precyclized chromophore of GFP.

of the hydrogen bond distance between N_{H2} of Arg96 guanidinium and carbonyl oxygen of Tyr66 is 3.22 Å, which is shorter than the corresponding distance in the crystal structure (see Table 1). The reason is that there is an

Table 1. Important Geometry Parameters (in Å) around Precyclized Chromophore^a

	MD	X-ray (2AWJ)	$\Delta = (X\text{-ray}) - (MD) $
Arg96-N _{H2} _Tyr66-O	3.22(0.34)	4.39	1.17
Tyr66-O _H _His148-N _{E2}	3.20(0.41)	3.07	0.13
Tyr66-O _H _Ser205-O _G	4.25(0.60)	5.52	1.27
Ser65-O_Val68-N	3.30(0.58)	2.88	0.42
Tyr66-C _A _W1042-O _I	4.07(0.37)	4.12	0.05
Tyr66-C _A _W1158-O ₂	5.05(0.41)	5.07	0.02
W1158-O ₂ _Gln69-N _{E2}	3.02(0.14)	3.18	0.16
Gly67-N_Ser65-C	3.16(0.15)	3.03	0.13

^aAverage distances are calculated over 10.0 ns of production runs, and standard deviations are given in parentheses.

electrostatic interaction between Arg96 guanidinium and carbonyl oxygen of Tyr66. For the hydrogen bond between

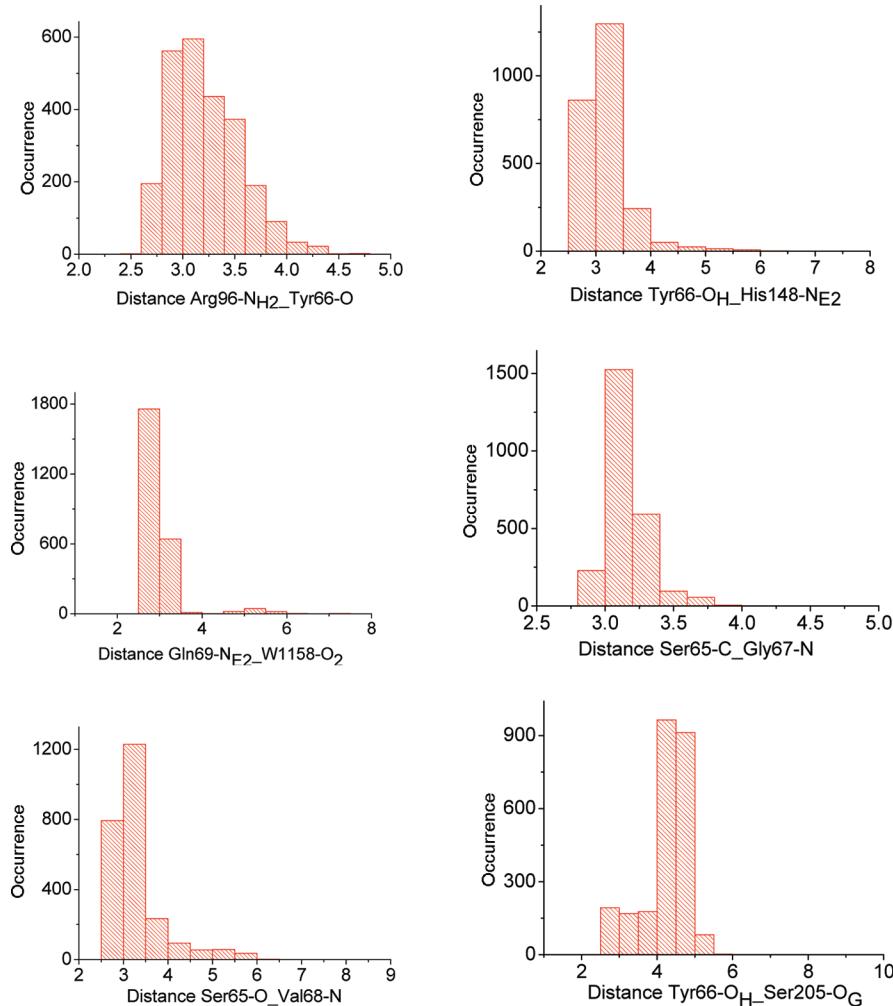


Figure 4. Histograms of frequency of occurrence of relevant distances (\AA) during MD simulation.

O_H of Tyr66 and O_G from Ser205, the average value of MD simulation is 4.25 \AA , and the experimental value is 5.52 \AA . The large difference between our result and experiment value is because there should be a crystal water molecule between O_H of Tyr66 and O_G of Ser205 as in the crystal structure of wild-type GFP (PDB entry 1GFL⁶⁶), but it does not exist in the precyclized crystal structure (PDB entry 2AWJ). So, the O_H of Tyr66 and O_G of Ser205 get close during MD simulation. For the hydrogen bond between $\text{N}_{\text{E}2}$ from His148 and O_H of Tyr66, the average value for the distance is 3.20 \AA , and the average value of the hydrogen bond between O_2 of W1158 (referring oxygen atom of W1158 as O_2) and $\text{N}_{\text{E}2}$ from Gln69 is 3.02 \AA . All these comparisons indicate that our MD results are consistent with experimental data (Figure 4 and Table 1).

In addition to the hydrogen bonding network described above, some other important hydrogen bonds near the precyclized chromophore are shown in Figure 5. The distance between atom O_G of Ser65 and atom $\text{O}_{\text{E}1}$ of Glu222 is around 2.75 \AA during the first 0.5 ns of MD simulations. The distance suddenly rises to about 3.25 \AA in the second 0.5 ns, and in the following 9.0 ns, the distance is around 5.00 \AA . Interestingly, the distance between atom O_G of Ser65 and atom $\text{O}_{\text{E}2}$ of Glu222 is around 3.25 \AA in the first 0.5 ns, and this distance suddenly reduces to about 2.75 \AA in the second 0.5 ns. In the following 9.0 ns, this distance is around 2.75 \AA (see Figure 5a). It can be inferred that at the time of 0.5 ns the carboxylate of

Glu222 inverted. This point is confirmed by Figure 5b and Figure 5c. In Figure 5b, the distance between atom O_G of Ser205 and atom $\text{O}_{\text{E}1}$ of Glu222 is around 2.65 \AA during the first 0.5 ns, and it is around 4.25 \AA during the second 0.5 ns. From 1.0 to 3.0 ns, this distance is around 3.5 \AA , and from 3.0 to 5.5 ns, this distance rises slowly to 4.0 \AA and is around this value until 10.0 ns. For the distance between O_G of Ser205 and atom $\text{O}_{\text{E}2}$ of Glu222, it is around 4.25 \AA during the first 0.5 ns, and it is around 2.65 \AA in the following 9.5 ns. In Figure 5c, the distance between atom O_1 of W1042 (referring oxygen atom of W1042 as O_1) and atom $\text{O}_{\text{E}1}$ of Glu222 carboxylate is around 4.5 \AA during the first 0.5 ns, and this distance is around 2.75 \AA during the second 0.5 ns. In the following 9.0 ns, this distance is around 4.25 \AA . By contrast, the distance between atom O_1 of W1042 and atom $\text{O}_{\text{E}2}$ of Glu222 carboxylate is around 2.75 \AA at the first 0.5 ns, and it is around 4.5 \AA at the second 0.5 ns. This distance is around 2.75 \AA in the following 9.0 ns. All these show that Glu222 has hydrogen bond interaction with Ser65, Ser205, and W1042, respectively, during the whole molecular dynamics process.

Figure 5d describes the distance between atom C_α of Tyr66 and atom O_1 of W1042 and atom O_2 of W1158, respectively, as a function of simulation time. In most cases, the former is around 4.0 \AA , and the latter is around 5.0 \AA , which is in line with the experiment value (see Table 1). So the hydrogen bonding network between C_α of Tyr66 to the Glu222 is one

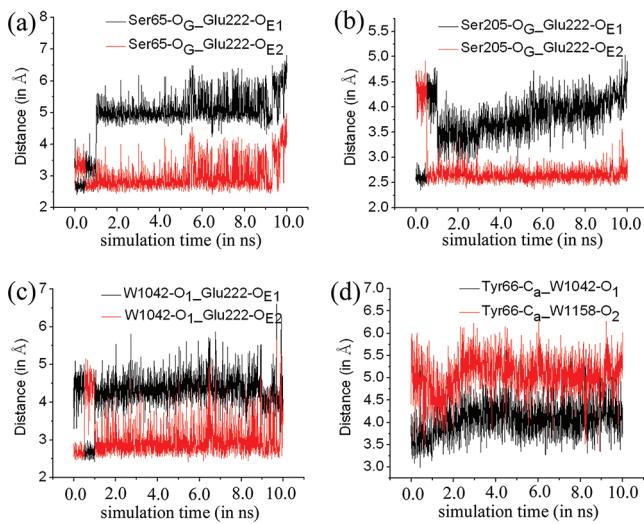


Figure 5. Geometric parameters during the molecular dynamics simulation. Distance between atom O_G of Ser65 and atom O_{E1} of Glu222 (black line) and atom O_{E2} of Glu222 (red line), respectively (a). Distance between atom O_G of Ser205 and atom O_{E1} of Glu222 (black line) and atom O_{E2} of Glu222 (red line), respectively (b). Distance between atom O₁ of W1042 and atom O_{E1} of Glu222 (black line) and atom O_{E2} of Glu222 (red line), respectively (c). Distance between atom C_a of Tyr66 and atom O₁ of W1042 (black line) and atom O₂ of W1158 (red line), respectively (d).

water molecule, which is W1042. Moreover, it can be seen that the distance between C_a of Tyr66 and atom O₁ of W1042 increases, and the distance between C_a of Tyr66 and atom O₂ of W1058 decreases at about 1.5 ns. These show that the hydrogen bonding network between C_a of Tyr66 and Glu222 may be two water molecules.

In our cluster model system, two mechanisms were investigated: cyclization initiated by deprotonation of the Tyr66 α-carbon or initiated by deprotonation of the Gly67 amide nitrogen. During our 10.0 ns MD simulation, no hydrogen bonding network was found between the Gly67 amide nitrogen and Glu222. Moreover, the Ser65, Tyr66, Gly67, and Val68 are composed of U-shaped structure, and the average distance between the carbonyl oxygen of Ser65 and amide nitrogen of Val68 is around 3.30 Å (see Table 1). In such a limited space, the chance of a water molecule seated between Gly67 amide nitrogen and Glu222 is small. So, the reaction of cyclization initiated by deprotonation of Gly67 amide nitrogen is unlikely to occur. There is a hydrogen bonding network between Tyr66 α-carbon and Glu222 which is consistent with experiment; however, in our cluster model, the energy barrier of the rate-limiting step in cyclization initiated by deprotonation of Tyr66 α-carbon is about 40.0 kcal/mol, which is too high to occur in vivo. However, our previous cluster model did not consider any protein environment. So, this time we investigate the cyclization initiated by deprotonation of Tyr66 α-carbon with the effect of protein environment.

3.2. QM/MM Results. Given the stability of the MD results, it is reasonable to choose two snapshots to investigate the mechanism of cyclization initiated by proton transfer from the Tyr66 α-carbon to Glu222 either via one or via two water molecules. For further details of this mechanism, please see Scheme 1 in our previous paper.⁴⁷

3.2.1. Proton Transfer via Two Water Molecules. Proton Transfer from Tyr66 α-Carbon to Glu222. In the ONIOM-ME

case, as for the geometry of reactant 2-Re, the H_{H2} of Arg96 guanidinium and Tyr66 carbonyl oxygen has a hydrogen bond interaction, and the distance between them is 1.62 Å (see Figure 6a). The distance between the α-carbon of Tyr66 and

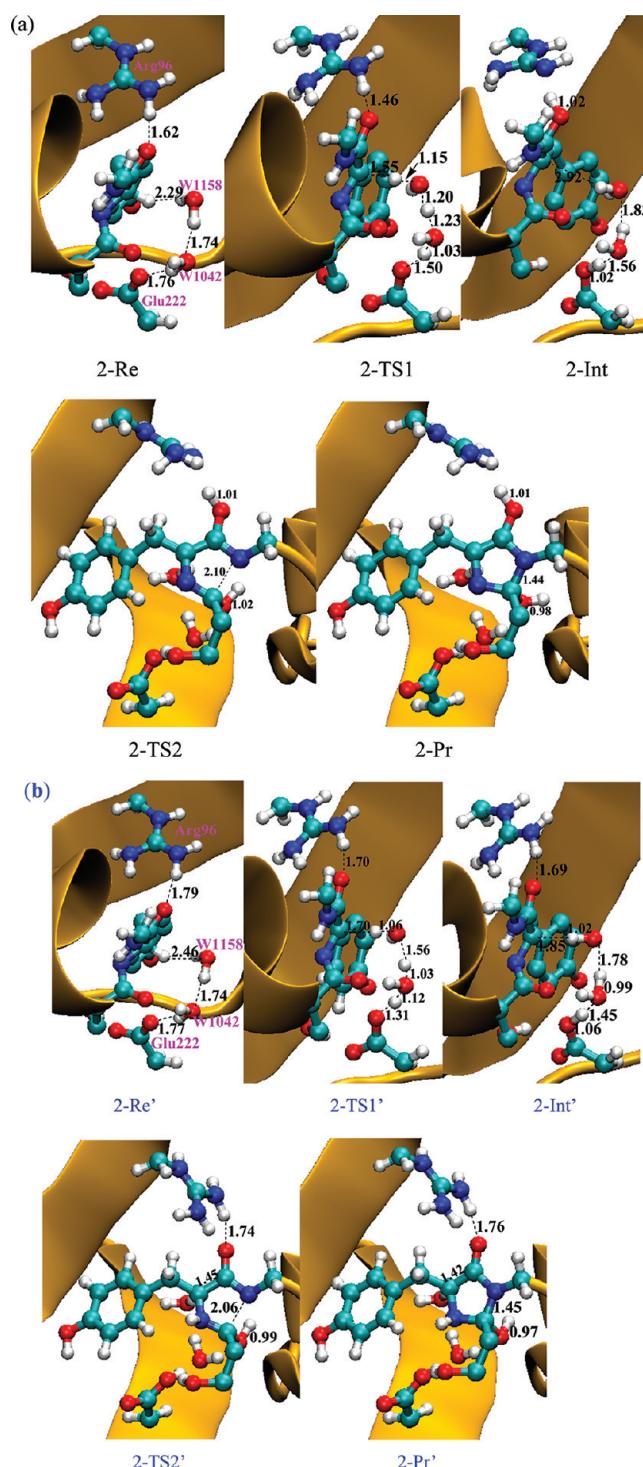


Figure 6. Geometries of reactant, transition state, and product optimized by ONIOM-ME (a) and ONIOM-EE (b), respectively, in the case of proton transfer via two water molecules.

the O₂ of W1158, the O₁ of W1042, and the O_{E2} of Glu222 carboxylate is 3.27, 2.73, and 2.72 Å, respectively. In the crystal structure, these three distances are 5.07, 2.88, and 3.20 Å, respectively, and in our cluster model system,⁴⁷ these three

distances are 3.24, 2.70, and 2.65 Å, respectively. The optimized geometries of proton transfer transition state **2-TS1** and the resulting intermediate **2-Int** are shown in Figure 6a. Our calculated results show that the proton transfer from α -carbon of Tyr66 to O_{E2} of Glu222 carboxylate via two water molecules is a concerted long-distance proton transfer process. The $C_\alpha-H_\omega$, O_2-H_2 , O_1-H_1 , and H_1-O_{E2} distances are 1.55, 1.20, 1.03, and 1.50 Å, respectively, in **2-TS1**, while in **2-Int**, the four distances are 2.92, 1.83, 1.56, and 1.02 Å, respectively. The distance between H_{H2} of Arg96 guanidinium and Tyr66 carbonyl oxygen is 1.46 and 1.02 Å, respectively, in **2-TS1** and **2-Int**. This shows that, accompanying the proton on Tyr66 α -carbon transferring to Glu222 carboxylate, the proton on Arg96 guanidinium transfers to Tyr66 carbonyl oxygen in the ONIOM-ME case. The O_G of Ser65 has hydrogen bond interaction with O_{E2} of Glu222 in **2-Int**, while it has hydrogen bond interaction with the O_{E1} of Glu222 in **2-Re** and **2-TS1**. This leads the carboxylate of Glu222 to tilt a little in **2-Int** relative to that in **2-Re** and **2-TS1**. Therefore, it can be seen that this is a conformation-coupled proton transfer process, which involves the coupling between proton transfer and conformational changes of the protein. In this case, the energies of proton transfer transition state **2-TS1** and product **2-Int** are calculated to be 19.9 and -0.2 kcal/mol, respectively, relative to the reactant **2-Re** (see Figure 8a). The corresponding relative energies of the QM part, ΔE_{QM} , for the proton transfer transition state **2-TS1** and product **2-Int** are 19.3 and 14.7 kcal/mol, respectively. So in the ONIOM-ME case, the protein effect is 0.6 and -14.9 kcal/mol, respectively, for **2-TS1** and **2-Int**. By observing the protein environment around the QM part, for example, the hydrogen bond interaction between the QM part and MM part, there is no big difference between **2-TS1** and **2-Int**. So, the large protein effect in **2-Int** is possibly because of MM electrostatic interactions.

In the ONIOM-EE case, as for the geometry of reactant **2-Re'**, the H_{H2} of Arg96 guanidinium and Tyr66 carbonyl oxygen has a hydrogen bond interaction, and the distance between them is 1.79 Å (see Figure 6b), which is much larger than the corresponding distance in the ONIOM-ME case. The distance between the α -carbon of Tyr66 and the O_2 of W1158, the O_1 of W1042, and the O_{E2} of Glu222 carboxylate is 3.41, 2.72, and 2.73 Å, respectively, which is much closer to the value in crystal structure than the corresponding distance in the ONIOM-ME case. The optimized geometries of proton transfer transition state **2-TS1'** and the resulting intermediate **2-Int'** are shown in Figure 6b. Our calculated results show that the proton transfer from the α -carbon of Tyr66 to O_{E2} of Glu222 carboxylate via two water molecules is a concerted long-distance proton transfer process. The $C_\alpha-H_\omega$, O_2-H_2 , O_1-H_1 , and H_1-O_{E2} distances are 1.70, 1.56, 1.12, and 1.31 Å, respectively, in **2-TS1'**. While in **2-Int'**, the four distances are 1.85, 1.78, 1.45, and 1.06 Å, respectively. The distance between H_{H2} of Arg96 guanidinium and Tyr66 carbonyl oxygen is 1.70 and 1.69 Å, respectively, in **2-TS1'** and **2-Int'**, which is different from the value in the ONIOM-ME case. This shows that in the ONIOM-EE case the proton on Arg96 guanidinium transferring to the Tyr66 carbonyl oxygen is not favorable by the electrostatic protein environment. The energies of **2-TS1'** and **2-Int'** are calculated to be 28.2 and 27.3 kcal/mol, respectively, higher than that of reactant **2-Re'** (see Figure 8b). It can be seen that there is a significant difference between the energy of **2-TS** and the energy of **2-TS'** and between the energy of **2-Int** and the energy of **2-Int'**, respectively. The reason is that the

proton transfer from Tyr66 α -carbon to Glu222 is a large distance charge transfer process (see Table 2), and the charge

Table 2. Atomic Mulliken Charges Are Summed for Different Moieties for Every Stationary Point along the Reaction Path in the Case of the Proton Transfer via Two Water Molecules

	2-Re	2-TS1	2-Int	2-TS2	2-Pr
Ser65	0.03	-0.02	0.00	0.26	0.28
Tyr66	-0.09	-0.55	0.03	0.11	-0.02
Gly67	-0.03	-0.07	-0.11	-0.45	-0.29
Arg96	0.88	0.84	0.10	0.11	0.12
Glu222	-0.71	-0.70	-0.03	-0.06	-0.07
W1042	-0.02	0.06	0.02	0.02	0.00
W1158	-0.04	0.49	0.04	0.03	0.03

distribution of the MM region has a large perturbation to the wave function of the QM part in this process. This can be seen from the value of ΔE_{QM} in the ONIOM-EE case. The ΔE_{QM} value for proton transfer transition state **2-TS1'** and product **2-Int'** is 37.0 and 34.2 kcal/mol, respectively, in the ONIOM-EE case, 17.7 and 19.5 kcal/mol, respectively, higher than that in the ONIOM-ME case. The wave function of the QM part in **2-Int'** is influenced much more by the electrostatic protein environment than that in **2-TS1'**. It can be seen that mechanical embedding largely neglects protein effects on the wave function of the QM part in this large distance charge transfer process, while electronic embedding qualitatively gives a better description of the electrostatic effects.

Nucleophilic Attack. The reaction of this step involves N of Gly67 attacking C of Ser65, resulting in the formation of the five-membered heterocycle. In ONIOM-ME case, the corresponding transition state **2-TS2** is shown in Figure 6a. In this process, simultaneous with formation of the C–N bond, a proton on N (H) transfers to carbonyl oxygen of Ser65 (O). The C–N distance is 2.10 Å, and the O–H distance is 1.02 Å in nucleophilic attack transition state **2-TS2**. Both distances are a little longer than the corresponding distances in the cluster model system.⁴⁷ In the product of nucleophilic attack **2-Pr**, the C–N and O–H distance are 1.44 and 0.98 Å, respectively, which show that the five-membered heterocycle ring has formed and H has transferred to O. The distance between H_{H2} of Arg96 guanidinium and Tyr66 carbonyl oxygen is 1.01 Å in both **2-TS2** and **2-Pr**. In this case, the nucleophilic attack barrier **2-TS2** is calculated to be 44.3 kcal/mol (i.e., 44.1 kcal/mol higher than that of **2-Re**), and **2-Pr** is found to lie 19.5 kcal/mol higher than **2-Re** (see Figure 8a). The energy of the QM part, ΔE_{QM} , for **2-TS2** and **2-Pr** is 62.4 and 40.6 kcal/mol, respectively, relative to **2-Re**. So in the ONIOM-ME case, the protein effect is -18.3 and -21.1 kcal/mol for **2-TS2** and **2-Pr**. Similarly, by observing the protein environment around the QM part, for example, the hydrogen bond interaction around the QM part, there is no big difference between **2-TS2**, **2-Pr**, and **2-Re**. So, the large protein effect in **2-TS2** and **2-Pr** is possibly because of MM electrostatic interactions, similar to **2-Int**.

In the ONIOM-EE case, the corresponding transition state **2-TS2'** is shown in Figure 6b. The C–N distance is 2.06 Å, and the O–H distance is 0.99 Å in nucleophilic attack transition state **2-TS2'**. Both distances are a little shorter than the corresponding distances in the ONIOM-ME case. In the product of nucleophilic attack **2-Pr'**, the C–N and O–H

distance are 1.45 and 0.97 Å, respectively. The distance between H_{H2} of Arg96 guanidinium and Tyr66 carbonyl oxygen is 1.74 and 1.76 Å, respectively, in 2-TS2' and 2-Pr', which is different from the value in the ONIOM-ME case. This confirms that in the ONIOM-EE case the proton on Arg96 guanidinium transferring to Tyr66 carbonyl oxygen is not favorable by the electrostatic protein environment. In this case, the nucleophilic attack barrier 2-TS2' is calculated to be 42.1 kcal/mol (i.e., 69.4 kcal/mol higher than that of 2-Re'), and 2-Pr' is 52.2 kcal/mol higher than that of 2-Re'. The relative energy of the QM part, ΔE_{QM} , for 2-TS2' and 2-Pr' is calculated to be 92.1 and 67.4 kcal/mol, respectively, in the ONIOM-EE case. It can be seen that the relative energy of 2-TS2' and 2-Pr' in the ONIOM-EE case is much higher than the corresponding energy in the ONIOM-ME case. Similarly, this is because a large distance charge transfer happened before nucleophilic attack, and charge distribution of the MM region has a large perturbation to the wave function of the QM part.

Comparison of the QM Region with the Previous Cluster Model System. In our ONIOM calculations, the geometries were optimized with a mechanical embedding scheme, in which the couplings between the QM region and MM region are treated with molecular mechanics calculation. In ONIOM calculations, one of its energies, ΔE_{QM} , is calculated independently. So, comparing the relative energy of the cluster model system with the corresponding energy of the QM part, ΔE_{QM} , is reasonable. The energy of 2-TS1 in the cluster model system is 18.4 kcal/mol, and the corresponding energy ΔE_{QM} is 19.3 kcal/mol; so the energies are close to each other. However, for the other three stationary points, the proton transfer product 2-Int, the transition state of nucleophilic attack 2-TS2, and the product of nucleophilic attack 2-Pr, the relative energies in the cluster model system are 2.1, 41.5, and 21.1 kcal/mol, respectively, and the corresponding values of the QM part, ΔE_{QM} , are 14.7, 62.4, and 40.6 kcal/mol, respectively. So, the difference between the relative energy of the cluster model system and ΔE_{QM} is big for the above three stationary points, which is probably due to the geometry differences between them. Figure 7 is the overlay of the cluster model system with the corresponding atoms in the QM part. It can be seen that the biggest difference between the cluster model system and the QM part appears in 2-Int, 2-TS2, and 2-Pr. First, with regard to the different geometries of Glu222, it can be seen that in 2-Int, 2-TS2, and 2-Pr (ONIOM system), because of protonation of Glu222, the H_G of Ser65 has a hydrogen bond interaction with O_{E2} of Glu222, unlike O_{E1} of Glu222 in 2-Re and 2-TS1. So, the carboxylate of Glu222 is a little tilting in 2-Int, 2-TS2, and 2-Pr relative to 2-Re. Second, regarding the different conformation of the Tyr66 phenol ring, in the cluster model system the phenol ring in 2-TS2 and 2-Pr rotates and is nearly perpendicular to the five-membered heterocycle. However, in the ONIOM system because of the hydrogen bond interaction between O_H from Tyr66 and N_{E2} from His148, O_G of Ser 205, respectively, the phenol ring is nearly in the same plane with the five-membered heterocycle. Third, regarding the relative position of W1024, for proton transfer product 2-Int the distance between Tyr66 C_α and O₁ of W1024 is 3.69 Å in the ONIOM system, 0.32 Å larger than the 3.37 Å in the cluster model system, which shows that the water molecule is released after proton transfer in the ONIOM system. Regarding the geometry in the cluster model system, to keep the calculated structures close to those obtained experimentally, some atoms were fixed to their X-ray structure

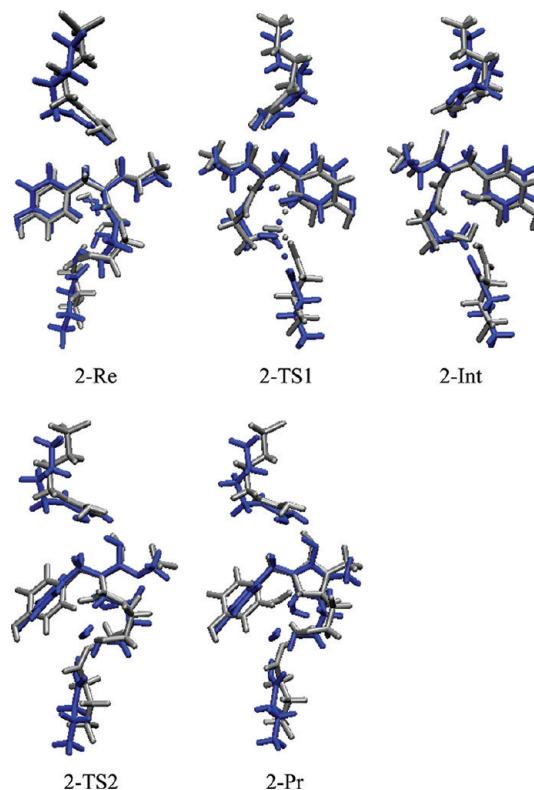


Figure 7. Overlay of the cluster model system (blue) with the corresponding atoms in the QM part (gray).

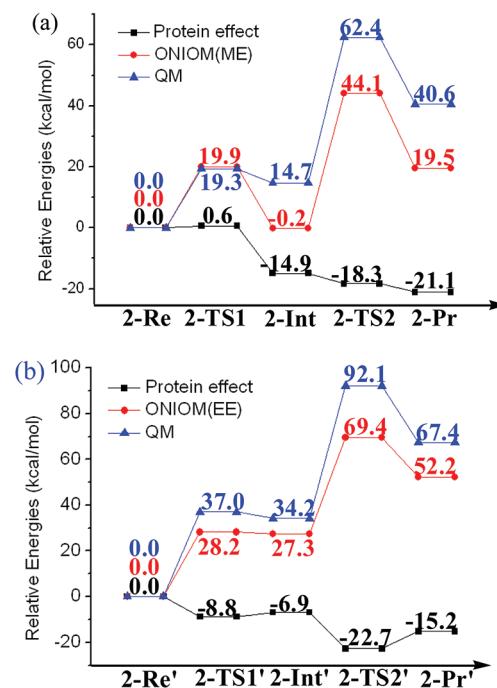


Figure 8. Potential energy diagram for the cyclization initiated by deprotonation of Tyr66 α -carbon via two water molecules. Energies of ONIOM (B3LYP/6-31G(d,p): Amber) and geometries optimized by mechanical embedding (a). Energies of ONIOM (B3LYP/6-31G(d,p): Amber) and geometries optimized by electronic embedding (b).

during geometry optimizations, precluding any large side chain movement. In QM/MM calculation, however, if any constituent atom lies within 20 Å of origin at atom C_α of Tyr66, the

whole residue will be allowed to relax. The QM/MM model made it easier to include residues close to the chromophore while still keeping the hydrogen bonds intact during five-membered heterocycle formation. The big energy difference between energy ΔE_{QM} in ONIOM calculations and the corresponding energy of the cluster model system illustrates the importance of including enough residues in the active site and a large QM part in QM/MM calculations.

3.2.2. Proton Transfer via One Water Molecule. Proton Transfer from Tyr66 α -Carbon to Glu222. In the ONIOM-ME case, for the optimized geometry of reactant **1-Re**, there is a hydrogen bond interaction between the H_{H2} of Arg96 guanidinium and Tyr66 carbonyl oxygen, and the distance between them is 1.58 Å (see Figure 9a). The distance between C_α of Tyr66 and the O_1 of W1024 is 3.30 Å, and the distance between C_α of Tyr66 and the O_2 of W1158 is 3.78 Å. So, the proton on the α -carbon of Tyr66 is transferred to O_{E2} of the Glu222 carboxylate via one water molecule under this circumstance, namely, via W1024. The three distances C_α – O_2 , O_2 – O_1 , and O_1 – O_E are 3.78, 2.81, and 2.75 Å, respectively, in **1-Re**. In the crystal structure, these three distances are 5.07, 2.88, and 3.20 Å, respectively. So, it can be seen that the geometries in the case of proton transfer via one water molecule are much closer to the crystal structure than that in the case of two water molecules. The C_α – H_w , O_1 – H_1 , and H_1 – O_E distances in reactant **1-Re** are 1.10, 0.98, and 1.78 Å, respectively, and 1.35, 1.50, and 1.06 Å, respectively, in proton transfer transition state **1-TS1** and 3.26, 1.66, and 1.00 Å, respectively, in proton transfer product **1-Int** (see Figure 9a). It can be seen that the C_α – H_α distance in **1-Int** is a little big, and the reason is that, when the new water is formed, the water rotated with H_α pointing to the phenol ring of Tyr66. This shows that it is a conformation-coupled proton transfer process, which involves the coupling between proton transfer and relocation of water molecules. About the energy of proton transfer from Tyr66 α -carbon to Glu222 via one water molecule, in the ONIOM-ME case, the barrier of proton transfer **1-TS1** is calculated to be 18.9 kcal/mol, and the energy of proton transfer product **1-Int** is –3.6 kcal/mol relative to the reactant **1-Re** (see Figure 10a). The corresponding energies of the QM part, ΔE_{QM} , are 27.0 and 19.0 kcal/mol, respectively. So in the ONIOM-ME case, the protein effect is –8.1 and –22.6 kcal/mol, respectively, for **1-TS1** and **1-Int**. As discussed before, the big protein effect is because of MM electrostatic interactions.

In the ONIOM-EE case, for the optimized geometry of reactant **1-Re'**, there is a hydrogen bond interaction between the H_{H2} of Arg96 guanidinium and Tyr66 carbonyl oxygen, and the distance between them is 1.83 Å (see Figure 9b). The distance between C_α of Tyr66 and the O_1 of W1024 is 4.05 Å, and the distance between C_α of Tyr66 and the O_2 of W1158 is 3.49 Å. So, the proton on the α -carbon of Tyr66 is transferred to O_{E2} of the Glu222 carboxylate via two water molecules in the ONIOM-EE case, which is different from the ONIOM-ME case. The C_α – H_w , O_2 – H_2 , O_1 – H_1 , and H_1 – O_E distances in reactant **1-Re'** are 1.10, 0.99, 0.98, and 1.77 Å, respectively, and 1.46, 1.56, 1.35, and 1.10 Å, respectively, in proton transfer transition state **1-TS1'** and 1.83, 1.77, 1.46, and 1.05 Å, respectively, in proton transfer product **1-Int'** (see Figure 9b). The relative energy of **1-TS1'** and **1-Int'** is 35.0 and 33.2 kcal/mol, respectively. It can be seen that there is a significant difference between the energy of **1-TS1** and the energy of **1-TS1'** and between the energy of **1-Int** and the energy of **1-Int'**.

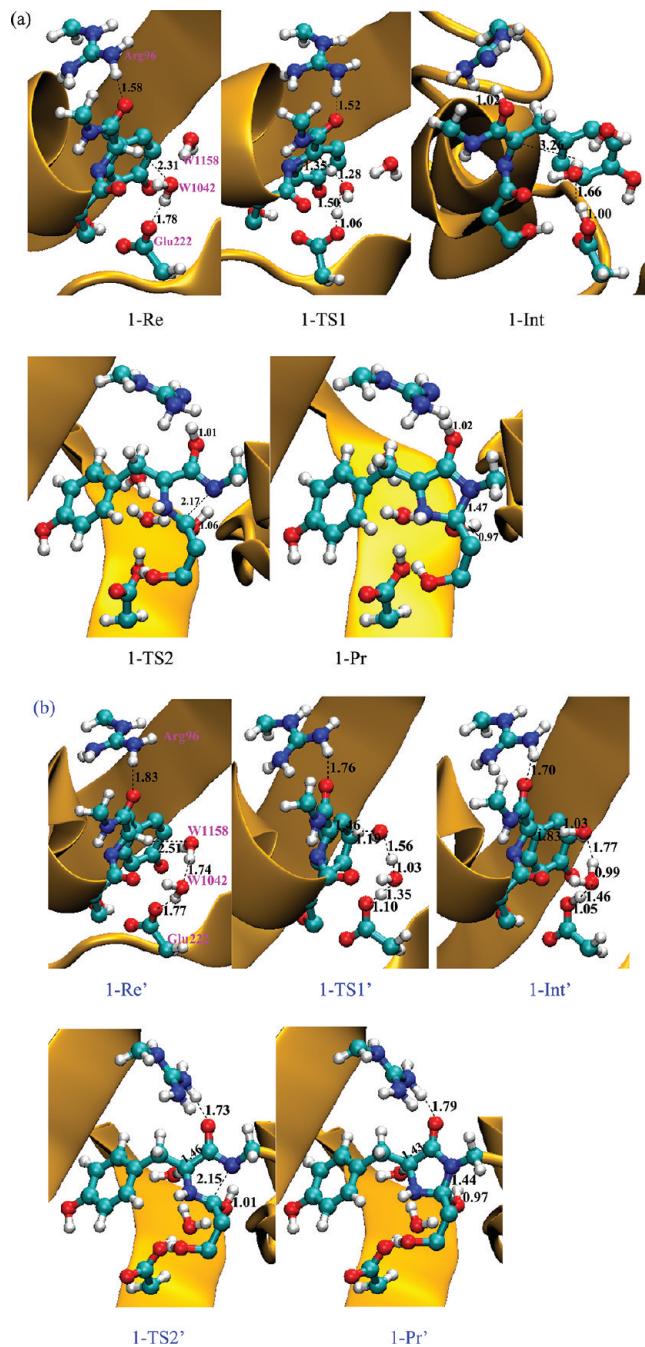


Figure 9. Geometries of reactant, transition state, and product optimized by ONIOM-ME (a) and ONIOM-EE (b) in the case of proton transfer via one water molecule.

The reason is that the proton transfer from Tyr66 α -carbon to Glu222 is a large distance charge transfer process (see Table 3), and the charge distribution of the MM region has a large perturbation to the wave function of the QM part in this process. This can be seen from the value of ΔE_{QM} in the ONIOM-EE case. The ΔE_{QM} value for **1-TS1'** and **1-Int'** is 39.3 and 39.5 kcal/mol, respectively, at the ONIOM-EE level, 12.3 and 20.5 kcal/mol, respectively, higher than that in the ONIOM-ME case. The wave function of the QM part in **1-Int'** is influenced much more by the electrostatic protein environment than that in **1-TS1'**. So, it can be seen that mechanical embedding largely neglects protein effects on wave function of the QM part in this large distance charge transfer

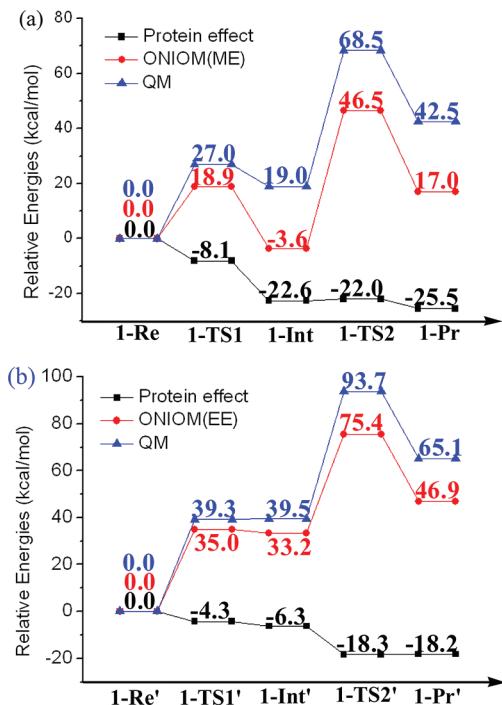


Figure 10. Potential energy diagram for the cyclization initiated by deprotonation of Tyr66 α -carbon via one water molecule. Energies of ONIOM (B3LYP/6-31G(d,p): Amber) and geometries optimized by mechanical embedding (a). Energies of ONIOM (B3LYP/6-31G(d,p): Amber) and geometries optimized by electronic embedding (b).

Table 3. Atomic Mulliken Charges Are Summed for Different Moieties for Every Stationary Point along the Reaction Path in the Case of the Proton Transfer via One Water Molecule

	1-Re	1-TS1	1-Int	1-TS2	1-Pr
Ser65	0.00	0.02	0.03	0.31	0.27
Tyr66	-0.04	-0.51	0.03	0.14	-0.01
Gly67	-0.03	-0.10	-0.14	-0.48	-0.29
Arg96	0.87	0.85	0.12	0.12	0.12
Glu222	-0.74	-0.10	-0.05	-0.07	-0.09
W1042	0.00	-0.08	0.05	0.05	0.06
W1158	-0.04	-0.07	-0.04	-0.04	-0.04

process, while electronic embedding qualitatively gives a better description of the electrostatic effects.

Nucleophilic Attack. In the ONIOM-ME case, under condition of cyclization initiated by proton transfer via one water molecule, for the geometry of nucleophilic attack transition state **1-TS2**, the C–N distance is 2.17 Å, and the O–H distance is 1.06 Å. Both distances are longer than the corresponding value in two-water-molecule conditions. For the product of nucleophilic attack **1-Pr**, the C–N distance is 1.47 Å, and the O–H distance is 0.97 Å, which shows that the proton on the Gly67 amide nitrogen has transferred to the carbonyl oxygen of Ser65. The relative energies of **1-TS2** and **1-Pr** are 46.5 and 17.0 kcal/mol, respectively, in the ONIOM-ME case (see Figure 10a). The corresponding energy of the QM part, ΔE_{QM} , is 68.5 and 42.5 kcal/mol, respectively. So in this case, the protein effect is -22.0 and -25.5 kcal/mol, respectively, for **1-TS2** and **1-Pr**. As discussed before, the large protein effect at the ONIOM-ME level is possibly from MM electrostatic interaction.

In the ONIOM-EE case, for the geometry of nucleophilic attack transition state **1-TS2'**, the C–N distance is 2.15 Å, and the O–H distance is 1.01 Å (see Figure 9b). Both distances are a little shorter than the corresponding distances in the ONIOM-ME case. For the product of nucleophilic attack **1-Pr'**, the C–N distance is 1.44 Å, and the O–H distance is 0.97 Å, which shows that the proton on the Gly67 amide nitrogen has transferred to the carbonyl oxygen of Ser65. In this case, the calculated energy of **1-TS2'** is 42.2 kcal/mol higher than that of **1-Int'** (i.e., 75.4 kcal/mol higher than that of **1-Re'**), and product **1-Pr'** is found to lie 46.9 kcal/mol higher than **1-Re'**. The relative energy of the QM part for **1-TS2'** and **1-Pr'**, ΔE_{QM} , is 93.7 and 65.1 kcal/mol, respectively. It can be seen that the relative energy of the nucleophilic attack transition state and product in the ONIOM-EE case is much higher than the corresponding energy in the ONIOM-ME case. Similarly, this is because large distance charge transfer happened before nucleophilic attack, and charge distribution of the MM region has a large perturbation to the wave function of the QM part.

To sum up, the barrier for the rate-limiting step in cyclization initiated by deprotonation of Tyr66 α -carbon in a large protein system is quite high, about 40.0 kcal/mol, not only in the case of proton transfer from the Tyr66 α -carbon to Glu222 via two water molecules but also in the case of proton transfer via one water molecule. The implication of these results is that cyclization may not be initiated by deprotonation of Tyr66 α -carbon. Combining this with the results of the MD simulation, we may plausibly infer that cyclization in chromophore maturation of wild-type GFP is neither initiated by deprotonation of Gly67 amide nitrogen nor initiated by deprotonation of Tyr66 α -carbon.

4. CONCLUSION

This paper is an extension of our previous work on cyclization in chromophore maturation of the GFP using a cluster model system, which suggested that deprotonation of Gly67 amide nitrogen is easier than deprotonation of α -carbon of Tyr66. In this paper, we investigated the two possible mechanisms of cyclization in chromophore maturation of GFP with protein environment by using MD and combined QM/MM calculations. Our results show that no hydrogen bonding network was found between Glu222 and the amide nitrogen of Gly67, indicating that deprotonation from the amide nitrogen of Gly67 may not be possible even though it is the lower-energy path way in the cluster model system. By contrast, there is a hydrogen bonding network between Tyr66 α -carbon and Glu222 which is in line with the experimental measurement. We studied the cyclization initiated by deprotonation of Tyr66 α -carbon in a large protein system in two cases: proton transfer via two water molecules and proton transfer via one water molecule. Our results show the cyclization in chromophore maturation of wild-type GFP is neither initiated by deprotonation of Gly67 amide nitrogen nor initiated by deprotonation of Tyr66 α -carbon.

Moreover, from the molecular dynamics simulation (see Table 1), the average distance between the amide nitrogen atom of Gly67 and the carbonyl carbon of Ser65 during the molecular dynamics simulations (3.16 Å) is shorter than the sum of their covalent radii (3.25 Å). The close interaction indicates the possibility that the amide nitrogen atom of Gly67 attacks the carbonyl carbon of Ser65 directly to complete the cyclization process. A detailed understanding of this step reaction requires further study. In all, this paper has provided

significant new insights toward understanding the mechanism of cyclization in chromophore maturation of GFP.

AUTHOR INFORMATION

Corresponding Author

*E-mail: jianguo_yu@bnu.edu.cn; smithsc@ornl.gov; q.sun@uq.edu.au.

ACKNOWLEDGMENTS

Y. Y. Ma thanks the Australian Institute for Bioengineering and Nanotechnology at The University of Queensland for facilitating a seven-month research visit (supported by Australian Research Council Discovery Project DP0771867) and the provision of computing resources from the Centre for Computational Molecular Science (funded in part by ARC LIEF project LE0882357) and the National Computational Infrastructure (NCI). This work was also supported by grants from the National Natural Science Foundation of China (Grant Nos. 20733002, and 20873008).

REFERENCES

- (1) Heim, R.; Prasher, D. C.; Tsien, R. Y. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12501.
- (2) Cubitt, A. B.; Heim, R.; Adams, S. R.; Boyd, A. E.; Gross, L. A.; Tsien, R. Y. *Trends Biochem. Sci.* **1995**, *20*, 448.
- (3) Orm, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J. *Science* **1996**, *273*, 1392.
- (4) Zimmer, M. *Chem. Rev.* **2002**, *102*, 759.
- (5) Giepmans, B. N. G.; Adams, S. R.; Ellisman, M. H.; Tsien, R. Y. *Science* **2006**, *312*, 217.
- (6) Niwa, H.; Inouye, S.; Hirano, T.; Matsuno, T.; Kojima, S.; Kubota, M.; Ohashi, M.; Tsuji, F. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13617.
- (7) Wang, S. F.; Smith, S. C. *Chem. Phys.* **2006**, *326*, 204.
- (8) Wang, S. F.; Smith, S. C. *J. Phys. Chem. B* **2006**, *110*, 5084.
- (9) Wang, S. F.; Smith, S. C. *Phys. Chem. Chem. Phys.* **2007**, *9*, 452.
- (10) Zhang, H.; Smith, S. C. *J. Theor. Comput. Chem.* **2007**, *6*, 789.
- (11) Zhang, H.; Wang, S. F.; Sun, Q.; Smith, S. C. *Phys. Chem. Chem. Phys.* **2009**, *11*, 8422.
- (12) Sun, Q.; Doerr, M.; Li, Z.; Smith, S. C.; Thiel, W. *Phys. Chem. Chem. Phys.* **2010**, *12*, 2450.
- (13) Vendrell, O.; Gelabert, R.; Moreno, M.; Lluch, J. M. *J. Chem. Theory Comput.* **2008**, *4*, 1138.
- (14) Vendrell, O.; Gelabert, R.; Moreno, M.; Lluch, J. M. *J. Phys. Chem. B* **2008**, *112*, 5500.
- (15) Vendrell, O.; Gelabert, R.; Moreno, M.; Lluch, J. M. *J. Am. Chem. Soc.* **2006**, *128*, 3564.
- (16) Olsen, S.; Smith, S. C. *J. Am. Chem. Soc.* **2008**, *130*, 8677.
- (17) Olsen, S.; Smith, S. C. *J. Am. Chem. Soc.* **2007**, *129*, 2054.
- (18) Olsen, S.; Smith, S. C. *Chem. Phys. Lett.* **2006**, *426*, 159.
- (19) Olsen, S.; Prescott, M.; Wilmann, P.; Battad, J.; Rossjohn, J.; Smith, S. C. *Chem. Phys. Lett.* **2006**, *420*, 507.
- (20) Reuter, N.; Lin, H.; Thiel, W. *J. Phys. Chem. B* **2002**, *106*, 6310.
- (21) Schäfer, L. V.; Groenhof, G.; Boggio-Pasqua, M.; Robb, M. A.; Grubmüller, H. *PLoS Comput. Biol.* **2008**, *4*, e1000034.
- (22) Schäfer, L. V.; Gerrit, G.; Klingen, A. R.; Ullmann, G. M.; Martial, B.-P.; Robb, M. A.; Helmut, G. *Angew. Chem., Int. Ed.* **2007**, *46*, 530.
- (23) Zhang, H.; Sun, Q.; Wang, S.; Olsen, S.; Smith, S. C. Theoretical studies of Green and Red Fluorescent Proteins. In *Hydrogen Bonding and Transfer in the Excited State*; John Wiley & Sons, Ltd: New York, 2011; p 815.
- (24) Sun, Q.; Wang, S. F.; Zhang, H.; Li, Z.; Pifisterer, C.; Fischer, S.; Nanbu, S.; Smith, S. C. *Aust. J. Chem.* **2010**, *63*, 363.
- (25) Andrensen, M.; Wahl, M. C.; Stiel, A. C.; Gräter, F.; Schäfer, L. V.; Trowitzsch, S.; Weber, G.; Eggeling, C.; Grubmüller, H.; Hell, S. W.; Jakobs, S. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13070.
- (26) Day, R. N.; Davidson, M. W. *Chem. Soc. Rev.* **2009**, *38*, 2887.
- (27) Lemay, N. P.; Morgan, A. L.; Archer, E. J.; Dickson, L. A.; Megley, C. M.; Zimmer, M. *Chem. Phys.* **2008**, *348*, 152.
- (28) Crameri, A.; W., E. A.; Tate, E.; Stemmer, W. P. C. *Nat. Biotechnol.* **1996**, *14*.
- (29) Iizuka, R.; Yamagishi-Shirasaki, M.; Funatsu, T. *Anal. Biochem.* **2011**, *414*, 173.
- (30) Reid, B. G.; F., G. C. *Biochemistry* **1997**, *36*, 6786.
- (31) Per, E. M. S.; Maria, W.; Marc, Z. *Int. J. Quantum Chem.* **2001**, *81*, 169.
- (32) Barondeau, D. P.; Hitomi, C.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D. *Biochemistry* **2005**, *44*, 16211.
- (33) Barondeau, D. P.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D. *J. Am. Chem. Soc.* **2007**, *129*, 3118.
- (34) Barondeau, D. P.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D. *J. Am. Chem. Soc.* **2006**, *128*, 4685.
- (35) Barondeau, D. P.; Putnam, C. D.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12111.
- (36) Barondeau, D. P.; Tainer, J. A.; Getzoff, E. D. *J. Am. Chem. Soc.* **2006**, *128*, 3166.
- (37) Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D. *Biochemistry* **2005**, *44*, 1960.
- (38) Huffman, H. A.; Phail, M. E.; Wachter, R. M. *Biochemistry* **2004**, *43*, 4464.
- (39) Patel, H. N.; Wachter, R. M. *Biochemistry* **2005**, *44*, 8303.
- (40) Sniegowski, J. A.; Lappe, J. W.; Patel, H. N.; Huffman, H. A.; Wachter, R. M. *J. Biol. Chem.* **2005**, *280*, 26248.
- (41) Sniegowski, J. A.; Phail, M. E.; Wachter, R. M. *Biochem. Biophys. Res. Commun.* **2005**, *332*, 657.
- (42) Wachter, R. M. *Acc. Chem. Res.* **2007**, *40*, 120.
- (43) Wachter, R. M. Mechanistic aspects of GFP chromophore biogenesis. *Genetically Engineered Probes for Biomedical Applications*; SPIE: San Jose, CA, USA, 2006.
- (44) Zhang, L.; Chan, N. H.; Dorrestein, P. C.; Wachter, R. M. *Biochemistry* **2008**, *47*, 10111.
- (45) Zhang, L.; Patel, H. N.; Lappe, J. W.; Wachter, R. M. *J. Am. Chem. Soc.* **2006**, *128*, 4766.
- (46) Craggs, T. D. *Chem. Soc. Rev.* **2009**, *38*, 2865.
- (47) Ma, Y. Y.; Sun, Q.; Zhang, H.; Peng, L.; Yu, J.-G.; Smith, S. C. *J. Phys. Chem. B* **2010**, *114*, 9698.
- (48) Guex, N.; Peitsch, M. C. *Electrophoresis* **1997**, *18*, 2714.
- (49) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- (50) A., C. D.; A., D. T.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, H.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. *AMBER 9*; University of California: San Francisco, 2006.
- (51) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170.
- (52) Hembel, S.; Sieber, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 1959.
- (53) Matsubara, T.; Sieber, S.; Morokuma, K. *Int. J. Quantum Chem.* **1996**, *60*, 1101.
- (54) Svensson, M.; Hembel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357.
- (55) Svensson, M.; Hembel, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 3654.
- (56) Vreven, T.; Morokuma, K. *J. Comput. Chem.* **2000**, *21*, 1419.
- (57) Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *THEOCHEM* **1999**, *461–462*, 1.
- (58) Vreven, T.; Byun, K. S.; Komáromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J. *J. Chem. Theory Comput.* **2006**, *2*, 815.
- (59) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718.
- (60) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.
- (61) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

- (62) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- (63) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- (64) Vreven, T.; Frisch, M. J.; Kudin, K. N.; Schlegel, H. B.; Morokuma, K. *Mol. Phys.* **2006**, *104*, 701.
- (65) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision A.02; Gaussian, Inc.: Wallingford CT, 2009.
- (66) Yang, F.; Moss, L. G.; Phillips, G. N. Jr. *Nat. Biotechnol.* **1996**, *14*, 1246.