Investigation on substrate specificity and catalytic activity of serine protease neuropsin

Masami Lintuluoto, Mitsumasa Abe, Yota Horioka, Yoshifumi Fukunishi, Hideki Tamura, Juha M. Lintuluoto

Graduate School of Life and Environmental Sciences, Kyoto Prefectural University, Kyoto 606-8522, Japan
Cellular and Molecular Biotechnology Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tokyo 135-0064, Japan
Laboratory of Biofunctional Science, Hoshi University School of Pharmacy and Pharmaceutical Sciences, Tokyo 142-8501, Japan
Institute for Advanced Life Sciences, Hoshi University School of Pharmacy and Pharmaceutical Sciences, Tokyo 142-8501, Japan
Graduate School of Engineering, Department of Synthetic and Biological Chemistry, Kyoto University, Kyoto 615-8530, Japan

Received March 31, 2022; Accepted September 21, 2022; Released online in J-STAGE as advance publication September 22, 2022
Edited by Akio Kitao

Neuropsin is one of serine proteases mainly found at the hippocampus and the amygdala, where it contributes to the long-term potentiation and memory acquisition by rebuilding of synaptic connections. Despite of the importance of neuropsin, the substrate specificity and regulation mechanisms of neuropsin have been unclear. Thus, we investigated the substrate specificity and the catalytic activity of neuropsin by the protein-ligand docking and molecular dynamics (MD) simulations and succeeded to reproduce the trend of the experimental results. Our study revealed that the substrate specificity and the activity of neuropsin depended on multiple factors: the substrate charge, the substrate orientation, the hydrogen bond network within the catalytic triad and the substrate, and the formation of the oxyanion hole. The apo neuropsin was not reactive without proper alignment of catalytic triad. The substrate binding induced the reactive alignment of catalytic triad. Then the substrate-neuropsin interaction forms the oxyanion hole that stabilizes the transition state and reduces the free-energy barrier of the following scission reaction.

Key words: KLK8, molecular dynamics simulation, oxyanion hole, protein-ligand docking

Neuropsin (KLK8) is involved in the signaling pathways related to the reconstruction of the synaptic connections and is related to the cognitive disorder and mental diseases. Since there is not enough 3D structural data for both apo- and holo-neuropsins, we investigated the catalytic activity and the substrate specificity by using the docking and MD simulations. The specific substrate binding induced the formation of catalytic triad, while the binding free energy did not correlate to the catalytic activity. The substrate specificity was proportional to the frequency of substrate binding inducing the formations of the catalytic triad and the oxyanion hole.
Neuropsin (also referred as kallikrein 8, KLK8) belongs to the family of KLK-related secreted proteases that are encoded by highly conserved multigene families [1,2]. Among the KLK families, neuropsin is one of the most interesting because it exhibits specific expression patterns in the central nervous system with high levels in the hippocampus and the amygdala [3], where it contributes to the memory acquisition and anxiety-like behavior through modulation of synaptic plasticity and structures [4,5]. Abnormal expression of the gene has also been implicated in cognitive disorder [6,7], mental diseases such as schizophrenia and bipolar disorder [8], and Alzheimer’s disease [9,10]. In humans, it has been reported that there are several isoforms of neuropsin (type 2-6) [11-14], what has not been identified in the mouse by alternative splicing. Since functional differences between the neuropsin’s splicing subtypes are still unclear, these variants are not analyzed in the present study. However, the KLKs are a group of the serine proteases, which share a high degree of structural homology [1,2]. Figure 1(a) shows the human neuropsin structure. Neuropsin has fourteen β-strands, three α-helices and a catalytic triad (His57, Asp102 and Ser195) which is located at the cleft formed by the seven β-strands [15]. The substrate pocket, where the catalytic triad is located at its bottom, is surrounded by the eight loops, Loops A ~ H as shown in Figure 1(a).

The serine protease is classified by the sequences of amino-acid residues and the structural similarities in two classes, subtilisin-like and chymotrypsin-like serine proteases. Trypsin, chymotrypsin and thrombin are included in chymotrypsin-like serine protease [16], and neuropsin shows the trypsin-like activity, which prefers Arg as the scission site. Chymotrypsin-like proteases have been suggested to share an identical catalytic mechanism, in which the acylation reaction starts by the nucleophilic attack of catalytic Ser to carbonyl of substrate scission site, and then the deacylation reaction follows by the water nucleophilic attack. In the acylation reaction, the catalytic His enhances the nucleophilicity of Ser by the proton abstraction from Ser-Oγ as a general base, and the catalytic Asp stabilizes the protonated His as shown in Figure 1 (b) [16-18]. A tetrahedral intermediate is formed via covalent bonding of SerOγ to the substrate carbonyl carbon. This tetrahedral intermediate is electrostatically stabilized through hydrogen bonds with the backbone amides of Gly193 and Ser195, together forming an “oxyanion hole” [16].

Serine proteases have some substrate selection mechanisms. One of the cation-selection interactions is a salt-bridge between the substrate and protease. Asp189 of trypsin locates at the bottom of the substrate pocket forms the salt bridge with Arg of the substrate, which is the scission site of the substrate (P1) [17]. The salt bridge between Asp and Arg was confirmed in the X-ray structure of neuropsin from *E. coli* with the leupeptin inhibitor complex (PDB ID: 5MS4) [19].

The two-step binding mechanism has been suggested for the substrate binding on the serine protease of the Chymotrypsin family from the MD [20] and computational methods [21,22] with the MEROPS databases [23]. The electrostatic recognition has been reported to be a major factor in the substrate recognition from the investigation applied to the 9 homogeneous Chymotrypsin-like serine proteases [22].

---

**Figure 1** (a) The structure of neuropsin ([PDB ID: 1NPM](https://pdb.org/pdb/explore/explore.do?structId=1NPM), UniProtKB: Q61955). Ball-and-stick models show the residues that form catalytic triad (His57, Asp102 and Ser195) and the key residue of ligand biding (Asp189). Red dashed circle represents the substrate binding cleft. (b) The catalytic triad, oxyanion hole (red dashed circle) and scission site (indicated by blue dashed arrow) in neuropsin. Two black arrows represent the proton relay.
Neuregulin-1 (NRG-1), fibronectin, vitronectin, synaptic adhesion molecule L1, and ephrin type B receptor 2 (EphB2) were reported as substrates of neuropsin [5,24,25]. *NRG-1 gene* generates multiple isoforms, which are classified into six protein types [26-29]. Most NRG-1 isoforms are synthesized as the membrane anchored precursors, so-called pro-NRG-1. The pro-NRG-1 is cleaved at the C-terminal end of EGF-like domain by the transmembrane proteases and releases matured NRG-1 (m-NRG-1), which interacts with heparan sulfate proteoglycan through its heparin binding (HB) domain [30,31]. The HB domain of m-NRG-1 is removed by neuropsin, resulting in the released ligand domain that binds ErbB4 receptor [24]. The ErbB4 is the only NRG-1-specific receptor and the dysregulation of NRG-1-ErbB4 signaling has been reported to contribute the schizophrenia [32].

Neuropsin has been reported to cleave m-NRG-1 at three specific sites between HB and EGF-like domains of m-NRG-1 [24]. Three distinct cleavage sites were KER18/GSG, SLR67/FKW and ELR98/INK, where the back slashes “/” represent the cleavage or scission sites [24]. The enzyme kinetics have been studied by using synthetic peptide substrates with the fluorescent 4-methylcoumaryl-7-amide (MCA) tag [24]. Ac-SLR-MCA and Ac-KDR-MCA have higher scission activities for mouse neuropsin, while the Ac-ELR-MCA, Ac-YGR-MCA and Ac-DVR-MCA have negligible activity, where “Ac” means the acetyl group [24].

We investigated the substrate specificity of neuropsin by using protein-ligand dockings and molecular dynamics (MD) simulations. We used mouse neuropsin and the above five substrates (Ac-SLR-MCA, Ac-KDR-MCA, Ac-ELR-MCA, Ac-YGR-MCA and Ac-DVR-MCA). First, a protein-ligand docking program generated multiple candidate complex structures, and we selected the three docking models after comparing to a human neuropsin with a small inhibitor complex structure. Second, the molecular dynamics simulations refined these docking model structures, and revealed the fluctuations of these model neuropsin-substrate structures. Third, we examined the occurrence of hydrogen bond networks that are essential to the scission reaction based on the MD trajectories and additional energy calculations to elucidate the important residues to substrate binding. The substrate specificity of neuropsin depended on multiple factors. Namely, the frequency of forming the appropriate conformation among the catalytic triad and the formation of oxyanion hole around Arg (P1) that was stabilized by the interaction between Arg (P1) of substrate and Asp189 of neuropsin.

**Materials and Methods**

**Building the Neuropsin-Substrate Complex Models**

We built the three mouse neuropsin-substrate complex model structures for each substrate based on the crystal structure of the mouse hippocampus neuropsin (PDB ID: 1NPM, UniProtKB: Q61955) [15], since the three-dimensional (3D) neuropsin-substrate complex structures with the five substrates in question were not available in the Protein Data Bank. The sequence identity and similarity between human and mouse neuropsins were 75% and 91%, respectively. This high similarity suggested that the 3D structure of human neuropsin should be very similar to that of mouse and the results of mouse might be applied to human. Since Arg76 and Asp77 have not been identified in 1NPM, the atomic coordinates of these amino acids were generated by using the Spanner homology modeling program [33] and the GASH program [34]. The protonated states of the charged residues (Asp, Glu, Arg, Lys, His) were prepared as the standard protonated states at pH7.

The neuropsin-substrate complex models were generated by Sievgene docking program [35] with the maximum volume overlap (MVO) method [36,37]. These programs are freely available from the web site (https://www.mypresto5.jp/). First, a protein-ligand docking program generated multiple candidate complex structures, and the comparison to the human neuropsin-Ac-Leu-Leu-Ar7 complex structure [19] (PDB ID: 5MS4), where Ar7 is a chemically modified Arg residue, the MVO program selected three docking models in which the substrate was well-overlapped to Ac-Leu-Leu-Ar7. The details are described in Supplementary Text S1. Sievgene could reproduce only 20-30% of correct receptor-ligand structures, while Sievgene can produce more than 100 candidate-complex structures. The MVO method can select more reliable structure among many other candidate-complex structures with comparison to the other experimentally correct complex structures, and 60% of the selected structures with the top-ranked score by the MVO method will be correct. Thus, we expected one or more correct complex structures existing among the selected three complex structures for each substrate. We selected the top three scoring conformations as the initial structures of substrate-protein complexes for each substrate. Thus, we prepared 15 initial complex structures for five substrates as shown in Supplementary Text S2 and Figure S1.

**Molecular Dynamics Simulations**

The canonical MD simulations were performed in the explicit water sphere of radius of 36 Å for all the fifteen neuropsin-substrate complex mode structures (=5 substrates x 3 docking poses) generated by Sievgene [35] docking calculations followed by the MVO method [36,37]. The complex structure was embedded in TIP3P [38] water molecules including Cl⁻ and Na⁺ ions at saline density, and the total charge of the systems was neutralized for each run.
neuropsin-substrate complex-model structure. The charges of protein atoms and the force field were originated from AMBER parm99 [39]. The RESP charges of atoms in the substrates were generated at HF/6-31G(d) with the structure optimized at B3LYP/6-31G(d) level by using Gaussian 09 [40], and GAFF [41,42] was used for the force field of the substrates. In all the MD simulations, the fast multipole method (FMM) [43,44] was used for the electrostatic interactions with a cut-off of 12 Å. The cut-off distance of the van der Waals interaction was 12 Å.

For all the fifteen neuropsin-substrate complex systems, we prepared the 3D-coordinates trajectories at 300K in the same manner described as follows. The steepest descent (1000 steps) energy minimization was performed with the distance restrain on the scission site (Arg (P1)) of the substrate and the reaction center (Asp189) of neuropsin, and the position restraints on the main-chain atoms of the neuropsin for all the fifteen model-complex structures. After the energy minimizations, the first equilibration runs were the short-time NVT MD (100 ps) performed at 300 K with the time step of 0.5 fs with the position restraints (the force constants were 10 kcal mol\(^{-1}\) Å\(^2\) on all the solute heavy atoms for equilibration of the solvent water molecules and the counter ions for each system. The second equilibration runs were the short-time NVT MD (1 ns) performed at 300 K with the time step was of 0.5 fs without the position restraints to determining the stable H-C, H-N, and H-O bond distances. The third equilibration runs were the 20 ns NVT simulations with the time step of 1.0 fs at 300 K with using the SHAKE algorithm to fix the H-C, H-N, and H-O distances [45]. Next, the production runs were the 30 ns NVT simulations with the time step of 1.0 fs at 300 K with using the SHAKE algorithm and the snapshot structures were obtained every 1 ps. Finally, 30,000 snapshot structures were obtained from each simulation, and altogether 450,000 snapshots (=30,000 snapshot structures x 3 candidate models x 5 substrates) were used for the analysis. The total NVT MD simulation time was 750 ns (= 50ns x 3 candidate models x 5 substrates). The concatenation of many randomly selected trajectories has been reported, and the results could be close to the canonical ensemble and the time-independent properties could be precise for the given force field [46].

All the MD simulations were performed by cosgene/myPresto program [47], pyMol software [48] was used for the visualization of snapshots from all trajectories, and the apo structure of neuropsin was used for mapping of the electrostatic potential surface of neuropsin with PDB2PQR [49].

**Estimation of the Binding Free Energies for the Peptide-Protein Complex Models**

The binding free energies of the substrates were calculated on the basis of the one-state end-point approach with the MM-Poisson-Bolzmann (PB) SA methods [50-55]. In addition to the binding energy estimations, to elucidate the key interactions among each residue in the ligand recognition, the binding free energy was decomposed into the interaction energy of single residue pair by MM-generalized-Born (GB) SA method [56-64]. The GB method is one of the MM-Poisson-Bolzmann (PB) SA methods [50-55]. In addition to the binding energy estimations, to elucidate the key electrostatic interactions for the given force field [46].

The definition of the receptor-ligand binding free energy (\(\Delta G_{\text{bind}}\)) is the free-energy difference between the complex system (\(G_{\text{complex}}^{\text{wat}}\)) and the summation of that of the isolated receptor (\(G_{\text{receptor}}^{\text{wat}}\)) and isolated ligand (\(G_{\text{ligand}}^{\text{wat}}\)) systems.

\[
\Delta G_{\text{bind}} = G_{\text{complex}}^{\text{wat}} - (G_{\text{receptor}}^{\text{wat}} + G_{\text{ligand}}^{\text{wat}})
\]  

The MM-PBSA method decomposes the free energy into the enthalpy and entropy terms. The change of the average potential energy term approximates the enthalpy (\(\Delta E\)) term, and the change of the water accessible-surface area of the receptor and ligand structures approximates the entropy change (\(\Delta S\)) term of water in the process of dehydration on the interface between the receptor and ligand and the reorganization of water solvent structure. In eq. 1, generally the normal mode analysis estimates the entropy term (\(T\Delta S_{\text{conf}}^{\text{wat}}\)). The uncertainty of the solute entropy estimation by the normal mode analysis has been reported [65].

\[
\Delta G_{\text{bind}}^{\text{MMPSA}} = \Delta E_{\text{wat}}^{\text{desolv}} + T\left(\Delta S_{\text{conf}}^{\text{wat}} + \Delta S_{\text{wat}}^{\text{wat}}\right) - (G_{\text{complex}}^{\text{wat}} - (G_{\text{receptor}}^{\text{wat}} + G_{\text{ligand}}^{\text{wat}}))
\]

In the current study, our MD simulation time (50 ns) was much shorter than the protein domain motions (~ micro seconds) and the molecular sizes of ligand molecules were similar to each other. Therefore, we ignored the conformational entropy changes of the receptor and ligand. Thus, our simplified binding free energy (\(\Delta G_{\text{bind}}^{\text{C}}\)) is given as follows.

\[
\Delta G_{\text{bind}}^{\text{C}} \approx G_{\text{complex}}^{\text{wat}} - (G_{\text{receptor}}^{\text{wat}} + G_{\text{ligand}}^{\text{wat}})
\]

This simplified method is so-called one-state end-point approach, and it has been successfully used for estimating
To estimate roughly the binding free energies and the pairwise interactions among the atoms of the substrate and neuropsin, we applied the MM-GBSA and MM-PBSA calculations to the 3D-coordinates trajectories by using the Ambertools15 [72]. Eq.3 was applied to the 3000 snapshot structures extracted every 10 ps from each trajectory for all complex models and the averaged $\Delta G_{\text{bind}}^S$ values over 3000 structures gave the $\Delta G_{\text{bind}}^S$ values discussed in the following sections. The water molecules and ions in each snapshot structure were removed in these calculations, since the PBSA and GBSA methods were assuming implicit solvent models.

Results and Discussion

Induced Catalytic Triad Formation by Substrate-Binding on Neuropsin

The values of Co-RMSDs of protein in complex models were stable in the last 30 ns simulations as shown in Supplementary Figure S2. The values of RMSDs of substrate in complex models were less than 4.0 Å, and were stable in the last 30 ns simulations as shown in Supplementary Figure S3. Therefore, we used the last 30 ns trajectories for the analysis and the evaluations for the binding free energy.

Figure 1(b) shows the definition of R1-R8 relating to the catalytic reaction and substrate recognition. The catalytic triad needs the formation of the R1-R2 hydrogen bonds pair or R1-R3 hydrogen bonds pair to form the reactive alignment of Ser195-His57-Asp102. The nucleophilic attack to the carbonyl carbon atom of Arg(P1) by Ser195Oγ starts the enzymatic reaction. The distance between the carbonyl carbon atom of Arg(P1) by Ser195Oγ was assigned to R4 as shown in Figure 1(b). The scission of substrate could occur when R4 and the reactive alignment of the catalytic triad are formed at the same time. The hydrogen bond formation R1-R3 was assigned for the distance below 2.5 Å, and the probability $P_1-P_3$ was estimated from the ratio (frequency) of the frame numbers with the hydrogen bond formation to the total number of frames. The probability $P_4$ was estimated from the ratio (frequency) of the frame numbers with $R_4 < 3.5$ Å to the total number of frames. Table 1 summarizes the averaged values of the distances of $R_1-R_4$ and the probabilities $P_1-P_4$.

| Substrate       | $\langle R_1 \rangle$ | $P_1$ | $\langle R_2 \rangle$ | $P_2$ | $\langle R_3 \rangle$ | $P_3$ | $\langle R_4 \rangle$ | $P_4$ | $P_{\text{tripad}}$ | $P_{\text{reactive}}$ |
|-----------------|----------------------|-------|-----------------------|-------|-----------------------|-------|-----------------------|-------|----------------------|----------------------|
| apo neuropsin   | 7.4 ± 0.7            | 0.0   | 2.1 ± 0.2             | 94.5  | 2.0 ± 0.2             | 98.4  | -                     | -     | 0.0                  | 0.0                  |
| Ac-SLR-MCA      | 4.0 ± 0.6            | 16.2  | 2.3 ± 0.2             | 81.5  | 1.9 ± 0.1             | 99.7  | 3.8 ± 0.3             | 55.5  | 16.2                 | 8.96                 |
| Ac-KDR-MCA      | 6.3 ± 0.9            | 3.4   | 2.1 ± 0.3             | 90.7  | 2.1 ± 0.2             | 92.7  | 5.3 ± 0.9             | 42.0  | 3.2                  | 1.32                 |
| Ac-ELR-MCA      | 4.0 ± 0.8            | 1.1   | 2.3 ± 0.2             | 73.6  | 2.0 ± 0.2             | 97.6  | 3.8 ± 0.5             | 18.1  | 1.1                  | 0.20                 |
| Ac-DVR-MCA      | 3.8 ± 0.8            | 31.0  | 2.4 ± 0.5             | 74.6  | 2.1 ± 0.3             | 91.5  | 3.8 ± 0.4             | 51.8  | 28.4                 | 14.69                |
| Ac-YGR-MCA      | 4.1 ± 0.6            | 5.6   | 3.1 ± 0.8             | 55.6  | 2.6 ± 0.7             | 79.4  | 3.9 ± 0.5             | 33.3  | 4.4                  | 1.48                 |

$^a$ $R_{1-4}$ represent the distances of Ser195Oγ-His57Nε, His57Nδ-His102Oγ1, His57Nδ-His102Oδ2, and Ser195Oγ-Arg(P1)C as shown in Figure 1 (b). Each value was averaged across the three substrate-protein complex models. The values after “±” symbols show the standard deviations.

The substrates are listed in the order of the experimental reactivity.

Thus, we approximated the probability of formation of catalytic triad ($P_{\text{tripad}}$) as

$$P_{\text{tripad}} = P_1 \cdot \max(P_2, P_3),$$

and the probability of formation of reactive conformation ($P_{\text{reactive}}$) as

$$P_{\text{reactive}} = P_{\text{tripad}} \cdot P_4.$$
where we assumed that all the events occur independently to each other. It should be noted that the trend of $P_{\text{reactive}}$
alone did not follow the experimental reactivity.

The averaged value of $R_1$ (the distance between His57N$^\varepsilon$ and Ser195O$^\gamma$) was 7.44 Å in apo neuropsin, meaning that the apo neuropsin did not form the catalytic triad. However, the substrate bindings reduced the $R_1$ value by 1-3 Å and increased $P_{\text{reactive}}$ value of apo up to 28%. These results suggested that the substrate binding is essential to activate the catalytic reactivity of neuropsin, and the substrate-recognition should depend on the substrate-binding energy.

This induced reactivity by substrate-binding feature made neuropsin different from the other digestive enzymes. Namely, digestive enzymes of the serine protease family form the reactive alignment of catalytic triad in apo form before the substrate bindings. The geometry around the catalytic triad of neuropsin was similar to those of the other chymotrypsin-like serine proteases, except $R_1$. The $R_1$ in the digestive enzymes of the serine protease family are ~ 3.0 Å [73-75]. The three catalytic residues are located at the bottom of the substrate pocket surrounding the eight flexible loops in KLK8 neuropsin, and the distances among the three catalytic residues may be affected by the fluctuation of the eight loops. The previous studies for the KLK protein family suggested that the alignments of these three catalytic residues may be disrupted due to the fluctuation of the eight loops in the KLK protein family [76].

### Substrate-Dependent Oxyanion Hole Formation

The hydrogen bonds in the oxyanion hole were suggested to stabilize both the ground and transition states of the ligand binding complex [77,78]. The nucleophilic attack by the Ser195O$^\gamma$ to the carbonyl carbon of substrate Arg (P1) should occur to form the tetrahedral intermediate at the early stage in the enzyme reaction of the serine protease as shown in Figure 1 (b). Thus, we next focused on the formation of the oxyanion hole at the substrate pocket. Figure 1(b) suggests that the formation of the three hydrogen bonds around the scission site creates the oxyanion hole. Table 2 summarizes the distances among the hydrogen donors and acceptors: the distances between carbonyl oxygen of substrate Arg (P1) and NH of Gly193 (R6), Asp194 (R7) and Ser195 (R8).

The probability of the formation of hydrogen bond was estimated in the same manner in the above section ($P_{R_6-R_7}$). Table 2 summarizes the probabilities of forming the hydrogen bonds that create the oxyanion hole. The last column shows the total frequency of each ligand. That is a simple product of the three frequencies, and assumes the three hydrogen bonds occurring independently. The trend of the total frequency reproduced the experimental reactivity of these five ligands.

#### Table 2

| Substrate     | $<R>$ (Å) | $P_6$ (%) | $<R>$ (Å) | $P_7$ (%) | $<R>$ (Å) | $P_8$ (%) | $P_{\text{oxyanion}}$ (%) | $P_{\text{total}}$ (%) | $k_{\text{cat}}/K_{\text{m}}$ (mM$^{-1}$min$^{-1}$) |
|---------------|-----------|-----------|-----------|-----------|-----------|-----------|--------------------------|------------------------|---------------------------------|
| Ac-SLR-MCA    | 2.0 ± 0.2 | 97.0      | 2.9 ± 0.3 | 22.7      | 2.7 ± 0.4 | 44.9      | 9.9                      | 0.89                   | 128                             |
| Ac-KDR-MCA    | 2.8 ± 0.2 | 65.1      | 3.7 ± 0.4 | 22.3      | 4.4 ± 0.4 | 32.1      | 4.7                      | 0.06                   | 31.2                            |
| Ac-ELR-MCA    | 3.9 ± 0.4 | 0.1       | 4.6 ± 0.4 | 0.0       | 4.6 ± 0.4 | 0.0       | 0.0                      | 0.00                   | 12.5                            |
| Ac-DVR-MCA    | 3.5 ± 0.4 | 24.4      | 4.4 ± 0.4 | 15.7      | 3.6 ± 0.4 | 31.7      | 1.2                      | 0.18                   | 3.54                            |
| Ac-YGR-MCA    | 2.1 ± 0.3 | 86.4      | 3.1 ± 0.3 | 6.2       | 2.9 ± 0.4 | 22.9      | 1.2                      | 0.02                   | 2.31                            |

$^a$ $R_6$-$8$ represent the distances of Gly193NH- Arg(P1)O, Asp194NH- Arg(P1)O and Ser195NH- Arg(P1)O as shown in Figure 1 (b). Each value was averaged across the three substrate-protein complex models. The values after “±” symbols show the standard deviations. $^b$ Reference 24

We approximated the probability of formation of oxyanion hole ($P_{\text{oxyanion}}$) as

$$P_{\text{oxyanion}} = P_6 \cdot P_7 \cdot P_8. \quad (6)$$

and the probability of occurrence of reaction ($P_{\text{total}}$) as

$$P_{\text{total}} = P_{\text{reactive}} \cdot P_{\text{oxyanion}}. \quad (7)$$
where we assumed that all the hydrogen bond formations occur independently to each other. Here, the trend of $P_{total}$ was almost proportional to that of the experimental specificity (the Pearson correlation coefficient was 0.95), and especially, the trend of $P_{oxygen}$ reproduced that of the experimental specificity (the Pearson correlation coefficient was 0.96), as summarized in Table 2 and Figure 2.

We counted the frame numbers that satisfied all the required conditions ($R_1$-$R_3 < 2.5 \text{ Å}, R_4 < 3.5 \text{ Å}$ and $R_5$-$R_8 < 2.5 \text{ Å}$) at the same time along the trajectories, and calculated the probability $P_{total}'$ estimated from the ratio (frequency) of the frame numbers to the total number of frames. The values of $P_{total}'$ were consistent with those of $P_{total}$ as shown in Figure 2 (a). Furthermore, the trend of $P_{total}'$ was almost proportional to that of the experimental specificity (the Pearson correlation coefficient was 0.94) as shown in Figure 2 (a). It seems that all the hydrogen bond events occur independently.

Calculated Affinities of Substrates in Complex Models

As mentioned above section, the formation of the hydrogen bond network among the reactive catalytic triad by the substrate binding was required for the nucleophilic attack to the scission site of substrate by Ser195, otherwise the nucleophilicity of Ser195 was remained to be not enough. The weak ligand binding free energy was not enough to induce the hydrogen bond network among the catalytic-triad, and Ser195 could not work effectively as a nucleophile. Therefore, we estimated the substrate-neuropsin binding free energy and examined the key-interaction between the substrate and the neuropsin.

Figure 3 shows the electrostatic potential surface of the last snapshot from the MD simulation of the apo neuropsin. The substrate pockets (cleft) including the catalytic triad residues and the Asp189 were negatively charged. In fact, the salt bridge $R_5$ between Arg(P1) and Asp189 of neuropsin (Figure 1(b)) was kept 97-100% in all five substrates during all the MD simulations.

Table 3 shows the calculated $\Delta G_{MM-PBSA}$ values based on the MD trajectories. The negative values of the electrostatic potentials around the substrate pocket, as shown in Figure 3, made the binding of positively charged substrates more favorable than neutral substrates. Our results are consistent with the results from the computational investigation with electrostatic molecular interaction fields (eMIFs), in which the four serine proteases among the nine Chymotrypsin-type serine proteases, especially kallikrein-1, favor the positive probe in their substrate pockets [22]. The trend of the $\Delta G_{MM-PBSA}$ values did not reproduce the substrate specificity. The $\Delta G_{MM-PBSA}$ value of Ac-SLR-MCA (-25.10 kcal/mol) was almost equal to that of Ac-KDR-MCA (-25.41 kcal/mol), but Ac-SLR-MCA has higher reactivity than Ac-KDR-MCA as shown in Table 2. The reason of the difference in reactivity should be due to the difference of binding poses of them. Both Arg(P1) of Ac-SLR-MCA and Ac-KDR-MCA formed the salt-bridges with Asp189 and hydrogen bonds with the oxygen-hole residues (Gly193, Asp194 and Ser195) as shown in Figures 4 and 5. The salt bridge was the main contributor to the binding energy, and the hydrogen bonds with the oxygen-hole residues also contributed to the...
binding energy. The difference of Ac-SLR-MCA and Ac-KDR-MCA was that the former weakly bound His57 but the latter did not. Ac-SLR-MCA showed slightly larger number of hydrogen bonds with the oxyanion-hole residues than Ac-KDR-MCA. Ac-KDR-MCA interacted with Loop E, while Ac-SLR-MCA did not. Thus, the binding affinities of Ac-SLR-MCA and Ac-KDR-MCA were almost the same, while the binding poses were different.

To proceed the catalytic reaction, Ser195, the substrate Arg(P1) and His57 should be close to each other. Ser195 and His57 were on the different loops as shown in Figure 1, and the distance between them was longer than the reactive conformation before the substrate binding. Ac-SLR-MCA interacted with both His57 and Ser195 more strongly compared to Ac-KDR-MCA. These interactions in Ac-SLR-MCA complex could reduce the distance between His57 and Ser195 to realize the more reactive conformation rather than Ac-KDR-MCA. These differences should be the reason why the frequencies of formation of oxyanion hole and catalytic triad of Ac-KDR-MCA were smaller than those of Ac-SLR-MCA. As the result, Ac-SLR-MCA should have higher reactivity than Ac-KDR-MCA. The details of difference of the receptor-ligand interactions between Ac-KDR-MCA and Ac-SLR-MCA are shown in Supplementary Text S2 and the next section “Substrate specificity” (Figures 4 and 5).

Substrate Specificity

Our computational results summarized in $P_{\text{reactive}}$, $P_{\text{oxyanion}}$ and $P_{\text{total}}$ reproduced the experimental trend of substrate reactivity. And our binding energy estimation suggested that neuropsin could bind and hold these substrates to align the catalytic triad into the reactive conformation, except for Ac-ELR-MCA. These results showed that our neuropsin-substrate complex structures and the simulation results were reasonable. Thus, we used the MD simulation to examine which residues were important to the substrate recognition (or substrate specificity) of the neuropsin.

We investigated the details of the important interaction in the substrate complex with the high specificity Ac-SLR-MCA and Ac-KDR-MCA models and negligible specificity Ac-YGR-MCA model, and estimated the contribution of each amino-acid residue for the binding free energy by using $\Delta G_{\text{MM-GBSA}}$ values. Figures 4-6 (a) show the binding conformations and the binding energy decompositions of each amino residue. The absolute values of the interaction energies which were > 2 kcal/mol were selected. The important interactions in the complexes are shown in Figures 4-6 (b).

The Asp189 and its neighbor residues like Thr190 were commonly main contributors to stabilize the complex in all complex models. Furthermore, the residues belonging to Loop H (Ser214, Trp215, Ser217 and so on), contributed to stabilize the complex. In addition, the backbone NH of the amino-acid residues of Gly193, Asp194 and Ser195 (Loop G) stabilized the complex by interacting with the carbonyl oxygen of substrate Arg (P1). These residues with the carbonyl oxygen of substrate were composed of the oxyanion hole in the substrate pocket. The hydrogen bonds in the

**Table 3** $\Delta G_{\text{MM-PBSA}}$ values (kcal mol$^{-1}$) and total charges of the five substrates.

| Substrate     | $\Delta G_{\text{MM-PBSA}}^{a}$ | Total charge of substrate |
|---------------|---------------------------------|---------------------------|
| Ac-SLR-MCA    | -25.10 ± 3.9                   | 1                         |
| Ac-KDR-MCA    | -25.41 ± 3.9                   | 1                         |
| Ac-ELR-MCA    | -6.21 ± 5.4                    | 0                         |
| Ac-DVR-MCA    | -19.11 ± 3.4                   | 0                         |
| Ac-YGR-MCA    | -28.01 ± 3.9                   | 1                         |

$^{a}$Each value was averaged across the three substrate-protein complex models. The values after “±” symbols show the standard deviations. The substrates are listed in the order of the experimental reactivity.
Figure 4 The interactions and the orientation of substrate to the protein in Ac-SLR-MCA model. (a) The interaction energy decomposed to the pair of residues-substrate. (b) The important residues for the interaction between the substrate and the protein. The tube model represents Ac-SLR-MCA molecule, and the red, blue and green circles represent the salt bridge, hydrogen bond and the hydrophobic interaction, respectively. (c) The oxyanion hole and the interactions in the complex model. The Arg (P1) in the substrate is only shown in the tube model. The broken red and pink lines represent the hydrogen bonds in the oxyanion hole and among the catalytic triad, respectively. (d) The substrate on the protein surface.

Figure 5 The interactions and the orientation of substrate to the protein in Ac-KDR-MCA model. (a) The interaction energy decomposed to the pair of residues-substrate. (b) The important residues for the interaction between the substrate and the protein. The tube model represents Ac-KDR-MCA molecule, and the red, blue and green circles represent the salt bridge, hydrogen bond and the hydrophobic interaction, respectively. (c) The oxyanion hole and the interactions in the complex model. The Arg (P1) in the substrate is only shown in the tube model. The broken pink lines represent the hydrogen bonds in the oxyanion hole and among the catalytic triad, respectively. (d) The substrate on the protein surface.

Figure 6 The interactions and the orientation of substrate to the protein in Ac-YGR-MCA model. (a) The interaction energy decomposed to the pair of residues-substrate. (b) The important residues for the interaction between the substrate and the protein. The tube model represents Ac-YGR-MCA molecule, and the red, blue and green circles represent the salt bridge, hydrogen bond and the hydrophobic interaction, respectively. (c) The oxyanion hole and the interactions in the complex model. The Arg (P1) in the substrate is only shown in the tube model. The broken pink lines represent the hydrogen bonds in the oxyanion hole and among the catalytic triad, respectively. (d) The substrate on the protein surface.
oxyanion hole were suggested to stabilize both the ground and transition states of the ligand binding complex [77,79]. The hydrogen bonds in the oxyanion hole seemed to be formed in the Ac-SLR-MCA, Ac-KDR-MCA and Ac-YGR-MCA models as shown in Figures 4 ~ 6 (c). Ser of the substrate (P3) interacted with Lys175 (Loop F) and Gly261 (Loop H) in the Ac-SLR-MCA model, and there was interaction between Leu12 and His57 as shown in Figure 4 (b). The substrate of Ac-SLR-MCA laid on the surface starting from the Loop H toward the Loop G as shown in Figure 4 (d). Lys of substrate (P3) in the Ac-KDR-MCA model interacted with the residues included in the Loop E as shown in Figure 5 (b), and the substrates of both the Ac-KDR-MCA and Ac-YGR-MCA models laid on the surface starting from the Loop E toward the Loop H via the Loop G as shown in Figures 5 (d) and 6 (d). The MCA tag interacted with the Loop H and the catalytic residues in the Ac-YGR-MCA model as shown in Figure 6 (b).

Conclusion

We investigated the substrate specificity and the catalytic activity of neuropsin by using the MD simulations with the fifteen neuropsin-substrate complex models (= 3 complex models for each substrate x 5 substrates). The substrate affinity depended on the charge of the substrate, and the more positively charged substrates have higher affinities due to the negative charges around the substrate pocket. Multiple interactions, such as interactions of substrate with the Loops G and H and the salt-bridge between substrate Arg (P1) and Asp189 produced sufficient stabilization of the complex and resulted in the local conformational change of the relative position among the catalytic triad. The substrates with higher specificity have more favorable hydrogen bond network among the catalytic triad due to the local conformational change. Furthermore, the substrates with the higher substrate specificity formed the oxyanion hole with Gly193 and Ser195.

Our results showed that the substrate affinity and the orientation of the substrate determined the catalytic activity, and we think that our results open a door to investigate the signaling mechanism beginning from the neuropsin-NGR-1.

Conflict of Interest

The authors declare no competing financial interest.

Author Contributions

L. M, M. A , and Y. H performed the overall simulations and analyses. All authors discussed the results and wrote the paper. All authors gave final approval of the version to be published.

Acknowledgements

The computations in this work were performed at the Research Center for Computational Science, Okazaki, Japan (20-IMS-C058). The present study was supported by the Project Focused on Developing Key Technology for Discovering and Manufacturing Drugs for Next-Generation Treatment and Diagnosis (2018-2021 and 2021-) from Japan Agency for Medical Research and Development (AMED).

References

[1] Prassas, I., Eissa, A., Poda, G., Diamandis, E.P. Unleashing the therapeutic potential of human kallikrein-related serine proteases. Nat. Rev. Drug Discov. 14, 183-202 (2015). https://doi.org/10.1038/nrd4534
[2] Guo, S., Skala, W., Magdolen, V., Brandstetter, H., Goettig, P. Sweetened kallikrein-related peptidases (KLKs): Glycan trees as potential regulators of activation and activity. Biol. Chem. 395, 959-976 (2014). https://doi.org/10.1515/hsz-2014-0140
[3] Chen, Z. L., Yoshida, S., Kato, K., Momota, Y., Suzuki, J., Tanaka, T., et al. Expression and activity-dependent changes of a novel limbic-serine protease gene in the hippocampus. J. Neurosci. 15, 5088-5097 (1995). https://doi.org/10.1523/JNEUROSCI.15-07-05088.1995
[4] Tamura, H., Ishikawa, Y., Hino, N., Maeda, M., Yoshida, S., Kaku, S., et al. Neuropsin is essential for early processes of memory acquisition and Schaffer collateral long-term potentiation in adult mouse hippocampus in vivo. J. Physiol. 570, 541-551 (2006). https://doi.org/10.1113/jphysiol.2005.098715
[5] Attwood, B. K., Bourgognon, J.-M., Patel, S., Mucha, M., Schiavon, E., Skrzypiec, A. E., et al. Neuropsin cleaves EphB2 in the amygdala to control anxiety. Nature 473, 372-375 (2011). https://doi.org/10.1038/nature09938
[6] Horii, Y., Yamasaki, N., Miyakawa, T., Shiosaka, S. Increased anxiety-like behavior in neuropsin (kallikrein-related peptidase 8) gene-deficient mice. Behav. Neurosci. 122, 498-504 (2008). https://doi.org/10.1037/0735-7044.122.3.498

[7] Tamura, H., Ishikawa, Y., Shiosaka, S. Does extracellular proteolysis control mammalian cognition? Rev. Neurosci. 24, 365-374 (2013). https://doi.org/10.1515/revneuro-2013-0007

[8] Izumi, A., Iijima, Y., Noguchi, H., Numakawa, T., Okada, T., Hori, H., et al. Genetic variations of human neuropsin gene and psychiatric disorders: Polymorphism screening and possible association with bipolar disorder and cognitive functions. Neuropsychopharmacology 33, 3237-3245 (2008). https://doi.org/10.1038/npp.2008.29

[9] Shimizu-Okabe, C., Yousef, G. M., Diamandis, E. P., Yoshida, S., Shiosaka, S., Fahnestock, M. Expression of the kallikrein gene family in normal and Alzheimer's disease brain. Neuroreport 12, 2747-2751 (2001). https://doi.org/10.1097/00001756-200108280-00031

[10] Herring, A., Munster, Y., Akkaya, T., Moghaddam, S., Deinsberger, K., Meyer, J., et al. Kallikrein-8 inhibition attenuates Alzheimer's disease pathology in mice. Alzheimers Dement 12, 1273–1287 (2016). https://doi.org/10.1016/j.jalz.2016.05.006

[11] Mitsui, S., Tsuruoka, N., Yamashiro, K., Nakazato, H., Yamaguchi, N. A novel form of human neuropsin, a brain-related serine protease, is generated by alternative splicing and is expressed preferentially in human adult brain. Eur. J. Biochem. 260, 627-634 (1999). https://doi.org/10.1046/j.1432-1327.1999.00213.x

[12] Magklara, A., Scorilas, A., Katakasos, D., Massobrio, M., Yousef, G. M., Fracchioli, Y. S., et al. The human KLK8 (neuropsin/ovasins) gene: Identification of two novel splice variants and its prognostic value in ovarian cancer. Clin. Cancer Res. 7, 806-811 (2001).

[13] Sher, Y.-P., Chou, C.-C., Chou, R.-H., Wu, H.-M., Chang, W.-S. W., Chen, C.-H., et al. Human kallikrein 8 protease confers a favorable clinical outcome in non-small cell lung cancer by suppressing tumor cell invasiveness. Cancer Res. 66, 11763-11770 (2006). https://doi.org/10.1158/0008-5472.CAN-06-3165

[14] Planque, C., Choi, Y.-H., Guyenet, S., Heuzé-Vourc'h, N., Briollais, L., Courty, Y. Alternative splicing variant of kallikrein-related peptidase 8 as an independent predictor of unfavorable prognosis in lung cancer. Clin. Chem. 56, 987-997 (2010). https://doi.org/10.1373/clinchem.2009.138917

[15] Kishi, T., Kato, M., Shimizu, T., Kato, K., Matsumoto, K., Yoshida, S., et al. Crystal structure of neuropsin, a hippocampal protease involved in kindling epileptogenesis. J. Biol. Chem. 274, 4220-4224 (1999). https://doi.org/10.1074/jbc.274.7.4220

[16] Hedstrom, L. Serine protease mechanism and specificity. Chem. Rev. 102, 4501-4524 (2002).

[17] Perona, J. J., Craik, C. S. Structural basis of substrate specificity in the serine proteases. Protein Sci. 4, 337-360 (1995). https://doi.org/10.1002/pro.5560040301

[18] Ishida, T., Kato, S. Role of Asp102 in the catalytic relay system of serine proteases: A theoretical study. J. Am. Chem. Soc. 126, 7111-7118 (2004). https://doi.org/10.1021/ja030405u

[19] Debela, M., Magdolen, V., Skala, W., Elsässer, B., Schneider, E. L., Craik, C. S., et al. Structural determinants of specificity and regulation of activity in the allosteric loop network of human KLK8/neuropsin. Sci. Rep. 8, 10705 (2018). https://doi.org/10.1038/s41598-018-29058-6

[20] Kahler, U., Kamenik, A. S., Waibl, F., Kraml, J., Liedl, K. R. Protein-protein binding as a two-step mechanism: Preselection of encounter poses during the binding of BPTI and trypsin. Biophys. J. 119, 652-666 (2020). https://doi.org/10.1016/j.bpj.2020.06.032

[21] Kraml, J., Kamenik, A. S., Waibl, F., Kranl, J., Liedl, K. R. Protein-protein binding as a two-step mechanism: Preselection of encounter poses during the binding of BPTI and trypsin. Biophys. J. 119, 652-666 (2020). https://doi.org/10.1016/j.bpj.2020.06.032

[22] Harrison, P. J., Law, A. J. Neuregulin 1 and schizophrenia: Genetics, gene expression, and neurobiology. Biol. Psychiatry 60, 132-140 (2006). https://doi.org/10.1016/j.biopsych.2005.11.002
[27] Meyer, D., Yamaai, T., Garratt, A., Riethmacher-Sonnenberg, E., Kane, D., Theill, L. E., et al. Isoform-specific expression and function of neuregulin. Development 124, 3575-3586 (1997). https://doi.org/10.1242/dev.124.18.3575

[28] Steinthorsdottir, V., Stefánsson, H., Ghosh, S., Birgisdottir, B., Bjornsdottir, S., Fasquel, A. C., et al. Multiple novel transcription initiation sites for NRG1. Gene 342, 97-105 (2004). https://doi.org/10.1016/j.gene.2004.07.029

[29] Tan, W., Wang, Y., Gold, B., Chen, J., Dean, M., Harrison, P. J., et al. Molecular cloning of a brain-specific, developmentally regulated neuregulin 1 (NRG1) isoform and identification of a functional promoter variant associated with schizophrenia. J. Biol. Chem. 282, 24343-24351 (2007). https://doi.org/10.1074/jbc.M702953200

[30] Ozaki, M., Itoh, K., Miyakawa, Y., Kishida, H., Hashikawa, T. Protein processing and releases of neuregulin-1 are regulated in an activity-dependent manner. J. Neurochem. 91, 176-188 (2004). https://doi.org/10.1111/j.1471-4159.2004.02719.x

[31] Loeb, J. A. Neuregulin: An activity-dependent synaptic modulator at the neuromuscular junction. J. Neurocytol. 32, 649-664 (2003). https://doi.org/10.1023/B:NEUR.0000020640.84708.35

[32] Mei, L., Xiong, W.-C. Neuregulin 1 in neural development, synaptic plasticity and schizophrenia. Nat. Rev. Neurosci. 9, 437-452 (2008). https://doi.org/10.1038/nrn2392

[33] Standley, D. M., Toh, H., Nakamura, H. ASH structure alignment package: Sensitivity and selectivity in domain classification. BMC Bioinformatics 8, 116 (2007). https://doi.org/10.1186/1471-2105-8-116

[34] Fukunishi, Y., Nakamura, H. Prediction of protein–ligand complex structure by docking software guided by other complex structures. J. Mol. Graph. Model. 26, 1030-1033 (2008). https://doi.org/10.1016/j.jmgm.2007.07.001

[35] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., Klein, M. L. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79, 926-935 (1983). https://doi.org/10.1063/1.445869

[36] Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., Baker, N. A. PDB2PQR: An automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. Nucleic Acids Res. 32, W665-W667 (2004). https://doi.org/10.1093/nar/gkh096

[37] Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., Baker, N. A. PDB2PQR: An automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. Nucleic Acids Res. 32, W665-W667 (2004). https://doi.org/10.1093/nar/gkh096
Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., et al. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. Acc. Chem. Res. 33, 889-897 (2000). https://doi.org/10.1021/ar000033j

Vorobjev, Y. N., Almagro, J. C., Hermans, J. Discrimination between native and intentionally misfolded conformations of proteins: ES/IS, a new method for calculating conformational free energy that uses both dynamics simulations with an explicit solvent and an implicit solvent continuum model. Proteins 32, 399-413 (1998). https://doi.org/10.1002/(SICI)1097-0134(19980901)32:4<399::AID-PROT1>3.0.CO;2-C

Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., Case, D. A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramide− DNA helices. J. Am. Chem. Soc. 120, 9401-9409 (1998). https://doi.org/10.1021/ja981844+

Fogolari, F., Brigo, A., Molinari, H. Protocol for MM/PBSA molecular dynamics simulations of proteins. Biophys. J. 85, 159-166 (2003). https://doi.org/10.1016/S0006-3495(03)74462-2

Wang, J., Morin, P., Wang, W., Kollman, P. A. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. J. Am. Chem. Soc. 123, 5221-5230 (2001). https://doi.org/10.1021/ja003834q

Massova, I., Kollman, P. A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. Perspect. Drug Discov. Des. 18, 113-135 (2000). https://doi.org/10.1021/ja00022a017

Yang, C.-Y., Sun, H., Chen, J., Nikolovska-Coleska, Z., Wang, S. Importance of ligand reorganization free energy in protein−ligand binding−affinity prediction. J. Am. Chem. Soc. 131, 13709-13721 (2009). https://doi.org/10.1021/ja9039373

Watanabe, Y. S., Fukunishi, Y., Nakamura, H. Generation of a flexible loop structural ensemble and its application to induced-fit structural changes following ligand binding. Biophysics 2, 1-12 (2006). https://doi.org/10.2142/biophysics.2.1

Watanabe, Y. S., Kim, J. G., Fukunishi, Y., Nakamura, H. Free energy landscapes of small peptides in an implicit solvent model determined by force-biased multicanonical molecular dynamics simulation. Chem. Phys. Lett. 400, 258-263 (2004). https://doi.org/10.1016/j.cplett.2004.10.118

Cramer, C. J., Truhlar, D. G. General parameterized SCF model for free energies of solvation in aqueous solution. J. Am. Chem. Soc. 113, 8305-8311 (1991). https://doi.org/10.1021/ja00022a017

Hawkins, G. D., Cramer, C. J., Truhlar, D. G. Pairwise solute descreening of solute charges from a dielectric medium. Chem. Phys. Lett. 246, 122-129 (1995). https://doi.org/10.1016/0009-2614(95)01082-K

Onufriev, A., Bashford, D., Case, D. A. Modification of the generalized Born model suitable for macromolecules. J. Phys. Chem. B 104, 3712-3720 (2000). https://doi.org/10.1021/jp994072s

Schaefer, M., Karplus, M. A comprehensive analytical treatment of continuum electrostatics. J. Phys. Chem. 100, 1578-1599 (1996). https://doi.org/10.1021/jp9521621

Still, W. C., Tempczyk, A., Hawley, R. C., Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. J. Am. Chem. Soc. 112, 6127-6129 (1990). https://doi.org/10.1021/ja00172a038

Lee, M. S., Olson, M. A. Calculation of absolute protein-ligand binding affinity using path and endpoint approaches. Biophys. J. 90, 864-877 (2006). https://doi.org/10.1529/biophysj.105.071589

Kuhn, B., Kollman, P. A. Binding of a diverse set of ligands to avidin and streptavidin: An accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. J. Med. Chem. 43, 3786-3791 (2000). https://doi.org/10.1021/jm000241h

Sa, R., Fang, L., Huang, M., Li, Q., Wei, Y., Wu, K. Evaluation of interactions between urokinase plasminogen and inhibitors using molecular dynamic simulation and free-energy calculation. J. Phys. Chem. A 118, 9113-9119 (2014). https://doi.org/10.1021/jp5064319

Tian, Z., Liu, J., Zhang, Y. Key residues involved in the interaction between Cydia pomonella pheromone binding protein 1 (CpomPBPI) and Codlemone. J. Agric. Food Chem. 64, 7994-8001 (2016). https://doi.org/10.1021/acs.jafc.6b02843

Duan, L., Liu, X., Zhang, J. Z. Interaction entropy: A new paradigm for highly efficient and reliable computation of protein−ligand binding free energy. J. Am. Chem. Soc. 138, 5722-5728 (2016). https://doi.org/10.1021/jacs.6b02682

Wan, S., Knapp, B., Wright, D. W., Deane, C. M., Coveney, P. V. Rapid, Precise, and Reproducible Prediction of Peptide–MHC Binding Affinities from Molecular Dynamics That Correlate Well with Experiment. J. Chem. Theory Comput. 11, 3346-3356 (2015). https://doi.org/10.1021/acs.jctc.5b00179
[71] Cao, R., Jin, Y., Xu, D. Recognition of cello-oligosaccharides by CBM17 from *Clostridium cellulovorans*: Molecular dynamics simulation. J. Phys. Chem. B 116, 6087-6096 (2012). https://doi.org/10.1021/jp3010647

[72] Case, D., Berryman, J., Betz, R., Cheatham III, T., Darden, T., et al. AMBER 2015. University of California, San Francisco, (2015).

[73] Frigerio, F., Coda, A., Pugliese, L., Lionetti, C., Menegatti, E., Amiconi, G., et al. Crystal and molecular structure of the bovine α-chymotrypsin-eglin c complex at 2.0 Å resolution. J. Mol. Biol. 225, 107-123 (1992). https://doi.org/10.1016/0022-2836(92)91029-o

[74] Howard, N., Abell, C., Blakemore, W., Chessari, G., Congreve, M., Howard, S., et al. Application of fragment screening and fragment linking to the discovery of novel thrombin inhibitors. J. Med. Chem. 49, 1346-1355 (2006). https://doi.org/10.1021/jm050850v

[75] Leiros, H.-K. S., Brandsdal, B. O., Andersen, O. A., Os, V., Leiros, I., Helland, R., et al. Trypsin specificity as elucidated by LIE calculations, X-ray structures, and association constant measurements. Protein Sci. 13, 1056-1070 (2004). https://doi.org/10.1110/ps.03498604

[76] Goettig, P., Magdolen, V., Brandstetter, H. Natural and synthetic inhibitors of kallikrein-related peptidases (KLKs). Biochimie 92, 1546-1567 (2010). https://doi.org/10.1016/j.biochi.2010.06.022

[77] Simon, L., Goodman, J. M. Enzyme catalysis by hydrogen bonds: The balance between transition state binding and substrate binding in oxyanion holes. J. Org. Chem. 75, 1831-1840 (2010). https://doi.org/10.1021/jo901503d

[78] Kraut, D. A., Sigala, P. A., Pybus, B., Liu, C. W., Ringe, D., Petsko, G. A., et al. Testing electrostatic complementarity in enzyme catalysis: Hydrogen bonding in the ketosteroid isomerase oxyanion hole. PLoS Biol. 4, e99 (2006). https://doi.org/10.1371/journal.pbio.0040099

[79] Di Cera, E. Serine proteases. IUBMB Life 61, 510-515 (2009). https://doi.org/10.1002/iub.186