THE CATALYTIC MECHANISM OF THE MARINE-DERIVED MACROCYCLASE, PATGmac

Natércia F. Brás[a], Pedro Ferreira[a], Ana R. Calixto[a], Marcel Jaspars[b], Wael Houssen [c,d], James H. Naismith[e], Pedro A. Fernandes[a] and Maria J.Ramos[a]*

The pharmaceutical industry has become excellent at developing small-molecule drugs (below 600 MW) targeting classical compact binding sites. However, many non-classical targets remain, with extended binding sites that cannot be modulated by using small molecules, but instead need biologics (antibodies or native peptides) to obtain a therapeutic effect. Of the top selling drugs on the market, 7 out of 10 are biologics aimed at complex diseases such as rheumatoid arthritis and cancer.[1] Macrocycles (500–2000 Da) are another class of therapeutics, which are able to modulate the same complex extended targets as biologics, but are easier to administer, hit intracellular targets, and may have a lower associated cost of goods.[2] Of the 68 approved macrocycle pharmaceuticals, 27 are cyclic peptides,[3] one of which can be orally administered. Cyclic peptides have a number of advantages over linear ones, including reduced susceptibility to metabolism, improved membrane permeability, and an entropic advantage on binding to a target. In the pharmaceutical industry, two aspects of macrocycles have proven problematic: their design and efficient production. Most processes to generate peptide macrocycles rely on the use of high-dilution conditions to prevent oligomerization, using controlled addition conditions making these approaches nonviable for large-scale synthesis.[4a],[4b] Alternative synthetic approaches have been developed, including on-bead macrocyclization, which requires attachment to the solid support via an amino acid side chain, and a 3-dimensional orthogonal protecting-group strategy.[5] Biological methods of macrocyclization include sortase-mediated ligation, but this results in the incorporation of the pentapeptide Leu-Pro-X-Thr-Gly (LPXTG) in the cyclic peptide where X is variable.[6]

Efficient formation of cyclic peptides without leaving a residual sequence in the final macrocycle is desirable. Natural cyclic peptides can be formed either by a nonribosomal peptide synthetase route or by the more recently discovered superfamily of ribosomally produced and posttranslationally modified peptides (RiPPs).[5] RiPPs are formed using a common biosynthesis, in which a precursor peptide, comprised of a leader sequence followed by a core peptide often flanked by signal sequences, is modified through the action of processing enzymes, which install posttranslational modifications in the core peptide.

The matured core peptide is then removed from the leader and signal sequences, liberating the mature, active peptide. In many cases, this last step results in the formation of a peptide macrocycle. Three RiPP macrocyclases have been defined; butelase-1, an asparagine/aspartate peptide ligase that is responsible for the formation of plant cyclic peptide cyclotides,[7] GmPOPB, a prolyl oligopeptidase involved in α-amanitin biosynthesis,[8] and PatG macrocyclase (PatGmac; Figure 1), which is involved in the biosynthesis of cyanobactins (cyanobacterially derived azole/azoline containing cyclic peptides).[9] The requirements of the PatGmac are a core peptide sequence of 6–11 residues which may include d-amino acids or unnatural amino acid residues, an AYD(GE) signal at the C terminus that is not incorporated in the final cyclic peptide and a cyclic residue (proline or thiazoline) at the C terminus of the core peptide that is included in the macrocycle.[10]

Our structural investigations have delineated the mode of action of PatGmac and have led to an understanding of the requirements described above.[9] PatGmac is a subtilisin-like serine protease produced by Prochloron sp., which is an obligate symbiont of the seasquirt Lissoclinum patella.[11] The more normal extended conformation of the peptide substrate is prevented by bulky enzyme residues (Met660/Arg686/Phe684) protruding into the binding groove as a consequence of a disulfide bridge formation between Cys685 and Cys724. As a result of this, only a bent substrate peptide can bind and this is facilitated by the conformational properties of the proline or thiazoline residue at the P1 position (P3, P2, P1 are the residues before the cleavage site and P1′, P2′ those after the cleavage site).

The unique structural feature of PatGmac is an insertion loop in the normal subtilisin sequence, which generates a helix–turn–helix motif that forms a protective lid over the active site. The AYD signal at the C terminus of the core peptide binds via the aspartate residue to basic residues in the helix–turn–helix motif (Arg589/Lys594/Lys598). After binding the substrate peptide, a normal serine protease mechanism ensues, but access by water is prevented by the strongly bound AYD signal, thus allowing the amino terminus of the core peptide to loop around and attack the acyl complex, resulting in cleavage of the recognition signal and cyclization of the core peptide.

In this work, we explored the mechanism of the reaction catalyzed by PatGmac using computational methods, to describe in atomistic detail the most probable catalytic mechanism of this enzyme. Considering the ability of the enzyme to macrocyclize an extensive range of nonactivated substrates, the data obtained in the present study has wide implications in a number of areas. Fully understanding the mechanism of PatGmac may lead to engineered analogues with improved properties that accept an even broader substrate range, thus increasing its utility in a range of pharmaceutical and biotechnological applications.

Results and Discussion

In previous work, some of us proposed a catalytic mechanism for PatGmac[9] (Scheme 1) based on the structure of the enzyme macrocyclase domain and biochemical characteristics. We suggested that the first step of the reaction is the nucleophilic attack of Ser783 aided by His618, which acts as a base, to the P1′ carbonyl group of the substrate, leading to the formation of an enzyme–substrate tetrahedral intermediate. After this first attack, the AYDG peptide is cleaved by protonation of the leaving group's N terminus by His618. Lastly, the N terminus of the substrate attacks the carbonyl group of the cysteine (P1) to form the macrocycle (Scheme 1). To test the proposed mechanism and provide a complete and detailed description of this catalytic mechanism, we performed QM/MM calculations.

First reaction step

In the optimized reactant structure, Ser783 is situated 3.89 Å from the carbon atom of P1 and correctly oriented to begin the attack. Asp548 is hydrogen-bonded to His618 (1.69 Å) and the latter is hydrogen-bonded to Ser783 (1.54 Å; Figure 2). The catalytic reaction begins with a typical proton transfer on the serine protease's catalytic triad, from Ser783 to His618. We defined the distance between Nδ1 of His618 and the Hγ of Ser783 as the putative reaction coordinate. As the proton of Ser783 was being transferred to His618, the proton of the His618 imidazole ring spontaneously moved to the Asp548 O− atom. We noticed that the structures of the reactants (Asp-COO−/His-H/Ser-OH; Figure 2) and of the generated intermediate (Asp-COOH/His-H/Ser-O−; Figure 3) are very close in energy (2.4 kcal mol−1). The free-energy barrier was calculated as 2.5 kcal mol−1. At the transition state structure, both the proton transferred from Ser783 to His618 and the one that was subsequently transferred from His618 to Asp548 assume an intermediate position between the respective residues (Figure 3).
The vibrational frequency analysis revealed an imaginary frequency corresponding to the reaction coordinate (an antisymmetric stretch involving both proton transfers) confirming the nature of the transition state (996.52 i cm−1). However, another small negative frequency (31.30 i cm−1) was found, as a consequence of the more relaxed geometry optimization.

Thermal corrections and zero-point energies were included in the free-energy calculation, despite the existence of a very small second imaginary constant. These corrections have a small contribution to the corresponding free-energy values in agreement to what has been reported in the literature regarding proton transfers between catalytic residues (ca. 1.2 kcal mol−1).[30] As the barrier was extremely shallow and not rate limiting, it was not optimized further.

The small free energy of reaction suggests that both structures may exist simultaneously with the latter being the most favorable to initiate the catalytic reaction. The possibility of the first step to begin with a direct nucleophilic attack of Ser783 on the substrate, with the proton transfer occurring during the attack was also considered. Defining the distance between Ser783 and P1 as the reaction coordinate, the energy barrier associated with this alternative was found to be very high (>30 kcal mol−1). Our proposal for the first step has a much smaller energy barrier (2.5 kcal mol−1), indicating that the proton transfer leading to an intermediate with enhanced nucleophilicity for attack at the substrate amide.

Second reaction step

The deprotonated Ser783 attacks the P1 carbonyl group, forming the enzyme-linked tetrahedral intermediate. A study of this step was made with a linear transit scan, with the aforementioned optimized structure after the proton transfer as reactants, and defining the distance between the Oγ of Ser783 and the carbonyl carbon of P1, which is initially 3.90 Å (Figure 3), as the reaction coordinate to be scanned.

A conformational change in His618 took place before the formation of the tetrahedral intermediate, with the approximation of Ser783 to the P1 carbonyl group, His618 changed its rotamer by rotating around the Cβ−Cγ bond, breaking the hydrogen bond with Ser783 hydroxy and forming another with the Ser783 carbonyl, while keeping the hydrogen bond with Asp548. Additionally, a hydrogen bond between the attacking oxygen of Ser783 and Asn717 amine was formed (Figure 4). This conformational change led to a stable intermediate (INT2) before Ser783 has completed the nucleophilic attack on the substrate. A free-energy barrier of 4.8 kcal mol−1 in relation to INT1 was found, characterized by one imaginary frequency (31.83 i cm−1) clearly corresponding to the reaction coordinate. We will refer to this point as TS2 (Figure 4). In that structure, His618 is situated between the Ser783 side chain and carbonyl oxygen atoms at distances of 2.22 Å and 2.28 Å respectively. Ser783, which initially was situated 3.90 Å from the P1 residue, stays at a distance of 3.57 Å in TS2 and at 3.61 Å in INT2.

Third reaction step

Starting from the optimized structure of INT2, the attack of Ser783 at the substrate was conducted resulting in formation of the tetrahedral intermediate, using the distance between the Ser783 deprotonated oxygen and the P1 carbonyl carbon (3.62 Å) as reaction coordinate (Figure 4). This attack had a free-energy barrier of 21.1 kcal mol−1 in relation to INT2. This rate-limiting transition state (TS3) was further freely optimized, having one imaginary frequency (744.16 i cm−1) that corresponds to the coordinated stretching of the bond between Ser783 oxygen and P1 carbonyl carbon and of the N−H bond of the terminal amino group of the substrate.

This proton, together with the Asn717 amine, constitutes the oxyanion hole that stabilizes the oxyanion of the acyl–enzyme complex. The interactions between them may consist of low-barrier hydrogen bonds (LBHBs) considering, particularly, their short length (<2.0 Å). The formation of LBHBs between the oxyanion and the oxyanion hole in serine proteases has previously been proposed,[31] although other reports considered them to be simple electrostatic interactions.[32] In this case, we can clearly see the transition from an electrostatic hydrogen bond in the reactants to a hydrogen bond where the proton is shared between both acceptors, commonly known as a single-well hydrogen bond, which corresponds to

an extreme case of a low-barrier hydrogen bond in which the barrier vanishes. However, we note that this structure, achieved at the transition state, is not a stable, long-lived interaction. Whether or not this kind of interaction is catalytic or anticatalytic is a matter of debate.[33]
Fourth reaction step

The conformational rearrangement that His618 makes it unlikely that it protonates the leaving group in the third step of the catalytic reaction, as initially proposed and which is typical for other serine proteases.[32] After the rearrangement, His618 remains at 5.97 Å from P1′ and not adequately oriented (Figure 5), the substrate's N terminus (charged) amine group would be the most suitable candidate to protonate P1′ amine.

.
Therefore, we conducted a linear transit scan using the distance between the NH3+ terminal proton and the P1′ amine group, which is 2.01 Å in the INT3 geometry, as reaction coordinate (Figure 5). A transition state (TS4) was then freely optimized and an imaginary frequency (1184.98 i cm−1) was found that corresponded to an asymmetric stretch between the transferred proton and the two nitrogen atoms (Figure 6). We found an energy barrier of 3.1 kcal mol−1 relatively to INT3. With the transfer of the proton, a very stable intermediate (INT4) is obtained [ΔG(step 4)=−23.8 kcal mol−1] and the P1−P1′ bond is cleaved but the AYDG tetrapeptide is retained on the active site of the enzyme, at 2.95 Å from P1 on the INT4 optimized structure (Figure 6).

To complete the catalytic mechanism of PatGmac, the substrate's N terminus attacks the carbon atom of P1 carbonyl group closing the macrocycle. We have used the distance between the N terminus and the P1 carbonyl carbon (4.99 Å) as reaction coordinate. We found that during this step the N terminus donates a proton to Ser783, consequently decomposing the enzyme–substrate tetrahedral structure by cleaving the bond between P1 and Ser783. The macrocycle is formed and the AYDG peptide, which was situated 2.95 Å from P1 at the beginning of this step, is further displaced. This step has a free-energy barrier of 19.8 kcal mol−1 in relation to INT4 and originates very stable products [ΔG(step 5)=−42.6 kcal mol−1; Figure 7], indicating that the displacement of the AYDG peptide is an important factor contributing to the great stability of the products. The freely optimized TS5 geometry shows one imaginary frequency (136.59 i cm−1), corresponding to the reaction coordinate. The present description of this step differs from the first proposed mechanism,[9] in that the substrate's N terminus donates a proton to Ser783 rather than His618.

The energetic pathway of the macrocyclization reaction catalyzed by PatGmac at the ONIOM(M06/6-311++G(2d,2p):Amber//B3LYP/6-31G(d):Amber) level is shown in Figure 8. According to these results, TS2 has the highest free energy. However, that is not the rate-limiting step, since both TS3 and TS5 have higher energy barriers in relation to the preceding intermediates, which are very stable. These TSs represent the formation of the enzyme–substrate complex (TS3) and the formation of the macrocycle and accompanying decomposition of the acyl–enzyme complex (TS5). TS3 has a slightly higher energy barrier than TS5 (21.1 kcal mol−1 and 19.8 kcal mol−1, respectively), but, given their proximity, both are relevant for the observed rate of this reaction (note that the difference between the two is very narrow and probably close to the accuracy of the methodology for relative energies between similar molecular structures). Both the formation and decomposition of the tetrahedral complex have previously been identified as slow steps,[32, 34] in accordance with our results reported herein and with our previous mass spectrometric detection of the acyl PatGmac intermediate. The potential energy surface (PES) of the reaction shows that it is strongly exothermic with the products 76.0 kcal mol−1 more stable than the

reactants. All mechanistic steps are exothermic with the exception of step 3 (formation of the acyl–enzyme intermediate).

.
The reaction rates reported for PatGmac are approximately 1 per day.[10, 35] Thus, the Gibbs free-energy barrier for the macrocyclization reaction may be estimated from the transition state theory, resulting in an observed experimental free energy of approximately 24 kcal mol−1, which is a comparable value to that obtained in the present work (21.1 kcal mol−1).

Conclusion

We have explored by computational approaches the macrocyclization reaction catalyzed by PatGmac. Our results showed that the mechanism followed by this enzyme is different to those typical of serine proteases, albeit with some similarities (Scheme 2). The typical proton transfer on the Ser–His–Asp catalytic triad occurs, as we described in the first reaction step, as well as the formation of the acyl–enzyme intermediate (second mechanistic step). We found, however, that the protonation of the AYDG leaving group of the substrate is most probably made by the substrate's NH_3^+ terminus and not by His618 (Scheme 2), due to steric impediments (Scheme 1). This enzyme differs from typical serine proteases in which deacylation of the enzyme–substrate complex occurs by attack of a water molecule regenerating the enzyme.[32] In PatGmac, the active site is shielded from water[9] and, because of that, deacylation is achieved by attack of the substrate's N terminus, which also protonates the acyl complex. The formation of the peptide bond is common to macrocyclization of other peptide substrates, namely other cyanobactins.[9, 36]

This new mechanism differs from that proposed previously (Schemes 1and 2);[9] the central difference is that His618 undergoes a conformational rearrangement and does not protonate the leaving group. Rather, it is the incoming substrate amino terminus that protonates the leaving group.

This study contributes to further understanding of the mechanism of macrocyclization of the PatG substrate. As cyclic peptides have great interest for industry, particularly pharmaceutical, more knowledge about their natural synthesis also contributes to improving the efficiency of large-scale production of such compounds. The findings reported herein suggest that adapting the enzyme process to utilize different substrates would need careful consideration of the pKa of the incoming nucleophile. Regarding the active site residues, the Asp–His–Ser triad is mandatory for the function of the enzyme and must be maintained in any engineered analogue, while Asn717 stabilizes the tetrahedral intermediate.

Viable options to increase the reaction rate have to focus on the rate-limiting step. For this reason, the most promising way to achieve catalysis would be to optimize the oxyanion hole. The presence of a positively charged residue at position 717 (i.e., Lys) would possibly further favor the nucleophilic attack of Ser783 on the substrate, which we found to be a rate-limiting step of the reaction. However, the size of Lys and its proximity to the positive N terminus are obstacles for the correct placement and protonation of the lysine.

Computational Methods

We started the modeling of the system by using an X-ray structure of the subtilisin-like domain of the PatGmac enzyme containing an analogue of the substrate in the active site (PDB:4AKT, 2.63 Å resolution). This structure had a mutation (His618Ala) on the catalytic triad of the active site (Asp548,

His618 and Ser783), which was reverted by superimposition with a structure of the free enzyme (PDB:4AKS, 2.19 Å resolution) that contained the original residue. The coordinates of Ala618 on the 4AKT structure were then replaced by those of His618 of the free enzyme structure. The hydrogen bonding network between the three residues of the catalytic triad and the proximity between the deprotonated nitrogen of the imidazole ring of His618 and the hydroxy group of Ser783, observed on the free enzyme structure, are mandatory to generate a productive conformation and to initiate the catalytic mechanism. Thus, we naturally assumed that the rotameric state and position of His618 is similar in the free enzyme and in the enzyme complexed with the substrate. Additionally, in the X-ray structure, a loop composed by residues 651–657 was missing and was modeled by using the program MODELLER.[12] This loop is located >10 Å away from the active site, thus, the modeling performed should not have a significant effect on the study of the catalytic mechanism. One natural substrate has the sequence Ile–MeOxH–Ala–ThH–Ile–OxH–Phe–ThH–Ala–Tyr–Asp–Gly, the sequence Ala–Tyr–Asp–Gly represent the recognition residues cleaved (at the ThH−Ala bond) during the reaction (ThH=thiazoline, OxH=oxazoline, MeOxH=methyl oxazoline). The mimic peptide on the 4AKT structure had a different sequence (Val–Pro–Ala–Pro–Ile–Pro–Phe–Pro–Ala–Tyr–Asp–Gly) in which the azoline heterocycles had been replaced by Pro. Hence, the heterocycles that on the X-ray structure were mimicked by prolines were corrected and missing parts were modeled in GaussView.[13] We have preserved the original coordinates for most of the residues of the substrate (Figure SI-1 in the Supporting Information shows the original precursor peptide and the modeled peptide). We used X-Leap[14] software to protonate the complex, assuming that all residues were in their physiological protonation states, and 23 Na^+ counter ions were added to neutralize the charge of the system. Additionally, we surrounded the system with 15 830 water molecules by using a truncated rectangular box of TIP3P water molecules with a minimum distance of 12 Å between any atom of the protein and the faces of the box.

We performed a two-step minimization by using the Amber 12[14] simulation package (parm 99SB force field) to relax the system by removing eventual tensions and clashes. Firstly, the water molecules and counter ion were minimized with the remainder of the system fixed (steepest descent algorithm for the first 500 cycles, and conjugate gradient algorithm for the last 1500 steps). Secondly, the position of all atoms (steepest descent algorithm for the first 5000 cycles, and conjugate gradient algorithm for the last 10 000 steps) of the model was minimized.

We ran molecular dynamics (MD) simulations, starting from the structure obtained after the minimization procedure, to see if the modeled structure was stable and preserved. First we warmed the system from 0 to 300 K in a 200 ps long simulation maintaining a constant volume and using periodic boundary conditions. Then, a 15 ns production run was conducted under periodic boundary conditions with the isobaric-isothermal ensemble defining a pressure of 1 atm and a temperature of 300 K, using the Langevin thermostat and the Berendsen barostat for that purpose. The cutoff for Lennard–Jones interactions was set to 10 Å. The Coulomb interactions were treated by using the Particle Mesh–Ewald (PME) method, with a cutoff of 10 Å for the real part of the sum. The time step of the simulation was 2 fs. The potential energy of the system was treated with the LeapFrog integration algorithm.

A structure of PatGmac complexed with substrate was taken from the MD simulation and was used as the starting point for the study of its reaction mechanism. We chose a structure of the system (after the equilibration) whose conformation was productive for the beginning of the catalytic cycle, which means that the distance between Ser783 and the substrate, and between the substrate's N terminus and P1 had to be small and appropriate for chemical reaction. The analysis of the trajectories has shown that, after the equilibration, these criteria are met in most of the structures, hence, the entropic cost for reaching a productive conformation should not be significant. To investigate the PES along the mechanism of the macrocyclization reaction, we performed QM/MM calculations, widely used in enzymatic studies[15-18] applying an ONIOM scheme,[19] as implemented in the Gaussian 09 software package.[20] The system,

containing a total of 5200 atoms, was divided into a “QM layer” containing 91 atoms (Figure 1) and an “MM layer,” which were treated at DFT and classical MM levels, respectively. The high layer includes the catalytic triad [Ser783, His618 (side chain until Cα), and Asp548 (side chain until Cβ)], Asn717 (side chain until Cγ) and P1 (Cys), P1′ (Ala), and P2′ (Tyr) residues, and the terminal Ile of the substrate (a list of the atoms in the QM layer is given in table SI-1 in the Supporting Information). For the QM layer, we employed the B3LYP functional using the 6-31G(d) basis set for geometry optimizations, which was shown to provide accurate results in previous studies,[21-24] whereas in the lower layer we used the AMBER parm 99SB force field. We used hydrogen atoms as link atoms where covalent bonds were between the two layers. The interaction between the two layers was treated by using an electrostatic embedding scheme. The study of each mechanistic step began by conducting linear scans along the reaction coordinates. These corresponded to specific interatomic distances that connected the reactants to the products of each hypothesized reaction step. The precise reaction coordinates that were assumed are described in the main text, where each reaction step is discussed. The structures of the reactants, intermediates, transition states (TS), and products were then fully optimized, starting with the guesses, taken from the linear transit scans, for the rate-limiting steps. In the case of chemical steps with very shallow barriers, we used the highest energy structures of the linear transit scan as very good transition state approximations. This procedure was motivated by the great complexity of performing free transition state geometry optimization in this very large and heterogeneous system. The differences in free energy that we derived when we compared the two procedures (i.e., taking the structure from the adiabatic mapping and making a free geometry optimization, both alternatives calculated for the rate-limiting steps) were 0.4 and 2.6 kcal mol−1. We performed vibrational frequency calculations for every stationary point, with those of the reactant, intermediates and product having no imaginary frequencies and those of the TS having just one, which in all cases was clearly related to the reaction coordinate. Even though the identities of the minima connected to each transition state were clear from the liner transit scans and were further supported by the observation of the relevant normal mode, we ran further internal reaction coordinate (IRC) calculations (albeit not to the full extent of the explored PESs), starting from the obtained TSs, to confirm that the minima that were connected to the TS were the ones that we were expecting. The final energies were obtained by conducting single-point (SP) energy calculations on the optimized geometries using different density functionals (M06, B1B95, and mPWB1K, in addition to B3LYP), known to have a good performance for thermodynamics and kinetics, and a larger basis set, 6-311++G(2d,2p), in the QM layer to improve the accuracy of the results (Table SI-2 in the Supporting Information shows the activation and reaction energies obtained with the different functionals). They provide a PES that translates into the same mechanism and are qualitatively equivalent. The values obtained with B3LYP seem somewhat elevated compared to those obtained with the other functionals. The energy barriers for each step are similar in all cases. However, with B3LYP and M06 the energy barrier to TS3 is slightly higher than the barrier to TS5, whereas with mPWB1K and B1B95 the opposite is observed. In the discussion we considered the results of the M06 functional because it has been shown to be the most appropriate for the description of the thermodynamics and kinetics of the chemistry of main-group elements.[25-27] We have calculated also the zero-point energy, the entropy and the thermal corrections to obtain Gibbs free energies at 298.15 K, which is a comparable temperature to that of the water where the Indo-Pacific seasquirt Lissoclinum patella naturally occurs.[28] To calculate the entropy and free energy, we employed the particle in a box/rigid rotor/harmonic oscillator formalism. This is a physically clear and rigorous formalism to calculate the entropy and free energy of a system within a single-conformation model.GD3 dispersion[29] was included in the single point calculations, as implemented in Gaussian 09 version D.01.

Acknowledgements

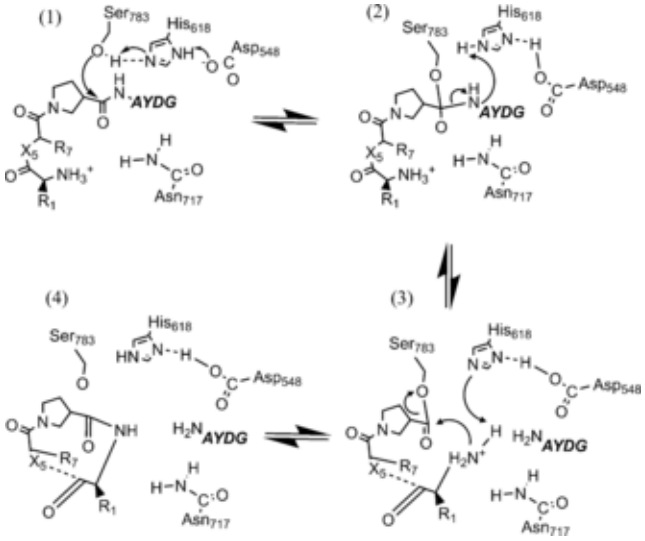
This work received financial support from the European Union (FEDER funds through COMPETE) and National Funds (Fundação para a Ciência e Tecnologia, FCT) through Projects EXCL/QEQ-

COM/0394/2012, EXCL-II/QEQ-COM/0394/2012, and Pest-C/EQB/LA0006/2013. N.F.B. would like to thank the FCT for her IF starting grant (IF/01355/2014).

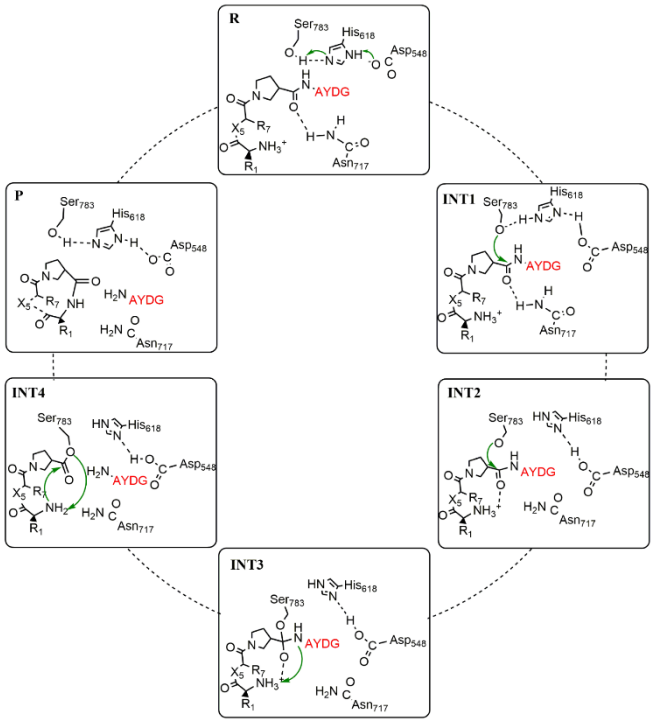
References

[1] in C&EN Supplement "The top 50 drugs of 2014", Vol. September 2014.
[2] E. A. Villar, D. Beglov, S. Chennamadhavuni, J. A. Porco, D. Kozakov, S. Vajda and A. Whitty, Nature Chemical Biology 2014, 10, 723-731.
[3] F. Giordanetto and J. Kihlberg, Journal of Medicinal Chemistry 2014, 57, 278-295.
[4] a) E. M. Driggers, S. P. Hale, J. Lee and N. K. Terrett, Nature Reviews Drug Discovery 2008, 7, 608-624; b) C. J. White and A. K. Yudin, Nature Chemistry 2011, 3, 509-524.
[5] P. G. Arnison, M. J. Bibb, G. Bierbaum, A. A. Bowers, T. S. Bugni, G. Bulaj, J. A. Camarero, D. J. Campopiano, G. L. Challis, J. Clardy, P. D. Cotter, D. J. Craik, M. Dawson, E. Dittmann, S. Donadio, P. C. Dorrestein, K. D. Entian, M. A. Fischbach, J. S. Garavelli, U. Goransson, C. W. Gruber, D. H. Haft, T. K. Hemscheidt, C. Hertweck, C. Hill, A. R. Horswill, M. Jaspars, W. L. Kelly, J. P. Klinman, O. P. Kuipers, A. J. Link, W. Liu, M. A. Marahiel, D. A. Mitchell, G. N. Moll, B. S. Moore, R. Muller, S. K. Nair, I. F. Nes, G. E. Norris, B. M. Olivera, H. Onaka, M. L. Patchett, J. Piel, M. J. T. Reaney, S. Rebuffat, R. P. Ross, H. G. Sahl, E. W. Schmidt, M. E. Selsted, K. Severinov, B. Shen, K. Sivonen, L. Smith, T. Stein, R. D. Sussmuth, J. R. Tagg, G. L. Tang, A. W. Truman, J. C. Vederas, C. T. Walsh, J. D. Walton, S. C. Wenzel, J. M. Willey and W. A. van der Donk, Natural Product Reports 2013, 30, 108-160.
[6] Z. M. Wu, X. Q. Guo and Z. W. Guo, Chemical Communications 2011, 47, 9218-9220.
[7] G. K. T. Nguyen, S. J. Wang, Y. B. Qiu, X. Hemu, Y. L. Lian and J. P. Tam, Nature Chemical Biology 2014, 10, 732-738.
[8] H. Luo, S. Y. Hong, R. M. Sgambelluri, E. Angelos, X. Li and J. D. Walton, Chemistry & Biology 2014, 21, 1610-1617.
[9] J. Koehnke, A. Bent, W. E. Houssen, D. Zollman, F. Morawitz, S. Shirran, J. Vendome, A. F. Nneoyiegbe, L. Trembleau, C. H. Botting, M. C. M. Smith, M. Jaspars and J. H. Naismith, Nature Structural & Molecular Biology 2012, 19, 767-772.
[10] J. A. McIntosh, C. R. Robertson, V. Agarwal, S. K. Nair, G. W. Bulaj and E. W. Schmidt, Journal of the American Chemical Society 2010, 132, 15499-15501.
[11] W. E. Houssen and M. Jaspars, Chembiochem 2010, 11, 1803-1815.
[12] A. Sali and T. L. Blundell, Journal of Molecular Biology 1993, 234, 779-815.
[13] R. K. Dennington, Todd; Millam, John in GaussView, Version 5, Vol. Semichem Inc., Shawnee Mission, KS, 2009.
[14] T. A. D. D.A. Case, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Götz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman in AMBER 12, Vol. University of California, San Francisco, 2012.
[15] M. E. Alberto, T. Marino, M. J. Ramos and N. Russo, Journal of Chemical Theory and Computation 2010, 6, 2424-2433.
[16] M. Cascella, C. Micheletti, U. Rothlisberger and P. Carloni, Journal of the American Chemical Society 2005, 127, 3734-3742.
[17] B. F. Ion, E. A. C. Bushnell, P. De Luna and J. W. Gauld, International Journal of Molecular Sciences 2012, 13, 12994-13011.
[18] R. Lonsdale and A. J. Mulholland, Current Topics in Medicinal Chemistry 2014, 14, 1339-1347.
[19] F. Maseras and K. Morokuma, Journal of Computational Chemistry 1995, 16, 1170-1179.
[20] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J.

Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox in Gaussian 09, Vol. Gaussian, Inc., Wallingford, CT, USA, 2009.
[21] N. F. Bras, M. J. Ramos and P. A. Fernandes, Physical Chemistry Chemical Physics 2012, 14, 12605-12613.
[22] A. R. Calixto, N. F. Bras, P. A. Fernandes and M. J. Ramos, Acs Catalysis 2014, 4, 3869-3876.
[23] G. P. Pinto, N. F. Bras, M. A. S. Perez, P. A. Fernandes, N. Russo, M. J. Ramos and M. Toscano, Journal of Chemical Theory and Computation 2015, 11, 2508-2516.
[24] A. J. M. Ribeiro, D. Santos-Martins, N. Russo, M. J. Rarnos and P. A. Fernandes, Acs Catalysis 2015, 5, 5617-5626.
[25] Y. Zhang, X. Xu and W. A. Goddard, Proceedings of the National Academy of Sciences of the United States of America 2009, 106, 4963-4968.
[26] Y. Zhao and D. G. Truhlar, Chemical Physics Letters 2011, 502, 1-13.
[27] Y. Zhao and D. G. Truhlar, Theoretical Chemistry Accounts 2008, 120, 215-241.
[28] L. Behrendt, A. W. D. Larkum, E. Trampe, A. Norman, S. J. Sorensen and M. Kuhl, Isme Journal 2012, 6, 1222-1237.
[29] S. Grimme, J. Antony, S. Ehrlich and H. Krieg, Journal of Chemical Physics 2010, 132.
[30] D. H. Wei, X. Q. Huang, J. J. Liu, M. S. Tang and C. G. Zhan, Biochemistry 2013, 52, 5145-5154.
[31] J. A. Gerlt and P. G. Gassman, Biochemistry 1993, 32, 11943-11952.
[32] L. Hedstrom, Chemical Reviews 2002, 102, 4501-4523.
[33] A. Warshel, P. K. Sharma, M. Kato, Y. Xiang, H. B. Liu and M. H. M. Olsson, Chemical Reviews 2006, 106, 3210-3235.
[34] J. Kraut, Annual Review of Biochemistry 1977, 46, 331-358.
[35] J. Lee, J. McIntosh, B. J. Hathaway and E. W. Schmidt, Journal of the American Chemical Society 2009, 131, 2122-+.
[36] W. Y. Xu, L. L. Li, L. C. Du and N. H. Tan, Acta Biochimica Et Biophysica Sinica 2011, 43, 757-762.



Scheme 1 Catalytic mechanism proposed for PatGmac.[9]



Scheme 2. General scheme for the catalytic mechanism of PatGmac predicted by earlier experiments and by present QM/MM calculations.

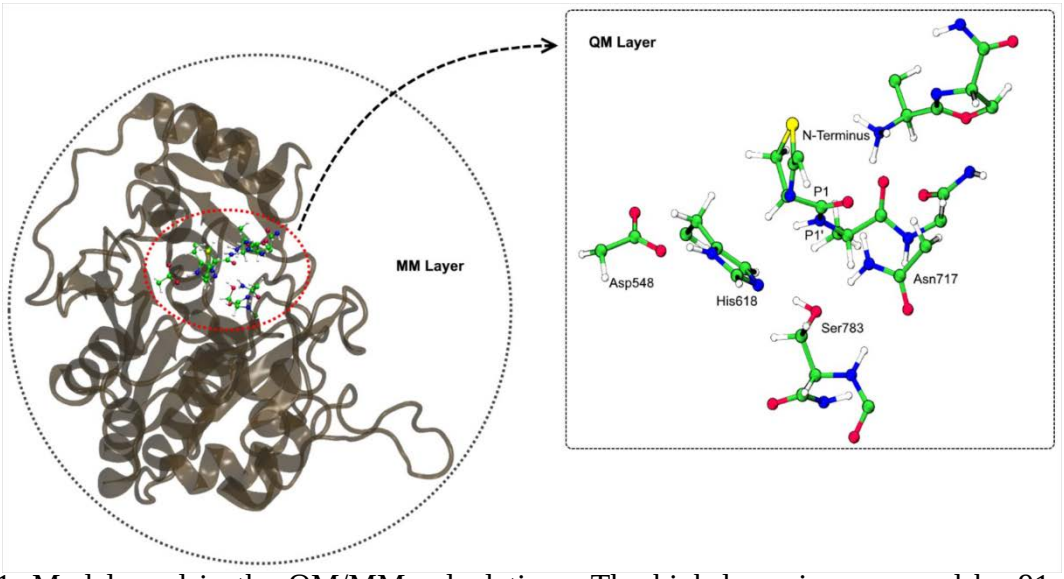


Figure 1. Model used in the QM/MM calculations. The high layer is composed by 91 atoms, which includes the residues of the catalytic triad (Ser783, His618 (side chain until Cα) and Asp548 (side chain until Cβ)), Asn717 (side chain until Cγ) and by P1(Cys), P1'(Ala), P2'(Tyr) residues and the terminal Ile of the substrate.

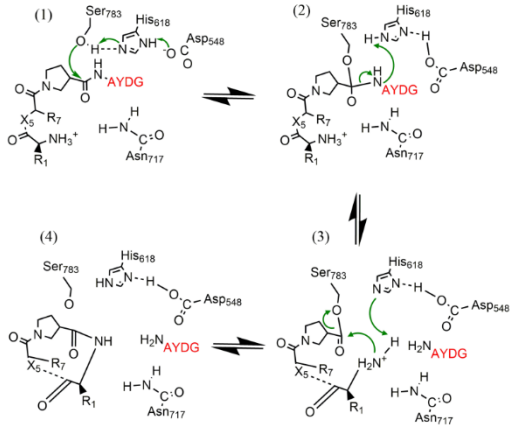


Figure 2. Representation of the catalytic mechanism proposed for PatGmac.[9]

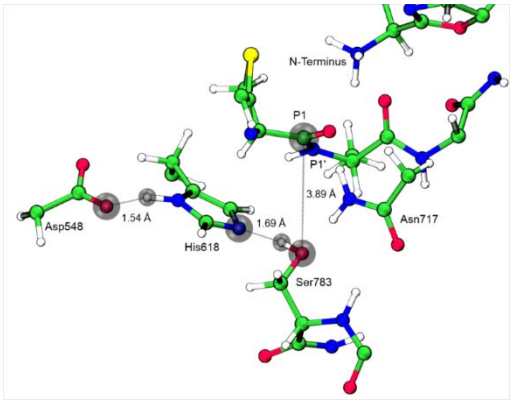


Figure 3. Structure of the reactants highlighting the most relevant distances for the first mechanistic step.

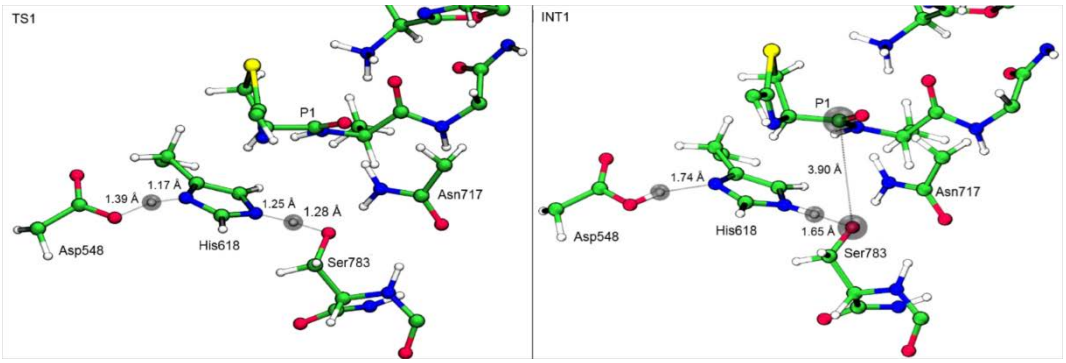


Figure 4. Representation of the structures of the first transition state (TS1) and the intermediate (INT1) on the first step.

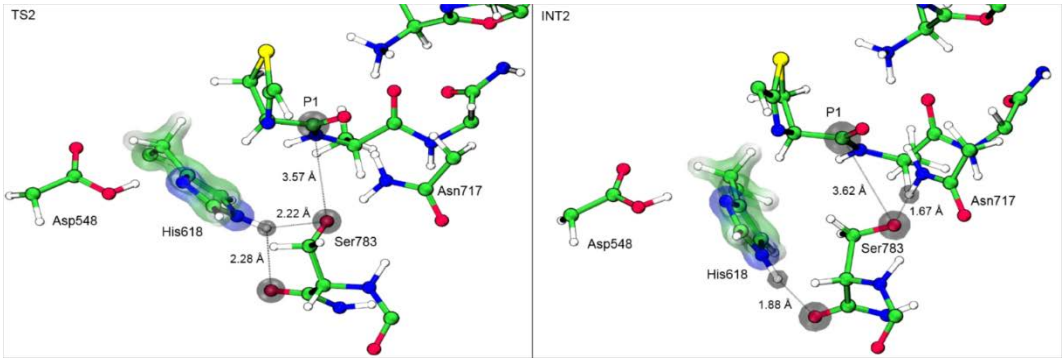


Figure 5. Representation of the structures of the transition state (TS2) and the intermediate (INT2) of the second step of the macrocyclization reaction

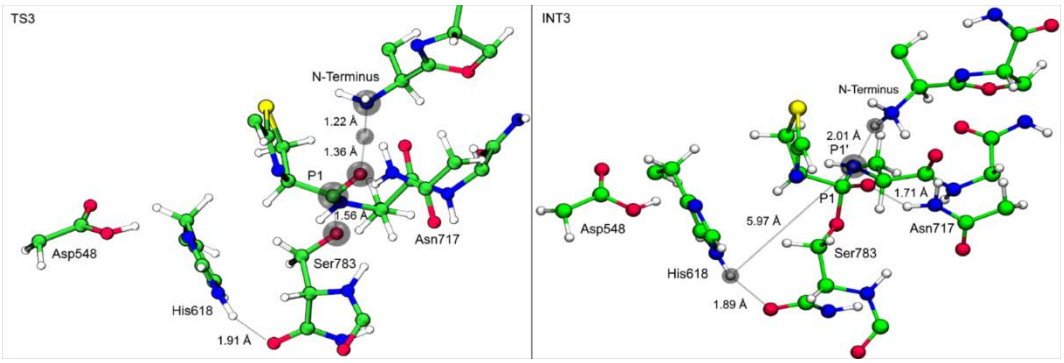


Figure 6. Representation of the structures of the transition state (TS3) and the intermediate (INT3) of the third step of the reaction.

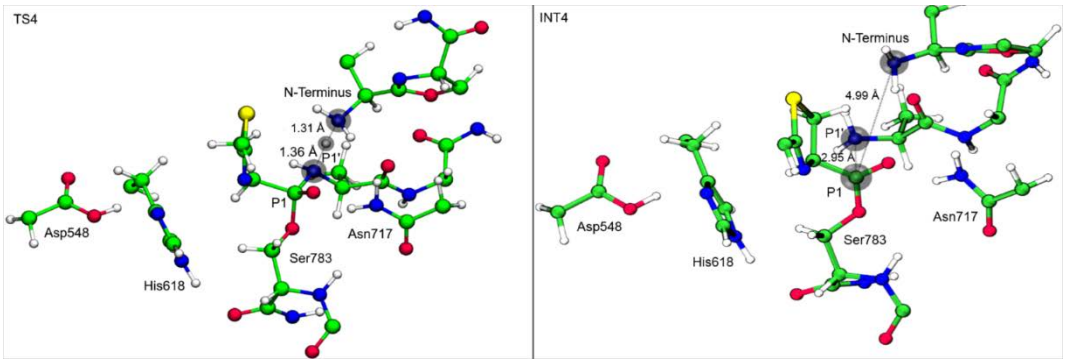


Figure 7. Representation of the structures of the transition state (TS4) and the intermediate (INT4) of the fourth step of the reaction

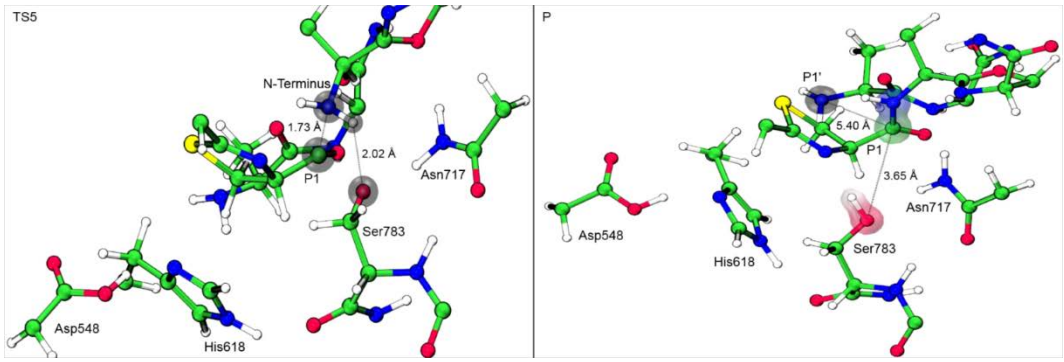


Figure 8. Representation of the structures of the fifth transition state (TS5) and products of the fifth step.

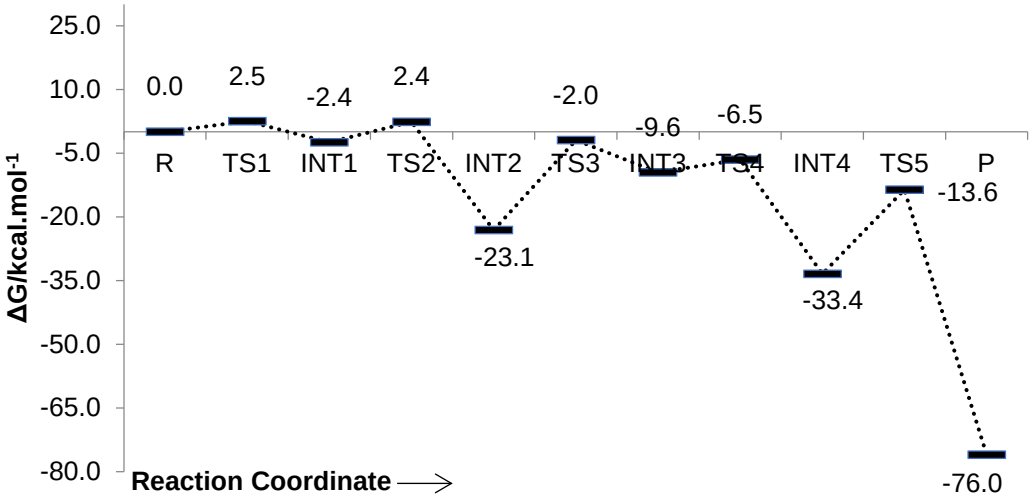


Figure 9. Potential energy surface (PES) for the macrocyclization reaction catalysed by PatGmac. The energies were obtained at ONIOM(M06/6-