Binding interactions of epididymal protease inhibitor and semenogelin-1: a homology modeling, docking and molecular dynamics simulation study

Changyu Shan[1], Hongwei Li[1], Yuping Zhang[2], Yuyan Li[3], Yingchun Chen[1] and Wei He[3]

[1] Department of Pharmaceutical Chemistry, The Third Military Medical University, Chongqing, China
[2] College of Pharmacy and Bioengineering, Chongqing University of Technology, Chongqing, China
[3] Department of Obstetrics and Gynecology, Southwest Hospital, The Third Military Medical University, Chongqing, China

ABSTRACT

Epididymal protease inhibitor (EPPIN) that is located on the sperm surface and specific to the male reproductive system is a non-hormonal contraceptive target, since the binding of EPPIN with the seminal plasma protein semenogelin-1 (SEMG1) causes a loss of sperm function. Here, we investigated the binding interactions between EPPIN and SEMG1 by homology modeling, docking and molecular dynamics simulation. Since no crystal structure was reported for EPPIN, its 3D structure was constructed by homology modeling and refined by dynamics simulation, illustrating the C-terminus domain of EPPIN could bind with its N-terminus domain through the residues 30–32 and 113–116. The binding interaction of SEMG1$^{10\text{-}8}$ peptide and EPPIN was investigated by Z-DOCK and dynamics simulation. After evaluating the models according to the calculated binding free energies, we demonstrated that C-terminus domain of EPPIN was important for the binding of SEMG1 via residues Tyr107, Gly112, Asn116, Gln118 and Asn122, while residue Arg32 in N-terminus domain also had contribution for their binding interaction. Additionally, the binding pocket of EPPIN was defined according to these key residues and verified by molecular docking with reported inhibitor EP055, suggesting that the pocket formed by Arg32, Asn114, Asn116, Phe117 and Asn122 could be important for the design of new ligands. This study might be helpful for the understanding of biological function of EPPIN and would encourage the discovery of non-hormonal contraceptive leads/drugs in the future.

INTRODUCTION

The use of condoms and the vasectomy operation were the classical methods for male contraception for many years (*Youssef, 1993*). The high use failure rate (18%) of the

condom and the irreversibility of vasectomy (*Guttmacher Institute, 2018*) have encouraged the pursuit of new targets for the development of male hormonal (*Chao & Page, 2016*) and non-hormonal (*Mannowetz, Miller & Lishko, 2017*; *O'Rand, Silva & Hamil, 2016*) contraceptives, including the inhibition of sperm function (*O'Rand, Silva & Hamil, 2016*) and physical blockage of the production and delivery of sperm (*Colagross-Schouten et al., 2017*; *Waller et al., 2016*).

The epididymal protease inhibitor (EPPIN), an example of a rat whey acidic protein four-disulfide core gene, is located on the surface of spermatozoa. In addition to its anti-bacterial activity (*Yenugu et al., 2004*) and its ability to modulate the proteolytic activity of a serine protease named prostate specific antigen (*O'Rand et al., 2006*), EPPIN inhibits sperm motility by binding to the seminal plasma protein, semenogelin-1 (SEMG1) (*Wang et al., 2005*). The binding of EPPIN and SEMG1 causes a loss of sperm function manifested as a rapid decrease in internal pH and calcium levels (*O'Rand & Widgren, 2012*). Because EPPIN is specific to the male reproductive system and its essential function on ejaculated spermatozoa could be reversibly blocked with easy access to the target on the sperm surface, this molecule was studied as a reasonable non-hormonal contraceptive target.

The EPPIN protein is a cysteine-rich protein, which contains both Kunitz (Asp26-Lys73) and WAP domains (Cys77-Cys127) (*Richardson et al., 2001*). Mutagenesis studies have demonstrated that Cys102, Tyr107, and Phe117 located at the Kunitz domain were of great significance for the binding affinity of EPPIN and SEMG1 (*Silva et al., 2012*). On the other hand, the truncation experiments with a series of SEMG1 peptides showed that the domain from 229 to 247 was considered as the EPPIN binding domain and Cys239 was the key residue (*Silva, Hamil & O'Rand, 2013*).

The crystalline structures of EPPIN and SEMG1 have not been reported. A computer model of EPPIN-SEMG1 was used to screen for hits, and a series of small organic compounds based on this strategy were found (*Hirano et al., 2001*; *Silva, Hamil & O'Rand, 2013*). In particular, **EP055** was reported as a potential male contraceptive since it should provide a reversible, short-lived pharmacological alternative (*O'Rand et al., 2018*) in the cynomolgus males. However, the detailed interactions of EPPIN and SEMG1 as well as the structural dynamic changes of EPPIN-SEMG1 complex, which should be helpful for understanding the mechanism of male contraception and developing new male contraceptives, have not been derived. Here, we reported a study on the binding interactions between EPPIN and SEMG1 by using homology modeling, molecular docking, and molecular dynamics (MD) simulation. The results suggested that the C- and N-terminus domains of EPPIN were important for the binding of SEMG1. Additionally, we found the pocket formed by Arg32, Asn114, Asn116, Phe117 and Asn122 could be important for the design of new ligands of EPPIN.

MATERIALS AND METHODS

Modeling of EPPIN and SEMG1

The full sequence of human EPPIN protein (EPPI_HUMAN or O95925, 133 residues) was retrieved from the UniProtKB (http://www.uniprot.org/uniprot/). After truncating the

signal peptide (Met1-Gly21), the N-terminus (Pro22-Asp71) and C-terminus (Lys73-Pro133) of EPPIN protein were aligned and constructed separately by using Discovery Studio 3.0 (*Chen & Weng, 2003*). After BLAST searching from the NCBI server, the known crystalline structures of neutrophil elastase (PDB ID: 2Z7F, chain A) (*Koizumi et al., 2008*) and carboxypeptidase inhibitor SmCI (PDB ID: 4BD9, chain B) (*Del Alonso et al., 2013*) showed the most similar sequences alignment to the EPPIN C-terminus (38.9%) and N-terminus (60.3%), respectively. Hence, we selected these two structures to build the C- and N-terminuses of EPPIN, respectively. According to the suggestion of Disulfide Bridge from UniProtKB, 7 disulfide bridge/bonds between Cys33-Cys61, Cys40-Cys65, Cys48-Cys60, Cys54-Cys69, Cys77-Cys127, Cys86-Cys110 and Cys102-Cys123 in EPPIN were patched (*Silva et al., 2012*). After the structures of N- and C-terminuses were homology modeled, they were linked and further refined by Discovery Studio 3.0 (*Chen & Weng, 2003*).

As for SEMG1, we selected a short-truncated fragment sequences flanking Cys239 residue, namely SEMG1$^{10\text{-}8}$, representing SEMG1$^{E229\text{-}Q247}$. As the BLAST searching result demonstrated low identities aligning with known crystal structures, the structure of this sequence was predicted and modeled by the I-TASSER server (*Yang et al., 2015*) and the one with the highest *C*-score was selected for further study, since the *C*-score is a confidence score for estimating the quality of predicted models by I-TASSER, where a *C*-score of higher value signifies a model with a high confidence and vice-versa (*Yang et al., 2015*).

Energy minimization and structure validation

After generating 3D models of EPPIN and SEMG1, we performed the energy minimizations using SYBYL-X 2.0 with the Powell method under AMBER7 FF99 force field and AMBER charges. The energy minimizations were terminated when the iterations reached 10,000 steps or the energy gradient less than 0.5 kcal/mol.

Structure evaluation and stereo-chemical analysis for the EPPIN models were performed by using proSA-web Z-scores (*Wiederstein & Sippl, 2007*) and PROCHECK Ramachandran plots (*Laskowski et al., 1993*). The visualization of the generated models was performed using the PyMOL program (*Schrodinger, LLC, 2015*).

Molecular dynamics simulation

We selected the EPPIN and SEMG1$^{10\text{-}8}$ to perform MD simulation studies. The residues were ionized within the physiological pH range (~7.40). To determine the protonation states for histidines and other residues, Discovery Studio 3.0 (*Chen & Weng, 2003*) was used to predict the pK values of the residues. As calculated pK values were lower than 7.40, all histidines were predicted not to be protonated. Sidechains of Asp, Glu, Arg and Lys were charged thus Asp$^-$, Glu$^-$, Arg$^+$ and Lys$^+$, respectively, according to all the simulations.

The solvent molecules and additional ions for simulations were added using the *tleap* (*Case et al., 2005*) module of AMBER14 under the ff14SB force field (*Maier et al., 2015*). The systems were first neutralized with Na$^+$ or Cl$^-$, then solvated in the TIP3P water

model (*Jorgensen et al., 1983*) and subsequently placed in a regular hexahedron box with a minimal distance of 12 Å for the solute from the box borders.

The systems were first minimized in the AMBER14 *pmemd.MPI* module in three stages. At the first two stages, the whole proteins or their backbone atoms were fixed by applying a harmonic force constant of 2 kcal/mol Å^2 respectively, thus making the water molecules and protein side chains free to move successively. In the following stage, the restraint strength was abolished, so that the entire systems would be able to move freely.

After the minimization, the systems were gradually heated from 0° K to 300° K for a total 50 pico-seconds (ps). This step was performed using the Langevin thermostat (*Izaguirre et al., 2001*) with a collision frequency of 2.0 ps^{-1}. Then, the pressure of systems was kept constant using a 50 ps simulation. Following this, in order to obtain a system equilibrium, a simulation of 0.5 nano-second (ns) was performed at 300° K, with constant pressure and without restriction.

Starting from the last frame of the equilibration, MD simulations for the different systems were performed, respectively. The MD simulations were ran by using the *pmemd. CUDA.MPI* module of AMBER 14. The electrostatics interaction was calculated using the particle mesh Ewald (*Darden, York & Pedersen, 1993*) method with an 8 Å non-bonded cutoff. The temperature and pressure of the system were kept constant during the whole MD simulations. The time interval for the MD was set as 2 femto-seconds (fs). The data were saved every 10 ps for analysis. Subsequently, 50 or more ns MD simulations were performed under 300° K and 1 atm.

The analyses of root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (R_g) and atom distances were carried out with the AMBER14 module *CPPTRAJ*, VMD and PyMOL programs (*Schrodinger, LLC, 2015*), respectively.

Conformations sampling for EPPIN and SEMG1

To select the most reasonable conformation of EPPIN and SEMG1 models for further studies, after the MD simulation reached equilibrium, some conformations were sampled by extracting one conformation per five ns. After initial energy minimizations using the methods introduced in the previous section ("Energy minimization and structure validation"), we selected the conformation with the lowest energy as the most reasonable model for further docking and MD simulations.

Molecular docking study of EPPIN-SEMG1 complexes

The docking study of the EPPIN-SEMG1[10-8] (protein-peptide) interaction was performed according to the ZDOCK module of Discovery Studio 3.0 (*Chen & Weng, 2003*). The selected 3D model of EPPIN and SEMG1[10-8] were used as inputs for the receptor and the ligand, respectively. According to the previous findings (*Silva et al., 2012*), we selected Tyr107 and Phe117 of EPPIN and Cys239 of SEMG1[10-8] as the binding site residues. After the predicted binding models (EPPIN-SEMG1[10-8]) were constructed by the ZDOCK module, we randomly chose three different poses from the top 50 ranks in the area with the densest docking cluster for further studies.

MM/PBSA calculations

To calculate the binding free energies between EPPIN and SEMG1[10-8] for three different binding conformations, 85 ns MD simulations were performed using the previous MD protocol, until the systems reached equilibrium. The binding free energies were calculated using the Molecular Mechanics Poisson–Boltzmann Surface Area (MM/PBSA) method (*Chen et al., 2016*, *Sun et al., 2014a*, *2014b*) implemented in AMBER 14. For each system, 100 snapshots of the equilibrium stage were used from the MD trajectory. For each snapshot, free energy was calculated for EPPIN, SEMG1[10-8] and the EPPIN-SEMG1[10-8] complex using a single trajectory approach. The total binding free energy can be calculated according to the following equation (*Fang et al., 2014*):

$$\Delta G_{bind} = \Delta G_{complex} - \Delta G_{receptor} - \Delta G_{ligand} \tag{1}$$

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{solv} - T\Delta S \tag{2}$$

$$\Delta E_{MM} = \Delta E_{vdw} + \Delta E_{ele} \tag{3}$$

$$\Delta G_{solv} = \Delta G_{PB} + \Delta G_{SA} \tag{4}$$

Where ΔE_{MM} denotes the gas-phase interaction energy between the receptor and the ligand (including van der Waals energy contribution (ΔE_{vdw}) and electrostatic energy contribution (ΔE_{ele})); ΔG_{PB} and ΔG_{SA} are the polar and nonpolar components of the de-solvation free energy, respectively; $T\Delta S$ represents the conformational entropy contribution at temperature T. Here, ΔG_{PB} was determined by the Poisson–Boltzmann approximation model, while ΔG_{SA} was estimated based on the solvent accessible surface area model by the method: $\Delta G_{SA} = \gamma \times SASA + \beta$, where the values of the constants γ and β were 0.00542 kcal·Å^{-2} and 0.92 kcal·mol^{-1}, respectively, (*Weiser, Shenkin & Still, 1999*). The solvent probe radius and ionic strength were set to be 1.4 Å and 0.1 mM, respectively. The interior and exterior dielectric constant of MM/PBSA calculation systems was 1.0 and 80.0.

Investigation of potential binding pockets by molecular docking studies

Surflex-Dock GeomX module in SYBYL-X 2.0 was used to investigate the potential binding pockets of EPPIN. During the progress of investigating these pockets, we selected the Multi-channel Surface mode and the important residues Gly109, Cys110, Gln111, Gly112, Asn113, Asn114 and Asn116 to generate and select binding pockets of EPPIN, respectively. The EPPIN ligand **EP055** (*O'Rand et al., 2018*) was docked into the binding pocket using the Surflex-Dock GeomX module. The visualization of the EPPIN-EP055 complex interaction was performed using PyMOL (*Schrodinger, LLC, 2015*) programs.

RESULTS AND DISCUSSION

Homological model, equilibration and stability of EPPIN

As the crystalline structure is not available for EPPIN, the 3D model of EPPIN was constructed by homological modeling. The neutrophil elastase (PDB ID: 2Z7F, chain A)
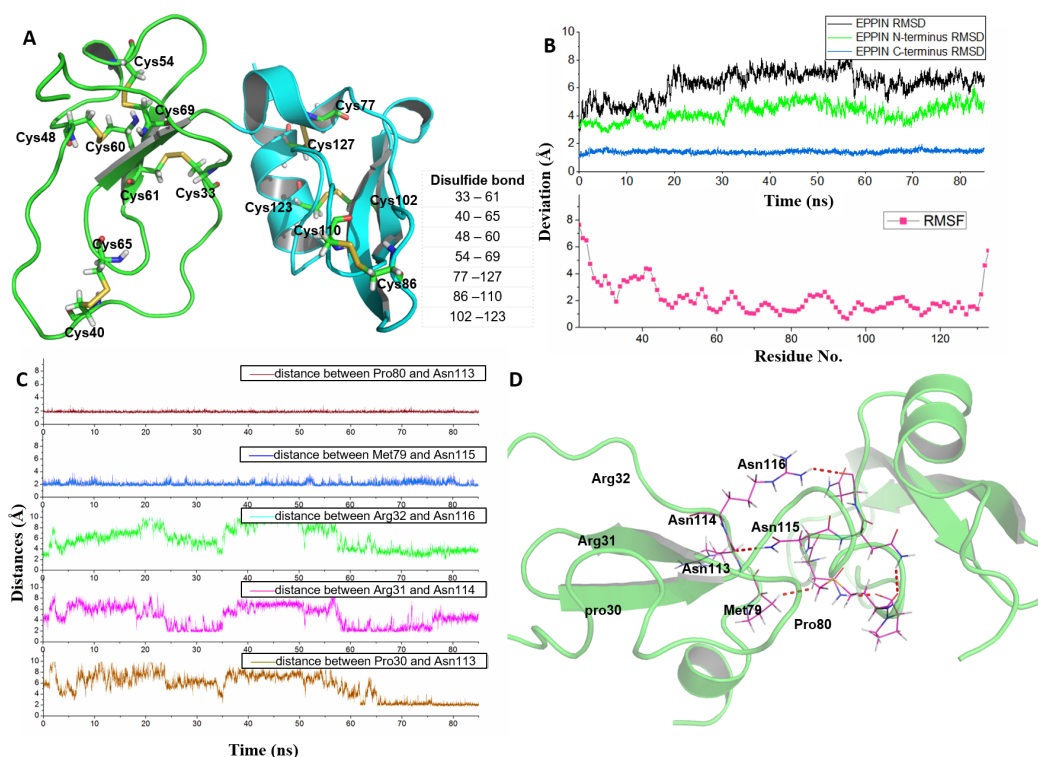
Figure 1 Homologous modeling structure of EPPIN after molecular dynamics stimulation. (A) The homology models of the N-terminus domain (23–77) and C-terminus domain (77–133) in EPPIN protein (disulfide bonds are highlighted as sticks); (B) RMSD (root mean square deviation, highlighted in black color) and RMSF (root mean square fluctuation, highlighted in red color) profile; (C) Time evolutions for the distances between the N-terminus (23–77) and C-terminus (77–133) of EPPIN, including Pro80 and Asn113, Met79 and Asn115, Arg32 and Asn116, Arg31 and Asn114, as well, Pro30 and Asn113. (D) The binding interaction between the N-terminus (22–77) and C-terminus (77–133) of EPPIN. Full-size ⤓ DOI: 10.7717/peerj.7329/fig-1

(*Koizumi et al., 2008*) was chosen as the reference structure of N-terminus because it had greater than 40% sequence identity. Similarly, the carboxypeptidase inhibitor SmCI (PDB ID: 4BD9, chain B) was selected as the template for C-terminus domain (77–133) (*Del Alonso et al., 2013*). The detailed sequence alignment of EPPIN C-terminus domain and its structure templates was shown in Fig. S1. As there are seven disulfide bonds in EPPIN, the alignment of the critical Cys amino acids for disulfide bonds in the candidate templates was also considered as another important factor for modeling criteria. We also used I-TASSER to build the model of each domain of EPPIN which was similar to the homological model.

There were two β-sheet domains and four disulfide bonds in N-terminus, while C-terminus had similar structures to reported model (*Silva et al., 2012*) with two α-helix and two β-sheet domains (Fig. 1A). The quality of the EPPIN model was predicted by ProSA-web with value of −5 (Fig. S1) and Ramachandran plots (Fig. S2), indicated that this structure employed the good-quality in the range of the theoretical protein structure models.

In order to rationalize the 3D structure of EPPIN, 85 ns MD simulation with water and ion was performed. We found that the RMSD of EPPIN was equilibrated after 20 ns and

was kept stable at 6 Å (in Fig. 1B). Because the disulfide bonds could stabilize the structure of EPPIN, the RMSF, a measurement of the average atomic mobility of their neighboring residues, had only a few changes with a deviation of 1–2 Å during the MD simulation (Fig. 1B). At the same time, the RMSF values of the residues in the N-terminus of EPPIN, especially, residues Gly23-Arg45, were much bigger than those in C-terminus, suggesting the C-terminus of EPPIN should have a smaller deviation. These results were consistent with the highly stable interactions between different residues in the C-terminus such as Pro80 and Asn113, as well Met79 and Asn115 (Figs. 1C and 1D) with distance of 2 Å.

Although the C-terminus is reported binding domain of SEMG1 (Silva et al., 2012), it is important to investigate the binding interactions between the C-terminus and N-terminus of EPPIN in order to understand the relationship between the structure and function among these domains. We found that some residues in the N-terminus, such as Arg31 and Arg32, were binding to Asn114 and Asn116 by hydrophobic interactions, respectively. In addition, there was a hydrogen bond between Pro30 and Asn113 with a distance of 2 Å. In brief, the disulfide bonds and the binding interactions between the C-terminus and N-terminus should stabilize the conformation of EPPIN.

The molecular docking of SEMG1 peptide to EPPIN

Seminal plasma protein semenogelin-1 has 462 residues, in which Cys239 is a critical amino acid for its binding to EPPIN. According to the truncation experimental data of the binding affinity with EPPIN (Silva, Hamil & O'Rand, 2013), the shortest SEMG1 peptide which could bind with EPPIN was SEMG1$^{10\text{-}8}$ for the sequence of Glu229-Gln247. Thus, we chose this short SEMG1 peptide and constructed its structure by I-TASSER servers (Yang et al., 2015) (Fig. 2A). The most desirable model of SEMG1$^{10\text{-}8}$ peptide got the C-score of −0.70. This score was typically in the range of −5–2, which is the quantitatively confidence of the models constructed by I-TASSER servers. A higher C-score signifies a model with a higher confidence and vice-versa (Yang et al., 2015). Moreover, this model displayed as an alpha helical structure, which was consistent with a previous report (O'Rand, Silva & Hamil, 2016).

Then, the 50 ns MD simulation for SEMG1$^{10\text{-}8}$ with water and ion was performed. We found that the RMSD was equilibrated and stable after 10 ns for this peptide (Fig. 2B). Notably, Cys239, the irreplaceable residue for the binding of SEMG1 and EPPIN showed the lowest RMSF, indicated the binding site of SEMG1$^{10\text{-}8}$ was more stable than the other part of this peptide (Fig. 2B).

The binding models of EPPIN with SEMG1$^{10\text{-}8}$ peptide were generated by ZDOCK. The binding positions were initially filtered according to scores and the interactions with the key residues, such as Cys102, Tyr107 and Phe117 in EPPIN and Cys239 in SEMG1$^{10\text{-}8}$. As shown in Fig. 2C, the most desirable binding center of SEMG1$^{10\text{-}8}$ was located in the cleavage structure formed by the N- and C-terminus of EPPIN. Three binding poses for EPPIN-SEMG1$^{10\text{-}8}$ were selected randomly in this most probable binding area for further MD studies (Fig. 2D). Further, the binding free energies calculated by MM/PBSA
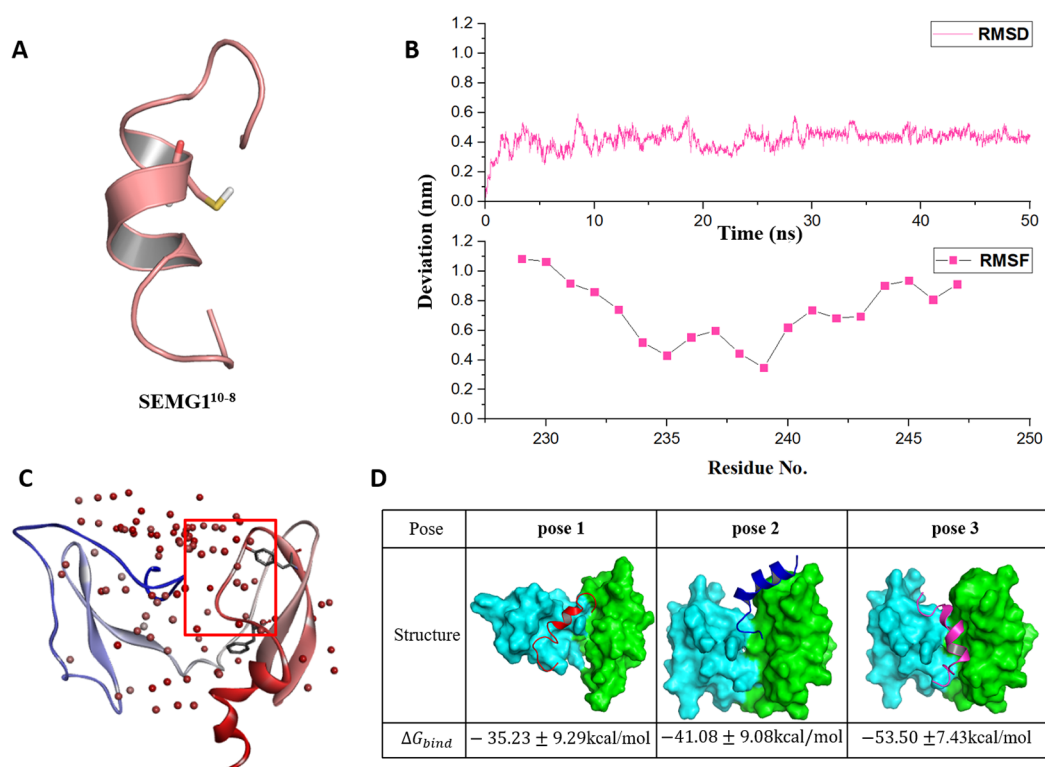
Figure 2 **The molecular dynamics optimized strucrues of SEMG1$^{10\text{-}8}$ peptide and the binding model of EPPIN and SEMG1$^{10\text{-}8}$.** (A) The 3D structure model of SEMG1$^{10\text{-}8}$ peptide, Cys239 was shown as stick. (B) The RMSD, RMSF profile of SEMG1$^{10\text{-}8}$ during the dynamic simulation. (C) The binding clusters of EPPIN. EPPIN was shown as cartoon and colored from blue (N-terminus domain) to red (C-terminus domain). Balls distributed over EPPIN indicates the binding cluster, the redder the balls, the more poses in the cluster. The red frame indicated the binding area, from which three binding poses were selected randomly. Tyr107 and Phe117 showed as sticks were utilized to filter the unwanted poses. (D) The three different poses for EPPIN and SEMG1$^{10\text{-}8}$ complex and their binding free energies. The EPPIN protein in each pose was illustrated as molecular surface and colored by cyan (N-terminus domain) and green (C-terminus domain). SEMG1$^{10\text{-}8}$ peptide was displayed as cartoon and shown as red, mazarine and megentas for each pose. Full-size ◨ DOI: 10.7717/peerj.7329/fig-2

was adopted to filter the model for EPPIN-SEMG1$^{10\text{-}8}$ complex. Besides considering the reported key residues of both SEMG1$^{10\text{-}8}$ peptide (Cys239) and EPPIN (Tyr107 and Phe117) (*Silva, Hamil & O'Rand, 2013*; *Silva et al., 2012*), the binding model with the lowest free energy was selected as the most plausible model. As shown in Fig. 2D, the binding interaction of C-terminus of EPPIN and SEMG1$^{10\text{-}8}$ was more reasonable because the corresponding **pose 2** and **pose 3** employed lower binding free energies (−41.08 kcal/mol and −53.50 kcal/mol, severally), while binding at N-terminus as **pose 1** resulted in −35.23 kcal/mol free energy (The binding free energy components was summarized in Table S1). These findings were consistent with previous reports (*Silva et al., 2012*), and further confirmed that SEMG1 possessed higher affinity to the C-terminus than N-terminus of EPPIN. Because **pose 3** had the lowest binding free energies, in which SEMG1$^{10\text{-}8}$ occupied the groove near EPPIN C-terminus, we finally chose **pose 3** for further investigating the particular binding interaction of EPPIN and SEMG1$^{10\text{-}8}$.
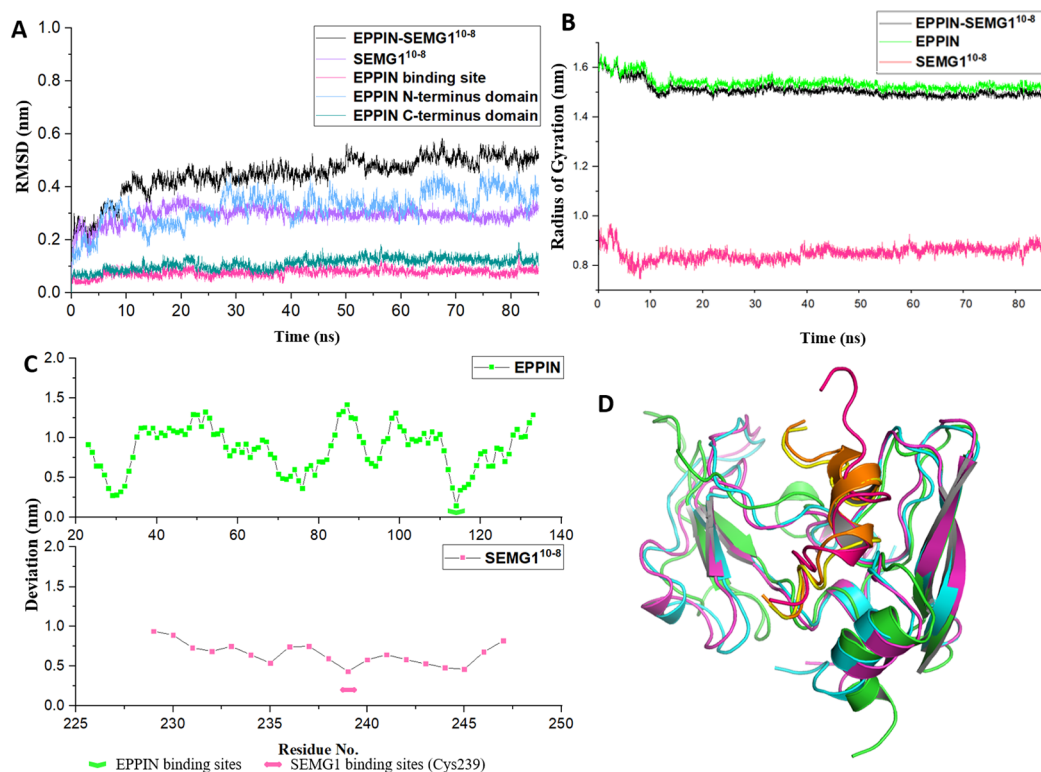
Figure 3 The molecular dynamic simulation process of EPPIN-SEMG1$^{10\text{-}8}$ complex. (A) The RMSD of EPPIN-SEMG1$^{10\text{-}8}$ complex, EPPIN N-terminus, EPPIN C-terminus, SEMG1$^{10\text{-}8}$ and EPPIN binding site (Gln108-Asn122); (B) The evolution of the radius of gyration (R_g) of the backbone ($C\alpha$) atoms over the course of the simulation for EPPIN-SEMG1$^{10\text{-}8}$ complex (black), EPPIN (green) and SEMG1$^{10\text{-}8}$ (red); (C) The RMSF of EPPIN-SEMG1$^{10\text{-}8}$ complex. The upper part (green dots line) indicated EPPIN component and the bottom part represented SEMG1$^{10\text{-}8}$ (red dots line), respectively. The binding sites of EPPIN and SEMG1 in each complex were showed as fold lines and double-headed arrows severally; (D) Initial configuration superimposed on the last and the average configuration during 85 ns of simulation for EPPIN-SEMG1$^{10\text{-}8}$ complex. Complexes are depicted as cartoons. For EPPIN, the initial, last and the average configuration were shown in green, magentas and cyan, respectively. For the SEMG1$^{10\text{-}8}$ peptide, the initial, last and the average configuration were shown in red, yellow and orange, respectively.

Full-size ◱ DOI: 10.7717/peerj.7329/fig-3

Equilibration and stability of EPPIN-SEMG1$^{10\text{-}8}$ complex

As shown in Fig. 3A, the structure of the EPPIN-SEMG1$^{10\text{-}8}$ complex gradually became stable during the first 20 ns simulation and the RMSD was about 5 Å after 20 ns. In the complex, the deviation of SEMG1$^{10\text{-}8}$ component was about 3 Å after 20 ns. Moreover, the RMSD value of EPPIN N-terminus domain was about 3 Å and showed mild fluctuate. However, EPPIN C-terminus domain exhibited a much more stable and lower RMSD value (about 1 Å) which signified this domain remained significantly steady in the MD simulation process. Impressively, RMSD of the binding site of EPPIN (Gln108-Asn122) was quite low with the value of 0.5–1 Å, suggesting that the structure of EPPIN binding area was stable during the simulation.

Next, to study the conformational flexibility and compactness of the peptides and protein, their R_g values were calculated. The R_g of a set of atoms are the mass-weighted root mean square distance of those atoms from their center of mass

(*Davoudmanesh & Mosaabadi, 2018*). The evolution of the R_g of the $C\alpha$ atoms in the process of the simulation was calculated for EPPIN, SEMG1$^{10-8}$ as well as their complex. According to Fig. 3B, the R_g value showed a declining trend for EPPIN, SEMG1$^{10-8}$ and their complex, indicating that the structural fluctuation of the protein complex became smaller along with the MD simulation time. Moreover, EPPIN presented slightly larger R_g values during the simulation than the EPPIN-SEMG1$^{10-8}$ complex, which might result from the constricted main-chain movement and increased stability through the interaction of EPPIN and SEMG1$^{10-8}$.

Further, the RMSFs of their $C\alpha$ atoms of the stabilized 60 ns were calculated to determine the conformational flexibility of EPPIN and SEMG1$^{10-8}$ peptides. The residues of EPPIN binding site (Gln108-Asn122) exhibited a lower RMSF with the value of 2–3 Å (Fig. 3C), suggesting the binding site was rigid during the MD simulation. The similar trend could also be observed in the SEMG1$^{10-8}$ peptide in the complex, residue Cys239, crucial for SEMG1 binding to EPPIN, represented the lowest RMSF (4 Å) (Fig. 3C). The fluctuations of the amino acids revealed that the binding sites were less flexible in both EPPIN and SEMG1$^{10-8}$, proving that the interaction of EPPIN-SEMG1$^{10-8}$ complex was stable during the MD simulation.

To elucidate the conformation changes during the simulation, the initial conformation of EPPIN-SEMG1$^{10-8}$ complex was superimposed on the last conformation and the average conformation over the course of MD process (Fig. 3D). The last conformation of the SEMG1$^{10-8}$ peptide was significantly different from its initial conformation, which exhibited a stable α-helix structure. Nevertheless, the last conformation was able to precisely align with the average conformation, especially for SEMG1$^{10-8}$ and C-terminus of EPPIN, indicated the complex was particularly stable during the MD simulation.

Interactions between EPIIN and SEMG1$^{10-8}$

Because the C-terminus domain was reported as the binding domain of SEGM1, the binding at this domain attracted most of the attention (*Silva et al., 2012*). As shown in Figs. 4A and 4B, SEMG1$^{10-8}$ could bind with the C-terminus domain of EPPIN, which was consistent with the experimental results (*O'Rand, Silva & Hamil, 2016*). The MD simulation result showed that five important residues at the C-terminus domain had strong interactions with SEMG1$^{10-8}$. It was found that these residues, including Tyr107, Gly112, Asn116, Gln118 and Asn122, stayed relatively close to the SEGM1$^{10-8}$ peptide by hydrogen bonds with a short distance of 2–4 Å during the MD simulation (Figs. 4C and 4D). For the SEMG1 peptide, residues such as Gln235, Thr236, Cys239 and Gln243 were important for the formation of hydrogen bonds. Among these residues, Cys239 formed two hydrogen bonds with Asn116 and Gln118 of EPPIN, respectively, indicating Cys239 played a crucial role in the interaction between EPPIN and SEMG1, which confirmed the former report (*Silva, Hamil & O'Rand, 2013*). Unexpectedly, Arg32 in the N-terminus domain, which is an important amino acid for the interaction between the C- and N-terminus domains, could interact with His242 of SEGM1$^{10-8}$.

In brief, we speculated that apart from the reported residues in C-terminus domain, such as Tyr107 and Gly112 (*Silva et al., 2012*), residues Asn116, Gln118 and Asn122 might
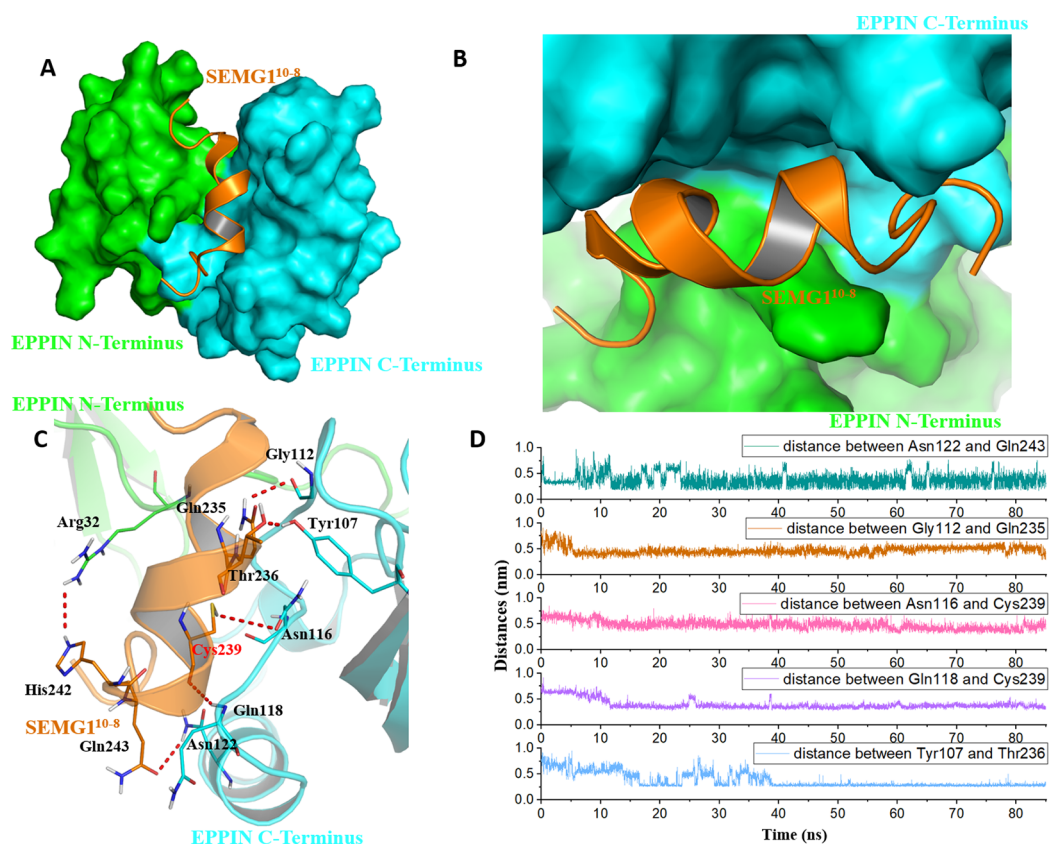
Figure 4 The binding interaction of EPPIN and SEMG1$^{10\text{-}8}$. (A and B) The structure of EPPIN-SEMG1$^{10\text{-}8}$ complex. EPPIN was depicted as protein surface in green (N-terminus domain) and cyan (C-terminus domai); (C) The binding interaction of EPPIN and SEMG1$^{10\text{-}8}$. The important residues for the EPPIN-SEMG1 interactions were shown as lines and the hydrogen bonds between EPPIN and SEMG1 were illustrated by red dashed lines; (D) Time evolutions for the distances between EPPIN and SEMG1$^{10\text{-}8}$, including Asn122 and Gln243, Gly112 and Gln235, Asn116 and Cys239, Gln118 and Cys239 as well as Tyr107 and Thr236. Full-size 🖼 DOI: 10.7717/peerj.7329/fig-4

play important roles for the binding of EPPIN. In addition, the interaction between the N-terminus domains of EPPIN with SEGM1 could not be ruled out due to these results.

Important residues in the potential binding pocket of EPPIN

A reasonable potential binding pocket was generated by the SYBYL 2.0 Docking Suite module using the evaluated residues in the C-terminus of EPPIN for the binding of SEMG1, such as Tyr107, Gly112, Asn116, Gln118 and Asn122. As expected, this pocket of EPPIN was located mainly in the C-terminus. The residues involved in the binding pocket for EPPIN were highlighted in green or blue sticks (Figs. 5B and 5C), including Arg32, Leu72, Asn114, Asn116, Phe117 and Asn122. Some residues of the predicted pocket, for instance Phe117, were supported by the reported mutation data (*Silva et al., 2012*). Interestingly, the residues Arg32 and Leu72, which was located in N-terminus, were also found around the binding pocket. Other important residues, such as Asn116 and Asn122 in the potential binding pocket were consistent with the results of the molecule docking and MD simulations.
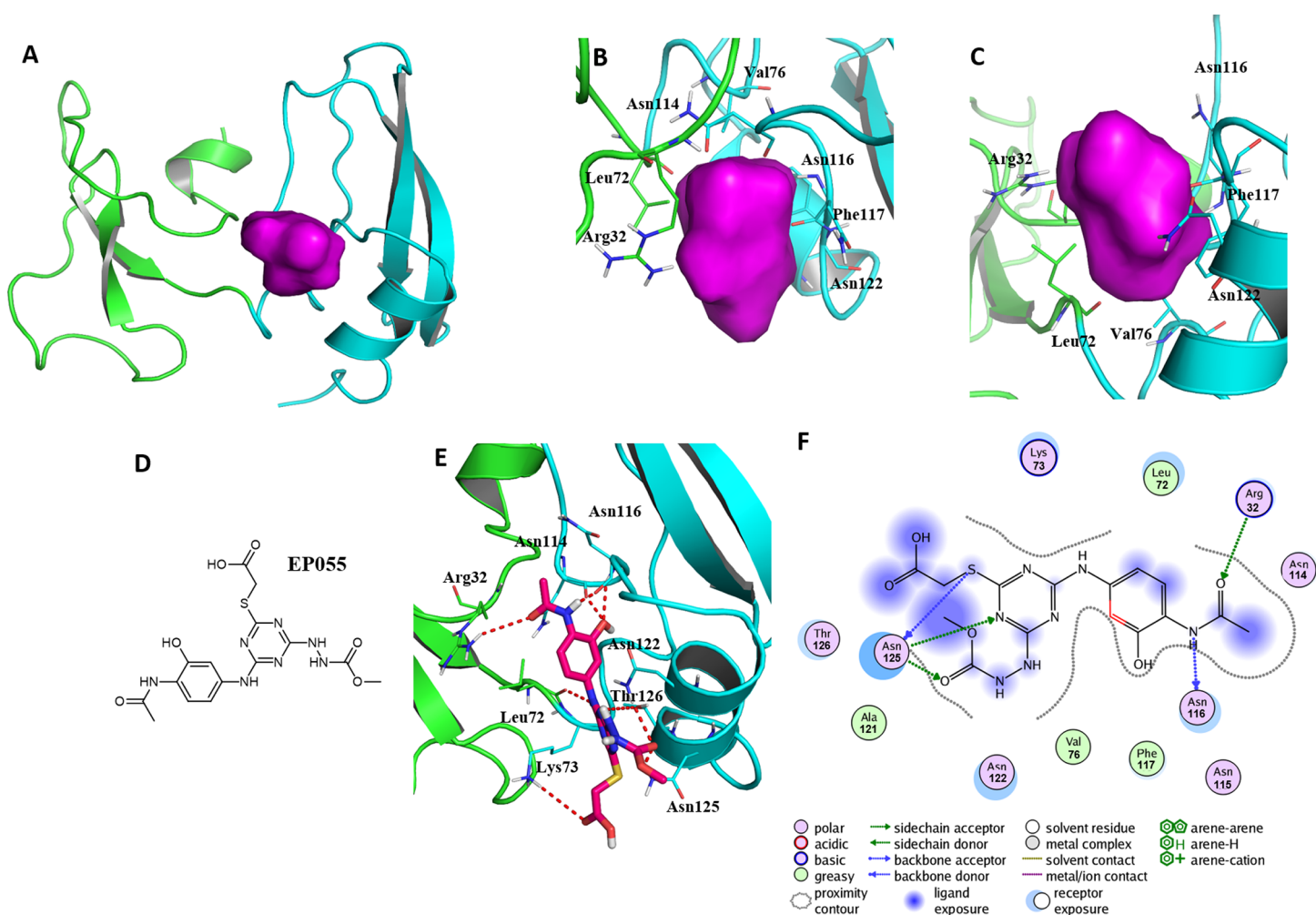
Figure 5 A reasonable potential binding pocket of EPPIN and the molecular docking of the reported ligand EP055 with EPPIN in this pocket.
(A) The predicted binding pocket of EPPIN; (B) The dorsal, and (C) ventral view of the binding pocket. The important residues involved in the binding pocket for EPPIN are highlighted in green or blue bars; (D) The structure of **EP055**; (E and F) The 3D and 2D visualization and interaction within the EPPIN-EP055 complex. The important residues for the EPPIN and **EP055** interactions were shown as lines and the hydrogen bonds between EPPIN and **EP055** were illustrated by a red dashed line. Full-size 🖼 DOI: 10.7717/peerj.7329/fig-5

To further verify this binding pocket, molecular docking of the reported ligand **EP055** (*O'Rand et al., 2018*) (Fig. 5D) and EPPIN was conducted (docking score = 7.24, meaning a relatedly strong binding interaction). The structure of the EPPIN-EP055 complex showed that the **EP055** could form hydrogen bonds with Lys73, Asn114, Asn116, Asn122, Thr126 and Asn125 in the C-terminus of EPPIN. Impressively, the residues Arg32 and Leu72 in the N-terminus also exhibited two obvious hydrogen bonds with **EP055** (Figs. 5E and 5F). These results suggested that the **EP055** could bind with both terminuses of EPPIN by multiple binding interactions.

CONCLUSIONS

The model of EPPIN was constructed by homological modeling based on the reported crystalline structures. The MD simulation results showed the structure of EPPIN was quite

stable due to its disulfide bonds and the binding interaction between residues 113–116 in C-terminus domain and 30–32 in N-terminus domain. The model of SEMG1$^{10\text{-}8}$ peptide, constructed by I-TASSER servers and rationalized by dynamics simulation, was docked with EPPIN. The binding free energy was calculated for different binding poses and the one with the lowest ΔG_{bind} was selected as EPPIN-SEMG1$^{10\text{-}8}$ complex model. The binding interaction of SEMG1$^{10\text{-}8}$ peptide and EPPIN was investigated by dynamics simulation. The results suggested that the C- and N-terminus domains of EPPIN were important for the binding of SEMG1. Additionally, we found the pocket formed by Arg32, Asn114, Asn116, Phe117 and Asn122 should be important for the design of new ligands of EPPIN. This detail binding interaction study might be helpful for the better understanding of the biological function of EPPIN and will encourage the discovery of non-hormonal contraceptive leads/drugs in the future.

ACKNOWLEDGEMENTS

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

Grant Disclosures

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Changyu Shan conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper.
- Hongwei Li performed the experiments.
- Yuping Zhang contributed reagents/materials/analysis tools.
- Yuyan Li analyzed the data.
- Yingchun Chen conceived and designed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Wei He conceived and designed the experiments, approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw data is available as a Supplemental File.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.7329#supplemental-information.

REFERENCES

Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ. 2005. The Amber biomolecular simulation programs. *Journal of Computational Chemistry* **26(16)**:1668–1688 DOI 10.1002/jcc.20290.

Chao JH, Page ST. 2016. The current state of male hormonal contraception. *Pharmacology & Therapeutics* **163**:109–117 DOI 10.1016/j.pharmthera.2016.03.012.

Chen F, Liu H, Sun H, Pan P, Li Y, Li D, Hou T. 2016. Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein–protein binding free energies and re-rank binding poses generated by protein–protein docking. *Physical Chemistry Chemical Physics* **18(32)**:22129–22139 DOI 10.1039/c6cp03670h.

Chen R, Weng Z. 2003. A novel shape complementarity scoring function for protein-protein docking. *Proteins: Structure, Function, and Genetics* **51(3)**:397–408 DOI 10.1002/prot.10334.

Colagross-Schouten A, Lemoy M-J, Keesler RI, Lissner E, VandeVoort CA. 2017. The contraceptive efficacy of intravas injection of Vasalgel^{TM} for adult male rhesus monkeys. *Basic and Clinical Andrology* **27(1)**:4 DOI 10.1186/s12610-017-0048-9.

Darden T, York DM, Pedersen LG. 1993. Particle mesh Ewald: an $N \log(N)$ method for Ewald sums in large systems. *Journal of Chemical Physics* **98(12)**:10089–10092 DOI 10.1063/1.464397.

Davoudmanesh S, Mosaabadi JM. 2018. Investigation of the effect of homocysteinylation of substance P on its binding to the NK1 receptor using molecular dynamics simulation. *Journal of Molecular Modeling* **24(7)**:177 DOI 10.1007/s00894-018-3695-7.

Del Alonso RM, Reytor ML, Trejo SA, Chavez MA, Aviles FX, Reverter D. 2013. A noncanonical mechanism of carboxypeptidase inhibition revealed by the crystal structure of the Tri-Kunitz SmCI in complex with human CPA4. *Structure* **21(7)**:1118–1126 DOI 10.1016/j.str.2013.04.021.

Fang J, Wu P, Yang R, Gao L, Li C, Wang D, Wu S, Liu A-L, Du G-H. 2014. Inhibition of acetylcholinesterase by two genistein derivatives: kinetic analysis, molecular docking and molecular dynamics simulation. *Acta Pharmaceutica Sinica B* **4(6)**:430–437 DOI 10.1016/j.apsb.2014.10.002.

Guttmacher Institute. 2018. Contraceptive use in the United States. *Available at https://www.guttmacher.org/fact-sheet/contraceptive-use-united-states*.

Hirano Y, Shibahara H, Obara H, Suzuki T, Takamizawa S, Yamaguchi C, Tsunoda H, Sato I. 2001. Relationships between sperm motility characteristics assessed by the computer-aided sperm analysis (CASA) and fertilization rates in vitro. *Journal of Assisted Reproduction and Genetics* **18(4)**:213–218 DOI 10.1023/a:1009420432234.

Izaguirre JA, Catarello DP, Wozniak JM, Skeel RD. 2001. Langevin stabilization of molecular dynamics. *journal of chemical physics* **114(5)**:2090–2098 DOI 10.1063/1.1332996.

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. 1983. Comparison of simple potential functions for simulating liquid water. *journal of chemical physics* **79(2)**:926–935 DOI 10.1063/1.445869.

Koizumi M, Fujino A, Fukushima K, Kamimura T, Takimoto-Kamimura M. 2008. Complex of human neutrophil elastase with 1/2SLPI. *Journal of Synchrotron Radiation* **15(3)**:308–311 DOI 10.1107/s0909049507060670.

Laskowski R, Macarthur MW, Moss DS, Thornton JM. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **26(2)**:283–291 DOI 10.1107/s0021889892009944.

Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KF, Simmerling C. 2015. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of Chemical Theory and Computation* **11(8)**:3696–3713 DOI 10.1021/acs.jctc.5b00255.

Mannowetz N, Miller MR, Lishko PV. 2017. Regulation of the sperm calcium channel CatSper by endogenous steroids and plant triterpenoids. *Proceedings of the National Academy of Sciences of the United States of America* **114(22)**:5743–5748 DOI 10.1073/pnas.1700367114.

O'Rand MG, Hamil KG, Adevai T, Zelinski M. 2018. Inhibition of sperm motility in male macaques with EP055, a potential non-hormonal male contraceptive. *PLOS ONE* **13(4)**:e0195953 DOI 10.1371/journal.pone.0195953.

O'Rand MG, Silva EJR, Hamil KG. 2016. Non-hormonal male contraception: a review and development of an Eppin based contraceptive. *Pharmacology & Therapeutics* **157**:105–111 DOI 10.1016/j.pharmthera.2015.11.004.

O'Rand MG, Widgren EE. 2012. Loss of calcium in human spermatozoa via EPPIN, the semenogelin receptor. *Biology of Reproduction* **86(2)**:55 DOI 10.1095/biolreprod.111.094227.

O'Rand MG, Widgren EE, Wang Z, Richardson RT. 2006. Eppin: an effective target for male contraception. *Molecular and Cellular Endocrinology* **250(1–2)**:157–162 DOI 10.1016/j.mce.2005.12.039.

Richardson RT, Sivashanmugam P, Hall SH, Hamil KG, Moore PA, Ruben SM, French FS, O'Rand M. 2001. Cloning and sequencing of human Eppin: a novel family of protease inhibitors expressed in the epididymis and testis. *Gene* **270(1–2)**:93–102 DOI 10.1016/s0378-1119(01)00462-0.

Schrodinger, LLC. 2015. *The PyMOL molecular graphics system*. Version 1.8. *Available at* https://pymol.org.

Silva EJR, Hamil KG, O'Rand MG. 2013. Interacting proteins on human spermatozoa: adaptive evolution of the binding of semenogelin I to EPPIN. *PLOS ONE* **8(12)**:e82014 DOI 10.1371/journal.pone.0082014.

Silva EJR, Hamil KG, Richardson RT, O'Rand MG. 2012. Characterization of EPPIN's semenogelin I binding site: a contraceptive drug target1. *Biology of Reproduction* **87(3)**:56 DOI 10.1095/biolreprod.112.101832.

Sun H, Li Y, Tian S, Xu L, Hou T. 2014a. Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set. *Physical Chemistry Chemical Physics* **16(31)**:16719–16729 DOI 10.1039/c4cp01388c.

Sun H, Li Y, Tian S, Xu L, Shen M, Pan P, Guan Y, Hou T. 2014b. Assessing the performance of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute dielectric constant MM/GBSA and MM/PBSA rescoring. *Physical Chemistry Chemical Physics* **16(40)**:22035–22045 DOI 10.1039/c4cp03179b.

Waller D, Bolick D, Lissner E, Premanandan C, Gamerman G. 2016. Azoospermia in rabbits following an intravas injection of Vasalgel. *Basic and Clinical Andrology* **26**:6 DOI 10.1186/s12610-016-0033-8.

Wang Z, Widgren EE, Sivashanmugam P, O'Rand MG, Richardson RT. 2005. Association of eppin with semenogelin on human spermatozoa. *Biology of Reproduction* **72(5)**:1064–1070 DOI 10.1095/biolreprod.104.036483.

Weiser J, Shenkin PS, Still WC. 1999. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *Journal of Computational Chemistry* **20(2)**:217–230 DOI 10.1002/(sici)1096-987x(19990130)20:2<217::aid-jcc4>3.0.co;2-a.

Wiederstein M, Sippl MJ. 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research* **35(Web Server)**:W407–W410 DOI 10.1093/nar/gkm290.

Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2015. The I-TASSER Suite: protein structure and function prediction. *Nature Methods* **12(1)**:7–8 DOI 10.1038/nmeth.3213.

Yenugu S, Richardson RT, Sivashanmugam P, Wang Z, O'Rand MG, French FS, Hall SH. 2004. Antimicrobial activity of human EPPIN, an androgen-regulated, sperm-bound protein with a whey acidic protein motif. *Biology of Reproduction* **71(5)**:1484–1490 DOI 10.1095/biolreprod.104.031567.

Youssef H. 1993. The history of the condom. *Journal of the Royal Society of Medicine* **86**:226–228.