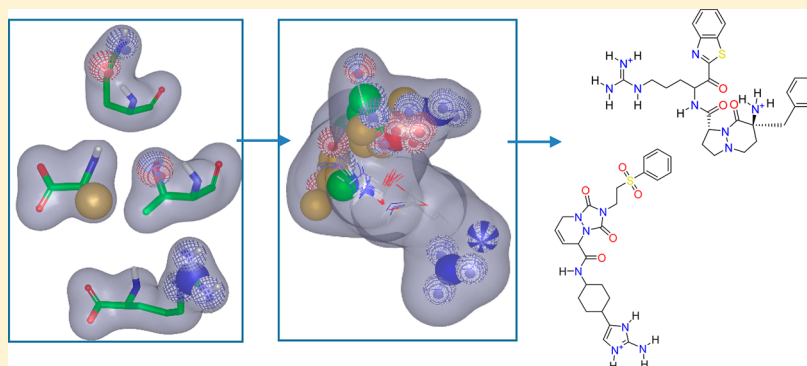


Protease Inhibitors in View of Peptide Substrate Databases

Birgit J. Waldner, Julian E. Fuchs, Michael Schauerperl, Christian Kramer,[†] and Klaus R. Liedl*

Institute of General, Inorganic and Theoretical Chemistry, University of Innsbruck, Innrain 82, 6020 Innsbruck, Austria

Supporting Information



ABSTRACT: Protease substrate profiling has nowadays almost become a routine task for experimentalists, and the knowledge on protease peptide substrates is easily accessible via the MEROPS database. We present a shape-based virtual screening workflow using vROCS that applies the information about the specificity of the proteases to find new small-molecule inhibitors. Peptide substrate sequences for three to four substrate positions of each substrate from the MEROPS database were used to build the training set. Two-dimensional substrate sequences were converted to three-dimensional conformations through mutation of a template peptide substrate. The vROCS query was built from single amino acid queries for each substrate position considering the relative frequencies of the amino acids. The peptide-substrate-based shape-based virtual screening approach gives good performance for the four proteases thrombin, factor Xa, factor VIIa, and caspase-3 with the DUD-E data set. The results show that the method works for protease targets with different specificity profiles as well as for targets with different active-site mechanisms. As no structure of the target and no information on small-molecule inhibitors are required to use our approach, the method has significant advantages in comparison with conventional structure- and ligand-based methods.

INTRODUCTION

Proteases are important targets in drug design, as they are part of numerous fundamental cellular processes.¹ There are seven distinct classes of proteases, which are classified according to the catalytic residue: serine, threonine, cysteine, aspartate, and glutamate proteases, metalloproteases, and asparagine peptide lyases.² Among each protease class, the reaction mechanism is highly conserved. In addition, proteases often have many closely related family members, and lead compounds often hit more than one target. Therefore, achieving target specificity when designing protease inhibitors still represents a difficult challenge.³

Current virtual screening strategies to find new small-molecule inhibitors can be divided into two groups: ligand-based approaches and structure-based approaches. To apply a ligand-based approach, information on one or more ligands that can bind to the target is required. From the set of known actives, structurally diverse compounds with similar bioactivity should be discovered.⁴

Structure-based methods require either an X-ray or NMR structure or a homology model of the target. Of the structure-based methods, docking and scoring is the most used method in virtual screening. However, finding the correct binding

conformation through a docking experiment remains a challenging task.⁵ Consideration of the flexibility of the protein and ligand is not easy to achieve, even with flexible docking methods.⁶ Another structure-based method is pharmacophore-based virtual screening.⁷ The “stripping” of functional groups has the advantage that scaffold hopping is possible if topological pharmacophores are used.⁸

Shape-based virtual screening with ROCS⁹ is an alternative to docking and pharmacophore-based virtual screening.¹⁰ Virtual screening results with ROCS show higher consistency than the results of docking strategies. Inclusion of the pharmacophore properties of the query molecule allows a combination of the chemical information and the information about the shape when screening for small-molecule inhibitors. Screening of the DUD database¹¹ using a combination of shape and pharmacophore properties revealed a superior performance of ROCS relative to docking approaches.¹²

With methods like proteomic identification of protease cleavage site specificity (PICS)¹³ and terminal isotopic labeling of substrates (TAILS)¹⁴ and the use of proteome-derived

Received: February 10, 2016

Published: June 1, 2016

substrate libraries,¹³ protease specificity profiles can be readily determined. In PICS, the carboxypeptide cleavage products of an oligopeptide library, consisting of natural biological sequences derived from human proteomes, are selectively isolated, and liquid chromatography–tandem mass spectrometry (LC–MS/MS) is used to identify the prime side sequences of the cleaved peptides. Nonprime side sequences are determined through automated database searches of the human proteome. PICS thus enables simultaneous determination of prime and nonprime side sequences of cleaved peptides.¹³ N-TAILS allows one to distinguish between N-termini of proteins and N-termini of protease cleavage products. Dendritic polyglycerol aldehyde polymers are used to remove tryptic and C-terminal peptides. Tandem mass spectrometry is used to analyze unbound naturally acetylated, cyclized, or labeled N-termini from proteins and their protease cleavage products.¹⁵ C-TAILS complements N-TAILS and represents an isotope-encoded quantitative C-terminomics strategy to identify neo-C-terminal sequences and protease substrates.¹⁴ With the availability of those efficient approaches for protease substrate profiling, the amount of information on protease peptide substrates is growing every day. With the cleavage entropy, a metric developed in our group, quantification of protease specificity and ranking of proteases according to specificity is possible.¹⁶ The MEROPS database represents the biggest collection of known protease peptide substrates, and it is constantly being improved and updated.² We have developed a virtual screening workflow based solely on the information on protease peptide substrate sequences present in the MEROPS database that can be used to find new small-molecule inhibitors. The types of possible interactions of the substrate peptides are the same as for small molecules. Therefore, it should be possible to find small molecules that form the same interactions with a protease as the corresponding peptide substrates. The idea of using an analysis of the protease peptide substrate space to find small-molecule inhibitors per se is not new. Recently it was shown in our group that proteases that are close in substrate space are often targeted by the same small molecules.¹⁷ Sukuru et al.¹⁸ developed a lead discovery strategy based on the similarity of proteases in the protease substrate space. They recovered the known inhibitors of proteases that are highly correlated. Their approach allows one to use a ligand-based approach to find inhibitors for proteases for which no ligands are known. However, information on small-molecule ligands for a protease that are similar in substrate space is needed in order to apply their method.

In developing a virtual screening workflow that transfers information on peptide substrate specificity to small-molecule specificity, we are faced with a complex three-dimensional problem. The relative positions of the features of the amino acid side chains in the peptide substrates and the overall shape of the bound peptide substrates are of high importance. In addition, the relative frequencies of amino acids in the peptide substrate sequences have to be considered. As a shape-based virtual screening method is most suited to address the problem and ROCS also offers the possibility to selectively weight pharmacophore features, shape-based virtual screening with ROCS is the method of choice for our virtual screening problem.

We tested our method on four targets, thrombin, factor Xa (fXa), factor VIIa (fVIIa), and caspase-3 (casp-3), which were selected according to substrate specificity profiles. In addition

to showing different substrate specificities, the proteases also have different catalytic mechanisms. Thrombin, fXa, and fVIIa are serine proteases, while casp-3 is a cysteine protease. Cleavage-site sequence logos for all four targets are shown in Figure 1. Protease subpockets are termed S4–S4' on the basis

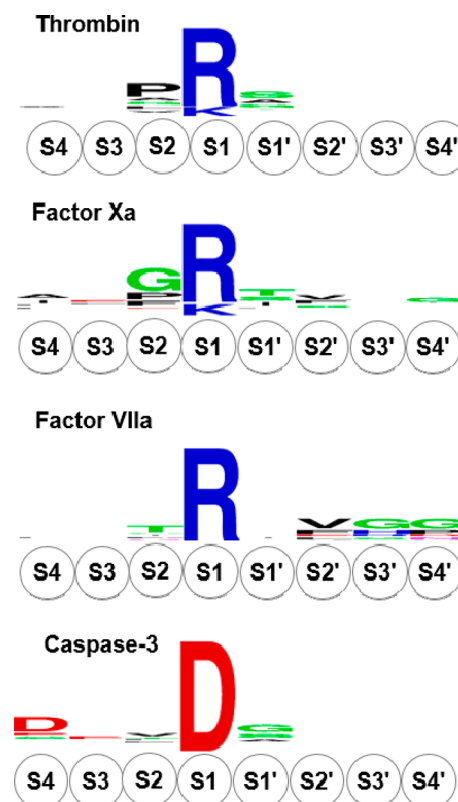


Figure 1. Cleavage-site sequence logos for thrombin (168 substrates), fXa (59 substrates), fVIIa (9 substrates), and casp-3 (651 substrates). The sequence logos were created with weblogo.²³ The height of the single-letter amino acid code indicates the preference for that amino acid in the respective subpocket. Thrombin, fXa, and fVIIa show almost the same substrate specificity in S1 but differ in other subpockets. Casp-3 shows the typical DEVD substrate specificity of caspases.²⁴

of the corresponding substrate positions P4–P4' according to the convention of Schechter and Berger.¹⁹ The peptide's scissile bond lies between P1 and P1'.

The developed workflow is schematically depicted in Figure 2.

METHODS

Preparation of Substrate Sequence Data. Substrate sequences were downloaded from the MEROPS database.² Substrate positions P3–P1 were considered in a first step, as most known inhibitors for the investigated proteases bind to the corresponding protease subpockets. For casp-3, tetrapeptides ranging from P4–P1 were also explored. Unique tri- or tetrapeptide sequences were downloaded from MEROPS.

Preparation of Substrate Structure Data. As MEROPS provides only substrate sequences but no information on substrate conformations, a way to convert the two-dimensional sequences into three-dimensional structures is needed. It is known that proteases universally recognize β -strands in their binding sites.²⁰ To obtain peptides in β -strand conformations,

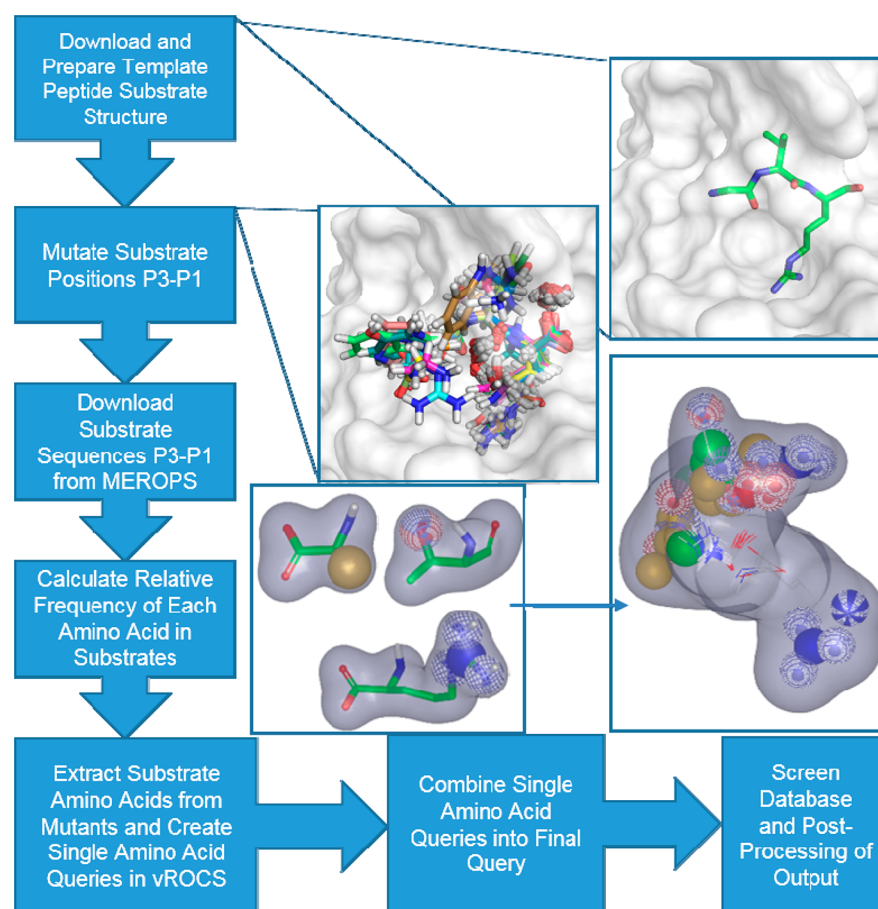


Figure 2. Workflow for shape-based virtual screening with vROCS using thrombin as an example. First, a suitable template peptide substrate structure has to be extracted from the X-ray structure of a peptide substrate complex or manually generated. Only the amino acid positions P3–P1 are kept. With a MOE residue scan, each position P3–P1 is mutated into each of the 20 natural amino acids, leading to a mutational space of 60. Through comparison with the substrate data downloaded from the MEROPS database, only the mutated amino acids present in the protease substrate sequences are kept. In vROCS, single amino acid queries without considering the backbone are first created. In a second step, the final query is created, and single amino acid queries for each position are combined according to the relative frequencies in the protease substrates. The final query is used to perform a virtual screening with vROCS.

we decided to use a mutation strategy based on a known X-ray structure of a protease–substrate complex downloaded from the Protein Data Bank (PDB).²¹ For fVIIa and fXa, no suitable complex structures could be found, so the same template was used for the three serine proteases fVIIa, fXa, and thrombin (PDB code 1FPH²²).

For casp-3, a different template was selected, as a template protease–substrate structure was available (PDB code 2DKO²⁵). The Molecular Operating Environment (MOE) software²⁶ was used for preparation of substrate conformations. Only the template peptide substrate positions P3–P1 or P4–P1 were kept. Mutations of the selected substrate positions were carried out using the residue scan functionality within the MOE software. The residue scan functionality allows one to perform single-point or multiple mutations within a peptide sequence.

Mutating each peptide position independently to all of the 20 amino acids leads to a mutational space of 60 for a tripeptide. Using the peptide substrate sequence lists, individual amino acids present in the peptide substrates listed in MEROPS were extracted from the 60 mutated sequences generated with the MOE residue scan for each position P3–P1 or P4–P1. The single amino acids for each substrate position were written to individual pdb files.

Preparation of Databases. We used the DUD-E database²⁷ for all four of our test cases. Database preparation was carried out with MOE. Duplicate entries were removed, and both the actives and decoys of all data sets were subjected to the MOE wash procedure to disconnect simple metal salts drawn in covalent notation, remove counterions and solvent molecules, add or remove explicit hydrogen atoms, and rebalance protonation states.

For shape-based virtual screening, 25 conformations for each active and decoy were created with OMEGA.^{28,29} The actives database for casp-3 required special attention because several entries contained not the bioactive but the prodrug form of the molecule. Prior to conformer generation, we manually hydrolyzed the lactones in the prodrug structures in MOE.

Potentially covalently bound molecules were kept, as the interactions directing the ligand into the subpockets should still be found.

Shape-Based Virtual Screening. To create the query for shape-based virtual screening, first each individual amino acid was loaded into vROCS, and the backbone features were disabled. For alanine, a hydrophobic feature was added because vROCS did not do this automatically as it did with the functionalities of the other amino acids. Each amino acid was then saved as a separate single amino acid query.

Table 1. ROCS Results^a

protease	AUC	EF ^{1%}	EF ^{2%}	EF ^{5%}	N _{substrates(MEROPS)}	N _{actives}	N _{decoys}
thrombin	0.66	20.36	19.51	9.65	168	369	25174
fXa	0.74	15.98	13.08	7.80	59	413	24893
fVIIa	0.84	4.30	2.79	8.19	9	68	1782
casp-3 (P3-P1)	0.75	0.50	0.50	1.69	651	199	10666
casp-3 (P4-P1)	0.74	1.99	2.48	2.58	651	199	10666

^aThe AUC values are better for fVIIa and casp-3 than for thrombin and fXa. However, thrombin and fXa show high early enrichment values. The results for fVIIa show that even without a low number of known substrate sequences, high AUC values and early enrichment can be achieved. The numbers of ranked actives and decoys in the data set are included, as the results for AUC and early enrichment are dependent on them.

To create a query correctly representing the relative frequencies of amino acid side chains in the preferred substrates of the corresponding protease, the relative frequencies were first calculated in the following way: Absolute frequencies were normalized according to the number of unique peptide substrate sequences and natural occurrence of amino acids. The normalization by the natural occurrence of the amino acids was needed to remove the bias in the experimental results³⁰ of the MEROPS peptide substrate sequences. As vROCS does not allow the number of times a feature should appear in the final query to be set, each individual amino acid query has to be loaded in according to the relative frequency in the protease peptide substrates. Since vROCS does not handle a large number of different amino acid queries to be loaded in a large number of times, we further normalized the frequencies in such a way that the most frequently occurring amino acid in the substrate has a frequency of 20. Tables with relative amino acid frequencies for each protease example can be found in Tables S1–S5 in the Supporting Information. To build the final query, each single amino acid query was loaded into vROCS according to the obtained frequency table. The query was then used in a ROCS validation run using the prepared actives and decoys data set. Of the 25 conformations for each active and decoy, only the highest ranked conformation was kept. Enrichment factors at X% (EF^{X%}) were calculated according to the following metric:³¹

$$EF^{X\%} = \frac{Actives_{sampled}}{N_{sampled}} \frac{N_{total}}{Actives_{total}}$$

where Actives_{sampled} is the number of actives found at X% of the screened database, N_{sampled} is the number of compounds at X% of the database, N_{total} is the number of compounds in the database, and Actives_{total} is the number of actives in the database.

RESULTS

The results of the shape-based virtual screening are summarized in Table 1. In addition to enrichment factors at 1 and 2% of the database screened, enrichment factors at 5% are also shown, as they might be more relevant for industry-scale applications. Figure 3 shows the receiver operating characteristic (ROC) curves for the results listed in Table 1. The results of performing the virtual screening using the query of one protease with the data set of the other protease are shown in Table 2.

DISCUSSION

Thrombin. The results for thrombin are lowest in terms of area under the curve (AUC) when screening the DUD-E database, but at the same time, the early enrichment is highest.

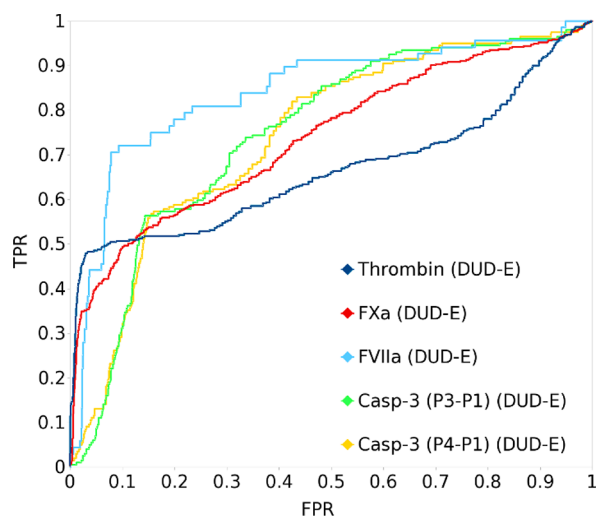


Figure 3. ROC curves for all examples and data sets. Early enrichment is high in all cases except for casp-3.

Table 2. Validation Runs for Shape-Based Virtual Screening^a

validation run	AUC
casp-3 with thrombin query	0.56
casp-3 with fXa query	0.63
casp-3 with fVIIa query	0.56
fXa with thrombin query	0.70
thrombin with fXa query	0.73

^aThe AUC values for the validation runs with the casp-3 data set show that actives are not found if the query from a different protease is used. The results for both the thrombin and fXa data sets show that it is not possible to differentiate between fXa and thrombin ligands, as their substrate specificities are too similar.

The highest-ranked decoys for thrombin all show the guanidine functionality at the P1 position, which is also fundamental for substrate recognition in the thrombin peptide substrates.³²

With regard to shape as well as chemical functionalities, the highest-ranked decoys look like classical thrombin inhibitors (Figure 4).³³ The lowest-ranked actives on the one side are smaller than the ROCS query, which leads to a penalty in volume overlap and thus to a lower ranking. In addition, most of them do not have the characteristic thrombin interacting groups and in general miss functional groups that allow for strong selective interactions with the binding site.

Factor Xa. In the same way as for thrombin, the highest-ranked decoys for fXa all contain the guanidine group and are shaped like classical fXa inhibitors. The number of peptides used for creation of the ROCS query for fXa is much lower than for thrombin, as in comparison there is little data in the

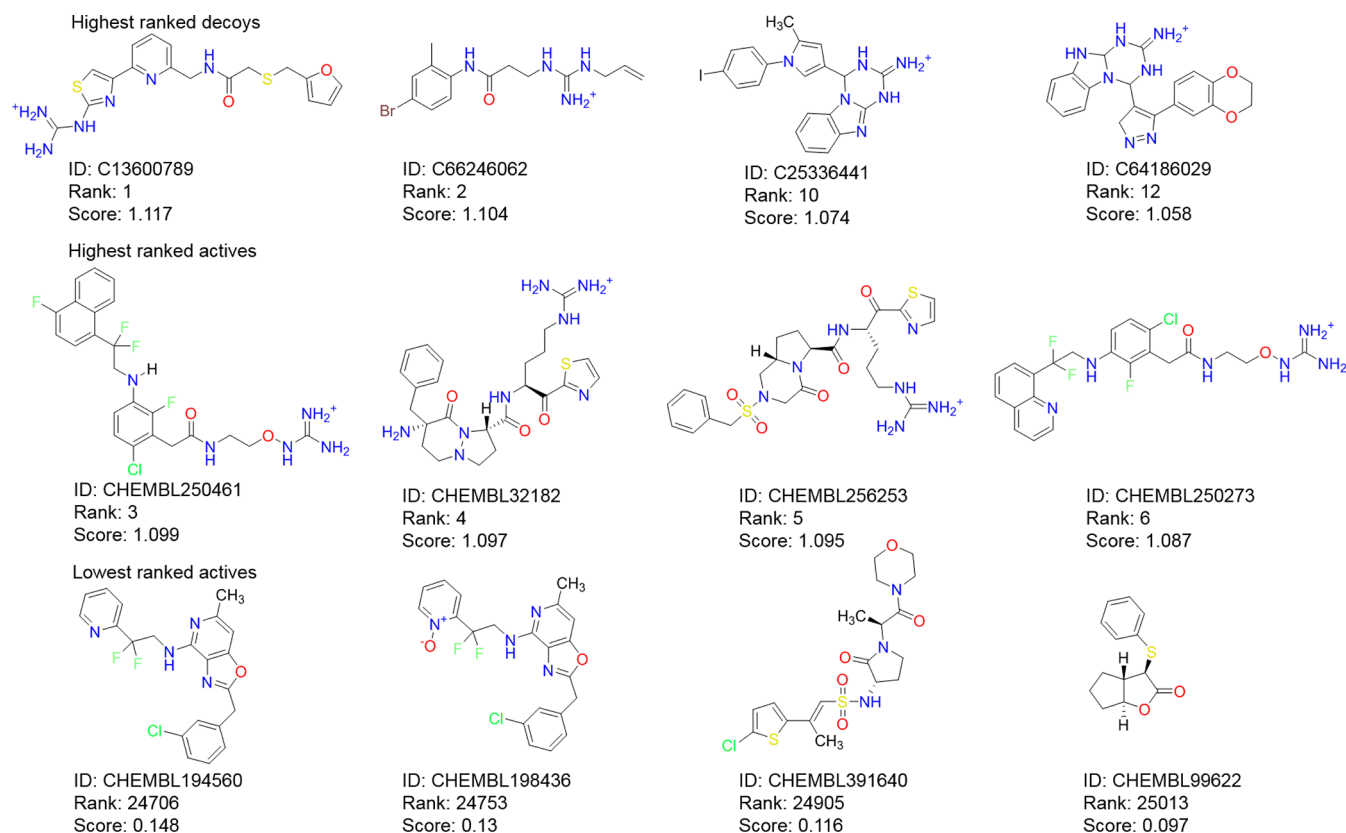


Figure 4. Highest-ranked decoys and actives and lowest-ranked actives when screening the DUD-E validation database. Both the highest-ranked actives and decoys possess guanidine-like functionalities at the P1 position, which are also preferred in the peptide substrates of thrombin. The lowest-ranked actives do not show the typical functional groups present in the thrombin peptide substrates. In addition, some of the inhibitors are a lot smaller than the query, which results in a large shape penalty.

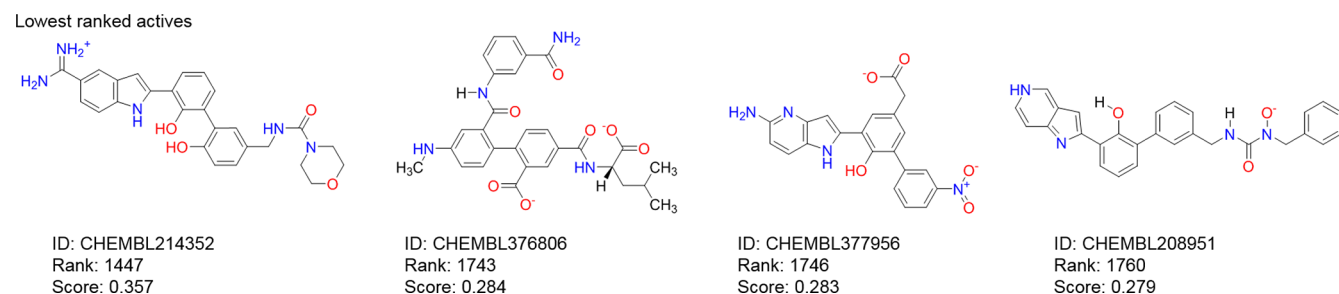


Figure 5. Lowest-ranked actives of fVIIa. Some actives are ranked low despite possessing the guanidine functionality important for binding to the S1 subpocket. This is probably caused by missing information in the vROCS query, as there are only nine substrates listed for fVIIa in the MEROPS database.

MEROPS database about fXa substrates. Despite the limited number of available substrates, the AUC values are quite high when screening the DUD-E database.

Factor VIIa. Also in fVIIa the highest-ranked actives and decoys all possess the guanidine functionality at the S1 binding position. In the case of fVIIa, the lowest-ranked actives miss the guanidine functionality. They even possess negatively charged groups in some cases, in contrast to the substrate specificity at the S1 position (Figure 5). For fVIIa there are only nine substrates listed in the MEROPS database, which is even fewer than for fXa. Therefore, for fVIIa the vROCS query might miss some important information because of incomplete substrate data. However, in view of the low number of known substrates, it is impressive how good the method performs in terms of AUC and early enrichment.

Caspase-3. In the case of casp-3, the carboxylate group at the S1 position seems to be required for the compound to be a high-ranked active or decoy (Figure 6). Interestingly, among the highest-ranked actives, several of them are prodrugs.³⁴ If the lactone functionality in the prodrugs is not opened and converted to the bioactive form, they are ranked lowest in the virtual screen. However, if used in their bioactive form, small molecules that are administered in prodrug form are among the highest-ranked actives. As casp-3 shows typical DEVD specificity²⁵ and thus also high specificity at S4, for casp-3 we used a model based on positions P3–P1 as well as a second model based on P4–P1. Using a broader substrate position range did not considerably improve the AUC and early enrichment. However, different actives and decoys were ranked highest, depending on how many substrate positions were used.

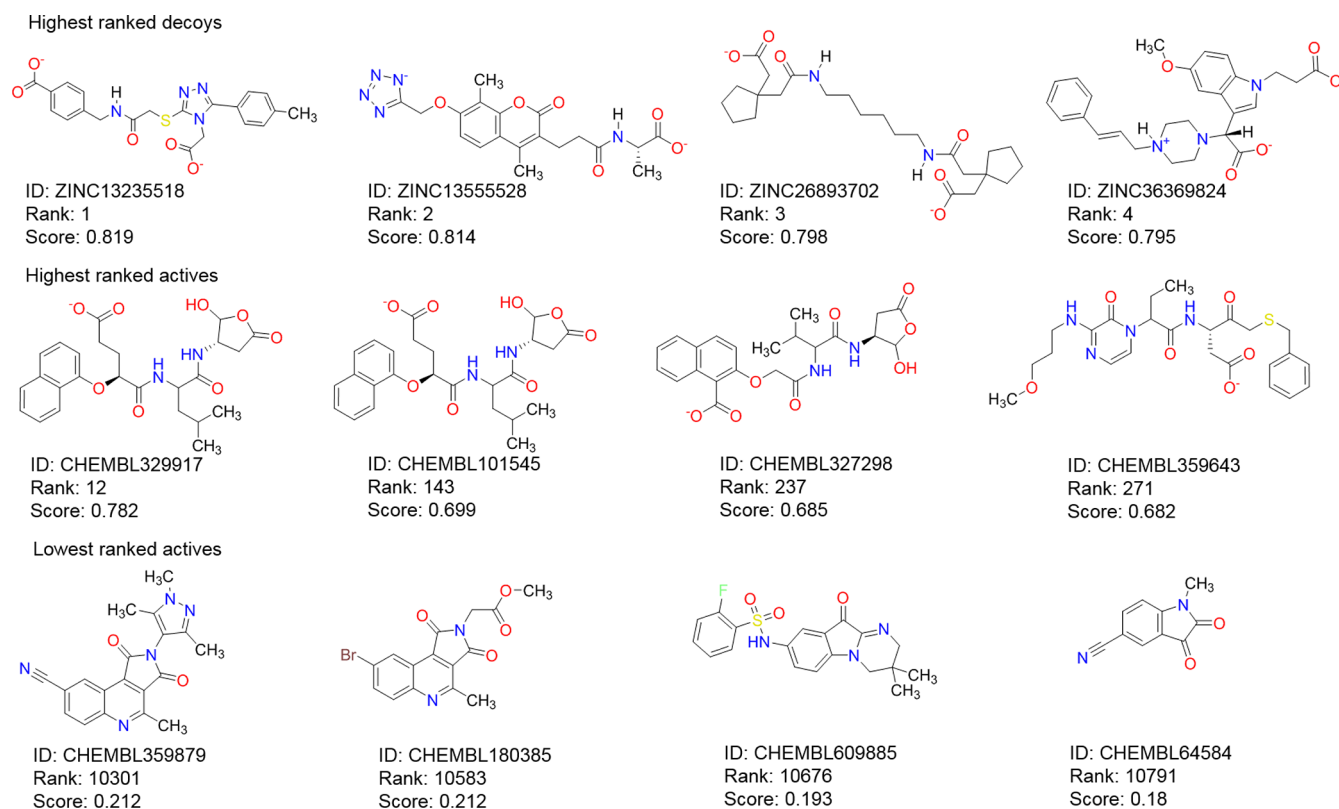


Figure 6. Examples of high-ranked actives and the lowest-ranked actives for casp-3 when substrate positions P3–P1 were used. It should be noted that CHEMBL329917, CHEMBL101545, and CHEMBL327298 were used in their bioactive form in the virtual screening experiments, which means that the lactone ring present in the prodrug form was opened. The lowest-ranked actives are small and do not possess the negatively charged group that could interact favorably with the S1 subpocket of casp-3. This leads to a high shape penalty in shape-based virtual screening and to a lack of matching functional groups with the vROCS query.

Table 3. Influence of Different Protease–Substrate Complex Templates^a

protease	AUC	EF ^{1%}	EF ^{2%}	$N_{\text{substrates(MEROPS)}}$	N_{actives}	N_{decoys}
fXa (thrombin protease–substrate complex template)	0.74	15.98	13.08	59	413	24893
fXa (casp-3 protease–substrate complex template)	0.68	6.54	10.66	59	413	24893
fVIIa (thrombin protease–substrate complex template)	0.84	4.23	2.79	9	68	1782
fVIIa (casp-3 protease–substrate complex template)	0.84	2.86	2.21	9	68	1782

^aThe AUC values and enrichment factors for the DUD-E screening for fXa decrease when a casp-3 protease–substrate complex template is used instead of a thrombin protease–substrate complex template. For fVIIa the choice of protease–substrate complex template has no effect on the AUC, while the early enrichment is slightly lower for the casp-3 protease–substrate complex template.

The lowest-ranked actives were similar for both substrate position ranges, however.

Importance of the Template Peptide. As the results of the shape-based virtual screening runs may very much depend on the query conformation, we investigated the importance of the template peptide. We compared the results of using either a thrombin protease–substrate complex as the template for the mutation strategy or a casp-3 protease–substrate complex for fXa and fVIIa, for which there are no protease–substrate complexes available in the PDB. The results in Table 3 show that for the fXa DUD-E validation runs, the results do get a little worse in terms of AUC and early enrichment when a casp-3 protease–substrate complex is used as the template for the mutation strategy instead of a thrombin protease–substrate complex. For fVIIa the AUC is not affected by using a different protease–substrate complex as the template for the mutation strategy. Only the early enrichment values decrease a little when the casp-3 protease–substrate complex is used as a template instead of the thrombin protease–substrate complex.

The results show that the mutation strategy works even when the peptide substrate sequences and the template peptide show low sequence identity. As long as a template peptide in an extended β -sheet conformation is available, our method can be applied.

Comparison with Alternative Virtual Screening Strategies. The main advantage of our method is that it does not require a structure or knowledge of any small-molecule ligands for a virtual screening to be performed when dealing with a protease target. Only information on protease substrate sequences is required. If there are no substrates for the desired protease target listed in the MEROPS database, substrate specificity profiling is done rather quickly, in comparison with generating a structure or finding small-molecule inhibitors.

The advantage compared with the method of Sukuru et al.¹⁸ is that we directly transfer the information about the known peptide substrates for a protease to the small-molecule space. Thus, to find new inhibitors no prior knowledge of small-molecule ligands is required.

As ROCS uses a very fast and efficient algorithm for the virtual screening runs, hundreds of thousands of molecules can be screened within hours. In combination with the easy accessibility of the data required for building the query, our method has significant advantages over docking and other structure-based methods as well as ligand-based approaches using small-molecule ligands as the basis for virtual screening experiments.

CONCLUSION

We have presented a method that enables the fast and efficient derivation of a model derived from protease peptide substrate data that can be readily applied to screen for small-molecule ligands. We have applied it to four different proteases that cover different active-site mechanisms, substrate specificities, and binding-site shapes. In all four cases, the method performed well in terms of AUC and early enrichment. Even in the case of fVIIa and fXa, where available substrate data is limited, the method successfully recovered actives from the very challenging data sets prepared from the DUD-E database. The workflow described herein represents the first approach to use protease substrate sequences as the training set for a virtual screening experiment. As the query creation in vROCS allows one to include information on the relative frequencies of amino acids of substrates in the respective subpockets and focus on the properties of side chains in substrates, scaffold hopping is made possible. The method can easily be applied to different protease systems. Thus, we believe it can also be applied to members of other enzyme types, such as kinases. In summary, we have developed a new tool to be used for rational drug design, allowing the huge amount of data on protease substrates to be used for finding new small-molecule inhibitors.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.6b00064.

Relative frequencies of amino acid substrate positions (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: Klaus.Liedl@uibk.ac.at.

Present Address

†C.K.: F. Hoffmann-La Roche AG, Grenzacherstrasse 124, 4070 Basel, Switzerland.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

B.J.W. is thankful to the Austrian Academy of Sciences for being a recipient of the DOC Grant. The authors also thank the Austrian Science Fund (FWF) for funding of Project P 23051.

ABBREVIATIONS

fVIIa, factor VIIa; fXa, factor Xa; casp-3, caspase-3

REFERENCES

(1) Drag, M.; Salvesen, G. S. Emerging Principles in Protease-Based Drug Discovery. *Nat. Rev. Drug Discovery* **2010**, *9*, 690–701.

(2) 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.

(3) Deu, E.; Verdoes, M.; Bogyo, M. New Approaches for Dissecting Protease Functions to Improve Probe Development and Drug Discovery. *Nat. Struct. Mol. Biol.* **2012**, *19*, 9–16.

(4) Geppert, H.; Vogt, M.; Bajorath, J. r. Current Trends in Ligand-Based Virtual Screening: Molecular Representations, Data Mining Methods, New Application Areas, and Performance Evaluation. *J. Chem. Inf. Model.* **2010**, *50*, 205–216.

(5) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat. Rev. Drug Discovery* **2004**, *3*, 935–949.

(6) Huang, S. Y.; Zou, X. Ensemble Docking of Multiple Protein Structures: Considering Protein Structural Variations in Molecular Docking. *Proteins: Struct., Funct., Genet.* **2007**, *66*, 399–421.

(7) Horvath, D. Pharmacophore-Based Virtual Screening. In *Cheminformatics and Computational Chemical Biology*; Springer: New York, 2011; pp 261–298.

(8) Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. Scaffold-Hopping” by Topological Pharmacophore Search: A Contribution to Virtual Screening. *Angew. Chem., Int. Ed.* **1999**, *38*, 2894–2896.

(9) ROCS, version 3.2.1.4; OpenEye Scientific Software, Santa Fe, NM; <http://www.eyesopen.com>.

(10) Hawkins, P. C.; Skillman, A. G.; Nicholls, A. Comparison of Shape-Matching and Docking as Virtual Screening Tools. *J. Med. Chem.* **2007**, *50*, 74–82.

(11) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **2006**, *49*, 6789–6801.

(12) Kirchmair, J.; Distinto, S.; Markt, P.; Schuster, D.; Spitzer, G. M.; Liedl, K. R.; Wolber, G. How to Optimize Shape-Based Virtual Screening: Choosing the Right Query and Including Chemical Information. *J. Chem. Inf. Model.* **2009**, *49*, 678–692.

(13) Schilling, O.; Overall, C. M. Proteome-Derived, Database-Searchable Peptide Libraries for Identifying Protease Cleavage Sites. *Nat. Biotechnol.* **2008**, *26*, 685–694.

(14) Schilling, O.; Barré, O.; Huesgen, P. F.; Overall, C. M. Proteome-Wide Analysis of Protein Carboxy Termini: C Terminomics. *Nat. Methods* **2010**, *7*, 508–511.

(15) Kleifeld, O.; Doucet, A.; auf dem Keller, U.; Prudova, A.; Schilling, O.; Kainthan, R. K.; Starr, A. E.; Foster, L. J.; Kizhakkedathu, J. N.; Overall, C. M. Isotopic Labeling of Terminal Amines in Complex Samples Identifies Protein N-Termini and Protease Cleavage Products. *Nat. Biotechnol.* **2010**, *28*, 281–288.

(16) 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.

(17) Fuchs, J. E.; von Grafenstein, S.; Huber, R. G.; Kramer, C.; Liedl, K. R. Substrate-Driven Mapping of the Degradome by Comparison of Sequence Logos. *PLoS Comput. Biol.* **2013**, *9*, e1003353.

(18) Sukuru, S. C. K.; Nigsch, F.; Quancard, J.; Renatus, M.; Chopra, R.; Brooijmans, N.; Mikhailov, D.; Deng, Z.; Cornett, A.; Jenkins, J. L.; Hommel, U.; Davies, J. W.; Glick, M. A Lead Discovery Strategy Driven by a Comprehensive Analysis of Proteases in the Peptide Substrate Space. *Protein Sci.* **2010**, *19*, 2096–2109.

(19) Schechter, I.; Berger, A. Protease Subsite Nomenclature. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157–162.

(20) Tyndall, J. D.; Nall, T.; Fairlie, D. P. Proteases Universally Recognize Beta Strands in their Active Sites. *Chem. Rev. (Washington, DC, U. S.)* **2005**, *105*, 973–1000.

(21) Berman, H.; Henrick, K.; Nakamura, H. Announcing the Worldwide Protein Data Bank. *Nat. Struct. Biol.* **2003**, *10*, 980–980.

(22) Stubbs, M. T.; Oschkinat, H.; Mayr, I.; Huber, R.; Anglikar, H.; Stone, S. R.; Bode, W. The Interaction of Thrombin with Fibrinogen. *Eur. J. Biochem.* **1992**, *206*, 187–195.

(23) Crooks, G. E.; Hon, G.; Chandonia, J.-M.; Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **2004**, *14*, 1188–1190.

(24) Talanian, R. V.; Quinlan, C.; Trautz, S.; Hackett, M. C.; Mankovich, J. A.; Banach, D.; Ghayur, T.; Brady, K. D.; Wong, W. W. Substrate Specificities of Caspase Family Proteases. *J. Biol. Chem.* **1997**, *272*, 9677–9682.

(25) Ganesan, R.; Mittl, P. R.; Jelakovic, S.; Grütter, M. G. Extended Substrate Recognition in Caspase-3 Revealed by High Resolution X-Ray Structure Analysis. *J. Mol. Biol.* **2006**, *359*, 1378–1388.

(26) *Molecular Operating Environment (MOE)*, version 2013.08; Chemical Computing Group: Montreal, QC, Canada, 2013.

(27) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55*, 6582–6594.

(28) Hawkins, P. C.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572–584.

(29) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. OMEGA, version 2.5.1.4; OpenEye Scientific Software: Santa Fe, NM; <http://www.eyesopen.com>.

(30) McCaldon, P.; Argos, P. Oligopeptide Biases in Protein Sequences and their Use in Predicting Protein Coding Regions in Nucleotide Sequences. *Proteins: Struct., Funct., Genet.* **1988**, *4*, 99–122.

(31) Good, A. C.; Oprea, T. I. Optimization of CAMD Techniques 3. Virtual Screening Enrichment Studies: A Help or Hindrance in Tool Selection? *J. Comput.-Aided Mol. Des.* **2008**, *22*, 169–178.

(32) Brandstetter, H.; Kühne, A.; Bode, W.; Huber, R.; von der Saal, W.; Wirthensohn, K.; Engh, R. A. X-ray Structure of Active Site-Inhibited Clotting Factor Xa Implications for Drug Design and Substrate Recognition. *J. Biol. Chem.* **1996**, *271*, 29988–29992.

(33) Winquist, J.; Geschwindner, S.; Xue, Y.; Gustavsson, L.; Musil, D.; Deinum, J.; Danielson, U. H. Identification of Structural–Kinetic and Structural–Thermodynamic Relationships for Thrombin Inhibitors. *Biochemistry* **2013**, *52*, 613–626.

(34) Rautio, J.; Kumpulainen, H.; Heimbach, T.; Oliyai, R.; Oh, D.; Järvinen, T.; Savolainen, J. Prodrugs: Design and Clinical Applications. *Nat. Rev. Drug Discovery* **2008**, *7*, 255–270.