Journal of Medicinal Chemistry

Brief Article

pubs.acs.org/jmc

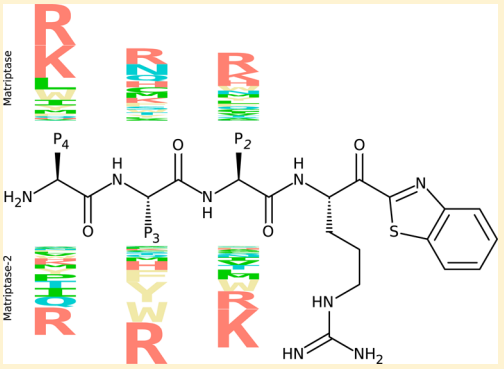
# Analysis of Subpocket Selectivity and Identification of Potent Selective Inhibitors for Matriptase and Matriptase-2

Dominic Duchêne,[†] Eloïc Colombo,[‡] Antoine Désilets,[‡] Pierre-Luc Boudreault,[‡] Richard Leduc,[‡] Eric Marsault,[‡] and Rafael Najmanovich*[,†]

[†]Departments of Biochemistry and [‡]Pharmacology, Faculty of Medicine and Health Sciences, Université de Sherbrooke, 3001 12e Avenue Nord, Sherbrooke, Quebec J1H 5N4, Canada

Supporting Information

**ABSTRACT:** We studied the factors affecting the selectivity of peptidomimetic inhibitors of the highly homologous proteases matriptase and matriptase-2 across subpockets using docking simulations. We observed that the farther away a subpocket is located from the catalytic site, the more pronounced its role in selectivity. As a result of our exhaustive virtual screening, we biochemically validated novel potent and selective inhibitors of both enzymes.




## INTRODUCTION

Proteases possess a broad range of functions in cellular processes, ranging from nonspecific digestive enzymes, maturation of peptidic hormone precursors or growth factors, tissue remodelling, or activation of specific zymogens. Therefore, deregulation or mutations of proteases can have critical consequences on physiological processes. Proteolytic enzymes have been classified in different families depending on their catalytic mechanism and homology.[1] Serine proteases account for almost one-third of all proteases.[2] The S1 serine protease family contains the catalytic triad histidine, aspartate, and serine. We focused on the S1A subfamily, which includes members of the novel family of type II transmembrane serine proteases (TTSPs).[3]

TTSPs are mostly found at the plasma membrane with an intracellular N-terminal domain, a transmembrane domain, and an exposed extracellular stem region composed of up to 11 structural domains terminated by a C-terminal catalytic domain.[3] TTSPs are further divided into four subfamilies: the human airway trypsin-like protease/differentially expressed in squamous cell carcinoma (HAT/DESC), hepsin/transmembrane protease/serine (TMPRSS), corin, and matriptase. Matriptase (also known as ST-14 for suppressor of tumorigenicity 14 protein) is one of the most studied members of his subfamily, followed by matriptase-2 (also known as TMPRSS6).

Matriptase is expressed in epithelial cells as a zymogen[4] and needs to be activated. It undergoes spontaneous hydrolysis of a peptide bond at the $Gly^{149}$ in the SEA domain. The protease is then autocatalytically processed and activated following cleavage at $Arg^{614}$ within the $RQAR^{614}$-VVGG activation sequence.[4] In physiological conditions, matriptase-2 is exclusively expressed in the liver initially under zymogen form. Matriptase-2 undergoes autocatalysis at $Arg^{576}$ within the $PSSR^{576}$-IVGG sequence located in the consensus activation site of its pro-domain.[5]

A tight regulation of matriptase by its endogenous inhibitor (such as the hepatocyte activation inhibitor 1, HAI-1), among others, plays a crucial role in placental development, neural tube closure,[6] and epithelial barrier homeostasis.[7] In past years, several studies have shown that a deregulation of matriptase can lead to various pathologies. Matriptase is thought to be implicated in osteoarthritis by initiating and inducting cartilage destruction.[8] Additionally, a mutation in matriptase has been shown to be implicated in autosomal recessive ichtyosis with hypotrichosis (ARIH),[9] a rare skin disease. Several studies have also shown that matriptase is upregulated in different types of epithelial tumors[10−12] but not mesenchymal-derived tumors. It has also been shown that matriptase can cause a malignant transformation in mice when orthotopically overexpressed.[13] Finally, it has been demonstrated that matriptase has the ability to cleave and activate the hemagglutinin precursor of the H1[14,15] and H9 subtypes[16] of influenza virus, which is an essential step in viral replication. Accordingly, treatment with matriptase inhibitors has been shown to reduce viral replication.[15,16]

Matriptase-2, on the other hand, is important for iron regulation.[17,18] The protease negatively regulates the expression of hepcidin, an important regulator of iron homeostasis[19] via

**Received:** September 22, 2014
**Published:** November 11, 2014

   dx.doi.org/10.1021/jm5015633 | J. Med. Chem. 2014, 57, 10198−10204

Table 1. Specificity and Selectivity of Compounds for Matriptase and Matriptase-2

| | $K_i$ (nM) | | | rank | |
|---|---|---|---|---|---|
| SEQ | M[a] | M2[b] | S[c] | M | M2 |
| RQAR | 0.011[d] ± 0.004[d] | 3.30[d] ± 1.00 | 300 | 2301 | 4033 |
| RQPR | 0.061[d] ± 0.004 | 7.80[d] ± 0.90 | 128 | 104 | 2312 |
| RQYR | 0.18[d] ± 0.04 | 0.74[d] ± 0.11 | 4 | 678 | 2754 |
| RQFR | 0.23[d] ± 0.02 | 1.40[d] ± 0.10 | 6 | 721 | 1168 |
| RYAR | 0.14[d] ± 0.02 | 6.30[d] ± 1.00 | 45 | 725 | 2298 |
| RRAR | 0.72[d] ± 0.20 | 15.40[e] ± 2.80 | 22 | 102 | 179 |
| LQAR | 0.15[d] ± 0.01 | 13.47[e] ± 2.66 | 90 | 659 | 908 |
| SQAR | 0.092[d] ± 0.023 | 34.80[e] ± 13.70 | 378 | 4028 | 3879 |
| WRER | 14.67[d] ± 1.87 | 46.78[e] ± 2.54 | 3 | 7 | 4184 |
| RNPR | 0.91[d] ± 0.11 | 47.77[e] ± 5.90 | 49 | 33 | 3690 |
| KNAR | 0.45[d] ± 0.11 | 24.84[e] ± 4.13 | 62 | 40 | 4084 |
| WCYR | 7.98[d] ± 1.27 | 1.03[d] ± 0.29 | 0.13 | 83 | 6561 |
| YYVR | 31.54[e] ± 7.14 | 2.56[d] ± 0.20 | 0.08 | 4227 | 23 |
| LWWR | 3.22[d] ± 0.37 | 1.98[d] ± 0.18 | 0.63 | 3244 | 31 |
| RLSR | 1.40[d] ± 0.21 | 2.48[d] ± 0.38 | 1.79 | 4741 | 52 |
| YKAR | 0.89[d] ± 0.29 | 20.79[d] ± 2.77 | 23 | 6226 | 71 |

[a]Matriptase. [b]Matriptase-2. [c]Selectivity defined as $K_{i,M2}/K_{i,M}$. [d]Tight-binding inhibitor [e]Classical inhibitor

cleavage of hemojuvelin from plasma membranes.[20] A number of studies have shown that mutations in the *TMPRSS6* gene lead to iron refractory iron deficiency anemia (IRIDA).[21] At the other side of the spectrum, iron overload is a condition defined by an increase in total body of iron of more than 5 g caused by a number of factors and leading to different illnesses.[22] Thus, modulation of matriptase-2 activity could be exploited as a therapy in these conditions.[23,24]

Considering these pathologies, both matriptase and matriptase-2 appear to be relevant therapeutic targets for a number of diseases. As a result, in recent years, several groups have reported different classes of inhibitors via diverse strategies[23−31] including virtual screening.[32] Among these, our group exploited the autocatalytic cleavage sequence of matriptase, flanked by a ketobenzothiazole serine trap, to produce high potency matriptase inhibitors.[15,28]

The optimization of a high selectivity profile against other proteases of the same family is often a major challenge when developing protease inhibitors. To achieve this goal, the molecule must exploit differences between the target binding site and those of other similar proteases.

We report herein the use of docking simulations to tackle the problem of selectivity between matriptase and matriptase-2 inhibitors. Toward this end, we performed a large scale docking experiment with 8000 different peptidomimetics against both proteases. The compounds are of the same type as our previously developed inhibitor[28] containing a tetrapeptide with arginine in position $P_1$ linked to a ketobenzothiazole serine trap group. The 20 natural amino acids were used to create all 8000 possible combinations at positions $P_4$, $P_3$, and $P_2$. Our docking results led to the identification of new potent matriptase-2 inhibitors with improved selectivity against matriptase, as well as novel matriptase inhibitors selective against matriptase-2. The docking results provided a tool to analyze selectivity between matriptase and matriptase-2, shedding light into the factors affecting molecular recognition in these two very similar proteases.

## RESULTS

Because there is no reported crystallographic structure of matriptase-2, a homology model was built using MODELER[33] and GROMACS[34] as described in the Methods section. The model is highly similar to the structure of matriptase (0.77 Å RMSD), which was used as a template (PDB ID 3NCL) given the high level (44%) of sequence identity of the two enzymes. In the substrate binding subpockets, only five residues differ between the two enzymes: Ser[800] to Ala[757] (subpocket S1), Phe[708] to His[665] (S2−S4), Tyr[755] to Glu[712] (S3), Phe[706] to Asp[663] (S4), and Asp[828] to Leu[785] (S4), respectively, in matriptase and matriptase-2 (Supporting Information (SI), Figure 1). The coordinates of the homology model are provided as SI.

It has been previously shown that protein flexibility plays a key role in docking simulations, especially in cross-docking studies and homology models.[35] Our docking method FlexAID[36] uses a probabilistic approach to account for rotamer changes upon ligand binding on-the-fly during each simulation using a rotamer library.[35] Accordingly, we included flexibility in various residues of the binding site. Residues Asp[705], Phe[708], Tyr[755], Gln[782], Gln[783], Gln[802], and Asp[828] of matriptase were found to be in at least one different rotamer conformation or able to undergo rotamer changes without producing steric clashes. The above residues and the corresponding residues in matriptase-2 (Asp[663], His[665], Glu[712], Tyr[739], Gln[740], Gln[759], and Leu[828]) were also set as flexible.

**Virtual Screening.** We have previously reported a peptidomimetic compound[28] with a $K_i$ of 0.011 ± 0.004 nM for matriptase[28] and 300-fold selectivity against matriptase-2 (sequence RQAR in Table 1). We performed docking simulations with a library of 8000 different compounds representing all the possible permutations of natural amino acids in positions P4, P3, and P2 (inset Figure 1), with the goal of identifying potent and selective inhibitors for matriptase-2 and to better rationalize the structural tenants of potency and selectivity between the two proteases. Additionally, it has been shown that arginine in $P_1$ is the favored amino acid in matriptase and matriptase-2,[28,37,38] for this reason the $P_1$ arginine moiety was kept constant. Given the probabilistic nature of the genetic algorithm based search procedure used in FlexAID, we made 20 independent docking simulations covering $10^6$ energy evaluations. This choice of parameters provided 93% accuracy in the prediction of binding poses in the


dx.doi.org/10.1021/jm5015633 | J. Med. Chem. 2014, 57, 10198−10204

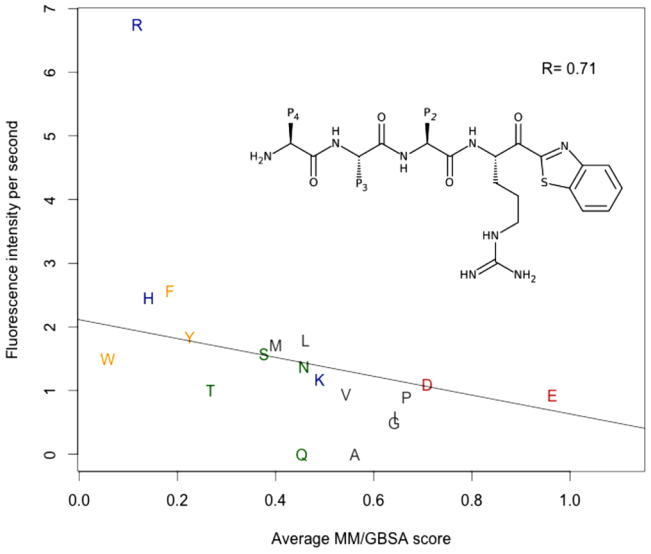



**Figure 1.** Comparison of experimental and computed average preferences at position $P_3$ in $IP_3XR$. Each point represents the average of all other 19 possibilities in position $P_2$. Colors represent amino acid types: aromatic (orange), positively (blue) or negatively (red) charged, polar (green), and hydrophobic (gray). The inset shows the general structure of the compounds used in this study.

presence of side chain flexibility in the target.[39] Each binding pose was rescored with MM/GBSA,[40] and the best one out of the 20 simulations was kept for further analysis. MM/GBSA rescoring was shown to yield better results than most scoring functions in finding the correct binding conformation and binding free energy.[41] The MM/GBSA scores of all 8000 compounds docked against matriptase and matriptase-2 can be found in SI.

**Agreement between Docking and Experimental Results.** Recently, Wysocka et al.[38] reported the substrate preferences of matriptase-2 using a fluorescence functional assay deconvoluting mixtures of fluorescent tetrapeptides. In particular, the authors measured the average activity of matriptase-2 for arrays of molecules of the form $IP_3XR$ for 19 different natural amino acids (all but Cys) in position $P_3$ averaged over the same 19 possibilities in position $P_2$ denoted by X. We extracted the fluorescence intensity values per second as reported by the authors (Figure 1B in Wysocka et al.[38]), then calculated the average MM/GBSA *Z*-scores over the same 19 amino acids for each possibility at position $P_3$ (Figure 1).

We obtained a statistically significant linear regression with $R = 0.71$ ($p$-value $2.16 \times 10^{-3}$). Somewhat surprisingly, while Gln and Ala behave in our docking simulations like other polar or hydrophobic residues, respectively, these two residues give no signal for Wysocka et al.,[38] irrespective of the amino acids present in position $P_2$. Lastly, Arg is the preferred residue on average at position $P_3$ in $IP_3XR$ but second best after Trp in our docking simulations. Considering the outlier nature of Arg and the uncertainty of the experimental results for Gln and Ala in $P_3$ as several compounds with these amino acids in position $P_3$ are active,[28,37] these three amino acids were not taken in consideration when calculating the above linear regression. The above correlation represents the agreement between 361 in silico and experimental results. The high correlation gives us confidence in the quality of the docking simulations.

**Average Subpocket Substrate Preferences.** To determine the binding preferences for matriptase and matriptase-2 in each subpocket, we looked at the average MM/GBSA *Z*-score for the 400 compounds with a conserved residue at each position in $P_4P_3P_2R$. For example, for a given residue at position $P_4$, we averaged over all 400 combinations at positions $P_3$ and $P_2$, or with a given residue at positions $P_3$ or $P_2$ instead and averaging over the remaining two positions in each case. For both matriptase and matriptase-2, we observed differences in the preferences across positions within each protein, as well as between proteins for given positions (SI, Figures 2 and 3, respectively). The observed amino acid preferences within $P_4$ (SI, Figure 3) are in agreement with the in vitro results of Wysocka et al.[38] Our results also agree with those of Béliveau et al.[37] for the residues in $P_4$ (Arg, Ala, Glu, Leu, and Tyr) within $P_4$QAR. However, our prediction that Arg is the preferred choice at $P_4$ agrees with that of Béliveau et al.[37] while Wysocka et al.[38] find Ile (second best choice in our study) as the best amino acid at that position. For position $P_3$, Wysocka et al.[38] only tested compounds with Ile in $P_4$, for which we have excellent agreement as shown in Figure 1. Our results also generally agree with those of Béliveau et al.[37] who tested Glu, Ala, Leu, Tyr, and Arg as part of $RP_3AR$. For position $P_2$, Wisocka et al.[38] showed a large preference for Ala and to a smaller extent for Met. In our case, a preference for Lys and Arg is similar to the results obtained by Béliveau et al.[37]

**Differences in Substrate Preferences between Matriptase and Matriptase-2.** The above average preferences can mask interesting differences that become apparent when looking at the preferred sequences of each protein. We compared the top 10% scoring compounds for matriptase and matriptase-2 and observed that there are distinct preferences among these (Figure 2).

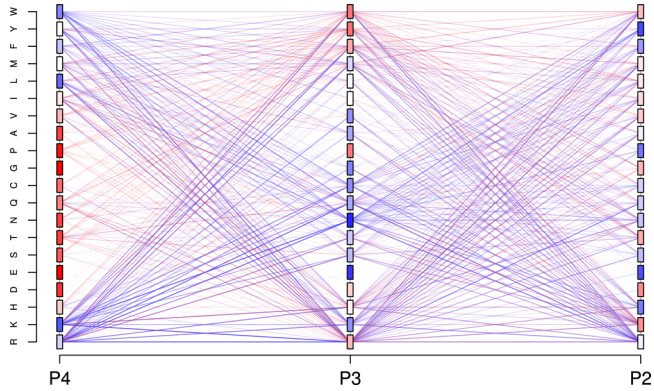



**Figure 2.** Analysis of subpocket preferences for top scoring compounds. The top scoring 800 compounds for matriptase (blue) and matriptase-2 (red) are shown in a parallel plot. Each line connecting the three subpocket-binding residues represents one compound. At each position, the color gradient corresponds to the fraction of top compounds that use the given amino acid at the particular position, white means equally distributed.

At each position a number of amino acids is more readily accepted in matriptase or matriptase-2. For example, in position $P_4$, Lys, Leu, Phe, and Trp are more prominent in top scoring compounds in matriptase. Differences are also clear for $P_3$ but less clear for $P_2$. In a recent study, Maurer et al.[42] mutated matriptase-2, reintroducing in its binding site four of the five amino acids observed in matriptase (all but D663F in S4) and analyzed their effect on binding a number of small-molecule inhibitors showing the importance of these mutations in the


dx.doi.org/10.1021/jm5015633 | J. Med. Chem. 2014, 57, 10198−10204

selectivity of compounds. The preference for hydrophobic and aromatic residues in P4 observed for matriptase in Figure 2 is in agreement with the results of Maurer et al.[42] and shows how the added hydrophobicity brought about by the particular amino acids in matriptase (Ser$^{800}$, Phe$^{708}$, Tyr$^{755}$, Phe$^{706}$, and Asp$^{828}$) compared to the respective residues in matriptase-2 (Ala$^{757}$, His$^{665}$, Glu$^{712}$, Asp$^{663}$, and Leu$^{785}$) may account for the observed differences in preferences.

**Shannon Entropy of Subpockets and Druggability.** The same analysis with a parallel plot can be performed to compare the top and bottom scoring compounds for matriptase (SI, Figure 4) and matriptase-2 (SI, Figure 5). We observed a larger variability in average MM/GBSA *Z*-scores for amino acids in $P_4$ than in $P_2$. These results point to differences across subpockets in terms of the effect of different amino acids with respect to discriminating top from bottom scoring compounds. This suggests that the $S_4$ subpocket has a higher potential to be exploited in drug design than $S_2$ as different choices in the compounds at position $P_4$ lead to differences in activity while in $P_2$ the choice of amino acid has less effect.

Shannon entropy[43] provides a way to measure a specificity profile for each subpocket. Ranging from 0 to 1, the value determines the tolerance for different amino acids at a given position. A subpocket with Shannon entropy 0 can only bind one specific amino acid (high selectivity for this amino acids), while a Shannon entropy of 1 means no preferences toward the 20 amino acids. To evaluate the binding selectivity of each subpocket, we clustered the 8000 compounds using MM/GBSA *Z*-score similarities. For each cluster, we calculated the Shannon entropy, as described by Fuchs et al.,[44] for the molecules in each cluster at each position. In Figure 3 we show such plot for matriptase-2.

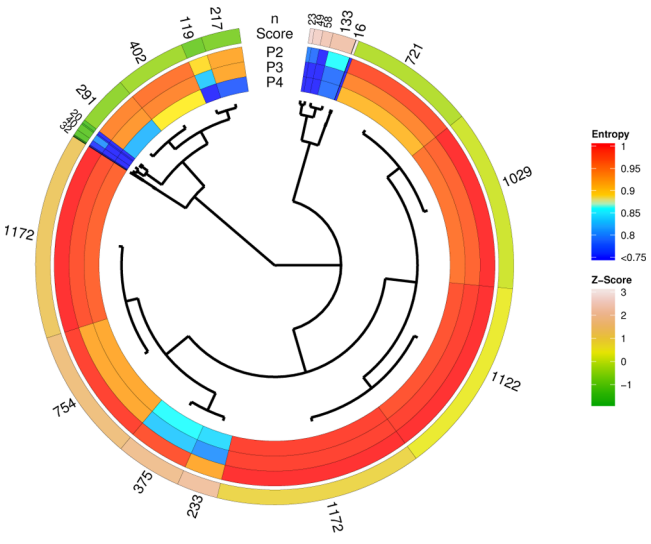



**Figure 3.** Shannon entropy of different compound clusters based on *Z*-score differences across different subpockets. The number of compounds in each cluster is shown next to the *Z*-score outer ring.

The three inner rings represent Shannon entropy values for each cluster at each respective subpocket as marked.

Molecule clusters split early into one cluster (11 o'clock arm) with good scoring molecules (outer ring) and the rest that splits into a cluster with poor scoring molecules (1 o'clock arm) and the remaining with average to poor scores. For all clusters, we have a decrease of Shannon entropy from $P_2$ to $P_4$, with the most pronounced decreases in Shannon entropy observed for good scoring clusters in the 11 o'clock arm of top scoring molecules. Similar results are obtained for matriptase (SI, Figure 6). These results show that for compounds with similar scores, the two proteases tend to be less restrictive and can accommodate more amino acids at position $P_2$, while there are fewer possible amino acids accepted at positions $P_3$ and $P_4$. In summary, we observe differences between matriptase and matriptase-2 in terms of their amino acid tolerances at each subpocket. Position $P_4$ followed by $P_3$ and $P_2$, respectively, have lower Shannon entropy, thus choices of amino acids at position $P_4$ and $P_3$ have a higher impact than at position $P_2$. Fuchs et al.[44] introduced the concept of information entropy above and discussed it in the context of experimental substrate promiscuity. Our results agree with those of Fuchs et al. but provide a more complete picture given the additional data available computationally. Fuchs observes that different protease families have different patterns of Shannon entropy variations (promiscuity) across subpockets. In the present case, by using MM/GBSA scores we observe that both matriptase and matriptase-2, two members of the TTSP SA1 matriptase subfamily, behave similarly, with Shannon entropy decreasing as distance from the catalytic site increases. In other words, within this subfamily, our results suggest that the larger the distance from the catalytic site, the higher the ability of choosing selectivity-conferring amino acids in the nonprimed direction. The differences between matriptase and matriptase-2 come about despite the fact that the existing amino acid differences between the two binding sites are distributed throughout all subpockets.

**Selective Inhibitors for Matriptase.** To develop more insight on the contribution of different positions in binding and further validate our virtual screening approach, we experimentally tested the activity of a number of compounds with good MM/GBSA scores for matriptase containing a single alteration compared to RQAR. Seven compounds were selected and synthesized, with single alterations in $P_2$ (RQPR, RQYR, and RQFR), $P_3$ (RYAR and RRAR), and $P_4$ (LQAR and SQAR) relative to RQAR. All compounds were tested for in vitro inhibition of matriptase and matriptase-2.

In all cases, the compounds have higher specificity for matriptase (Table 1). Among these compounds we find two new compounds, RQPR and SQAR that have very low $K_i$ for matriptase and 128- or 378-fold selectivity defined as the ratio of the $K_i$ for matriptase-2 by that of matriptase.

Another four compounds (WRER, RNPR, KNAR, and WCYR) were selected entirely based on having better MM/GBSA ranks for matriptase (top 10%) compared to matriptase-2 (bottom 50%). Two of these compounds are selective for matriptase against matriptase-2 with a selectivity of 49 for RNPR and 60 for KNAR. While their $K_i$ are slightly worse than RQAR, they are comparable in both specificity and selectivity to many compounds related to RQAR by single alterations and provide novel venues for chemical modifications in the pursuit of lead compounds.

**Selective Inhibitors for Matriptase-2.** Four compounds (YYVR, LWWR, RLSR, and YKAR) with predicted higher specificity for matriptase-2 (top 10% rank) than for matriptase (bottom 50% rank) were experimentally tested. Interestingly, we found one compound (YYVR) that is 13 times more selective for matriptase-2 than matriptase and a second compound (LWWR) with 2-fold selectivity. One compound selected for matriptase is 8 times more selective for matriptase-


dx.doi.org/10.1021/jm5015633 | J. Med. Chem. 2014, 57, 10198−10204

2 (WCYR). YYVR and WCYR along with recent data on recognition of substrate residues in the P and P′ positions by these proteases[45] will serve as starting points to develop selective compounds of the TTSP family.

Docking methods are generally applied to three distinct but inter-related tasks. Namely, the prediction of binding modes (the structure of the ligand-protein complex), the discrimination of binders from nonbinders (virtual screening), and last, the prediction of binding affinities. Docking methods are notoriously bad at tackling satisfactorily these three problems simultaneously.[46] It is therefore interesting to note that there is a qualitative agreement (as shown in Table 1) between virtual screening ranks and binding affinities as well as quantitative agreement between averaged MM/GBSA scores and activities (as shown in Figure 1). In conclusion, the virtual screening led to the identification of potent and selective inhibitors of matriptase and matriptase-2 despite a high level of homology between the two binding pockets and provided insights into the subpocket preferences of each enzyme.

## METHODS

**Preparation of Ligand Structures.** The 3D coordinates of the ligands with correct chirality were generated with Open Babel 2.3.0[47] using the SMARTS input format. The coordinates were exported in PDB format. The orientation of the ketobenzothiazole and the arginine in $P_1$ were fixed in a conformation mimicking that found in the thrombin inhibitor (0ZE) in PDB structure 1B5G for every compound.

**Preparation of Protein Targets.** The structure of the protease domain of matriptase (PDB code 3NCL) was obtained from the RCSB Protein Data Bank.[48] Hydrogen atoms and water molecules were removed from the structure. A homology model of matriptase-2 was built with MODELER.[33] Twenty different models were generated using basic parameters. The conformation with the best DOPE[33,49] score was used.

**Docking Parameters.** Molecules were docked in each protein with the small-molecule protein docking program FlexAID.[36,39] FlexAID uses a genetic algorithm to optimize the relative position of the ligand with respect to the protein as well as dihedral angles of the ligand and flexible residues (optimized using an exhaustive search with a rotamer library). The population size and number of generations used in the genetic algorithm were both set at 1000 for a total of $10^6$ energy evaluations. The ligands were set as semiflexible. Arg in $P_1$ and the ketobenzothiazole group were set as rigid as well as all the peptide bonds. All side chain dihedral bonds were set as flexible. The carbon making the reversible covalent bond with the protein was fixed at a distance of 3.0 ± 0.1 Å of the catalytic $Ser^{805}$ to emulate the effect that the covalent bond would have in constraining the position of the ketobenzothiazole group within the binding site. Whereas this procedure only approximates covalent binding as it ignores the influence of developing a covalent bond on the structure of the binding site, it has been shown to be successful in a number of protein families and types of covalent ligands.[50] Target flexibility was restricted to binding-site side chains. Flexible residues were located based on the superimposition and detection of residues seen in different conformations in the binding sites of other matriptase structures: 1EAW with bovine pancreatic trypsin inhibitor (BPTI), 1EAX with the small molecule inhibitor benzamidine, 2GV6 with the small molecule inhibitor CJ-730, 2GV7 with the small molecule inhibitor CJ-672, and 3BN9 with a noncompetitive FAB antibody inhibitor. For each of the 8000 docked compounds, 20 different simulations were performed using our volunteer-based grid computing platform NRG@Home using the BOINC platform.[51] Each docking result was rescored with the MM/GBSA method as described below. FlexAID and the NRGsuite PyMOL plugin to perform and view on-the-fly docking simulations for the Window, Linux, and MacOS platforms can be obtained free of charge at http://bcb.med.usherbrooke.ca/flexaid.

**MM/GBSA Rescoring of Docking Conformations.** All protein−ligand complexes were first energy-minimized with GROMACS[52] 4.5.1 using the conjugate gradient algorithm in a TIP3P water box. For the proteins, AMBER03 force field parameters were used. For the ligands, general AMBER force field generated in Antechamber and converted to GROMACS formats by ACPYPE[53] were used. Partial charges were calculated with the Partial Equalization of Orbital Electronegativities (PEOE) algorithm by Gaisteiger and Marsilli.[54] All poses were then rescored with MM/GBSA using the HCT model effective Born radii.

**Synthesis of Inhibitors.** Inhibitors were synthesized using a combination of solid phase and solution synthesis methodologies. Intermediate **2** (SI, Scheme 1) was prepared as reported earlier.[28] Analogues **3a−s** (SI, Scheme 2) were synthesized by standard solid phase peptide synthesis (SPPS) using the 4-chlorotrityl chloride resin.[28] Briefly, C-terminal anchoring of the first amino acid on the resin was performed in dichloromethane (DCM) in the presence of diisopropylethylamine (DIPEA). Tripeptide synthesis were pursued with standard Fmoc protected amino acid coupling, deprotected in basic conditions (20% piperidine/DMF) and in order to selectively couple the modified arginine **2** (SI, Scheme 1), last coupling was done using N-Boc-protected amino acid. Analogues **3a−s** were cleaved from resin under mild acidic conditions (DCM/TFE or HFIP, see note in SI, Scheme 2). Tripeptides **3a−s** and intermediate **2** were engaged in peptidic coupling reaction, followed by oxidization and acidolysis to give tetrapeptide compounds **5a−s** based on the synthetic strategy described by Colombo et al.[28] Generally, compounds were obtained as a 3:1 mixture of epimers at the P1 position,[55] which were purified and separated by reverse-phase preparative high-performance liquid chromatography (prep-HPLC) above 95% purity (UV). Compound purity and identity were confirmed by UPLC-MS.

**Inhibition Assays of Matriptase and Matriptase-2.** Active recombinant human matriptase (amino acids 596−655) and matriptase-2 (amino acids 78−811) were expressed and purified as previously described.[27,37] Proteases were active-site titrated with 4-methylumbelliferyl-*p*-guanidino benzoate (MUGB) prior to inhibition assays. Enzymatic assays and $K_i$ determination were performed at room temperature in an assay buffer containing 50 mM Tris-HCl, 15 mM NaCl, and 500 at pH 7.4 as described previously.[28] Then 1 nM of protease was added to a reaction buffer containing 0 nM, 10 nM, or 10 mM of inhibitors and 100 mM of a fluorogenic substrate (Boc-Gln-Ala-Arg-AMC), and proteolytic activity was monitored by measuring the release of fluorescence (excitation, 360 nm; emission, 441 nm) in a FLX800 TBE microplate reader (Bio-Tek Instruments, Winooski, VT, USA). If substantial inhibition occurred using a ratio $I/E = 10$, compound were treated as tight-binding inhibitors, and if inhibition occurs only at $I/E > 10$, compounds were treated as classical reversible inhibitors. For $K_i$ determination of tight-binding inhibitors, plots of enzyme velocity as a function of inhibitor concentration were fitted by nonlinear regression analysis to the Morrison $K_i$ equation.[56] For $K_i$ determination of classical reversible inhibitors, plots of enzyme velocity as a function of substrate concentration at several inhibitor concentrations were fitted by nonlinear regression to equations describing different model of reversible inhibition (competitive, uncompetitive, noncompetitive, and mixed model), and the preferred model was used for $K_i$ determination. All assays were performed at least three times in duplicates, and data are presented as mean ± standard deviation (SD). Nonlinear regression and statistical analysis were performed using GraphPad Prism version 6.02 for Windows (GraphPad Software, San Diego, CA, USA).

## ASSOCIATED CONTENT

### Supporting Information

Solution and solid-phase synthesis, materials and equipment; synthetic procedures; chemistry; progress curves; docking results for the 8000 compounds (CSV) and the coordinates of the model structure of matriptase-2 (PDB). This material is available free of charge via the Internet at http://pubs.acs.org.

## AUTHOR INFORMATION


### Corresponding Author

*E-mail: rafael.najmanovich@usherbrooke.ca. Phone: +1 (819) 820-6868 Ext. 12374.


### Author Contributions

D.D., E.C., A.D., R.L., E.M., and R.J.N. designed the experiments, D.D., E.C., A.D., and P.L.B. performed the experiments. D.D., E.C., A.D., and R.J.N. wrote the manuscript. All authors reviewed and gave approval to the final version of the manuscript.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We acknowledge the input of Francis Gaudreault in the use of FlexAID for the docking simulations and all NRG@Home volunteers for their CPU time donated to this project. R.J.N., E.M., and R.L. are part of CR-CHUS and the Institute of Pharmacology of Sherbrooke. R.J.N. and E.M. are members of PROTEO (the Québec network for research on protein function, structure and engineering). R.J.N. is a member of GRASP (Groupe de Recherche Axé sur la Structure des Protéines). This work was supported by the Canadian Institutes for Health Research (CIHR). E.C. was partly founded by a Proteo student fellowship.

## ABBREVIATIONS USED

MMGB/SA, Molecular Mechanics/Generalized Born Surface Area

## REFERENCES


(1) Rawlings, N. D.; Waller, M.; Barrett, A. J.; Bateman, A. MEROPS: the Database of Proteolytic Enzymes, Their Substrates and Inhibitors. *Nucleic Acids Res.* **2014**, *42*, D503−D509.

(2) Page, M. J.; Di Cera, E. Serine Peptidases: Classification, Structure and Function. *Cell. Mol. Life Sci.* **2008**, *65*, 1220−1236.

(3) Bugge, T. H.; Antalis, T. M.; Wu, Q. Type II Transmembrane Serine Proteases. *J. Biol. Chem.* **2009**, *284*, 23177−23181.

(4) Oberst, M. D.; Williams, C. A.; Dickson, R. B.; Johnson, M. D.; Lin, C.-Y. The Activation of Matriptase Requires Its Noncatalytic Domains, Serine Protease Domain, and Its Cognate Inhibitor. *J. Biol. Chem.* **2003**, *278*, 26773−26779.

(5) Jiang, J.; Yang, J.; Feng, P.; Zuo, B.; Dong, N.; Wu, Q.; He, Y. *N*-Glycosylation Is Required for Matriptase-2 Autoactivation and Ectodomain Shedding. *J. Biol. Chemistry* **2014**, *289*, 19500−19507.

(6) Szabo, R.; Hobson, J. P.; Christoph, K.; Kosa, P.; List, K.; Bugge, T. H. Regulation of Cell Surface Protease Matriptase by HAI2 Is Essential for Placental Development, Neural Tube Closure and Embryonic Survival in Mice. *Development* **2009**, *136*, 2653−2663.

(7) List, K.; Haudenschild, C. C.; Szabo, R.; Chen, W.; Wahl, S. M.; Swaim, W.; Engelholm, L. H.; Behrendt, N.; Bugge, T. H. Matriptase/MT-SP1 Is Required for Postnatal Survival, Epidermal Barrier Function, Hair Follicle Development, and Thymic Homeostasis. *Oncogene* **2002**, *21*, 3765−3779.

(8) Milner, J. M.; Patel, A.; Davidson, R. K.; Swingler, T. E.; Désilets, A.; Young, D. A.; Kelso, E. B.; Donell, S. T.; Cawston, T. E.; Clark, I. M.; Ferrell, W. R.; Plevin, R.; Lockhart, J. C.; Leduc, R.; Rowan, A. D. Matriptase Is a Novel Initiator of Cartilage Matrix Degradation in Osteoarthritis. *Arthritis Rheum.* **2010**, *62*, 1955−1966.

(9) Basel-Vanagaite, L.; Attia, R.; Ishida-Yamamoto, A.; Rainshtein, L.; Ben Amitai, D.; Lurie, R.; Pasmanik-Chor, M.; Indelman, M.; Zvulunov, A.; Saban, S.; Magal, N.; Sprecher, E.; Shohat, M. Autosomal Recessive Ichthyosis with Hypotrichosis Caused by a Mutation in ST14, Encoding Type II Transmembrane Serine Protease Matriptase. *Am. J. Hum. Genet.* **2007**, *80*, 467−477.

(10) Bergum, C.; List, K. Loss of the Matriptase Inhibitor HAI-2 During Prostate Cancer Progression. *Prostate* **2010**, *70*, 1422−1428.

(11) Jin, X.; Yagi, M.; Akiyama, N.; Hirosaki, T.; Higashi, S.; Lin, C.-Y.; Dickson, R. B.; Kitamura, H.; Miyazaki, K. Matriptase Activates Stromelysin (MMP-3) and Promotes Tumor Growth and Angiogenesis. *Cancer Sci.* **2006**, *97*, 1327−1334.

(12) Köbel, M.; Kalloger, S. E.; Boyd, N.; McKinney, S.; Mehl, E.; Palmer, C.; Leung, S.; Bowen, N. J.; Ionescu, D. N.; Rajput, A.; Prentice, L. M.; Miller, D.; Santos, J.; Swenerton, K.; Gilks, C. B.; Huntsman, D. Ovarian Carcinoma Subtypes Are Different Diseases: Implications for Biomarker Studies. *PLoS Med.* **2008**, *5*, e232.

(13) List, K.; Kosa, P.; Szabo, R.; Bey, A. L.; Wang, C. B.; Molinolo, A.; Bugge, T. H. Epithelial Integrity Is Maintained by a Matriptase-Dependent Proteolytic Pathway. *Am. J. Pathol.* **2009**, *175*, 1453−1463.

(14) Hamilton, B. S.; Gludish, D. W. J.; Whittaker, G. R. Cleavage Activation of the Human-Adapted Influenza Virus Subtypes by Matriptase Reveals Both Subtype and Strain Specificities. *J. Virol.* **2012**, *86*, 10579−10586.

(15) Beaulieu, A.; Gravel, É.; Cloutier, A.; Marois, I.; Colombo, E.; Désilets, A.; Verreault, C.; Leduc, R.; Marsault, E.; Richter, M. V. Matriptase Proteolytically Activates Influenza Virus and Promotes Multicycle Replication in the Human Airway Epithelium. *J. Virol.* **2013**, *87*, 4237−4251.

(16) Baron, J.; Tarnow, C.; Mayoli-Nüssle, D.; Schilling, E.; Meyer, D.; Hammami, M.; Schwalm, F.; Steinmetzer, T.; Guan, Y.; Garten, W.; Klenk, H.-D.; Böttcher-Friebertshäuser, E. Matriptase, HAT, and TMPRSS2 Activate the Hemagglutinin of H9N2 Influenza a Viruses. *J. Virol.* **2013**, *87*, 1811−1820.

(17) Ramsay, A. J.; Hooper, J. D.; Folgueras, A. R.; Velasco, G.; López-Otín, C. Matriptase-2 (TMPRSS6): A Proteolytic Regulator of Iron Homeostasis. *Haematologica* **2009**, *94*, 840−849.

(18) Du, X.; She, E.; Gelbart, T.; Truksa, J.; Lee, P.; Xia, Y.; Khovananth, K.; Mudd, S.; Mann, N.; Moresco, E. M. Y.; Beutler, E.; Beutler, B. The Serine Protease TMPRSS6 Is Required to Sense Iron Deficiency. *Science* **2008**, *320*, 1088−1092.

(19) Silvestri, L.; Pagani, A.; Nai, A.; De Domenico, I.; Kaplan, J.; Camaschella, C. The Serine Protease Matriptase-2 (TMPRSS6) Inhibits Hepcidin Activation by Cleaving Membrane Hemojuvelin. *Cell Metab.* **2008**, *8*, 502−511.

(20) Silvestri, L.; Guillem, F.; Pagani, A.; Nai, A.; Oudin, C.; Silva, M.; Toutain, F.; Kannengiesser, C.; Beaumont, C.; Camaschella, C.; Grandchamp, B. Molecular Mechanisms of the Defective Hepcidin Inhibition in TMPRSS6 Mutations Associated with Iron-Refractory Iron Deficiency Anemia. *Blood* **2009**, *113*, 5605−5608.

(21) Finberg, K. E.; Heeney, M. M.; Campagna, D. R.; Aydinok, Y.; Pearson, H. A.; Hartman, K. R.; Mayo, M. M.; Samuel, S. M.; Strouse, J. J.; Markianos, K.; Andrews, N. C.; Fleming, M. D. Mutations in TMPRSS6 Cause Iron-Refractory Iron Deficiency Anemia (IRIDA). *Nature Genet.* **2008**, *40*, 569−571.

(22) Kew, M. C. Hepatic Iron Overload and Hepatocellular Carcinoma. *Liver Cancer* **2014**, *3*, 31−40.

(23) Riba, M.; Rausa, M.; Sorosina, M.; Cittaro, D.; Garcia Manteiga, J. M.; Nai, A.; Pagani, A.; Martinelli-Boneschi, F.; Stupka, E.; Camaschella, C.; Silvestri, L. A Strong Anti-Inflammatory Signature Revealed by Liver Transcription Profiling of Tmprss6−/− Mice. *PLoS One* **2013**, *8*, e69694.

(24) Maurer, E.; Gütschow, M.; Stirnberg, M. Hepatocyte Growth Factor Activator Inhibitor Type 2 (HAI-2) Modulates Hepcidin Expression by Inhibiting the Cell Surface Protease Matriptase-2. *Biochem. J.* **2013**, *450*, 583−593.

(25) Galkin, A. V.; Mullen, L.; Fox, W. D.; Brown, J.; Duncan, D.; Moreno, O.; Madison, E. L.; Agus, D. B. CVS-3983, a Selective Matriptase Inhibitor, Suppresses the Growth of Androgen Independent Prostate Tumor Xenografts. *Prostate* **2004**, *61*, 228−235.

(26) Stoop, A. A.; Craik, C. S. Engineering of a Macromolecular Scaffold to Develop Specific Protease Inhibitors. *Nature Biotechnol.* **2003**, *21*, 1063−1068.

(27) Désilets, A.; Longpré, J.-M.; Beaulieu, M.-E.; Leduc, R. Inhibition of Human Matriptase by Eglin C Variants. *FEBS Lett.* **2006**, *580*, 2227−2232.
(28) Colombo, E.; Désilets, A.; Duchêne, D.; Chagnon, F.; Najmanovich, R. J.; Leduc, R.; Marsault, E. Design and Synthesis of Potent, Selective Inhibitors of Matriptase. *ACS Med. Chem. Lett.* **2012**, *3*, 530−534.
(29) Quimbar, P.; Malik, U.; Sommerhoff, C. P.; Kaas, Q.; Chan, L. Y.; Huang, Y.-H.; Grundhuber, M.; Dunse, K.; Craik, D. J.; Anderson, M. A.; Daly, N. L. High-Affinity Cyclic Peptide Matriptase Inhibitors. *J. Biol. Chem.* **2013**, *288*, 13885−13896.
(30) Fittler, H.; Avrutina, O.; Glotzbach, B.; Empting, M.; Kolmar, H. Combinatorial Tuning of Peptidic Drug Candidates: High-Affinity Matriptase Inhibitors Through Incremental Structure-Guided Optimization. *Org. Biomol. Chem.* **2013**, *11*, 1848−1857.
(31) Goswami, R.; Mukherjee, S.; Ghadiyaram, C.; Wohlfahrt, G.; Sistla, R. K.; Nagaraj, J.; Satyam, L. K.; Subbarao, K.; Palakurthy, R. K.; Gopinath, S.; Krishnamurthy, N. R.; Ikonen, T.; Moilanen, A.; Subramanya, H. S.; Kallio, P.; Ramachandra, M. Structure-Guided Discovery of 1,3,5 Tri-Substituted Benzenes as Potent and Selective Matriptase Inhibitors Exhibiting in Vivo Antitumor Efficacy. *Bioorg. Med. Chem.* **2014**, *22*, 3187−3203.
(32) Sisay, M. T.; Steinmetzer, T.; Stirnberg, M.; Maurer, E.; Hammami, M.; Bajorath, J.; Gütschow, M. Identification of the First Low-Molecular-Weight Inhibitors of Matriptase-2. *J. Med. Chem.* **2010**, *53*, 5523−5535.
(33) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M.-Y.; Pieper, U.; Sali, A. Comparative Protein Structure Modeling Using Modeller. *Current Protocols in Bioinformatics*; Baxevanis, A. D., Ed.: Wiley: New York, 2006, Chapter 5, Unit 5.6−5.6.30.
(34) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845−854.
(35) Gaudreault, F.; Chartier, M.; Najmanovich, R. J. Side-Chain Rotamer Changes Upon Ligand Binding: Common, Crucial, Correlate with Entropy and Rearrange Hydrogen Bonding. *Bioinformatics* **2012**, *28*, i423−i430.
(36) Najmanovich, R. J. Side Chain Flexibility Upon Ligand Binding: Docking Predictions and Statistical Analysis. Ph.D. Thesis. Feinberg Graduate School, Weizmann Institute of Sciences *arXiv:13014564*, ,2004.
(37) Béliveau, F.; Désilets, A.; Leduc, R. Probing the Substrate Specificities of Matriptase, Matriptase-2, Hepsin and DESC1 with Internally Quenched Fluorescent Peptides. *FEBS J.* **2009**, *276*, 2213−2226.
(38) Wysocka, M.; Gruba, N.; Miecznikowska, A.; Popow-Stellmaszyk, J.; Gütschow, M.; Stirnberg, M.; Furtmann, N.; Bajorath, J.; Lesner, A.; Rolka, K. Substrate Specificity of Human Matriptase-2. *Biochimie* **2014**, *97*, 121−127.
(39) Gaudreault, F.; Najmanovich, R. J. FlexAID: Revisiting Docking on Non Native-Complex Structures. Unpublished results.
(40) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2011**, *51*, 69−82.
(41) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of the Molecular Mechanics/Poisson Boltzmann Surface Area and Molecular Mechanics/Generalized Born Surface Area Methods. II. The Accuracy of Ranking Poses Generated From Docking. *J. Comput. Chem.* **2011**, *32*, 866−877.
(42) Maurer, E.; Sisay, M. T.; Stirnberg, M.; Steinmetzer, T.; Bajorath, J.; Gütschow, M. Insights Into Matriptase-2 Substrate Binding and Inhibition Mechanisms by Analyzing Active-Site-Mutated Variants. *ChemMedChem* **2012**, *7*, 68−72.
(43) Shannon, C. E. A Mathematical Theory of Communication. *Bell System Tech. J.* **1948**, *27*, 379−423.
(44) Fuchs, J. E.; von Grafenstein, S.; Huber, R. G.; Margreiter, M. A.; Spitzer, G. M.; Wallnoefer, H. G.; Liedl, K. R. Cleavage Entropy as Quantitative Measure of Protease Specificity. *PLoS Comput. Biol.* **2013**, *9*, e1003007.
(45) Barré, O.; Dufour, A.; Eckhard, U.; Kappelhoff, R.; Béliveau, F.; Leduc, R.; Overall, C. M. Cleavage Specificity Analysis of Six Type II Transmembrane Serine Proteases (TTSPs) Using PICS with Proteome-Derived Peptide Libraries. *PLoS One* **2014**, *9*, e105984.
(46) Huang, S.-Y.; Grinter, S. Z.; Zou, X. Scoring Functions and Their Evaluation Methods for Protein−Ligand Docking: Recent Advances and Future Directions. *Phys. Chem. Chem. Phys.* **2010**, *12*, 12899−12908.
(47) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: an Open Chemical Toolbox. *J. Cheminform.* **2011**, *3*, 33.
(48) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucl. Acids. Res.* **2000**, *28*, 235−242.
(49) Shen, M.-Y.; Sali, A. Statistical Potential for Assessment and Prediction of Protein Structures. *Protein Sci.* **2006**, *15*, 2507−2524.
(50) London, N.; Miller, R. M.; Krishnan, S.; Uchida, K.; Irwin, J. J.; Eidam, O.; Gibold, L.; Cimermančič, P.; Bonnet, R.; Shoichet, B. K.; Taunton, J. Covalent Docking of Large Libraries for the Discovery of Chemical Probes. *Nature Chem. Biol.* **2014**, *10*, 1066−1072.
(51) Anderson, D. P. BOINC: a System for Public-Resource Computing and Storage. . In Fifth IEEE/ACM International Workshop on Grid Computing, Pittsburgh, PA, November 8, 2004; IEEE: Los Angeles, 2004; pp 4−10.
(52) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(53) Sousa da Silva, A. W.; Vranken, W. F. ACPYPE—AnteChamber PYthon Parser interfacE. *BMC Res. Notes* **2012**, *5*, 367.
(54) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity—A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219−3228.
(55) Costanzo, M. J.; Almond, H. R.; Hecker, L. R.; Schott, M. R.; Yabut, S. C.; Zhang, H.-C.; Andrade-Gordon, P.; Corcoran, T. W.; Giardino, E. C.; Kauffman, J. A.; Lewis, J. M.; de Garavilla, L.; Haertlein, B. J.; Maryanoff, B. E. In-Depth Study of Tripeptide-Based Alpha-Ketoheterocycles as Inhibitors of Thrombin. Effective Utilization of the S1′ Subsite and Its Implications to Structure-Based Drug Design. *J. Med. Chem.* **2005**, *48*, 1984−2008.
(56) Williams, J. W.; Morrison, J. F. The Kinetics of Reversible Tight-Binding Inhibition. *Methods Enzymol.* **1979**, *63*, 437−467.