Organic & Biomolecular Chemistry

Dynamic Article Links

Cite this: *Org. Biomol. Chem.*, 2012, **10**, 1207

www.rsc.org/obc

PAPER

Mechanism of C-terminal intein cleavage in protein splicing from QM/MM molecular dynamics simulations†

Jon I. Mujika,*[a] Xabier Lopez[a] and Adrian J. Mulholland[b]

Received 23rd August 2011, Accepted 31st October 2011
DOI: 10.1039/c1ob06444d

Protein splicing is a post-translational process in which a biologically inactive protein is activated by the release of a segment denoted as an intein. The process involves four steps. In the third, the scission of the intein takes place after the cyclization of the last amino acid of the segment, an asparagine. Little is known about the chemical reaction necessary for this cyclization. Experiments demonstrate that two histidines (the penultimate amino acid of the intein, and a histidine located 10 amino acids upstream) are relevant in the cyclization of the asparagine. We have investigated the mechanism and determinants of reaction in the GyrA intein focusing on the requirements for asparagine activation for its cyclization. First, the influence that the protonation states of these two histidines have on the orientation of the asparagine side chain is investigated by means of molecular dynamics simulation. Molecular dynamics simulations using the CHARMM27 force field were carried out on the three possible protonation states for each of these two histidines. The results indicate that the only protonation state in which the conformation of the system is suitable for cyclization is when the penultimate histidine is fully protonated (positively charged), and the upstream histidine is in the His_ε neutral tautomeric form. The free energy profile for the reaction in which the asparagine is activated by a proton transfer to the upstream histidine is presented, computed by hybrid quantum mechanics/molecular mechanics (QM/MM) umbrella sampling molecular dynamics at the SCCDFTB/CHARMM27 level of theory. The calculated free energy barrier for the reaction is 19.0 kcal mol^{-1}. B3LYP/6-31+G(d) QM/MM single-point calculations give a qualitatively a similar energy profile, although with somewhat higher energy barriers, in good agreement with the value derived from experiment of 25 kcal mol^{-1} at 60 °C. QM/MM molecular dynamics simulations of the reactant, activated reactant and intermediate states highlight the importance of the Arg181-Val182-Asp183 segment in catalysing the reaction. Overall, the results indicate that nucleophilic activation of the asparagine for its cyclization by the upstream histidine acting as the base is a plausible mechanism for the C-terminal cleavage in protein splicing.

Introduction

Many proteins are synthesized in biologically inactive forms, and are activated in post-translational processes such as proteolytic cleavage. This process is usually catalyzed by external proteins, but some proteins are able to self-catalyze the reactions without the need for any other protein or cofactor: protein splicing; the autoprocessing of *hedgehog* proteins; the autocleavage of amidohydrolase precursors; and the formation of pyruvoyl enzymes are examples.[1] All these processes are initiated by a N–S or N–O acyl rearrangement of a peptide bond involving the amino group of cysteine, serine, or threonine residue.[1] In protein splicing, a segment of an inactive protein, the intein (internal protein), is excised from the rest of the protein, and the two flanking domains, the C- and N-exteins, join each other, forming a biologically active protein.[2,3] An analogous process is RNA splicing, in which an intervening sequence (the 'intron') is excised and the flanking regions (the 'exons') are joined in order to yield a functional messenger RNA molecule. Protein splicing is also an important tool in biotechnology. The inteins (or introns) may harbour not only the splicing domain, but also the homing endonuclease domain, the DNA sequence of which ensures the mobility of the intein gene within the host gene.[4] In some cases, the intein is the product of two different genes (split inteins). The first evidence of protein splicing was found in 1990 in the *Saccharomyces cerevisiae* VMA1 gene product.[5,6]

[a]*Kimika Fakultatea, Euskal Herriko Unibertsitatea and Donostia International Physics Center (DIPC, P. K. 1072, 20080, Donostia, Euskadi, Spain. E-mail: joni.mujika@ehu.es; Fax: 0034943015341*

[b]*Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol, UK, BS8 1TS. E-mail: Adrian.Mulholland@bristol.ac.uk; Fax: +44 (0)117 925 1295*

† Electronic supplementary information (ESI) available: RMSD of MM and QM/MM molecular dynamics; trends of critical distances during MM molecular dynamics; harmonic restraints applying during MM molecular dynamics; SCCDFTB/CHARMM27 and B3LYP/6-31+G(d) potential energy surfaces calculated for SCCDFTB/CHARMM27 structures; and two dimensional potential energy surfaces characterized with SCCDFTB/CHARMM. See DOI: 10.1039/c1ob06444d

Inteins can be classified based on either the domains they harbour or their sequence. According to the first type of classification, three classes of natural inteins with splicing activity have been identified:[7] maxi-inteins (harbouring both splicing and homing endonuclease domains), minimal or mini-inteins (only the splicing domain is present) and trans-splicing inteins (the protein junction belongs to different proteins). Alternatively, classification based on sequence leads to three classes of inteins:[8,9] the standard Class 1 inteins present a Cys, Ser or Thr as the first residue at their N-terminal, while in Class 2 and Class 3 inteins this residue is an alanine. In addition, Class 3 inteins contain a WCT triplet.[8,9]

The widespread importance of inteins, in biology and biotechnology, makes understanding of the splicing mechanism important. *In vitro* observations,[10,11] in which the intein was isolated and inserted between two foreign proteins, have elucidated the main features of the mechanism of protein splicing: (i) only amino acids located in the intein (and the first residue of the C-extein) are directly involved in the process; (ii) it is an intramolecular process; (iii) it requires no other proteins or external sources of metabolic energy such as ATP. Moreover, sequence analysis of known inteins reveals the conservation of several blocks of amino acids, such as the A and B blocks located in the N-terminal splicing domain, the C, D, E and H blocks in the endonuclease domain, and the F and G blocks in the C-terminal splicing domain.[7] Among the conserved amino acids, of particular importance are the first amino acids of the Class I intein and of the C-extein, which have a thiol or hydroxyl side chain (Cys, Ser or Thr), and the last amino acid of the intein, which is an asparagine or glutamine (Asn or Gln). The generally accepted overall mechanism (see Fig. 1) in standard Class I inteins is composed of four steps (the Class 2 and Class 3 inteins follow a different mechanism):[9,12] (i) N–O or N–S acyl rearrangement to form a linear ester intermediate; (ii) transesterification to form a branched ester intermediate; (iii) cyclization of an Asn (or Gln), coupled with the cleavage of the intein–C-extein peptide bond, forming an aminosuccinimide residue and an ester; and (iv) spontaneous hydrolysis of the aminosuccinimide residue and S–N or O–N rearrangement of the ester between the exteins to form a more stable amide (peptide) bond. Thus, the first two steps involve chemical reactions to join the two extein segments, and it is in the third step when the intein is released. The kinetics of these steps has been determined individually for the *Pyrococcus abyssi* PoIII intein. In the same study, it was shown that the C-terminal scission can take place in a broad range of pH values.[13] More specifically, Wood *et al.* showed that the rate of C-terminal cleavage increases on going from pH 7.5 to 6.0.[14,15] Gangopadhyay *et al.* inserted the RecA intein of *Mycobacterium tuberculosis* into a green fluorescent protein (GFP) to study the process.[16] They took advantage of the fact that the GFP-intein fusion protein formed a fluorescent chromophore. They found that the optimum pH for the entire protein splicing process is 6.5.[16]

Although the overall mechanism is well established, understanding of the chemical reactions involved is still limited. In the third step, for instance, it is known that amino acids located in block G are involved, that is, the last two amino acids of the intein (asparagine preceded by an histidine) and the first C-extein amino acid (Ser, Thr or Cys) and that the C-terminal cleavage involves the cyclization of the asparagine. However, how the protein catalyzes the cyclization of the asparagine remains unclear. Two alternative reaction pathways can potentially lead to asparagine cyclization: deamidation or cleavage.[1] In deamidation, the backbone amide nitrogen atom attacks the side chain carbonyl carbon, while in the cleavage mechanism, the amide side chain attacks the backbone carbonyl carbon atom. Several experiments have revealed a C-terminal aminosuccinimide residue after the third step,[17,18] suggesting that in protein splicing, the asparagine cyclization follows the cleavage mechanism. In addition, some mutagenesis experiments suggest that the penultimate histidine plays a role in asparagine cyclization,[19,20] although absence of this amino acid gives contradictory results: while in some studies this residue was found to be indispensable for the process,[21] in others the protein spliced.[20,22,23] A crystal structure of the pre-splicing state of the GyrA intein from *Mycobacterium xenopi* indicates that this histidine may be hydrogen-bonded to the main chain carbonyl oxygen of the crucial asparagine (or glutamine) residue,[24] and the authors suggested that it could also protonate the scissile peptide bond. On the other hand, amino acids located in block F have been proposed to be essential for the C-terminal cleavage of the intein.[25] A histidine located in this block is highly conserved in intein sequences,[26] and its mutation prevents the formation of the splicing products,[27,28] which may indicate that this histidine is involved in catalyzing the asparagine cyclization.

Molecular modelling and simulation provide good routes to analyse mechanisms of chemical reactions in proteins.[29] The two reaction mechanisms that lead to Asn cyclization, *i.e.* deamidation or cleavage, have been studied computationally using small models.[30–35] Ruiz-Lopez *et al.*[33,34] investigated both mechanisms by gas-phase optimization of a 26-atom model at the B3LYP/6-31+G(d,p) level, with solvent effects studied by single-point calculations at B3LYP/6-31++G(d,p) with the IEF-PCM (polarizable continuum model) method. They found that the deamidation mechanism has a lower activation energy than the cleavage reaction,[34] although the difference is not very significant.

Two alternative mechanisms for the N- and C-terminal cleavages (steps 1 and 3 in Fig. 1) have previously been studied computationally. On one hand, Shemella *et al.*[36] proposed a mechanism for low pHs where the reaction is activated upon an initial protonation of the peptide bond N atom by a hydroxonium. They employed a cluster model formed of 25 atoms to study the C-terminal cleavage, using the PM3 semiempirical method to optimize the structure and the B3LYP DFT functional, including thermal and entropic contributions from harmonic frequencies and solvent effects by polarizable continuum model, to calculate the free energies in solution. With this scheme, a free energy barrier of 25 kcal mol^{-1} was obtained for the asparagine cyclization. A similar activation mechanism was proposed by Johansson *et al.*[37,38] for the N–O acyl shift reaction in SEA, a multidomain protein that undergoes an autoproteolysis reaction. This reaction is analogous to the N-terminal cleavage in protein splicing. Molecular dynamics simulations of SEA indicated[37] that the secondary structure of the protein observed experimentally is better described when the uncleaved peptide bond is in the *cis* conformation instead of the more stable *trans* conformation. Then, the nucleophilic attack of the serine side chain onto the protonated peptide bond was studied using the Gly-Ser dipeptide as a model.[38] The calculations were carried out combining the B3LYP functional and the charge polarizable continuum model (CPCM). The results suggested that the peptide bond strain could make the peptide bond N atom more labile to be protonated; once protonated, a lower activation energy for the

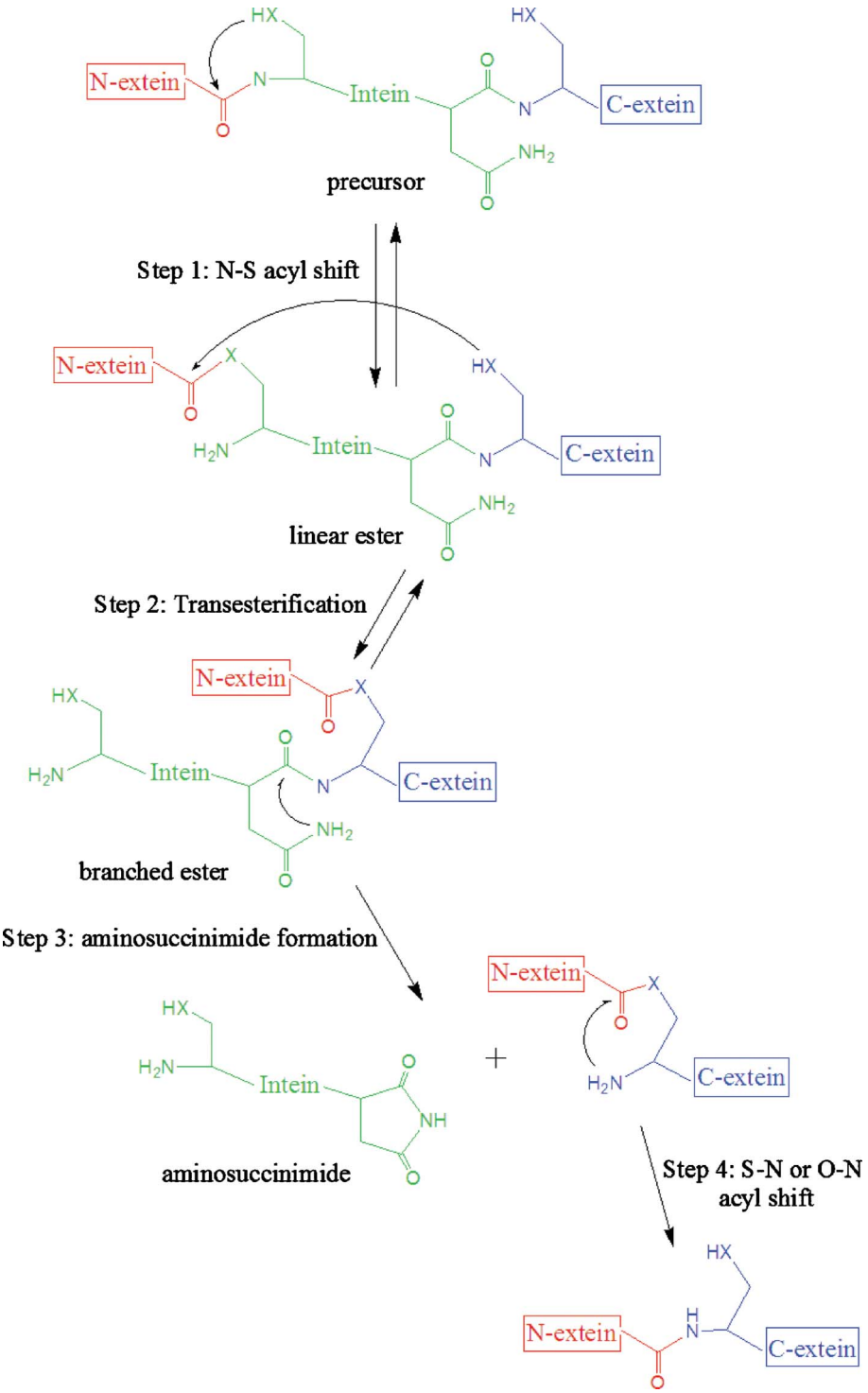

Fig. 1 Schematic illustration of proposed four steps for protein splicing.

nucleophilic attack was found, although a high energy barrier (40 kcal mol^{-1}) was obtained with the cluster model employed.

A different mode of activation of asparagine cyclization was proposed by Ding *et al.*[27] Based on the crystal structure of a mini-intein derived from the intein encoded by the DNA helicase gene of the cyanobacterium *Synechocystis* sp. PCC6803, those authors proposed a mechanism for asparagine cyclization in which the histidine located in block F acts as a general base, accepting a proton from the asparagine side chain and thus activating its cyclization. We have previously investigated this mechanism in quantum chemical calculations on small models.[39] We used density functional theory (with the B3LYP functional) in conjunction with a continuum (PCM) treatment of (aqueous) solution, to characterize structurally and energetically the main stationary points along this reaction pathway. We found that the most efficient way to activate the Asn side chain is by eliminating a

proton from its side chain N_δ atom. A reaction pathway was characterized in the gas phase, in which a proton of the asparagine was transferred to the upstream histidine, activating its cyclization. In addition, the penultimate histidine was found to stabilize the oxyanion intermediate and later to protonate the scissile peptide bond. The cyclization of the activated Asn was found to be the rate-limiting step, with a free energy barrier of 29.3 kcal mol^{-1} in solution. At around the same time, a study combining NMR pK_a measurement and quantum calculations was published by Du *et al.*[28] First, the role of a histidine located in block B and involved in the N-terminal splicing was analyzed. The pK_a value of this residue was experimentally measured in the *Mycobacterium tuberculosis* RecA intein. A significant pK_a shift was quantified, going to a pK_a values of 7.3 ± 0.6 in the splicing precursor to <3.5 in the product. Based on these results, calculations with ONIOM (inner part described at the B3LYP/6-31G(d,p) level and the outer part with the AMBER force field) were employed to characterize a reaction pathway in which the B-block histidine acts as the general base activating the serine side chain for its nucleophilic attack onto the scissile N-terminal peptide bond. A similar reaction mechanism was suggested by the authors for the C-terminal cleavage, in which the F-block histidine would activate the side chain of the asparagine. This suggestion was supported by a high pK_a value of this histidine measured experimentally in this work and by the fact that the mutation of this amino acid caused a reduction in the amount of C-terminal cleavage compared with wild type.

Thus, these studies suggest that the Asn side chain activation for its cyclization by upstream histidine seems a plausible mechanism, but a further investigation is required. In particular, little is know about the protonation states of the two histidines (the upstream histidine located in block F and the penultimate histidine of the intein in block G) and their role during the Asn cyclization. The pK_a values of these two residues are very close to physiological pH (pK_a values of 6.6 and 6.2 were predicted with the PROPKA method[40] at the 1AM2 crystal structure), and therefore little information about their protonation states can be extracted from them. In the present study, we analyze the dynamics of the GyrA intein, with the aim of investigating which is the most likely protonation states of these two residues. First, nine different molecular dynamics simulations of 2 ns production each were carried out considering the three possible protonation states of the two histidines. Their structural analysis allows the determination of the most suitable system for the cyclization of Asn. In the second part of the work, departing from a representative geometry, we determined the free energy profile for the nucleophilic activation of the Asn side chain by F-block histidine and subsequent cyclization of Asn. Finally, the main stationary points located in the free energy profile are further investigated by QM/MM molecular dynamics simulations, to analyse the role of the intein. These results, which are consistent with experimental findings, give a structurally detailed picture of the asparagine cyclization reaction in the protein, and shed light on the protonation states of the two histidines required for splicing activity.

Methodology

System preparation

The X-ray crystal structure of GyrA intein from *Mycobaterium xenopi* (2.20 Å resolution, PDB code 1AM2) was chosen as the starting structure for simulations here.[24] In order to model the Intein–C-extein peptide bond, a serine was manually inserted as the first amino acid of the C-extein. Hydrogen atoms were added with the HBUILD routine in CHARMM. Conventional protonation states at pH = 7 were assigned to all amino acids, except for His187 and His197, for which various protonation states were tested (see below). A neutral tautomer was assigned for the remaining histidine residues. The total charge of the system without the charge of these two histidines was 0. All protein residues and crystal water molecules containing at least one atom within 25 Å of the system centre (C atom of Asn198) were included in the model. The system was solvated by superimposing a 25 Å radius sphere of pre-equilibrated CHARMM-type TIP3P (TIPS3P) water molecules.[41,42] Any water molecule with its oxygen atom within 2.6 Å of a non-hydrogen atom was deleted. The positions of the added water molecules were optimized by energy minimization (500 steps steepest descent (SD), 1500 steps Adopted Basis Newton Raphson (ABNR)), keeping all other atoms fixed. In order to ensure full hydration, a second solvation of the system was performed and the water molecules were again subject to energy minimization, followed by 20 ps of Langevin dynamics (at 300 K) for all water molecules, keeping again other atoms frozen. Energy minimization with all non-water atoms fixed was then carried out first with SD (500 steps) and then with 2500 steps of Adopted Basis Newton–Raphson (ABNR). Finally, the full system, without fixing any atoms, was fully minimized (1500 steps SD, 5000 steps ABNR). A gradient tolerance convergence criterion of 0.01 kcal mol^{-1} was used. A non-bonded cut-off of 13 Å was defined.

Molecular dynamics simulations

The CHARMM program (version c31b1)[43,44] was employed to carry out the MM molecular dynamics simulations using the CHARMM27 all-atom force field.[45] Stochastic boundary conditions,[46] which have provided satisfactory results in previous studies of other proteins,[47,48] were applied by dividing the system into two regions: the reaction region, consisting of all atoms within 21 Å from the system centre, and the buffer region, containing all other atoms in the 25 Å sphere. Charged groups in the buffer region were made artificially neutral (retaining atomic charges designed to show realistic H-bonding behaviour).[49] As is standard in the stochastic boundary molecular dynamics approach, the reaction region was treated with Newtonian dynamics and the buffer region with Langevin dynamics, at a temperature of 300 K. A deformable boundary potential[46] was imposed on the water oxygen atoms to prevent them from diffusing away. In the buffer region, the Langevin equations of motion impose a friction coefficient and a random force on all heavy atoms, which allows the buffer region to act as a simplified heat bath. Friction coefficients of 250 ps^{-1} for non-hydrogen protein atoms and 62 ps^{-1} on water oxygen atoms were used.[50] Moreover, non-solvent heavy atoms were harmonically restrained to their crystal coordinates with force constants based on model average β factors.[50] These restraints were scaled (linearly and in four steps) from zero at 21 Å from the centre of the system, to a maximum at 25 Å.[51] The force constants defined for these harmonic restraints are presented in Table S1 of the Supporting Information.† The SHAKE algorithm[52] was applied to fix the length of all MM bonds involving hydrogen atoms, and a 1 fs time step was used. The system was initially

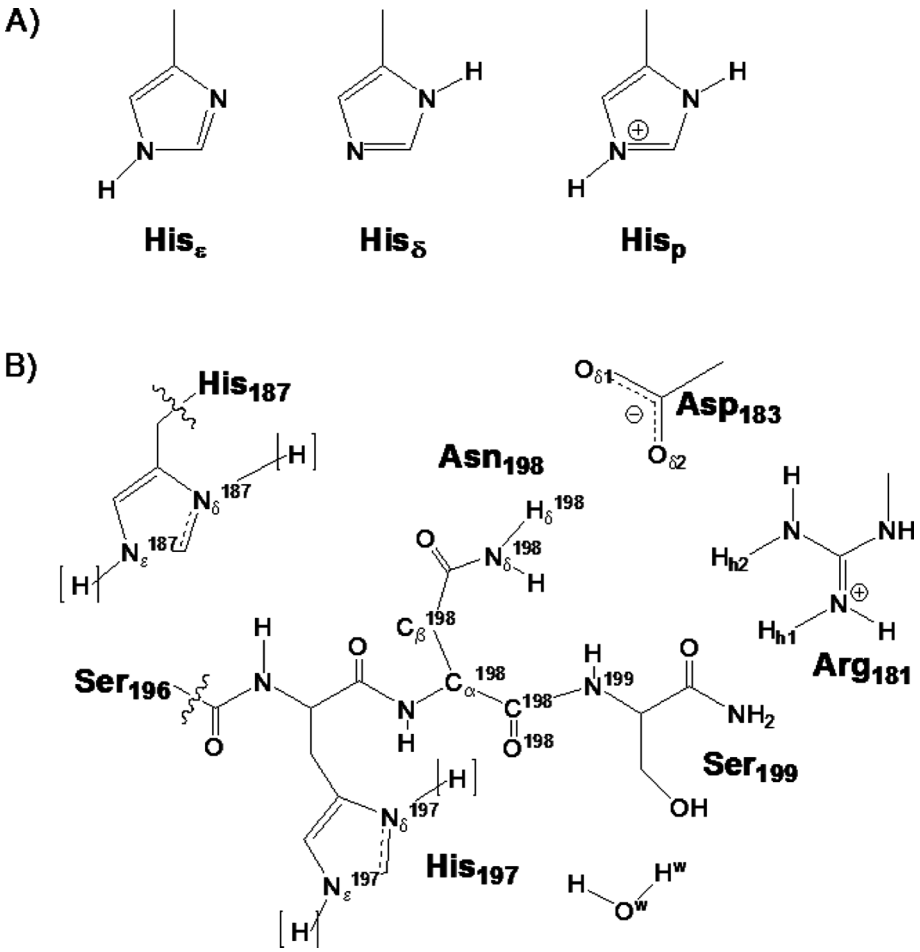

Fig. 2 Above, the three possible protonation states (neutral tautomers and protonated form) of a histidine side chain. Below, the atom labelling used in the text. The wavy lines indicate the position of the two link atoms in QM/MM calculations.

heated from 10 to 300 K in three 20 ps intervals: first from 10 to 100 K, then from 100 to 200 K, and, finally, from 200 to 300 K. In the next stage, the whole system was equilibrated for 100 ps. 2 ns of production were carried out for each of the nine simulations performed.

The histidine amino acid side chain has three likely protonation states (see Fig. 2A): two tautomers for the neutral state (His_ε and His_δ) and one positively charged (His_p). Nine different molecular dynamics simulations were carried out considering each possible protonation state for the two histidines studied. The different simulations are terms as MD_{XY}, where X and Y stand for the His187 and His197 protonation states following the terminology presented in Fig. 2A, that is E, D or P for His_ε, His_δ and His_p, respectively. For instance, in the MD_{DE} systems His187 is present in the His_δ form, and His197 in the His_ε protonation state. The labelling of atoms involved in the third step of protein splicing is presented in Fig. 2B.

Potential of mean force and QM/MM MD calculations

The free energy profile of the Asn198 cyclization reaction was determined by computing the potential of mean force (PMF) by means of umbrella sampling[53] QM/MM[54] molecular dynamics simulations at 300 K. This technique allows the sampling of many conformations during the reaction, for which a reaction coordinate must be defined.[51] At each value of this reaction coordinate a sampling window is created, from which an appropriate analysis tool calculates the free energy profile. A representative snapshot of MD_{EP} simulation was chosen as the initial geometry, and based on previous calculations on this reaction in gas phase, two steps were characterized:[37,39] (i) activation of Asn198 side chain by a proton transfer to His187; and (ii) Asn198 cyclization.

The CHARMM program[41,42] was used for the umbrella sampling MD calculations. The QM part was treated with the self-consistent charge-density functional tight binding (SCC-DFTB) method[55] and included the side chains of His187, all atoms of Asn198 and Ser199, and the C and O atoms of His197 (see Fig. 2B). In total, 43 atoms were treated as QM atoms. In order to fulfil the bonding requirements of the QM system, where a covalent bond separated the QM and MM regions, a hydrogen 'link atom' was added as the boundary of the QM region.[56,57] Two link atoms were introduced: between the C_α and C_β atoms of His187 and between the C and C_α atoms of Ser196. Note that due to the particular nature of the system in which the intein itself is the "substrate" of the reaction, it was necessary to place a link atom at the backbone of the intein. Of the three type of bonds found in a protein backbone, the bond formed by the C_α and carbonyl C atoms was chosen as the most suitable one to place a link atom.[57]

The QM/MM nonbonded interactions between the link atom and the host group of the first MM atom bound to the link atom were excluded by the use of EXGR keyword. This scheme has been found to give reliable results for enzymatic reactions.[58] All other atoms not included into the QM region were treated by molecular mechanics (MM) using the CHARMM27 all-atom force field.[45]

In the first step, the proton transfer from Asn198 side chain to His187, the reaction coordinate was defined as the $r(N_\delta^{198}–H_\delta^{198})–r(H_\delta^{198}–N_\varepsilon^{187})$ distance combination, while in the second step, the $(N_\delta^{198}–C^{198})$ distance was chosen as the reaction coordinate for asparagine cyclization. In order to confirm that the reaction is well reproduced considering two reaction steps, the same two reaction coordinates were combined to characterize the three-dimensional potential energy surface of the reaction (included in the Supporting Information†). The surface shows that the reaction proceeds in a stepwise mechanism. Moreover, similar reaction coordinates have been applied successfully for similar processes in various other enzymes.[59,60] A harmonic restraint of 100 kcal mol^{-1} $Å^{-2}$ was applied to both reaction coordinates. Each window of the simulation was sampled with 10 ps of equilibration followed by 30 ps of production dynamics, analogous to other invesigations.[61] The weighted histogram analysis method (WHAM)[62] was used to compute the free energy profile.

Results

Influence of His187 and His197 protonation states

All possible combinations of protonation states of His187 and His197 were modelled in order to evaluate the influence of their protonation states on the conformation adopted by the intein. In total, nine molecular dynamics simulations were carried out. The time evolutions of the root-mean square deviation (RMSD) are presented in Fig. S1 of the Supporting Information.† The RMSDs of the nine simulations were calculated taking the initial X-ray structure as the reference and including only the heavy atoms of the intein. The results indicate that the nine models are in a reasonably well-equilibrated state during the analyzed trajectories (2 ns of production each). The evolutions of the RMSD for the nine MD simulations were similar, the final RMSD value being below 2 Å in all nine simulations.

As explained in the Introduction, the third step of protein splicing involves the release of the intein due to the cyclization of the last amino acid of the intein, Asn198. In this section, the influence of the protonation state of His187 and His197 on the conformation adopted by Asn198 is analysed, as well as the interaction of this residue with the two histidines. Therefore, four geometrical parameters were monitored during the nine MD simulations: (i) distance between the N_δ^{198} and C^{198} atoms; (ii) the dihedral angle of the Asn198 side chain ($N_\delta^{198}–C_\beta^{198}–C_\alpha^{198}–C^{198}$ angle); (iii) the distance between N_δ^{198} and His187; and (iv) the distance between N_δ^{197} and O^{198}. The mean values for these parameters are collected in Table 1 and their trends illustrated in the Supplementary Information.†

Position and interactions of Asn198. The mean values of the $N_\delta^{198}–C^{198}$ distance are presented in Table 1, and their trends illustrated in Fig. S2.† For seven of the nine simulations, the average values are around or above 4.5 Å. In contrast, for the remaining two systems the average distance is shorter than 4 Å. These two systems are MD_{EP} and MD_{ED}, where the mean values of the $N_\delta^{198}–C^{198}$ distance are 3.86 ± 0.32 Å and 3.69 ± 0.38 Å, respectively. Note that in these two systems the protonation state of His187 is the same (His_ε), and that in both systems the N_δ atom of His197 is protonated. They differ at N_ε^{197}, which is protonated in MD_{EP} and unprotonated in MD_{ED}. During these two simulations, after a few picoseconds of production where the $N_\delta^{198}–C^{198}$ distance is similar to the remaining systems and therefore lies around 5 Å, there is a conformational change that decreases the distance to a range of 3.5–4 Å. In the case of MD_{ED}, this distance remains approximately constant until the end of the simulation, while in MD_{EP} there is another small change after 1 ns of simulation.

The Asn198 side-chain position in the MD_{ED} and MD_{EP} systems is notably different from the other seven simulations. This is confirmed by analysis of the $N_\delta^{198}–C_\beta^{198}–C_\alpha^{198}–C^{198}$ dihedral angle of Asn198. This dihedral angle is a good indicator of the Asn198 side chain orientation with respect to its peptide bond. A value close to 0° means that the N_δ^{198} atom is above the C^{198} atom and therefore the Asn198 side chain is in a good position for the nucleophilic attack and concomitant cyclization; on the other hand, a value close to 180° means that the N_δ^{198} is orientated

Table 1 Average values and standard deviation (in brackets) of geometrical parameters relevant for the Asn198 cyclization reaction, from MM (CHARMM27) MD simulations of the nine systems (2 ns of production MD each). These different systems represent the three possible protonation states of His187 and His197 (see Fig. 2A). The values in the crystal structure (PDB code 1AM2)[24] are also shown. All distances are in Å and angles in degrees. Atom labels are shown in Fig. 2B

			Distances			Angles	
Simulation	His^{187}	His^{197}	$N_\delta^{198}–C^{198}$	$N_\delta^{198}–N^{187}$[a]	$N_\delta^{197}–O^{198}$	$N_\delta^{198}–C_\beta^{198}–C_\alpha^{198}–C^{198}$	$N_\delta^{197}–H_\delta^{197}–O^{198}$
MD_{PP}	His_P	His_P	4.72 (0.22)	—	3.42 (0.56)	163.7 (18.0)	111.5 (27.2)
MD_{DP}	His_δ	His_P	4.86 (0.08)	5.43 (0.47)	3.10 (0.63)	171.5 (15.4)	136.8 (27.4)
MD_{EP}	His_ε	His_P	3.86 (0.32)	3.76 (1.29)	3.09 (0.25)	44.5 (33.9)	102.9 (23.3)
MD_{PD}	His_P	His_δ	4.47 (0.14)	—	3.20 (0.48)	160.5 (11.0)	126.1 (31.8)
MD_{DD}	His_δ	His_δ	4.32 (0.42)	8.00 (0.92)	3.96 (0.58)	141.5 (52.5)	90.9 (21.7)
MD_{ED}	His_ε	His_δ	3.69 (0.38)	6.05 (1.09)	3.48 (0.38)	76.8 (32.0)	81.7 (23.9)
MD_{PE}	His_P	His_ε	4.51 (0.37)	—	4.14 (0.69)	168.9 (56.6)	—
MD_{DE}	His_δ	His_ε	4.53 (0.23)	7.03 (0.93)	3.61 (0.34)	166.2 (11.4)	—
MD_{EE}	His_ε	His_ε	4.45 (0.14)	6.56 (0.57)	3.60 (0.41)	163.8 (15.6)	—
1AM2	—	—	4.76	2.74	2.90	165.8	—

[a] Distance between N_δ^{198} and the closest unprotonated N atom of His187.

opposite to C^{198} disfavouring the cyclization. The trends of this angle in the simulations are presented in Fig. S3.†

A direct correlation between the N_δ^{198}–C^{198} distance and the N_δ^{198}–C_β^{198}–C_α^{198}–C^{198} dihedral angle is observed (see Fig. S2 and S3†): the larger the angle, the larger the distance. Thus, in all systems except MD_{EP} and MD_{ED} the mean values of this dihedral angle are similar, being in the range of 140–170°. That means that in all these systems the Asn198 side chain is orientated opposite to the scissile peptide bond. In the other two systems, the average values of the dihedral angle are 44.5 ± 33.9° for MD_{ED} and 76.8 ± 32.0° for MD_{EP}. It can be seen in Fig. S3† that the trend of this dihedral angle in both simulations correlates with the N_δ^{198}–C^{198} distance, and after the conformational change that takes place after the first few picoseconds in both simulations, the value of this angle decreases to less than 40°. However, as was the case for the N_δ^{198}–C^{198} distance, while in MD_{ED} the value of the angle remains constant until the end, during the MD_{EP} simulation there is a conformational change after 1 ns of production where the dihedral angle increases to *ca.* 100°, that is, the Asn198 side chain is finally positioned perpendicular to the plane formed by the other three atoms.

In our previous work, a reaction pathway in which His187 activates the Asn198 side chain for cyclization was characterized, in a small model.[39] This mechanism was previously proposed[27,28] and involves a proton transfer from N_δ^{198} to His187, which implies that at least one of the N atoms of His187 imidazole group must be unprotonated. In order to investigate if this reaction pathway is feasible, the distance between the unprotonated N atom of His187 and N_δ^{198} were measured in the six simulations in which His187 presents an unprotonated N atom (His_δ or His_ε protonation states). The evolution of this distance is presented in Fig. S4.†

The average values presented in Table 1 show that in five of the six simulations this distance is above 6 Å, and only in the MD_{EP} system is this value below 4 Å, namely 3.76 ± 1.29 Å. Moreover, the trend of this distance is equilibrated during all the simulation (see Fig. S4†). These results confirm that the Asn198 side chain activation by His187 is geometrically possible, but it is only feasible when His187 and His197 are in the His_ε and His_p protonation states, respectively. On the other hand, in the other system where the N_δ^{198}–C^{198} distance was below 4 Å and therefore in a good position for the Asn cyclization, the MD_{ED} system, the average value of the N_δ^{198}–N_δ^{187} distance is significantly larger (6.05 ± 1.09 Å) and it seems unlikely that His187 can act as a base to accept a proton from the Asn198 side chain with this protonation state.

Position of His197. The role of His197 in splicing is still under debate. X-Ray crystal structures demonstrate[24,27] that this amino acid may form a hydrogen bond with the scissile peptide bond O atom through the protonated N_δ^{197} atom. For instance, in the 1AM2 crystal structure this distance is 2.9 Å.[24] The N_δ^{197}–O^{198} distance was measured during the MD simulations, and their trends are presented in Fig. S5.† The average value of the N_δ^{197}–O^{198} distance is in the range of 3–4 Å for all the systems. As expected, the highest values correspond to the simulations where His197 is present in the His_ε protonation state, since the N atom (N_δ^{197}) closest to O^{198} is unprotonated and therefore is unable to form a hydrogen bond with the peptide bond O atom. In contrast, when His197 is fully protonated (His_p protonation state), the mean value of the N_δ^{197}–O^{198} distance is shorter (around 3.1 Å, except for the MD_{PP} system (3.42 ± 0.56 Å)). Interestingly, the MD_{EP} system shows one of the lowest mean values for this distance.

The analysis of the trajectories indicates that during all the simulations where N_δ^{197} is protonated (His_p and His_δ protonation states) the N_δ^{197}–H_δ^{197}–O^{198} angle is in the range of 100–140 degrees. In the case of MD_{EP}, the average value of this angle is 102.9 ± 23.3°. These values are typical for a carbonyl group forming two hydrogen bond interactions.[47] During the simulations, the carbonyl group of Asn198 interacts with His197 and with a solvent water molecule.

Relevance to the splicing reaction. All the results presented above lead to the conclusion that the MD_{EP} system has the most suitable conformation for Asn198 cyclization in protein splicing. Of the nine systems studied, only during the MD_{EP} simulation does the Asn198 side-chain adopt the right orientation for its cyclization, as indicated by the low average value computed for the N_δ^{198}–C_β^{198}–C_α^{198}–C^{198} dihedral angle. This good disposition is also reflected by the shorter N_δ^{198}–C^{198} distance, less than 4.0 Å. A representative snapshot from this simulation is illustrated in Fig. 3A, showing this good, reactive conformation. On the other hand, for the 7 of the remaining 8 systems the Asn198 side chain is orientated in the opposite direction (as an example, a snapshot of the MD_{PP} simulation is shown in Fig. 3B), which would make

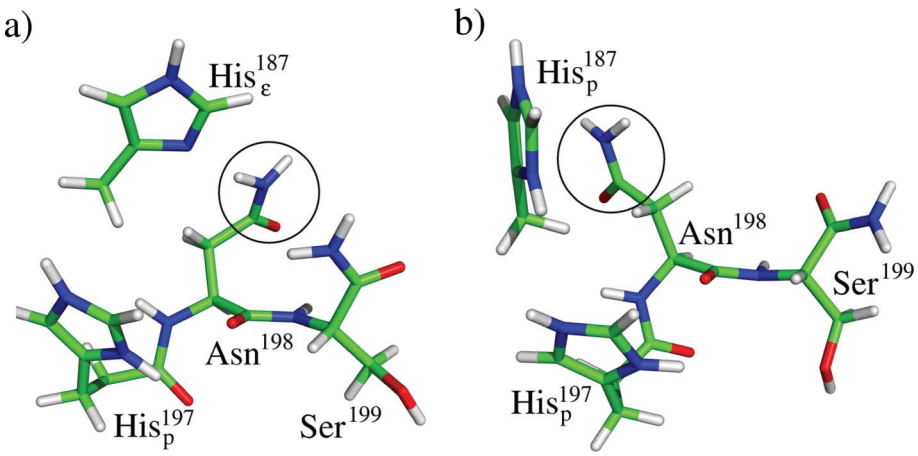

Fig. 3 (a) a representative snapshot from the MD_{EP} simulation where the Asn198 side chain is well-positioned for its cyclization; (b) a representative snapshot from the MD_{PP} simulation, where the Asn198 side chain is in orientated opposite to its backbone.

the Asn198 cyclization difficult. In the MD_{ED} simulation, Asn198 also adopts a reasonable position for its cyclization. However, it is difficult to imagine a feasible reaction pathway, considering its protonation state. In the different crystal structures His197 is well positioned to interact with O^{198}, suggesting that this residue plays an important role in stabilizing the negative charge of O^{198} in the tetrahedral intermediate by a charge-relay system. Even though during the MD_{ED} simulation His197 is in a good position to form this interaction, there would be significantly less stabilization with a neutral formal charge. On the other hand, during the MD_{ED} simulation His187 is far away from the Asn198 side chain. Our previous work demonstrated[39] that the most efficient way to activate this cyclization is by removal of a proton from N_{δ}^{198}. The nine different simulations indicate that His187 is the most (or the only) suitable candidate to accept the proton. However, during the MD_{ED} simulation this residue is too far away from the Asn198 side chain for this proton transfer to be possible. Nevertheless, these two amino acids interact during the entire MD_{EP} simulation, suggesting that this proton transfer can readily take place in this system. Thus, from all these considerations, the MD_{EP} system appears to be the only one in which asparagine cyclization and peptide bond cleavage is likely to occur.

The protonation state of the MD_{EP} system is consistent with available experimental data. Du *et al.* estimated the pK_a values of the F-block and G-block histidines in the *Mtu* recA intein,[28] that is, residues analogous to His187 and His197. The F-block histidine has pK_a values of 7.9 ± 0.6 in the spliced precursor and 8.9 ± 0.2 in the product; the values decrease to 7.3 ± 0.6 in the precursor and 6.2 ± 0.2 in the product for His197. Several experiments have determined the optimum pH for protein splicing to be in the range of 6.0-7.5.[14–16] The high pK_a value of the F-block histidine suggests that this residue may act as a base to accept a proton from the asparagine side chain. On the other hand, the pK_a value of His197 lies in the optimum pH range, and it is difficult to predict its protonation state. However, based on the results presented here, and crystal structures where this residue is in good disposition to form a hydrogen bond interaction with the asparagine's carbonyl group,[24,63] it is reasonable to conclude that this residue exists predominantly as the positively charged His_p tautomer, or at least that this is the form responsible for activity.

Free energy profile for Asn198 activation and cyclization

The QM/MM free energy profile (Fig. 4) for asparagine activation and subsequent cyclization is presented in this section. The same reaction pathway has been investigated previously,[39] but those calculations did not take into account explicitly the influence of the enzymatic environment or the effects of protein dynamics. The reaction takes place in two steps: (i) activation of the Asn198 side chain by a proton transfer to the neutral His187; and (ii) cyclization of the activated Asn. A snapshot of the MD_{EP} simulation was taken as the initial geometry, and because His187 was in a very good position to accept a proton from the Asn198 side chain (H_{δ}^{198}–N_{ε}^{187} = 1.97 Å), the direct proton transfer was modelled. In previous modelling of the reaction in the gas phase,[39] a water molecule was included to assist the proton transfer. This difference may reflect the importance of the enzymatic environment in providing a suitable arrangement for the reaction. The second step of the modelled reaction mechanism involves the Asn198 cyclization.

The calculated free energy barrier for proton transfer from Asn198 to His187 is 16.8 kcal mol^{-1}. The activated reactant produced lies 13.8 kcal mol^{-1} above the initial reactant. It is a zwitteronic species in which the N_{δ}^{198} atom bears a negative charge and His187 a positive one. The subsequent cyclization of the activated Asn198 presents a free energy barrier of 19.0 kcal mol^{-1} (5.2 kcal mol^{-1} above the intermediate), and therefore this is the rate-limiting step in the modelled reaction. This step was also found to be the rate-limiting step from the potential energy surface of the small cluster model studied in the gas-phase and including solvent effects with the PCM continuum model.[39] The final minimum is a succinimide intermediate, in which Asn198 has cyclized. This stationary point lies 15.9 kcal mol^{-1} above the initial reactant. Experimentally, the kinetics of each of the four steps in protein splicing have been studied for the *Pyrococcus abyssi* PolII intein.[13] For the third step, a rate constant (k_{cat}) of 2.8×10^{-4} s^{-1} has been measured at pH = 7.0 and 60 °C.[13] Applying the Eyring equation, this equates to a free energy barrier of 25.3 kcal mol^{-1}. Thus, the overall energy barrier (19.0 kcal mol^{-1}) computed in the present work is in reasonable agreement with the experimental value. It must be pointed out that in these calculations, quantum tunnelling and zero-point vibrational energy corrections are not

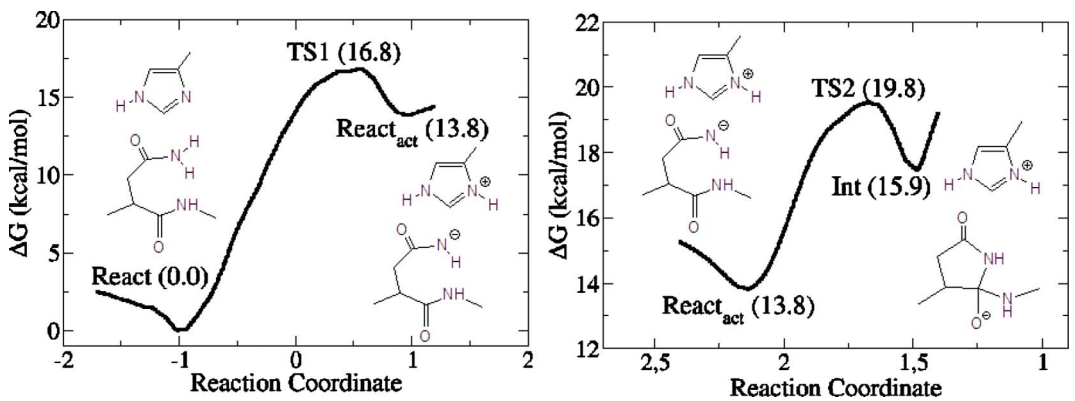

Fig. 4 Free energy profile computed at the SCCDFTB/CHARMM27 level of QM/MM theory for the third step of protein splicing, considering two steps: nucleophilic activation of Asn198 side chain (on the left) and subsequent Asn198 cyclization (on the right). Energies of important species in the reaction are indicated in parentheses (in kcal mol^{-1}), relative to the reactant state.

included, although they probably only have a small effect on the barrier. In the calculations carried out in the gas-phase,[39] for instance, the inclusion of ZPVE decreases the relative energies of the transition states by *ca.* 1 kcal mol^{-1}.

In order to check if SCCDFTB is a reliable method to characterize the cyclization of Asn198 in protein splicing, the potential energy surface of the reaction was also characterized at the SCCDFTB/CHARMM27, and the energies of these structures calculated with a higher level DFT QM/MM method for comparison (shown in Table S2†). The Gaussian09[64] program was employed to perform single point calculations on the SCCDFTB/CHARMM27 structures, treating the QM part with the B3LYP/6-31+(d) level of theory and including the MM point charges; the MM energies were added to get the final QM/MM energies. A similar protocol has been previously employed in order to test the accuracy of SCCDFTB.[65] The final snapshot of the window ascribed to the initial reactant was taken from the umbrella sampling simulation as the initial geometry, and the same reaction coordinates as employed in the PMF calculations were defined to compute the SCCDFTB/CHARMM27 potential energy surface. The first step, *i.e.*, the proton transfer from the Asn198 side chain to His187 shows an activation energy of 16.2 kcal mol^{-1} at the SCCDFTB/CHARMM27 QM/MM level, with the activated reactant 15.6 kcal mol^{-1} above the initial reactant. The cyclization of Asn198 requires overcoming the TS2 transition state, with an energy of 21.5 kcal mol^{-1}, and finally the tetrahedral intermediate is reached, with a relative energy of 18.2 kcal mol^{-1}. The relative energies calculated with B3LYP/6-31+G(d)/CHARMM27 are 24.2 and 23.2 kcal mol^{-1} for TS1 and the activated reactant, and 33.3 and 28.0 kcal mol^{-1} for TS2 and the intermediate. Thus, the B3LYP/6-31+G(d)/CHARMM27//SCCDFTB/CHARMM27 energies of all stationary points are 8–10 kcal mol^{-1} higher (relative to the reactants) than the SCCDFTB/CHARMM27 energies. Qualitatively both potential energy surfaces present the same profile, with TS2 as the rate-limiting step. It is worth noting that the DFT energies were calculated on structures optimized with SCCDFTB and therefore the stationary points are not real minima on the B3LYP/CHARMM27 potential energy surface; geometry optimization at the higher level would be likely to change the relative energies somewhat. Furthermore, some examples where the B3LYP functional shows limitations have been reported.[66–68] The B3LYP method often underestimates barriers compared to higher level calculations or experiments, but also significantly overestimates barriers in some cases. Quantum effects such as zero-point energy will also reduce the effective barrier a little. It should also be remembered that the apparent experimental value was from experiments at 60 °C (the apparent experimental barrier at 300 K is likely to be higher), and the free energy barriers here were calculated from simulations at 300 K. Altogether, taking into account that the SCCDFTB method probably underestimates the energy barriers for the reaction somewhat, the calculated free energy barriers appear consistent with experiment, supporting the mechanism modelled here.

QM/MM molecular dynamics simulations of key structures

Longer QM/MM MD simulations were run to investigate the structures of the reactant, activated reactant and intermediate. A schematic representation of these three states is shown in Fig. 4. Unlike MM MD simulations, a QM/MM method allows the topology (and electronic distribution) of the system to change, and therefore bonds can potentially form or break. Taking the initial geometries and velocities from the umbrella sampling simulations, the three states cited above were modelled by QM/MM MD simulations. In the QM/MM MD of the intermediate here, a harmonic restraint was applied to the N_{δ}^{198}–C^{198} distance to avoid its opening (force constant of 50 kcal mol^{-1} Å^{-2}). The QM part included the same atoms as in the umbrella sampling simulations, and it was again treated with the SCCDFTB method.[54,69] Also as above, the MM part was modelled with the CHARMM27 force field.

Each QM/MM MD simulation was run for 500 ps. The RMSD of the three QM/MM MD simulations (shown in the Supporting Information† and calculated taking the initial geometry as a reference and including all heavy atoms of the intein) show low values, and therefore the systems can be considered stable and equilibrated for structural analysis. The average and standard deviation values of selected distances are presented in Table 2, and a representative snapshot from the simulation of each state is shown in Fig. 5. Based on an experiment using the trans-splicing intein from the *dnaE* gene of *Synechocystis sp.* PCC6803,[25] in which the mutations of residues located in block F to alanine inhibit the C-terminal cleavage, it is thought that their role is to form a hydrophobic surface at the back of the Asn198 residue. Therefore, we analysed particularly the behaviour of residues located in block F.

Reactant. In the SCCDFTB/CHARMM27 MD simulation of the reactant, the intein showed the same conformation as during the MD_{EP} simulation. The Asn198 side chain is in close proximity to His187 (H_{δ}^{198}–N_{ε}^{187} distance of 2.08 ± 0.28 Å) and therefore in good position for the proton transfer. The N_{δ}^{198}–C^{198} distance is 4.04 ± 0.20 Å, slightly longer than during the MD_{EP} simulation and the penultimate His197 residue interacts directly with the carbonyl O atom of Asn198. In the reactant state, Arg181 is far away from the Asn198 side chain (N_{δ}^{198}–N_{h1}^{181} distance of 7.03 ± 1.39 Å) and from Asp183.

Activated reactant. In this state, a proton (H_{δ}^{198}) has been transferred from Asn198 to His187. Therefore, the Asn198 side chain bears a negative charge and His187 is positively charged. This protonation state is maintained during the entire simulation. The average distances of H_{δ}^{198} from N_{δ}^{198} and N_{ε}^{187} are 2.15 ± 0.17 Å and 1.02 ± 0.02 Å, respectively. The results indicate

Table 2 Average values and standard deviation (in brackets) of some important interatomic distances from SCCDFTB/CHARMM27 QM/MM MD simulations (over 500 ps) for the initial reactant, activated reactant, and intermediate states. All values in Å. Atom labels are shown in Fig. 2B

Distances	Reactant	Activated reactant	Intermediate
N_{δ}^{198}–C^{198}	4.04 (0.20)	3.88 (0.07)	1.49 (0.03)
H_{δ}^{198}–N_{δ}^{198}	1.02 (0.01)	2.15 (0.17)	3.99 (0.77)
H_{δ}^{198}–N_{ε}^{187}	2.08 (0.28)	1.02 (0.02)	1.03 (0.01)
H_{δ}^{198}–O^{182}	4.06 (0.37)	3.28 (0.44)	2.80 (0.62)
H_{δ}^{197}–O^{198}	2.05 (0.72)	1.75 (0.14)	1.61 (0.03)
N_{δ}^{198}–N_{h1}^{181}	7.03 (1.39)	4.99 (1.03)	8.23 (0.45)
N_{h1}^{181}–$O_{\delta2}^{183}$	11.89 (0.60)	7.27 (0.95)	2.84 (0.45)
N_{h2}^{181}–$O_{\delta1}^{183}$	11.79 (0.59)	8.61 (0.85)	2.87 (0.60)

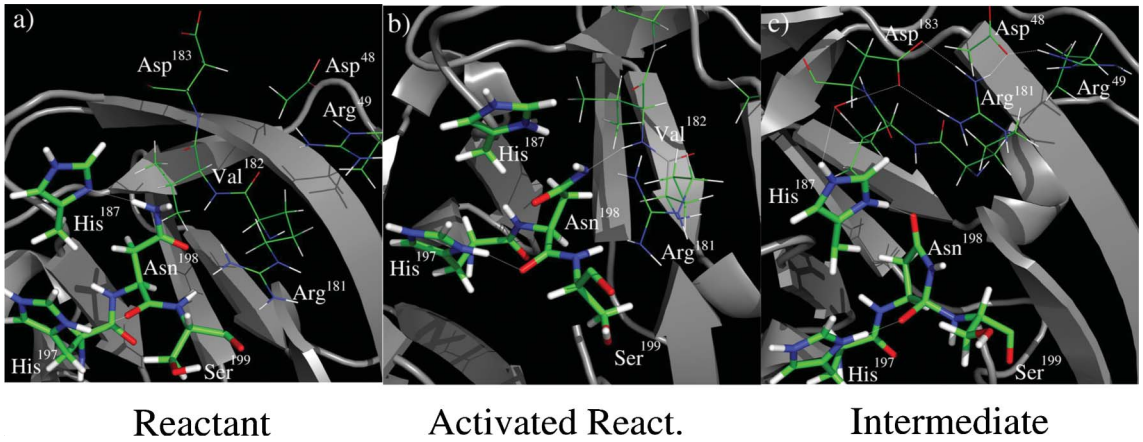

Fig. 5 Representative structures of the active site during the three SCCDFTB/CHARMM MD simulations carried out over 500 ps of production to model the main states located in the PMF simulation: (a) initial reactant; (b) activated reactant; and (c) intermediate. The conformation of the protein adapts to the electronic structure shown by the reactive zone at each stationary point (see body text).

that deprotonation of Asn198 does not lead to its immediate cyclization, *i.e.* the deprotonated ('activated') state of asparagine is stable and does not react on the timescale of these simulations. The N_{δ}^{198}–C^{198} distance (3.88 ± 0.07 Å) is only slightly shorter than in the reactant. Interestingly, the new charge distribution triggers some conformational changes of the amino acids located in block F, especially in the position of Arg181. This residue has approached the Asn198 side chain, and the N_{δ}^{198} –N_{h1}^{181} distance (4.99 ± 1.03 Å) is significantly shorter than it was in the reactant. In addition, the Val182 amide backbone interacts weakly with N_{δ}^{198} (see Fig. 5B) and N_{ε}^{187} through its amide N and O atoms, respectively. On the other hand, Arg181 is far away from Asp183.

Intermediate. In this state, Asn198 has cyclized (N_{δ}^{198}–C^{198} distance of 1.49 ± 0.03 Å) to form a succinimide tetrahedral intermediate. The negative charged borne by O^{198} is mainly stabilized by His197, as shown by the fact that the H_{δ}^{197}–O^{198} distance is reduced to 1.61 ± 0.03 Å, 0.4 Å shorter than in the reactant. His187 no longer interacts with the Asn198 side chain in this state; instead, one of the protonated N atom of His187 forms a hydrogen bond with the oxygen atom of a water molecule (see Fig. 5C), which in turn interacts with Asp183. Thus, we can infer that Asp183 is involved in stabilizing the new charged borne by His187. Interestingly, the neutralization of the negative charge of Asn198 due to its cyclization has triggered the movement of Arg181 away from Asn198: the average N_{δ}^{198}–N_{h1}^{181} distance is 8.2 ± 0.4 Å, more than 3 Å longer than in the activated reactant structure. In this state, Arg181 forms two stable hydrogen bonds with the two O_{δ} atoms of Asp183. In this way, in the tetrahedral intermediate a hydrogen bond network is formed by Arg181, Asp183, Asp48 and Arg49 (see Fig. 5C).

Role of the environment. The QM/MM MD calculations indicate the importance of the protein environment, especially the Arg181-Val182-Asp183 segment, in assisting the reaction mechanism. Arg181 is mobile, and adopts a different position in different stages of the reaction. The positive charge of this residue may be important in stabilizing the negative charge located on the Asn198 side chain after one of its protons has been transferred to His187. Then, once Asn198 has cyclized to form the tetrahedral intermediate, Arg181 moves away from Asn198 to interact with Asp183 and be part of a hydrogen bond network. It is also worth noticing that, unlike in the reactant, Asp183 interacts weakly with His187 through a bridging water molecule. Many inteins contain an aspartate or glutamate residue analogous to Asp183 in GyrA. For instance, Glu424 in the *Mycobacterium tuberculosis* recA mini-intein (of which a crystal structure has the PDB code 2IN0),[63] or Asp438 in PI-SceI (PDB code 1EF0),[70] are located 5 and 4 amino acids before the upstream histidine, respectively. The presence of an aspartate at this position makes this residue a good candidate to act as a base group accepting a proton from the Asn198 side chain. However, the pK_a value of Asp183 predicted by PROPKA[40] is 4.7. Although this value is slightly larger than the value for a standard aspartic acid (3.8), it is still predicted to be negatively charged under physiological conditions, and it seems unlikely that this amino acid would be involved in the reaction. In addition, Asp183 is far away from the side chain of Asn198 (the N_{δ}^{198}–$O_{\delta 1}^{183}$ distance is 7.4 Å in the 1AM2 crystal structure),[24] and only when His187 bears a positive charge (after the proton has been transferred from Asn198) does Asp183 approach (see above). Therefore, it is unlikely that Asp183 activates Asn198 side chain by abstracting a proton.

Conclusions

We present a computational study of the C-terminal peptide cleavage in the GyrA intein, including explicitly the entire intein and the solvent. MM molecular dynamics simulations demonstrate that the most suitable protonation state for protein splicing is that in which His187 and His197 adopt the His_{ε} (neutral) and His_{p} (positively charged) protonation states, respectively. These protonation states are very likely at the pH range (6.0 to 7.5) in which protein splicing takes place.[14–16]

A reaction pathway in which His187 acts as a base to activate (deprotonate) the Asn198 side chain for its subsequent cyclization has been characterized here by QM/MM umbrella sampling molecular dynamics simulations. The SCCDFTB/CHARMM27 free energy profile (potential of mean force), shows a free energy barrier of 19.0 kcal mol^{-1}, in reasonable agreement with the experimental value of 25.3 kcal mol^{-1}.[13] Higher level (density functional theory) QM/MM single-point calculations indicate

that this calculated value may be underestimated somewhat, although qualitatively the same energy profile is observed. It should also be taken into account that while the experiments were carried out at 60 °C, a temperature of 25 °C was considered in the simulations presented here. QM/MM molecular dynamics simulations show the importance of the amino acids located in block F during the reaction, especially the Arg181-Val182-Asp183 segment. They also show that His197 forms a strong hydrogen bond with O^{198} and stabilizes the negative charge of this atom at the tetrahedral intermediate.

In summary, this work provides structural details (including the effects of conformational dynamics) for the entire intein of the key processes of splicing in the hydrolysis of its C-junction. In particular, the results clarify the role played by two histidines, and support the mechanism in which the F-block histidine accepts a proton from the Asn198 side chain, in agreement with previous studies.[27,28] A similar mechanism has been proposed for the N-terminal splice junction, in which a histidine located in block B is thought to act as a base to deprotonate the nucleophile, a cysteine.[28] In addition, the results shed light on the catalytic role of the intein, and highlight the importance of the residues located in the F-block during this reaction mechanism.

Acknowledgements

J. I. M. and X. L. thank Eusko Jaurlaritza and the Spanish Office for funding. A. J. M. is an EPSRC Leadership Fellow (grant number EP/G007705/1).

Notes and references

1 H. Paulus, *Chem. Soc. Rev.*, 1998, **27**, 375.
2 H. Paulus, *Annu. Rev. Biochem.*, 2000, **69**, 447.
3 C. J. Noren, J. Wang and F. B. Perler, *Angew. Chem., Int. Ed.*, 2000, **39**, 450.
4 F. S. Gimble and J. Thorner, *Nature*, 1992, **357**, 301.
5 P. M. Kane, C. T. Yamashiro, D. F. Wolczyk, N. Neff, M. Goebl and T. H. Stevens, *Science*, 1990, **250**, 651.
6 R. Hirata, Y. Ohsumk, A. Nakano, H. Kawasaki, K. Suzuki and Y. Anraku, *J. Biol. Chem.*, 1990, **265**, 6726.
7 T. C. Evans and M. Q. Xu, *Chem. Rev.*, 2002, **102**, 4869.
8 K. Tori, B. Dassa, M. A. Johnson, M. W. Southworth, L. E. Brace, Y. Ishino, S. Pietrokovski and F. B. Perler, *J. Biol. Chem.*, 2009, **285**, 2515.
9 K. Tori and F. B. Perler, *J. Bacteriol.*, 2011, **193**, 2035.
10 F. B. Perler, D. G. Comb, W. E. Jack, L. S. Moran and B. Quiang, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 5577.
11 M. Kawasaki, S. I. Makino, H. Matsuzawa, Y. Satow, Y. Ohya and Y. Anraku, *Biochem. Biophys. Res. Commun.*, 1996, **222**, 827.
12 M. Q. Xu and F. B. Perler, *EMBO J.*, 1996, **15**, 5146.
13 K. V. Mills, D. M. Dorval and K. T. Lewandowski, *J. Biol. Chem.*, 2004, **280**, 2714.
14 D. W. Wood, V. Derbyshire, W. Wu, M. Chartrain and M. B. G. Belfort, *Biotechnol. Prog.*, 2000, **16**, 1055.
15 D. W. Wood, W. Wu, G. Belfort, V. Derbyshire and M. Belfort, *Nat. Biotechnol.*, 1999, **17**, 889.
16 J. P. Gangopadhyay, S. Q. Jiang and H. Paulus, *Anal. Chem.*, 2003, **75**, 2456.
17 Y. Shao, M. Q. Xu and H. Paulus, *Biochemistry*, 1995, **34**, 10844.
18 M. Q. Xu, D. G. Comb, H. Paulus, C. J. Noren, Y. Shao and F. B. Perler, *EMBO J.*, 1994, **13**, 5517.
19 A. A. Cooper, Y. J. Chen, M. A. Lindorfer and T. H. Stevens, *EMBO J.*, 1993, **12**, 2575.
20 L. Chen, J. Benner and F. B. Perler, *J. Biol. Chem.*, 2000, **275**, 20431.
21 S. Wang and X. Q. Liu, *J. Biol. Chem.*, 1997, **272**, 11869.
22 C. P. Scott, E. Abel-Santos, M. Wall, D. C. Wahnon and S. J. Benkovic, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 13638.
23 F. B. Perler, *Nucleic Acids Res.*, 2000, **28**, 344.
24 T. Klabunde, S. Sharma, A. Telenti, W. R. Jacobs and J. C. Sacchettini, *Nat. Struct. Biol.*, 1998, **5**, 31.
25 I. Ghosh, L. Sun and M. Q. Xu, *J. Biol. Chem.*, 2001, **276**, 24051.
26 S. Pietrokovski, *Protein Sci.*, 1994, **3**, 2340.
27 Y. Ding, M. Q. Xu, I. Ghosh, X. Chen, S. Ferrandon, G. Lesage and Z. Rao, *J. Biol. Chem.*, 2003, **278**, 39133.
28 Z. Du, P. T. Shemella, Y. Liu, S. A. McCallum, B. Pereira, S. K. Nayak, G. Belfort, M. Belfort and C. Wang, *J. Am. Chem. Soc.*, 2009, **131**, 11581.
29 R. Lonsdale, K. E. Ranaghan and A. J. Mulholland, *Chem. Commun.*, 2010, **46**, 2354.
30 B. Peters and B. L. Trout, *Biochemistry*, 2006, **45**, 5384.
31 F. A. Konuklar, V. Aviyente and M. F. Ruiz-Lopez, *J. Phys. Chem. A*, 2002, **106**, 11205.
32 F. A. Konuklar, V. Aviyente, T. Z. Sen and I. Bahar, *J. Mol. Model.*, 2001, **7**, 147.
33 S. Catak, G. Monard, V. Aviyente and M. F. Ruiz-López, *J. Phys. Chem. A*, 2006, **110**, 8354.
34 S. Catak, G. Monard, V. Aviyente and M. F. Ruiz-López, *J. Phys. Chem. A*, 2008, **112**, 8752.
35 I. Kaliman, A. Nemukhin and S. Varfolomeev, *J. Chem. Theory Comput.*, 2010, **6**, 184.
36 P. Shemella, B. Pereira, Y. Zhang, P. V. Roey, G. Belfort, S. Garde and S. K. Nayak, *Biophys. J.*, 2007, **92**, 847.
37 D. G. A. Johansson, B. Macao, A. Sandberg and T. Hard, *J. Mol. Biol.*, 2008, **377**, 1130.
38 D. G. A. Johansson, G. Wallin, A. Sandberg, B. Macao, J. Aqvist and T. Hard, *J. Am. Chem. Soc.*, 2009, **131**, 9475.
39 J. I. Mujika, X. Lopez and A. J. Mulholland, *J. Phys. Chem. B*, 2009, **113**, 5607.
40 H. M. M. Olsson, C. R. Søndergard, M. Rostkowski and J. H. Jensen, *J. Chem. Theory Comput.*, 2011, **7**, 525.
41 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926.
42 K. E. Shaw, C. J. Woods and A. J. Mulholland, *J. Phys. Chem. Lett.*, 2010, **1**, 219.
43 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187.
44 www.charmm.org.
45 A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586.
46 C. L. Brooks and M. Karplus, *J. Chem. Phys.*, 1983, **79**, 6312.
47 M. W. Van der Kamp, F. Perruccio and A. J. Mulholland, *Proteins: Struct., Funct., Bioinf.*, 2007, **69**, 521.
48 L. Masgrau, K. E. Ranaghan, N. S. Scrutton, A. J. Mulholland and M. J. Sutcliffe, *J. Phys. Chem. B*, 2007, **111**, 3032.
49 K. E. Ranaghan, L. Ridder, B. Szefczyk, W. A. Sokalski, J. C. Hermann and A. J. Mulholland, *Mol. Phys.*, 2003, **101**, 2695.
50 C. L. Brooks, III and M. Karplus, *J. Mol. Biol.*, 1989, **208**, 159.
51 L. Ridder, I. Rietjens, J. Vervoort and A. J. Mulholland, *J. Am. Chem. Soc.*, 2002, **124**, 9926.
52 J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327.
53 G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187.
54 A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227.
55 M. Elstner, D. Porezag, G. Jungnickel, T. Frauenheim, S. Suhai and G. Seifert, in *Tight-Binding Approach to Computational Materials Science*, ed. P. E. A. Turchi, A. Gonis and L. Colombo, 1998, vol. 491, pp. 131-136.
56 N. Reuter, A. Dejaegere, B. Maigret and M. Karplus, *J. Phys. Chem. A*, 2000, **104**, 1720.
57 A. Rodriguez, C. Oliva, M. Gonzalez, M. W. Van der Kamp and A. J. Mulholland, *J. Phys. Chem. B*, 2007, **111**, 12909.
58 P. H. Konig, M. Hoffmann, T. Frauenheim and Q. Cui, *J. Phys. Chem. B*, 2005, **109**, 9082.
59 A. L. Bowman, I. M. Grant and A. J. Mulholland, *Chem. Commun.*, 2008, 4425.
60 D. Xu and H. Guo, *J. Am. Chem. Soc.*, 2009, **131**, 9780.
61 Q. Xu, H. B. Guo, A. Wlodawer, T. Nakayama and H. Guo, *Biochemistry*, 2007, **46**, 3784.

62 S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman and J. M. Rosenberg, *J. Comput. Chem.*, 1992, **13**, 1011.
63 P. Van Roey, B. Pereira, Z. Li, K. Hiraga, M. Belfort and V. Derbyshire, *J. Mol. Biol.*, 2007, **367**, 162.
64 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. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. 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. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *GAUSSIAN 09*, Gaussian, Inc., Wallingford, CT, 2009.
65 L. Capoferri, M. Mor, J. Sirirak, E. Chudyk, A. J. Mulholland and A. Lodola, *J. Mol. Model.*, 2011, **17**, 2375.
66 F. Claeyssens, J. N. Harvey, F. R. Manby, R. A. Mata, A. J. Mulholland, K. E. Ranaghan, M. Schutz, S. Thiel, W. Thiel and H. J. Werner, *Angew. Chem., Int. Ed.*, 2006, **45**, 6856.
67 R. Lonsdale, J. N. Harvey and A. J. Mulholland, *J. Phys. Chem. Lett.*, 2010, **1**, 3232.
68 M. W. Van der Kamp, J. Zurek, F. R. Manby, J. N. Harvey and A. J. Mulholland, *J. Phys. Chem. B*, 2010, **114**, 11303.
69 Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim and M. Karplus, *J. Phys. Chem. B*, 2001, **105**, 569.
70 B. W. Poland, M. Q. Xu and F. A. Quiocho, *J. Biol. Chem.*, 2000, **275**, 16408.

