Supporting Information

Distinct roles of the catalytic cysteine and histidine in the protease and ligase mechanisms of human legumain as revealed by DFT based QM/MM simulations

*Brigitta Elsässer, Florian B. Zauner, Johann Messner, Wai Tuck Soh, Elfriede Dall,

Hans Brandstetter

***Corresponding Author:** brigitta.Elsässer@sbg.ac.at

Computational Details and Methods

Structural Model

The high resolution crystallographic structure of *human* legumain (1.8 Å) complexed with cystatin E/M was used as starting structure for our calculations (PDB code 4N6O[1]). Cystatin is legumain inhibitor, which can bind canonical to legumain with its reactive center loop. Due to the large size of the substrate and in order to make it more substrate-like only a GSNSI pentamer located at the active site (i.e. P3 – P2´ residues) was applied for the simulations. The C– and N–terminus were capped by ACE and NME to represent the former protein-like environment. In order to study the cleavage procedure the complex was titrated at pH 4.0 using MOE2016.08[2] to generate the starting protonation state. In addition MOE2016.08 has been used to calculate all relevant pK_a values mentioned in the manuscript. Five chloride ions served as counter ions to maintain neutrality in the system and the complex was solvated in an 80 Å cubic box of water (containing 40388 water molecules) centered on the enzyme. According to *in silico* titration experiments (using the H++ server[3] and MOE2016.08[2]) both Cys189 and His148

were kept neutral. MD simulations followed by QM/MM optimization showed that at pH 4.0 the sulfur of Cys189 flips over and points toward the carbonyl of the scissile peptide bond.

(a) Quantum Mechanical Region

A total of 100 atoms were treated quantum mechanically (QM region). The forces in the QM subsystem were calculated at the DFT/B3LYP level using Ahlrichs-pVDZ basis set, which demonstrated to provide reliable results for enzyme reactions [4-9]. These comprise the active site residues: SNN147, His148, Gly149, Cys189, Wat305 (that water that potentially participates in the hydrolysis) and fragments of Thr146, Ser150, Ala188, Glu190. In addition the substrate residues: Asn302, Ser302 and fragments of Ser301 and Ile304 to represent a reliable electronic description of the QM core.

(b) Mobile Region

The remaining protein atoms (~4100 atoms), counter ions and the solvent molecules of the water box were included in the molecular mechanics region (MM region) and computationally treated by AMBER99 force field.

(c) Water Molecules

All water molecules present in the crystal structure were included in the calculations. The location of water molecules that can hydrogen bond to a proton donor and/or acceptor is likely to be important for the stabilization of the transition states and intermediates.

MD and QM/MM calculations

The bonds between the QM and MM subsystems were capped with H atoms.[10] In the first step of the calculation the entire solvent-enzyme-substrate structure was equilibrated by performing a series of molecular dynamics annealing runs for 100 ps at temperatures 50 K, 150 K, 200 K, 250 K and for 100 ns at 298.15 K with fixed positions of the atoms in the QM region. Afterwards, the atomic positions in the MM region were fixed and the atomic positions in the QM region were optimized at the B3LYP level and the wave-functions were expanded using the Ahlrichs-pVDZ basis set, since previous calculations [5,11] on enzymatic systems have shown good performance using Ahlrichs-pVDZ basis set by describing the geometry and energetics. Then the resulting equilibration stage of the structure was optimized using multi-region optimization algorithm as implemented in NWChem[12]. This method performs a sequence of alternating optimization cycles of the QM and MM regions. The effective charges were recalculated in each optimization cycle by fitting the electrostatic field outside the QM region to that produced by the full electron density representation. Several cycles of this optimization were carried out until convergence was obtained.

Spring Method

Starting from the legumain-cystatine E/M complex a sequence of constrained optimizations was performed to study the cleavage step. Harmonic restraints between SG(Cys189) and the peptide carbonyl C(Asn302) (1.8 Å) and/or between HG(Cys189) and the peptide nitrogen N(Ser303) (1.0 Å) were imposed to drive the system over the transition states and reaction barriers to the intermediate state while at the same time allowing the MM system (initially equilibrated to the

reactant structure) to adjust to the changes. When a reasonable estimate of the acyl-enzyme formation was obtained, the constraints were lifted, and the system was optimized using a sequence of optimizations and dynamical relaxation steps similarly to those discussed above. Similarly, to simulate the water attack a spring between OW(Wat305) and the carbonyl C(Asn302) was used to generate the second intermediate. The product state was reached by a constrain between HW(Wat305) and SG(Cys189) followed by optimization and relaxation steps.

NEB (nudged elastic band) calculations and NMA (normal mode analysis)

To obtain an unbiased view of the reactive process, we have used the implementation of the NEB approach[13] within the NWChem QM/MM module. The NEB method optimizes the trial reaction pathway between two fixed points. In our simulations we calculated 15-15 replicas connected by harmonic spring forces between reactant and INT1, INT1 and INT2 and between INT2 and product states. To ensure full relaxation of the protein environment, the first NEB optimization pass was followed by 20 ps of molecular dynamics equilibration at room temperature of the MM region for each bead along the pathway. After this equilibration, another round of NEB optimization was performed. The structures of the transition states has also been characterized by frequency calculations (NMA) using NWChem [12].

Free Energy Calculations

The free energy profile over the NEB optimized pathway was obtained by calculating free energy differences between the consecutive NEB beads. This approach is similar to the

multilevel perturbation methodology, which has been used by Valiev et al.[14] for reactions in solutions.

The calculated product state of the ligase reaction fits very well with the reactant state of the protease action (Figure S1).

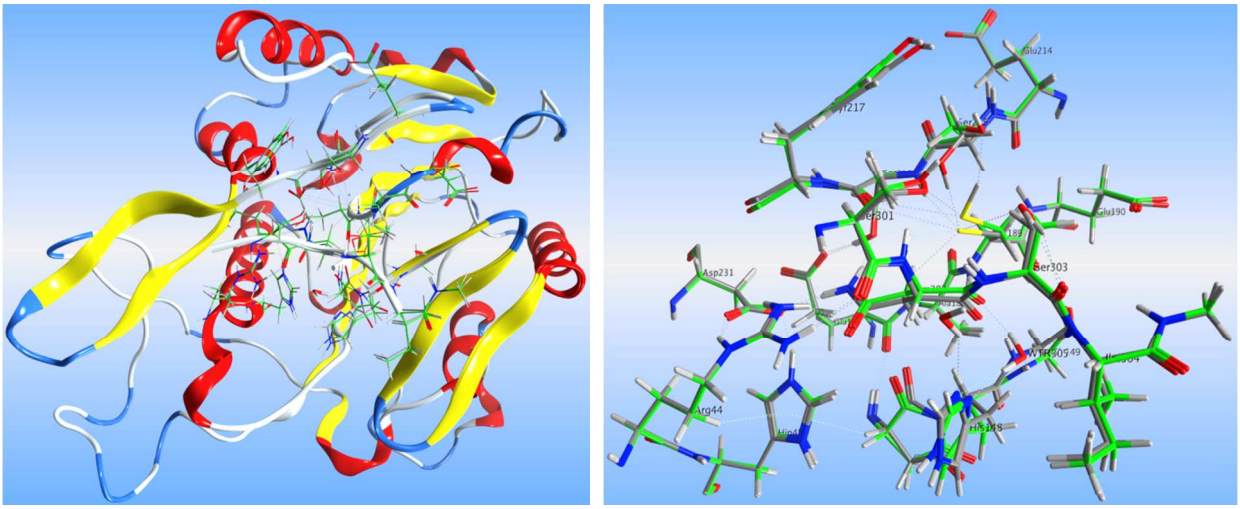

Figure S1. Superposition of the reactant state of the cleavage step (green carbons) with the product state of the ligase step (grey carbons), RMSD over 267 residues 0.089 Å. (right) The active site is represented by sticks and the rest of the enzyme is depicted as cartoon. (left) Zoom up of the active site.

Preparation of C189M-legumain and testing its ligase activity

The point mutant C189M-prolegumain was cloned, expressed and purified following a protocol described earlier[15]. Briefly, the Cys189 to Met189 mutation was introduced using the inverse-PCR method[16]. Protein expression was performed by using the *Leishmania tarentolae* expression system (LEXSY, Jena Bioscience, Germany). Inactive proenzyme was purified from the LEXSY supernatant via Ni^{2+}-purification. Subsequently, the protein was concentrated using an Amicon Ultra centrifugal filter unit (MWCO: 10 kDa; Millipore) and buffer exchanged to 20 mM

Tris pH 7.5, 20 mM NaCl and 5 mM DTT via a PD-10 column (GE-Healthcare). To convert the zymogen to the mature asparaginyl endopeptidase domain, the C-terminal prodomain was removed by addition of human cathepsin-S in a 1:20 molar ratio at pH 4.0 and 37 °C[17]. After 5 h incubation the reaction was stopped by the addition of 1 µM E-64. The *in trans* activated protein was then further purified by size exclusion chromatography utilizing an Äkta FPLC system equipped with a Superdex 75 10/300 GL column (GE Healthcare) equilibrated in a buffer composed of 20 mM citric acid pH 4.0, 50 mM NaCl and 2 mM DTT.

To test the ligase activity of the *in trans* activated C189M-legumain variant, we incubated wild-type legumain with a two-fold excess of human cystatin E at pH 4.0 and 37 °C. This lead to the conversion of intact cystatin E to Asn39-cleaved cystatin E as monitored by SDS-PAGE[18]. Subsequently, C189M-legumain was added to the reaction, which resulted in the conversion of free (not bound to wild-type legumain), Asn39-processed cystatin E to intact cystatin E.

Generation of the product state of the ligation using MMTS-blocked Cys189 and Met189

Further support and confirmation of the calculated ligation mechanism arises from calculations with MMTS-blocked Cys189 (Cys189–S–CH_3) and C189M mutant enzyme. Since experimental studies of Dall et al. have shown that the ligase activity remains upon oxidation of Cys189[1] or mutation to Met189, as next the proposed mechanism was also simulated with disulfide Cys189–S–CH_3 and Met189 variant, separately. Using the "Protein Builder" tool of MOE2016.08 Cys189 was mutated to Met189 and the mixed disulfide was generated using the "Builder" function. In both cases the enzyme-substrate complex was optimized as described in the *Computational Methods* section. The only difference between the native and modified enzyme

is the position of the proton of the substrate C-terminus carbonyl. Namely, in case of the Cys189–S–CH_3 and Met189 mutant Ser303 still remains neutral; however the catalytic histidine (His148) becomes positively charged and rather builds up a zwitterion with the C-terminus carbonyl (Figure S4 left). This might be due to the lack of negative charge on residue 189. Though, the calculated reaction proceeds similarly, with the only difference that an additional proton transfer (from NE2(His148) to OD2(Asn302)), which is necessary to eliminate the catalytic water (Figure S2, S3).

The resulting product (Figure S4 right) is analogue to the product of the native enzyme. The carbonyl of the scissile peptide bond is buried in the oxy anion hole of the peptide nitrogens N(Gly149) and N(Met189), furthermore the catalytic water is located between NE2(His148) and N(Ser303).

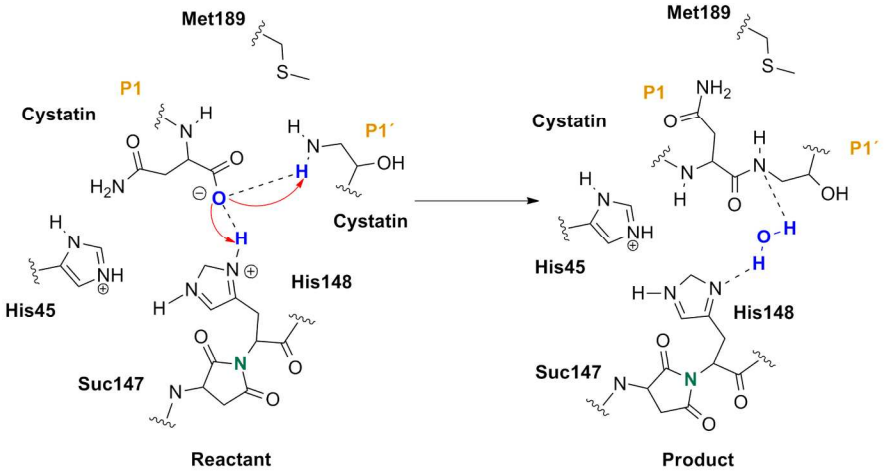

Figure S2. Calculated ligation mechanism with Met189.

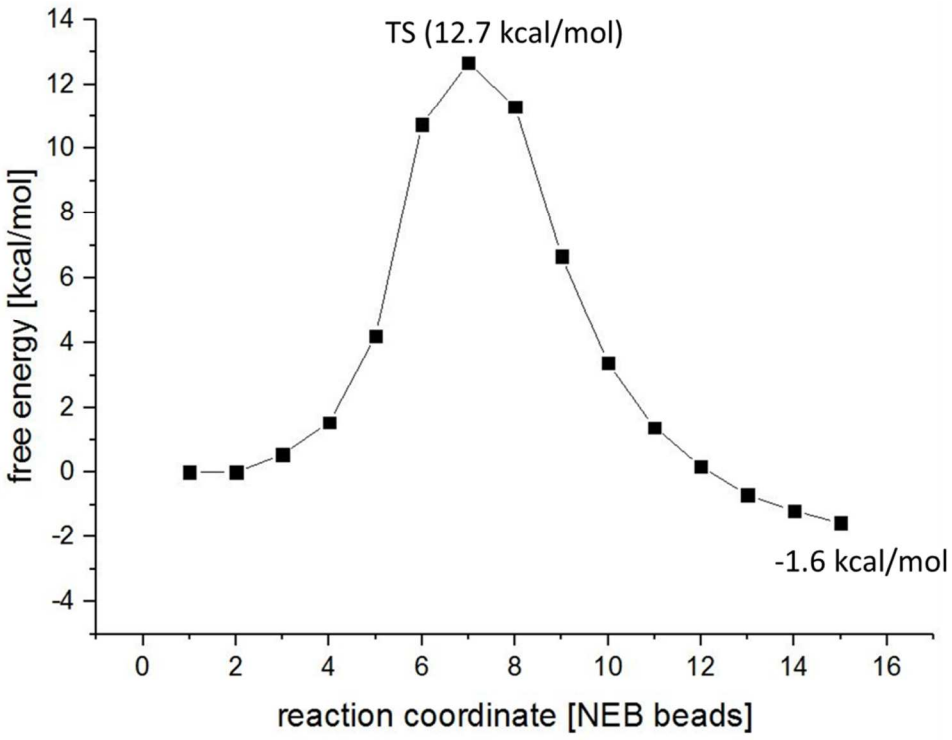

Figure S3. Calculated free energy profile for the ligation with C189M mutant.

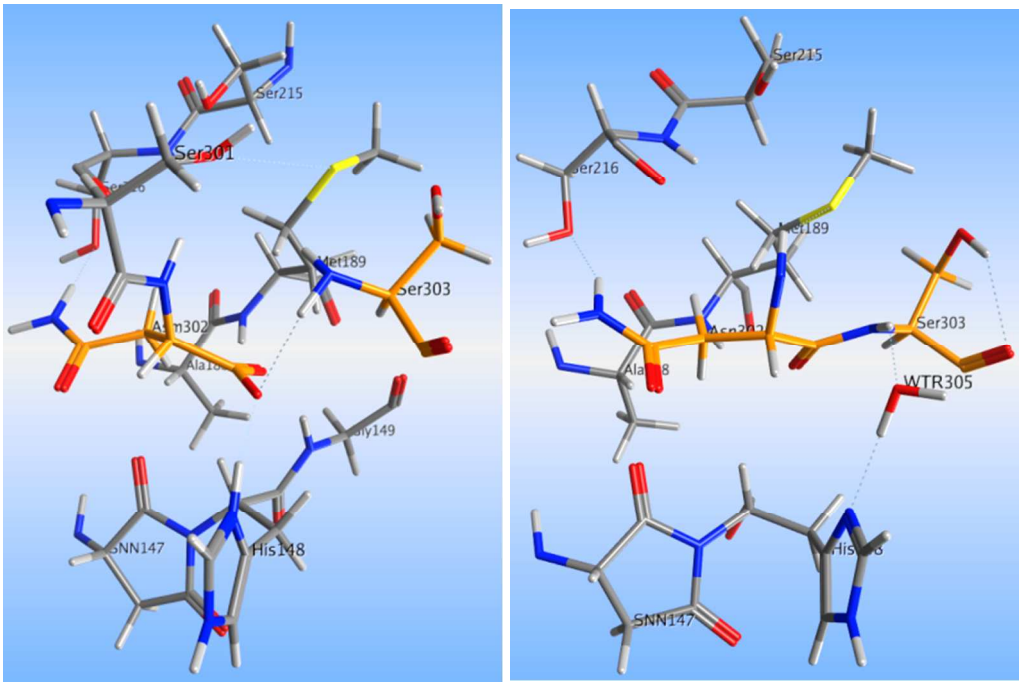

Figure S4. Zoom up of the active site region for reactant (left) and product (right) state of the Met189 mutant. Orange carbons depict the substrate and grey carbons designate the enzyme.

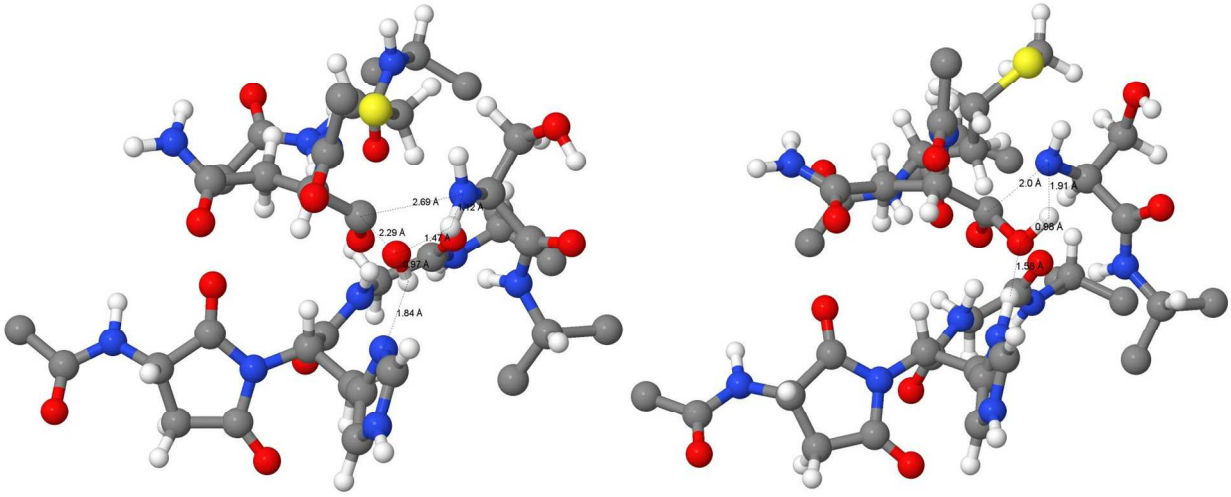

Figure S5. Transition state structures of the ligation pathway for the wild type (left) and C189M mutant (right).

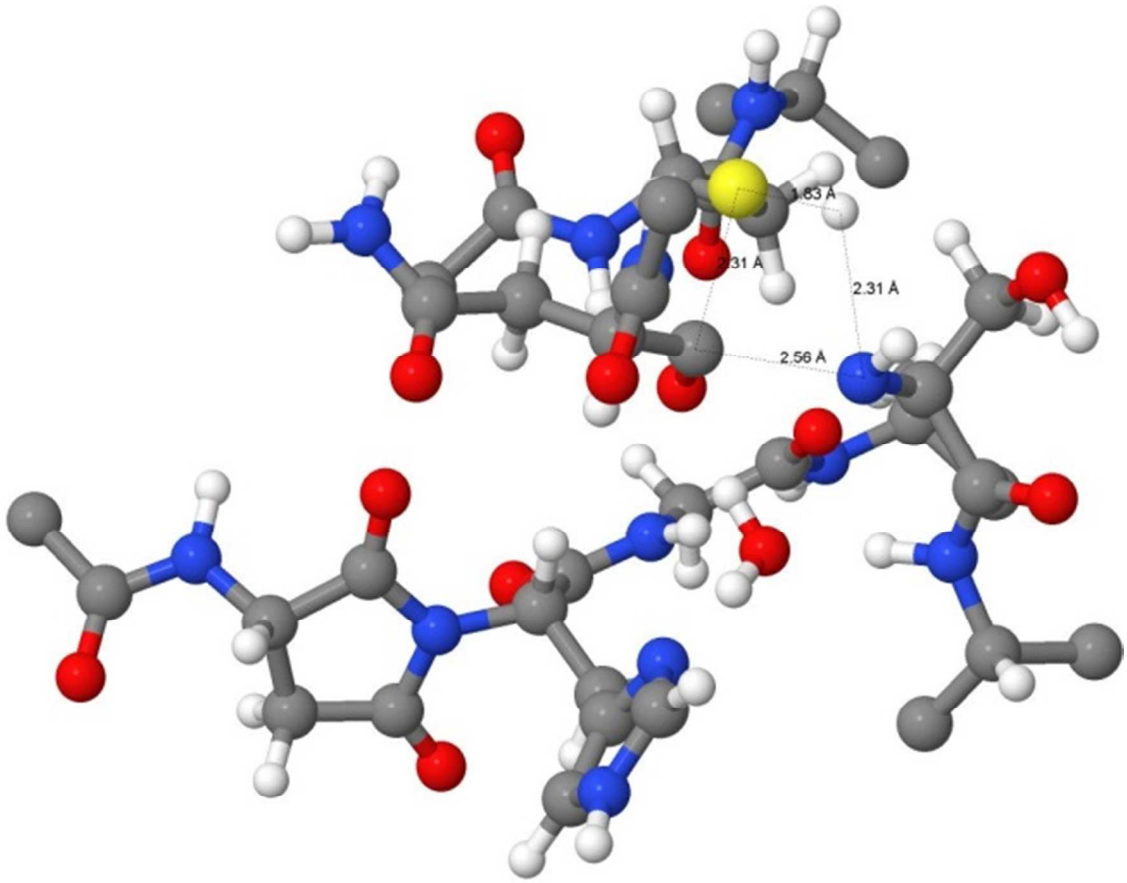

Figure S6. Structure of TS1

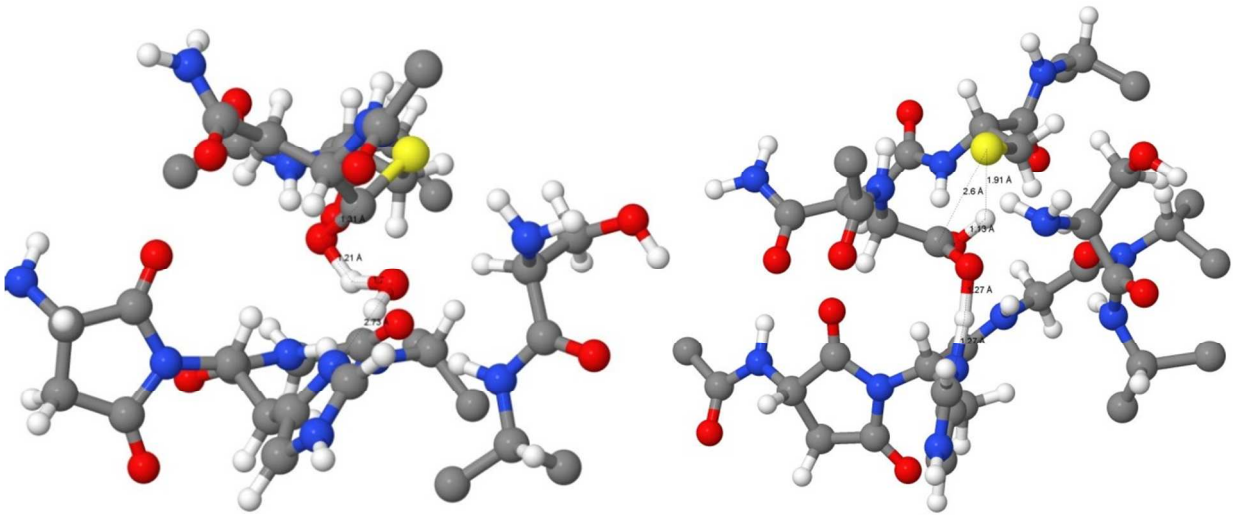

Figure S7. (left) TS2 (right) TS3

Proteolysis with E190K

Experimental findings of Dall et al.[19] revealed that the E190K mutant shows a 4-fold increase in k_{cat} at pH 5.5 compared with the wild type, whereas K_M remains the same. pKa calculations result that the E190K mutant lowers the local pKa of Cys189. Thereby, the deprotonation of Cys189 is encouraged and leads to a faster thioester formation. Reaction path calculations using the E190K yielded the same reaction path, however the reaction barrier for the first step of the reaction is much lower (12.6 kcal/mol) in case of the Lys190 variant (Table S1). Transition state search and free energy calculations display a similar shape for the energy curve of the formation of the first intermediate. In addition, the transition state structure is also very likely to the one as in case of the native enzyme (Figure S8). Also the second step of the reaction proceeds by the same manner via a late TS2. In closing step the energy curve shows a small plateau at TS3, but comparable barrier.

Table S1. Comparison of the calculated relative energies (in kcal/mol) over the proteolysis pathway of the native and E190K mutated human legumain.

	native enzyme	E190K mutant
RS→INT1	19.3	12.6
INT1→INT2	7.0	6.7
INT2→PS	7.1	6.6

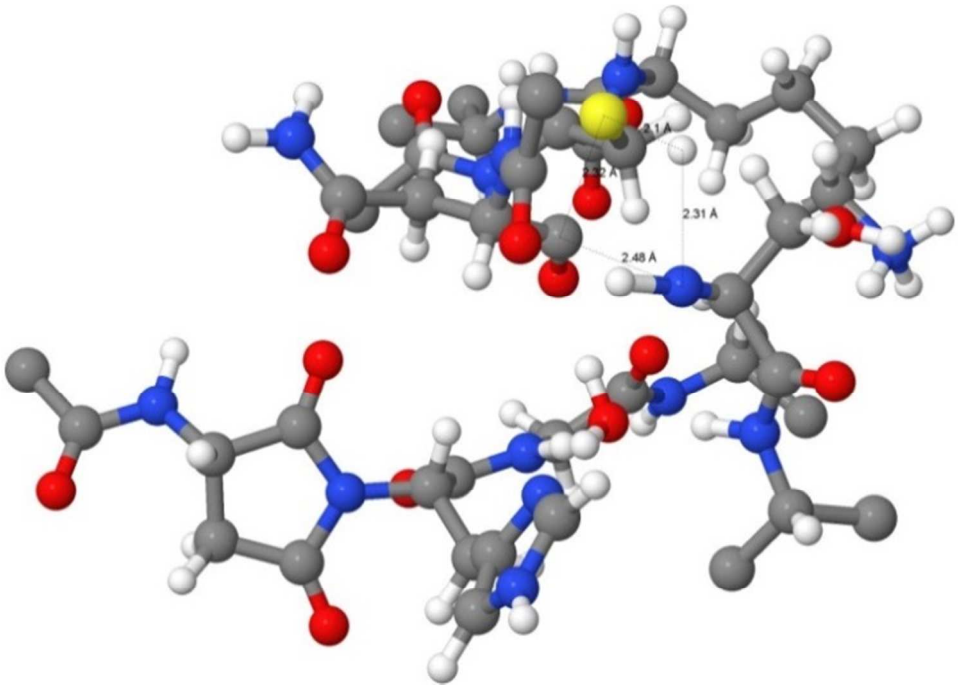

Figure S8. TS1 of the calculated reaction pathway for the E190K mutant.

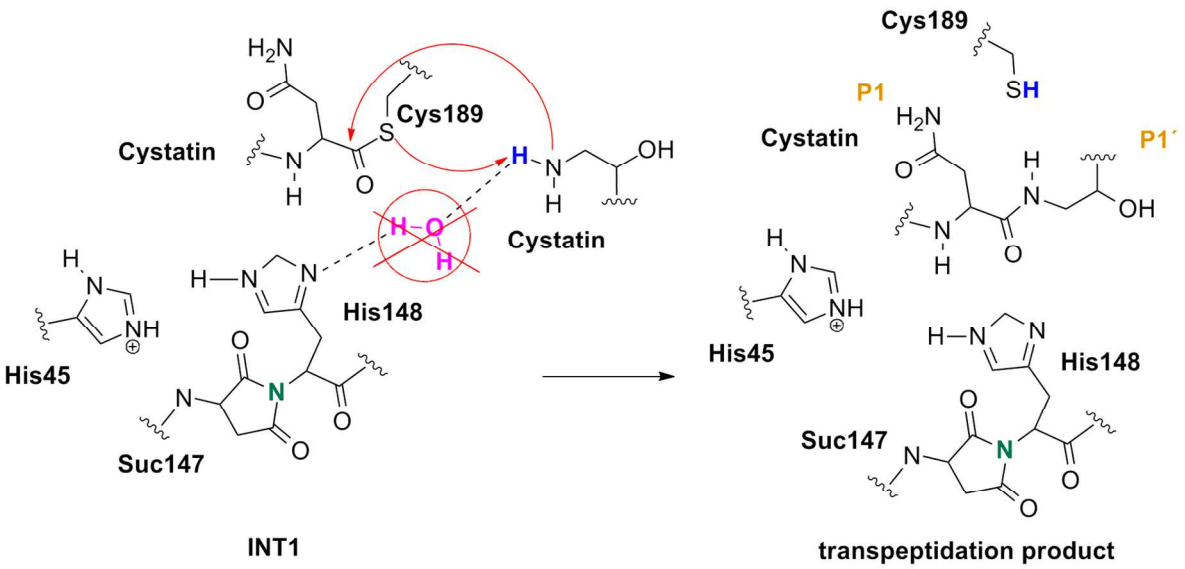

Figure S9. Reaction mechanism of transpeptidation

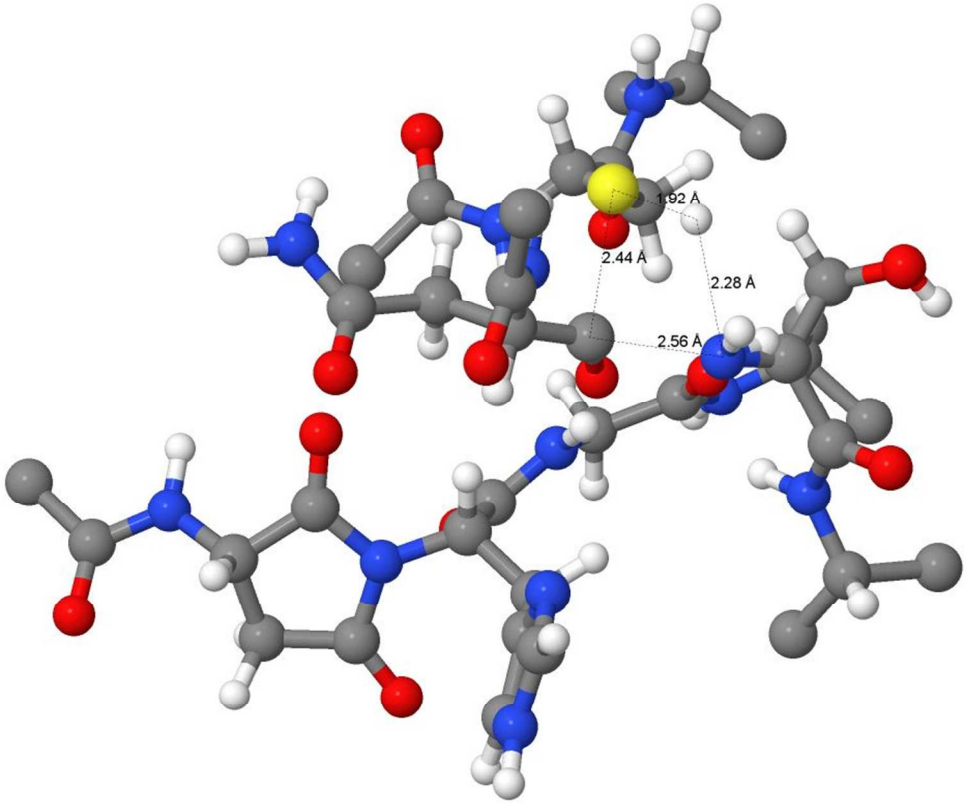

Figure S10. The structure of the transition state along the transpeptidation path.

References

(1) Dall, E.; Fegg, J. C.; Briza, P.; Brandstetter, H. *Angew Chem Int Ed.* 2015, *54*, 2917.

(2) *Molecular Operating Environment (MOE)*, 2016.08; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2017.

(3) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. *Nucleic Acids Res* 2012, *40*, W537.

(4) Baum, I.; Elsässer, B.; Schwab, L. W.; Loos, K.; Fels, G. *ACS Catal* 2011, *1*, 323.

(5) Elsässer, B.; Valiev, M.; Weare, J. H. *J Am Chem Soc* 2009, *131*, 3869.

(6) Elsässer, B.; Dohmeier-Fischer, S.; Fels, G. *J Mol Model* 2012, *18*, 3169.

(7) Elsässer, B.; Fels, G. *Phys Chem Chem Phys* 2010, *12*, 11081.

(8) Elsässer, B.; Fels, G.; Weare, J. H. *J Am Chem Soc* 2014, *136*, 927.

(9) Elsässer, B.; Schoenen, I.; Fels, G. *ACS Catal* 2013, *3*, 1397.

(10) Valiev, M.; Garrett, B. C.; Tsai, M. K.; Kowalski, K.; Kathmann, S. M.; Schenter, G. K.; Dupuis, M. *J Chem Phys* 2007, *127*, 51102.

(11) Elsässer, B.; Fels, G. *Phys Chem Chem Phys* 2010, *12*, 11081.

(12) Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; de Jong, W. A. *Comput Phys Commun* 2010, *181*, 1477.

(13) Henkelman, G.; Jonsson, H. *J Chem Phys* 2000, *113*, 9978.

(14) Valiev, M.; Yang, J.; Adams, J. A.; Taylor, S. S.; Weare, J. H. *J Phys Chem B* 2007, *111*, 13455.

(15) Dall, E.; Brandstetter, H. Acta Crystallogr Sect F Struct Biol Cryst Commun 2012, 68, 24.

(16) Wang, J.; Wilkinson, M. F. *Biotechniques* 2001, *31*, 722.

(17) Freier, R.; Dall, E.; Brandstetter, H. *Scientific reports* 2015, *5*, 12707.

(18) Dall, E.; Fegg, J. C.; Briza, P.; Brandstetter, H. *Angew Chem Int Ed Engl* 2015, *54*, 2917.

(19) Dall, E.; Brandstetter, H. *P Natl Acad Sci USA* 2013, *110*, 10940.