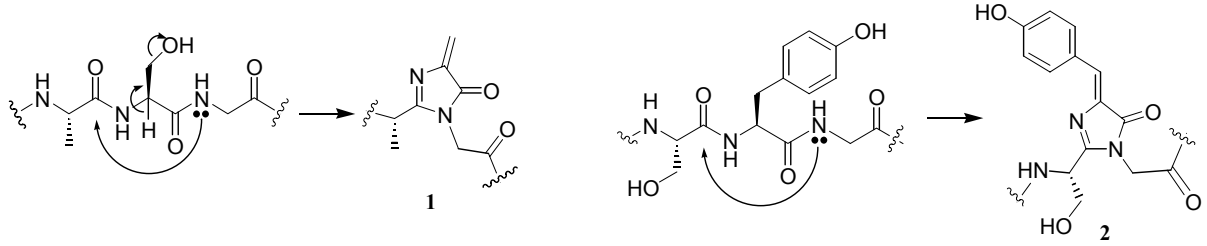
## FORMATION AND PROPERTIES OF IMIDAZOLONES IN PEPTIDES AND PROTEINS

Reported by Heather A. Relyea November 7, 2002

### INTRODUCTION

A variety of post-translational modifications have been identified in proteins. One unusual autocatalytic modification is the cyclization of tripeptides to form imidazolones. These structures have been found only as the electrophile **1** in the enzyme histidine ammonia lyase (HAL) and its analogs and as a fluorophore **2** in the green fluorescent protein (GFP) family of proteins (Scheme 1). These two groups of proteins are otherwise unrelated. HAL and its analog phenylalanine ammonia lyase (PAL) catalyze the non-oxidative deamination of aromatic amino acids, and have been employed in the asymmetric synthesis of non-natural aromatic amino acids[1] and in resolution procedures.[2] GFP and its analog DsRed have no catalytic function; however, it was recently suggested that they function as photoprotective agents.[3] These proteins are widely employed as biological imaging tools.[4]

**Scheme 1. Imidazolone moieties in HAL and GFP.**




While there does not appear to be an evolutionary link between these families of proteins, the two prototypes do generate their prosthetic groups through analogous pathways, and this process has been studied using both chemical and biochemical methods. Additionally, the mechanism of HAL has been investigated via physical organic techniques, and the fluorescent proteins provide an excellent example of how chemical knowledge can be used to re-engineer natural species for use in research.

### HISTIDINE AMMONIA LYASE

#### Synthetic Applications

Because HAL and PAL catalyze reversible reactions, they have synthetic utility for the preparation of enantiopure amino acids. While HAL is highly substrate-specific, PAL has been shown to work with a variety of phenylalanine derivatives and can therefore be used to synthesize non-natural aromatic amino acids.[1] Alternatively, HAL and PAL have both recently been shown to facilitate selective transport of one enantiomer across a membrane.[2] In this application,

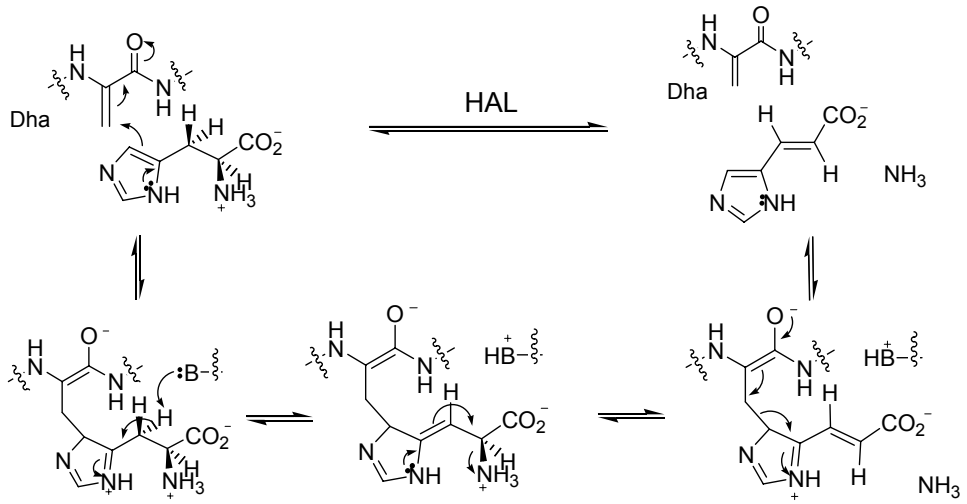
 

specific mutations are made that maintain substrate binding but block catalysis, thus rendering the enzymes as receptor/transporter entities.

**Mechanistic Studies of HAL**

Histidine ammonia lyase (HAL) deaminates histidine by abstracting the non-acidic β proton in the presence of a more acidic $-NH_3^+$ group. The mechanism was assumed to require an electrophile in the active site to enhance the acidity of the β-proton. The first major mechanistic studies, performed by Wickner in 1969,[5] concluded that the active site electrophile was a dehydroalanine (ΔAla). The site of modification was later determined to be serine residue 143.[6] In 1995 Rétey and coworkers proposed a mechanism for the deamination that invokes dehydroalanine (Scheme 2).[7] This scheme proceeds through attack of the enone β-carbon of ΔAla on the imidazole ring of histidine in a Friedel-Crafts type reaction, followed by $C_\beta$ deprotonation and an $E1_{cb}$-like elimination of ammonia. A similar mechanism is proposed for the analogous enzyme phenylalanine ammonia lyase (PAL), in which the phenyl ring of the substrate is attacked by ΔAla in the same manner.[8]

**Scheme 2. Proposed Mechanism of Histidine Deamination by HAL.**




**Discovery of MIO**

In 1999, Rétey and coworkers solved the crystal structure of mature HAL,[9] revealing a surprising structural moiety, 4-methylidine-imidazol-5-one (MIO) **1**, in the active site, rather than a dehydroalanine residue. While this cofactor did indeed arise from the serine at position 143, it also incorporated the alanine and glycine residues at the respective adjacent positions 142 and 144. MIO is the product of the Gly144 nitrogen attacking the carbonyl carbon of Ala142 followed by two dehydration steps (Scheme 1).

The authors proposed a revised mechanism of action for HAL which parallels their earlier dehydroalanine mechanism, and in fact only utilizes the dehydroalanine moiety within MIO. However, they note that MIO is a "modified dehydroalanine" with enhanced electrophilicity because the nitrogen of residue 143 can no longer contribute electron density to the enone system as it could in a linear system. Perhaps a more convincing argument for the necessity of MIO is that the imidazolone ring aromatizes[10] upon addition to the methylidine unit, thus compensating for the simulataneous loss of aromaticity in the imidazole ring of the substrate histidine (Figure 1).

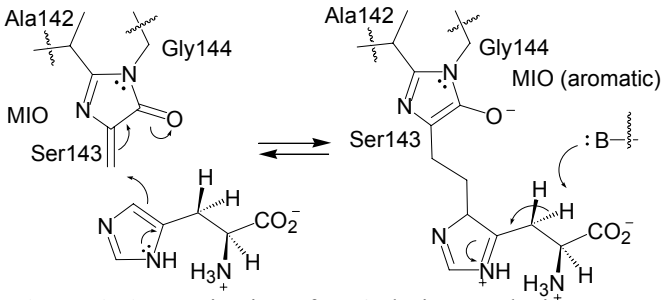


**Figure 1.** Aromatization of MIO during catalysis.

**Biogenesis of the Prosthetic Group**

Site-directed mutagenesis studies have shown that while modest substitutions in the [142]Ala-Ser-Gly tripeptide are tolerated lead to a MIO-like cyclized ene-imine,[6,10] changing the residues immediately adjacent to these prevents MIO formation (Table 1).[10,11] These data indicate that the structure and conformation of the immediate environment of the tripeptide have a profound effect on the ability of HAL to form MIO.[10] It has also been found that Phe329, which has been proposed to stabilize the aromatic form of MIO, is essential in MIO formation from the unmodified peptide.[11] MIO similarly cannot be formed when Asp141 (adjacent to the catalytic tripeptide) is replaced with the much smaller alanine, leading to the conclusions that (1) cyclization is a result of steric constraint and pre-organization (i.e., peptide deformation) in the active site and (2) cyclization occurs prior to dehydration of the central serine residue.[11]

**Table 1. Effects of Substitution on Catalytic Activity of HAL.**

| mutation | relative activity | MIO formation | ref. # |
|---|---|---|---|
| none | 1.00 | yes | 9 |
| Gly141Ala | 0.01 | N. A. | 10 |
| Ala142Gly | 0.93 | yes | 10 |
| Ser143Cys | 1.00 | yes | 6 |
| Gly144Ala | 0.37 | yes | 10 |
| Asp145Ala | -- | no | 11 |
| Phe329Ala | <0.01 | yes | 11 |
| Phe329Gly | -- | no | 11 |

In contrast to the cyclization-dehydration pathway, computational studies favor a dehydration-cyclization pathway.[12] These calculations suggest that dehydration of the central serine to form dehydroalanine brings the bond-forming atoms 144N and 142C closer together by inducing a tight turn conformation (Table 2). The dehydration step may then provide the

**Table 2. Distance between N and C in various peptides.**

| *type* | *C-N distance (Å)* |
|---|---|
| Average in GlySerAla | 3.44 |
| Average in GlyΔAlaAla | 3.1 |
| In HAL | 2.89 |

preorganization previously thought to arise from bulky adjacent residues. Additionally, thermodynamic factors favor the dehydration-cyclization pathway by approximately 6 kcal/mol.[12]

## GREEN FLUORESCENT PROTEIN

### Characterization of the Fluorophore

The peptide cyclization involved in biogenesis of MIO is analogous to the modification of the $^{65}$Ser-Tyr-Gly tripeptide **2** found as the fluorophore of green fluorescent protein (GFP) in jellyfish *Aequorea victoria*.[13] GFP is the best studied of the 27 known proteins in the GFP family, which includes proteins from reef coral. The chromophores in each are believed to arise from similar biogenesis pathways. It is believed that small differences in the environment around the fluorophore result in the observed variety of colors.[14]

Wild type GFP has excitation maxima at 395 nm and 475 nm, with an emission maximum at 508 nm.[4] Random mutagenesis was performed on this protein in an effort to obtain a chromophore with a single strong absorption maxima that could be selectively excited. One of the new proteins, Blue Fluorescent Protein (BFP) had a single excitation maxima at 382 nm; its emission at 448 nm results in a distinct and very strong blue fluorescence. This difference arose from substituting the tyrosine residue in the fluorophore with histidine. The protein was remarkably robust to this change, and the fluorophore was still formed autocatalytically.[15] Thus, although the structure of the chromophore has been altered, the machinery responsible for its biogenesis was not affected. This finding prompted efforts to develop other GFP mutants with different fluorescence emission wavelengths for use in *in vivo* labeling experiments.[16]

### Biogenesis of the GFP Fluorophore

In 1996, two independent crystal structures of GFP were solved, revealing that the fluorophore is isolated from the environment within an 11-stranded β-barrel structure (Figure 2).[17] The crystal structure confirmed the conclusion of Cody *et al.*, who in 1993 had used mass spectrometry and two-dimensional NMR techniques to elucidate the structure of the chromophoric peptide obtained from GFP by proteolysis.[13]

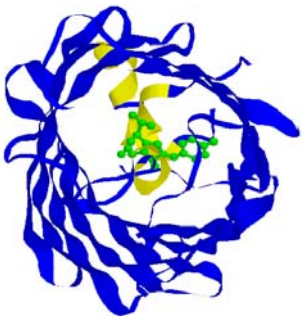
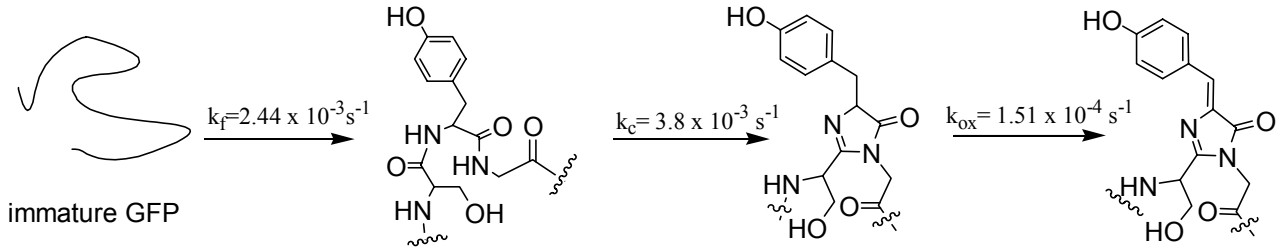


**Figure 2.** Structure of GFP.

As with MIO formation in HAL, fluorophore formation in GFP is believed to occur after protein folding, with cyclization promoted by steric constraints within the protein's cavity. On the basis of the crystal structure, it is believed that a nearby arginine residue (Arg96) induces a "tight turn" conformation in the tripeptide unit as is observed with MIO.[17] The computational studies that suggest

dehydration is the first step in MIO formation in HAL also suggest that dehydrogenation of tyrosine at the $C_\alpha$-$C_\beta$ bond is the first step in GFP fluorophore formation, again followed by cyclization.[12] However, *in vitro* kinetic studies show three distinct steps for fluorophore formation in the following order: (1) protein folding, $k_f = 2.44 \times 10^{-3}\ s^{-1}$, (2) cyclization and dehydration of the tripeptide, $k_c = 3.8 \times 10^{-3}\ s^{-1}$, and (3) oxidation of the cyclized peptide, $k_{ox} = 1.51 \times 10^{-4}\ s^{-1}$ (Scheme 3).[18] It is of note that an oxidation of tyrosine must occur to form the GFP fluorophore rather than simple dehydration as required for MIO formation, and molecular oxygen is believed to be the stoichiometric oxidant.[16]

**Scheme 3. Kinetics of Fluorophore Formation in GFP.**




## RATIONAL DESIGN OF ENGINEERED FLUORESCENT PROTEINS

### Development of Red-Shifted Analogs of GFP

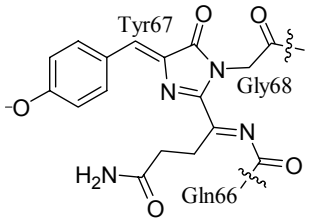
Early studies of GFP and BFP showed that it would indeed be possible to make fluorescent proteins in a variety of colors while maintaining the autocatalytic nature of fluorophore formation, thus allowing for genetically encoded *in vivo* labeling.[15] As GFP is small, non-reactive, and independent of other proteins, it is an ideal genetic fusion tag for tracking other proteins, both *in vitro* and in cells. The next step in developing robust and useful systems for *in vivo* sensing was to create proteins in which the chromophore would mature at a faster rate and which could exhibit red-shifted absorption and emission spectra, such that low-energy irradiation could be used to induce fluorescence, and fluorescence resonance energy transfer (FRET) with longer-wavelength fluorophores would be possible.

Based on the crystal structure of GFP, the hydroxyl group of a threonine residue adjacent to the chromophore (Thr203) was proposed to destabilize the aromatic species. This threonine was systematically replaced with the aromatic residues histidine, tryptophan, and tyrosine in an effort to stabilize the chromophore and thereby induce lower energy (red-shifted) emissions. These attempts were successful, and emission maxima >520 nm were attained,[17] compared to 508 nm for wild type GFP. The variant Thr203Tyr was optimized through directed evolution and dubbed Yellow Fluorescent Protein (YFP). The structural basis of the red shift was rationalized through π-stacking of the newly

introduced tyrosine residue with the aromatic chromophore, as was revealed through crystallographic studies.[19]

**Discovery and Characterization of DsRed**

In a search for natural variants of GFP, Matz *et al.* screened other marine life forms for genetic homologs.[20] A red-emitting analog found in coral absorbs at 558 nm and emits at 583 nm in its mature form, these being far lower energies than GFP and its variants to date. This protein was named DsRed and is now commercially available from CLONTECH. The tripeptide sequence which is cyclized in this analog is Gln-Tyr-Gly, and in its mature form the fluorophore **3** is related to that of GFP, but has an extra acyl imine (from the Gln residue) that is conjugated to the rest of the system.[21] This extended conjugation is believed to give rise to the longer wavelength emission, as supported by recent model fluorophore studies.[22] The fluorophore structure was confirmed by two independent crystal structures at 1.9 and 2.0 Å resolution.[23,24] When the final oxidative step, presumably dependent on oxygen, is inhibited, DsRed exhibits the same spectral characteristics as GFP.[21,25] This protein has an unusual *cis* peptide bond leading into the tripeptide sequence, which is believed to promote the final slow oxidation step that shifts the emission maxima from green to red (the analogous linkage in GFP is in the standard *trans* conformation). The overall monomeric structure of DsRed is identical to GFP; however, in its native form DsRed is a homotetramer.[23]



**3**

## *IN VIVO* APPLICATIONS OF FLUORESCENT PROTEINS

**Production of Viable *In Vivo* Fluorescent Labels**

While DsRed exhibits highly desirable spectral properties for FRET and other applications, its *in vivo* use was initially hindered by its propensity to form a tetramer, as well as by its slow maturation. The tetramerization of DsRed is problematic because it can cause non-native aggregation of its fusion protein during tagging experiments.[26] Tsien and coworkers were able to overcome DsRed aggregation by combining directed and random mutagenesis techniques.[27] They were able to first create a dimer by changing specific hydrophobic residues on one plane of contact in the tetramer. Several libraries were made to create a monomeric species, and these were then subjected to directed evolution to optimize fluorescence. The final protein contained 33 mutations as compared to the wild type.[27]

The fluorophore maturation process has been accelerated through rational engineering of the proteins.[28] A recent study suggests that there is a structural basis for the rate of fluorophore development from green to red.[29] Comparisons of several variants of DsRed with a range of maturation

rates led to the conclusion that increased space around the fluorophore accelerates the rate at which red fluorescence develops and reduces residual green emission. This seems to contradict earlier reports that steric strain actually *promotes* fluorophore formation. However, these conclusions refer to residues in different regions of the protein with respect to the tripeptide responsible for fluorophore formation.

**Recent Applications in Imaging**

An interesting offshoot of the attempts to produce a faster-maturing red fluorophore is the "Fluorescent Timer", which evolves from green to red fluorescence at a known rate.[30] When fused with a target protein, the up- or down-regulation of that target in living systems in response to stimuli can be monitored in a time-dependent fashion. Since fluorescence only appears when the target protein is being produced, relative intensities of green and red emissions can be used to determine when production began (or stopped) relative to the time at which a stimulus was administered.

In a slightly different application, GFPs can be used to image cell compartmentalization and determine regulatory features. Monomeric cyan fluorescent protein (CFP) and YFP are used as markers and attached to various lipid modifiers.[31] Researchers were able to directly determine the type of lipid modification that causes protein aggregation to form "lipid rafts" in cell membranes. By measuring the intensity of FRET between the cyan and yellow FPs, it was possible to map out areas of protein aggregation without destroying the cells as had previously been necessary.

While fluorescent proteins have been used for some time as $Ca^{2+}$ indicators,[4] a more advanced BFP sensor has been developed that can detect both $Zn^{2+}$ and $Cu^{2+}$, giving different fluorescence characteristics for each. Crystallographic studies indicate that the fluorophore itself is a ligand for both metals, but the differing coordination geometries lead to differing wavelengths and emission intensities.[32] However, these sensors currently have only micromolar sensitivity, while *in vivo* applications would require nanomolar sensitivity.

## CONCLUSION

There are numerous applications of fluorescent proteins, and the utility of ammonia lyases has been demonstrated as well. The unique properties of each of these proteins can be attributed to their cyclic prosthetic groups. Each is formed autocatalytically, initiated by specific peptide conformations. Studies of the formation of these cyclic moieties and how they function when matured have led to a variety of rational engineering designs. These studies also exhibit the use of organic chemistry techniques to understand and manipulate biological systems.

**REFERENCES**


(1) Liu, W. **1999** U.S. Patent 5981239
(2) Skolaut, A.; Retey, J. *Angew. Chem. Intl. Ed.* **2002**, *41*, 2960-2962.
(3) Salih, A.; Larkum, A.; Cox, G.; Kuhl, M.; Hoegh-Guidberg, O. *Nature* **2000**, *408*, 850-853.
(4) Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509-544.
(5) Wickner, R. B. *J. Biol. Chem.* **1969**, *244*, 6550-2.
(6) Langer, M.; Lieber, A.; Retey, J. *Biochemistry* **1994**, *33*, 14034-8.
(7) Langer, M.; Pauling, A.; Retey, J. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 1464-5.
(8) Rother, D.; Poppe, L.; Morlock, G.; Viergutz, S.; Retey, J. *Eur. J. Biochem.* **2002**, *269*, 3065-3075.
(9) Schwede, T. F.; Retey, J.; Schulz, G. E. *Biochemistry* **1999**, *38*, 5355-5361.
(10) Baedeker, M.; Schulz, G. E. *Eur. J. Biochem.* **2002**, *269*, 1790-1797.
(11) Baedeker, M.; Schulz, G. E. *Structure* **2002**, *10*, 61-67.
(12) Donnelly, M.; Fedeles, F.; Wirstam, M.; Siegbahn, P. E.; Zimmer, M. *J. Am. Chem. Soc.* **2001**, *123*, 4679-4686.
(13) Cody, C. W.; Prasher, D. C.; Westler, W. M.; Prendergast, F. G.; Ward, W. W. *Biochemistry* **1993**, *32*, 1212-18.
(14) Labas, Y. A.; Gurskaya, N. G.; Yanushevich, Y. G.; Fradkov, A. F.; Lukyanov, K. A.; Lukyanov, S. A.; Matz, M. V. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 4256-4261.
(15) Wachter, R. M.; King, B. A.; Heim, R.; Kallio, K.; Tsien, R. Y.; Boxer, S. G.; Remington, S. J. *Biochemistry* **1997**, *36*, 9759-9765.
(16) Heim, R.; Prasher, D. C.; Tsien, R. Y. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 12501-4.
(17) Ormo, M.; Cubitt, B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J. *Science* **1996**, *273*, 1392-1395.
(18) Reid, B. G.; Flynn, G. C. *Biochemistry* **1997**, *36*, 6786-6791.
(19) Wachter, R. M.; Elsliger, M.-A.; Kallio, K.; Hanson, G. T.; Remington, S. J. *Structure* **1998**, *6*, 1267-1277.
(20) Matz, M. V.; Fradkov, A. F.; Labas, Y. A.; Savitsky, A. P.; Zaraisky, A. G.; Markelov, M. L.; Lukyanov, S. A. *Nat. Biotech.* **1999**, *17*, 969-73.
(21) Gross, L. A.; Baird, G. S.; Hoffman, R. C.; Baldridge, K. K.; Tsien, R. Y. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 11990-11995.
(22) He, X.; Bell, A. F.; Tonge, P. J. *Org. Lett.* **2002**, *4*, 1523-1526.
(23) Wall, M. A.; Socolich, M.; Ranganathan, R. *Nature Structural Biology* **2000**, *7*, 1133-1138.
(24) Yarbrough, D.; Wachter, R. M.; Kallio, K.; Matz, M. V.; Remington, S. J. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 462-467.
(25) Baird, G. S.; Zacharias, D. A.; Tsien, R. Y. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 11984-11989.
(26) Mizuno, H.; Sawano, A.; Eli, P.; Hama, H.; Miyawaki, A. *Biochemistry* **2001**, *40*, 2502-2510.
(27) Campbell, R. E.; Tour, O.; Palmer, A. E.; Steinbach, P. A.; Baird, G. S.; Zacharias, D. A.; Tsien, R. Y. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 7877-7882.
(28) Bevis, B. J.; Glick, B. S. *Nat. Biotech.* **2002**, *20*, 83-87.
(29) Terskikh, A. V.; Fradkov, A. F.; Zaraisky, A. G.; Kajava, A. V.; Angres, B. *J. Biol. Chem.* **2002**, *277*, 7633-7636.
(30) Terskikh, A.; Fradkov, A.; Ermakova, G.; Zaraisky, A.; Tan, P.; Kajava, A. V.; Zhao, X.; Lukyanov, S.; Matz, M.; Kim, S.; Weissman, I.; Siebert, P. *Science* **2000**, *290*, 1585-1588.
(31) Zacharias, D. A.; Violin, J. D.; Newton, A. C.; Tsien, R. Y. *Science* **2002**, *296*, 913-916.
(32) Barondeau, D. P.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D. *J. Am. Chem. Soc.* **2002**, *124*, 3522-3524.