

---

# Theoretical Study of the Mechanism of Peptide Ring Formation in Green Fluorescent Protein

---

PER E. M. SIEGBAHN,<sup>1</sup> MARIA WIRSTAM,<sup>1</sup> MARC ZIMMER<sup>2</sup>

<sup>1</sup>*Department of Physics, Stockholm University, Box 6730, S-113 85 Stockholm, Sweden*

<sup>2</sup>*Chemistry Department, Connecticut College, New London, Connecticut 06320*

*Received 8 May 2000; revised 26 May 2000; accepted 20 July 2000*

---

**ABSTRACT:** Density functional calculations using hybrid functionals (B3LYP) have been performed to study the mechanism of peptide ring formation in green fluorescent protein (GFP). Several different chemical models were used ranging from a minimal model of the ring formation to a model including the full side chains of the groups involved in forming the peptide ring. The surrounding protein was described using a dielectric cavity model. The previously most accepted mechanism was found to lead to an endothermic cyclization of about 10 kcal/mol, independent of chemical model used. The formation of the required dihydro-imidazolone intermediate was found to be even more endothermic with 16–18 kcal/mol. In contrast, another mechanism where the dehydration of residue 66 precedes cyclization was found to be exothermic by 1.9 kcal/mol and to go over an endothermic intermediate of only 6.7 kcal/mol. Correcting these results using the more accurate G2-M scheme leads to an intermediate with an energy of only +3.7 kcal/mol and an overall exothermicity of 4.7 kcal/mol. Possible transition states involving proton transfer steps were also investigated. Comparisons are made to the similar and more well-known deamination reaction of Asn-Gly sequences in peptides, for which good agreement is obtained with experiments. The results are discussed with respect to available experiments. © 2001 John Wiley & Sons, Inc. *Int J Quantum Chem* 81: 169–186, 2001

**Key words:** B3LYP; G2; mechanism; enzyme

---

## Introduction

**G**reen fluorescent proteins (GFP) are of great interest due to their application to *in situ* monitoring of gene expression, protein movement, and

cell development [1–4]. GFP-tagged proteins can be monitored noninvasively in living cells by flow cytometry, fluorescence microscopy [5], or macroscopic imaging methods. Chromophore formation in GFP does not require any additional factors, the protein does not interfere with cell growth and function, and it is not toxic to cells. GFP has been expressed in a wide variety of organisms such as

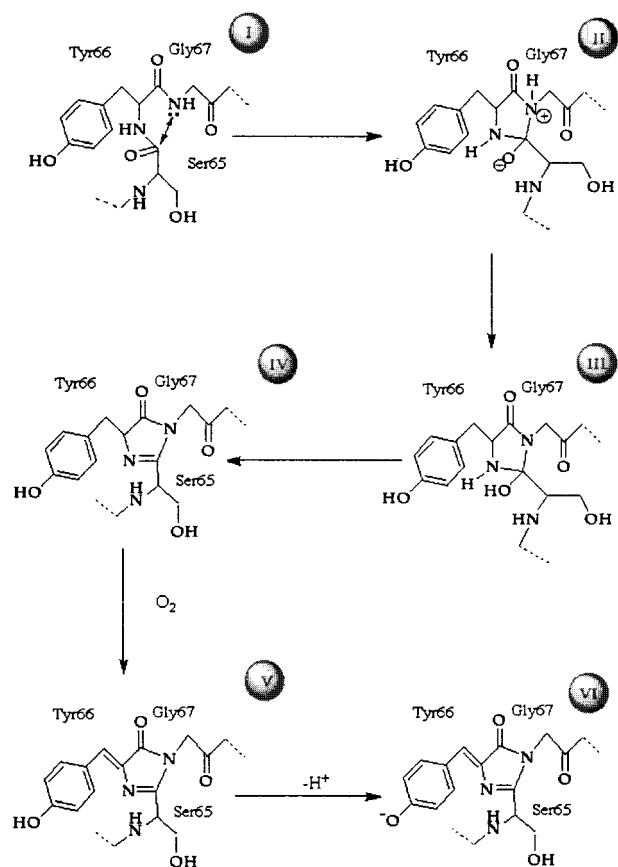
*Correspondence to:* P. E. M. Siegbahn; e-mail: ps@physto.se.

*Escherichia coli*, *Caenorhabditis elegans* [6], *Xenopus laevis*, *Drosophila melanogaster* [2], zebra fish [7], plants [8], human embryonic kidney cells [9], transgenic mice [10], and in other mammalian cells [11]. Some applications of GFP that have been reported are as a simple and rapid immunoassay, a versatile reporter for bioprocess monitoring, genetic probes, a signal for protein-protein interactions, a monitor for the expression of a mammalian odorant receptor, and an aid in transcriptional regulatory analysis in embryos and larvae.

The green lining along the edges of the jellyfish *Aequorea victoria* is due to the fluorescence of GFP. The green light is produced when calcium binds aequorin, another bioluminescent protein in *Aequorea victoria*, which transfers energy indirectly to GFP. GFP has become a commonly used tool in cell biology because it can be expressed in a variety of organisms where radiation with long wavelength ultraviolet (UV) light can transfer energy to GFP in place of aequorin.

Proteolytic treatment of GFP, which is composed of 238 amino acids, demonstrated that the chromophore was contained in a stable hexapeptide fragment, 64FSYGVQ69 [12]. In the intact GFP, the intrinsic chromophore is formed by autocatalytic internal cyclization of the tripeptide 65SYG67. GFP fluorescence is not observed until 90 min to 4 h after protein synthesis [13, 14]. Apparently, the protein folds quickly, but the subsequent fluorophore formation and oxidation is slow [15]. GFP refolding from an acid-, base-, or guanidine HCl-denatured state (chromophore containing but nonfluorescent) occurs with a half-life of between 24 s [16] and 5 min [17], and the recovered fluorescence is indistinguishable from that of native GFP [18].

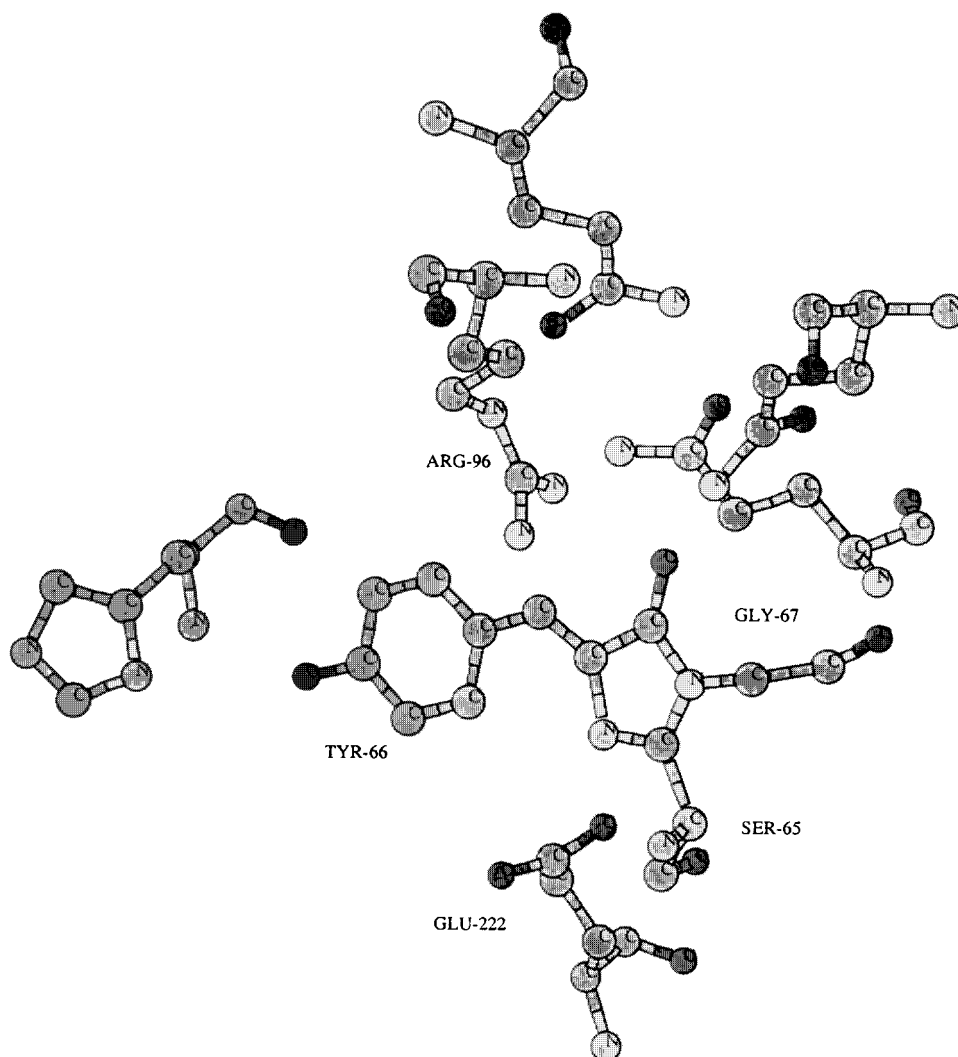
An interesting feature of GFP is that its function is based on a chromophore formed through a rarely observed posttranslational cyclization of a peptide from its own backbone structure. Posttranslational cyclizations also occur in microcin B17 and other natural products containing heterocycles derived from serine and cysteine residues. However, GFP is the only protein in which this process is autocatalytic and does not involve external enzymes. The detailed mechanism for the formation of the chromophore is unknown. However, Heim et al. [14] and Cubitt et al. [19] have proposed the autocatalytic biosynthetic mechanism shown in Figure 1. The scheme accounts for the spontaneous chromophore formation in a variety of GFP-expressing organisms that are unlikely to contain the same specific catalysts for the process.



**FIGURE 1.** Proposed mechanism for the chromophore formation in GFP [I, linear peptide (precyclization, immature); III, imidazolone ring; IV, dihydro-imidazolone; V, dihydrotyrosine (mature)].

The crystal structures of wild-type GFP as both a dimer [20] and a monomer [21], and the solid-state structures of several mutants [22, 23], have been reported. The chromophore is located in the center of a can consisting of 11  $\beta$  sheets. The can is a nearly perfect cylinder with a height of 42 Å and a radius of 12 Å. The chromophore region is shown in Figure 2. By enclosing the chromophore in the can, it may be protected from quenching by oxygen [24] and attack by hydronium ions [17]. Deletion mapping experiments [25] have shown that nearly the entire structure (residues 7–229) is required for chromophore formation and/or fluorescence.

Based on the computational analysis of the hexapeptide FSYGVQ, which we completed prior to the publication of the crystal structure of GFP, we proposed [26–28] that the posttranslational chromophore formation occurs due to the presence of low-energy conformations that have very short intramolecular distances between the carbonyl carbon



**FIGURE 2.** X-ray structure of the chromophore region of *A. victoria* GFP.

of Ser65 and the amide nitrogen of Gly67 (in I, Fig. 1). We also suggested that an arginine side chain, such as Arg73, may hydrogen bond to the carbonyl oxygen of Ser65 activating the carbonyl carbon of Ser65 for attack by the lone pair of the Gly67 amide nitrogen. The close proximity of an arginine to the chromophore, namely Arg96, has since been confirmed by the GFP crystal structures.

Following the publication of the solid-state GFP structures, we used the wild-type GFP structure as a starting point for the conformational analysis of the chromophore-forming region [26–28]. All the solid-state structures contain the fully cyclized chromophore (mature), and the only way to get information about the linear precatalytic form (immature) is by computational methods. This analysis

showed that the chromophore-forming residues in immature GFP are preorganized in a tight turn with the carbonyl carbon of Ser65 very close to the amide nitrogen of Gly67, and that the GFP barrel also tremendously restricts the conformational space of the chromophore-forming region, so that the residues are kept in place for autocatalytic cyclization, a slow step ( $t_{1/2} \approx 5$  min) in chromophore formation.

Because the autocatalytic posttranslational reaction that occurs to form the chromophore in GFP is so unusual and because understanding the mechanism of the reaction will hopefully facilitate the design of mutants that fluoresce quicker and at higher temperatures, we have studied the mechanism more closely. A hybrid density functional

theory (DFT) method (B3LYP) [29] with large basis sets and molecular mechanics was used to investigate the mechanism presented in Figure 1 and alternative mechanisms. The first step in such an investigation is to calculate the relative thermodynamic stabilities of the intermediates, and this will be the main focus of the present work. The very accurate G2 method [30] was used as a benchmark test on one of the smaller models to calibrate the somewhat less accurate B3LYP method in this particular problem. The general accuracy obtainable using the B3LYP method has been demonstrated in benchmark tests comprising 55 common molecules composed of first and second row atoms [31]. Using slightly larger basis sets than used here, an average absolute deviation compared to experiments of 2.2 kcal/mol was obtained for the atomization energies, of 0.013 Å for the bond distances, and of 0.62 degrees for the bond angles. The present accuracy should be almost as high as in this benchmark test.

The present study leads to a suggestion of a different mechanism for peptide cyclization in which oxidation of Tyr66 precedes cyclization, which ought to be carefully examined experimentally. Possible transition states involving both Arg96 and Glu222 as catalysts for the chromophore formation were studied. In particular, concerted proton transfer mechanisms of a type previously found to be likely reaction steps in ribonucleotide reductase (RNR) [32] and enzyme-coenzyme-B<sub>12</sub> [33] substrate reactions, were investigated. However, due to the complexity and the many possibilities available for these types of processes, this part of the present study should only be regarded as preliminary.

---

## Computational Details

The DFT calculations were performed in three steps. Following an optimization of the geometry using medium size basis sets, the energy was evaluated using large basis sets. In the third step the effect of the polarized surrounding was evaluated. All these steps were made at the B3LYP level [29, 34] using the Gaussian-94 program [35].

The B3LYP functional can be written as

$$F^{\text{B3LYP}} = (1 - A)F_x^{\text{Slater}} + AF_x^{\text{HF}} + BF_x^{\text{Becke}} + CF_c^{\text{LYP}} + (1 - C)F_c^{\text{VWN}},$$

where  $F_x^{\text{Slater}}$  is the Slater exchange,  $F_x^{\text{HF}}$  is the Hartree-Fock exchange,  $F_x^{\text{Becke}}$  is the gradient part of

the exchange functional of Becke [29],  $F_c^{\text{LYP}}$  is the correlation functional of Lee, Yang, and Parr [36], and  $F_c^{\text{VWN}}$  is the correlation functional of Vosko, Wilk, and Nusair [37].  $A$ ,  $B$ , and  $C$  are the coefficients determined by Becke [29] using a fit to experimental heats of formation, where the correlation functionals of Perdew and Wang [38] were used instead of  $F_c^{\text{VWN}}$  and  $F_c^{\text{LYP}}$  in the expression above.

For most models the B3LYP energy calculations were done using the large 6-311+G(2d,2p) basis sets in the Gaussian-94 program. This basis set has two sets of polarization functions on all atoms and also diffuse functions. For the systems containing a phenol ring, this basis set turned out to give rise to linear dependencies that led to divergence of the energies. To correct this problem, the diffuse functions were therefore removed for these systems. The effects of the diffuse functions were checked on some of the other systems that did not have the linear dependency problem and were shown to change the relative energies by only a few tenths of a kilocalorie/mole. For the largest models used to study the actual reactions, including the transition states, the second set of polarization functions was removed but the diffuse functions were kept. The effect of the second set of polarization functions was also shown to be only a few tenths of a kilocalorie/mole on the smaller systems. In the B3LYP geometry optimizations a much smaller basis set, the d95 set of the Gaussian-94 program, was used, since it has been shown that the final energies are extremely insensitive to the quality of the geometry optimization [31, 39, 40]. Based on previous basis set investigations, it is concluded that the basis sets used are close to saturated for the energies of the present reactions. Any inaccuracies of the results should therefore come from the chemical models used or from the use of the B3LYP functional.

For most model systems zero-point vibrational effects were calculated at the Hartree-Fock (HF) level using the d95 basis set and were then scaled by 0.9 as usual. To calculate these effects at the HF level is one order of magnitude cheaper than to obtain them at the B3LYP level at present. In benchmark tests it has not been found to be more accurate to obtain zero-point effects at the more expensive B3LYP level [31], and in a few test cases the same experience was obtained here. The fact that B3LYP optimized geometries are used for the zero-point determinations rather than HF optimized geometries, should have only quite small effects on the results as tested in previous studies [41]. For the present model reactions the relative effects of zero-

point vibration are furthermore found to be quite small, in the range of 0–2 kcal/mol. All energies discussed below include zero-point effects. Thermal effects are found to be less important and are not discussed here. For the transition states obtained, the HF Hessians were found to yield only one imaginary vibrational frequency.

The dielectric effects from the surrounding protein were obtained using the self-consistent isodensity polarized continuum model (SCI-PCM) as implemented in the Gaussian-94 program [35, 42]. This is a simple model for treating long-range solvent effects and considers the solvent as a macroscopic continuum with a dielectric constant  $\epsilon$  and the solute as filling a cavity in this continuous medium. The cavity is defined self-consistently in terms of a surface of constant charge density for the solute molecule. The default isodensity value of  $0.0004 e/B^3$  was used, which has been found to yield volumes close to the observed molar volumes. The solvent effect is derived from the interactions of the surface potential with the dielectric continuum. The dielectric constant ( $\epsilon$ ) of the protein is the main empirical parameter of the model, and it was chosen to be equal to 4 in line with previous suggestions for proteins. This value corresponds to a dielectric constant of about 3 for the protein itself and of 80 for the water medium surrounding the protein. This choice has been found to give very good agreement with experiment for two different electron transfer processes in the bacterial photosynthetic system [43]. In the present case where neutral models are chosen throughout, the details of how the dielectric effects are computed are quite unimportant since the effects are found to be extremely small. In one case the geometry was reoptimized including the reaction field, but this gave quite insignificant changes to the structure.

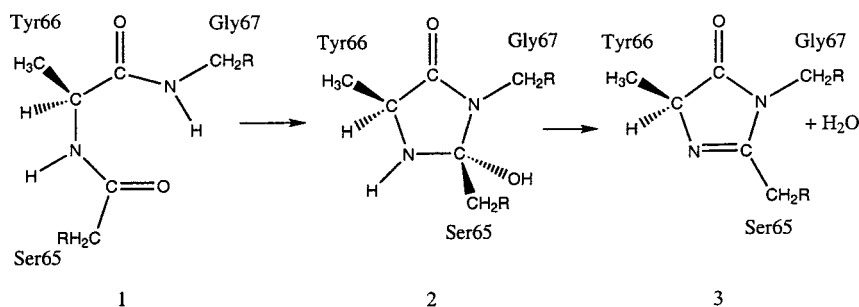
Even though quite accurate results were obtained for complicated electron transfer processes [43],

where the description of the electrostatic interaction between the solute and the solvent is much more critical than in the present study, it may still be questioned how accurately a continuous dielectric medium represents the protein environment. To further address this question, some test calculations were made for two of the intermediates using a different approach. The method used was the protein dipoles Langevin dipoles (PDL) model as implemented in the Polaris program package [44]. The PDL model uses the protein coordinates and places solvent dipoles on a spherical grid. The average polarization of each dipole is then approximated by a Langevin dipole. The dipole moments of the protein atoms interact with the point charges the Langevin dipoles and also with each other. The interaction energy is determined self-consistently.

As long as a similar protein environment is considered, the energetic effects obtained using PDL depend mainly on the charge distribution among the quantum chemical atoms. Since the results obtained using the SCI-PCM method also depend strongly on the atomic charges, the effects on the relative energies obtained using PDL are likely to be small whenever the effects obtained using the SCI-PCM method are small. The intermediates tested were **2** and **3** of Scheme 1 using the min model. The X-ray structure used for these calculations is the one obtained for *A. Victoria* at 1.90 Å resolution [20]. Using the PDL model, the surrounding protein was found to stabilize **3** by 0.8 kcal/mol relative to **2**, which indeed is a very small value.

In order to test the accuracy of the B3LYP method for the present systems, a few calculations were done at a higher level for some of the smallest models used. The method used for these tests was the G2MP2 method [45], where the energy is written as

$$E = E(\text{CCSD}(T)/S) + E(\text{MP2}/L) - E(\text{MP2}/S) - A n \alpha - B n \beta, \quad (1)$$



SCHEME 1.

where  $S$  and  $L$  denote a small and a large basis set, respectively.  $A$  and  $B$  are empirical parameters that do not enter into the present relative energies. Equation (1) can be seen simply as an extrapolation of the CCSD(T) energy from the small to the large basis. In the present case somewhat smaller basis sets were used than those originally suggested for the G2MP2 method. The small basis  $S$  used is the 6-31G\* basis and the large basis  $L$  the 6-311+G(2df,2p) basis, which have been demonstrated to give almost the same accuracy as the use of the original larger basis sets [46]. The B3LYP optimized geometries were used. In order not to confuse the present scheme with the original G2MP2 scheme, the results discussed below are denoted G2-M.

In order to investigate the effect of the helix strain on the mechanism, Monte Carlo simulations were performed. The coordinates of the wild-type GFP solid-state structure (1GFL) were obtained from the Protein Data Bank; hydrogen atoms were added to protein and solvent atoms as required. The chromophore was graphically restored to the polypeptide as it would be prior to the autocatalytic cyclization. Crystallographically determined water molecules were incorporated in the calculations. A "not" sphere of 12 Å from residues 65–67, with a secondary constrained sphere extending a further 3 Å, was used in all simulations and conformational searches. The AMBER\* force field as implemented in MacroModel v6.5 [47] was used for all molecular modeling.

Dihedral Monte Carlo multiple minimum searches [48] of the chromophore-forming region in reduced and oxidized GFP were undertaken with the closure bonds between residues 63 and 64, as well as between residues 69 and 70. All flexible dihedral angles in residues 64–69 were rotated. A minimum ring closure distance of 1.00 Å and a maximum of 4.00 Å was used. Each Monte Carlo (MC) step varied between 1 and 15 of the rotatable dihedral angles by between 0 and 180°. During the search procedure, minimization continued until convergence was reached or until 1000 iterations had been performed. The Polak–Ribiere conjugate-gradient minimization mode was used *in vacuo* with a derivative convergence criterion of 0.05 kJ/mol. Structures within 50 kJ/mol of the lowest energy minimum were kept and a usage-directed method [49] was used to select structures for subsequent MC steps. Six thousand MC steps were taken and all conformations within 50 kJ/mol of the lowest energy conformation were combined and subjected to a further 10,000

iterations with the multiconformer minimization mode of MacroModel. All unique conformations within 50 kJ/mol of the global minimum structure were kept. Structures were considered unique when the least-squares superimposition of all the pairs of related nonhydrogen atoms in residues 64–69 found no pair that was separated by less than 0.025 Å. Residues 62–70 of the global minima of both the oxidized and reduced precyclized forms of GFP determined in the manner described above were removed from the protein. Hydrogens were added and the dihedral strain energy was calculated without minimizing the structures.

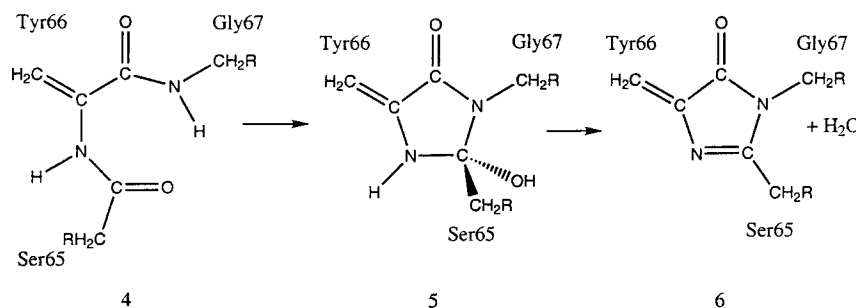
---

## Relative Thermodynamic Stabilities

The first step in a theoretical investigation of the feasibility of a mechanism such as the one in Figure 1 is to determine the relative thermodynamic stabilities of the intermediates. In the present study this has been done at two different levels presented in two different subsections below. First, a sequence of models of increasing chemical complexity were studied using the efficient B3LYP method. This study was undertaken to investigate the convergence of the results with respect to the chemical model. As will be shown below, this convergence is very fast, and close to converged results are obtained even for the smallest model. Secondly, for the smallest model the very accurate and time-consuming G2-M method was used to confirm the reliability of the B3LYP results. As shown by the benchmark tests described in the introduction, highly accurate results are expected also at the B3LYP level for these types of systems, but exceptions may of course in principle exist. Two mechanisms for peptide cyclization were studied. In the first one cyclization precedes oxidation as in Figure 1, which will be termed the reduced mechanism, and in the second one the order is reversed, which will be termed the oxidized mechanism. The reduced mechanism is schematically illustrated in Scheme 1 and the oxidized mechanism in Scheme 2.

### DIFFERENT CHEMICAL MODELS AT THE B3LYP LEVEL

The chemical models used to test different mechanisms are derived from the general structures of 1–6. R in these structures denotes the continuation of the peptide chain. Three different amino acids are involved in cyclization. They are residues Ser65,



SCHEME 2.

Tyr66, and Gly67. The five different models used are described briefly in Table I. In the model denoted "min" the smallest possible model for the side chains of these residues is used, which means a hydrogen atom for Ser65 and Gly67 and a methyl group for Tyr66. A methyl group is required for the tyrosine since for this side chain a hydrogen atom should be removed in the oxidized structure. Even though the min model represents the simplest possible model for the side chains, an even smaller chemical model is possible to use to form the ring structures. In this model denoted "min-" the  $\alpha$ -carbons connecting Ser65 and Gly67 are replaced by hydrogen atoms while Tyr66 is replaced by a methyl group as in the min model. This is the model used for the G2-M calculations described in the next subsection. The third model, denoted "vin," is the same as the min model except that the methyl group is replaced by a vinyl alcohol group to represent the conjugation effects on the rings that might appear due to the phenol ring of the Tyr66 residue. The model denoted "vin-ser" is the same as the vinyl model except that the actual methanol side chain is used for Ser65. In the model denoted "vin-ser+"

used in a few calculations, HCO is used to extend the peptide chain at Gly67 and  $\text{NH}_2$  to extend the chain at Ser65. Finally, in the largest model denoted "tyr-ser" the vinyl alcohol in the vin-ser model is replaced by a phenol group. In this model there is thus no model approximation apart from the one that the peptide is cut off before Ser65 and after Gly67.

The fragments used in the DFT calculations and described in Table I are approximations of the chromophore-forming region. In order to examine the effect of the helix strain on the mechanism, residues 62–70 were removed from the minimized precyclized reduced and oxidized GFP structures (see computational details), and the strain dihedral energy of these nonapeptides was calculated. The oxidized conformation had 8.8 kcal/mol less dihedral strain than the reduced form [50].

Before the results of the model calculations are discussed, mutation experiments performed on the three side chains in the structures 1–6 will be summarized. Gly67 is absolutely essential for the peptide cyclization and cannot be replaced by any other amino acid. This is most likely due to steric and not to chemical effects. Tyr66 has been mutated to His and some other aromatic residues, but they only form a small percentage of the chromophore. Mutation to a nonaromatic residue appears not to give any peptide cyclization and/or fluorescence. Ser65 has been mutated to Ala, Cys, Gly, Leu, Arg, Asn, Asp, Phe, and Trp. All mutants form the chromophore but result in different absorption spectra. In summary, these results strongly indicate that the only requirements in the tripeptide region for the actual peptide cyclization are a sterically very small residue at the Gly67 position and possibly an aromatic side group at the Tyr66 position. It is therefore highly unlikely that the models used in the present study should not be reasonably good representations of this reaction. In fact, it can already at the

TABLE I  
Models used in the present study.<sup>a</sup>

Model	Ser65	Tyr66	Gly67
min-	H(-)	CH <sub>3</sub>	H(-)
min	H	CH <sub>3</sub>	H
vinyl	H	CH <sub>2</sub> (C <sub>2</sub> H <sub>2</sub> OH)	H
vin-ser	CH <sub>2</sub> OH	CH <sub>2</sub> (C <sub>2</sub> H <sub>2</sub> OH)	H
vin-ser+	CH <sub>2</sub> OH(+)	CH <sub>2</sub> (C <sub>2</sub> H <sub>2</sub> OH)	H(+)
tyr-ser	CH <sub>2</sub> OH	CH <sub>2</sub> (C <sub>6</sub> H <sub>4</sub> OH)	H

<sup>a</sup> The different models used for Ser65, Tyr66, and Gly67 are given in each column. (-) and (+) means that the peptide chain has been somewhat shortened and extended, respectively, with respect to the ones in Schemes 1 and 2, see text.

**TABLE II**  
Relative energies for the reactants, intermediate, and products for the different models (see Table I) of the reduced mechanism in Scheme 1.

Model	Linear 1	Dihydro-imidazolone 2	Imidazolone 3
min-	0.0	16.0	10.7
min	0.0	16.5	11.5
vinyl	0.0	15.6	11.5
vin-ser	0.0	16.2	9.2
vin-ser+	0.0	17.7	—
tyr-ser	0.0	16.2	10.6

onset be expected that the results for most of the models should be rather similar to each other and to experiments.

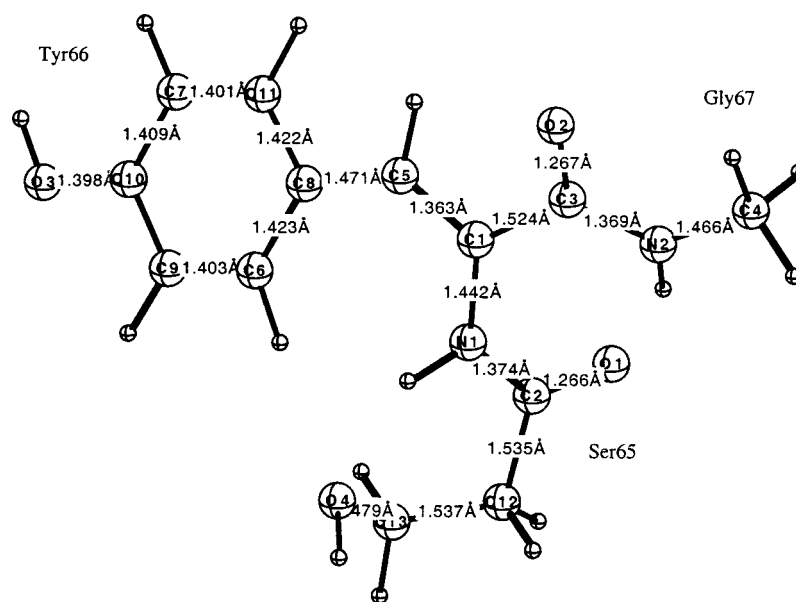
The results for the reduced mechanism (Scheme 1) are shown in Table II and for the oxidized mechanism (Scheme 2) in Table III. As examples of the structures obtained, the largest tyr-ser models for the oxidized scheme are shown in Figures 3–5. The calculated results for the reduced mechanism, here discussed first, are quite striking and do not support the mechanism in Figure 1. First, and most importantly, the peptide cyclization leading to the imidazolone (3) structure is strongly endothermic by 9–12 kcal/mol in all models used. Even if it would be kinetically possible to perform

**TABLE III**  
Relative energies for the reactants, intermediate and products for the different models of the oxidized mechanism in Scheme 2.

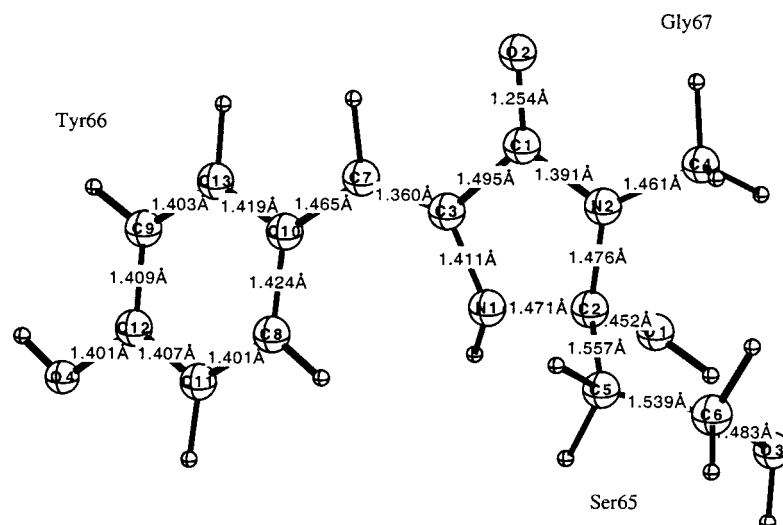
Model	Linear 4	Dihydro-imidazolone 5	Imidazolone 6
min-	0.0	7.6	4.1
min	0.0	8.6	3.6
vinyl	0.0	8.3	0.4
vin-ser	0.0	7.8	-3.3
tyr-ser	0.0	6.7	-1.9

the cyclization, this process would thus quickly go back to the thermodynamically more stable linear form. There is furthermore no indication from the results of the different models that an even more realistic model would make the cyclization favorable. If the experimental interpretation leading to the mechanism in Figure 1 is still correct, it means that there is an unusually large error in the B3LYP method and that a large correction therefore needs to be added to the results in Table II. This will be tested in the next subsection.

Also the results for the dihydro-imidazolone (2) structure using the reduced mechanism are in conflict with the mechanism in Figure 1. As seen in Table II, the formation of this ring is endothermic by 16–18 kcal/mol, which makes it unlikely as



**FIGURE 3.** Largest tyr-ser model for the linear peptide (4) of the oxidized scheme.

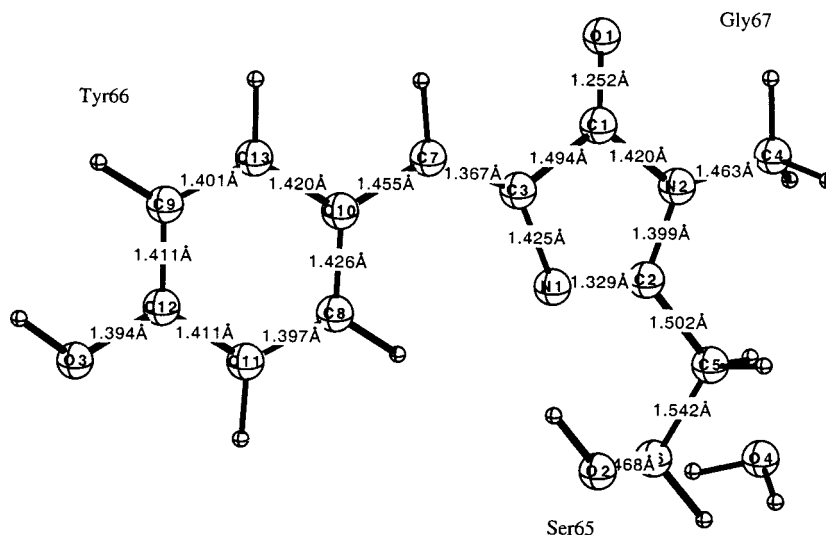


**FIGURE 4.** Largest tyr-ser model for the dihydro-imidazolone (5) of the oxidized scheme.

an intermediate in the peptide cyclization process. However, it is energetically possible at room temperature to reach this structure provided that the barriers are very small. As will be shown this does not turn out to be the case.

An argument for the reduced mechanism [19] is that it is closely analogous to the first step in the known deamination reaction of Asn-Gly sequences in peptides and proteins, in which the amino group of the Gly attacks the side chain amide of Asn, also forming a five-membered ring [51]. In order to see if the B3LYP method is systematically incorrect for this type of processes, a few calculations were done

also on this deamination process. Using a model essentially equivalent to the min model, the B3LYP method gives an exothermicity for this process of 2.4 kcal/mol well in line with the known experimental result and in contrast to the endothermicity of 11.5 kcal/mol obtained at the same level for the GFP cyclization process using the reduced mechanism. A calculation was also performed on the possible carbinolamine intermediate, corresponding to the dihydro-imidazolone (2), and it was found to be 11.3 kcal/mol higher in energy than the linear Asn-Gly peptide. This energy does not exclude it as a possible intermediate, but this still appears very



**FIGURE 5.** Largest tyr-ser model for the imidazolone (6) of the oxidized scheme.

unlikely. The most probable reaction path is the direct transfer of a hydrogen from the amino group of Gly to the side chain nitrogen of Asn, which would avoid the carbinolamine intermediate.

The results in Table II for the reduced mechanism are quite stable with respect to the model used, as expected based on chemical grounds. Two possible weak trends can be identified. First, the endothermicity for formation of the imidazolone (3) goes down by about 2 kcal/mol as the methanol is introduced as a model for Ser65. This is due to the formation of a hydrogen bond between the water produced in the process and the OH group of the Ser65 alcohol. This hydrogen bond is apparently 2 kcal/mol stronger than the ones possible without the presence of the alcohol. It is somewhat unclear what this means for the actual protein where all hydrogen bonds should be close to saturated before the reaction. Also, as mentioned above, mutating Ser65 to another residue without this possibility for hydrogen bonding seems to have almost no effect on the efficiency of the peptide cyclization. The second trend seen is a possible weak tendency for an increased endothermicity of 2 kcal/mol for the dihydro-imidazolone (2) formation due to increased steric repulsion as the peptide is extended in the vin-ser+ model. However, this effect needs a more careful examination with larger models to be made certain.

The results for the oxidized mechanism (Scheme 2) shown in Table III are much more in line with a possible peptide cyclization. Both the dihydro-imidazolone (5) and the imidazolone (6) structures are significantly easier to access in the oxidized than in the reduced mechanism. Most importantly, once a conjugated model is introduced for Tyr66, the entire peptide cyclization forming the imidazolone (6) is found to be almost thermoneutral. With the uncertainties of the B3LYP method the results must be interpreted as leaving the oxidized mechanism as an open possibility at this stage of testing.

According to the results in Table II, the formation of the dihydro-imidazolone (5) is still endothermic by about 8 kcal/mol using the oxidized mechanism. This is significantly less than the 16–18 kcal/mol found for the reduced mechanism, which at least makes the dihydro-imidazolone (5) a plausible unstable intermediate in the cyclization process if the system is oxidized prior to cyclization. It is difficult to envision a possible formation of the imidazolone (6) without forming the dihydro-imidazolone (6) as an intermediate, but possibilities

exist and are tested here. It can finally be noted that the energy of the dihydro-imidazolone (5) is very stable with respect to model used. This is somewhat different from the results for the imidazolone (6) where the introduction of a conjugated group as a model for Tyr66 has a stabilizing effect of about 4 kcal/mol. This effect is due to the presence of the double bond in the imidazolone (6), which increases the number of resonance possibilities with the conjugated side group.

As described above all energies discussed here include a simple estimate of the protein surrounding. This estimate is made using the SCI-PCM method, which uses a dielectric cavity shaped according to the model system. A dielectric constant of 4 was chosen for the surrounding protein. The effect of the surrounding dielectric is extremely important in cases where the model system changes its charge in a process, as shown, for example, by model calculations for the charge separation in photosynthesis [43]. The effect can also be significant for charged model systems even if the charge does not change. For neutral model systems, like the present ones, large effects are quite rare. In fact, if the dielectric effect is found to be large for a neutral model system, this indicates a deficiency in the model such as missing strong hydrogen bonds. For the present mechanisms all models show quite small dielectric effects in the range 0–2 kcal/mol for the relative energies, which means that the models are consistently chosen. This also has the advantage that the results are insensitive to the precise dielectric constant chosen, which is normally the main uncertainty of this type of simple protein model. The conclusion that only small dielectric effects are involved are further supported by the results obtained using the PDL model for structures 2 and 3 of Scheme 1. Using this approach, 3 was found to be stabilized by only 0.8 kcal/mol relative to 2.

#### CALIBRATION CALCULATIONS FOR THE SMALLEST MODEL

The results described in the preceding subsection strongly indicate that the mechanism in Figure 1 should not be thermodynamically feasible. This is mainly because the formation of the imidazolone (3) is found to be about 10 kcal/mol above the linear peptide but also because the intermediate dihydro-imidazolone (2) is quite unstable by 16–18 kcal/mol with respect to the linear peptide, which makes it an unlikely intermediate. These conclusions are based on the B3LYP results, which are expected to

**TABLE IV**  
**Benchmark calibration results for the reduced mechanism using the smallest min– model. S and L are the small and large basis sets, respectively.**

Method	Basis	Linear	Dihydro-	Imidazolone
		1	2	3
B3LYP		0.0	16.0	10.7
CCSD(T)	S	0.0	17.3	13.2
MP2	S	0.0	17.7	13.5
MP2	L	0.0	13.7	9.1
G2-M	S,L	0.0	13.3	8.8

be quite accurate based on previous benchmarks; see the introduction. In order to test this, the even more accurate G2-M method was used for the smallest min– model.

The G2-M results are given in Tables IV and V together with the B3LYP and some intermediate results. These results show quite clearly that the B3LYP method is sufficiently accurate also for the present problem. For the reduced mechanism the G2-M relative energies are for the dihydro-imidazolone (2) +13.3 kcal/mol and for the imidazolone (3) +8.8 kcal/mol, compared to the B3LYP results of +16.0 kcal/mol and +10.7 kcal/mol, respectively. For the oxidized mechanism the relative G2-M energies are +4.6 kcal/mol and +1.3 kcal/mol, respectively, while the B3LYP results are +7.6 kcal/mol and +4.1 kcal/mol.

The difference between the G2-M results for the small min– model can be used to correct the B3LYP results for the most realistic model used, the tyrosine model. This leads to the best predictions that can be made based on the present calculations with relative energies for the dihydro-imidazolone and imidazolone for the reduced mechanism of

**TABLE V**  
**Benchmark calibration results for the oxidized mechanism using the smallest min– model. S and L are the small and large basis sets, respectively.**

Method	Basis	Linear	Dihydro-	Imidazolone
		4	5	6
B3LYP		0.0	7.6	4.1
CCSD(T)	S	0.0	9.2	6.0
MP2	S	0.0	9.0	6.2
MP2	L	0.0	4.4	1.5
G2-M	S,L	0.0	4.6	1.3

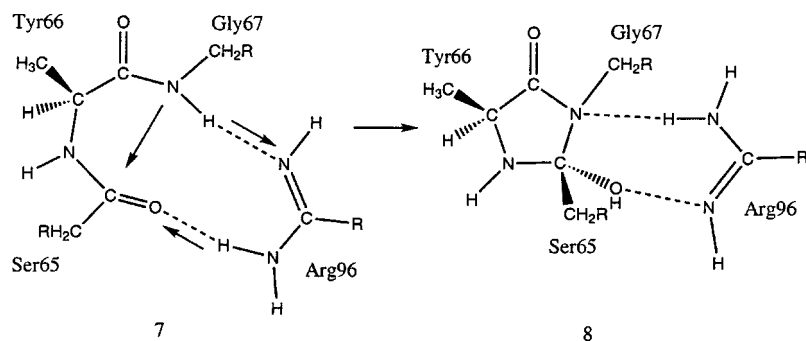
+13.5 kcal/mol and +8.7 kcal/mol, and for the oxidized mechanism of +3.7 kcal/mol and –4.7 kcal/mol. The conclusion is that a mechanism where the oxidation of the peptide is made prior to cyclization is energetically feasible provided that the barriers are small enough. If this oxidation is made only after the cyclization as in the reduced mechanism, on the other hand, this is not energetically feasible.

As a by-product of the G2-M calculations large basis MP2 results and small basis CCSD(T) results are available. These are also given in Table IV and show that the results do not vary very much with the method. Coincidentally, the large basis MP2 results are very close to the G2-M results, differing from these by at most 0.4 kcal/mol. The small basis CCSD(T) results are all 4–5 kcal/mol higher than the G2-M results.

All the above results indicate that the oxidized mechanism, Scheme 2, leads to feasible energetics for the ring closure. A remaining question is if this mechanism also fulfills the requirement of an exothermic oxidation by O<sub>2</sub>, which should then occur for the linear structure 1 going to 4. The sum of the two O–H bond strengths in H<sub>2</sub>O<sub>2</sub> is 134.3 kcal/mol experimentally [30]. The two C–H bond strengths for the hydrogens removed from 1 must therefore be smaller than this value to make the oxidation exothermic. At the B3LYP level the sum of these C–H bond strengths is 133.6 kcal/mol, and at the G2-M level 131.9 kcal/mol, thus fulfilling the energetic requirement.

## Barriers for the Chromophore Formation

To determine the barrier heights for the present reactions is naturally much more difficult than to obtain the relative stabilities of the intermediates, in particular since the size of the models used must be considered as quite large by quantum chemical standards. Nevertheless, some attempts were made to locate plausible transition states. The types of reaction mechanisms investigated primarily are of the type found to be important in some previous studies on similar enzyme reactions. These are concerted proton transfer reactions mediated in this case either by an arginine or by a glutamic acid. In these reactions one proton of the substrate is accepted by the nitrogen (or oxygen) of an Arg (or Glu) amino acid and another proton is donated from the other nitrogen (or oxygen) of the Arg (or Glu) to a different

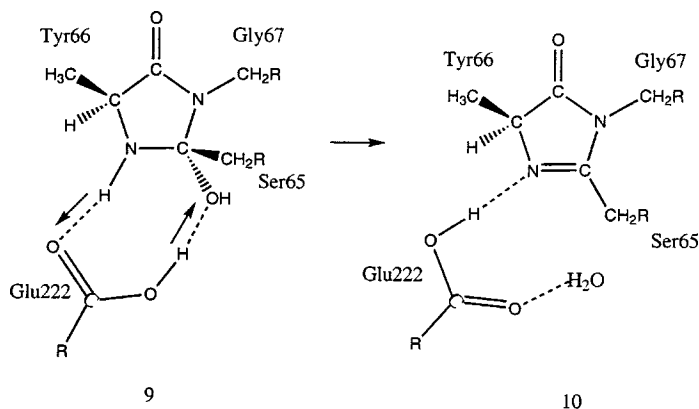


SCHEME 3.

position of the substrate; see Schemes 3 and 4. In this way the amino acids can help bridge sites at some distance from each other, where otherwise a steric constraint would prevent the transfer reaction. In these mechanisms the charge separation is kept at a minimum, which can be a problem when separate donating and accepting residues are required. It should be added that in the models below, both arginine and glutamic acid are assumed neutral. Since there is no experimental evidence that these residues are in fact neutral from the beginning, a few kilocalories/mole should probably be added to the computed barriers below.

The proton transfer reaction required to form the dihydro-imidazolone in Scheme 1 or 2 is one where a proton is moved from the peptide amino group of Gly67 to the peptide carboxyl group of Ser65; see Scheme 3. To perform this transfer a neutral arginine is ideally suited and should be placed with its protonated nitrogen end with a hydrogen bond to the Ser65 carboxyl group and with its unprotonated nitrogen end with a hydrogen bond to the Gly67 amino group. A requirement for this mechanism is that the arginine is able to trans-

fer a proton to a nearby base with only a minor cost prior to the reaction. The essential role of an arginine, Arg73, has been suggested based on previous molecular mechanics (MM) studies. These studies showed that it is indeed possible to place the arginine in the desired position. The shortest distance between the amino nitrogen of Gly67 that releases the proton, as shown in Scheme 3, and the accepting nitrogen of Arg96 in any of the reported crystal structures is 4.14 Å (Protein Data Bank reference code 2emo) and the longest distance is 4.94 Å (Protein Data Bank reference code 1bfp). These solid-state distances show that Arg96 is close to the chromophore-forming region and could be involved in a proton transfer reaction as depicted in Scheme 3. However, since the distance between Arg96 and Gly67 in the immature GFP (as in structure 7) is more relevant to the mechanism, we reanalyzed the results from our conformational analysis of the chromophore-forming region in immature GFP [28] and undertook a new conformational analysis of the chromophore-forming region in immature GFP with neutral Arg96 and Glu222 residues. In order to find the low-energy



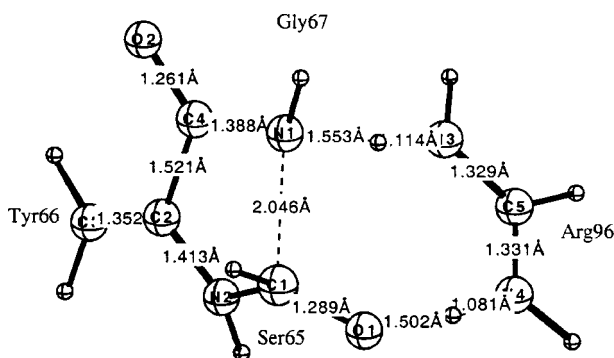
SCHEME 4.

conformations all the variable dihedral angles in the chromophore-forming region were varied as well as those in residue 96. The shortest distance between the amino nitrogen of Gly67 that releases the proton, as shown in Scheme 3, and the accepting nitrogen of Arg96 in any of the conformations found within 10 kcal/mol of the lowest energy conformation is 3.99 Å and the longest distance is 6.44 Å. The shortest distance between the proton-accepting carbonyl oxygen of Ser65 and the proton-donating nitrogen of Arg96 is 2.84 Å and the longest is 5.80 Å.

To go from the dihydro-imidazolone to the imidazolone in Schemes 1 and 2, a proton transfer is required between the amino group of Ser65 to the hydroxyl group formed at Ser65 in the dihydro-imidazolone formation; see Scheme 4. The hydroxyl group and the proton transferred will then form a water molecule. This type of process is quite similar to one of the steps suggested to occur in the RNR substrate reactions [32]. In that case a carboxylate group is ideally positioned to perform the transfer and form water. This carboxylate group in RNR has been shown by mutations to be essential for this step. In GFP the carboxylate group of Glu222 is quite well positioned to catalyze a similar reaction.

Two of the models described in Table I and discussed in the previous section were used to determine the transition states. The first one is the smallest model, termed min-, which was also used to calibrate the accuracy of the calculations; see above. For this model both the glutamic acid and arginine models were included simultaneously in the transition state determinations, but for the oxidized mechanism a determination of the first transition state was also made without the glutamic acid present. The largest substrate model used here is the vin-ser model described in Table I. As shown by the results discussed in the previous section, this model contains the main chemical effects necessary to allow the chromophore formation. The energetic differences compared to the largest model used, which includes the full side chains of all three amino acids, the tyr-ser model, are quite insignificant. Transition state determinations were only made with both glutamic acid and arginine present, not with any one of them absent as for the smaller model. Most calculations were done for the first transition state, the one between the linear peptide and the dihydro-imidazolone structure, since this is expected to be the rate-limiting reaction step.

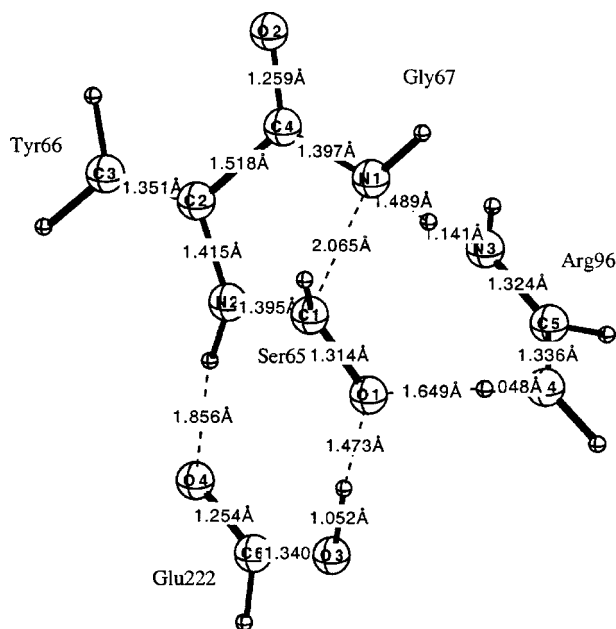
Starting with the smallest model and the model with only arginine present, the calculated barrier for the first transition state of the oxidized mechanism



**FIGURE 6.** Transition state between the linear peptide and the dihydro-imidazolone (5) for the oxidized scheme using the smallest min- model only, including Arg96 as additional amino acid.

is found to be quite high, 26.9 kcal/mol including substantial effects of both zero-point vibration,  $-2.6$  kcal/mol, and dielectric effects,  $-3.8$  kcal/mol. Large effects from zero-point vibration are expected for these types of reactions where the transition states have partly broken X-H bonds. The origin of the large dielectric effect is that at the transition state there is substantial contributions of an  $\text{ArgH}^+$  structure, as can be seen in Figure 6. This means that charge has been transferred from the substrate to arginine, and in any such charge transfer process dielectric effects are expected to be important. The large dielectric effect may also indicate that the protein effects are being underestimated by the simple dielectric cavity model; see further below. Even if the dihydro-imidazolone (5) formation is considered to be rate limiting for chromophore formation, a barrier of 26.9 kcal/mol is still somewhat too high. Since GFP fluorescence is observed 90 min to 4 h after protein synthesis, this indicates a rate-limiting barrier of about 23 kcal/mol.

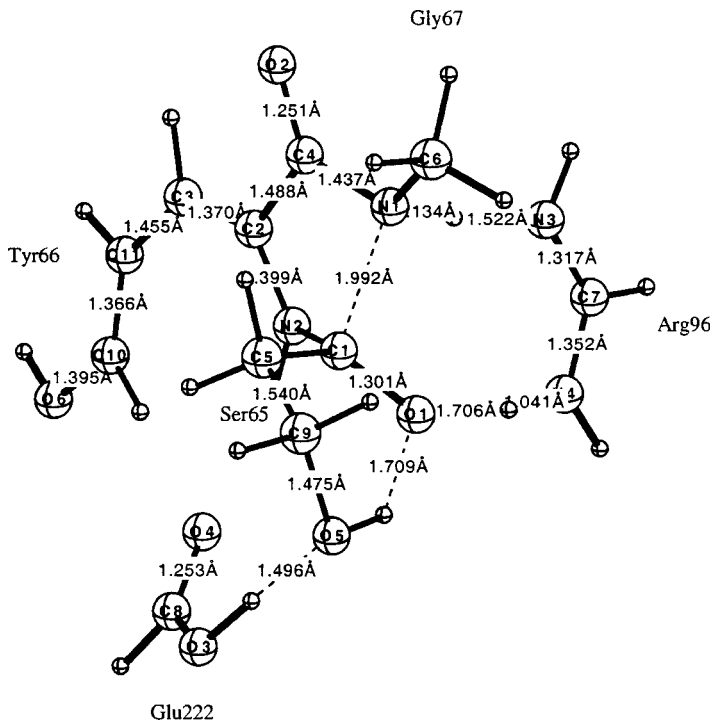
The next step in increasing the model is to include the glutamic acid model. Even though this residue is not directly involved in this proton transfer, it could have a stabilizing effect on the transition state, in particular on the hydroxyl group formed. The transition state is shown in Figure 7. Indeed, a small stabilizing effect of  $-3.4$  kcal/mol is seen bringing down the barrier for the oxidized mechanism to 23.5 kcal/mol in good agreement with the measured rate. The barrier for the reduced mechanism using this model is 26.1 kcal/mol, which is higher than the oxidized mechanism as expected based on the results of the preceding section. The difference of only 2.6 kcal/mol is surprisingly small, which would not be enough to discard this mech-



**FIGURE 7.** Transition state between the linear peptide and the dihydro-imidazolone (5) for the oxidized scheme using the smallest min- model, including both Arg96 and Glu222 as additional amino acids.

anism if this was the only reason (see below for the larger model, however). The strongest argument against the reduced mechanism is instead that the chromophore formation is predicted to be endothermic by 8.7 kcal/mol in contrast to the oxidized mechanism, which is predicted to be exothermic by 4.7 kcal/mol; see above.

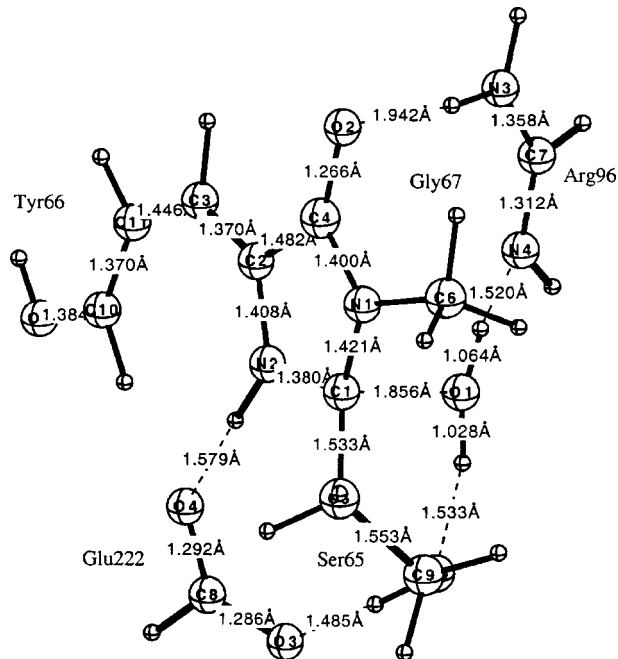
The results for the first transition state of the oxidized mechanism using the largest model tried here, the vin-ser model in Table I, do not change much compared to the min- model. The barrier for the vin-ser model is 24.3 kcal/mol compared to 23.5 kcal/mol for the min- model. The structure of the transition state is shown in Figure 8. Even though the stability of the barrier height with respect to the model used may not seem very surprising considering the small changes found for the dihydro-imidazolone structure in Table III for the different models, there are some striking differences for the structures of the transition states. For the transition states in Figures 6 and 7 the proton has already left the peptide nitrogen of Gly67, while this is not the case for the transition state in Figure 8. One reason for this is connected to the change in



**FIGURE 8.** Transition state between the linear peptide and the dihydro-imidazolone (5) for the oxidized scheme using the vin-ser model, including both Arg96 and Glu222 as additional amino acids.

the model made by replacing the hydrogen by a methyl group, modeling the Gly67 peptide, at this nitrogen. This leads to a significant change in the N–H bond strength for the proton moving over to the arginine of about 6 kcal/mol. This estimate was obtained from a calculation comparing the N–H bond strengths in  $\text{NH}_2(\text{CH}_3)$  with  $\text{NH}(\text{CH}_3)_2$  using a large basis set. The model used in Figure 8 is obviously more realistic than the ones in Figures 6 and 7 and should therefore give the most reliable structure of the transition state. Due to cancellation of effects, of steric and resonance origin, the change in transition state structure does not cause any change in barrier height for the oxidized model, but this is not so for the reduced model; see below.

For the vin-ser model the second transition state for the oxidized mechanism, the one between dihydro-imidazolone (5) and imidazolone (6), was also located. This transition state structure is shown in Figure 9. The barrier height with respect to the dihydro-imidazolone (5) is 10.4 kcal/mol. This means that the total barrier height with respect to the linear peptide reactant should be about 14.1 kcal/mol, where the best predicted endothermicity for formation of the dihydro-imidazolone (5) of 3.7 kcal/mol has been added. This places this



**FIGURE 9.** Transition state between the dihydro-imidazolone (5) and the imidazolone (6) for the oxidized scheme using the vin-ser model, including both Arg96 and Glu222 as additional amino acids.

transition state significantly below the first transition state, and the second reaction step is therefore not rate limiting.

In the model used for the second transition state, Glu222 is used to transfer the proton. However, it has been experimentally shown that Glu222 can be replaced by a glycine, and the chromophore will still be formed [52]. In light of the rather low barrier found for the second step of only 14.1 kcal/mol, this experimental finding is not difficult to rationalize. Replacing the active glutamate in Figure 9 by a glutamine, Gln94, for example, which is well positioned in the X-ray structure, would probably not raise the barrier by more than a few kilocalories/mole. To avoid confusion it should be noted that for a glutamine only the amino group of the side chain takes part in the proton transfer. Even with this replacement, it would still make this step faster than the first step, so it would not be rate limiting. In the first step Glu222 is even less important as shown above, and it could again be replaced by a glutamine.

For the reduced mechanism only the first transition state was located using the vin-ser model. The calculated barrier height is 34.4 kcal/mol, which is 10.1 kcal/mol higher than for the oxidized mechanism. This result is in line with the difference in stability of the dihydro-imidazolone (5) discussed in the previous section and is therefore more expected than the result for the smallest min– model. There are several reasons for the change of barrier height for the different models of the reduced mechanism. One reason was discussed above and is connected to the rather large change of N–H bond strength of the Gly67 peptide amino group obtained by replacing the hydrogen atom by the more realistic methyl group in the larger model. It should also be pointed out that for the endothermicity of the ring closing step, the effect of the change in N–H bond strength is partly canceled by the formation of the N–C bond of the ring, but this cancellation does not occur for the transition state in the two models since the C–N bond is not yet formed. The larger model should be the more reliable one and thus gives an indication that the ring formation has a much too high barrier for the reduced mechanism in line with the large endothermicity found for this mechanism; see above.

A different type of mechanism for chromophore formation was finally tested for the reduced mechanism. In this model the substrate is initially protonated from an external residue at the Ser65 peptide carbonyl. This means that the substrate in this model has a positive charge unlike the neutral mod-

els discussed above. The proton added to the substrate could, for example, come from a protonated Arg96. Only the smallest min– model was used to study this mechanism, which turns out to be quite unfavorable for ring formation. The dihydroimidazolone, which is the same as structure 2 for the neutral min– model discussed above with an additional proton at the Gly67 peptide amino group, is unstable with respect to the linear peptide by as much as 30.0 kcal/mol. Several attempts were made to avoid the formation of the unfavorable dihydroimidazolone (2) and go directly to the formation of a protonated imidazolone (3). However, only very high-energy pathways were found, and this type of mechanism was therefore considered much less plausible than the neutral mechanisms discussed above.

In earlier work, using MM, we have shown that the chromophore-forming residues of GFP are pre-organized in a tight turn with the carbonyl carbon of Ser65 2.87 Å from the amide nitrogen of Gly67. Furthermore we have also shown that the  $\beta$  barrel surrounding the chromophore tremendously restricts the conformations available to residues in the chromophore-forming region. Both these factors aid in chromophore formation. The starting point of these calculations was the structure of immature precatalytic GFP (species I in Fig. 1), which was obtained by graphic conversion from the solid-state structure of mature (species V in Fig. 1) GFP. We have repeated the calculations with an oxidized form of the immature precatalytic GFP (the GFP equivalent of structure 4 in Scheme 2). Oxidizing Tyr66 prior to cyclization does not change the conclusions reached in our previous work. The distance between the carbonyl carbon of Ser65 and the amide nitrogen of Gly67, for all the low-energy conformations within 10 kcal/mol from the lowest energy conformation found for the oxidized form of the immature precatalytic GFP, is between 2.83 and 2.99 Å. A Ramachandran plot of the  $\phi$  and  $\psi$  dihedral angles of Tyr66 was obtained and shown to be very similar to that of the immature precatalytic GFP without prior oxidation of Tyr66.

---

## Conclusions

We initially undertook this computational analysis of the chromophore-forming reaction of GFP to establish the role of Arg96 in the autocatalytic cyclization. However, it soon became apparent that the calculated energetics of the cyclization, inde-

pendent of the chemical model used, were not in agreement with the currently accepted mechanism shown in Figure 1. We therefore proceeded to study a modification of this mechanism. According to the new mechanism, which is shown in Scheme 2, the chromophore-forming cyclization is preceded by the dehydration of Tyr66 to dehydrotyrosine. The calculated energetics of this mechanism are much more favorable, especially if one considers the possibility of proton transfer reactions facilitated by Arg96 and Glu222 or Gln94 (in the case of the E222G mutant), as shown in Schemes 3 and 4. Unfortunately our new mechanism is not in agreement with the most commonly accepted interpretations of certain experiments and therefore requires that either these experiments be reinterpreted or that the calculations have an unusually large error or both.

To reconcile the present results with experimental observations, there are essentially two alternatives that have both advantages and disadvantages. The first alternative is that the mechanism shown in Figure 1 is correct. According to this mechanism the first step in chromophore formation is the folding of the protein. The kinetics of this step were measured by monitoring the folding of the denatured mature GFP back to the fluorescent form of GFP [15]. An interesting observation is that a time lag was observed in *de novo* folding that is not observed in renaturing. The second step is the cyclization of the tripeptide and the formation of the reduced chromophore. The evidence for this step is the finding that anaerobically expressed GFP decrease by  $1 \pm 4$  Da after exposure to air and not by the 20-Da loss predicted for cyclization [19]. The final step is presumed to be the oxidation of Tyr66. The evidence for the last step is circumstantial—anaerobically expressed GFP does not fluoresce and upon addition of molecular oxygen it begins to fluoresce ( $t_{1/2} = 76$  min) [14]. The main drawback of this mechanism is that it appears energetically improbable in light of the results of the present calculations. If the experimental mechanism were right, this would either mean that there is an unusual error in the present methods or that a very important energetic effect of the surrounding protein has been left out of the model, both of which would be surprising and also quite interesting from a modeling and understanding perspective. In that case this error would occur although the same methods give results in very good agreement with experimental evidence for the quite similar and much better known deamination reaction of Asn-Gly sequences in peptides and proteins. It should furthermore be noted that the same

methods and models have been used successfully to explain experimental observations in the case of the substrate reactions in RNR [32], which have many similarities to the present reactions.

The second alternative is that the oxidation of residue 66 is a prerequisite to chromophore formation (Scheme 2) and that no cyclization occurs prior to the slow oxidation step. This would mean that the mass spectrometry data and other experimental data were either misinterpreted or incorrect [19].

Advantages to the second alternative are that they are in line with the calculated energetics from the present calculations. If the final slow step is not the oxidation of cyclized Tyr66 to dehydrotyrosine, but rather the conversion of the linear peptide to imidazolone, because the oxidation of the imidazolone occurs at residue 65, it would explain why the S65T mutant attains fluorescence quicker than wild-type GFP.

Our DFT calculations have suggested that the commonly accepted mechanism for chromophore formation is strongly endothermic and unlikely to occur. We have proposed an alternative mechanism, which we have called the oxidized mechanism which is energetically viable, that requires the reinterpretation of a number of experimental observations.

#### ACKNOWLEDGMENTS

We thank Arieh Warshel and Zhen Tao Chu for supplying the POLARIS program to us and assisting us in using this program.

#### References

- Santa Cruz, S.; Chapman, S.; Roberts, A. G.; Roberts, I. M.; Prior, D. A. M.; Oparka, K. J. *Proc Natl Acad Sci* 1996, 93, 6286.
- Gerdes, H.-H.; Kaether, C. *FEBS Lett* 1996, 389, 44.
- Tannahill, D.; Bray, S.; Harris, W. A. *Dev Biol* 1995, 168, 12501.
- Wang, S.; Hazelrigg, T. *Nature* 1994, 369, 400.
- Gilroy, S. *Annu Rev Plant Physiol Mol Bio* 1997, 48, 165.
- Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C. *Science* 1994, 263, 802.
- Amsterdam, A.; Lin, S.; Hopkins, N. *Dev Biol* 1995, 171, 123.
- Chiu, W.-L.; Niwa, Y.; Zeng, W.; Hirano, T.; Kobayashi, H.; Sheen, J. *Curr Biol* 1996, 6, 325. Heimlein, M.; Epel, B. L.; Padgett, H. S.; Beachy, R. N. *Science* 1995, 270, 1983.
- Marshall, J.; Molloy, R.; Moss, G. W. J.; Howe, J. R.; Hughes, T. E. *Neuron* 1995, 14, 211.
- Chiocchetti, A.; Tolosano, E.; Hirsh, E.; Silengo, L.; Alturda, F. *Biochim Biophys Acta* 1997, 1352, 193.
- Ludin, B.; Doll, T.; Meill, R.; Kaech, S.; Matus, A. *Gene* 1996, 173, 113.
- Cody, C. W.; Prasher, D. C.; Westler, W. M.; Pendergast, F. G.; Ward, W. W. *Biochemistry* 1993, 32, 1212.
- Cramer, A.; Whitehorn, E. A.; Tate, E.; Stemmer, W. P. C. *Natl Biotech* 1996, 14, 315-319.
- Heim, R.; Prasher, D. C.; Tsien, R. Y. *Proc Natl Acad Sci USA* 1994, 91, 12501.
- Reid, B. G.; Flynn, G. C. *Biochemistry* 1997, 36, 6786.
- Makino, Y.; Amada, K.; Taguchi, H.; Yoshida, M. *J Bio Chem* 1997, 272, 12468.
- Ward, W. W.; Bokman, S. H. *Biochemistry* 1982, 21, 4535.
- Bokman, S. H.; Ward, W. W. *Biochem Biophys Res Commun* 1981, 101, 1372.
- Cubitt, A. B.; Heim, R.; Adams, S. R.; Boyd, A. E.; Gross, L. A.; Tsien, R. Y. *Trends Biochem Sci* 1995, 20, 448.
- Yang, F.; Moss, L. G.; Phillips, Jr., G. N. *Natl Biotech* 1996, 14, 1246.
- Brejč, K.; Sixma, T. K.; Kitts, P. A.; Kain, S. R.; Tsien, R. Y.; Ormö, M.; Remington, S. J. *Proc Natl Acad Sci* 1997, 94, 2306.
- Ormö, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J. *Science* 1996, 273, 1392.
- Palm, G.; Zdanov, A.; Gaitanaris, G. A.; Stauber, R.; Pavlakis, G. N.; Wlodawer, A. *Natl Struct Bio* 1997, 4, 361.
- Nageswara Rao, B. D.; Kempel, M. D.; Pendergast, F. G. *Bio-phys J* 1980, 32, 630.
- Dopf, J.; Horiagon, T. M. *Gene* 1996, 173, 39.
- Zimmer, M.; Branchini, B.; Lusins, J. O. *Bioluminescence and Chemiluminescence, Proc. 9th Int. Symp.*; Hastings, J. W.; Kricka, L. J.; Stanley, P. E., Eds.; Wiley: New York, 1996; p. 407.
- Branchini, B. R.; Lusins, J. O.; Zimmer, M. *J Biomol Struct Dyn* 1997, 14, 441.
- Branchini, B. R.; Nemser, A. R.; Zimmer, M. *J Am Chem Soc* 1998, 120, 1.
- Becke, A. D. *Phys Rev* 1988, A38, 3098. Becke, A. D. *J Chem Phys* 1993, 98, 1372. Becke, A. D. *J Chem Phys* 1993, 98, 5648.
- Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J Chem Phys* 1991, 94, 7221.
- Bauschlicher, Jr., C. W.; Ricca, A.; Partridge, H.; Langhoff, S. R. In *Recent Advances in Density Functional Methods, Part II*; Chong, D. P., Ed.; World Scientific: Singapore, 1997.
- Siegbahn, P. E. M. *J Am Chem Soc* 1998, 120, 8417-8429.
- Glusker, J. P.; Bock, C. W.; Siegbahn, P. E. M.; George, P. *J Phys Chem B* 1999, 103, 7531-7541.
- Stevens, P. J.; Devlin, F. J.; Chabrowski, C. F.; Frisch, M. J. *J Phys Chem* 1994, 98, 11623.
- Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94, Revision B.2*; Gaussian: Pittsburgh, 1995.

36. Lee, C.; Yang, W.; Parr, R. G. *Phys Rev* 1988, B37, 785.
37. Vosko, S. H.; Wilk, L.; Nusair, M. *Can J Phys* 1980, 58, 1200.
38. Perdew, J. P.; Wang, Y. *Phys Rev B* 1992, 45, 13244; Perdew, J. P. In *Electronic Structure of Solids*; Ziesche, P.; Eischrig, H., Eds.; Akademie: Berlin, 1991. Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys Rev B* 1992, 46, 6671.
39. Siegbahn, P. E. M. *Adv Chem Phys*, Vol. XCIII; Prigogine, I.; Rice, S. A., Eds.; Wiley: New York, 1996.
40. Prabhakar, R.; Blomberg, M. R. A.; Siegbahn, P. E. M. *Theor Chem Acc*, to appear.
41. Pavlov, M.; Siegbahn, P. E. M.; Sandström, M. *J Phys Chem A* 1998, 102, 219–228.
42. Wiberg, K. B.; Rablen, P. R.; Rush, D. J.; Keith, T. A. *J Am Chem Soc* 1995, 117, 4261. Wiberg, K. B.; Keith, T. A.; Frisch, M. J.; Murcko, M. *J Phys Chem* 1995, 99, 9072.
43. Blomberg, M. R. A.; Siegbahn, P. E. M.; Babcock, G. T. *J Am Chem Soc* 1998, 120, 8812–8824.
44. Lee, F. S.; Chu, Z. T.; Warshel, A. *J Comput Chem* 1993, 14, 161–185.
45. Curtiss, L. A.; Raghavachari, K.; Pople, J. A. *J Chem Phys* 1993, 98, 1293.
46. Froese, D. J.; Humbel, S.; Svensson, M.; Morokuma, K. *J Phys Chem* 1997, 101, 227.
47. Mohamadi, F.; Richards, N. G. F.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caulfield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J Comput Chem* 1990, 11, 440.
48. Chang, G.; Guida, W. C.; Still, W. C. *J Am Chem Soc* 1989, 111, 4379.
49. Saunders, M.; Houk, K. N.; Wu, Y.-D.; Still, W. C.; Lipton, M.; Chang, G.; Guida, W. C. *J Am Chem Soc* 1990, 112, 1419.
50. We would like to thank the referee for suggesting these calculations.
51. Wright, H. T. *Crit Rev Biochem Mol Biol* 1991, 26, 1.
52. Ehrig, T.; O'Kane, D. J.; Pendergast, F. G. *FEBS Lett* 1995, 367, 163.