


## RESEARCH ARTICLE

# Electrostatic recognition in substrate binding to serine proteases

Birgit J. Waldner<sup>1</sup> | Johannes Kraml<sup>1</sup> | Ursula Kahler<sup>1</sup> | Alexander Spinn<sup>1</sup> | Michael Schauer<sup>1</sup> | Maren Podewitz<sup>1</sup> | Julian E. Fuchs<sup>1†</sup> | Gabriele Cruciani<sup>2</sup> | Klaus R. Liedl<sup>1</sup> 

<sup>1</sup>Institute of General, Inorganic and Theoretical Chemistry, and Center for Molecular Biosciences Innsbruck (CMBI), University of Innsbruck, Innsbruck, Austria

<sup>2</sup>Laboratory of Chemometrics, Department of Chemistry, University of Perugia, Perugia, Italy

**Correspondence**

Klaus R. Liedl, Institute of General, Inorganic and Theoretical Chemistry, and Center for Molecular Biosciences Innsbruck (CMBI), University of Innsbruck, Innrain 80-82, A-6020 Innsbruck, Austria.  
Email: klaus.liedl@uibk.ac

**Present Address**

Julian E. Fuchs, Medicinal Chemistry/ Structural Research, Boehringer Ingelheim RCV GmbH & Co. KG, Vienna, Austria

**Funding information**

Austrian Science Fund (FWF), Grant/Award Number: P30565; Austrian Academy of Sciences, Grant/Award Number: DOC Grant to BJW

**Abstract**

Serine proteases of the Chymotrypsin family are structurally very similar but have very different substrate preferences. This study investigates a set of 9 different proteases of this family comprising proteases that prefer substrates containing positively charged amino acids, negatively charged amino acids, and uncharged amino acids with varying degree of specificity. Here, we show that differences in electrostatic substrate preferences can be predicted reliably by electrostatic molecular interaction fields employing customized GRID probes. Thus, we are able to directly link protease structures to their electrostatic substrate preferences. Additionally, we present a new metric that measures similarities in substrate preferences focusing only on electrostatics. It efficiently compares these electrostatic substrate preferences between different proteases. This new metric can be interpreted as the electrostatic part of our previously developed substrate similarity metric. Consequently, we suggest, that substrate recognition in terms of electrostatics and shape complementarity are rather orthogonal aspects of substrate recognition. This is in line with a 2-step mechanism of protein-protein recognition suggested in the literature.

**KEYWORDS**

electrostatic similarity, encounter complex, molecular interaction fields, protease, substrate, substrate recognition

## 1 | INTRODUCTION

In humans, more than 560 different genes code for proteases<sup>1</sup> constitute nearly 3% of all approximately 19 000 human genes<sup>2</sup>; therefore, proteases have been investigated as drug targets extensively.<sup>3</sup> Many protease inhibitors are already very successful drugs, eg, to control blood coagulation,<sup>4</sup> to treat hypertension and diabetes,<sup>5</sup> to fight cancer, and to combat viral diseases like HIV<sup>6</sup> and Hepatitis C.<sup>7</sup>

However, protease promiscuity and overlapping specificity profiles are major problems that have to be overcome in most protease inhibitor drug design efforts.<sup>8,9</sup>

Proteases catalyze the hydrolysis of peptide bonds through acceleration of the nucleophilic attack on the peptide amide group, which would otherwise be kinetically stable.<sup>10</sup> To fulfill their biological function, proteases have to bind their native substrates and stabilize the transition state for hydrolysis of the peptide bond.<sup>11</sup> Promiscuity and specificity of this recognition process not only exhibit striking differences between different proteases, but also show large

Birgit J. Waldner and Johannes Kraml has contributed equally.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2018 The Authors Journal of Molecular Recognition Published by John Wiley & Sons Ltd

variations within the binding cleft of an individual protease.<sup>12–14</sup> Obviously, evolution had to tackle quite different tasks—on the one hand, designing proteases that are able to digest more or less every peptide that they encounter, and on the other hand, designing proteases within a signaling cascade, that should specifically recognize the subsequent member of the signaling chain to ensure the proper transmission of the signal.<sup>15</sup> This evolutionary pressure led to proteases ranging from highly promiscuous to extremely specific. Surprisingly, these extremes can occur within the same family of evolutionarily related proteases, eg, the Chymotrypsin family of serine proteases.<sup>16</sup>

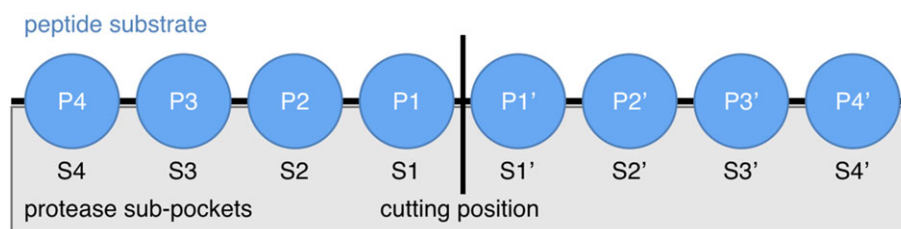
The promiscuity and specificity of substrate recognition are very often not spread evenly along the binding cleft. Proteases often prefer certain substrate amino acids in given distances to their catalytic center. Half a century ago, Schechter and Berger<sup>17</sup> suggested a convention to denote the peptide substrate amino acids (P4 to P4') and the subpockets (S4 to S4') within the binding cleft around the scissile bond (cf. Figure 1).

Several methods have been published to describe and localize promiscuity and specificity of the protease binding interface, thus facilitating a comparison of individual proteases.<sup>18–21</sup> Our cleavage entropy metric<sup>14,22</sup> is based on substrate data deposited in the MEROPS database<sup>23</sup> and quantifies the specificity of peptide recognition in each subpocket. To compare proteases based on their substrate recognition, we developed a metric that considers the positional abundance of individual amino acids.<sup>24</sup> These methods allow the investigation of localized regions of promiscuity and specificity in the binding interface of proteases and the analysis of their thermodynamic properties. Thus, different subpockets of the same protease as well as binding clefts of different proteases can be studied and compared.

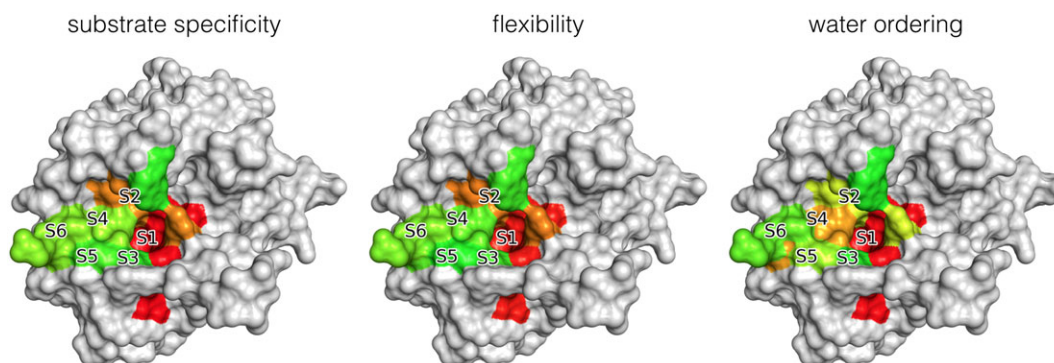
Research on snake venom metalloproteases revealed strong hints that their promiscuity is linked to their flexibility.<sup>25</sup> Likewise, we found that Caspases<sup>26</sup> and Thrombin<sup>27</sup> display strong correlations between flexibility and promiscuity. For Thrombin, this correlation translates into ordering processes of water molecules in the binding interface. Regions of specificity show ordered water molecules in the interface, whereas regions of promiscuity tend to have more disordered water molecules in the first solvation sphere (Figure 2). On the other hand, enthalpic contributions to hydration of the S1 and S4 to S6 are almost identical. Thus, dynamic, and therefore entropy of hydration, contributes strongly to the recognition of substrates in Thrombin.

Electrostatic interactions are quite different from other contributions of substrate recognition as they are long-range interactions that change little with small differences in distance.<sup>28,29</sup> This has several important consequences. Obviously, a more continuous distance dependence varies less with conformational changes, much in contrast to shape-dependent recognition like van der Waals interactions and recognition that relies on precise exit vectors like hydrogen bonds. On the contrary, due to the long-range character of electrostatic interactions, assigning them to specific subpockets is more challenging.

Calculating differences in electrostatic molecular interaction fields (eMIFs) of proteins is a rather challenging task. Many different approaches exist, and all of them have a significant impact on the result. Differences in the handling of the solvent and the solute, either implicit as a continuum or explicit, can yield highly different results.<sup>30</sup> When using an implicit model, it is also not trivial to assign each point on the grid a certain value for the dielectric constant. This problem is irrelevant for high distances to the solute but can yield errors for points close to it.<sup>31</sup> Furthermore, differences in handling multipoles will also introduce differences in the results.<sup>32</sup> The biggest error, however, is included when using different protonation states for the



**FIGURE 1** Peptide substrate amino acid (Pi and Pi') and protease subpocket enumeration (Si and Si') with respect to the cutting position (vertical line). The N-terminal side of the substrate is located on the left



**FIGURE 2** Correlation of substrate specificity with backbone flexibility and orientational ordering of water molecules in the non-prime site (S6–S1) of Thrombin's binding cleft (ranging from red—specific, rigid and ordered, via yellow to green—promiscuous, flexible and disordered)<sup>27</sup>

model, as introducing an extra charge, or removing one, changes the entire electrostatic field significantly.

In a previous study, we used GRID-probes that test van der Waals interactions and electrostatics simultaneously. Even taking into account conformations extracted from molecular dynamics trajectories, we could only achieve limited correlation with substrate recognition.<sup>33</sup>

To predict the specificity of proteases, Pethe et al<sup>34</sup> used a structure-based approach that ranks possible substrates according to interaction energies and reorganization penalties. Their scheme outperforms conventional methods that focus solely on knowledge-based prediction of substrate preferences.

Okun and Chen compared proteases with a statistical model. They calculated electrostatic similarities using a volumetric overlay of isopotentials.<sup>35</sup>

In PIPSA,<sup>36</sup> the Hodgkin index is used to compare different Molecular Interaction Fields of proteins. The program was also used by Henrich et al<sup>37</sup> to compare the electrostatic similarity of the 3 proteases, Trypsin, Thrombin, and uPA.

Various approaches are already available that compare binding sites,<sup>38</sup> often for the purpose of off-target prediction and drug repurposing.<sup>39</sup> Such methods rely on molecular interaction fields (MIFs), eg, BioGPS<sup>40,41</sup> and IsoMIF,<sup>42</sup> on shape and physicochemical properties of the surface, eg, protein functional surfaces,<sup>43</sup> on graphs representing the 3D atomic similarities, eg, IsoCleft<sup>44</sup> or on fingerprints describing the binding sites, eg, PocketMatch.<sup>45</sup> In several data bases, properties of binding sites are stored for comparison, such as pseudocenters with projected physicochemical properties in CavBase,<sup>46-48</sup> in CavSimBase<sup>49</sup> and in SiteEngine,<sup>50</sup> sequence and structural similarity in CPASS,<sup>51</sup> position of functional groups in SuMo,<sup>52</sup> or surface geometrics and electrostatics in eF-site.<sup>53</sup>

However, most of these methods are not meant to compare structurally very similar cavities as found in our set of chymotrypsin-like proteases (Figure 3), or define the binding site without ligand information. Therefore, we chose to implement our own method optimized to compare similar binding interfaces and able to compare these interfaces based on a distance criterion to a ligand as described in the methods section below in detail.

## 2 | METHODS

### 2.1 | Electrostatic substrate preferences

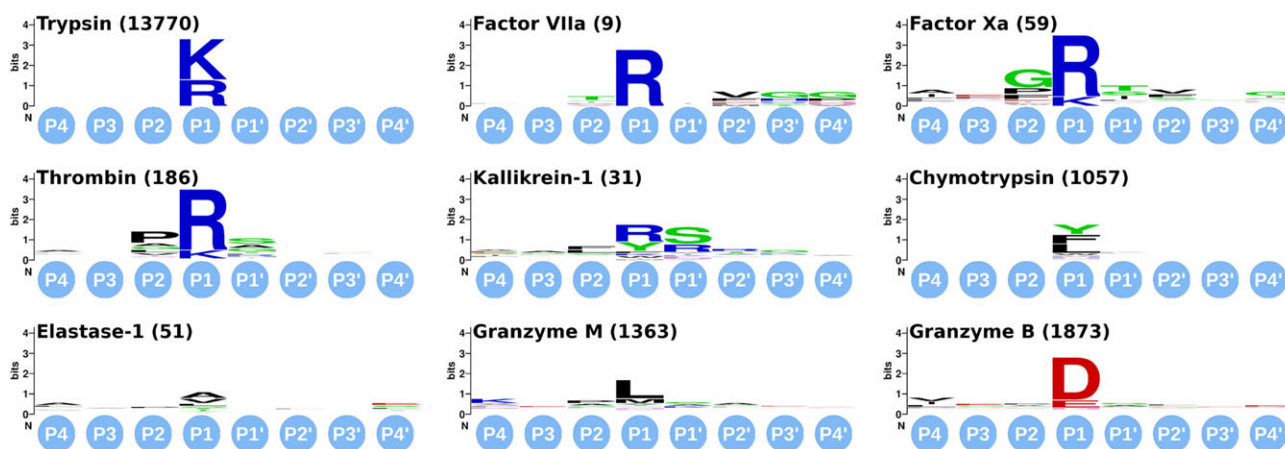
We extracted and isolated the electrostatic contributions and the shape recognition contributions in the substrate preference similarity metric that we defined previously.<sup>24</sup> We achieved this goal by binning the amino acid residues according to their electrostatic properties into positively charged (K, R, H), negatively charged (D, E), and neutral amino acids (G, P, A, V, L, I, M, F, Y, W, S, T, C, N, Q). In this way, we split off shape-dependent and size-dependent aspects of substrate recognition and focus solely on electrostatic recognition and specificity. The shape-dependent aspects of substrate recognition can be studied individually for each of the 3 bins, especially within the bin of neutral amino acids. However, this aspect is beyond the scope of the current study. Considering histidine as a positively charged amino acid is somewhat arbitrary, but in line with the usual classification in sequence logos.<sup>54</sup> Nevertheless, histidine is rather underrepresented in substrate data (1.7%); thus, this choice does not influence the analysis significantly (correlation data is shown in the Supporting Information).

For each of the 9 proteases under investigation, we extracted substrate data from the MEROPS database<sup>23</sup> (Trypsin S01.151, Factor VIIa S01.215, Factor Xa S01.216, Thrombin S01.217, Kallikrein-1 S01.160, Chymotrypsin S01.001, Elastase-1 S01.153, Granzyme M S01.139, Granzyme B S01.010) for substrate residues P4 to P4'.

For each substrate position, amino acids are assigned to 1 of the 3 bins according to their electrostatic properties. For each substrate position, a vector is constructed via Equation 1.

$$v_{Pi} = \left( N_{positive}^{Pi} = \frac{\sum_{a=1}^3 n_a^{Pi}}{\sum_{a=1}^3 p_a}, N_{negative}^{Pi} = \frac{\sum_{a=1}^2 n_a^{Pi}}{\sum_{a=1}^2 p_a}, N_{neutral}^{Pi} = \frac{\sum_{a=1}^{15} n_a^{Pi}}{\sum_{a=1}^{15} p_a} \right) \quad (1)$$

In Equation 1,  $N_{bin}^{Pi}$  is the score in 1 of the 3 electrostatic bins (positive, negative, or neutral) of substrate position  $P_i$ ,  $a$  the index of the amino acids in 3 different bins,  $n_a^{Pi}$  the number of occurrences of amino acid  $a$  in the substrate position  $P_i$ , and  $p_a$  the natural



**FIGURE 3** Cleavage site sequence logos of substrate data used for generating substrate preference similarity metric. Numbers in brackets indicate the number of substrates filed in the MEROPS database for each protease. The logos were generated with WebLogo<sup>54</sup> (the underlying data is supplied in the Supporting Information)

occurrence of amino acid *a*. This weighting with the natural occurrence of the amino acids in human proteins<sup>55</sup> results in an intrinsic normalization of the bins.

Concatenation of vectors  $\mathbf{v}_{p_i}$  for substrate positions P4 to P4' and bin is summarized into a vector of dimension  $3 \cdot 8 = 24$ . This vector is subsequently normalized to 1 and contains the electrostatic substrate preferences for each protease.

$$\mathbf{v}_{\text{Protease}} = \frac{(\mathbf{v}_{P4}, \dots, \mathbf{v}_{P4'})}{\|(\mathbf{v}_{P4}, \dots, \mathbf{v}_{P4'})\|} \quad (2)$$

The electrostatic substrate similarity of 2 proteases is calculated by forming the scalar product between the respective electrostatic substrate preference vectors. A scalar product of 0 indicates orthogonal substrate preferences. A scalar product of 1 implies identical electrostatic substrate preferences, as found when comparing an individual protease with itself, ie, the electrostatic substrate preference. Additionally, the product of every single line of the vector can be interpreted as contribution corresponding to either positive, negative, or neutral substrate residues at every single substrate position. The sum of individual substrate positions indicates the overall contribution of positive, negative, and neutral substrate residues.

## 2.2 | Electrostatic molecular interaction fields (eMIFs)

X-ray structures for all 9 proteases were downloaded from the PDB (Trypsin 1PQ7,<sup>56</sup> Factor VIIa 1KLI,<sup>57</sup> Factor Xa 1C5M,<sup>58</sup> Thrombin 4AYY,<sup>59</sup> Kallikrein-1 1SPJ,<sup>60</sup> Chymotrypsin 4CHA,<sup>61</sup> Elastase-1 1QNJ,<sup>62</sup> Granzyme M 2ZGH,<sup>63</sup> Granzyme B 1FQ3<sup>64</sup>). For Thrombin and for Granzyme M, structures with a ligand were chosen to ensure a conformation of the active form.<sup>65,66</sup> For structures containing a ligand, the ligand was removed.

The structures were aligned with respect to their Ca atoms, and all structures were protonated with Protonate3D<sup>67</sup> and prepared with MOE.<sup>68</sup> The catalytic histidine (His-57) was chosen to be uncharged in all structures.

Using the program GRID,<sup>69,70</sup> 3 molecular interaction fields (MIFs) were calculated on a grid for the entire binding interface of the proteases. For each probe, a grid spacing of 1 Å was used. For the first MIF, a hydrophobic H-probe was used in order to characterize the shape of the binding cleft. It was further restricted by a distance criterion (5 Å) to ligands (P4 to P4') in aligned peptide complex structures, ie, 1DE7, 3LU9 (Thrombin), and 2ZGH (Granzyme M). Only grid points that fulfill the distance criterion and show favorable interactions (<0 kcal/mol) with the H-probe were used in the further calculations of the eMIFs. In order to minimize van der Waals interactions and focus on electrostatic contributions alone, the eMIFs were calculated with user-defined GRID probes with charges of +1 and -1. Both, the van der Waals radius of the probes and the cutoff for the van der Waals interactions, were set to their smallest allowed input values of 0.01 and 3 Å, respectively, basically switching of the van der Waals interactions. Only points of the eMIFs that show favorable interactions (<0 kcal/mol) were kept for further calculations. Obviously, the points of favorable interactions for the positive probe are points of unfavorable interactions for the negative probe and vice

versa. Thus, the final eMIFs are represented by a grid, of which the points need to have a negative energy (favorable interaction) and furthermore have to be a subset of the previously selected grid points of the H-probe (proximity to the ligand, no overlap with the protease itself).

The proteases and their eMIFs were realigned slightly by overlaying the weighted center of grid points of the H-probe MIFs and aligning the first eigenvector<sup>71</sup> of the H-probe MIFs tensor of inertia, ie, the one corresponding to the largest eigenvalue. Due to the high similarity of the binding clefts, this procedure resulted in an excellent alignment. Thus, the second and third eigenvalues were not used for further refinement, avoiding problems due to their near degeneration. The resulting alignment was consequently used for analyses with the electrostatic GRID probes.

The overlap of the eMIFs corresponding to the same charge was calculated using spherical Gaussian functions with a  $\sigma$  of 2 Å centered at the grid points, according to Equation 3. We tested several options for the width of the Gaussian function, ranging from 1 to 2 Å but found surprisingly little impact on the resulting correlation with experimental substrate data, which is in line with the long-range nature of electrostatic interactions, the results of these calculations are summarized in the supporting info.

$$O = N_A N_B \cdot \left( \frac{\pi \sigma_A^2 \sigma_B^2}{2(\sigma_A^2 + \sigma_B^2)} \right)^{\frac{3}{2}} \cdot \exp \left( -\frac{(\mathbf{r}_A - \mathbf{r}_B)^2}{2(\sigma_A^2 + \sigma_B^2)} \right) \quad (3)$$

In Equation 3,  $N_{A,B}$  is the height of the spherical Gaussian function,  $\sigma_{A,B}^2$  is the variance of the spherical Gaussian function, and  $\mathbf{r}_{A,B}$  is the center of the spherical Gaussian function, for A and B, respectively.

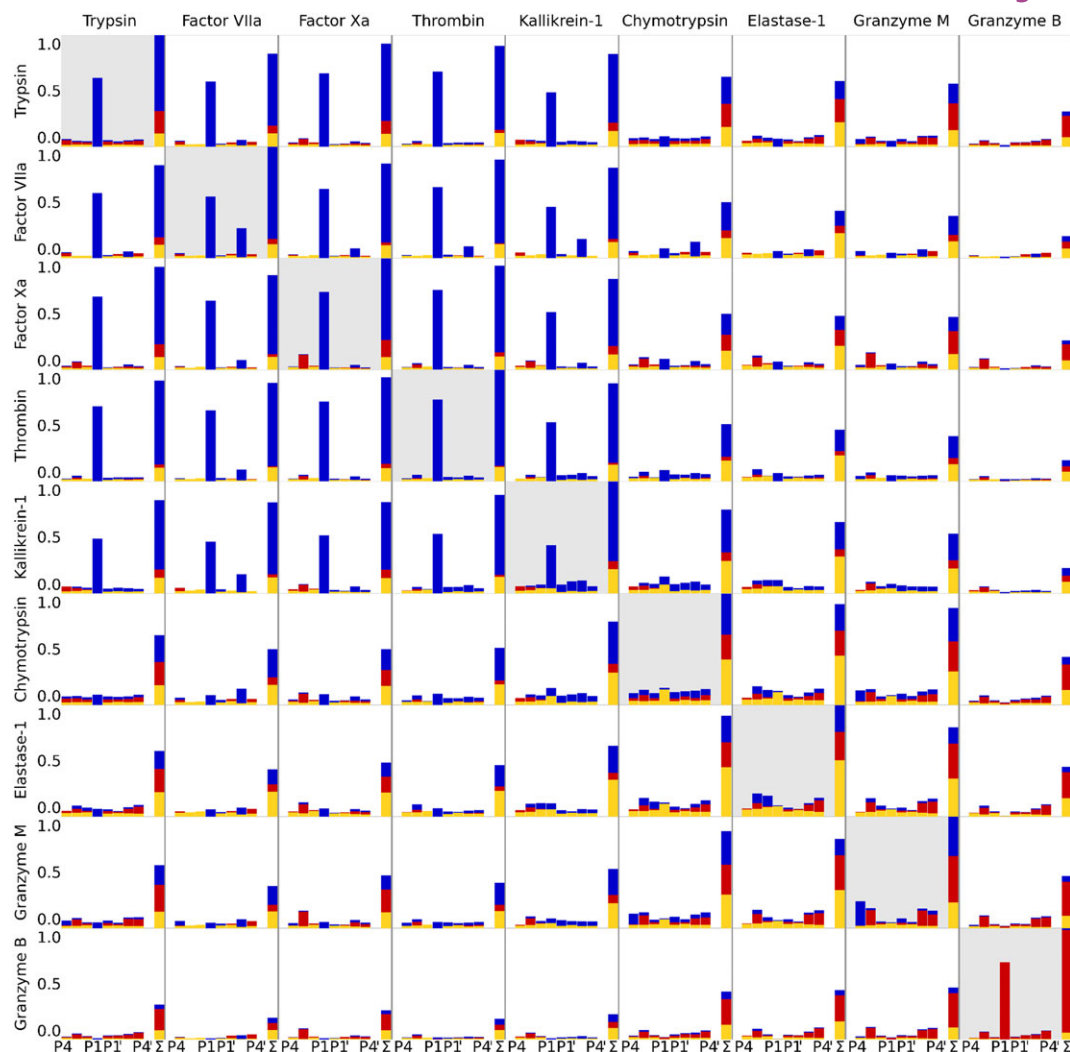
## 3 | RESULTS

### 3.1 | Electrostatic substrate similarities

Electrostatic substrate similarities show whether 2 proteases cleave similar substrates or whether they have opposing substrate preferences. The substrates are only characterized by their charge—positive, neutral, or negative. The shape of the amino acids is neglected, so, eg, the recognition of glutamate and aspartate is considered to be exactly equivalent. The electrostatic substrate similarities can be broken down to the contributions of the single substrate positions that sum up to the total value. The contributions to self-similarities reflect the electrostatic substrate preference of single proteases and are normalized to give a total of 1.

Figure 4 gives an overview of the similarities between 2 proteases in the substrate space after binning the amino acids, as described earlier. Our metric detects prominent similarities among the proteases with high preference for positively charged residues in the S1 subpocket, ie, Trypsin, Factor VIIa, Factor Xa, Thrombin, and Kallikrein-1. The same holds true for the proteases reading primarily neutral amino acids, ie, Chymotrypsin, Elastase 1, and Granzyme M. Granzyme B, the only protease of our set that predominantly reads negatively charged amino acids, is singled out by our metric and shows low similarities to all other proteases.

Electrostatic substrate preferences are not only highlighted for the S1 subpocket, but also for all other subpockets. For example, the



**FIGURE 4** The heights of the bars indicate the electrostatic substrate similarities between all 9 investigated proteases, ranging from P4 to P4', and the resulting sum ( $\Sigma$ ) on the right. Blue represents favoring of positively charged amino acids, yellow neutral ones, and red negatively charged ones. The self-similarities are depicted as diagonal entries and placed on a gray background in the symmetric matrix

electrostatic substrate preferences of Granzyme M reveal a propensity for negatively charged residues in the subpockets S3, S3', and S4' that is hardly identifiable in the cleavage site sequence logos (Figure 3). Among the proteases that prefer positively charged amino acids in S1, Factor VIIa and Kallikrein-1 are quite different in the substrate preferences in that subpocket, yet they share a preference for positively charged substrate amino acids in S3'. Granzyme B, favoring negatively charged substrate residues over large parts of its binding site, shows generally only minimal similarity with the proteases that prefer positive residues in the S1 subpocket. Still within these, the largest electrostatic substrate similarity with Granzyme B is determined for Trypsin. While Trypsin is very specific for positive amino acids in S1, in remote subpockets, it accepts negatively charged residues. This peculiarity is highlighted when compared with Granzyme B.

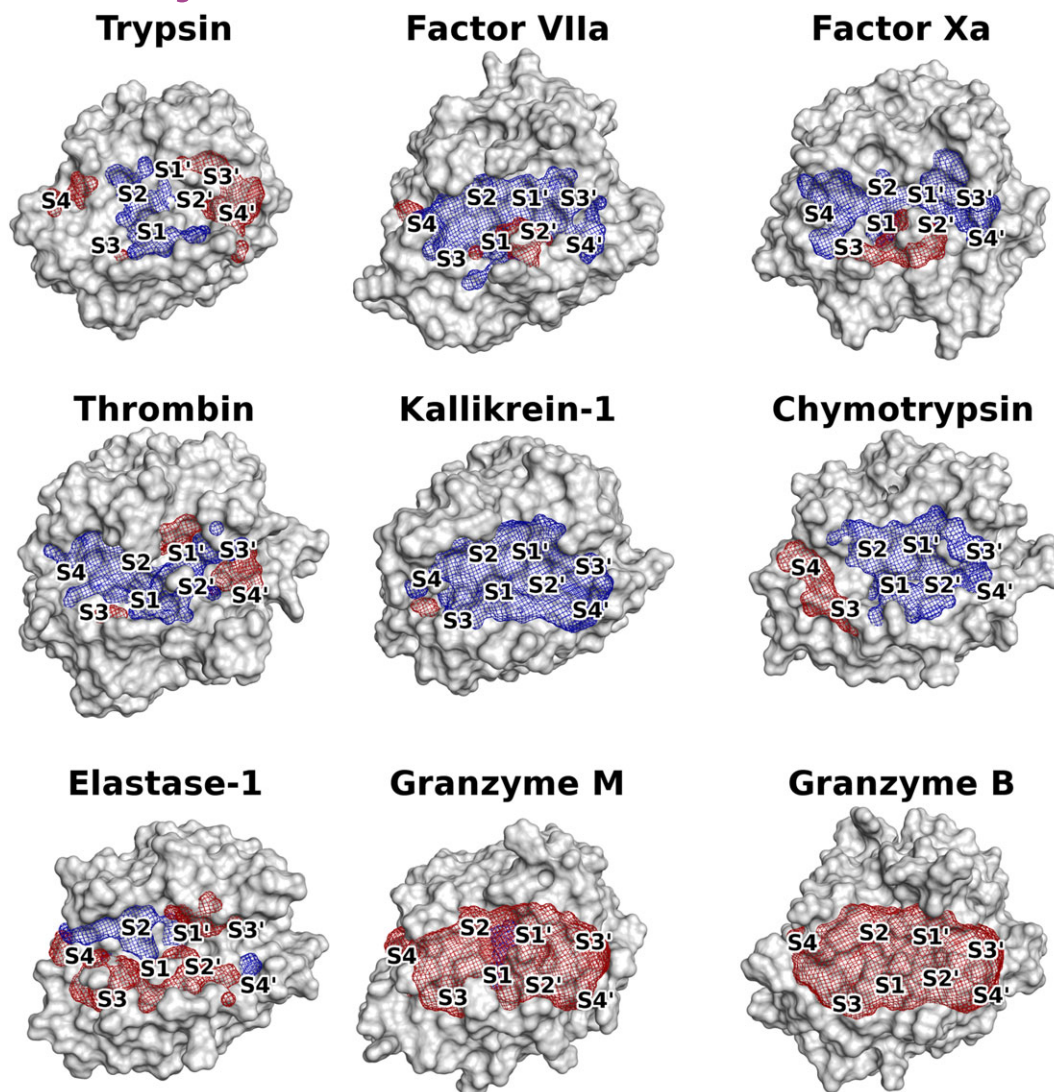
### 3.2 | Electrostatic molecular interaction fields (eMIFs)

With the negatively and the positively charged GRID probes, the eMIFs of the proteases can be determined (Figure 5).

Trypsin shows favorable interactions with the positive probe in its S1 subpocket, while in the more peripheral S4 and S4' subpockets, it prefers the negative probe. Factor VIIa, Factor Xa, Thrombin, and Kallikrein-1 favor the positive probe in large parts of their binding clefts. Chymotrypsin has a dyadic eMIF as in the prime site it interacts favorably with the positive probe and towards the outer non-prime site with the negative probe. In the S2 and the S4' subpocket, Elastase-1 favors the positive probe, while in the rest of the prime site and in S3 it favors the negative probe. Both Granzymes seem to show a completely negative eMIF, but on closer inspection Granzyme M favors the positive probe in S1', the corresponding eMIF is hidden below a layer of grid points favoring interactions with the negative probe.

### 3.3 | Electrostatic substrate preferences and electrostatic molecular interaction fields

A joint view of electrostatic substrate preferences and eMIFs of a protease demonstrates the presence of similar patterns in both metrics. In subpockets that are associated with a substrate readout preference for positively charged residues, strong interactions with



**FIGURE 5** The electrostatic molecular interaction fields (eMIFs) are shown for the 9 investigated proteases. The interactions with the positive probe are depicted in blue, whereas the interactions with the negative probe are depicted in red. A cutoff of  $-3$  kcal/mol was used for the visualization of the fields

the positive probe are present, while in subpockets with substrate readout preference for negatively charged amino acid residues the negative interactions predominate. Commonly, where no strong electrostatic substrate preferences are detected, no strong electrostatic interactions can be found.

The eMIFs of Granzyme B show the same pattern as the electrostatic substrate preference (Figure 6), ie, favorable interactions with the negative probe are visible over the entire binding site. In the electrostatic substrate preference, the same trend can be observed.

For Kallikrein-1, primarily favorable interactions with the positive probe are visible. Substantial areas where the negative probe is favored can be found only in the outer regions of the non-prime site, ie, the S3 subpocket. Again, this is very well reflected by the electrostatic substrate preference, as in the S3 and S4 subpockets a propensity for negatively charged amino acids can be seen.

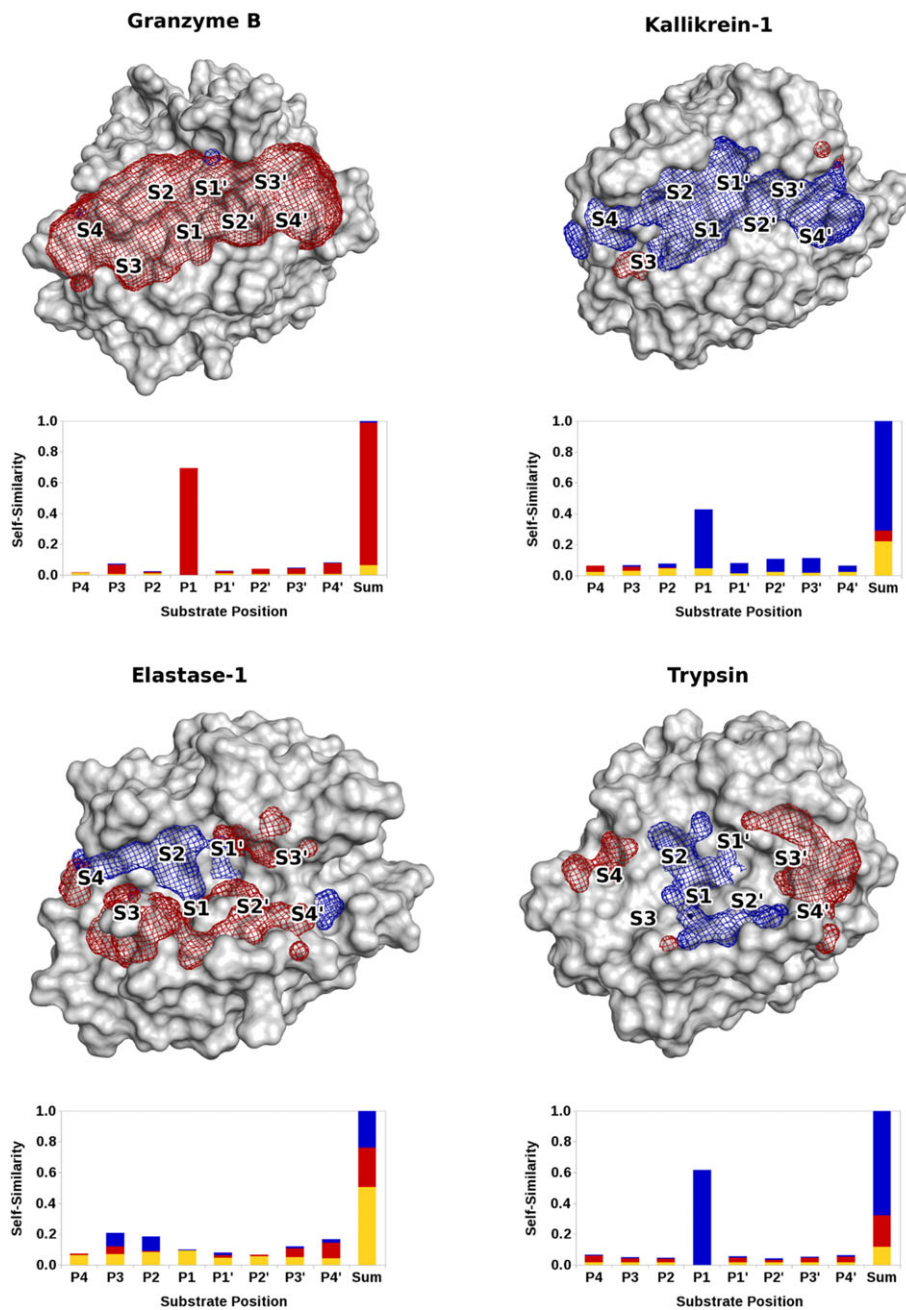
For Elastase-1, the eMIFs reflect the electrostatic substrate preference very well. As the S1 subpocket is specific for neutral amino acids, practically no electrostatic interactions are visible. The prime site substrate self-similarities show preferences for the negative

probe, whereas the non-prime site varies more in electrostatic preferences. The eMIFs correspond very well with this. In the S3 subpocket, which is to be rather unspecific in terms of electrostatics, both interactions can be observed, although the positive eMIF at that position is barely visible because it is hidden behind the negative one.

In the S1 subpocket of Trypsin, positively charged amino acids are strongly favored, which is also visible in the eMIF. On the periphery of the binding site, the protease starts favoring negatively charged amino acids, which is also mirrored very well by the calculated eMIFs, where the negative eMIF starts to dominate around the S3' and S4' subpockets.

### 3.4 | Electrostatic substrate similarities and eMIF overlaps

Calculation of the overlap between the eMIFs of different proteases establishes areas where the same interactions prevail and facilitates the comparison with the electrostatic substrate similarities. Some representative examples of this comparison between eMIF overlap and electrostatic substrate similarities are shown in Figure 7.



**FIGURE 6** The eMIFs of Granzyme B, Kallikrein-1 (top), Elastase-1 and trypsin (bottom) for the positive (blue) and the negative probe (red), using a cutoff of  $-3$  kcal/mol and the electrostatic substrate preference for positive amino acids (blue), neutral amino acids (yellow), and negative amino acids (red). Substrate self-similarities are in general well reflected by the eMIFs

Considering Chymotrypsin and Kallikrein-1 (Figure 7), the differences in substrate similarity can easily be explained by the eMIFs. Both proteases favor positively charged and neutral amino acids in nearly the entire binding cleft. However, on the non-prime site, Kallikrein-1 favors negative amino acids in the S3 subpocket. This hotspot is also visible in the overlap eMIF of the 2 proteases. On the entire prime site, the eMIF overlap shows favorable positive interactions, as does the substrate similarity.

Thrombin and Factor Xa are the most similar ones in their electrostatic substrate readout among our set of proteases. Their eMIFs vary only very little (Figure 7). There are small differences in the subpockets on the outer regions of the prime site as well as farther

away on the non-prime site. But near the S1 subpocket, the differences of the eMIFs are negligible.

An overview of the similarities of the 9 proteases in positive substrate readout and positive eMIF overlap is given in the upper part of Figure 8. The lower part of Figure 8 is the corresponding equivalent showing the negative part of the substrate similarities and the eMIF overlaps from the negative GRID probe. The matrices of the substrate similarities match the electrostatic parts of the total substrate similarities already shown in Figure 4.

At first glance, proteases that are similar in electrostatic substrate readout are also similar in their eMIF similarity. To quantify the correlation of the similarity in substrate space and the eMIF similarity,





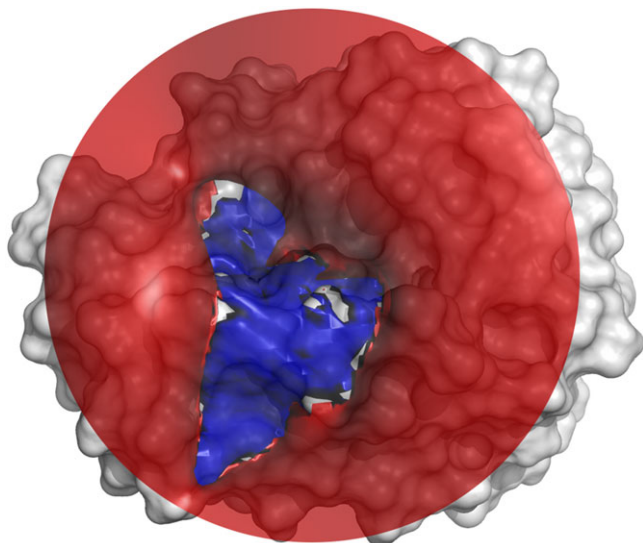
	Trypsin	Factor VIIa	Factor Xa	Thrombin	Kallikrein-1	Chymotrypsin	Elastase-1	Granzyme M	Granzyme B
<b>Substrate Readout positive</b>									
Trypsin	0.680	0.647	0.693	0.753	0.615	0.240	0.162	0.175	0.038
Factor VIIa	0.647	0.827	0.707	0.751	0.640	0.252	0.135	0.169	0.047
Factor Xa	0.693	0.707	0.734	0.776	0.602	0.188	0.124	0.125	0.031
Thrombin	0.753	0.751	0.776	0.860	0.712	0.292	0.190	0.198	0.052
Kallikrein-1	0.615	0.640	0.602	0.712	0.710	0.379	0.244	0.233	0.067
Chymotrypsin	0.240	0.252	0.188	0.292	0.379	0.368	0.238	0.301	0.066
Elastase-1	0.162	0.135	0.124	0.190	0.244	0.238	0.238	0.144	0.046
Granzyme M	0.175	0.169	0.125	0.198	0.233	0.301	0.144	0.352	0.049
Granzyme B	0.038	0.047	0.031	0.052	0.067	0.066	0.046	0.049	0.013
<b>eMIF Overlap positive</b>									
Trypsin	6.14	6.19	6.00	6.16	6.23	5.39	5.23	5.31	3.69
Factor VIIa	6.19	6.46	6.19	6.23	6.50	5.76	5.68	5.48	4.17
Factor Xa	6.00	6.19	6.37	6.20	6.43	5.75	5.22	5.31	4.12
Thrombin	6.16	6.23	6.20	6.34	6.36	5.53	5.17	5.39	3.74
Kallikrein-1	6.23	6.50	6.43	6.36	6.95	6.10	5.78	5.63	4.57
Chymotrypsin	5.39	5.76	5.75	5.53	6.10	5.95	5.16	4.97	3.56
Elastase-1	5.23	5.68	5.22	5.17	5.78	5.16	5.59	4.49	3.79
Granzyme M	5.31	5.48	5.31	5.39	5.63	4.97	4.49	5.36	3.57
Granzyme B	3.69	4.17	4.12	3.74	4.57	3.56	3.79	3.57	4.74
<b>Substrate Readout negative</b>									
Trypsin	0.202	0.068	0.117	0.028	0.075	0.207	0.207	0.242	0.191
Factor VIIa	0.068	0.049	0.024	0.008	0.027	0.069	0.065	0.057	0.062
Factor Xa	0.117	0.024	0.152	0.038	0.075	0.143	0.144	0.205	0.147
Thrombin	0.028	0.008	0.038	0.010	0.017	0.036	0.039	0.054	0.046
Kallikrein-1	0.075	0.027	0.075	0.017	0.071	0.077	0.061	0.075	0.055
Chymotrypsin	0.207	0.069	0.143	0.036	0.077	0.223	0.223	0.269	0.230
Elastase-1	0.207	0.065	0.144	0.039	0.061	0.223	0.255	0.315	0.235
Granzyme M	0.242	0.057	0.205	0.054	0.075	0.269	0.315	0.416	0.307
Granzyme B	0.191	0.062	0.147	0.046	0.055	0.230	0.235	0.307	0.924
<b>eMIF Overlap negative</b>									
Trypsin	5.68	4.92	4.51	5.26	3.70	5.22	5.51	6.04	6.24
Factor VIIa	4.92	5.03	4.54	3.90	1.39	4.82	5.10	5.59	5.75
Factor Xa	4.51	4.54	5.04	3.86	1.56	4.35	5.16	5.46	5.71
Thrombin	5.26	3.90	3.86	5.65	1.95	4.26	4.99	5.69	5.94
Kallikrein-1	3.70	1.39	1.56	1.95	3.95	3.73	4.45	4.67	4.51
Chymotrypsin	5.22	4.82	4.35	4.26	3.73	5.70	5.38	5.83	5.80
Elastase-1	5.51	5.10	5.16	4.99	4.45	5.38	6.01	6.25	6.43
Granzyme M	6.04	5.59	5.46	5.69	4.67	5.83	6.25	6.84	6.88
Granzyme B	6.24	5.75	5.71	5.94	4.51	5.80	6.43	6.88	7.35

**FIGURE 8** Substrate similarity and eMIF overlap for positive (top) and negative (bottom) substrate space and probe are compared. High similarity and overlap are shaded in dark blue (positive) and dark red (negative), respectively, while low similarity and overlap are depicted on a white background (both). For readability purposes, the eMIF overlap was scaled logarithmically

apportioning the binding cleft into subpockets. Still, for an efficient substrate prediction, which however is beyond the scope of this study, such a partitioning of electrostatic contributions to subpockets of the binding cleft would be highly desirable.

This is in line with the notion that protein-protein recognition follows a 2-step mechanism. Firstly, an initial encounter complex forms when enzyme and substrate meet. The association rates for this initial encounter complex are largely governed by electrostatics.<sup>80</sup> An energy funnel pulls substrate and enzyme together and directs the substrate towards the binding site<sup>81-84</sup> (Figure 9). In a second step, conformational changes lead to the formation of a compatible binding interface.<sup>85,86</sup> Here, shape complementarity and flexibility are crucial to enable weak van der Waals interactions and to avoid clashes.

Electrostatics and shape complementarity in context of substrate recognition can be considered rather orthogonal properties resulting in different aspects of substrate recognition.<sup>87</sup> Thus, we believe that these 2 aspects of substrate recognition can be separated very efficiently by our knowledge-based approach to analyze substrate readout data as presented in this study. Electrostatic substrate preferences can be characterized very well by binning substrate residues according to their charge. On the other hand, we expect that shape complementarity can be characterized by analyzing substrate recognition within the 3 bins, especially within the neutral bin comprising 15 different neutral amino acid residues. If we can describe the contributions of electrostatics and shape complementarity in a solely physics-based way, it will be possible to predict the localized



**FIGURE 9** The binding interface of trypsin depicted with the eMIFs of the positive (blue) and negative (red) probe. An energy cutoff of  $-0.5$  kcal/mol was used for the visualization of far-reaching electrostatic interactions. The eMIF forms a funnel-like long-range interaction profile that presumably guides substrates towards an initial encounter complex

specificity and promiscuity of proteases and most likely also of other biomolecular interfaces.

## 5 | CONCLUSIONS

A knowledge-based approach to characterize differences in electrostatic substrate preferences is introduced and applied on 9 homologous serine proteases of the chymotrypsin family. The approach bins known substrate residues into positively charged, negatively charged, and neutral amino acids. Thus, electrostatic preferences in substrate recognition are quantified within subpockets of the binding cleft of the 9 serine proteases and can be compared between different proteases. Similarities and differences in electrostatic preferences can easily be identified on a localized subpocket level but also globally for the complete binding cleft.

On the other hand, eMIFs are calculated in a physics-based way studying X-ray structures using the program GRID in combination with user-defined probes that focus on electrostatics. The binding cleft within the X-ray structures is delimited by a proximity criterion to known ligands. Calculating the overlap between eMIFs results again in similarities and differences in electrostatic preferences.

Comparing the knowledge-based and physics-based similarities and differences in electrostatic preferences, a high correlation between the 2 totally different approaches is found. This implies that the electrostatic part of substrate recognition and substrate specificity can be explained very well by eMIFs.

Due to the long-range nature of electrostatics, we assume that these electrostatic molecular interactions fields determine the formation of an initial encounter complex between substrates and proteases.

## ACKNOWLEDGEMENTS

This work was supported by the Austrian Science Fund (FWF) [project P 30565]. B.J.W. is a recipient of the Austrian Academy of Sciences DOC Grant.

## ORCID

Klaus R. Liedl  <http://orcid.org/0000-0002-0985-2299>

## REFERENCES

- Puente XS, Sánchez LM, Gutiérrez-Fernández A, Velasco G, López-Otín C. A genomic view of the complexity of mammalian proteolytic systems. *Biochem Soc T*. 2005;33(2):331-334. <https://doi.org/10.1042/BST0330331>
- Ezkurdia I, Juan D, Rodríguez JM, et al. Multiple evidence strands suggest that there may be as few as 19000 human protein-coding genes. *Hum Mol Genet*. 2014;23(22):5866-5878. <https://doi.org/10.1093/hmg/ddu309>
- Southan C. A genomic perspective on human proteases as drug targets. *Drug Discov Today*. 2001;6(13):681-688. [https://doi.org/10.1016/S1359-6446\(01\)01793-7](https://doi.org/10.1016/S1359-6446(01)01793-7)
- Straub A, Roehrig S, Hillisch A. Oral, direct thrombin and factor Xa inhibitors: the replacement for warfarin, leeches, and pig intestines? *Angew Chem Int Edit*. 2011;50(20):4574-4590. <https://doi.org/10.1002/anie.201004575>
- Zaman MA, Oparil S, Calhoun DA. Drugs targeting the renin-angiotensin-aldosterone system. *Nat Rev Drug Discov*. 2002;1(8):621-636. <https://doi.org/10.1038/nrd873>
- Lu Z, Chu Y, Wang Y. HIV protease inhibitors: a review of molecular selectivity and toxicity. *HIV AIDS (Auckl)*. 2015;7:95-104. <https://doi.org/10.2147/HIV.S79956>
- De Clercq E. Development of antiviral drugs for the treatment of hepatitis C at an accelerating pace. *Rev Med Virol*. 2015;25(4):254-267. <https://doi.org/10.1002/rmv.1842>
- Turk B. Targeting proteases: successes, failures and future prospects. *Nat Rev Drug Discov*. 2006;5(9):785-799. <https://doi.org/10.1038/nrd2092>
- Drag M, Salvesen GS. Emerging principles in protease-based drug discovery. *Nat Rev Drug Discov*. 2010;9(9):690-701. <https://doi.org/10.1038/nrd3053>
- Radzicka A, Wolfenden R. Rates of uncatalyzed peptide bond hydrolysis in neutral solution and the transition state affinities of proteases. *J Am Chem Soc*. 1996;118(26):6105-6109. <https://doi.org/10.1021/ja954077c>
- Tyndall JDA, Nall T, Fairlie DP. Proteases universally recognize Beta strands in their active sites. *Chem Rev*. 2005;105(3):973-1000. <https://doi.org/10.1021/cr040669e>
- Perona JJ, Craik CS. Structural basis of substrate specificity in the serine proteases. *Protein Sci*. 1995;4(3):337-360. <https://doi.org/10.1002/pro.5560040301>
- Hedstrom L. Serine protease mechanism and specificity. *Chem Rev*. 2002;102(12):4501-4524. <https://doi.org/10.1021/cr000033x>
- Fuchs JE, Grafenstein S, Huber RG, et al. Cleavage entropy as quantitative measure of protease specificity. *PLoS Comput Biol*. 2013;9(4):e1003007. <https://doi.org/10.1371/journal.pcbi.1003007>
- Hedstrom L. Introduction: proteases. *Chem Rev*. 2002;102(12):4429-4430. <https://doi.org/10.1021/cr0101712>
- Perona JJ, Craik CS. Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold. *J Biol Chem*. 1997;272(48):29987-29990. <https://doi.org/10.1074/jbc.272.48.29987>
- Schechter I, Berger A. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Co*. 1967;27(2):157-162. [https://doi.org/10.1016/S0006-291X\(67\)80055-X](https://doi.org/10.1016/S0006-291X(67)80055-X)

18. López-Otín C, Overall CM. Protease degradomics: a new challenge for proteomics. *Nat Rev Mol Cell Biol.* 2002;3(7):509-519. <https://doi.org/10.1038/nrm858>
19. Salisbury CM, Maly DJ, Ellman JA. Peptide microarrays for the determination of protease substrate specificity. *J Am Chem Soc.* 2002;124(50):14868-14870. <https://doi.org/10.1021/ja027477q>
20. Schilling O, Overall CM. Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat Biotechnol.* 2008;26(6):685-694. <https://doi.org/10.1038/nbt1408>
21. Lange PF, Overall CM. Protein TAILS: when termini tell tales of proteolysis and function. *Curr Opin Chem Biol.* 2013;17(1):73-82. <https://doi.org/10.1016/j.cbpa.2012.11.025>
22. Schauerl M, Fuchs JE, Waldner BJ, Huber RG, Kramer C, Liedl KR. Characterizing protease specificity: how many substrates do we need? *PLOS one.* 2015;10(11):e0142658. <https://doi.org/10.1371/journal.pone.0142658>
23. Rawlings ND, Barrett AJ, Finn R. Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* 2016;44(D1):D343-D350. <https://doi.org/10.1093/nar/gkv1118>
24. Fuchs JE, Grafenstein S, Huber RG, Kramer C, Liedl KR. Substrate-driven mapping of the Degradome by comparison of sequence logos. *PLoS Comput Biol.* 2013;9(11):e1003353. <https://doi.org/10.1371/journal.pcbi.1003353>
25. Wallnoefer HG, Lingott T, Gutiérrez JM, Merfort I, Liedl KR. Backbone flexibility controls the activity and specificity of a protein-protein interface: specificity in snake venom metalloproteases. *J Am Chem Soc.* 2010;132(30):10330-10337. <https://doi.org/10.1021/ja909908y>
26. Fuchs JE, von Grafenstein S, Huber RG, Wallnoefer HG, Liedl KR. Specificity of a protein-protein interface: local dynamics direct substrate recognition of effector caspases. *Proteins.* 2014;82(4):546-555. <https://doi.org/10.1002/prot.24417>
27. Fuchs JE, Huber RG, Waldner BJ, et al. Dynamics govern specificity of a protein-protein interface: substrate recognition by thrombin. *PLOS ONE.* 2015;10(10):e0140713. <https://doi.org/10.1371/journal.pone.0140713>
28. Honig B, Nicholls A. Classical electrostatics in biology and chemistry. *Science.* 1995;268(5214):1144-1149. <https://doi.org/10.1126/science.7761829>
29. Sheinerman FB, Norel R, Honig B. Electrostatic aspects of protein-protein interactions. *Curr Opin Struct Biol.* 2000;10(2):153-159. [https://doi.org/10.1016/S0959-440X\(00\)00065-8](https://doi.org/10.1016/S0959-440X(00)00065-8)
30. Warshel A, Sharma PK, Kato M, Parson WW. Modeling electrostatic effects in proteins. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics.* 2006;1764(11):1647-1676. <https://doi.org/10.1016/j.bbapap.2006.08.007>
31. Schutz CN, Warshel A. What are the dielectric "constants" of proteins and how to validate electrostatic models? *Proteins.* 2001;44(4):400-417. <https://doi.org/10.1002/prot.1106>
32. Rasmussen TD, Ren P, Ponder JW, Jensen F. Force field modeling of conformational energies: importance of multipole moments and intramolecular polarization. *Int. J. Quantum Chem.* 2007;107(6):1390-1395. <https://doi.org/10.1002/qua.21278>
33. Waldner BJ, Fuchs JE, Huber RG, et al. Quantitative correlation of conformational binding enthalpy with substrate specificity of serine proteases. *J Phys Chem B.* 2016;120(2):299-308. <https://doi.org/10.1021/acs.jpcc.5b10637>
34. Pethe MA, Rubenstein AB, Khare SD. Large-scale structure-based prediction and identification of novel protease substrates using computational protein design. *J Mol Biol.* 2017;429(2):220-236. <https://doi.org/10.1016/j.jmb.2016.11.031>
35. Okun RY, Chen BY. A statistical model of electrostatic isopotential variation in serine protease binding cavities 2015. *IEEE Int Conf Bioinforma and Biomed (BIBM).* 2015;1246-1252. <https://doi.org/10.1109/BIBM.2015.7359859>
36. Wade RC, Gabdoulline RR, Rienzo FD. Protein interaction property similarity analysis. *Int J Quantum Chem.* 2001;83(3-4):122-127. <https://doi.org/10.1002/qua.1204>
37. Henrich S, Feierberg I, Wang T, Blomberg N, Wade RC. Comparative binding energy analysis for binding affinity and target selectivity prediction. *Proteins.* 2010;78(1):135-153. <https://doi.org/10.1002/prot.22579>
38. Henrich S, Salo-Ahen OMH, Huang B, Rippmann FF, Cruciani G, Wade RC. Computational approaches to identifying and characterizing protein binding sites for ligand design. *J Mol Recognit.* 2010;23(2):209-219. <https://doi.org/10.1002/jmr.984>
39. Ehrh C, Brinkjost T, Koch O. Impact of binding site comparisons on medicinal chemistry and rational molecular design. *J Med Chem.* 2016;59(9):4121-4151. <https://doi.org/10.1021/acs.jmedchem.6b00078>
40. Siragusa L, Cross S, Baroni M, Goracci L, Cruciani G. BioGPS: navigating biological space to predict polypharmacology, off-targeting, and selectivity. *Proteins.* 2015;83(3):517-532. <https://doi.org/10.1002/prot.24753>
41. Wu C, Jin X, Tsueng G, Afrasiabi C, Su AI. BioGPS: building your own mash-up of gene annotations and expression profiles. *Nucleic Acids Res.* 2016;44(D1):D313-D316. <https://doi.org/10.1093/nar/gkv1104>
42. Chartier M, Najmanovich R. Detection of binding site molecular interaction field similarities. *J Chem Inf Model.* 2015;55(8):1600-1615. <https://doi.org/10.1021/acs.jcim.5b00333>
43. Binkowski TA, Joachimiak A. Protein functional surfaces: global shape matching and local spatial alignments of ligand binding sites. *BMC Struct Biol.* 2008;8(1):45. <https://doi.org/10.1186/1472-6807-8-45>
44. Najmanovich R, Kurbatova N, Thornton J. Detection of 3D atomic similarities and their use in the discrimination of small molecule protein-binding sites. *Bioinformatics.* 2008;24(16):i105-i111. <https://doi.org/10.1093/bioinformatics/btn263>
45. Yeturu K, Chandra N. PocketMatch: a new algorithm to compare binding sites in protein structures. *BMC Bioinformatics.* 2008;9(1):543. <https://doi.org/10.1186/1471-2105-9-543>
46. Schmitt S, Kuhn D, Klebe G. A new method to detect related function among proteins independent of sequence and fold homology. *J Mol Biol.* 2002;323(2):387-406. [https://doi.org/10.1016/S0022-2836\(02\)00811-2](https://doi.org/10.1016/S0022-2836(02)00811-2)
47. Glinca S, Klebe G. Cavities tell more than sequences: exploring functional relationships of proteases via binding pockets. *J Chem Inf Model.* 2013;53(8):2082-2092. <https://doi.org/10.1021/ci300550a>
48. Krotzky T, Fober T, Hüllermeier E, Klebe G. Extended graph-based models for enhanced similarity search in Cavbase. *IEEE/ACM TCBB.* 2014;11(5):878-890. <https://doi.org/10.1109/TCBB.2014.2325020>
49. Leinweber M, Fober T, Strickert M, et al. CavSimBase: a database for large scale comparison of protein binding sites. *IEEE T Knowl Data En.* 2016;28(6):1423-1434. <https://doi.org/10.1109/TKDE.2016.2520484>
50. Shulman-Peleg A, Nussinov R, Wolfson HJ. SiteEngines: recognition and comparison of binding sites and protein-protein interfaces. *Nucleic Acids Res.* 2005;33(Web Server):W337-W341. <https://doi.org/10.1093/nar/gki482>
51. Powers R, Copeland JC, Stark JL, Caprez A, Guru A, Swanson D. Searching the protein structure database for ligand-binding site similarities using CPASS v.2. *BMC Res Notes.* 2011;4(1):17. <https://doi.org/10.1186/1756-0500-4-17>
52. Jambon M, Andrieu O, Combet C, Deléage G, Delfaud F, Geourjon C. The SuMo server: 3D search for protein functional sites. *Bioinformatics.* 2005;21(20):3929-3930. <https://doi.org/10.1093/bioinformatics/bti645>
53. Kinoshita K, Nakamura H. Identification of protein biochemical functions by similarity search using the molecular surface database eF-site. *Protein Sci.* 2003;12(8):1589-1595. <https://doi.org/10.1110/ps.0368703>
54. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. *Genome Res.* 2004;14(6):1188-1190. <https://doi.org/10.1101/gr.849004>

55. Kozłowski LP. Proteome-pl: proteome isoelectric point database. *Nucleic Acids Res.* 2017;45(D1):D1112-D1116. <https://doi.org/10.1093/nar/gkw978>
56. Schmidt A, Jelsch C, Østergaard P, Rypniewski W, Lamzin VS. Trypsin revisited crystallography at (Sub) atomic resolution and quantum chemistry revealing details of catalysis. *J Biol Chem.* 2003;278(44):43357-43362. <https://doi.org/10.1074/jbc.M306944200>
57. Sichler K, Banner DW, D'Arcy A, et al. Crystal structures of uninhibited factor VIIa link its cofactor and substrate-assisted activation to specific interactions. *J Mol Biol.* 2002;322(3):591-603. [https://doi.org/10.1016/S0022-2836\(02\)00747-7](https://doi.org/10.1016/S0022-2836(02)00747-7)
58. Katz BA, Mackman R, Luong C, et al. Structural basis for selectivity of a small molecule, S1-binding, submicromolar inhibitor of urokinase-type plasminogen activator. *Chem Biol.* 2000;7(4):299-312. [https://doi.org/10.1016/S1074-5521\(00\)00104-6](https://doi.org/10.1016/S1074-5521(00)00104-6)
59. Hilpert K, Ackermann J, Banner DW, et al. Design and synthesis of potent and highly selective thrombin inhibitors. *J Med Chem.* 1994;37(23):3889-3901. <https://doi.org/10.1021/jm00049a008>
60. Laxmikanthan G, Blaber SI, Bennett MJ, Scarisbrick IA, Juliano MA, Blaber M. 1.70 Å X-ray structure of human apo kallikrein 1: structural changes upon peptide inhibitor/substrate binding. *Proteins.* 2005;58(4):802-814. <https://doi.org/10.1002/prot.20368>
61. Tsukada H, Blow DM. Structure of alpha-chymotrypsin refined at 1.68 Å resolution. *J Mol Biol.* 1985;184(4):703-711. [https://doi.org/10.1016/0022-2836\(85\)90314-6](https://doi.org/10.1016/0022-2836(85)90314-6)
62. Würtele M, Hahn M, Hilpert K, Höhne W. Atomic resolution structure of native porcine pancreatic elastase at 1.1 Å. *Acta Crystallogr D Biol Crystallogr.* 2000;56(4):520-523. <https://doi.org/10.1107/S0907444900000299>
63. Wu L, Wang L, Hua G, et al. Structural basis for proteolytic specificity of the human apoptosis-inducing Granzyme M. *J Immunol.* 2009;183(1):421-429. <https://doi.org/10.4049/jimmunol.0803088>
64. Estébanez-Perpi á E, Fuentes-Prior P, Belorgey D, et al. Crystal structure of the caspase activator human Granzyme B, a proteinase highly specific for an Asp-P1 residue. *Biol Chem.* 2005;381(12):1203-1214. <https://doi.org/10.1515/BC.2000.148>
65. Huntington JA, Esmo CT. The molecular basis of thrombin allostery revealed by a 1.8 Å structure of the "slow" form. *Structure.* 2003;11(4):469-479. [https://doi.org/10.1016/S0969-2126\(03\)00049-2](https://doi.org/10.1016/S0969-2126(03)00049-2)
66. Pozzi N, Vogt AD, Gohara DW, Di Cera E. Conformational selection in trypsin-like proteases. *Curr Opin Struct Biol.* 2012;22(4):421-431. <https://doi.org/10.1016/j.sbi.2012.05.006>
67. Labute P. Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins.* 2009;75(1):187-205. <https://doi.org/10.1002/prot.22234>
68. ULC CCG. Molecular Operating Environment (MOE) (2013.08); 2017.
69. Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem.* 1985;28(7):849-857. <https://doi.org/10.1021/jm00145a002>
70. Goodford P. The basic principles of GRID. In: Cruciani G, ed. *Molecular Interaction Fields: Applications in Drug Discovery and ADME Prediction*. Vol. 27. Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA; 2006:3-26.
71. Guennebaud G, Jacob B, Avery P, Bachtch A, Barthelemy S. Eigen v3; 2010; <http://eigen.tuxfamily.org>.
72. Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 1967;27(2):209-220.
73. Oksanen J, Blanchet FG, Friendly M, et al. vegan: community ecology package. R Package (24-2); 2017; <https://CRAN.R-project.org/package=vegan>.
74. Richter S, Wenzel A, Stein M, Gabdoulline RR, Wade RC. webPIPSA: A web server for the comparison of protein interaction properties. *Nucleic Acids Res* 2008;36(Web Server issue):W276-280. <https://doi.org/10.1093/nar/gkn181>. (<http://pipsa.h-its.org/pipsa/>)
75. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences* 2001;98(18):10037-10041. <https://doi.org/10.1073/pnas.181342398>
76. Huntington JA. Molecular recognition mechanisms of thrombin. *J Thromb Haemost.* 2005;3(8):1861-1872. <https://doi.org/10.1111/j.1538-7836.2005.01363.x>
77. Batra J, Szabo A, Caulfield TR, Soares AS, Sahin-Toth M, Radisky ES. Long-range electrostatic complementarity governs substrate recognition by human chymotrypsin C, a key regulator of digestive enzyme activation. *J Biol Chem.* 2013;288(14):9848-9859. <https://doi.org/10.1074/jbc.M113.457382>
78. Gabb HA, Jackson RM, Sternberg MJE. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J Mol Biol.* 1997;272(1):106-120. <https://doi.org/10.1006/jmbi.1997.1203>
79. Sinha N, Smith-Gill S. Electrostatics in protein binding and function. *Curr Protein Pept Sc.* 2002;3(6):601-614. <https://doi.org/10.2174/1389203023380431>
80. Schreiber G, Haran G, Zhou H-X. Fundamental aspects of protein-protein association kinetics. *Chem Rev.* 2009;109(3):839-860. <https://doi.org/10.1021/cr800373w>
81. Sinha N, Mohan S, Lipschultz CA, Smith-Gill SJ. Differences in electrostatic properties at antibody-antigen binding sites: implications for specificity and cross-reactivity. *Biophys J.* 2002;83(6):2946-2968. [https://doi.org/10.1016/S0006-3495\(02\)75302-2](https://doi.org/10.1016/S0006-3495(02)75302-2)
82. Tworowski D, Feldman AV, Safran MG. Electrostatic potential of aminoacyl-tRNA synthetase navigates tRNA on its pathway to the binding site. *J Mol Biol.* 2005;350(5):866-882. <https://doi.org/10.1016/j.jmb.2005.05.051>
83. Campbell B, Petukh M, Alexov E, Li C. On the electrostatic properties of homodimeric proteins. *J Theor Comput Chem.* 2014;13(3):1440007. <https://doi.org/10.1142/S0219633614400070>
84. Li L, Chakravorty A, Alexov E. DelPhiForce, a tool for electrostatic force calculations: applications to macromolecular binding. *J Comput Chem.* 2017;38(9):584-593. <https://doi.org/10.1002/jcc.24715>
85. Dagliyan O, Proctor Elizabeth A, D'Auria Kevin M, Ding F, Dokholyan Nikolay V. Structural and dynamic determinants of protein-peptide recognition. *Structure.* 2011;19(12):1837-1845. <https://doi.org/10.1016/j.str.2011.09.014>
86. Kozakov D, Li K, Hall DR, et al. Encounter complexes and dimensionality reduction in protein-protein association. *eLife.* 2014;3:e01370. <https://doi.org/10.7554/eLife.01370>
87. Sulea T, Purisima EO. Profiling charge complementarity and selectivity for binding at the protein surface. *Biophys J.* 2003;84(5):2883-2896. [https://doi.org/10.1016/S0006-3495\(03\)70016-2](https://doi.org/10.1016/S0006-3495(03)70016-2)

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Waldner BJ, Kraml J, Kahler U, et al. Electrostatic recognition in substrate binding to serine proteases. *J Mol Recognit.* 2018;31:e2727. <https://doi.org/10.1002/jmr.2727>