Computational Analysis of the Autocatalytic Posttranslational Cyclization Observed in Histidine Ammonia-Lyase. A Comparison with Green Fluorescent Protein

Maria Donnelly,[†] Flavia Fedeles,[†] Maria Wirstam,[‡,§] Per E. Siegbahn,*[,‡] and Marc Zimmer*[,†]

Contribution from the Chemistry Department, Connecticut College, 270 Mohegan Avenue, New London, Connecticut 06320, and Department of Physics, Stockholm University, Box 6730, S-113 85 Stockholm, Sweden

Received November 20, 2000. Revised Manuscript Received March 20, 2001

Abstract: Density functional calculations using hybrid functionals (B3LYP) have been performed to study the mechanism of the autocatalytic posttranslational cyclization observed in histidine ammonia-lyase. Two mechanisms were analyzed, the commonly accepted mechanism in which cyclization precedes dehydrogenation (reduced mechanism) and a mechanism in which dehydrogenation precedes cyclization (oxidized mechanism). The reduced pathway is not supported by the calculations, while the alternative oxidized mechanism where a dehydration occurs prior to the formation of the ring yields reasonable energetics for the system. Database searches showed that the oxidative mechanism in which the formation of the dehydro amino acids in residue $i + 1$ precedes the cyclization is also structurally advantageous as it results in shorter distances between the carbonyl carbon of residue i and the amide nitrogen of residue $i + 2$ and, therefore, preorganizes the protein for cyclization. Conformational searches showed that these distances were also unusually short and exhibited very little variation in the Δ-Ala143 HAL tetramer, indicating that like GFP the tetrameric form of HAL is rigidly preorganized for cyclization. The monomeric form of HAL is less preorganized than the tetrameric form of HAL. Dehydro amino acids aid in the preorganization, but the main driving force in the rigid tight turn formation is the influence of the surrounding protein.

Introduction

The recent sequencing of the human genome has produced "the most important, most wondrous map ever produced by humankind".[1] To utilize this information and to better understand protein structure, a large effort is underway to use the amino acid sequence to determine the three-dimensional structure of proteins. We have been particularly interested in how specific protein sequences can in rare instances lead to autocatalytic posttranslational intrachain cyclizations[2] and are responsible for increasing structural and functional diversity.[3]

Autocatalytic intrachain posttranslational cyclizations have only been reported for green fluorescent protein (GFP), GFP homologues from nonbioluminescent Anthozoa species, and histidine ammonia-lyase (HAL).

The bioluminescence of green fluorescent protein is due to the excitation of a cyclic chromophore, which is formed posttransationally from its own backbone structure through the attack of the amide nitrogen of Gly67 on the carbonyl carbon of Ser65, Figure 1.[4−6] This process is unusual because, while most posttranslational protein ring formations are produced only with the assistance of one or more catalysts, the cyclization in GFP occurs autocatalytically.[2,4,6] Density functional calculations

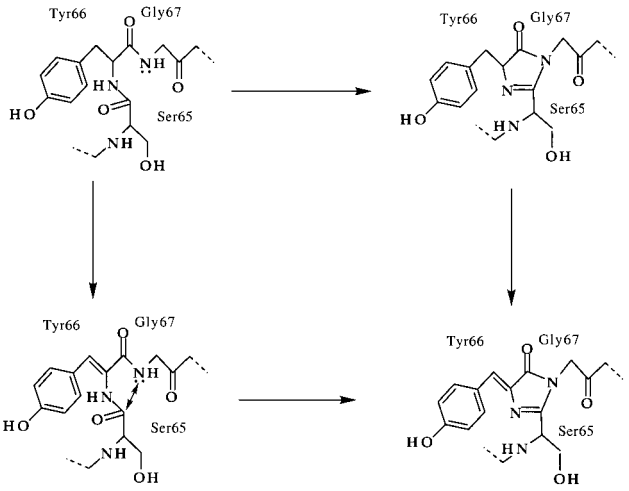

Figure 1. Proposed reaction sequence for the formation of the chromophore in GFP. The sequence going through the top right is the conventional reduced mechanism, while that going through the bottom left is the oxidized mechanism predicted by DFT calculations.

[†] Connecticut College.

[‡] Stockholm University.

[§] Current address: Department of Chemistry, Columbia University, 3000 Broadway, MC 3131, New York, NY 10027.

(1) Clinton, W. J. B. III. White House, June 26, 2000.

(2) Branchini, B. R.; Nemser A. R.; Zimmer, M. *J. Am. Chem. Soc.* **1998**, *120*, 1. Branchini, B. R.; Lusins, J. O.; Zimmer, M. *J. Biomol. Struct., Dyn.* **1997**, *14*, 441. Branchini, B. R.; Lusins, J.; Zimmer, M. In *Proceedings of the 9th International Symposium on Bioluminescence and Chemiluminescene*; Hastings, J. W., Kricka, L. J., Stanley, P. E., Eds.; John Wiley & Sons: Chichester, U.K., 1996; p 407.

(3) Okeley, N. M.; van der Donk, W. A. *Chem., Biol.* **2000**, *7*, R159.

(4) *Green Fluorescent Protein Properties, Applications, and Protocols*; Chalfie, M., Kain, S., Eds.; Wiley-Liss: New York, 1998.

(5) Heim, R.; Prasher, D. C.; Tsien, R. Y. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12501.

(6) Cubitt, A. B.; Heim, R.; Adams, S. R.; Boyd, A. E.; Gross, L. A.; Tsien, R. Y. *Trends Biochem. Sci.* **1995**, *20*, 448.

10.1021/ja004009c CCC: $20.00
Published on Web 05/01/2001

have led to the proposal that dehydration might precede cyclization in GFP.[7] Initially it was believed that the autocatalytic GFP cyclization was unique, but recent research has indicated that a family of enzymes, including histidine ammonia-lyase (HAL) and the closely related phenylalanine ammonia-lyase (PAL), also contains posttranslational ring formations that occur autocatalytically through the attack of the protein backbone on itself.[8] The apparent rarity of these autocatalytic cyclization processes and their presence in biologically active proteins has made them a very interesting area of study.

It has been suggested that the autocatalytic posttranslational cyclization observed in GFP is due to the fact that the residues involved in the cyclization are preorganized into a tight turn conformation[2] within a rigid β-can; see Figure 2.[9]

Molecular biological methods were used to find GFP homologues in brightly colored corals by searching for proteins that have sequence homology with the residues in the turns between the β-sheets of GFP and the α-helix sequences that hold the chromophore in place in the center of the GFP β-can.[10] Six GFP homologues were found in the corals. They are remarkably similar in size and have the same overall β-can structure of GFP, as evidenced by the similar distances found between the tight turns separating the β-sheets. The top (residues 82−91) and the bottom (residues 129−140) of the cans are also extremely well conserved. Both Tyr66 and Gly67, which form the chromophore in GFP, are conserved in wild-type GFP and in all six anthozoa species.

The recently published crystal structure of histidine ammonia-lyase revealed that it has undergone a posttranslational autocatalytic cyclization.[8] However, the site of the cyclization is not located in the center of a rigid β-can structure; in fact, it is not far from the surface of the protein (see Figure 2). Histidine ammonia-lyase, which is commonly referred to as histidase or HAL, catalyzes the nonoxidative β-elimination of the α-amino group of L-histidine to produce $\alpha-\beta$-unsaturated *trans*-urocanate and ammonia in the first step of histidine metabolism.[8,11,12] It has been suggested that urocanate acts as a solar radiation blocker in the human skin.[8,13] In its native state, histidase is a tetrameric protein consisting of four identical 53 000 Da chains.[12,14] The β-elimination that histidase catalyzes is very interesting because it requires an electrophilic group to abstract a very nonacidic proton that is bound to the carbon β to the extremely acidic α-amino group.[8,11,15,16] In order for it to provide a good leaving group, the α-amino moiety must remain protonated throughout this reaction.[8] Early studies based on the identification of derivatives of inactivated histidase indicated that the electrophilic site was a dehydroalanine residue.[12,14,17,18] The proposed existence of a dehydroalanine residue was based on indirect evidence, and the origin and location of the residue

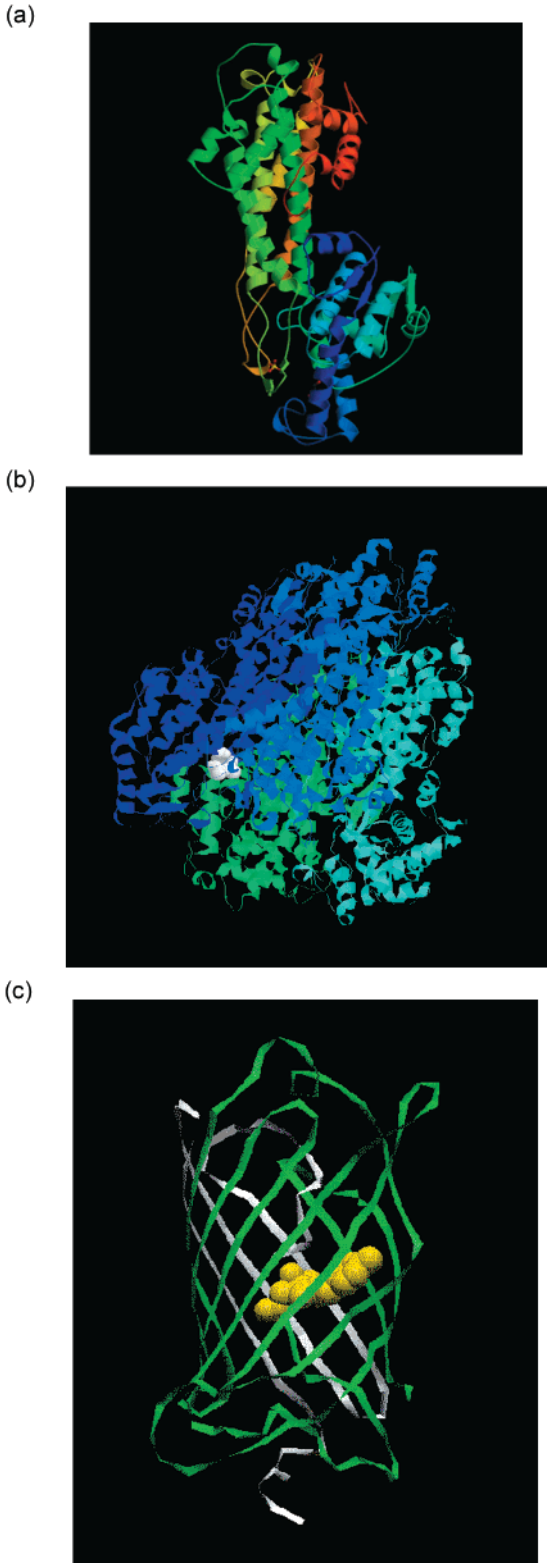

Figure 2. (a) Crystal structure of the HAL monomer. (b) Crystal structure of the HAL tetramer. The white spheres in the tetramer indicate the cyclization site. (c) In GFP, the chromophore is represented by the yellow spheres; it is located in the center of the canlike structure.

(7) Siegbahn, P. E. M.; Wirstam, M.; Zimmer M. *Int. J. Quantum Chem.* **2001**, *81*, 169.
(8) Schwede, T. F.; Rétey, J.; Schulz, G. E. *Biochemistry* **1999**, *38*, 5355.
(9) Yang, F.; Moss, L. G.; Phillips, G. N., Jr. *Nat. Biotechnol.* **1996**, *14*, 1246. Brejc, K.; Sixma, T. K.; Kitts, P. A.; Kain, S. R.; Tsien, R. Y.; Ormoe, M.; Remington, S. J. *Proc. Nat. Acad. Sci. U.S.A.* **1997**, *94*, 2306. Ormoe, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J. *Science* **1996**, *273*, 1392.
(10) Matz, M. V.; Fradkov, A. F.; Labas, Y. A.; Savitisky, A. P.; Zaraisky, A. G.; Markelov, M. L.; Lukyanov, S. A. *Nature Biotechnol.* **1999**, *17*, 969.
(11) Galpin, J. D.; Ellis, B. E.; Tanner, M. E. *J. Am. Chem. Soc.* **1999**, *121*, 10840.
(12) Hernandez, D.; Stroh, J. G.; Phillips, A. T. *Arch. Biochem. Biophys.* **1993**, *307*, 126−132.
(13) Morrison, H.; Bernasconi, C.; Pandey, G. *Photochem. Photobiol.* **1984**, *40*, 549.
(14) Consevage, M. W.; Phillips, A. T. *Biochemistry* **1985**, *24*, 301−308.
(15) Anderson, V. E. *Comprehensive Biological Catalysis*; Academic Press: San Diego, 1998; Chapter 23.
(16) Hanson, K. R.; Havir, E. A. In *The Enzymes*; Boyer, P. D., Ed.; Academic Press: New York, 1972; Vol. 7, pp 75−166.
(17) Wickner, R. B. *J. Biol. Chem.* **1969**, *244*, 6550−6552.
(18) Givot, I. L.; Smith, T. A.; Albeles, R. H. *Arch. Biophys. Biochem.* **1969**, *244*, 6341−6353.

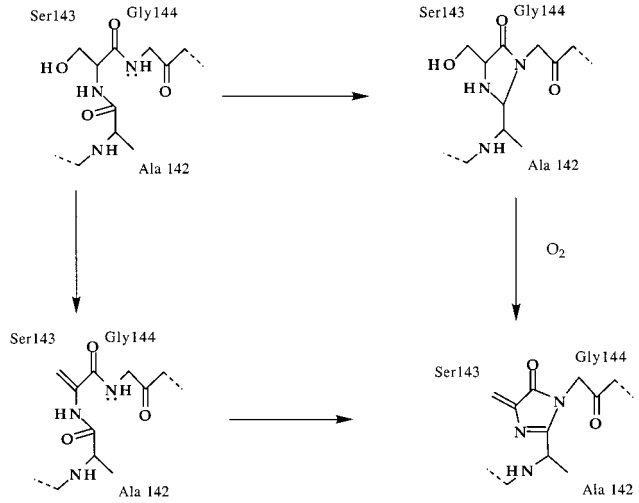

Figure 3. Proposed reaction sequence for the ring formation in HAL. The sequence going to the top right is the conventional reduced mechanism, while that going through the bottom left is the oxidized mechanism predicted by DFT calculations.

in the amino acid sequence could not be determined.[12] In 1993, Hernandez et al. published a paper showing the presence of an excess mass of 184 Da with a UV absorption at 340 nm in L-cysteine-modified histidase. The UV absorption and molecular weight were not consistent with a simple derivative of dehydroalanine.[12] Then, in 1999, Schwede et al. published the crystal structure of histidase.[8] The crystal structure of HAL indicates that its electrophilic functionality is a 4-methylideneimidazol-5-one (MIO) moiety[8,11] rather than a dehydroalanine residue. This ring structure is formed through an autocatalytic posttranslational modification of the residue sequence Ala142-Ser143-Gly144. A proposed reaction sequence for this cyclization is shown in Figure 3.[8] The reaction sequence is very similar to the chromophore-forming reactions of GFP, Figure 1.

Histidase is closely related to phenylalanine ammonia-lyase, or PAL, an enzyme that mediates the production of (*E*)-cinnamic acid through the reversible elimination of ammonia from phenylalanine[19] in a manner that is very similar to the breakdown of histidine into urocanic acid by HAL. Like HAL, PAL was believed to have a dehydroalanine residue, which had been produced autocatalytically from a serine,[8] at its active site.[19] Although the crystal structure of phenylalanine ammonia-lyase has not yet been solved, the recent discoveries about the MIO functionality in HAL, and the similarities between the catalytic processes of the two enzymes, seem to indicate that PAL is likely to contain a similar autocatalytically formed ring system.

Even though GFP and HAL have different tertiary structures (see Figure 2), they have many similarities. The chromophore of GFP and the MIO moiety of HAL are chemically similar (see Figures 1 and 3). In both HAL and GFP the residue following the cyclization site is a glycine, presumably due to its minimal steric hindrance. They are also both members of a small group of proteins in which an autocatalytic intrachain posttranslational cyclization occurs.

In this paper, we have attempted to determine whether the rigid tight turn preorganization found around the cyclization site of GFP is a prerequisite to posttranslational autocatalytic cyclization reactions. We have also used MM and DFT calculations to determine whether the oxidized mechanism we have suggested for GFP[7] is a low-energy pathway in the cyclization of HAL. In our GFP calculations,[2,7] we did not examine the effect the dehydro amino acids have on the precyclization tight turn conformation. Here we describe our findings that the formation of the dehydro amino acid can aid in the preorganization of the cyclization site.

Experimental Section

The DFT calculations were performed in three steps using a procedure similar to that described for our GFP calculations.[7] 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[20,21] using the Gaussian-94 program.[22]

For all 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. In the B3LYP geometry optimizations, a much smaller basis set, the d95 set of the Gaussian-94 program, was used. The use of such a small basis was tested on the GFP systems studied in the previous investigation.[7] Increasing the basis in the geometry optimization by adding polarization functions changed the relative final energies (calculated using the big basis above) by at most a few tenths of a kilocalorie per mole. This is in line with previous experience, which shows that the final relative energies are extremely insensitive to the quality of the geometry optimization.[23,24] To clarify, this means that even when the geometries optimized with the small basis sets are used, relative energies are accurately reproduced. This is true even though the larger basis sets change C−C, C−N, and C−O distances by two to three hundreds of an angstrom. On the basis of 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.

All energies discussed below include zero-point effects, calculated at the level used for geometry optimization. For the present model reactions, the relative effects of zero-point vibration are found to be quite small, in the range 0−3 kcal/mol. Thermal effects are found to be less important and are not discussed here.

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.[22,25] A dielectric constant of 4.0 was used in these calculations. In the present case where neutral models are chosen throughout, the details of how the dielectric effects are computed are quite unimportant since these effects are found to be very small, with a maximum effect of 1.15 kcal/mol and an average absolute effect of only 0.68 kcal/mol on the relative energies reported for HAL in Tables 1 and 2. Positively charged models were not used in the present study but were investigated in the previous study of GFP.[7] Even for these charged models, the relative dielectric effects tend to be small as long as the charge is the same from model to model.

(19) Lewandowicz, A.; Jemielity, J.; Kanska, M.; Zon, J.; Paneth, P. *Arch. Biochem. Biophys.* **1999**, *370*, 216−221.

(20) 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.

(21) Stevens, P. J.; Devlin, F. J.; Chablowski, C. F.; Frisch, M. J. *J. Phys Chem.* **1994**, *98*, 11623.

(22) 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 Inc.: Pittsburgh, PA, 1995.

(23) Bauschlicher, C. W., Jr.; Ricca, A.; Partridge, H.; Langhoff, S. R. In *Recent Advances in Density Functional Methods, Part II*; Chong, D. P., Ed.; World Scientific Publishing Co.: Singapore, 1997; p 165.

(24) Siegbahn, P. E. M. In *Advances in Chemical Physics*; Prigogine, I., Rice, S. A., Eds.; J. Wiley: New York, 1996; Vol. XCIII, p 333.

(25) 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.

Table 1. Relative Energies (in kcal/mol) for the Reduced Mechanism Investigated for HAL[a] and GFP[b]

	linear (Figure 4a)	dihydroimidazolone (Figure 4b)	imidazolone (Figure 4c)
HAL	0.0	15.6	4.1
GFP	0.0	16.2	10.6

[a] The optimized structures are shown in Figure 4. [b] The values are taken from ref 7.

Table 2. Relative Energies (in kcal/mol) for the Oxidized Mechanism Investigated for HAL[a] and GFP[b]

	linear (Figure 5a)	dihydroimidazolone (Figure 5b)	imidazolone (Figure 5c)
HAL	0.0	4.5	−2.1
GFP	0.0	6.7	−1.9

[a] The optimized structures are shown in Figure 5. [b] The values are taken from ref 7.

Molecular Mechanics. The coordinates for the crystal structure of histidine ammonia-lyase were obtained from the file 1B8F[8] in the Protein Data Bank. MacroModel7.0[26] was used to graphically replace the posttranslationally formed ring with the precyclized amino acid sequence. Hydrogen atoms were added to both protein atoms and solvent atoms as required.

A 12-Å "hot" sphere from residues 142−144 with a secondary constrained sphere of 3 Å was used in all Monte Carlo conformational searches[27] performed on the monomer. The backbone and side-chain torsion angles of residues 142−144 were randomly varied between 0 and 180°. The amide backbone bonds of residues 140 and 146 were designated as closure bonds with minimum and maximum closure distances of 1 and 4 Å, respectively. The closure bonds anchor the residues 141−145 by acting as "rubber bands" between the protein fragment that is distorted in each MC step and the remainder of the protein. After each MC step, closure bonds between 1.00 and 4.00 Å are minimized like any other bond using the appropriate AMBER* force constant. The side-chain torsion angles of all residues located in a sphere of 8 Å around the site of the posttranslational cyclization were also randomly varied between 0 and 180°. The chirality of all backbone carbons was maintained. The crystallographically determined water molecules were incorporated into all calculations. All conformations within 50 kJ/mol of the global minimum were saved.

The coordinates for the monomer were converted into the tetrameric structure with the IMB Jena Image Library (http://www.imb-jena.de/IMAGE.html).[28] Monte Carlo conformational searches of the tetrameric structure used a 10-Å "hot" sphere from residues 142−144 of chain A, with a secondary constrained sphere of 3 Å. All the torsion angles in the backbone of residues 142−144 and all the side-chain torsion angles in an 8-Å sphere around the cyclization site were randomly varied between 0 and 180°. The same closure bonds were used as in the monomeric structures. The chirality of the protein backbone was maintained. All conformations within 50 kJ/mol of the global minimum were saved.

All Monte Carlo conformational searches performed on HAL used the AMBER* force field. Each calculation consisted of 10 000 MC steps with 500 iterations per step. Structures were considered to be unique when a pair of related atoms was separated by more than 0.025 Å after a least-squares superposition of all non-hydrogen atoms.

The molecular dynamics simulations used a 10 Å + 3 Å sphere centered at the cyclization site. A 10-ps preequilibrium run was started from the fully minimized lowest energy structure found in the Monte Carlo search with an initial temperature of 300 K and time steps of 1.5 fs. The actual simulations were run at 300 K for 1 ns with 1.5-fs time steps. All molecular dynamics calculations used the AMBER* force field and SHAKE[29,30] constrained hydrogens.

PDB searches were used to locate other proteins that contained the sequences VGASG, GASGD, GASG, and ASGD. The Cambridge-Structural Database (CSD) version 5.19, which contains 215 000 structures and was last updated in April 2000, was searched for the substructures shown in Figure 7.

Results and Discussion

Mechanism. Density functional calculations using hybrid functionals (B3LYP) have been performed to study the mechanism of the autocatalytic posttranslational cyclization observed in histidine ammonia-lyase.

In our earlier work, several different models and methods were used to investigate the autocatalytic cyclization of GFP.[7] The results obtained from these investigations indicate that the most commonly suggested mechanism (referred to as the reduced mechanism) is not energetically favorable, but leads to large endothermicities for formation of the imidazolone ring and the required intermediate, irrespective of what chemical models and computational methods used. This led to a proposal of an alternative mechanism where the cyclization is preceded by the dehydration of Tyr66 to dehydrotyrosine, which was found to be much more probable from an energetics point of view, resulting in a less unstable intermediate and reaction energies close to thermoneutral. This is the oxidized mechanism (see Figure 1). One argument that has been used for the reduced mechanism is the analogy to the well-established deamination step of Asn-Gly sequences in peptides and proteins.[31] However, when exactly the same computational methods were employed for the first step in the deamination process, it was found to be exothermic. The exothermicity was found to be 2 kcal/mol. These results thus show that the two chemical systems actually are significantly different.

One purpose of the quantum chemical calculations performed in the present study is to investigate whether the oxidized mechanism suggested for GFP[7] is also applicable for HAL. As shown in Figures 4 and 5, the quality of the chemical model used for the calculations on HAL corresponds to that of the largest model used for GFP. The crystal structures of GFP and HAL reveal that two of the three amino acids directly involved in the ring formation are different in the two enzymes. While both systems contain a glycine at a similar position, the Tyr66 site in GFP is occupied by a serine (Ser143) in HAL. Furthermore, the position of Ser65 of GFP is occupied by an alanine (Ala142) in HAL. Consequently, the phenol group used as a model for Tyr66 in the previous study of GFP[7] is replaced by the hydroxyl group of Ser143 in the present investigation of HAL. Likewise, the hydroxyl group used as a model for Ser65 in GFP is substituted by a hydrogen atom in the present study.

Tables 1 and 2 show the energetics obtained for HAL for the oxidized and the reduced mechanism. In addition, the corresponding results for GFP are given for comparison. In Figures 4 and 5, the optimized structures obtained for the model calculations on HAL are shown.

As shown in Tables 1 and 2, the energetics obtained are rather similar for the two systems. For the reduced mechanism, the HAL and GFP models both yield an endothermicity of ∼16 kcal/mol for formation of the required dihydroimidazolone intermediate, shown for HAL in Figure 4b. However, the

(26) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, C.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440.

(27) Chang, G.; Guida W. C.; Still, W. C. *J. Am. Chem. Soc.* **1989**, *111*, 4378. Chang, G.; Guida W. C.; Still, W. C. *J. Am. Chem. Soc.* **1990**, *112*, 1419.

(28) Sühnel, J. *Comput. Appl. Biosci.* **1996**, *12*, 227−229.

(29) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

(30) Ryckaert, J. P. *Mol. Phys.* **1985**, *55*, 549.

(31) Cubitt, A. B.; Heim, R.; Adams, S. R.; Boyd, A. E.; Gross, L. A.; Tsien, R. Y. *Trends Biochem, Sci*. **1995**, *20*, 448.

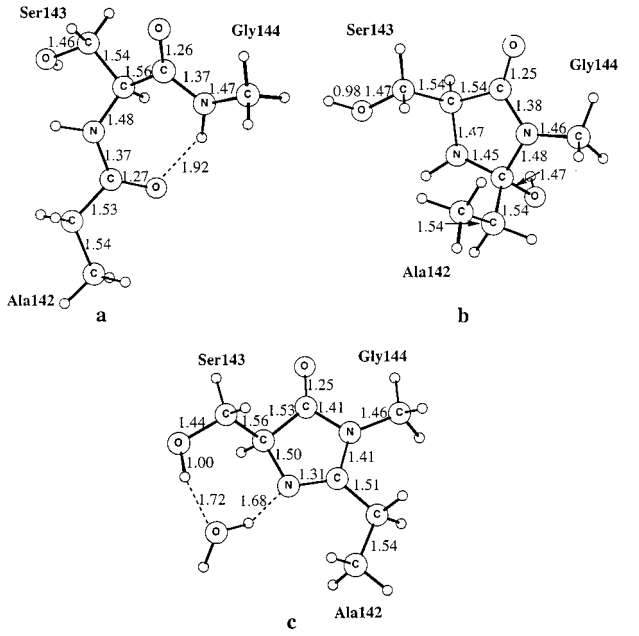

Figure 4. Fully optimized model structures for intermediates involved in the reduced mechanism investigated for HAL. Distances are given in angstroms. The relative energies are given in Table 1.

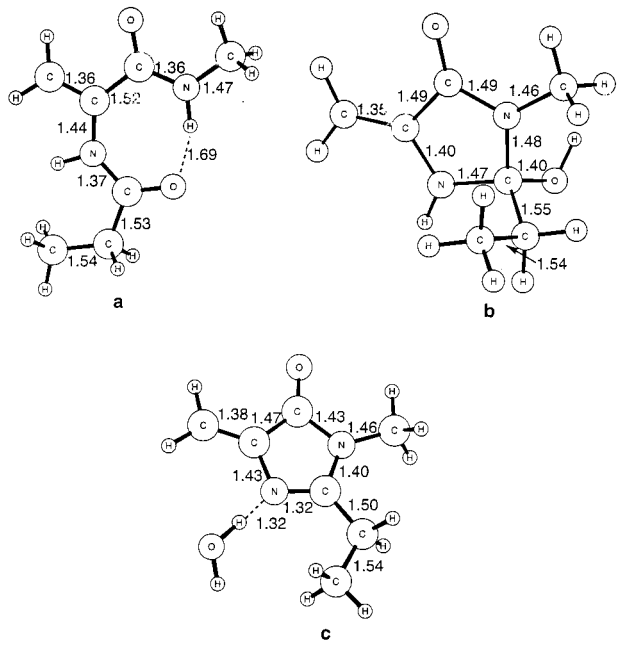

Figure 5. Fully optimized model structures for intermediates involved in the oxidized mechanism investigated for HAL. Distances are given in angstroms. The relative energies are given in Table 2.

calculated endothermicity for formation of the imidazolone ring differs somewhat between the two systems. While an endothermicity of 10.6 kcal/mol is obtained for GFP, the corresponding value obtained for the HAL model shown in Figure 4c is only 4.1 kcal/mol. This difference between the HAL and GFP models can be explained by the number of hydrogen bonds formed for the water produced in the reaction in the two systems. While the water present in the GFP model only forms one hydrogen bond, the HAL model structure automatically leads to formation of two hydrogen bonds (see Figure 4c), which stabilizes the imidazolone ring formed in the HAL model system relative to the imidazolone ring formed in the GFP model. These hydrogen bonds are additional hydrogen bonds compared to the reactant. Several attempts were made to obtain convergence to a structure for HAL with only one hydrogen bond to water, but these were all unsuccessful. Since the hydroxyl group of the tyrosine in GFP is further away from the ring, there was no such problem for GFP. In the real enzyme, it appears more likely that only one additional hydrogen bond is actually formed, since the serine hydroxyl group most likely already has a hydrogen bond to a solvent water for the reactant, indicating that the reaction endothermicity for the reduced mechanism of HAL should in reality be 8−10 kcal/mol rather than 4 kcal/mol as obtained in the calculations. These minor uncertainties in the HAL reaction energy have no consequence for the present conclusions since the intermediate dihydroimidazolone is already so high in energy. With any reasonable barrier in the first step for the formation of the dihydroimidazolone (several estimates were obtained in the GFP study), the reduced mechanism must also be considered energetically very unlikely for HAL.

In Table 2, the calculated relative energies for the oxidized mechanism where the oxidation precedes the cyclization are given. For this mechanism, the results for the two enzymes GFP and HAL are even more similar. The reaction step leading to the dihydroimidazolone shown in Figure 5b is slightly less endothermic for HAL with a value of 4.5 kcal/mol compared to the value of 6.7 kcal/mol obtained for GFP. The oxidized mechanism leads to exothermic reactions both for the GFP and the HAL models. Since for the oxidized mechanism, in contrast to the reduced mechanism, the number of hydrogen bonds formed to water is the same, the exothermicities are also very similar, with values of −2.1 kcal/mol for HAL and −1.9 kcal/mol for GFP.

To summarize, the energy profiles for HAL and GFP are found to be very similar for two different pathways investigated. While the pathway termed the reduced mechanism is not supported by the calculations for GFP or for HAL, the alternative oxidized mechanism where a dehydration occurs prior to the formation of the imidazolone ring yields reasonable energetics for both enzyme systems. The small energetic difference found for the product of the reduced mechanism between the HAL and GFP models is mainly due to different types of hydrogen bonding between the water produced in the reaction and the model cluster and should to a major extent be regarded as due to artifacts of the models rather than as due to chemical differences between the actual enzymes.

Preorganization. Given the above-mentioned results of our DFT calculations, we examined the structural effects of dehydroalanine on the precyclization geometry of HAL using conformational searches, molecular dynamics simulations, and database analyses.

It has been shown that immature, or precyclized, GFP is preorganized for an autocatalytic cyclization in a tight turn conformation that results in very short intramolecular distances, between the carbonyl carbon of Ser65 and the amide nitrogen of Gly67.[2] The tight turn conformation holds the carbonyl carbon and amide nitrogen involved in the cyclization at a distance of 2.87 Å, which is shorter than the corresponding distance in 98.62% of the proteins that contain the FSYGVQ hexapeptide fragment.[2] This unusually short interatomic distance allows the nucleophilic attack to occur rapidly without the presence of a catalyst. Is a similar preorganization present in HAL?

Do the Residues Adjacent to the Cyclization Site Preorganize HAL? Because the active site of histidine ammonia-lyase is located in a VGASGD sequence, a PDB search was performed to locate proteins containing VGASG, GASGD, GASG, or ASGD sequences. Each protein was only counted once, even if there were numerous structures and mutants.

Table 3. Results from the pdb Search for Proteins with Sequences Similar to the Cyclization Site of HAL (VGASGD)[a]

sequence	no. of proteins	average distance (Å)	minimum distance (Å)
VGASG	2	4.07	4.04
GASGD	2	3.44	3.56
ASGD	14	3.92	3.11
GASG	27	3.90	3.04

[a] The *i* carbonyl carbon to *i* + 2 amide nitrogen distances are listed and shown in Figure 6.

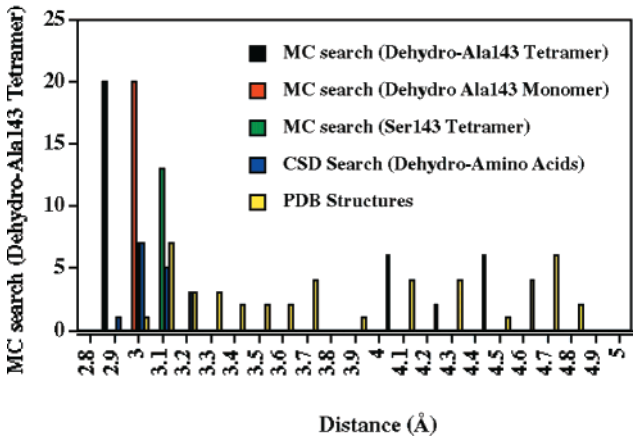

Figure 6. Histogram containing the amide nitrogen-to-carbonyl carbon distances for the 20 lowest energy conformations found in the Monte Carlo searches performed on the Δ-143Ala HAL monomer and tetramer and the Ser143 Hal tetramer, as well as the distances that were located in the PDB searches listed in Table 3 and the distances found in the CSD search for the substructures shown Figure 7A and listed in Table 4.

However, if the protein was composed of more than one strand that contained the signature sequence, the distance was measured and counted for each strand. The distances between the glycine amide nitrogen and the alanine carbonyl carbon that were obtained from this searches are listed in Table 3 and shown in Figure 6. This distance is a good measure of the extent of precyclic preorganization. The average distances are all significantly longer than the sum of the van der Waals radii of nitrogen and carbon (3.25 Å).[32]

A conformational search of the VGASGD hexapeptide fragment confirmed these observations. Although there are some low-energy tight turn conformations available to the VGASGD fragment, numerous low-energy extended conformations were also found. We conclude that the residues adjacent to the cyclization site do not preorganize HAL for cyclization.

Do Dehydro Amino Acids Lead to the Formation of Tight Turns? Dehydro (Δ) residues are found in peptides from microbial sources where they result in well-defined structural motifs[33] and in increased resistance to degradation.[34] Conformational studies have shown that ΔPhe strongly favors the formation of β-turns in short peptides and 3_{10}-helical structures if more than one ΔPhe is present.[33] A Cambridge Structural

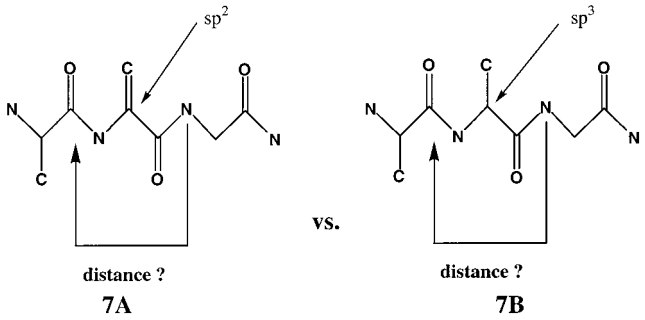

Figure 7. Substructures used in the Cambridge Structural Database search. The structures found with fragment A are listed in Table 4.

Table 4. All Dehydro Amino Acids Found in the Cambridge Structural Database with the Substructure Shown in Figure 7A and the Distance between the *i*th Carbonyl Carbon and *i* + 2th Amide Nitrogen for Each Structure

CSD reference code	N-to-C distance shown in Figure 7 (Å)
JIYHII[37]	3.044
JOWYUP[38]	3.176
JUCSIJ[39]	3.087
KUSHAH[40]	3.007; 3.067
PEVJOP[41]	2.988
VOZBAN[42]	3.123; 3.215
WEVJEM[43]	3.258; 3.295
YAZMOB[44]	3.039; 3.110
ZEYTUS[45]	3.116
ZOYJUS[46]	3.013
ZUBWUO[47]	3.078
ZUNJIB[48]	3.175

Database search for the substructure shown in Figure 7A found 10 structures with an average distance of 3.1 Å from the carbonyl carbon of residue *i* to the amide nitrogen of residue *i* + 2 (Table 4). This is 0.1 Å shorter than the sum of the van der Waals radii.[32] An identical search for the substructure shown in Figure 7B had 432 hits with an average distance of 3.44 Å, which is significantly longer than that found for the dehydro structures. The decrease in the distance is in part due an increase in the conjugation length and its associated increase in planarity. These searches show that the oxidative mechanism in which the formation of the ΔAla precedes the cyclization is structurally advantageous as it results in shorter carbonyl carbon of residue *i* to the amide nitrogen of residue *i* + 2 distances and therefore preorganizes the protein for cyclization.

Effect of the Protein Environment of HAL on the Precyclization Geometry. Monte Carlo conformational searches were performed on the precyclized ΔAla (bottom left Figure 3) in the HAL monomer as described in the Experimental Section. While conformational searches of the VGASGD and VGΔASGD hexapeptide fragments found many conformations

(32) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.
(33) Jain, R. M.; Rajashankar, K. R.; Ramakumar, S.; Chauhan, V. S. *J. Am. Chem. Soc*. **1997**, *119*, 3205.
(34) English, M. L.; Stammer, C. H. *Biochem. Biophys. Res. Commun*. **1978**, *83*, 1464.
(35) Ciajolo, M. R.; Tuzi, A.;.Pratesi, C. R.; Fissi, A.; Pieroni, O. *Biopolymers* **1990**, *30*, 911,
(36) Piazzesi, A. M.; Bardi, R.; Crisma, M.; Bonora, G. M.; Toniolo, C.; Chauhan, V. S.; Kaur, P.; Uma, K.; Balaram P. *Gazz. Chim. Ital.* **1991**, *121*, 1,
(37) Chauhan, V. S.; Bhandary K. K. *Int. J. Pept. Protein Res.* **1992**, *39*, 223.
(38) Rajashankar, K. R.; Ramakumar, S.; Chauhan, V. S. *J. Am. Chem. Soc.* **1992**, *114*, 9225.
(39) Padmanabhan, B.; Singh, T. P. *Biopolymers* **1993**, 33, 613.
(40) Busetti, V.; Crisma, M.; Toniolo, C.; Salvadori, S.; Balboni G. *Int. J. Biol. Macromol*. **1992**, *14*, 23.
(41) Rajashankar, K. R.; Ramakumar, S.; Mal, T. K.; Chauhan V. S. *Angew. Chem., Int. Ed. Engl*. **1994**, *33*, 970.
(42) Tuzi, A.; Ciajolo M. R.; Guarino, G.; Temussi, P. A.; Fissi, A.; Pieroni, O. *Biopolymers* **1993**, *33*, 1111.
(43) Rajashankar, K. R.; Ramakumar, S.; Jain, R. M.; Chauhan V. S. *J. Am. Chem. Soc*. **1995**, *117*, 10129.
(44) Rajashankar, K. R.; Ramakumar, S.; Jain, R. M.; Chauhan V. S. *J. Biomol. Struct. Dyn.* **1996**, *13*, 641.
(45) Rajashankar, K. R.; Ramakumar, S.; Jain, R. M.; Chauhan V. S. *J. Am. Chem. Soc.* **1995**, *117*, 11773.
(46) Rajashankar, K. R.; Ramakumar, S.; Mal, T. K.; Jain, R. M.;.Chauhan, V. S. *Angew. Chem., Int. Ed. Engl*. **1996**, 35, 765.

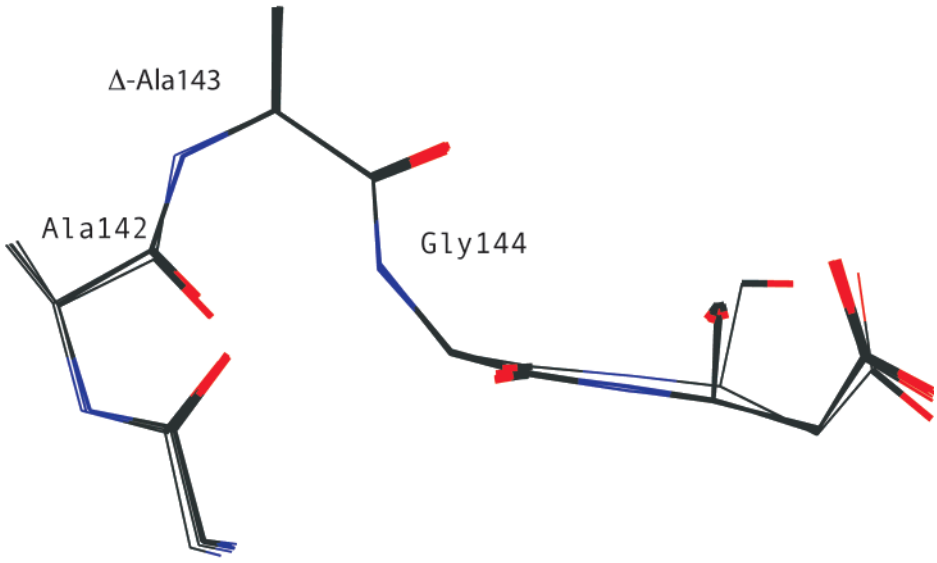

Figure 8. Superimposition of the cyclization site in the 20 lowest energy structures that were located by the Monte Carlo conformational search performed on the ΔAla143 HAL tetramer.

ranging from tight turn structures to extended conformations, the conformational search of the same fragment in HAL found that there was little variation in the Gly144 nitrogen to Ala142 carbonyl carbon distances (see Figure 6).

In its native state, histidase is found as a tetramer composed of four identical peptide chains rather than as a monomer. The tetrameric form is significant because it places the cyclization site inside the protein and close to three different chains (Figure 2). As a result, the environment surrounding the area of the ring formation could be different than that found in the monomer, where residues 142−144 are found near the surface of the protein. To have a more realistic environment and to determine the effect of the tetramer, conformational searches of the cyclization site in the tetramer were conducted as described in the Experimental Section. Figure 6 shows that as expected there was very little variation in the distance between the amide nitrogen of Gly144 and the carbonyl carbon of Ala142 that it attacks in the cyclization process. None of the conformations obtained in these calculations showed distances of greater than 2.885 Å; this is slightly shorter than the distances found for the monomer and is 0.45 Å less than the sum of the van der Waals radii of nitrogen and carbon. Interatomic distances of less than 2.885 Å are also significantly shorter than the corresponding distances that were located in the PDB search of proteins containing the same residue sequence as that present at the histidase cyclization site, none of which were less than 3.00 Å (Figure 6 and Table 3). They are also significantly shorter than the average distance between any *i* carbonyl carbon and *i* + 2 amide nitrogen in the backbone structure of a protein. In fact, an examination of 50 representative proteins in the PDB showed that only 0.89% of such distances were less than 2.90 Å.[2]

Not only were all the amide nitrogen-to-carbonyl carbon distances very short, but the conformations adopted by the six residues surrounding the site of MIO formation were also very similar in all of the low-energy conformations (Figure 8).

The results from the conformational searches were confirmed by molecular dynamics simulations, which showed that residues involved in the cyclization were held in a tight turn conformation.

The unusually short atomic distances seen for the HAL tetramer, and the lack of variation in these distances, indicates that, like GFP, the tetrameric form of the enzyme is rigidly preorganized for cyclization. The data also suggest that preorganization may be present in the monomer of histidase but that it is less significant than that found in the tetramer.

Conclusion

Autocatalytic posttranslational intrachain modifications are extremely rare. We have shown that there are many similarities in the cyclization observed in GFP and HAL; in both cases, we suggest that a dehydro amino acid is formed prior to cyclization and the protein forms a tight turn that is rigidly preorganized for cyclization.

The calculated energetics of the cyclization observed in HAL are not in agreement with the currently accepted mechanism in which cyclization precedes dehydrogenation. The "oxidized" mechanism which we first proposed for GFP[7] is also energetically more favorable in HAL. Database searches showed that the oxidative mechanism in which the formation of the dehydro amino acids in residue *i* + 1 precedes the cyclization is also structurally advantageous as it results in shorter distances between the carbonyl carbon of residue *i* and the amide nitrogen of residue *i* + 2 and, therefore, preorganizes the protein for cyclization. The unusually short distances between the carbonyl carbon of Ser142 and the amide nitrogen of Gly144 seen for the HAL tetramer, and the lack of variation in these distances, indicates that, like GFP, the tetrameric form of the enzyme is rigidly preorganized for cyclization. The monomeric form of HAL is less tightly preorganized than the tetrameric HAL. Dehydro amino acids aid in preorganization but the main driving force in the rigid tight turn formation is the influence of the surrounding protein.

The mechanisms have been differentiated solely on the basis of thermodynamics. Both mechanisms are required to be exothermic; however, the reduced one was found to be quite endothermic and was therefore discounted. The calculation of the transition states is not trivial. Since additional residues must be involved in transferring the required protons, substantially larger models are required for the transition states. In our previous paper on GFP,[7] transition states were determined using models of ~40 atoms (more than 25 heavy atoms). We felt that these calculations failed to provide additional information on

the preferred mechanism in the case of GFP, and therefore, we did not do transition-state calculations for HAL.

Acknowledgment. M.Z. is a Henry Dreyfus Teacher−Scholar and thanks the NIH (Grant GM59108-01) and the Research Corp. for financial support. The work of M.W. was supported by STINT (The Swedish Foundation for International Cooperation in Research and Higher Education).

JA004009C