

# Exploring Proprotein Convertase Subtilisin/Kexin 9 (PCSK9) Autoproteolysis Process by Molecular Simulations: Hints for Drug Design

Jacopo Sgrignani,<sup>[a]</sup> Enrico M. A. Fassi,<sup>[b]</sup> Carmen Lammi,<sup>[b]</sup> Gabriella Roda,<sup>[b]</sup> and Giovanni Grazioso\*<sup>[b]</sup>

<sup>[a]</sup> Istituto di Ricerca in Biomedicina (IRB), Università della Svizzera Italiana (USI), Via V. Vela 6, CH-6500, Bellinzona, Switzerland.

<sup>[b]</sup> Dipartimento di Scienze Farmaceutiche, Università degli Studi di Milano, Via L. Mangiagalli 25, 20133 Milan, Italy.

## Abstract

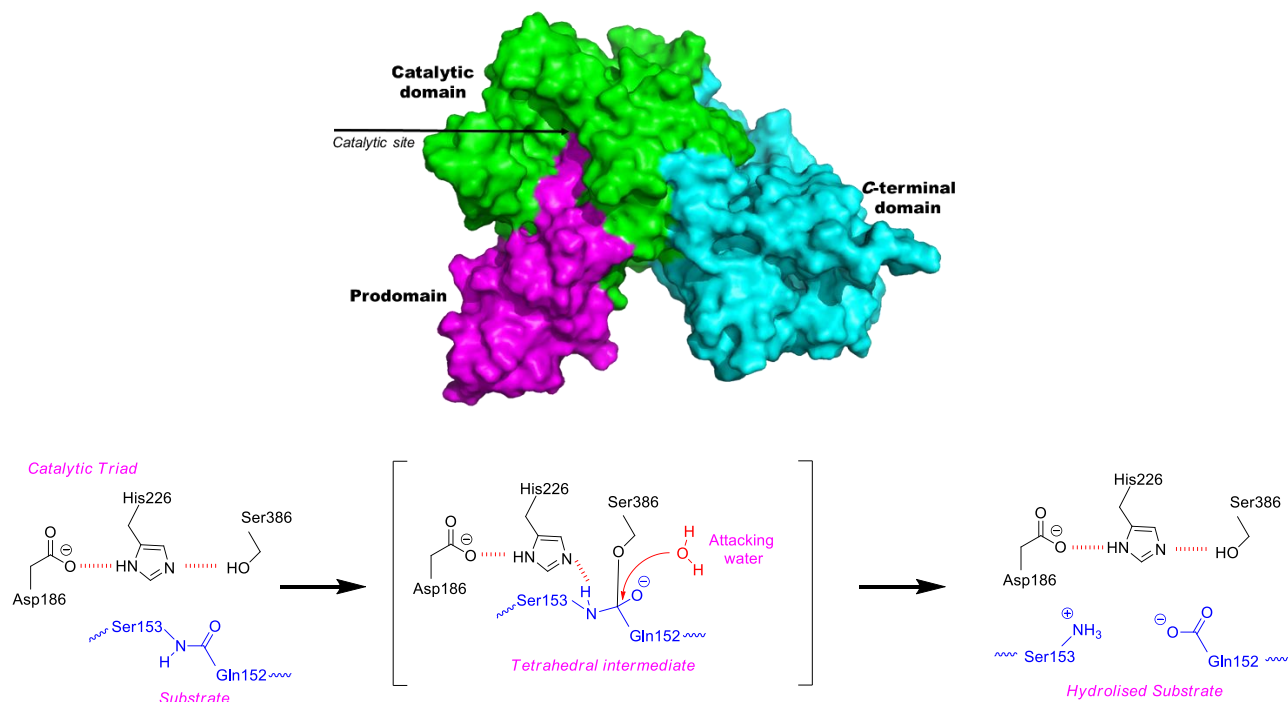
Proprotein convertase subtilisin/kexin 9 (PCSK9) is a notable target for the treatment of hypercholesterolemia because it regulates the population of the low-density lipoprotein receptor (LDLR) on liver cells. The PCSK9 zymogen is a serine protease that spontaneously undergoes a double self-cleavage step. The available X-ray structures depict the PCSK9 mature state, but the atomic details of the zymogen state of the enzyme are still unknown. Additionally, why the protease activity of PCSK9 is blocked after the second autoprocessing step remains unclear, as this deviates from other members of the PCSK family. By performing constant-pH molecular dynamics (MD) simulations, we investigated the protonation state of the catalytic triad of PCSK9 and found that it strongly influences the catalytic properties of the enzyme. Moreover, we determined the final step of the maturation process by classical and steered MD simulations. This study could facilitate the identification of ligands capable of interfering with the PCSK9 maturation process.

## INTRODUCTION

Proprotein convertase subtilisin/kexin 9 (PCSK9) regulates plasma low-density lipoprotein cholesterol (LDL-c) levels by direct interaction with the LDL receptor (LDLR).<sup>[1]</sup> Their protein–protein interaction (PPI) on the surface of the hepatocytes directs the LDLR to degradation via the lysosome pathway. As a consequence, the physiological activity of PCSK9 leads to an increment in the circulating cholesterol levels by decreasing the LDLR population on liver cell membranes. For this reason, PCSK9 inhibition represents a valuable therapeutic strategy for the treatment of hypercholesterolemia and consequent coronary heart disease.<sup>[2,3]</sup> Evolocumab<sup>[4]</sup> and alirocumab<sup>[5]</sup>, two monoclonal antibodies (mAbs) endowed with PCSK9 inhibitory activity, have been released on the market to treat statin-resistant hypercholesterolemia. However, this type of therapy is characterized by high cost and low patient compliance; therefore, both industry and academic research activities are focused on the development of alternative therapeutic strategies like the usage of small molecules,<sup>[6]</sup> peptides,<sup>[7-11]</sup> or peptidomimetics<sup>[12,13]</sup> capable of impairing the PCSK9/LDLR PPI. Additionally, another promising strategy to decrease the circulating level of PCSK9 is inhibition of PCSK9 synthesis by means of small interfering RNA, such as inclisiran,<sup>[14]</sup> or the use of small molecules capable of blocking the maturation process of the enzyme. PCSK9 is a 72-kDa zymogen comprised of 692 amino acids, which are synthesized in the endoplasmic reticulum as a glycoprotein and are unable to bind to LDLR. The PCSK9 zymogen is constituted by a signal sequence (residues 1–30), an *N*-terminal domain (prodomain, residues 31–152), a catalytic domain (residues 153–425), and a *C*-terminal (or *V*) domain (residues 426–692) (**Figure 1**).<sup>[15]</sup> By the hydrolysis of the signal peptide (first autocatalytic step), the zymogen is converted into proPCSK9. Then, by a second autoprocessing step in which the amide bond between Gln152 and Ser153 is hydrolyzed, the proPCSK9 is converted into its mature state (PCSK9) and is secreted outside the liver cells.<sup>[16]</sup> Considering its zymogenic nature and the competitive inhibition of its catalytic activity by small molecules, the consequent impairment of the maturation process represents one possible drug design strategy. In 2016, Salowe *et al.*<sup>[17]</sup> reported a biochemical method for the discovery of covalent autocleavage inhibitors, describing the kinetic properties of a lactone tool compound. To the best of our knowledge, no other irreversible inhibitors of PCSK9 have been reported in the literature. Thus, in order to design new compounds, insights into the autoprocessing mechanism are still needed. In light of these results, new strategies to target PCSK9 should be identified.

The available X-ray structures depict PCSK9 (**Figure 1**, residues 61–692) in the mature state, but no structural information about the other states is available. Curiously, all the available X-ray structures display the prodomain still bound to the catalytic domain and to the prodomain *C*-terminal end (Gln152) at a distance of about 20 Å from the catalytic domain *N*-term (Ser153). Both are spaced out by a hairpin created by residues 366–383. Since Gln152 and Ser153 are covalently bound in the proPCSK9 state, a huge conformational modification of the enzyme could be expected throughout the second maturation step. Moreover, unlike other enzymes belonging to the same family of subtilisin/kexin proteins, the protease activity of PCSK9 is interrupted after the second phase of self-catalysis.

This avoids a third autocatalytic cleavage and protects the surrounding proteins from the protease activity of PCSK9. However, while this last aspect could be explained by the fact that the prodomain occupies the catalytic site, preventing access to other proteins, it is unclear why a third step of hydrolysis does not occur.<sup>[18]</sup>



**Figure 1.** Graphical representation of the proPCSK9 model and hydrolysis reaction performed by the proPCSK9 zymogen.<sup>[19]</sup>

Therefore, we report on computational studies aimed at predicting the conformational modifications of PCSK9 after the second auto-maturation process, providing mechanistic insights into the Ser153 *N*-term removal from the catalytic site. Additionally, constant-pH molecular dynamics (CpHMD) simulations were performed because we predict that a change in the protonation state of the charged amino acids located in the PCSK9 catalytic site could affect the progression of the self-proteolysis process. Our computational results could suggest strategies to design new PCSK9 inhibitors using an approach that does not target the PCSK9/LDLR PPI.<sup>[7-9,12]</sup>

## COMPUTATIONAL METHODS

**ProPCSK9 models setup.** The low-resolution X-ray structures disclose atomistic details of PCSK9 from residue 61 to 682.<sup>[11,20,21]</sup> These studies revealed the conformation of PCSK9 in the mature state, in which the prodomain was already cleaved from the catalytic domain, and the Ser153 *N*-terminus was a distance of about 20 Å from the Gln152 *C*-terminus. Furthermore, the presence of highly flexible loops in the PCSK9 sequence led to incomplete X-ray structures (i.e., sequence 173–178 has not been solved). Therefore, the Modeller9.21 algorithm<sup>[22]</sup> was employed, and the human PCSK9 sequence stored in the UniProt database with code Q8NBP7 and the PCSK9 X-

ray crystal structure stored in Protein Data Bank with code 2P4E<sup>[20]</sup> were used as templates to build a reliable and complete proPCSK9 model. In this model (named proPCSK9), the Gln152 and Ser153 were still bound by the amide bond, representing a model of the immature state of the enzyme. One hundred models were produced by Modeler9.21<sup>[22]</sup> and scored by the atomic distant-dependent potential known as discrete optimized protein energy (DOPE).<sup>[23]</sup> The best quality model attained from Modeler was geometrically optimized and equilibrated by 750 ns of MD simulations using Amber19<sup>[24]</sup> following the protocols described in our previous works.<sup>[8,9]</sup> Moreover, in order to gain insights into the biochemical process occurring immediately before the catalytic cleavage of proPCSK9, an additional model (named proPCSK9-cut) was created. This model, in which Gln152-Ser153 had just been cleaved using the molecular modeling tools available on the Maestro software,<sup>[25]</sup> was used to predict the PCSK9 conformational modifications occurring immediately after the second self-cleavage step. In other words, energy minimization and MD simulations in the proPCSK9-cut model allowed us to determine the enzyme maturation process. This model also served to predict the pKa values of the amino acids that play a critical role in the catalytic site.

**Steered MD simulations.** Steered MD (SMD) simulations<sup>[26]</sup> are a well-validated tool in structure-based drug design. They can provide the atomistic description and potential of mean force (PMF) profile of the simulated process (i.e., the channel opening and ligand unbinding process). Here, we performed SMD simulations to estimate the PMF profile of the 366–383 hairpin opening because we believe this motion is essential for the maturation of the enzyme. A harmonic time-dependent potential was added to the standard Hamiltonian of a collective variable controlling the phenomenon under investigation. In these calculations, the distance between the enzyme core (center of mass of the C $\alpha$  atoms of Gln190, Arg194, and Arg237) and the center of mass of the C $\alpha$  atoms of residues Ser373–Ser376 belonging to the 366–383 hairpin was used as descriptor. To avoid any protein distortions, the pulling velocity was set to 0.006 Å/ps, with a spring constant of 0.5 kcal/mol·Å<sup>2</sup>. The hairpin opening was simulated from 9.3 to 40.3 Å. The applied force and the external work performed on the system were calculated by applying the Amber scripts “ASMD.py” to determine which of the 50 replicas of the SMD simulations’ work values were the closest to the Jarzynski average. Additionally, “create.PMF.py” was used to depict the PMF profile. SMD simulations were conducted on the proPCSK9 model in which Gln152 and Ser153 were still bound.

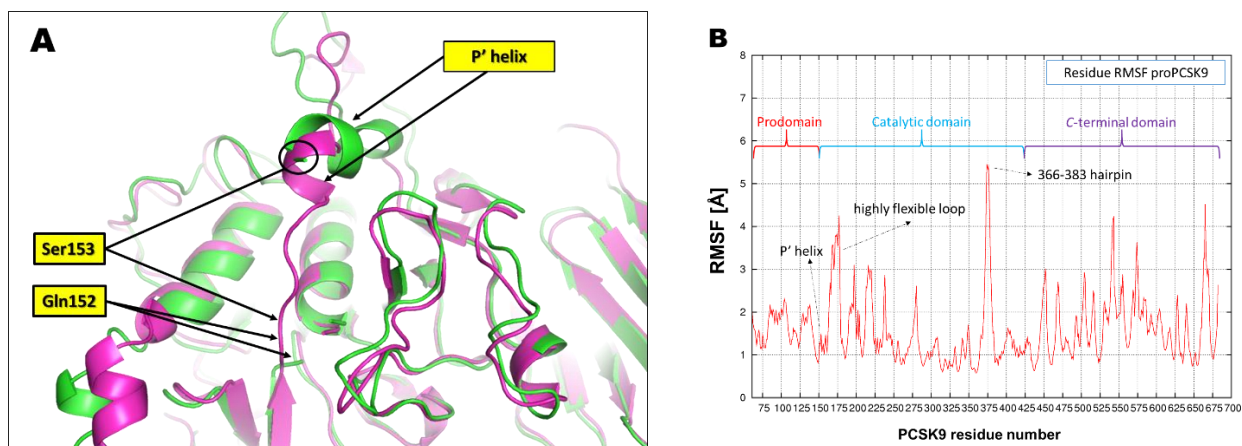
**Umbrella sampling simulations.** The energy barrier between the closed and supposedly open states of the 366–383 hairpin was investigated by performing umbrella sampling (US) simulations using the pmemd.cuda tool of the Amber 2019 package.<sup>[24]</sup> The distance between the two mass centers used in the SMD simulations was chosen as a collective variable (CV). Ten windows of CV values from 9.15 to 22.00 Å were explored (**Figure S1**, Supporting Information), constraining the system to the specific CV values by harmonic constant ranging from 1.5 to 2.5 kcal mol<sup>-1</sup> Å<sup>-2</sup>. Each US window ran for 50 ns, reaching a total simulation time of 500 ns. The free energy profile was estimated by the weighted histogram analysis method, as implemented in the WHAM program

developed by Grossfield Laboratory.<sup>[27]</sup> The statistical uncertainty of the free energy profiles was estimated by Monte Carlo bootstrap error analysis.

**CpHMD simulations.** CpHMD simulations<sup>[28]</sup> were conducted using the ff14SB Amber force field<sup>[29]</sup> to represent the protein atoms and the TIP3P model<sup>[30]</sup> to simulate the water environment. Titration curves were produced by performing 12 simulations at pH values from 0 to 12, for a total of 600-ns-long CpHMD simulations. During these, Asp186, His226 (both members of the catalytic triad), and the residues in their proximity (Asp224 and His229) were titrated. Moreover, to better characterize the protonation state of His226, 250-ns-long MD simulations were conducted at a constant pH of 7.2 in both the proPCSK9 and proPCSK9-cut models. The Amber script “cphstats” was used to parse the cpout files and to predict the residue protonation state and pKa values, creating files with the titration statistics for all simulations. For greater depth, “cphstats” were also used to compute the fraction of snapshots that the system spent in each state for each residue titrated during the simulations, resulting in a file named “populations.dat.”

## RESULTS AND DISCUSSION

**proPCSK9 models.** To sample the conformational space of PCSK9 in the state preceding the second self-cleavage process, the proPCSK9 model—in which the prodomain was still bound to the PCSK9 catalytic and V domains (sequence 153–682)—was created by means of the Modeller algorithm (see Methods section).<sup>[22]</sup> The loops not solved by X-ray studies were properly sculpted inside the modeling procedure and, among 100 putative proPCSK9 coarse models, the best model was optimized by 750-ns-long unbiased MD simulations. Strikingly, the obtained proPCSK9 model displayed an overall significant conformational stability, as explicated from the root mean square deviation (RMSD) plot (**Figure S2**, Supporting Information). Moreover, the visual inspection of the model revealed that the sequence shaping the P' helix (residues 153–166) in the mature state (**Figure 2A**) was partially folded below the 366–383 hairpin. This conformation permitted the close contact of Gln152-Ser153 with the catalytic triad (Asp186, His226, and Ser386) responsible for the amide bond hydrolysis during the second self-cleavage step. Moreover, the analysis of the C $\alpha$  residue root mean square fluctuation (RMSF) plot (**Figure 2B**) evidenced that three proPCSK9 areas possessed the highest conformational flexibility: 1) the loop shaped by residues 169–178, 2) the 366–383 hairpin, and 3) the C-terminal domain.

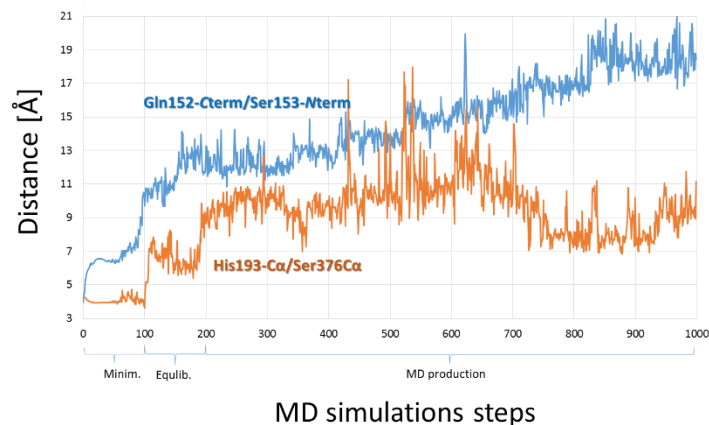


**Figure 2.** A) Predicted conformation of the P' helix in the proPCSK9 model (magenta) and in the X-ray structure 2P4E (green). The 366–383 hairpin is not shown for clarity. B) RMSF plot of the proPCSK9 C $\alpha$  atoms during the unbiased MD simulations.

The conformational mobility of some PCSK9 areas observed on the RMSF plot can be rationalized considering that the 169–178 residues belong to an unsolved region of the X-ray (indicating an area endowed with a high conformational freedom)<sup>[18,31]</sup> and that the flexibility of the C-terminal domain does not greatly influence the catalytic domain activity since it is not in close contact with the residues involved in the serine-proteases activity of PCSK9. In contrast, the high conformational mobility shown by the 366–383 hairpin was unexpected. However, considering that it sizes the catalytic site, we suppose that its conformational freedom is needed to assist the P' helix relocation subsequent to the auto-proteolysis process.

To validate this hypothesis, unbiased MD simulations were performed on a proPCSK9 model in which the Gln152-Ser153 peptide bond had just been cleaved (the proPCSK9-cut model). In these simulations, the 366–383 hairpin conformational states were sampled, and the P' helix relocation was determined.

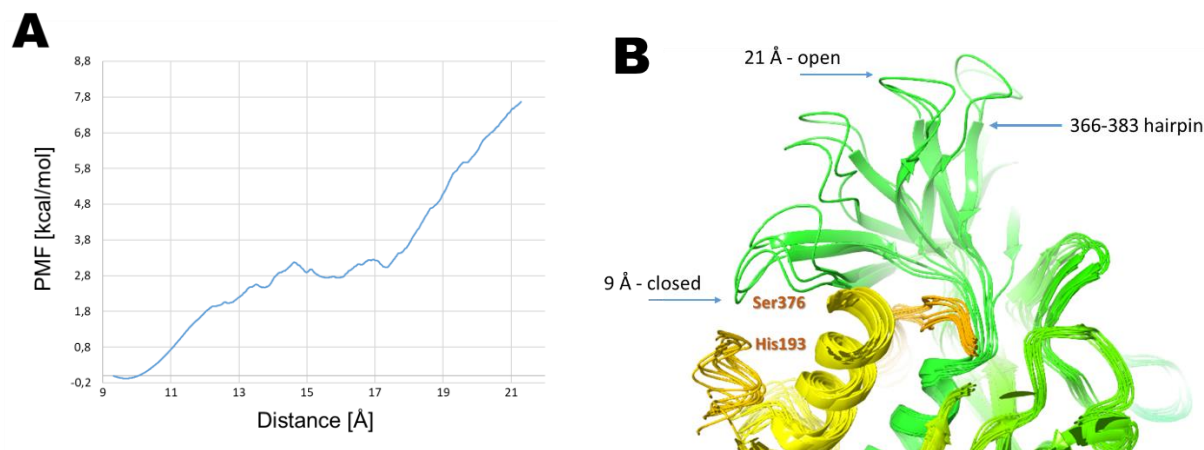
**MD simulations on the proPCSK9-cut model.** The 1- $\mu$ s-long MD simulation of the proPCSK9-cut model showed that despite the simulations starting from a conformation of Gln152-Ser153 (in which these residues were already close to each other during the minimization steps), the distance between the domain terms (Gln152-CO and Ser153-NH) increased to about 7 Å. Then, during 500-ns-long MD simulations, this distance progressively increased to the final crystallographic value of about 19 Å. Once this distance was reached, no further increases were observed during MD simulations (cyan line in **Figure 3**). Interestingly, these simulations revealed the P' segment progressive folds as  $\alpha$ -helix and the Ser153-NH relocation from the catalytic site by a slight opening of the 366–383 hairpin (orange line in **Figure 3**). In fact, as depicted in **Figure 3**, the His193C $\alpha$ -Ser376C $\alpha$  distance progressively increased together with the Gln152-Ser153 distance. Then, once Ser153 reached the optimal distance (Step 800 in **Figure 3**; about 19 Å), the 366–383 hairpin fluctuated at a value of 3 Å lower than the previous one.



**Figure 3.** The orange line depicts the 366–383 hairpin opening to allow the Ser153 relocation. The distance between the Ser376-C $\alpha$  (in the center of the hairpin) and the His193-C $\alpha$  (PCSK9 core) atoms was monitored during all steps of the 500-ns-long MD simulations. The blue line shows the progressive removal of the Ser153 N-term from the catalytic site.

However, the visual inspection of the MD trajectories permitted us to observe that the 366–383 hairpin opening is needed to allow the Ser153 N-term removal from the catalytic site. Consequently, to identify the presence of new 366–383 hairpin conformations besides the one depicted in X-ray studies, the conformational states of the 366–383 hairpin were sampled by SMD and US simulations. We hypothesize that an undescribed metastable conformational state of the enzyme could represent a target for structure-based drug design of a new generation of PCSK9 inhibitors.

**SMD and US simulations of the 366–383 hairpin.** SMD simulations enable sampling of rare events as biomolecular conformational changes on time scales accessible to MD simulations. Here, we sampled the 366–383 hairpin opening process, estimating the PMF of the system as a function of the distance between the hairpin and the PCSK9 core. This distance (see Methods for details) was progressively increased from the initial distance of 9.3 Å to 40.3 Å at a speed of 0.006 Å/ns. To acquire a satisfactory sampling, 50 SMD simulation replicas were run, and the Jarzynski equation<sup>[32]</sup> was applied to calculate the PMF associated with the hairpin opening progression (**Figure 4**).

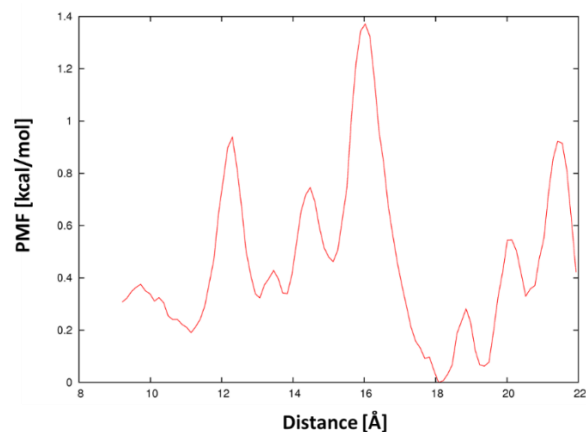


**Figure 4.** A) Plot of the PMF versus the distance between the hairpin and the PCSK9 core. B) Representative conformations of the 366–383 hairpin.

As shown in the PMF plot (**Figure 4A**), the presence of a saddle at the distance of 15–17 Å suggests that the hairpin could assume a metastable conformational state. In other words, the 366–383 hairpin spontaneously adopts the closed conformation, while a slightly open state could be shaped to allow for Ser153 *N*-term migration. Therefore, to evaluate the stability of the hairpin at the distance found in the saddle points, 500-ns-long unbiased MD simulations were performed on a trajectory frame showing the 366–383 hairpin at a distance of 16 Å from the enzyme core. Interestingly, these simulations displayed progressive hairpin closure, leading us to conclude that without any external bias (such as the one imposed by SMD), the hairpin assumed a stable closed conformation, as found in the mature (and crystallized) state of the enzyme. Consequently, the metastable conformational state identified on the PMF plot (**Figure 4A**) could represent only a transient hairpin conformation, allowing for the PCSK9 maturation (**Figure 4B**).

Finally, it is well-known that the results of SMD depend on factors such as the pulling velocity or the value of  $k$ . Therefore, to verify the reliability of our findings, we also performed US calculations.<sup>[27,33]</sup> Consistent with the SMD outcomes, US simulations indicated that no stable hairpin open state conformation was identified (**Figure 5**) since the barrier between the closed (distance 9 Å in Figure 5) and the open (distance ≈18 Å in Figure 5) state of the hairpin is close to 1.4 kcal/mol.



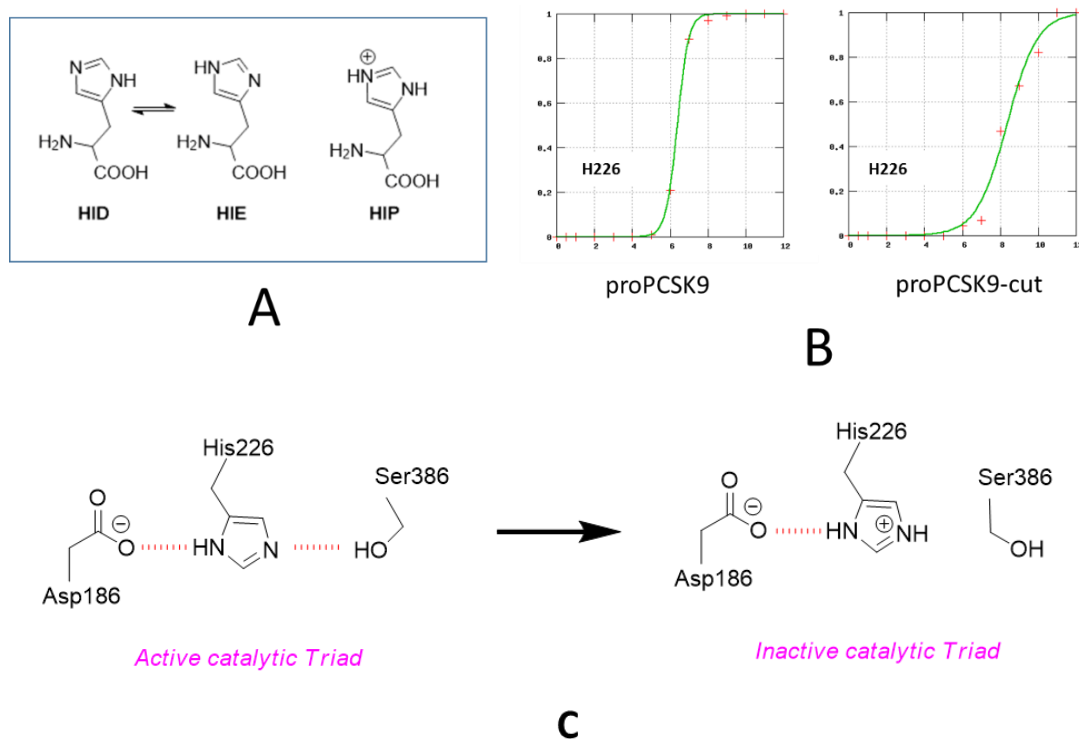


**Figure 5.** PMF obtained from US calculations using the hairpin distance from the enzyme *core* as the CV.

Nevertheless, another question remains unanswered in the maturation progression of PCSK9. Unlike other enzymes belonging to the same class as PCSK9, it is not clear why the protease activity of this enzyme does not proceed after the second maturation step. In other words, the reason why Gln152 is not cleaved from the prodomain during an additional (third) cleavage process remains undiscovered. We predicted that the answer to this question could be found by analyzing the protonation state of the catalytic triad (Asp186, His226, and Ser386) since the catalytic site, becoming accessible to the water environment after Ser153 relocation, could be strongly influenced by the presence of surrounding water molecules. To verify this hypothesis, CpHMD simulations were run.

**CpHMD simulations.** It is well-known that the protonation state of residues significantly affects the protein physiological functions. Experimental evidence of PCSK9 has confirmed this belief; in fact, PCSK9 shows five times greater affinity to LDLR at pH 5.5 than at pH 7.4.<sup>[18,20,34]</sup> Moreover, Ultsch *et al.*<sup>[36]</sup> recently proved that the 31–60 sequence of PCSK9 displays a pH-sensitive folding. In this study, we conducted CpHMD to predict the protonation states of the titrable residues located in the catalytic site of PCSK9 residues. CpHMD simulations were performed at pH 7.2—the pH value of the endoplasmic reticulum in which PCSK9 undergoes the auto-maturation process—for both the proPCSK9 and proPCSK9-cut models. In particular, we focused on the protonation state of Asp186, His226 (both members of the catalytic triad), and some residues in their proximity (Asp224 and His229). Our simulations suggest that the majority of the inspected residues retained the expected/starting protonation state in both models. In particular, Asp186 and Asp224 were negatively charged, and His229 was monoprotonated. In contrast, it was interesting to note that only His226 changed its protonation state during the transition from the proPCSK9 to proPCSK9-cut models. It is well-known that histidine can exist in three equivalent tautomeric forms, distinguished into HID (Amber residues nomenclature, **Figure 6A**), when the hydrogen atom belongs to the imidazole  $\delta$ -nitrogen atom; HIE, when the hydrogen is on the imidazole  $\epsilon$ -nitrogen atom; and HIP in an acidic environment, in which both nitrogen atoms are protonated (**Figure 6A**). In the HIP state, the residue is positively

charged. The pKa value of the imidazole ring is 6.00, but depending on the protonation/solvation state of amino acids in the surrounding environment, this value can be dramatically different. Intriguingly, in our simulations of the proPCSK9 model, His226 exhibited the HID/HIE state for 92% of the simulation time and the HIP state for 8% of the time. The calculated pKa was 6.14 (**Figure 6B**). In contrast, when CpHMD simulations were performed on the proPCSK9-cut model, His226 exhibited the HIP state for 97% of the simulation time and the HID state for the remaining time (less than 3%). In this case, a pKa of 8.7 was estimated (**Figure 6B**).



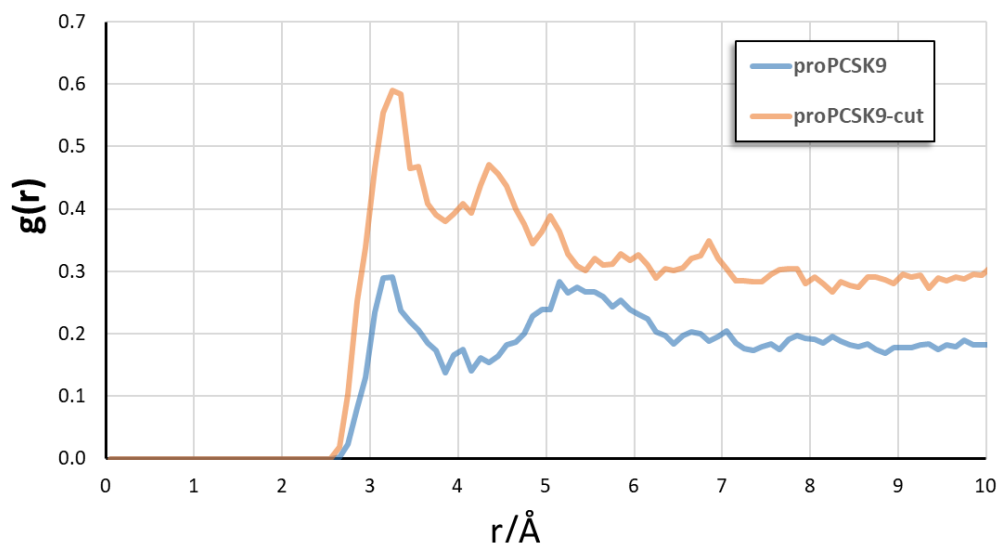
**Figure 6.** **A)** Protonation state of histidine and Amber nomenclature. **B)** Titration curves of His226 from pH 0 to 12. Calculations were performed on the proPCSK9 model, in which Gln152 and Ser153 were still bound, and on the proPCSK9-cut model, in which the amide bond joining both residues had just been hydrolyzed. **C)** Catalytic triad alignment before and after the Gln152-Ser153 hydrolysis.

Consequently, we can conclude that in the proPCSK9 state, His226 exists as an HID tautomer, allowing for the creation of an H-bond with the hydroxyl group of Ser386, while the hydrogen atom on the  $\delta$ -nitrogen establishes an H-bond with the carboxylic group of Asp186 (**Figure 6C**). By this H-bond network, the catalytic triad is perfectly aligned, and the serine-protease activity of PCSK9 is fully active. Then, once the Ser153 is cleaved (simulations performed on the proPCSK9-cut model), the *N*-terminus relocation from the catalytic crevice exposes the catalytic triad to the water environment, leading to the variation in the protonation state of His226 from HID/HIE to HIP. In the conformational state in which PCSK9 is ready to bind to LDLR, His226 cannot create a stable H-bond with Ser386, the catalytic triad is not aligned, and consequently, PCSK9 cannot exert further serine-protease activity. Moreover, in our simulations, it was observed that His226 created a salt bridge with the charged

Gln152 C-terminal end, contributing to the anchoring of the prodomain on the PCSK9 catalytic domain. The creation of this salt bridge, together with the changes in the protonation state of His226, explains why the PCSK9 prodomain is protected from further self-hydrolysis reactions, contributing to making PCSK9 safe for other proteins present in the intracellular environment.

To better understand the reasons for the shift in the pKa value of His226, we analyzed the simulation outputs with a focus on the microenvironment surrounding the catalytic triad, particularly the solvent accessibility of the single residue. For histidine residues, many studies have demonstrated the existence of a relationship between the pKa and its solvent accessibility, expressed as solvent accessible surface area (SASA).<sup>[37-39]</sup> The results of these investigations have shown that a higher solvent accessibility corresponds to a higher pKa of the basic residue. Consequently, we computed the average SASA of His226 on the proPCSK9 and proPCSK9-cut models by `gmx_sasa` script (implemented in the GROMACS package) using 100 frames extracted from the CpHMD simulations. The results indicated that the His226 SASA on the cleaved enzyme (proPCSK9-cut model) was 32 Å<sup>2</sup> higher than that of the zymogen proPCSK9. In fact, SASA values of 43.5 Å<sup>2</sup> (mean err. = 2.3) and 11.3 Å<sup>2</sup> (mean err. = 1.6) were predicted for the proPCSK9-cut and proPCSK9 models, respectively. These data further confirm that in the cleaved enzyme, the His226 is more accessible to the solvent and therefore more prone to assume the HIP state.

Moreover, the spherical atomic radial distribution function  $g(r)$  was calculated between the nitrogen atoms of the His226 imidazole ring and the oxygen atoms of the solvent molecules over the MD trajectories of CpHMD simulations of both PCSK9 maturation states (Figure 7). The peaks in Figure 7 reveal how the density of water varies as a function of distance from the reference atoms (the nitrogen atoms of the His226 imidazole ring). The attained results highlight that the water density around His226 in the proPCSK9-cut model was double that predicted for proPCSK9, additionally confirming that the auto-catalytic cut of the enzyme exposes the catalytic triad to the solvent.<sup>[40]</sup>



**Figure 7.** Radial pair distribution function data for the proPCSK9 (orange line) and proPCSK9-cut (cyan line) models. The  $g(r)$  between the His226 side chain and the oxygen atoms of the solvent molecules was selected to acquire this plot. These data were attained by means of the VMD  $g(r)$  script.<sup>[41]</sup>

## CONCLUSIONS

PCSK9 plays a crucial role in the regulation of cholesterol homeostasis, constituting one of the most promising targets for the identification of drugs against hypercholesterolemia. For this reason, Xu *et al.*<sup>[5]</sup> described PCSK9 as “a Holy Grail in the world of lowering LDL-C.” Until now, several inhibition strategies have been adopted to inhibit PCSK9, although the actual drugs in commerce are expensive mAbs. In this paper, we performed MD simulations (classical, SMD and US) to shed light on the PCSK9 maturation process with the goal of identifying new PCSK9 features that can be targeted by rational drug design. Starting from a complete PCSK9 model in the immature state (proPCSK9), we determined that PCSK9 conformational modification leads to the mature state of the enzyme, in which the P’ helix is shaped, and Gln152 and Ser153 are 20 Å apart. We speculate that the maturation process involves the putative opening of the 366–383 hairpin that, immediately after the Ser153-*N*-term passage, returns to a closed conformation. This finding could open the way for the design of drugs capable of stabilizing the open state of the enzyme. In fact, a new generation of PCSK9 inhibitors capable of stabilizing the 366–383 hairpin in the open conformation could avoid the PCSK9/LDLR interaction, triggering the hypocholesterolemic effect. In this framework, the transient conformational state of proPCSK9 could be used as a target for the molecular docking of a virtual library of compounds (small molecules as well as small peptides). In these calculations, the centroid of the Ser153, Pro155, His229, and Ser225 residues should be the center of the box used for docking.

Additionally, our findings suggests that His226 assumes the HIP state during molecular simulations of the mature state of the PCSK9. However, by CpHMD simulations, we predicted that His226 plays a key role in the catalytic activity of the enzyme. In fact, in the proPCSK9, His226 can assume the  $N\delta$ -protonated (HID) state, allowing for perfect alignment of the catalytic triad side chains. Then, in the mature state, the pKa of His226 was calculated 2 units higher, leading to the hypothesis that His226 can adopt the double protonated (HIP) state. Thus, the explanation for why PCSK9 does not possess any serine-protease activity in the mature state is because it avoids the alignment of the catalytic triad.

## ACKNOWLEDGEMENTS

We would like to thank INDACO, CINECA, and the Regione Lombardia award of the LISA initiative for providing high-performance computing resources and support. GG gratefully acknowledges the support of NVIDIA Corporation for donating the Titan Xp GPU, which was utilized for the majority of the calculations in this study.

## CONFLICT OF INTEREST

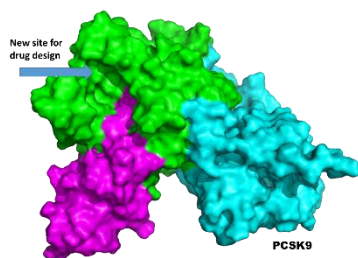
The authors declare no conflict of interest.

## REFERENCES

- [1] T. A. Lagace, D. E. Curtis, R. Garuti, M. C. McNutt, S. W. Park, H. B. Prather, *J. Clin. Invest.* **2006**, *116*, 2995-3005.
- [2] A. L. Catapano, A. Pirillo, G. D. Norata, *Vasc. Health Risk Manag.* **2017**, *13*, 343-351.
- [3] N. G. Seidah, *Expert Opin. Ther. Targets* **2009**, *13*, 19-28
- [4] F. Raal, R. Scott, R. Somaratne, I. Bridges, G. Li, S. M. Wasserman, E. A. Stein, *Circulation* **2012**, *126*, 2408-2417.
- [5] E. A. Stein, S. Mellis, G. D. Yancopoulos, N. Stahl, D. Logan, W. B. Smith, E. Lisbon, M. Gutierrez, C. Webb, R. Wu, Y. Du, T. Kranz, E. Gasparino, G. D. Swergold, *N. Engl. J. Med.* **2012**, *366*, 1108-1118.
- [6] S. Xu, S. Luo, Z. Zhu, J. Xu, *Eur. J. Med. Chem.* **2019**, *162*, 212-233
- [7] G. Grazioso, C. Bollati, J. Sgrignani, A. Arnoldi, C. Lammi, *J. Agric. Food Chem.* **2018**, *66*, 10552-10557.
- [8] C. Lammi, J. Sgrignani, G. Roda, A. Arnoldi, G. Grazioso, *ACS Med. Chem. Lett.* **2019**, *10*, 425-430.
- [9] C. Lammi, C. Zanoni, G. Aiello, A. Arnoldi, G. Grazioso, *Sci. Rep.* **2016**, *6*, 29931.
- [10] Y. Zhang, M. Ultsch, N. J. Skelton, D. J. Burdick, M. H. Beresini, W. Li, M. Kong-Beltran, A. Peterson, J. Quinn, C. Chiu, Y. Wu, S. Shia, P. Moran, P. Di Lello, C. Eigenbrot, D. Kirchhofer, *Nat. Struct. Mol. Biol.* **2017**, *24*, 848-856.
- [11] Y. Zhang, C. Eigenbrot, L. Zhou, S. Shia, W. Li, C. Quan, J. Tom, P. Moran, P. Di Lello, N. J. Skelton, M. Kong-Beltran, A. Peterson, D. Kirchhofer, *J. Biol. Chem.* **2014**, *289*, 942-955.
- [12] M. Stucchi, G. Grazioso, C. Lammi, S. Manara, C. Zanoni, A. Arnoldi, G. Lesma, A. Silvani, *Org. Biomol. Chem.* **2016**, *14*, 9736-9740.
- [13] C. Lammi, J. Sgrignani, A. Arnoldi, G. Lesma, C. Spatti, A. Silvani, G. Grazioso, *J. Med. Chem.* **2019**, *62*, 6163-6174.
- [14] K. K. Ray, U. Landmesser, L. A. Leiter, D. Kallend, R. Dufour, M. Karakas, T. Hall, R. P. T. Troquay, T. Turner, F. L. J. Visseren, P. Wijngaard, R. S. Wright, J. J. P. Kastelein, *N. Engl. J. Med.* **2017**, *376*, 1430-1440.
- [15] A. S. Peterson, L. G. Fong, S. G. Young, *J. Lipid Res.* **2008**, *49*, 1152-1156.

- [16] S. Benjannet, D. Rhainds, R. Essalmani, J. Mayne, L. Wickham, W. Jin, M. C. Asselin, J. Hamelin, M. Varret, D. Allard, M. Trillard, M. Abifadel, A. Tebon, A. D. Attie, D. J. Rader, C. Boileau, L. Brissette, M. Chretien, A. Prat, N. G. Seidah, *J. Biol. Chem.* **2004**, *279*, 48865-48875.
- [17] S. P. Salowe, L. Zhang, H. J. Zokian, J. J. Gesell, D. L. Zink, J. Wiltsie, X. Ai, M. Kavana, S. Pinto, *J. Biomol. Screen* **2016**, *21*, 1034-1041.
- [18] H. J. Kwon, T. A. Lagace, M. C. McNutt, J. D. Horton, J. Deisenhofer, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 1820.
- [19] L. Polgar, *Cell. Mol. Life Sci.* **2005**, *62*, 2161-2172.
- [20] D. Cunningham, D. E. Danley, K. F. Geoghegan, M. C. Griffor, J. L. Hawkins, T. A. Subashi, A. H. Varghese, M. J. Ammirati, J. S. Culp, L. R. Hoth, M. N. Mansour, K. M. McGrath, A. P. Seddon, S. Shenolikar, K. J. Stutzman-Engwall, L. C. Warren, D. Xia, X. Qiu, *Nat. Struct. Mol. Biol.* **2007**, *14*:413-419.
- [21] E. N. Hampton, M. W. Knuth, J. Li, J. L. Harris, S. A. Lesley, G. Spraggon, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 14604-14609.
- [22] B. Webb, A. Sali, *Curr. Protoc. Protein Sci.* **2016**, *86*, 2.9.1-2.9.37.
- [23] M.-Y. Shen, A. Sali, *Protein Sci.* **2006**, *15*, 2507-2524.
- [24] I. Y. B.-S. D. A. Case, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, V. W. D. Cruzeiro, T. A. Darden, D. G. R. E. Duke, G. Giambasu, T. Giese, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R. Harris, Y. H. N. Homeyer, S. Izadi, A. Kovalenko, R. Krasny, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, T. L. J. Liu, R. Luo, V. Man, D. J. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H., A. O. Nguyen, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L., J. S. Simmerling, J. Swails, R. C. Walker, J. Wang, H. Wei, L. Wilson, R. M. Wolf, X. Wu, L. Xiao, Y., D. M. Y. a. P. A. K. Xiong, Amber2019, University of California, San Francisco, **2019**.
- [25] Schrödinger Inc., Release 2019-4: Maestro, Schrödinger, New York, **2019**.
- [26] S. Park, F. Khalili-Araghi, E. Tajkhorshid, K. Schulten, *J. Chem. Phys.* **2003**, *119*, 3559-3566.
- [27] B. Roux, *Comput. Phys. Commun.* **1995**, *91*, 275-282.
- [28] J. Mongan, D. A. Case, J. A. McCammon, *J. Comput. Chem.* **2004**, *25*, 2038-2048.
- [29] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J. Chem. Theory Comput.* **2015**, *11*, 3696-3713.
- [30] Y. Mao, Y. Zhang, *Chem. Phys. Lett.* **2012**, *542*, 37-41.
- [31] C. I. Schroeder, J. E. Swedberg, J. M. Withka, K. J. Rosengren, M. Akcan, D. J. Clayton, N. L. Daly, O. Cheneval, K. A. Borzilleri, M. Griffor, I. Stock, B. Colless, P. Walsh, P. Sunderland, A. Reyes, R. Dullea, M. Ammirati, S. Liu, K. F. McClure, M. Tu, S. K. Bhattacharya, S. Liras, D. A. Price, D. J. Craik, *Chem. Biol.* **2014**, *21*, 284-294.
- [32] C. Jarzynski, *Phys. Rev. E* **1997**, *56*, 5018-5035.
- [33] J. Sgrignani, G. Grazioso, M. De Amici, *Biochemistry* **2016**, *55*, 5191-5200.
- [34] D. W. Zhang, T. A. Lagace, R. Garuti, Z. Zhao, M. McDonald, J. D. Horton, J. C. Cohen, H. H. Hobbs, *J. Biol. Chem.* **2007**, *282*, 18602-18612.
- [35] S. Benjannet, Y. G. Luna Saavedra, J. Hamelin, M.-C. Asselin, R. Essalmani, A. Pasquato, P. Lemaire, G. Duke, B. Miao, F. Duclos, R. Parker, G. Mayer, N. G. Seidah, *J. Biol. Chem.* **2010**, *285*, 40965-40978.
- [36] M. Ultsch, W. Li, C. Eigenbrot, P. Di Lello, M. T. Lipari, S. Gerhardy, A. P. AhYoung, J. Quinn, Y. Franke, Y. Chen, M. Kong Beltran, A. Peterson, D. Kirchhofer, *J. Mol. Biol.* **2019**, *431*, 885-903.
- [37] C. N. Schutz, A. Warshel, *Proteins: Struct. Funct. Bioinf.* **2001**, *44*, 400-417.
- [38] A.-S. Yang, M. R. Gunner, R. Sampogna, K. Sharp, B. Honig, *Proteins: Struct. Funct. Bioinf.* **1993**, *15*, 252-265.
- [39] S. Pahari, L. Sun, E. Alexov, *Database* **2019**, *2019*, 1-7.
- [40] S. H. Lapidus, N. N. Rajput, X. Qu, K. W. Chapman, K. A. Persson, P. J. Chupas, *Phys. Chem. Chem. Phys.* **2014**, *16*, 21941-21945.
- [41] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33-38.

## TABLE OF CONTENTS



*By constant-pH MD simulations, the protonation state of the catalytic triad of PCSK9 was investigated. Moreover, the last step of the maturation process of the enzyme was figured out by MD simulations. This study paves the way to the design of ligands capable to interfere with the PCSK9 maturation process.*

**Keywords:** drug design • molecular dynamics • pKa prediction • PCSK9 • constant-pH molecular dynamics simulations