Stepwise Simulation of 3,5-Dihydro-5-methylidene-4H-imidazol-4-one (MIO) Biogenesis in Histidine Ammonia-lyase

Sánchez-Murcia, Pedro A.; Bueren-Calabuig, Juan A.; Camacho-Artacho, Marta; Cortés-Cabrera, Álvaro; Gago, Federico

Published in:
Biochemistry

DOI:
10.1021/acs.biochem.6b00744

Publication date:
2016

Document Version
Peer reviewed version

Stepwise simulation of 4-methylidene-imidazole-5-one (MIO) biogenesis in histidine ammonia lyase

*Pedro A. Sánchez-Murcia[1], Juan A. Bueren-Calabuig[1†], Marta Camacho-Artacho[2], Álvaro Cortés-Cabrera[1], Federico Gago[1]**

[1]*Área de Farmacología, Departamento de Ciencias Biomédicas, Unidad Asociada al IQM-CSIC, Universidad de Alcalá, E-28805 Alcalá de Henares, Spain. E-mail: federico.gago@uah.es*

[2] *Structural Biology Department, Centro Nacional de Investigaciones Oncológicas (C.N.I.O.), E-28029 Madrid, Spain*

KEYWORDS. Histidine ammonia-lyase · MIO · enzymatic mechanism · post-translational modification · chemical pathway

ABSTRACT

A 4-methylidene-imidazole-5-one (MIO) electrophilic moiety is post-translationally and autocatalytically generated in homotetrameric histidine ammonia-lyase (HAL) and other enzymes containing the tripeptide Ala-Ser-Gly in a suitably positioned loop. The backbone cyclization step is identical to that taking place during fluorophore formation in green fluorescent protein (GFP) from the tripeptide Ser-Tyr-Gly but dehydration, rather than dehydrogenation by molecular oxygen, is the reaction that gives rise to the mature MIO ring system. To gain additional knowledge about this unique process and shed light on some still unresolved issues we have made use of extensive molecular dynamics (MD) simulations and hybrid quantum mechanics/molecular mechanics (QM/MM) calculations implementing the self-consistent-charge density-functional tight-binding (SCC-DFTB) method on a fully solvated tetramer of *Pseudomonas putida* HAL. Our results strongly support that mechanical compression of the reacting loop by neighboring protein residues in the precursor state is absolutely required to prevent formation of inhibitory main-chain hydrogen bonds and to enforce proper alignment of donor and acceptor orbitals for bond creation. The consideration of the protein environment in our computations shows that water molecules, which have been mostly neglected in previous theoretical work, play a highly relevant role in the reaction mechanism, and more importantly, that backbone cyclization resulting from the nucleophilic attack of the Gly amide lone pair to the π^* orbital of the Ala carbonyl precedes side-chain dehydration of the central serine.

The lack of electrophiles in the side chains of proteinogenic amino acids would limit the feasibility of carrying out some chemical reactions in the active site of many enzymes were it not for the presence of bound coenzymes (e.g. pyridoxal phosphate) or metal ions such as Mg^{2+} and Zn^{2+}. Remarkably, in a few of these natural catalysts, an alternative solution has evolved that consists of creating an endogenous electrophile 'in situ' out of a small portion of the protein backbone and a serine side chain. Thus, the 4-methylidene-imidazole-5-one (MIO) heterocycle[1, 2] is post-translationally and autocatalytically generated from the tripeptide Ala-Ser-Gly (Scheme 1) located in a suitably positioned loop in homotetrameric histidine ammonia-lyases (HAL)[3] and phenylalanine ammonia-lyases (PAL).[4] These enzymes catalyze the non-oxidative deamination of L-His and L-Phe by promoting the abstraction of the non-acidic β-protons of these two α-amino acids to yield *trans*-urocanic and *trans*-cinnamic acids, respectively.[5] A similar post-translational autocatalytic intrachain cyclization event, which needs oxygen for its completion, leads to the 4-(*p*-hydroxybenzylidene)-imidazolidin-5-one fluorophore naturally present in green fluorescent protein (GFP)[6] and its variants.

Since catalytically proficient HAL is also formed when the protein is synthesized *in vitro*, the information for MIO biosynthesis must be encoded in the amino acid sequence itself and depend on the proper folding and local environment of a short stretch of the polypeptide chain. Elucidation of the nature of the electrophilic moiety in HAL took almost half a century. Until 1999, when Schulz and colleagues solved the crystal structure of mature HAL from *Pseudomonas putida* (Pp-HAL),[7] it was thought that a dehydroalanine (ΔAla) was present in the active site of both HAL[3, 8] and PAL[9] because treatment with the reducing agent $NaBH_4$ resulted in enzyme inactivation and [^{3}H]-Ala was the only labeled material found in acid hydrolysates of borotritide-reduced HAL.[10]

Once the presence and structure of MIO were demonstrated, several laboratories tried to gain insight into both the biogenesis of this novel prosthetic group and the details of the catalytic mechanism of these enzymes. Knowledge about the former unique process was obtained over the years by (i) biochemical, biophysical and X-ray crystallographic analyses of several wild-type[7] and site-point variants of HAL[11] and PAL enzymes[12], (ii) 'grafting' of the MIO-forming tripeptide into the GFP sequence by means of structure-guided protein engineering,[13] and (iii) quantum mechanical (QM) calculations using reduced model systems.[14]

In their seminal report on the crystal structure of mature Pp-HAL, Schulz *et al*.[7] postulated a mechanism (Scheme 1, A) that was analogous to that proposed for fluorophore biosynthesis in *Aequorea victoria* GFP (Av-GFP) by Tsien and co-workers a few years earlier.[15] In both proteins, the imidazolidin-5-one heterocycle would result from the attack of the backbone nitrogen of a glycine (Gly144 in Pp-HAL and Gly67 in Av-GFP) onto the carbonyl carbon of the amino acid two positions before in the primary sequence (Ala142 in Pp-HAL and Ser65 in Av-GFP). Thus, an initial cyclization step ($\mathbf{I_A}$), followed by the elimination of a water molecule, would initially lead to the imidazolidin-5-one ring system ($\mathbf{II_A}$). Then the new N=C double bond would promote the dehydration of the Ser143 side chain to give rise to the formation of the conjugated double bond at position 4 in MIO. Two years later, however, Zimmer and colleagues[14] published that this mechanism was not supported by their QM calculations on a reduced model system, which consisted of the Ala-Ser-Gly tripeptide and a crystallographic water molecule, using density functional theory (B3LYP) and a 6-311+G(2d,2p) basis set. In analogy to what they had proposed earlier for the autocatalytic cyclization of GFP,[16] they advanced an alternative mechanism (Scheme 1, B) consisting of the initial exocyclic dehydration of Ser to ΔAla ($\mathbf{I_B}$), the subsequent cyclization of which would give rise to $\mathbf{II_B}$. In their view, the

nucleophilic attack in this second step would be favored over that initially originating $\mathbf{I_A}$ because the reacting groups would be in closer proximity.

[Scheme 1]

The crystal structures[11] and active site characterizations[17] of several HAL variants have had a major impact on our understanding of the role played by certain residues in MIO biogenesis and catalysis (Table S1). In summary, Phe329Gly and Asp145Ala Pp-HAL variants (cf. Figure 1) have been shown to be defective in MIO formation and to lack any measurable enzymatic activity whereas others (e.g. Phe329Ala and Tyr280Phe) do contain MIO but are virtually inactive. In addition, some other HAL variants resulting from site-directed mutagenesis lose >99.9% of activity (e.g. Asn195Ala) but evidence for the presence of MIO remains lacking. Complementarily, in another series of ingenious experiments, Getzoff and colleagues engineered GFP to contain the Ala-Ser-Gly tripeptide of HAL in place of the native Ser-Tyr-Gly sequence with a view to converting the naturally occurring fluorophore into a MIO moiety.[13] Indeed, the crystallographic analysis of the so-called Ser65Ala and Tyr66Ser GFPhal variants revealed a planar five-membered MIO. It was then hypothesized that the GFP and HAL scaffolds promote backbone cyclization by aligning in close proximity the Gly(i) amide lone pair and the π orbital of the residue *i–2* carbonyl and preventing the involvement of these main chain groups in intramolecular hydrogen bonding interactions in the precursor state.

Therefore, available evidence strongly indicates that (i) some steric strain on the flexible loop containing the reactive tripeptide and a precise pre-organization of the surrounding residues are required for MIO formation, (ii) the carboxylate of Asp145 is essential for cyclization, and (iii) the nature of the residue at position 329 is important for both MIO formation and catalysis.

Nonetheless, the role of Asn195 has passed largely unnoticed and earlier calculations using reduced model systems may have been suboptimal for providing a comprehensive theoretical rationale for the preferred reaction pathway. Moreover, despite all of the aforementioned results, the sequential order of the reactions and the structural factors responsible for the enhanced reactivity of the attacking Gly's amide nitrogen have not been completely clarified. For these reasons, we decided to revisit the issue of MIO biogenesis in HAL taking into account the impact on the reaction mechanism of both the protein environment and the aqueous medium. To this end, we made use of extensive molecular dynamics (MD) simulations in explicit solvent of the whole Pp-HAL tetramer during the course of which hybrid quantum mechanics/molecular mechanics (QM/MM) calculations in the active site were performed to drive bond formation and breakage employing the Density Functional-based Tight Binding (DFTB) method.[18] The probability distributions along the coordinate of the cyclization reaction in the two alternate mechanisms (A and B in Scheme 1) were determined using umbrella sampling and pieced together by means of the two-dimensional variational free energy profile (2D-vFEP) method.[19, 20] This procedure allowed us to characterize the geometrical features of all possible intermediates and to propose a definite sequence of events for MIO biogenesis in HAL that is supported by the calculated energy profiles using higher level QM methods and found to be in agreement with available experimental data.

METHODS

Model systems and molecular dynamics simulations. The initial Cartesian coordinates of the wild-type tetramer were taken from the crystal structure of the Phe329Gly HAL variant (PDB id. 1GK2) upon replacement of Gly329 by Phe using PyMOL.[11] In addition, the Phe329Gly HAL

tetramer itself and another variant containing Ala at position 329 were modeled and simulated for comparative purposes. The protonation states of all titratable residues (His, Glu, Asp, Lys and Arg) were assigned on the basis of pK_a calculations at a pH of 7.0 using the H++ program.[21] In each of the four active sites the crystallographically found sulfate was replaced with a more physiologically relevant phosphate anion for which atom-centered point charges were calculated using the restrained electrostatic potential (RESP) method as implemented in the RED server.[22] Each macromolecular system was immersed in a box of TIP3P water molecules[23] that extended 15 Å away from any solute atom and 22 Na^+ ions were placed at random locations around it to ensure electrical neutrality. The cutoff distance for the nonbonded interactions was 10 Å and periodic boundary conditions were used. Electrostatic interactions were treated by using the smooth particle mesh Ewald (PME) method with a grid spacing of 1 Å. The SHAKE algorithm was applied to all bonds involving hydrogen atoms and an integration step of 2.0 fs was employed throughout. The molecular dynamics (MD) simulation protocol made use of the *pmemd_cuda.SPFP* module in the AMBER12 suite of programs.[24] Firstly, solvent molecules and counterions were relaxed by energy minimization and allowed to redistribute around the positionally restrained solute atoms (25 Kcal mol^{-1} $Å^{-2}$) at constant temperature and pressure (1 atm), essentially as described previously.[25] These harmonic restraints were gradually reduced until they were completely removed. The resulting system was then heated from 100 to 300 K during 20 ps, equilibrated at 300 K for 100 ps and further simulated under the same conditions up to a total time of 60 ns, during which system coordinates were collected every 20 ps for further analysis. All the MD simulations were performed using one (Phe329Ala) or four nodes (wild-type and Phe329Gly variant) of NVIDIA GPUs.

Hybrid quantum mechanics / molecular mechanics (QM/MM) calculations. All the QM/MM calculations were performed within the *sander* module of AMBER12, which also includes a complete treatment of long-range electrostatic interactions. The QM region encompassed the active site region where bonds are broken and formed whereas the MM region included all the remaining protein and solvent atoms as well as the counterions. Care was taken not to cut any polar bonds when defining the QM/MM boundary. Thus, the QM region contained (i) all the atoms of residues Ala142, Ser143 and Gly144 capped at the ends with "linking atoms" from residues 141 and 145, (ii) the side chains (from Cβ) of Asn195, Tyr280' and Glu414 (from Cγ), and (iii) water molecules WAT_A and WAT_B. The Density Functional-based Tight Binding (DFTB) method was employed. A 30-ps unrestrained MD equilibration was first performed to relax the system while treating the QM part with the DFTB method (geometry $\mathbf{SP_1}$). Each reaction coordinate specified throughout the text took place over a 10-ps period and was driven by a harmonic force constant of 750 kcal mol^{-1} $Å^{-2}$: (1) for the *cyclization reaction*, the N(Gly144)-C(Ala142) (d_1) and H(Gly144)-O(Ala142) (d_2) distances were gradually shortened; (2) the *endocyclic dehydration* was achieved by defining a collective reaction coordinate consisting of the simultaneous elongation of the Ala142 C=O bond and the shortening of the H(Ala142)-OG(Ser143) distance; and (3) to enforce the Tyr280'-catalyzed *exocyclic dehydration* three distances were defined as reaction coordinate: the shortening of the distances HH(Tyr280')-OH(Ser143) and OH(Tyr280')-HA(Ser143) and the enlargement of Cβ-OG(Ser143) bond. The non-catalyzed exocyclic dehydration was explored in two steps: generation of the enol tautomer with a reaction coordinate consisting in the shortening of the distance HA(Ser143)-O(Gly144), and a second step where the distances HA(Ser143)-OH(Ser143) and the bond Cβ-OG(Ser143) bond were shortened or enlarged, respectively. In all

cases, the starting and final product geometries for each QM/MM reaction were optimized following a hybrid QM/MM protocol of 5000 cycles of steepest descendent followed by 5000 steps of conjugate gradient energy minimization.

Umbrella sampling and two-dimensional variational free energy profile (2D-vFEP) method.

The probability distributions along the first reaction coordinate were determined using umbrella sampling and pieced together by means of 2D-vFEP method.[19] The cyclization reactions for proposals A and B were explored by defining windows of 0.10 Å along the two independent reaction coordinates that specified the shortening of N(AlaX)-SG(CysY) (d_1) and H(AlaX)-OH(CysY) (d_2) distances. For each window the system was left to oscillate for 20 ps around the restrained distance. In the proximity of the transition state (TS), window separation was reduced to 0.05 Å. The free energy 2D surfaces were calculated using the 2D-vFEP method[19] after taking the last 15ps of each windows and plotted by means of GNUPLOT 4.6.[27]

Quantum mechanical calculations

QM geometry optimization at a higher level of theory was achieved using program Gaussian09[28] and the B3LYP/6-31G* method on the geometries obtained from the QM/MM simulations. The molecular systems studied were the same that comprised the QM region in the QM/MM studies. Carbon atoms were restrained to ensure that their positions were the same as found within the protein tetramer whereas heteroatoms as well as hydrogen atoms were allowed to move freely. Vibrational frequencies were computed on each stationary point for identification as either a TS (one imaginary frequency corresponding to the intramolecular bond to be created) or an energy minimum (all frequencies positive) on the potential energy surface.

RESULTS

Pre-reactive conformation and structural aspects

Standard MD simulations on a solvated HAL monomer proved to be inadequate for capturing the intrinsic dynamics of this protein. Moreover, test calculations involving a dimer to gain additional structural support showed that this enlarged system was also insufficient to achieve a suitable near-attack conformation in the Ala-Ser-Gly loop that might lead to cyclization (data not shown). In the end, we found it necessary not only to use the tetramer as the biologically relevant macromolecule but also to introduce a phosphate ion in the active site of the enzyme as a more physiological mimetic of the sulfate that is present in most X-ray crystal structures as a surrogate of the carboxylate moiety of the substrate. By simulating the tetramer under these conditions we improved the sampling as well because the evolution of the reactive loop could be studied concomitantly in each of the four monomeric subunits (Figure 1A and Figure S1).

Tetramer stabilization, as assessed in terms of both root-mean-square deviation (RMSd) from the initial structure and radius of gyration, was achieved after ~10 ns of simulation (Figure S2). As regards the MIO-forming loop, a well-defined network of favorable hydrogen-bonding interactions was observed (principally in site 4, made up by subunit D and Tyr280' from subunit A) that allowed the Asp-Ser-Gly tripeptide to adopt a potentially reactive conformation (Figures 1A and S1). Thus, the two complementary intramolecular hydrogen bonds between the amide functionalities of residues 142 and 144 help bring into close proximity the nucleophilic (amidic nitrogen of Gly144) and electrophilic (carbonyl carbon of Ala142) atoms in each of the four active sites (average distance ~3 Å, Figure S1A). An additional interaction that helps activate the reacting amides is the hydrogen bond established between the carbonyl oxygen of Ser143 and the

side-chain carboxamide of Asn195. On the other hand, the binding to O(Ala142) and long residence time of a water molecule (WATA) contributes to increasing the electrophilicity of the carbonyl carbon of Ala142. The orientation of WATA, which was firstly proposed to act as a potential acid/base catalyst for cyclization by Schultz et al. ('Wat197') 11, is mostly fixed by the Glu414 side-chain carboxylate, which contributes to its polarization, and some surrounding water molecules, in particular WATB (Figure 1A), which is strongly hydrogen bonded to O(Gly144). Complementarily, the hydroxyl group on the side chain of Ser143 interacts with the carboxylate moiety of Glu414 and occasionally with the hydroxyl group of Tyr280' from the neighboring subunit (Figures S1C and S1D). This hydrogen-bonding network promotes the subsequent exocyclic dehydration of the loop, as suggested previously and further discussed below.[11]

To evaluate the effect that the presence of the different substituents on the side chain of the residue at position 329 may have on the conformational arrangement of the reactive loop, we performed two additional MD simulations, under identical conditions as before, for the Phe329Gly and Phe329Ala HAL variants (Figure S1). As observed for the wild-type enzyme, the distance between the two reactive atoms for the cyclization step remained consistently short in both proteins over the whole MD trajectories (Figure 1B, light bars). However, significant differences were found in the mean distance between H(Gly144) and the competing carbonyl oxygen of residue 329 (Figure 1B, dark bars). Thus, whereas in the proteins that contain a Cβ at this position (wild-type and Phe329Ala) this distance is always longer than 3.0 Å in the four centers, in the Phe329Gly variant it is much shorter and similar to that between the two bond-forming atoms H(Gly144) and O(Ala142). Thereby, placement of just a methyl group on this glycine (Phe329Ala variant) pushes the reacting NH of Gly144 away from the carbonyl oxygen

of residue 329, avoiding the formation of the competing hydrogen bond with the latter atom. This result is consistent with the experimental finding that MIO is formed in Pp-HALF329A but not in Pp-HALF329G.

[Figure 1]

QM/MM study of the mechanism of MIO biogenesis

The two main mechanisms depicted in Scheme 1 were studied, step by step, using the hybrid QM/MM method during the course of the MD simulation of the wild-type tetramer. To this end, a QM region was defined containing the atoms displayed in Figure 1A, namely all those belonging to the Ala142-Ser143-Gly144 loop plus those from the side chains of Asn195, Glu414 and Tyr280', which are directly involved in hydrogen-bonding interactions with the MIO-forming tripeptide, as well as the two conserved water molecules named above (WAT_A and WAT_B).

Nucleophilic attack: effect of a previous exocyclic dehydration

The cyclization event is the first step of proposal A. The attack of the Gly144 amide nitrogen onto the Ala142 carbonyl leads from the near-attack conformation of the reactive loop ($\mathbf{SP_1}$) to the cyclic intermediate $\mathbf{I_A}$. The corresponding 2D free energy map obtained by means of umbrella sampling is shown in Figure 2.

[Figure 2]

According to the results shown in Figure 2, this cyclization process is endergonic and the proton transfer step is an early event in the nucleophilic attack (Figure 3). As the Gly144 nitrogen in the initial acyclic configuration approaches the Ala142 amide carbon (starting point $\mathbf{SP_1}$), it transfers its bonded hydrogen to O(Ala142) in a process that is assisted by WAT_B, as shown at point $\mathbf{TS_{SP1\square IA}}$. This proton transfer event is facilitated by the neighboring O(Gly144).

This scenario is similar to that reported in the non-enzymatic deamidation reactions of Asn and Gln residues in proteins where the nucleophilic attack of the nitrogen atom onto the carbonyl carbon of the Asn/Gln side chain gives rise to a cyclic tetrahedral intermediate.[29, 30]

Regarding the topicity of this nucleophilic addition, the previous N □ O proton transfer induces a change in the angle of attack of N(Gly144) (Figure 3). For the approach of a nucleophile (Nu) to a carbonyl, whose geometry is planar trigonal, it has been demonstrated that the Nu 'prefers' a Nu-C-O angle, or Bürgi-Dunitz angle (BDA)[31], of 107°, a value that is close to the classical tetrahedral geometry in the addition product (109º). This angle measures the "offset" of the Nu's approach to the electrophile. In our case, the previous proton transfer of the amide hydrogen induces a change in the BDA from ~87º (**SP$_1$**, Figure 3) to a canonical value of ~107º after proton transfer. As a consequence, empty and filled orbitals are optimally aligned for formation of the N-C bond (Figure 3). This narrow value observed in **SP$_1$** is due to the compression exerted by the environment on the three residues located in the reacting loop by means of (i) the "pulling" attraction of the Asn195 carboxamide on O(Gly144), (ii) the bridging of O(Gly144) and O(Ala142) by the conserved WAT$_B$, and (iii) the "pushing" repulsion on the loop by the phenyl ring of Phe329, which promotes a suitable geometry for proton transfer prior to the nucleophile attack.

[Figure 3]

The energy minimized **I$_A$** intermediate shows a canonical tetrahedral geometry at the carbonyl carbon of Ala142 (Figure 2). The N(Gly144)–C(Ala142) distance is 1.46 Å and the length of the H–O(Ala142) bond is 0.99 Å. In addition, the hydrogen bonds between the Ser143 hydroxyl group and the side chains of Glu414 and Tyr280' are maintained, indicating that these two amino acids may assist or are involved in the subsequent exocyclic dehydration reaction.

Proposal B (Scheme 1), in contrast, entails an initial exocyclic dehydration (Scheme 1 and Table 1). In the starting geometry $\mathbf{SP_1}$ the side chain of Ser143 is activated by two hydrogen bonds with the side chains of both Glu414 and Tyr280'. This helps to orientate and stabilize the Ser143 side-chain hydroxyl group as a formal hydroxide anion. Accordingly, we planned the exocyclic elimination step as a one-step acid/base Tyr-catalyzed process, in which the phenol group of Tyr280' protonates the OH(Ser143), with the ensuing abstraction of Hα(Ser143) by the phenolate (Scheme 2). In addition, we also probed the alternative intramolecular Tyr-unassisted process in two steps through the keto-enol tautomerism of $\mathbf{SP_1}$. The QM-computed activation energy (Table 1) shows that the assistance of Tyr280' reduces by ~59% the energy barrier that has to be overcome to yield $\mathbf{I_B}$ (Table 1). For the unassisted alternative, it is noteworthy that intermediate $\mathbf{I_{B,ke}}$, whose formation is endergonic, can easily evolve to $\mathbf{I_B}$ by overcoming an energy barrier of less than 9 kcal mol^{-1} (Table 1). The proposed catalytic role of Tyr280' in MIO formation is at odds with the finding of MIO in the crystal structure of the Pp-HAL_{Y280F} variant. However, to the best of our knowledge, no experimental evidence of possible differences in the kinetics of this process is available.

[Scheme 2]

Next, we explored the cyclization reaction leading from $\mathbf{I_B}$ to $\mathbf{II_B}$ by using umbrella sampling (pathway B, Scheme 1). The starting QM/MM-optimized geometry of $\mathbf{I_B}$ shows that the distances (d_1,d_2) between the atoms involved in the cyclization step (H(Gly144)-O(Ala142) and N(Gly144)-C(Ala142)) are slightly shorter than those measured in $\mathbf{SP_1}$ (1.85, 3.06 and 1.88 and 2.97, respectively), in analogy with Zimmer *et al.*'s description of intermediate **5a** in their proposed 'oxidized mechanism'.[14] However, our data taking into account the protein environment clearly indicate that the pathway connecting intermediates $\mathbf{I_B}$ and $\mathbf{II_B}$ crosses an

activation energy barrier that is higher than that calculated for the cyclization step in proposal A ($\mathbf{TS_{SP1\square IA}}$: 37 kcal mol^{-1} *vs*. $\mathbf{TS_{IB\square IIB}}$: 49 kcal mol^{-1}, Figures 2 and 4).

[Figure 4]

Evolution of $\mathbf{I_A}$ to MIO

Upon formation of $\mathbf{I_A}$, two successive dehydration steps lead to **MIO** (Figure 5). The endocyclic water elimination proceeds through the removal of the Ala142 carbonyl oxygen and the Ser143 amide proton. Whereas in the initial $\mathbf{SP_1}$ reacting conformation both atoms are disposed *trans*-coplanar (as in most peptide bonds), it is noteworthy that both groups are arranged *syn*-periplanar in $\mathbf{I_A}$ (Figures 5 and S3). Indeed, this CO(Ala142)-NH(Ser143) geometry was found to be stable during the course of an unrestrained MD simulation in which this region was treated quantum mechanically in the absence of any distance or angle restraints (Figure S3). In addition, the energy difference between $\mathbf{I_A}$ and $\mathbf{II_A}$ was found to be less than 5 kcal mol^{-1} (Table 1), showing that the endocyclic dehydration step gives rise to a quasi-reversible potential of mean force (PMF) energy profile that is in consonance with previous experimental results reported by Getzoff and colleagues.[13] These authors found, by means of UV-vis spectroscopic measurements, that engineered GFPhal variants consist of a mixed population containing aromatic and nonaromatic MIO species, which are interchangeable throughout a quasi-reversible dehydration/hydration reaction.

The final exocyclic dehydration from $\mathbf{II_A}$ to afford **MIO** was modeled as before for pathway B taking into account both the catalyzed and the non-catalyzed processes (Figure 5). Again, the energies obtained show that a general acid-base catalytic role of Y280' may considerably reduce the activation barrier of the exocyclic reaction. The entirety of quantum mechanically computed enthalpies and activation energies for reactions in pathways A and B are summarized in Table 1.

In all cases, the rate-limiting step in the wild-type enzyme (i.e. the process with the higher activation energy) is the cyclization step. Consistently, no advantageous effects are observed on either $\Delta E^{\ddagger}$ on ΔH_{298} if the exocyclic dehydration precedes cyclization. In addition, it is remarkable that the relative orientation of the two crystallographic water molecules, WAT_A and WAT_B, has a deep effect on the computed energies. Therefore, it becomes clear that the protein environment of the reacting residues is critical to suitably position these water molecules in a configuration that makes the overall reaction possible.

[Figure 5]

[Table 1]

Discussion

Two possible mechanisms for the biogenesis of the electrophile MIO in HAL have been proposed on the basis of X-ray crystal structures of wild-type and mutant Pp-HAL (Table S1) and related lyases (e.g. PAL) as well as theoretical calculations. Intrigued by the factors promoting the extraordinary reactivity of an amide group that is able to attack another peptide bond in its proximity to generate MIO, we decided to revisit the proposed mutually exclusive mechanisms by considering not only a reduced representation of the active site, as done previously, but also the solvated protein environment. In this regard, recent progress in the implementation of hybrid QM/MM methods[32-35] has allowed us to address several biologically relevant questions concerning reactivity in macromolecules by taking into account the full protein environment during a chemical reaction.[36-39]

The present results strongly support the view that the protein environment in tetrameric HAL is crucial for exerting the required compression on the reactive Ala142-Ser143-Gly144 loop and bringing into close proximity residues 142 and 144 in a suitable orientation (Figure 1A). MD

simulations showed that the near-attack conformation for this loop is achieved only when the full tetramer in explicit water is employed and not when a single HAL monomer or a HAL dimer are considered. We were able to ascertain that these three residues are well positioned and activated, in terms of reactivity, by Tyr280', Glu414 and Asn195 side chains together with two water molecules, herein labeled as WAT_A and WAT_B (Figure 1A). Previous site-directed mutagenesis studies have shown that side-chain replacement at those critical positions (i.e. Tyr280'Phe, Glu414Ala and Asn195Ala) abolishes HAL catalytic activity (Table S1). Whereas the roles of the two former residues in MIO biogenesis have already been outlined in earlier publications, we now report that Asn195 may play an active part in the polarization of electron density on the Gly144 amide nitrogen by hydrogen bonding with O(Ser143) thus facilitating attack by Ala142. This would then be reminiscent of the role ascribed to Arg96 as an activator in the formation of the GFP chromophore through an interaction of its guanidinium group with the Tyr66 carbonyl oxygen that leads to the Gly67 α-enolate tautomer (i.e. the nucleophile in the cyclization step).[40] Thus, on the basis of our data, we propose that the lack of activity in the Asn195Ala variant is most likely due to its inability to form MIO. Regrettably, we have not been able to generate or find any experimental biophysical evidence supporting the absence or presence of MIO in this variant protein.

The experimentally demonstrated requirement for MIO formation in Pp-HAL of at least one methyl group in the side chain of amino acid 329 can be explained as a consequence of the need to favor a suitable orientation of the attacking NH(Gly144) with respect to the carbonyl CO(Ala142) (Figure 1B). In this respect, whereas in the simulation of the MIO-defective Phe329Gly variant a large fluctuation of the Ser143 ψ angle is observed (Figure S1), resulting in distances between H(Gly144) and O(Ala142) atoms that are longer than at the starting time

(Figure 1B), in the wild-type protein this distance remains unperturbed at ~2 Å. An intermediate situation is observed for the Phe329Ala variant. Thus, this difference in the relative arrangement of donor and acceptor orbitals for bond creation in the cyclization reaction appears to be responsible for the lack of loop reactivity and absence of MIO formation when a β-carbon is missing at position 329.

For MIO biogenesis our calculations strongly support the sequence of events depicted in proposal A (**SP_1** □ **I_A** □ **II_A** □ **MIO**) over that shown in alternative B. Accordingly, the amide nitrogen of Gly144 would first attack the carbonyl carbon of Ala142 in a process that requires (mechanical) energy to afford intermediate **I_A** (Figure 2). An early transfer of H(Gly144) to O(Ala142), mediated by WAT_B, would precede formation of the N-C bond that induces a change in the Bürgi-Dunitz angle of attack of N(Gly144) to an optimal value. This result is partially in consonance with the hypothesis put forward by Schulz *et al*. that a water molecule plays the role of an acid/base catalyst in the cyclization reaction[11] but these authors pointed to WAT_A ('WAT_{187}') instead. In this regard, it is noteworthy that the polypeptide environment is thought to help to deprotonate the amide in the non-enzymatic deamidation reactions of Asn and Gln residues in proteins[29, 30] where the nucleophilic attack of the nitrogen atom onto the carbonyl carbon of the Asn side chain gives rise to a cyclic tetrahedral intermediate. Moreover, it has been proposed, on the basis of QM calculations, that proton transfer between proximal carbonyl atoms in peptide models is mediated with the assistance of a bridging water.[41] Complementarily, although the biogenesis of the fluorophore in GFP requires oxygen as the oxidant, the current experimental evidence also points to the cyclization step as the first event.[40, 42]

Intermediate **I_A** would then evolve to MIO through two consecutive dehydration steps. In terms of energy and geometrical considerations, our results support the endocyclic dehydration

as the first elimination event through the elimination of H(Ser143) and O(Ala142) from the same face of the loop due to their *syn*-periplanar geometry (Figures 5 and S3). The subsequent exocyclic dehydration would be assisted by residues Tyr280' and Glu414 to afford MIO. In this regard, it has been experimentally shown that these two residues are essential for HAL catalysis: if Glu414 is replaced by Gln (the positionally equivalent residue in PAL) or by Ala, a dramatic loss in enzymatic activity is observed. Moreover, when a Phe residue is introduced at position 280' in place of Tyr, the enzymatic activity drops by two orders of magnitude. Nonetheless, the facts that MIO is found in the crystal structure of Tyr280Phe HAL and that Gln naturally replaces Glu in PAL lead us to conclude that none of these residues (Tyr280' and Glu414) is strictly required for MIO biogenesis but their absence may have an impact on the kinetics of the process. We propose the role of a general acid/base catalyst for Tyr280' but a Lewis acid role for Glu414. These findings are in contrast with the biogenesis of the fluorophore in GFP where the conserved Glu222 has been invoked as a general base that may abstract the Gly67 amide proton or the Tyr66 Hα proton.[42]

Finally, by taking into account the fully solvated tetramer environment when computing the QM energies, in contrast to earlier pioneering theoretical studies, we have been able to show the crucial importance and specific roles of the two water molecules (WAT_A and WAT_B) that were detected and highlighted in the seminal X-ray crystallographic work.

CONCLUSIONS

We made use of a computational chemistry setup encompassing MD simulations in explicit solvent and QM/MM calculations on the crucial parts of the trajectories to shed light on the still unresolved issue of MIO biogenesis in HAL. The two alternate reaction pathways that have been

proposed over the years were simulated and we have found strong evidence in support of one mechanism over the other. Importantly, we now show that for a realistic simulation of MIO formation in HAL it is imperative to take into account the whole tetrameric enzyme, the solvent environment, and a crucial phosphate anion.

According to our results, cyclization resulting from covalent bond formation between N(Gly144) and C(Ala142) is the first reaction step in MIO biogenesis (proposal A in Scheme 1). The distance between these atoms for nucleophilic attack is optimal in the reactive conformation $\mathbf{SP_1}$ leading to intermediate $\mathbf{I_A}$ but increases by ~1 Å and higher barrier energies are found if intermediate $\mathbf{I_B}$ is first formed instead.

To activate the cyclization step a proper environment that imposes a 'forced' geometry on the Ala142-Ser143-Gly144 tripeptide region is strictly required: on the one hand, negative electron density is polarized on the N(Gly144) nucleophile by the hydrogen bond established between the side-chain amide of Asn195 and the carbonyl oxygen of Ser143, an interaction that has passed largely unnoticed until now; on the other hand, the C(Ala142) electrophile is activated by formation of a hydrogen bond between the carbonyl oxygen of this residue and the amide hydrogen of Gly144. In addition, a positionally restrained water molecule (the highly conserved WAT_B in the crystallographic structures) assists the transfer of the amide proton from Gly144 to O(Ala142).

Although an active role in MIO formation has been previously suggested for residues Tyr280' and Glu414, the importance of Asn195 has been explained only in the context of HAL enzymatic activity mostly concerning the binding of the substrate. Our results strongly support that Asn195 can activate the Gly144 nitrogen as a nucleophile through the anchoring of its side chain with the carbonyl oxygen of Ser143 in a way reminiscent to that played by Arg96 in GFP

chromophore biosynthesis. This Ser143-Asn195 interaction was shown to be stable throughout our extensive MD simulation times and also in our QM/MM calculations of the cyclization step. Glu414 and Tyr280' play a role in the exocyclic dehydration step, the former as a Lewis acid catalyst and the latter as a general acid/base catalyst, in agreement with previous suggestions. This conclusion is in consonance with the experimental data although no evidence is available regarding the effect of amino acid replacement in these two positions on the kinetics of MIO biogenesis, an information that would undoubtedly help to further understand the role of these two residues.

Taken together, our results strongly support that (i) cyclization resulting from the nucleophilic attack of the Gly amide lone pair to the π^* orbital of the Ala carbonyl precedes dehydration, (ii) mechanical compression of the loop by neighboring protein residues enforces the reaction by eliminating inhibitory main-chain hydrogen bonds in the precursor state, and (iii) water molecules play an important role in the reaction mechanism.

Our results also highlight the dangers associated to the use, for theoretical calculations, of reduced model systems that cannot capture all the subtleties of complex reactive sites within biological macromolecules.

FIGURES

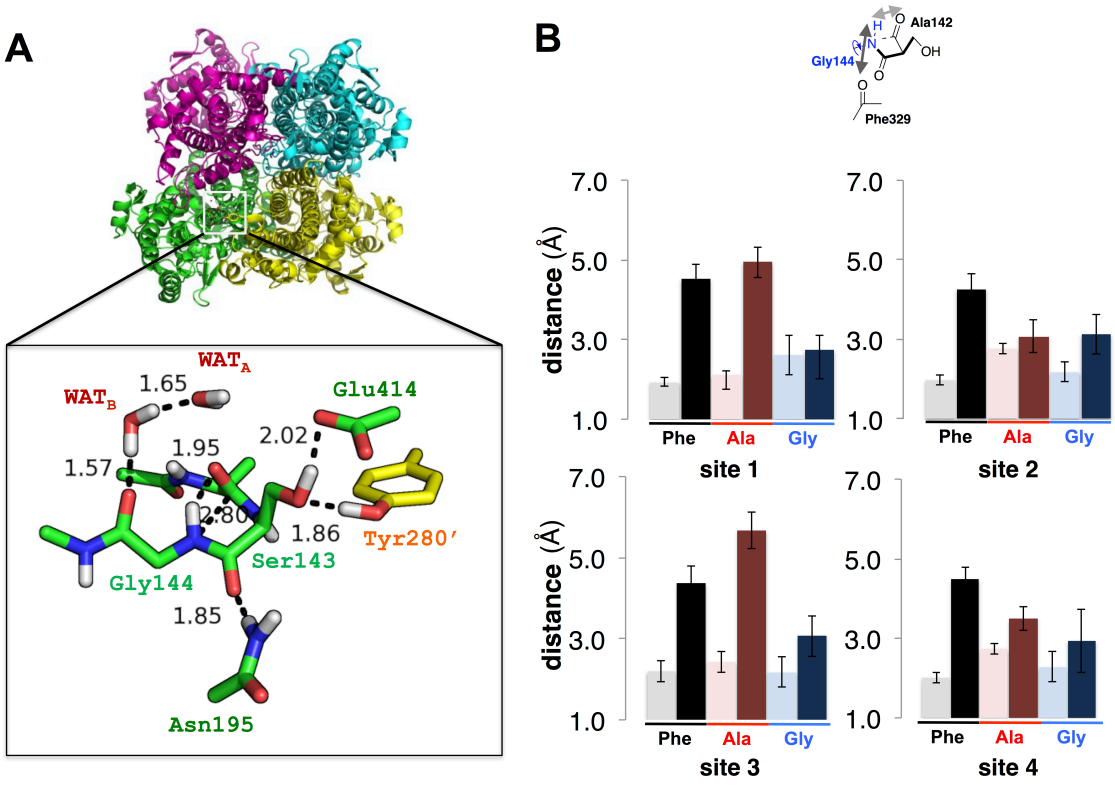


Figure 1. (A) Schematic view of homotetrameric HAL and detail of the pre-reactive conformation of the catalytic site (quantum region) identified by means of MD simulation. (B) Effect of the side chain of residue 329 on the mean distance between the Gly144 amide hydrogen and the Ala142 carbon (light colored bars) and between the Gly144 amide hydrogen and the oxygen of residue 329 (dark colored bars) in the wild-type, Phe329Ala and Phe329Gly variants in the four active centers (site 1-4) present in the homotetramer. Note the consistently shorter N(Gly144)-O(Gly329) distance relative to that between N(Gly144) and either O(Phe329) or O(Ala329).

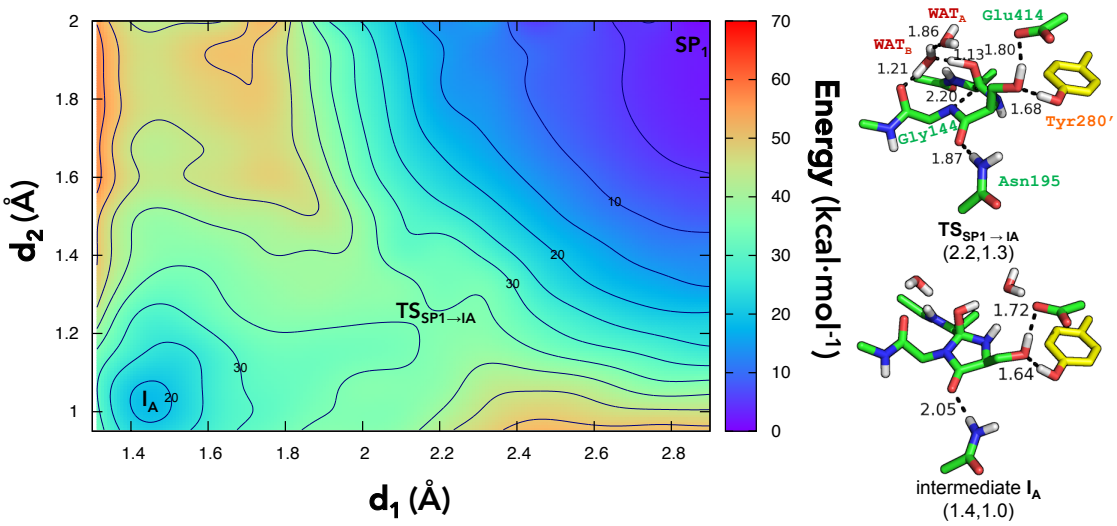


Figure 2. *Left*. Free energy (kcal mol^{-1}) surface for the nucleophilic addition following the pathway depicted in proposal A (Scheme 1). *Right*. QM/MM-identified geometries of transition state ($\mathbf{TS_{SP1\rightarrow IA}}$) and intermediate $\mathbf{I_A}$. The distances (Å) defining starting and final target values in the two reaction coordinates for each species are indicated as (d_1,d_2).

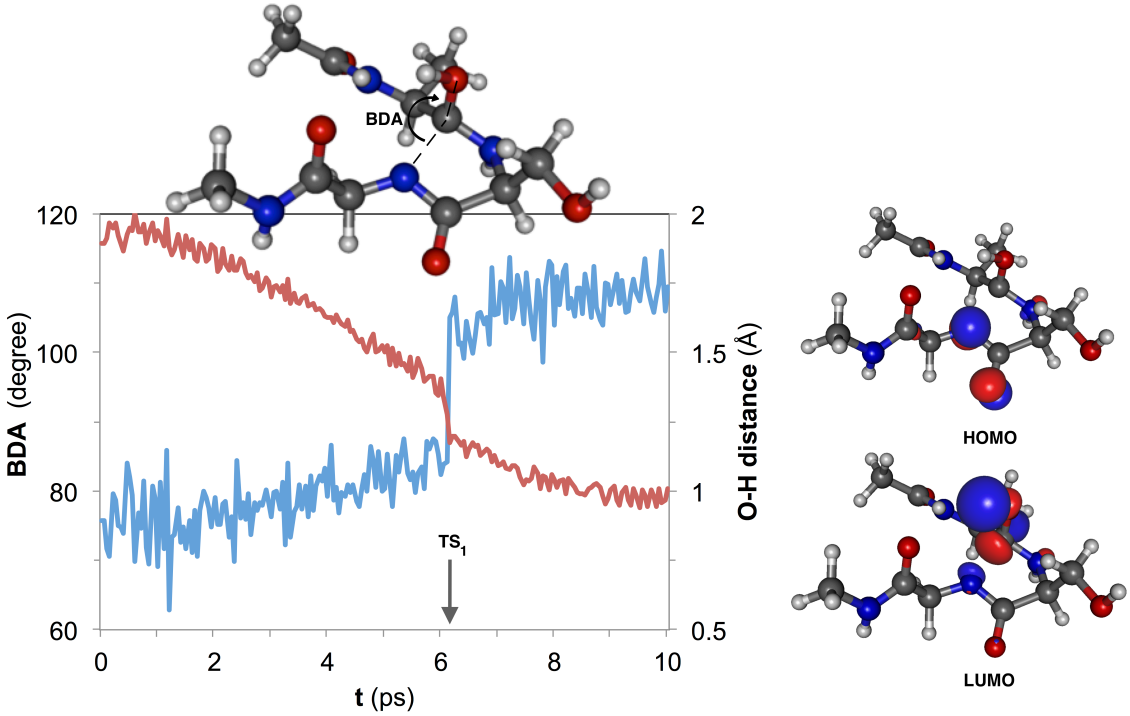


Figure 3. *Left*. Evolution of the Bürgi-Dunitz angle (BDA), defined by N(Gly144)–C(Ala142)–O(Ala142) (blue), and H(Gly144)–O(Ala142) distance (red) along the QM reaction coordinate from $\mathbf{SP_1}$ to $\mathbf{I_A}$. The time at which proton transfer occurs is indicated with an arrow. *Right*. Representation of the HOMO and LUMO in TS_1 (isovalue = 0.8 $Å^{-3}$).

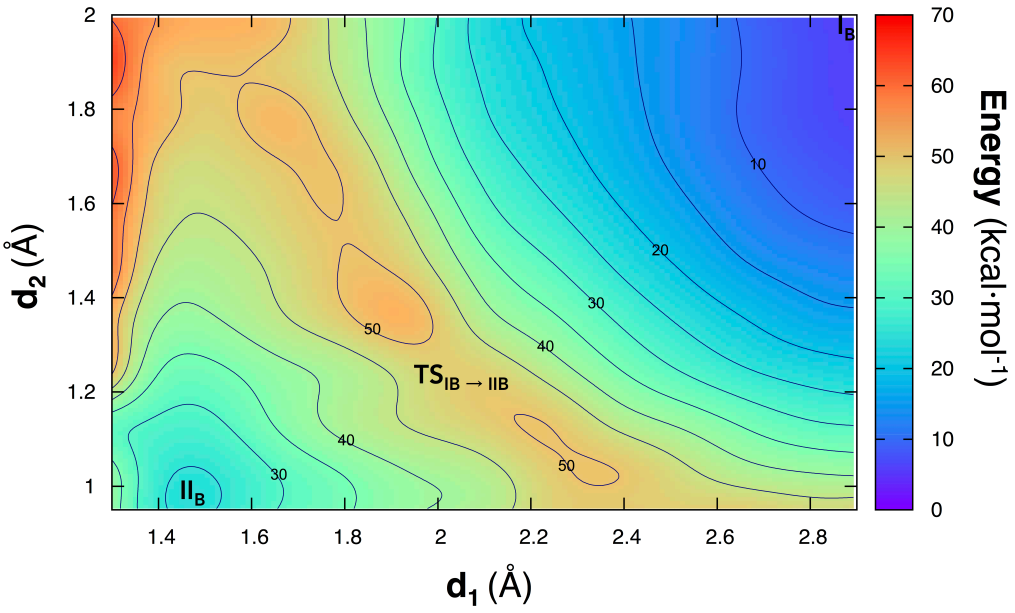


Figure 4. Free energy surface for the nucleophilic addition taking place at the second step of proposal B in Scheme 1.

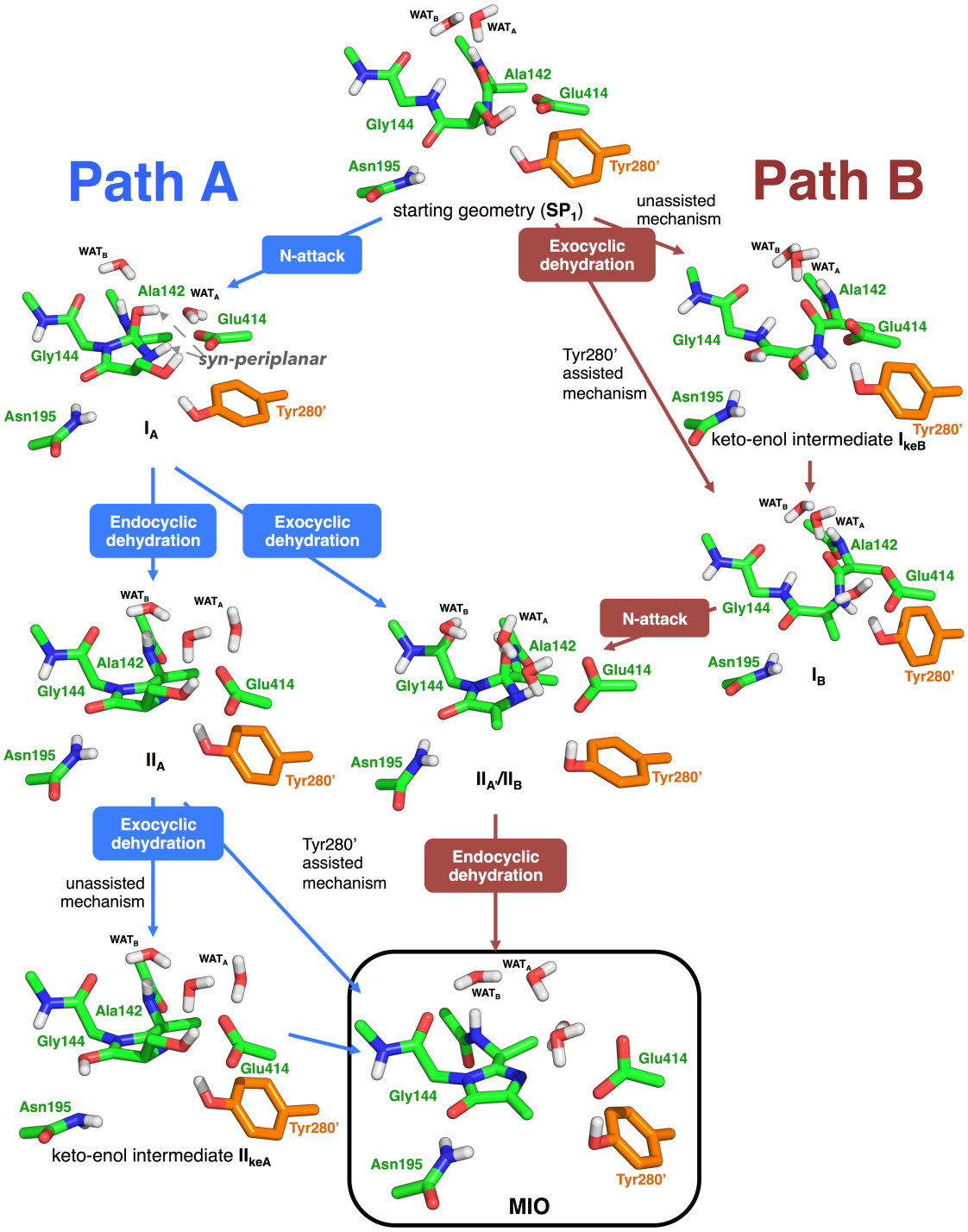
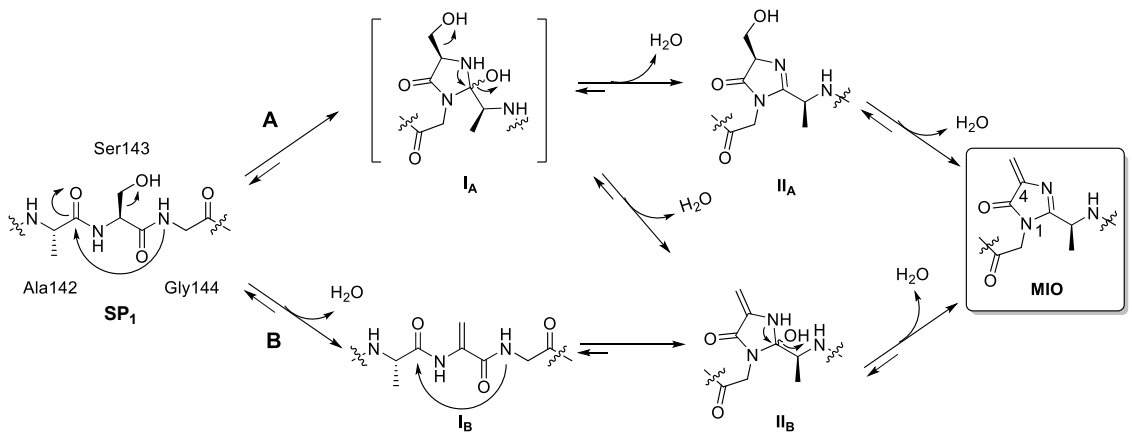
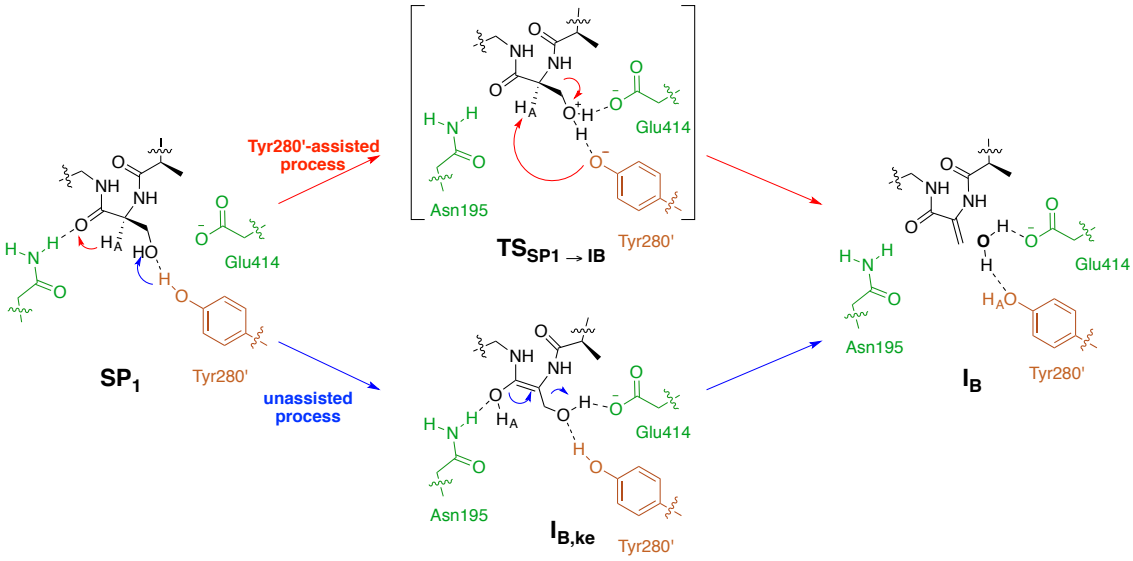


Figure 5. QM-optimized geometries for both reaction mechanism proposals A and B.

SCHEMES



Scheme 1. The two existing proposals for MIO biogenesis mainly differ in the temporal sequence of the individual events: (A) nucleophilic attack leading to cyclization followed by two dehydration steps; (B) *exo*-dehydration, nucleophilic attack, and *endo*-dehydration. Note the elimination of two water molecules in both pathways.



Scheme 2. The two proposals posited for the exocyclic elimination process: a Tyr-assisted mechanism (top pathway, in red) and the intramolecular unassisted keto-enol tautomerism of **SP1** (bottom pathway, in blue). *ke = keto-enol*.

Table 1. Computed DFT enthalpy variations (ΔH_{298}) and activation energies ($\Delta E‡$) (kcal mol^{-1})* for the chemical pathways shown in Figure 5, which contains all the possible entities considered in the present work.

Intermediate	ΔH_{298} (kcal mol^{-1})	$\Delta E^{‡}$ (kcal mol^{-1})
Starting geometry $\mathbf{SP_1}$	0.0	0.0
PATH A		
$\mathbf{I_A}$	3.6	56.8
$\mathbf{II_A}$	7.9	44.5
$\mathbf{II_{ke,A}}$	59.1	73.1, 5.5 (**)
MIO	1.3	41.4
PATH B		
$\mathbf{I_B}$	18.8	29.0
$\mathbf{I_{ke,B}}$	29.5	70.9, 8.6 (**)
$\mathbf{II_B}$	8.1	73.0
MIO	1.8	67.5

***** B3LYP/6-31G* including the zero-point correction. (**) This second energy corresponds to the evolution of the keto-enol form to either **MIO** or $\mathbf{II_B}$.

Supporting Information. The Supporting Information is available free of charge on the

ACS Publications website at DOI:

Figure S1. Schematic view of the starting structure used for the QM/MM calculations (center) and evolution of (A) N(Gly144) – C(Ala142), (B) O(Gly144) – ND(Asn195), (C) OG(Ser143) – OH(Tyr280'), and (D) OG(Ser143) – OE1(Glu414) distances in the four active sites of HAL along the exploratory MD simulation of the tetramer (60 ns). C atoms in loop residues and surrounding amino acids are colored in green and yellow, respectively.

Figure S2. Root-Mean-Square deviation (RMSd, Å) of non-hydrogen atoms and radius of gyration (degree) of wild-type (A), Phe329Ala (B), and Phe329Gly (C) variants during the 60-ns MD simulations.

Figure S3. Evolution of the CO(Ala142)-NH(Ser143) angle in **IA** during a 10-ps unrestrained QM/MM simulation.

Table S1. Summary of *P. putida* histidine ammonia-lyase variants, crystallographic evidence of MIO presence and relative enzymatic activities.

AUTHOR INFORMATION

Corresponding Author

*Prof. Federico Gago, Área de Farmacología, Departamento de Ciencias Biomédicas, Unidad Asociada al IQM-CSIC, Universidad de Alcalá, E-28805 Alcalá de Henares, Spain. E-mail: federico.gago@uah.es.

Present Addresses

† Present address: *Computational Biology, School of Life Sciences, School of Science and Engineering, University of Dundee, Dundee, DD1 5EH, United Kingdom*.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding Sources

This work was supported in part by grants SAF2012-39760-C02-02 from Ministerio de Economía y Competitividad and S2010/BMD-2457 BIPEDD2 from Comunidad Autónoma de Madrid.

ACKNOWLEDGMENT

We are grateful to Prof. José Manuel Sánchez Ruiz (University of Granada, Spain) for bringing the subject of MIO biogenesis to our attention.

ABBREVIATIONS

GFP, green fluorescent protein; HAL, histidine ammonia-lyase; MD, molecular dynamics; MIO, 4-methylidene-imidazole-5-one; PAL, phenylalanine ammonia-lyase; QM/MM, quantum mechanics/molecular mechanics; SCC-DFTB, self-consistent-charge density-functional tight-binding.

REFERENCES

[1] Poppe, L. (2001) Methylidene-imidazolone: a novel electrophile for substrate activation, *Curr. Opin. Struct. Biol. 5*, 512-524.
[2] Rétey, J. (2003) Discovery and role of methylidene imidazolone, a highly electrophilic prosthetic group, *Biochim. Biophys. Acta Prot. Proteom. 1647*, 179-184.
[3] Rechler, M. M. (1969) The purification and characterization of L-histidine ammonia-lyse (*Pseudomonas*), *J. Biol. Chem. 244*, 551-559.
[4] Langer, B., Langer, M., and Rétey, J. (2001) Methylidene-Imidazolone (MIO) from histidine and phenylalanine ammonia-lyase, In *Adv. Protein Chem.* (Judith P. Klinman, J. E. D., Ed.), pp 175-214, Academic Press.
[5] Poppe, L., and Retey, J. (2005) Friedel-Crafts-type mechanism for the enzymatic elimination of ammonia from histidine and phenylalanine, *Angew. Chem. Int. Ed. 44*, 3668-3688.
[6] Craggs, T. D. (2009) Green fluorescent protein: structure, folding and chromophore maturation, *Chem. Soc. Rev. 38*, 2865-2875.

[7] Schwede, T. F., Rétey, J., and Schulz, G. E. (1999) Crystal structure of histidine ammonia-lyase revealing a novel polypeptide modification as the catalytic electrophile, *Biochemistry 38*, 5355-5361.
[8] Givot, I. L., Smith, T. A., and Abeles, R. H. (1969) Studies on the mechanism of action and the structure of the electrophilic center of histidine ammonia lyase, *J. Biol. Chem. 244*, 6341-6353.
[9] Hanson, K. R., and Havir, E. A. (1969) Reduction of the active site of L-phenylalanine ammonia lyase, *Fed. Proc., Fed. Proc. Am. Soc. Exp. Biol. 28*, 602.
[10] Wickner, R. B. (1969) Dehydroalanine in histidine ammonia lyase, *J. Biol. Chem. 244*, 6550-6552.
[11] Baedeker, M., and Schulz, G. E. (2002) Autocatalytic peptide cyclization during chain folding of histidine ammonia-lyase, *Structure 10*, 61-67.
[12] Calabrese, J. C., Jordan, D. B., Boodhoo, A., Sariaslani, S., and Vannelli, T. (2004) Crystal structure of phenylalanine ammonia lyase:□ multiple helix dipoles implicated in catalysis, *Biochemistry 43*, 11403-11416.
[13] Barondeau, D. P., Kassmann, C. J., Tainer, J. A., and Getzoff, E. D. (2005) Understanding GFP chromophore biosynthesis:□ controlling backbone cyclization and modifying post-translational chemistry, *Biochemistry 44*, 1960-1970.
[14] Donnelly, M., Fedeles, F., Wirstam, M., Siegbahn, P. E., and Zimmer, M. (2001) Computational analysis of the autocatalytic posttranslational cyclization observed in histidine ammonia-lyase. A comparison with green fluorescent protein, *J. Am. Chem. Soc. 123*, 4679-4686.
[15] Heim, R., Prasher, D. C., and Tsien, R. Y. (1994) Wavelength mutations and posttranslational autoxidation of green fluorescent protein, *Proc. Natl. Acad. Sci. USA 91*, 12501-12504.
[16] Siegbahn, P. E. M., Wirstam, M., and Zimmer, M. (2001) Theoretical study of the mechanism of peptide ring formation in green fluorescent protein, *Int. J. Quantum Chem 81*, 169-186.
[17] Röther, D., Poppe, L., Viergutz, S., Langer, B., and Rétey, J. (2001) Characterization of the active site of histidine ammonia-lyase from *Pseudomonas putida*, *Eur. J. Biochem. 268*, 6011-6019.
[18] Cui, Q., Elstner, M., Kaxiras, E., Frauenheim, T., and Karplus, M. (2000) A QM/MM implementation of the Self-Consistent Charge Density Functional Tight Binding (SCC-DFTB) method, *J. Phys. Chem. B 105*, 569-585.
[19] Lee, T.-S., Radak, B. K., Huang, M., Wong, K.-Y., and York, D. M. (2013) Roadmaps through free energy landscapes calculated using the multidimensional vFEP approach, *J. Chem. Theory Comput. 10*, 24-34.
[20] Lee, T.-S., Radak, B. K., Pabis, A., and York, D. M. (2012) A new maximum likelihood approach for free energy profile construction from molecular simulations, *J. Chem. Theory Comput. 9*, 153-164.
[21] Anandakrishnan, R., Aguilar, B., and Onufriev, A. V. (2012) H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations, *Nucleic Acids Res. 40*, W537-W541.
[22] Vanquelef, E., Simon, S., Marquant, G., Garcia, E., Klimerak, G., Delepine, J. C., Cieplak, P., and Dupradeau, F.-Y. (2011) R.E.D. Server: a web service for deriving RESP and

ESP charges and building force field libraries for new molecules and molecular fragments, *Nucleic Acids Res. 39*, W511-W517.

[23] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys. 79*, 926-935.

[24] Case, D. A., Darden, T. A., Cheatham, T. E., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Goetz, A. W., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Wolf, R. M., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Hsieh, M. J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Salomon-Ferrer, R., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., and Kollman, P. A. (2012) AMBER 12, University of California, San Francisco.

[25] Bueren-Calabuig, J. A., Coderch, C., Rico, E., Jiménez-Ruiz, A., and Gago, F. (2011) Mechanistic insight into the catalytic activity of ββα-metallonucleases from computer simulations: *Vibrio vulnificus* periplasmic nuclease as a test case, *ChemBioChem 12*, 2615-2622.

[26] Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., and Kollman, P. A. (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method, *J. Comput. Chem. 13*, 1011-1021.

[27] Racine, J. (2006) Gnuplot 4.0: a portable interactive plotting utility, *J. Appl. Econom. 21*, 133-141.

[28] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery Jr., J. A., Peralta, J. E., Ogliaro, F., Bearpark, M. J., Heyd, J., Brothers, E. N., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A. P., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, N. J., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, Ö., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. (2009) Gaussian 09, Gaussian, Inc., Wallingford, CT, USA.

[29] Kosky, A. A., Razzaq, U. O., Treuheit, M. J., and Brems, D. N. (1999) The effects of alpha-helix on the stability of Asn residues: Deamidation rates in peptides of varying helicity, *Protein Sci. 8*, 2519-2523.

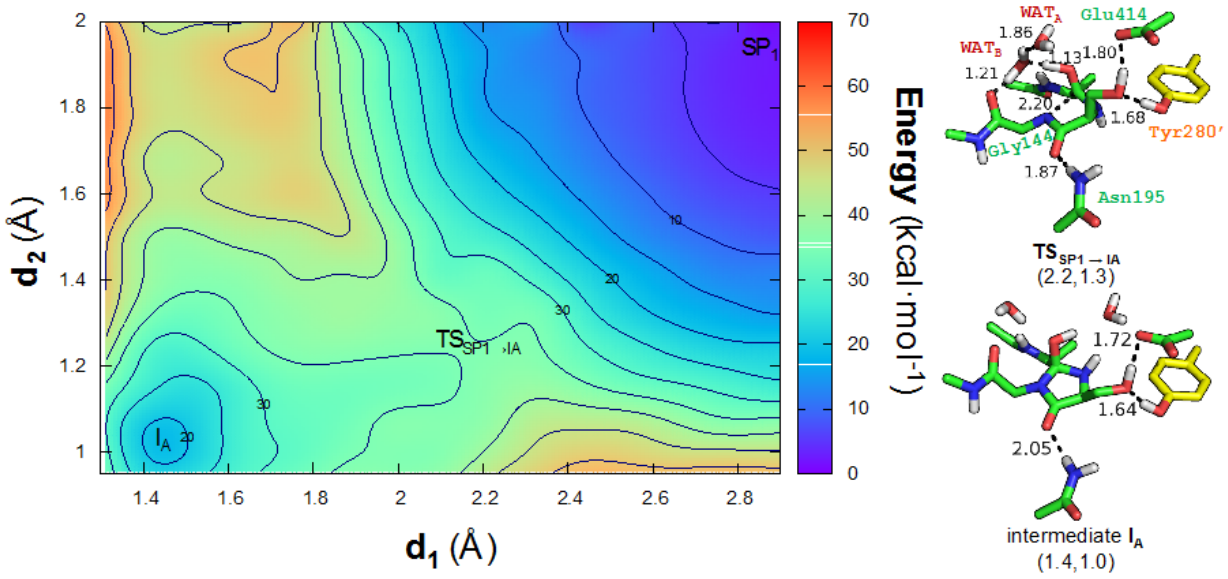
[30] Capasso, S., Mazzarella, L., Sica, F., Zagari, A., and Salvadori, S. (1993) Kinetics and mechanism of succinimide ring formation in the deamidation process of asparagine residues, *J. Chem. Soc. Perkin Trans. 2*, 679-682.

[31] Bürgi, H. B., Dunitz, J. D., Lehn, J. M., and Wipff, G. (1974) Stereochemistry of reaction paths at carbonyl centres, *Tetrahedron 30*, 1563-1572.

[32] Quesne, M. G., Borowski, T., and de Visser, S. P. (2016) Quantum mechanics/molecular mechanics modeling of enzymatic processes: caveats and breakthroughs, *Chem. Eur. J. 22*, 2562-2581.

[33] Senn, H. M., and Thiel, W. (2009) QM/MM methods for biomolecular systems, *Angew. Chem. Int. Ed. 48*, 1198-1229.
[34] Borowski, T., Quesne, M., and Szaleniec, M. (2015) QM and QM/MM methods compared: case studies on reaction mechanisms of metalloenzymes, In *Adv. Protein Chem. Struct. Biol.* (Tatyana, K.-C., Ed.), pp 187-224, Academic Press.
[35] Steinbrecher, T., and Elstner, M. (2013) QM and QM/MM simulations of proteins, In *Biomolecular Simulations: Methods and Protocols* (Monticelli, L., and Salonen, E., Eds.), pp 91-124, Humana Press, Totowa, NJ.
[36] Bueren-Calabuig, J. A., and Michel, J. (2016) Impact of Ser17 phosphorylation on the conformational dynamics of the oncoprotein MDM2, *Biochemistry 55*, 2500-2509.
[37] Bueren-Calabuig, J. A., Pierdominici-Sottile, G., and Roitberg, A. E. (2014) Unraveling the differences of the hydrolytic activity of *Trypanosoma cruzi* trans-sialidase and *Trypanosoma rangeli* sialidase: a quantum mechanics-molecular mechanics modeling study, *J. Phys. Chem. B 118*, 5807-5816.
[38] Bueren-Calabuig, J. A., Coderch, C., Rico, E., Jimenez-Ruiz, A., and Gago, F. (2011) Mechanistic insight into the catalytic activity of betabetaalpha-metallonucleases from computer simulations: Vibrio vulnificus periplasmic nuclease as a test case, *Chembiochem 12*, 2615-2622.
[39] Bueren-Calabuig, J. A., Negri, A., Morreale, A., and Gago, F. (2012) Rationale for the opposite stereochemistry of the major monoadducts and interstrand crosslinks formed by mitomycin C and its decarbamoylated analogue at CpG steps in DNA and the effect of cytosine modification on reactivity, *Org. Biomol. Chem. 10*, 1543-1552.
[40] Wachter, R. M. (2007) Chromogenic cross-link formation in green fluorescent protein, *Acc. Chem. Res. 40*, 120-127.
[41] Chen, P.-T., Wang, C.-C., Jiang, J.-C., Wang, H.-K., and Hayashi, M. (2011) Barrierless proton transfer within short protonated peptides in the presence of water bridges. A density functional theory study, *J. Phys. Chem. B 115*, 1485-1490.
[42] Sniegowski, J. A., Lappe, J. W., Patel, H. N., Huffman, H. A., and Wachter, R. M. (2005) Base catalysis of chromophore formation in Arg96 and Glu222 variants of green fluorescent protein, *J. Biol. Chem. 280*, 26248-26255.

TOC graphic


SP1
TSSP1→IA
IA
d2 (Å)
d1 (Å)
Energy (kcal·mol-1)
WATA
Glu414
WATB
Tyr280'
Gly144
Asn195
TSSP1→IA
(2.2,1.3)
intermediate IA
(1.4,1.0)