KDCR homework

Nicolás Rojo Esteban

1 Introduction

Hybrid quantum mechanics/molecular mechanics (QM/MM) simulation is a powerful approach that combines the accuracy of quantum chemistry with the efficiency of classical force fields en.wikipedia.org . By treating only the region of interest (such as an enzyme's active site) with a quantum mechanical method and the rest of the system with molecular mechanics, QM/MM enables the study of chemical reactions in large biomolecular systems that would be imposible to model with full quantum mechanics[1]. This multiscale technique is especially well-suited for enzymatic reactions, where covalent bond-breaking/forming events are localized to the active site, and it has become a standard tool for exploring enzyme mechanisms. In our QM/MM implementation, we employ the additive scheme with electrostatic embedding (the quantum region is polarized by the classical environment). The system of interest in this work is the NS3/NS4A protease of the Hepatitis C Virus (HCV), this protease is essential for protein processing and replication. NS3/4A is a protease containing a catalytic triad (His-57, Asp-81, and Ser-139) responsible for cleaving the peptide bonds at specific junctions of the HCV polyprotein. The catalytic mechanism of NS3/4A follows a two-step acylation–deacylation pathway observed in other serine proteases

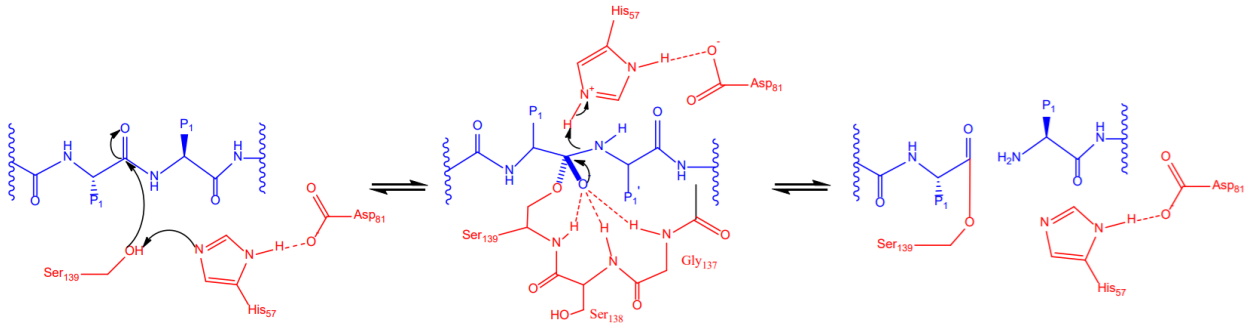


Figure 1: Proposed mechanism for the acylation process catalyzed by the hepatitis C virus NS3/NS4A protease with the NS5A/5B substrate.[2]

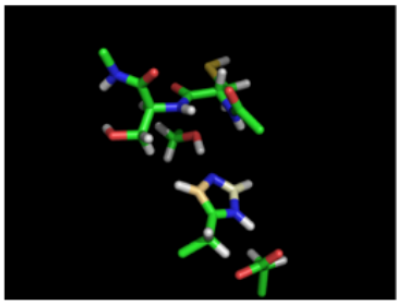
In the acylation step (which is the focus of this work), His-57 acts as a general base (assisted by Asp-81) to deprotonate the Ser-139 hydroxyl, thereby activating Ser-139 as a nucleophile. The Oγ of Ser-139 attacks the carbonyl carbon of the peptide bond in the substrate, leading to formation of a tetrahedral intermediate stabilized by the enzyme's oxyanion hole [3]. Then, this intermediate collapses with the cleavage of the peptide bond, releasing the N-terminal fragment of the substrate and forming a covalent acyl-enzyme intermediate in which the substrate's carbonyl is esterified to Ser-139. Later, in the deacylation step, this acyl-enzyme intermediate would be hydrolyzed by a water molecule to regenerate the free enzyme, but in this study we restrict our analysis to the acylation half-reaction of the protease.

2 Computational details

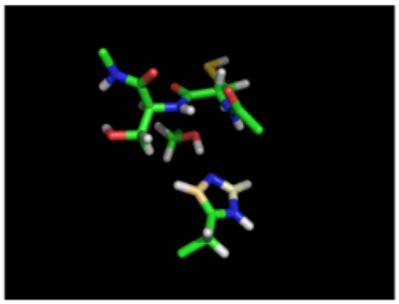
In this work, we investigate the acylation step of NS3/4A protease with one of its natural substrates (a peptide representing a native cleavage site) using hybrid QM/MM reaction simulations. All simulations were carried out with the CHARMM[4] program using its additive QM/MM scheme with full electrostatic embedding of the QM region in the MM environment. The quantum subsystem was treated with the self-consistent charge density-functional tight-binding (SCC-DFTB) method[5], an approximate density functional theory approach that offers a favorable accuracy-to-cost ratio. The remainder of the protease–substrate system was described by the CHARMM22[6] molecular mechanics force field for proteins. To properly handle covalent bonds crossing the QM/MM boundary (for example, bonds connecting a QM region atom to an MM region atom), we employed the generalized hybrid orbital (GHO) technique[7]. The GHO method effectively saturates the QM–MM boundary with a hybrid orbital construction, enabling a seamless coupling of the quantum and classical regions without disrupting the covalent connectivity of the protein.

3 Results and discussion

A key aspect of our study is examining how the choice of the QM region affects the simulated reaction. It is generally desirable to include all key catalytic groups (such as the full Ser–His–Asp triad in serine proteases) in the QM region to capture their electronic contributions; however, as the QM region is increased, the computational cost also increases. To check which is the ideal QM region, we carried out four separate potential energy scans for the acylation step, each with a different QM region size.



QM/MM 1



QM/MM 2

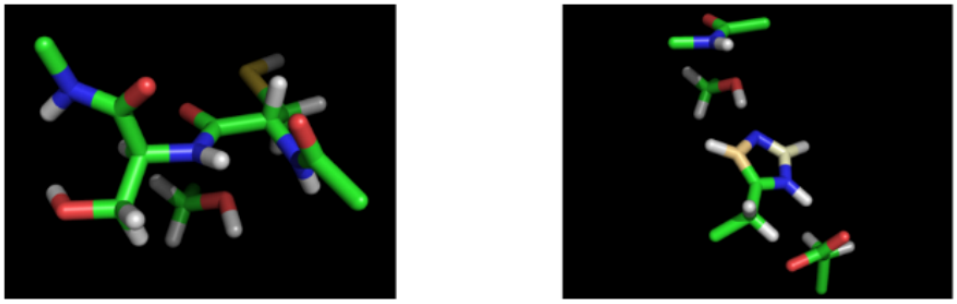


Figure 2: Regions of the system included in the QM zone in each scan

1. Scan 1: QM region includes the full catalytic triad Ser-139, His-57, and Asp-81 along with the adjacent substrate residues around the peptide bond. This represents the most extensive QM region of the four, incorporating all catalytic residues and relevant substrate atoms.

2. Scan 2: Only Ser-139 and His-57 from the triad are treated with QM, while Asp-81 is kept in the MM region. This test examines the effect of omitting the Asp acid from the quantum treatment, to see if its influence can be adequately represented by the force field alone.

3. Scan 3: QM region includes only the catalytic Ser-139 side chain and the substrate fragment containing the scissile peptide bond. In this QM setup, neither the base (His-57) nor the acid (Asp-81) are in the QM region.

4. Scan 4: QM region is similar in the scan 1 having the whole catalytic triad (Ser-139, His-57, and Asp-81) but the difference is that the part of the substrate that is included in the QM zone is reduced to the atoms that form the peptide bond exclusively.

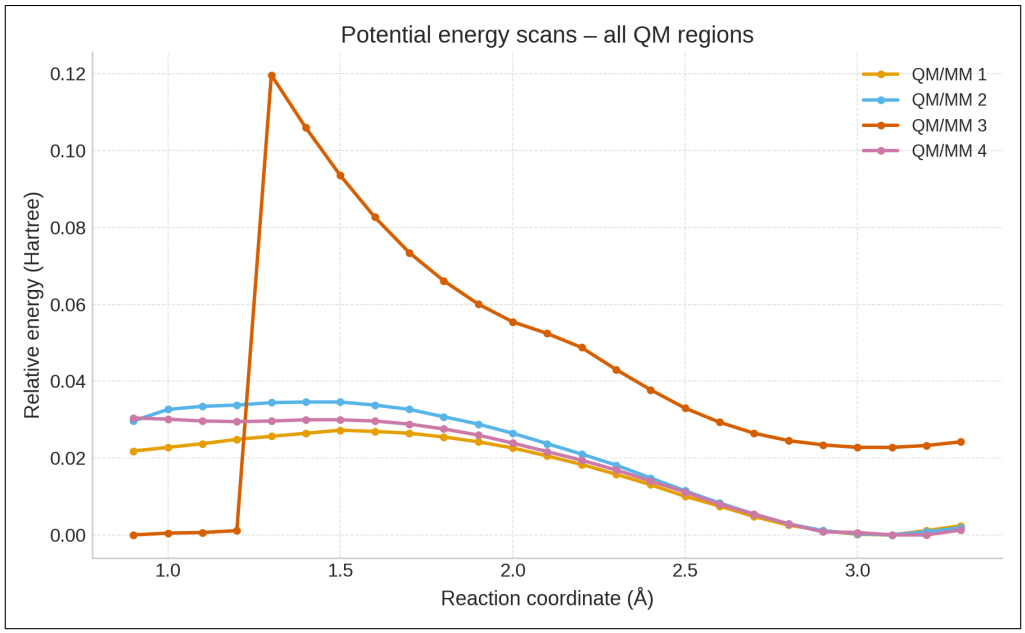


Figure 3: Potential energy scans of the acylation step for all four QM regions.

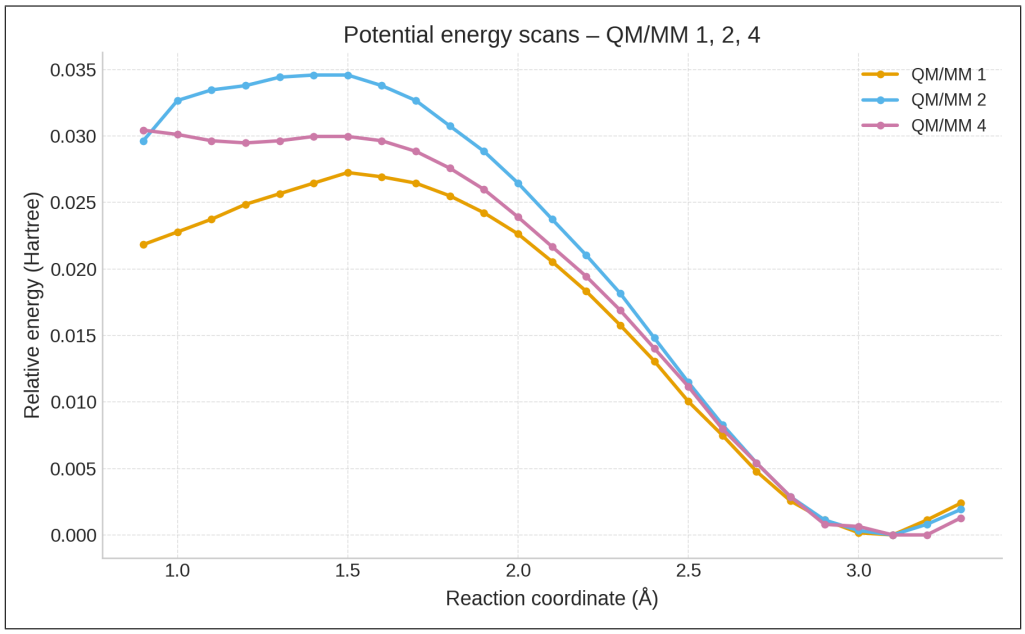


Figure 4: Scans including only QM/MM 1, 2 and 4. Scan 3 is omitted for clarity because its barrier is an order of magnitude larger.

Table 1: Summary of the reaction barriers obtained from the potential energy scans

Scan	QM residues	$\Delta E^{\ddagger}$ / Hartree
QM/MM 1	Ser–139, His–57, Asp–81 + substrate	17
QM/MM 2	Ser–139, His–57 (Asp in MM)	22
QM/MM 3	Ser–139 + substrate only	75
QM/MM 4	Minimal substrate atoms	19

The catalytic triad of the NS3/NS4A protease must be included entirely in the quantum region to obtain a correct description of the acylation step. When Asp-81 is described classically (scan 2) the barrier increases by ~ 5 kcal mol^{-1} because the anionic Asp$^-$ can no longer polarise His–57 during proton abstraction. If His-57 is not included in the QM region (scan 3) the reaction profile exhibits a wall of ~ 75 kcal mol^{-1}. A minimal QM model of the substrate (scan 4) lowers the barrier but it produces an artificial reaction intermediate. The most reliable description is therefore *scan 1*, where Ser–139, His–57, Asp–81 and a big region of the substrate including the the peptide bond are all treated at the SCC–DFTB level: the resulting potential barrier (~ 17 kcal mol^{-1}) is in the expected enzymatic range.

When performing a series of QM/MM energy evaluations, each computed energy E_i contains an arbitrary constant offset arising from the internal Hamiltonian reference energies. As a result, the energies cannot be plotted directly to obtain the ΔE barriers or to compare across the different scans.

To eliminate these offsets, we apply the FORTRAN program `setzero.e`, which executes the following steps:

1. Reads the list of QM/MM energies $\{E_i\}$ at each coordinate.
2. Choose a reference point j whose energy E_j defines the zero of the relative energies.
3. Computes corrected energies
$$E_i^{\mathrm{rel}} = E_i - E_j$$
4. Write out the adjusted energy file.

By plotting E_i^{rel} instead of the raw E_i, all scans share a common zero and the barrier heights $\max(E_i^{\mathrm{rel}})$ directly reflect the true relative energy differences.

The progress of the reaction is represented by four distances, two for the nucleophilic attack/proton-abstraction step and two for the peptide-bond cleavage/proton-donation step. They are defined as:

$$\chi_1 = d\big(O_s, C_p\big) \qquad \text{(Ser-139 O}_\gamma\text{ – carbonyl C)}\,, \tag{1}$$
$$\chi_2 = d\big(H_s, N_\varepsilon\big) \;-\; d\big(O_s, H_s\big) \qquad \text{(His-57 N}_\varepsilon\text{ – Ser-139 OH proton)}\,, \tag{2}$$
$$\chi_3 = d\big(C_p, N_p\big) \qquad \text{(carbonyl C – peptide N)}\,, \tag{3}$$
$$\chi_4 = d\big(H_s, N_p\big) \;-\; d\big(H_s, N_\varepsilon\big) \qquad \text{(proton to leaving-group N)}\,. \tag{4}$$

- Single distances (e.g. $\chi_1 = d(O_s, C_p)$, $\chi_3 = d(C_p, N_p)$) directly track bond formation or cleavage.

- Proton transfers involve the coordinates of two changing bonds (breakage of one H–X bond and formation of another). By subtracting distances,

$$\chi_2 = d(H_s, N_\varepsilon) - d(O_s, H_s)\,, \quad \chi_4 = d(H_s, N_p) - d(H_s, N_\varepsilon)\,,$$

  we construct an antisymmetric coordinate that evolves smoothly from negative values (proton on the donor) to positive values (proton on the acceptor). This eliminates the need for separate potentials on each bond and ensures a smooth reaction path.

4 Bibliography

References

[1] A. Warshel and M. Levitt, *J. Mol. Biol.* **103**, 227–249 (1976).

[2] J. Á. Martínez-González, M. González, L. Masgrau, and R. Martínez, *ACS Catal.* **5**, 246–255 (2015).

[3] V. L. Schramm, *Curr. Opin. Struct. Biol.* **21**, 668–673 (2011).

[4] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, *et al.*, *J. Comput. Chem.* **30**, 1545–1614 (2009).

[5] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, *Phys. Rev. B* **58**, 7260–7268 (1998).

[6] A. D. MacKerell Jr., D. Bashford, M. Bellott, R. L. Dunbrack Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, *et al.*, *J. Phys. Chem. B* **102**, 3586–3616 (1998).

[7] J. Gao, P. Amara, C. Alhambra, and M. J. Field, *Mol. Phys.* **101**, 2695–2714 (2003).