Current Protein and Peptide Science, 2017, *18*, 905-913

REVIEW ARTICLE

Determinants of Macromolecular Specificity from Proteomics-Derived Peptide Substrate Data

Julian E. Fuchs[a,*], Oliver Schilling[b] and Klaus R. Liedl[c]

[a]*Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom;* [b]*Institute of Molecular Medicine and Cell Research, University of Freiburg, Stefan-Meier-Str. 17, D-79104 Freiburg, Germany and BIOSS Centre for Biological Signaling Studies, University of Freiburg, D-79104 Freiburg, Germany;* [c]*Institute of General, Inorganic and Theoretical Chemistry, Center for Molecular Biosciences Innsbruck (CMBI), University of Innsbruck, Innrain 80/82, A-6020 Innsbruck, Austria*

ARTICLE HISTORY

Received: February 25, 2017
Revised: March 30, 2017
Accepted: April 15, 2017

DOI:
10.2174/1389203717666160724211231

Abstract: ***Background:*** Recent advances in proteomics methodologies allow for high throughput profiling of proteolytic cleavage events. The resulting substrate peptide distributions provide deep insights in the underlying macromolecular recognition events, as determinants of biomolecular specificity identified by proteomics approaches may be compared to structure-based analysis of corresponding protein-protein interfaces.

Method: Here, we present an overview of experimental and computational methodologies and tools applied in the area and provide an outlook beyond the protein class of proteases.

Results and Conclusion: We discuss here future potential, synergies and needs of the emerging overlap disciplines of proteomics and structure-based modelling.

Keywords: Macromolecular recognition, molecular modelling, peptide binding, protein-protein-interface, protease substrate profiling, specificity.

1. INTRODUCTION

Proteolytic enzymes (proteases, proteinases, peptidases) cleave peptidic bonds in substrate proteins and peptides and fulfil multiple central roles in living organisms [1]. Therefore, specialized proteases with diverse functions and distinct catalytic classes and fold types have evolved [2]. Some enzymes process well-defined substrates at specific sites and are involved in signalling cascades, *e.g.,* the blood clotting cascade [3], the apoptosis pathway [4] or the complement cascade [5]. Other proteases cleave a variety of substrates and are required for digestion of nutrition proteins [6] as well as degradation of extracellular matrix proteins [7]. Therefore, the range of substrates ("degradome") defines the biological function of proteolytic enzymes [8] and turns them into attractive drug targets [9].

Proteases are not only an interesting protein class in terms of their biological functions but also as prototypes of multi-specific protein-protein interfaces [10]. A multitude of protease substrate sequences has been reported in scientific literature [11] and gathered in publicly available online databases (MEROPS [12], CutDB [13], PMAP [14], DegraBase [15], TopFIND [16]). Information content of MEROPS, its access and utilization, also in respect of protease substrate specificity, has recently been reviewed by the curators of the database [17]. Consensus substrate sequences in the P4-P4' amino acid positions [18] flanking the scissile bond of protease substrates are often depicted as heat maps [19], sequence logos [20], or iceLogos [21] (see Fig. **1** for an example sequence logo for the serine protease factor Xa generated with Weblogo [22]).

Recently, the Skylign web server was launched to facilitate generation and interactive manipulation of sequence logos [22]. As of December 2014 MEROPS lists 13,768 substrates for the unspecific serine protease trypsin-1 only, with the vast majority stemming from proteomics-based identification techniques [23, 24]. Several other proteolytic enzymes spanning different catalytic types are characterized with at least 1,000 annotated targets.

These innovative experimental methodologies allow for rapid identification of proteolytic events at the proteome level using mass spectrometry and therefore increasingly broaden the range of available peptide substrate data [25-32]. The gathered amount of substrate data allows for quantification and direct comparison of protease specificity [33]. In combination with structure-based techniques, molecular determinants of macromolecular specificity and promiscuity can be identified and generalized from proteases to general protein-protein interfaces [34]. In the following review, we

*Address correspondence to this author at the Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom; Tel: +44-1223-763854;
Fax: +44-1223-763076; E-mail: Julian.Fuchs@uibk.ac.at

1875-5550/17 $58.00+.00

will outline technologies used on both the experimental and computational side and aim to judge future potential and challenges for this emerging field at the interface of proteomics and structural bioinformatics.

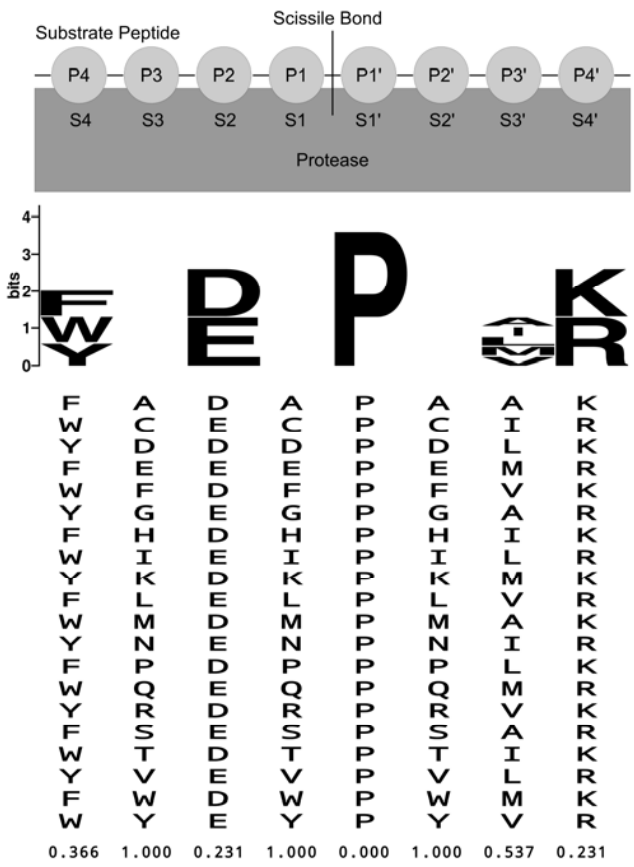

Fig. (1). Protease cleavage site sequence logos: Schematic representation of a protease binding cleft (dark grey) and its subpockets S4-S4' flanking the scissile bond. The substrate peptide P4-P4' is represented as light grey spheres. The specificity pattern for a hypothetical protease is shown as sequence logo and raw sequence data for 20 peptides with corresponding cleavage entropy values S on bottom. The example protease shows a highly complex cleavage pattern: P4 accepts aromatic residues, P2 negatively charged residues, P1' only tolerates proline, whilst S3' prefers hydrophobic and S4' positively charged amino acids. Other pockets S3, S1, and S2' show no substrate readout and thus have constant cleavage entropies of 1.

2. DEGRADOMICS METHODS AND DATA

Several approaches for the specificity profiling of proteases have been established. Importantly, the different strategies have particular advantages and should be considered as being highly complementary. Determination of protease specificity is a fundamental step in their biochemical characterization and provides the basis for the design of specific probes and inhibitors. For yet uncharacterized so-called "novel" proteases, powerful specificity profiling approaches enable rapid de-orphanizing and establishing of robust activity assays. As outlined in the present review, the combination of positional specificity profiles with structural investigations and modern computational techniques are exceptionally powerful in providing a molecular understanding of peptide substrate recognition by proteolytic enzymes. On a basic level, protease specificity can be investigated with a small number of peptidic substrates. This is exemplified by an early study on matrix metalloproteases, in which a set of 16 synthetic octapeptides were used to assess specificity of skin fibroblast collagenase [35]. The sequences of these peptides represent variations of known collagenase cleavage sites in proteins. However, usage of only a few peptidic substrates severely limits coverage of sequences diversity and is intrinsically biased. Phage display is a powerful technique for the profiling of protein-peptide interactions. Phage display has been adopted to identify preferred peptidic substrates for proteolytic enzymes [36]. Randomized, genetically encoded sequences are expressed as protease-sensitive linkers between an affinity domain and a truncated form of the gene III protein. Each phage particle encodes one substrate sequence. Efficient cleavage of the substrate sequence in the linker region removes the affinity domain, thus enabling separation of phage particles with cleavable sequences. Substrate phage display enables extensive coverage of sequence diversity and iterative refinement of protease specificity profiles. The method has been widely adopted. Further developments include bacterial display in combination with fluorescence-activated cell sorting [37] and automated platforms for increased throughput, enabling profiling of entire protease families, such as matrix metalloproteases [38].

While phage display is a genetic approach to generate sequence diversity, complimentary techniques have been developed based on synthetic peptide libraries. Here, three approaches are particularly outstanding: positional scanning synthetic combinatorial libraries (PS-SCLs), peptide nucleic acids (PNA) arrays, and mixture-based oriented peptide libraries. The PS-SCL strategy employs peptide libraries in which one position (e.g. P1) is occupied with a defined residue while the other positions are randomized [39]. The aim is to profile specificity of the defined position without interference of the randomized other positions. Important refinements of the PS-SCL strategy include a more versatile chemical synthesis, allowing for randomization of the P1 residue together with more sensitive fluorescence detection. The typical design of a PS-SCL experiment focuses on either prime or non-prime specificity. However, some proteases such as matrix metalloproteases (MMPs) have a specificity profile that spans across the scissile peptide bond. Turk *et al.* [40] designed a two-step strategy to tackle such cases using synthetic peptide libraries. Firstly, prime-site specificity is profiled by an N-terminally protected degenerate peptide library. Cleavage products possess free N-termini and are analyzed by Edman sequencing, while chemical protection of uncleaved peptides prevents Edman degradation. The mixed signals stemming from Edman sequencing of the library pool are translated into subsite preferences. The preferred prime site motif is used in a second step. Here, the library consists of fixed prime-site and randomized non-prime site residues. The library has free N-termini and C-terminal biotin tags. The fixed prime-site motif serves as an anchor to define the scissile peptide bond. Proteolysis releases the non-prime cleavage product while uncleaved peptides and prime-site cleavage products are captured by biotin affinity chromatography. Non-prime cleavage products are

analyzed by Edman sequencing, yielding positional preferences. The approach has been particularly useful in characterizing metalloproteases [40-42]. PNAs employ peptidic substrates that are tagged to specific nucleic acid sequences. Proteolysis removes a terminal fluorophore. Subsequently, PNAs are hybridized to a complementary, spatially deconvoluting microarray [43]. Lack of fluorescence at a specific position indicates preferential proteolysis of that specific peptide sequence. The approach typically employs hundreds to thousands of different peptide sequences [44].

In contrast to synthetic peptide libraries, proteome-derived peptide libraries employ natural sequence diversity. Protease specificity profiling using proteome-derived peptide libraries is highlighted in (Fig. **2A**) [24]. Peptide libraries are generated by endoproteolytic digestion of proteomes such as cell lysates, thereby representing natural sequence diversity. After inactivation of the digestion protease and dimethylation of α– and ε–amines, peptide libraries are incubated with a test protease. Prime-site cleavage products possess free N-termini, allowing biotinylation and specific retrieval, followed by mass-spectrometry based identification. The corresponding non-prime site sequences are inferred from proteome sequence databases. Thus, this approach profiles prime- and non-prime specificity in a single experiment, directly identifies the position of the scissile peptide bond, and retrieves individual cleavage sequences rather than pooled preferences. The method often identifies hundreds of cleavage sequences for a test protease and has been used for serine-, cysteine-, metallo- and aspartyl-proteases [24, 45] as well as for the deorphanizing of previously uncharacterized proteases [46]. Identification of large arrays of cleavage sequences enables investigation of subsite cooperativity, as highlighted for HIV protease 1 (see Fig. **2B**): here, presence of a "large" residue in P1 favors acceptance of a "small" residue in P3 and vice-versa This effect has been originally described by Ridky *et al.* [47] and is correctly portrayed by specificity profiling with proteome-derived peptide libraries [24]. Notably, few other methods enable direct assessment of protease subsite cooperativity. Adaptions of the technique include multiplexed stable isotope tagging for kinetic investigations [48]. Proteome-derived peptide libraries have also been used to investigate the specificity of carboxypeptidases and Nα-acetyltransferases [49, 50]. In both cases, sophisticated chromatographic strategies were employed to retrieve modified peptides. Mass-spectrometry based identification of cleavage products has also been used in combination with synthetic rather than proteome-derived peptide libraries [51].

Importantly, proteome-derived peptide libraries do not indicate processing of native substrates. To this end, powerful approaches exist that identify and quantify protease cleavage sites in cells and tissues. Since identification of proteineous substrates in physiologic or pathologic settings exceeds the scope of the present review, we refer to two other reviews for a discussion of these degradomic techniques [52, 53]. However, detailed knowledge of protease specificity is useful in delineating candidate proteases that are able to mediate cleavage events that were identified *in vivo*. This concept has already been implemented for peptidomic analyses [54].

3. QUANTIFYING, MAPPING AND COMPARING SPECIFICITY

Experimental peptide substrate data can be utilized to quantify protease specificity. Based on probabilities for each amino acid in each protease subpocket, an information entropy "cleavage entropy" can be calculated [30]. These continuous values between zero (completely unspecific) and one (perfectly specific) highlight positions of specificity in protease substrate recognition and can be readily projected onto

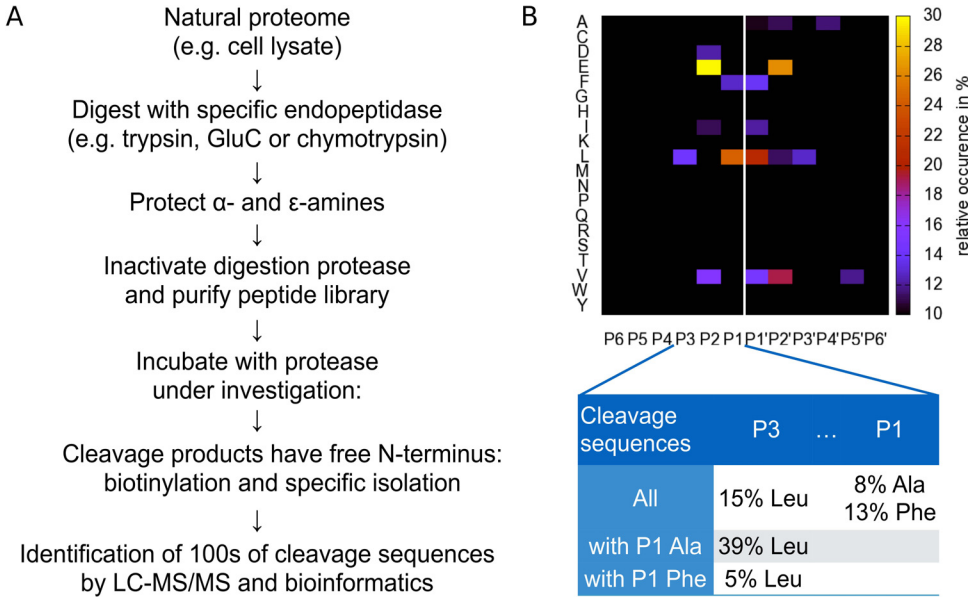

Fig. (2). Experimental techniques for protease substrate profiling: **A**) Schematic work-flow for protease substrate characterization using proteome-derived peptide libraries and LC-MS/MS technology for identification of cleaved peptides. **B**) Exemplary heat map of amino acid frequencies in HIV protease 1 substrates at positions P6-P6' reflecting preference of hydrophobic residues [23]. Interestingly, closer inspection of the ensemble of substrate sequences unveils cooperativity of substrate positions P3 and P1, where presence of a larger residue in one pocket limits size for the bound amino acid in the adjacent subsite.

the binding site where protein-substrate co-crystal structures are available (see Fig. **3A**). Because most proteases show a canonical orientation for the substrate peptide in extended beta conformation around the scissile for at least 6 amino acids (P3-P3') [55, 2], amino acid side-chains explore overlapping regions of the substrate binding cleft. Effects on substrate turn-over have been demonstrated for regions far beyond this central region have been shown recently in human and mouse granzyme B [56].

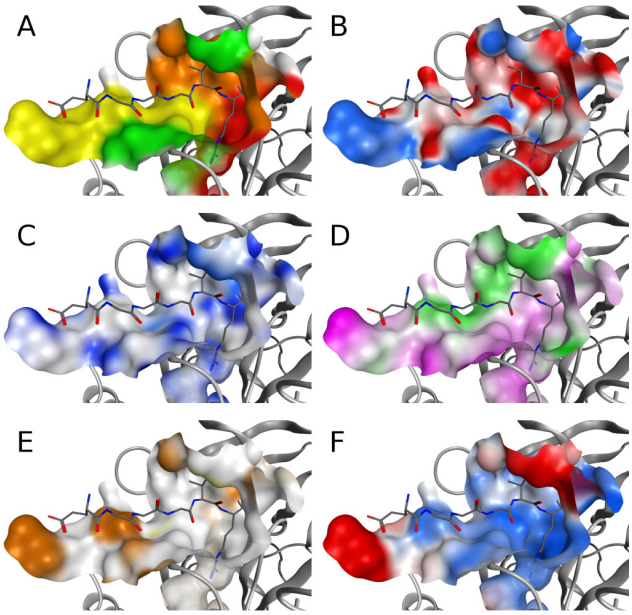

Fig. (3). Mappings to the binding site of thrombin: P1-P6 residues of the fibrinogen-derived peptide are shown as sticks, whilst the protease is depicted as grey cartoon. **A**) Specificity quantified as cleavage entropy S is projected to the molecular surface of thrombin on a colour gradient from red (specific) *via* yellow to green (unspecific). The deep S1 pocket determines substrate readout of thrombin (S_{S1}=0.176), whilst S2, S4, and S6 show preferences to variable extent (S_{S2}=0.635, S_{S4}=0.892, S_{S6}=0.899). S3 and S5 are almost complete non-specific: S_{S3}=0.971, S_{S5}=0.959. **B**) Binding site electrostatics mapped on a gradient from red (negative) over white (neutral) to blue (positive). **C**) Hydrogen bonding potential mapped on a gradient from white to blue (strong donor or acceptor). **D**) Lipophilicity mapped from pink (polar) *via* white (neutral) to green (apolar). **E**) Exposure of the pocket region mapped from white (enclosed) to orange (accessible). **F**) Flexibility as measured by crystallographic B-factors on a gradient from blue (rigid) *via* white to red (flexible).

Specificity landscapes form the basis for investigation of biomolecular recognition processes and can even be utilized to rationalize virtual screening results targeting proteases [57]. Alternatively, protease substrate data may be exploited to map distances between proteolytic distances (the degradome) in substrate space [33]. Thereby, similarities in substrate recognition can be identified in the absence of protease sequence or structural similarity. Here, peptide substrate data shows a similar information content as structural information [58]. Substrate-derived protease similarities have also been successfully employed as a lead discovery technique for novel protease inhibitors [59]. Substrate-driven mapping strategies have recently been explored similarly for kinase specificity mapping [60].

4. CHARACTERIZING PROTEIN STRUCTURE AND DYNAMICS

Structural data for proteases and their complexes with substrates and inhibitors has increased dramatically in recent years. Currently, the Protein Data Bank [61] contains in total 105,212 entries (database accession 27.11.2014). There are 8,901 of these structures annotated as peptidases (with enzyme classification (EC) number 3.4), thus representing 8.5% of the whole database and 14.6% of all enzymes with annotated EC code (total 61,014 entries). These structural data form the basis for a molecular understanding of protease-substrate interactions as a model for biomolecular recognition.

Binding sites of proteases may be computationally characterized in terms of static molecular properties like electrostatics [62], hydrophobicity [63] or cooperative hydration networks [64]. Size and surface properties of binding cavities can be similarly explored automatically [65, 66]. Different molecular probes can be used for the detection of interaction hot spots and their characterization [67]. See (Fig. **3B-F**) for example mappings to the binding site of the serine protease thrombin [68]. Molecular properties were calculated and mapped using Molecular Operating Environment (MOE) [69].

All aforementioned approaches treat proteins and their binding sites as rigid bodies. By contrast, proteins are inherently flexible entities and explore a range of conformational states in physiological conditions [70, 71]. Therefore, static enthalpic driving forces of molecular recognition are complemented by entropic factors arising from the dynamics of the system [72]. Molecular dynamics simulations allow exploration of conformational dynamics of biological systems at atomistic level in silico [73] with increasing time scales [74] and accuracy [75]. Generated conformational ensembles of proteins can be utilized to characterize binding sites in terms of global and local flexibility and their respective time scales [76]. Conformational entropies calculated from positional fluctuations [77] or dihedral distributions [78, 79] allow to identify flexible regions in binding sites. Additionally, state-of-the-art simulations are performed in presence of water boxes surrounding the simulated systems (explicit solvation), thus enthalpy and entropy of water molecules bound to binding site regions can be explored [80-83]. Hydration effects are especially interesting for proteases as the hydration is known to be key for protein stability and function [84, 85]. Recently, technologies using mixed solvents to probe binding site preferences for particular molecular fragments and allowing direct calculation of binding free energies have been developed [86, 87].

On the other hand experimental data allows insights into binding site flexibility for a limited set of proteolytic enzymes. Nuclear magnetic resonance (NMR) spectroscopy has been successfully employed to characterize solution dynamics in bacterial subtilisins [88], thrombin [89], and HIV protease [90]. On the other hand, ensemble refinement of crystal structures has been utilized recently to investigate binding site conformations of complement factor D [91].

Future broader utilization of noise in electron densities for identification of alternate protein conformations will help to characterize macromolecular flexibility based on crystallographic data [92].

5. SPECIFICITY OF PROTEASES

Substrate specificity of proteases has long been thought to be driven by static structural features only. Within the class of chymotrypsin-like serine proteases single amino acids directly contacting the substrate in the S1 pocket were thought to explain specificity completely [93]. This class of proteases shares a common catalytic triad and protein framework with mostly unspecific pockets except for S1-P1 interactions [94]. Therefore, the presence of an Asp residue in the S1 of trypsin explains its specificity for Arg and Lys at P1, whilst chymotrypsin and elastase show a preference for hydrophobic amino acids at the same position [95].

Hedstrom *et al.* attempted to prove the simple paradigm that single residues direct specificity by attempting to convert trypsin to chymotrypsin by S1-directed mutagenesis [96]. Even after replacing the whole S1 pocket of trypsin with the corresponding residues in chymotrypsin, the specificity could not be exchanged entirely. Therefore, simple static effects are not sufficient to explain protease specificity and attention was pointed towards adjacent surface loops [97]. Similarly, no unique solution was later found to exchange the specificity of trypsin and elastase and it was hence concluded that protease specificity is both difficult to rationalize and transfer [98]. The situation was even more complicated by the discovery that the S1 pocket in elastase communicates with other subpockets [99]. Therefore, factors for protease specificity were summed by Hedstrom as follows [1]: "Catalysis and specificity are not simply controlled by a few residues, but are rather a property of the entire protein framework, controlled *via* the distribution of charge across a network of hydrogen bonds and perhaps also by the coupling of domain motion to the chemical transformation".

All the aforementioned findings point towards more complex origins of proteases specificity beyond static structural factors. Indeed, dynamic contributions have been described in the recognition of almost all catalytic classes ranging from the serine proteases subtilisins [88] and α-lytic protease [100] *via* the aspartic HIV protease [101] to snake venom metalloproteases [102]. Recently, quantitative metrics for substrate specificity allowed for direct correlation of binding site flexibility and substrate promiscuity. Thereby, a direct interplay of dynamics and substrate recognition was shown for effector caspases [34]. Here, correlations between receptor backbone dynamics and specificity were shown to be stronger than between hydrogen bonding occupancy and specificity. Recently, a propagation of protease dynamics to the first hydration layers that might explain the specificity profile has been shown for thrombin [103]. This highlights that dynamics are key to a molecular understanding of protease-substrate peptide interactions and macromolecular binding events in general. Flexible binding sites adopt more diverse conformations which leads to promiscuity when following a binding paradigm of conformational selection [104, 105].

6. IMPLICATIONS FOR GENERAL MACROMOLECULAR BINDING EVENTS BEYOND PROTEASES

Protein-protein interfaces are of highest interest for both structural biology [106] as well as drug design [107]. The interface between proteolytic enzymes and their substrates is a well-studied example and therefore offers peptide data sets suitable for statistical analysis. In addition to the described analysis of the protease side of the protein-protein interface, the substrate side may be investigated by similar means raising the question which structural properties turn proteins into protease substrates. Here, broad proteolysis corresponding to non-specificity has been successfully used as a probe for thermal unfolding [108, 109], indicating strong links to local flexibility and accessibility [110]. Thereby, differences in protein dynamics and stability caused by mutations have been linked to different so-called conformational diseases [111-113]. Statistical analyses of glutamyl endopeptidase and caspase-3 cleavage sites revealed independence of cleavages sites from local secondary structure [114]. Exposure appears to be more crucial than flexibility and local interactions to allow proteolysis [115]. Recently, limited proteolysis was coupled with targeted proteomics to describe conformational changes in proteins [116]. On the other hand, local unfolding events are required for some proteolytic events, thus preventing their direct identification by fluctuations around the native state [117, 118].

Proteomics techniques allow profiling of substrate spectra of enzyme classes beyond proteases. Recently, several techniques assessing substrate preferences of kinases by phosphoproteomics have been developed [119, 120]. Still, the data basis here is not yet as broad as for peptides binding to PDZ domains, where again correlations between receptor promiscuity and flexibility have been reported [121]. Similarly, the binding specificity of ephrins to the Eph receptor appears to be coupled to intrinsic dynamics [122] as well as the binding specificity of ubiquitin [123]. Small molecule selectivities of G protein-coupled receptors have recently been successfully modelled and predicted based on structural descriptors [124]. Here, the number of disulfide bonds in the extracellular region seems to govern receptor promiscuity by determining the maximum ligand size in the entrance pathway. Comparably, affinity maturation of antibodies leading to proteins with increased affinity and selectivity is paralleled by a loss of flexibility [125]. Recently, even antibody degradation sites quantified *via* mass spectrometry data have been linked to local flexibility [126]. Intrinsically disordered proteins (IDPs) demonstrate the extreme case: Here, extreme flexibility leads to almost complete non-specific binding [127] and also higher susceptibility to proteolysis [128]. This extreme binding promiscuity is central for the function of IDPs as mediators for many cellular interaction networks in parallel [129].

CONCLUSION AND PERSPECTIVES

We have demonstrated how molecular determinants of specificity may be deduced from proteomics-derived peptide data. Joined forces of proteomics and structure-based modelling approaches allow tackling of topical questions of molecular biosciences and provide further insights into protein-protein interactions at molecular level.

As for all data-driven modelling techniques data accessibility and careful database curation are a prerequisite for statistical analyses. We therefore encourage the scientific community to support and make use of public data resources and associated tools [130]. For the described studies, required data spans from crystal structures from the Protein Data Bank [61], *via* protease classification schemes from MEROPS [12], their sequences from UniProt [131], to proteomics-derived peptide data. Here, the PRoteomics IDEntifications (PRIDE) database [132] allow uploading and annotation of large proteomics data sets that may be shared for further analysis *via* ProteomeXchange [133, 134].

Together with more and more sophisticated experimental techniques to detect proteolytic events and their kinetics [135], increasingly broader and more detailed analyses of their molecular origins can be performed. Thus, we are convinced that collaborations between experts in proteomics and structural bioinformatics will lead to a new understanding of macromolecular interactions and in turn to exciting novel opportunities for structure-based drug design.

CONSENT FOR PUBLICATION

Not applicable.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

JEF thanks the Austrian Academy of Sciences (DOC fellowship at University of Innsbruck) and the UK Medical Research Council (grant number MR/K020919/1) for funding. OS is supported by grants of the Deutsche Forschungsgemeinschaft (DFG) (SCHI 871/2 and SCHI 871/5) and the SFB850 (Project B8), a starting grant of the European Research Council (Programme "Ideas" - Call identifier: ERC-2011- StG 282111-ProteaSys), and the Excellence Initiative of the German Federal and State Governments (EXC 294, BIOSS). KRL acknowledges support from the Austrian Science Fund (projects P23051 and P26997).

REFERENCES

[1] Hedstrom, L. Serine protease mechanism and specificity. *Chem. Rev*., **2002**, *102*, 4501-4524.

[2] Madala, P.K.; Tyndall, J.D.; Nall, T.; Fairlie, D.P. Update 1 of: Proteases universally recognize beta strand in their active sites. *Chem. Rev*., **2010**, *110*, PR1-PR31.

[3] Davie, E.W.; Fujikawa, K.; Kisiel, W. The coagulation cascade: initiation, maintenance, and regulation. *Biochemistry*, **1991**, *30*, 10363-10370.

[4] Hengartner, M.O. The biochemistry of apoptosis. *Nature*, **2000**, *407*, 770-776.

[5] Sim, R.B.; Tsiftsoglu, S.A. Protease sof the complement system. *Biochem. Soc. Trans*., **2004**, *32*, 21-27.

[6] Carginale, V.; Trinchella, F.; Capasso, C.; Scudiero, R.; Riggio, M.; Parisi, E. Adaptive evolution and functional divergence of pepsin gene family. *Gene*, **2004**, *333*, 81-90.

[7] Overall, C.M.; Kleifeld, O. Tumour microenvironment – opinion: validating matrix metalloproteases as drug targets and anti-targets for cancer therapy. *Nat. Rev. Cancer*, **2006**, *6*, 227-239.

[8] Lopez-Otin, C.; Overall, C.M. Protease degradomics: a new challenge for proteomics. *Nat. Rev. Mol. Cell Biol*., **2002**, *3*, 509-519.

[9] Turk, B. Targeting proteases: successes, failures and future prospects. *Nat. Rev. Drug Discov*., **2006**, *5*, 785-799.

[10] Otlewski, J.; Jelen, F.; Zakrzewska, M.; Oleksy, A. The many faces of protease-protein inhibitor interaction. *EMBO J*., **2005**, *24*, 1303-1310.

[11] Rawlings, N.D. A large and accurate collection of peptidase cleavages in the MEROPS database. *Database*, **2009**, *2009*, bap015.

[12] 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.

[13] Igarashi, Y.; Eroshkin, A.; Gramatikova, S.; Gramatikoff, K.; Zhang, Y.; Smith, J.W.; Osterman, A.L.; Dodzik, A. CutDB: a proteolytic event database. *Nucleic Acids Res*., **2007**, *35*, D546-D549.

[14] Igarashi, Y.; Heureux, E.; Doctor, K.S.; Talwar, P.; Grammatikova, S.; Grammatikoff, K.; Zhang, Y.; Blinov, M.; Ibragimova, S.S.; Boyd, S.; Ratnikov, B.; Cieplak, P.; Godzik, A.; Smith, J.W.; Osterman, A.L.; Eroshkin, A.M. PMAP: databases for analyzing proteolytic events and pathways. *Nucleic Acids Res*., **2009**, *37*, D611-D618.

[15] Crawford, E.D.; Seaman, J.E.; Agard, N.; Hsu, G.W.; Julian, O.; Mahrus, S.; Nguyen, H.; Shimbo, K.; Yoshihara, H.A.I.; Zhuang, M.; Chalkley, R.J.; Wells, J.A. The DegraBase: A database of proteolysis in healthy and apoptotic human cells. *Mol. Cell. Proteomics*, **2013**, *12*, 813-824.

[16] Fortelny, N.; Yang, S.; Pavlidis, P.; Lange, P.F.; Overall, C.M. proteome TopFIND 3.0 with TopFINDer and PathFINDer: database and analysis tools for the association of protein termini to pre- and post-translational events. *Nucleic Acids Res*., **2015**, *43*, D290-D297.

[17] Rawlings, N.D.; Barrett, A.J.; Bateman, A. Using the MEROPS database for proteolytic enzymes and their inhibitors and substrates. *Curr. Protoc. Bioinform*., **2014**, *48*, 1.25.1-1.25.33.

[18] Schechter, I.; Berger, A. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun*., **1967**, *27*, 157-162.

[19] Schilling, O.; auf dem Keller, O.; Overall, C.M. Factor Xa subsite mapping by proteome-derived peptide libraries improved using WebPICS, a resource for proteomic identification of cleavage sites. *Biol. Chem*., **2011**, *392*, 1031-1037.

[20] Schneider, T.D.; Stephens, R.M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res*., **1990**, *18*, 6097-6100.

[21] Colaert, N.; Helsens, K.; Martens, L.; Vandekerckhove, J.; Gevaert, K. Improved visualization of protein consensus sequence by iceLogo. *Nat. Methods*, **2009**, *6*, 786-787.

[22] Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. WebLogo: A sequence logo generator. *Genome Res*., **2004**, *14*, 1188-1190.

[23] Wheeler, T.J.; Clements, J.; Finn, R.D. Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. *BMC Bioinform*., **2014**, *15*, 7.

[24] Schilling, O.; Overall, C.M. Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat. Biotechnol*., **2008**, *26*, 685-694.

[25] Van Damme, P.; Maurer-Stroh, S.; Plasman, K.; Van Durme, J.; Colaert, N.; Timmerman, E.; De Bock, P.J.; Goethals, M.; Rousseau, F.; Schymkowitz, J.; Vandekerckhove, J.; Gevaert, K. Analysis of protein processing by N-terminal proteomics reveals novel species-specific substrate determinants of granzyme B orthologs. *Mol. Cell. Proteomics*, **2009**, *8*, 258-272.

[26] Boyd, S.E.; Pike, R.N.; Rudy, G.B.; Whisstock, J.C.; Garcia de la Banda, M. PoPS: a computational tool for modeling and predicting protease specificity. *J. Bioinform. Comput. Biol*., **2005**, *3*, 551-585.

[27] Barkan, D.T.; Hofstetter, D.R.; Mahrus, S.; Pieper, U.; Wells, J.A.; Craik, C.S.; Sali, A. Prediction of protease substrates using sequence and structure features. *Bioinformatics*, **2010**, *26*, 1714-1722.

[28] Song, J.; Tan, H.; Perry, A.J.; Akutsu, T.; Webb, G.I.; Whisstock, J.C.; Pike, R.N. PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites. *PLoS One*, **2012**, *7*, e50300.

[29] Fortelny, N.; Cox, J.H.; Kappelhoff, R.; Starr, A.E.; Lange, P.F.; Pavlidis, P. Overall, C.M. Network analyses reveal pervasive functional regulation between proteases in the human protease web. *PLoS Biol*., **2014**, *12*, e1001869.

[30] 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**, e1003007.

[31] Rogers, L.D.; Overall, C.M. Proteolytic post-translational modifications in proteins: proteomic tools and methodology. *Mol. Cell. Proteomics*, **2013**, *12*, 3532-3542.

[32] Schilling, O.; auf dem Keller, U.; Overall, C.M. Protease specificity profiling by tandem mass spectrometry using proteome-derived peptide libraries. *Methods Mol. Biol*., **2011**, *753*, 257-272.

[33] 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.

[34] Fuchs, J.E.; von Grafenstein, S.; Huber, R.G.; Wallnoefer, H.G.; Liedl, K.R. Specificity of protein-protein interfaces: local dynamics direct substrate recognition of effector caspases. *Proteins*, **2014**, *82*, 546-555.

[35] Fields, G.B.; Van Wart, H.E.; Birkedal-Hansen, H. Sequence specificity of human skin fibroblast collagenase. Evidence for the role of collagen structure in determining the collagenase cleavage site. *J. Biol. Chem*., **1987**, *262*, 6221-6226.

[36] Matthews, D.J.; Wells, J.A. Substrate phage: selection of protease substrates by monovalent phage display. *Science*, **1993**, *260*, 1113-1117.

[37] Boulware, K.T.; Daugherty, P.S. Protease specificity determination by using cellular libraries of peptide substrates (CLiPS). *Proc. Natl. Acad. Sci. USA,* **2006**, *103*, 7583-7588.

[38] Ratnikov, B.I.; Cieplak, P.; Grammatikoff, K.; Pierce, J.; Eroshkin, A.; Igarashi, Y.; Kazanov, M.; Sun, Q.; Godzik, A.; Osterman, A.; Stec, B.; Strongin, A.; Smith, J.W. Basis for substrate recognition and distinction by matrix metalloproteinases. *Proc. Natl. Acad. Sci. USA,* **2014**, *111*, 4148-4155.

[39] Rano, T.A.; Timkey, T.; Peterson, E.P.; Rotonda, J.; Nicholson, D.W.; Becker, J.W.; Chaapman, K.T.; Thornberry, N.A. A combinatorial approach for determining protease specificities: application to interleukin-1beta converting enzyme (ICE). *Chem. Biol*., **1997**, *4*, 149-155.

[40] Turk, B.E.; Huang, L.L.; Piro, E.T.; Cantley, L.C. Determination of protease cleavage site motifs using mixture-based oriented peptide libraries. *Nat. Biotechnol*., **2001**, *19*, 661-667.

[41] Caescu, C.I.; Jeschke, G.R.; Turk, B.E. Active-site determinants of substrate recognition by the metalloproteinases TACE and ADAM10. *Biochem. J*., **2009**, *424*, 79-88.

[42] Turk, B.E.; Lee, D.H.; Yamakoshi, Y.; Klignenhoff, A.; Reichenberger, E.; Wright, J.T.; Simmer, J.P.; Komisarof, J.A.; Cantley, L.C.; Bartlett, J.D. MMP-20 is predominantely a toots-specific enzyme with a deep catalytic pocket that hydrolyzes type V collagen. *Biochemistry*, **2006**, *45*, 3863-3874.

[43] Winssinger, N.; Damoiseaux, R.; Tully, D.C.; Geierstanger, B.H.; Burdick, K. Harris, J.L. PNA-encoded protease substrate microarrays. *Chem. Biol*., **2004**, *11*, 1351-1360.

[44] Harris, J.; Mason, D.E.; Li, J.; Burdick, K.W.; Backes, B.J.; Chen, T.; Shipway, A.; Van Heeke, G.; Gough, L.; Ghaemmaghami, A.; Shakib, F.; Debaene, F.; Winssinger, N. Activity profile of dust mite allergen extract using substrate libraries and functional proteomic microarrays. *Chem. Biol*., **2004**, *11*, 1361-1372.

[45] Biniossek, M.L.; Nagler, D.K.; Becker-Pauly, C.; Schilling, O. Proteomic identification of protease cleavage sites characterizes prime and non-prime specificity of cysteine cathepsins B, L, and S. J. *Proteome Res*., **2011**, *10*, 5363-5373.

[46] Perera, N.C.; Schilling, O.; Kittel, H.; Back, W.; Kremmer, E.; Jenne, D.E. NSP4, an elastase-related protease in human neutrophils with arginine specificity. *Proc. Natl. Acad. Sci. USA,* **2012**, *109*, 6229-6234.

[47] Ridky, T.W.; Cameron, C.E.; Cameron, J.; Leis, J.; Copeland, T.; Wlodawer, A.; Weber, I.T.; Harrison, R.W. Human immunodeficiency virus, type 1 protease substrate specificity is limited by interactions between substrate amino acids bound in adjacent enzyme subsites. *J. Biol. Chem*., **1996**, *271*, 4709-4717.

[48] Jakoby, T.; van den Berg, B.H.; Tholey, A. Quantitative protease cleavage site profiling using tandem-mass-tag labeling and LC-MALDI-TOF/TOF MS/MS analysis. *J. Proteome Res*., **2012**, *11*, 1812-1820.

[49] Van Damme, P.; Evjenth, R.; Foyn, H.; Demeyer, K.; De Bock, P.J.; Lillehaug, J.R.; Vandekerckhove, J.; Arnesen, T.; Gevaert, K. Proteome-derived peptide libraries allow detailed analysis of the substrate specificities of N(alpha)-acetyltransferases and point to hNaa10p as the post-translational actin N(alpha)-acetyltransferase. *Mol. Cell. Proteomics*, **2011**, *10*, M110.004580.

[50] Tanco, S.; Lorenzo, J.; Garcia-Pardo, J.; Degroeve, S.; Martens, L.; Aviles, F.X.; Gevaert, K.; Van Damme, P. Proteome-derived peptide libraries to study the substrate specificity profiles of carboxypeptidases. *Mol. Cell. Proteomics*, **2013**, *12*, 2096-2110.

[51] O'Donoghue, A.J.; Eroy-Reveles, A.A.; Knudsen, G.M.; Ingram, J.; Zhou, M.; Statnekov, J.B.; Greninger, A.L.; Hostetter, D.R.; Qu, G.; Maltby, D.A.; Anderson, M.O.; Derisi, J.L.; McKerrow, J.H.; Burlingame, A.L.; Craik, C.S. Global identification of peptidase specificity by multiplex substrate profiling. *Nat. Methods*, **2012**, *9*, 1095-1100.

[52] Petrera, A.; Lai, Z.W.; Schilling, O. Carboxyterminal protein processing in health and disease: key actors and emerging technologies. *J. Proteome Res*., **2014**, *13*, 4497-4504.

[53] Lai, Z.W.; Petrera, A.; Schilling, O. Protein amino-terminal modifications and proteomic approaches for N-terminal profiling. *Curr. Opin. Chem. Biol*., **2015**, *24*, 71-79.

[54] Klein, J.; Ealer, J.; Zurbig, P.; Vlahou, A.; Mischak, H.; Stevens, R. Proteasix: a tool for automated and large-scale prediction of proteases involved in naturally occurring peptide generation. *Proteomics*, **2013**, *13*, 1077-1082.

[55] Tyndall, J.D.; Nall, T.; Fairlie, D.P. Proteases universally recognize beta strand in their active sites. *Chem. Rev*., **2005**, *105*, 973-999.

[56] Van Damme, P.; Plasman, K.; Vandemoortele, G.; Jonckheere, V.; Maurer-Stroh, S.; Gevaert, K. Importance of extended protease substrate recognition motifs in steering BNIP-2 cleavage by human and mouse granzymes B. *BMC Biochem*., **2014**, *15*, 21.

[57] Duchene, D.; Colombo, E.; Desilets, A.; Boudreault, P.L.; Leduc, R.; Marsault, E.; Najmanovich, R. Analysis of Subpocket Selectivity and Identification of Potent Selective Inhibitors for Matriptase and Matriptase-2. *J. Med. Chem*., **2014**, *57*, 10198-10204.

[58] Fuchs, J.E.; Liedl, K.R. Substrate sequences tell similar stories as binding cavities. *J. Chem. Inf. Model*., **2013**, *53*, 3115-3116.

[59] Sukuru, S.C.; Nigsch, F.; Quancard, J.; Renatus, M.; Chopra, R.; Broojimans, 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.

[60] Forti, F.L. Combined experimental and bioinformatics analysis for the prediction and identification of VHR/DUSP3 nuclear targets to DNA damage and repair. *Integr. Biol*., **2015**, *7*, 73-89.

[61] Berman, H.M.; Kleywegt, G.J.; Nakamura, H.; Markley, J.L. The Protein Data Bank archive as an open data resource. *J. Comput. Aided Mol. Des*., **2014**, *28*, 1009-1014.

[62] Miteva, M.A.; Tuffery, P.; Villoutreix, B.O. PCE: web tools to compute protein continuum electrostatics. *Nucleic Acids Res*., **2005**, *33*, W372-W375.

[63] Hlevnjak, M.; Zirkovic, G.; Zagrovic, B. Hydrophobicity matching – a potential prerequisite for the formation of protein-protein complexes in the cell. *PLoS One*, **2010**, *5*, e11169.

[64] Kuhn, B.; Fuchs, J.E.; Reutlinger, M.; Stahl, M.; Taylor, N.R. Rationalizing tight ligand binding through cooperative interaction networks. *J. Chem. Inf. Model*., **2011**, *51*, 3180-3198.

[65] Schmidtke, P.; Le Gouilloux, V.; Maupetit, J.; Tuffery, P. fpocket: online tools for protein ensemble pocket detection and tracking. *Nucleic Acids. Res*., **2010**, *38*, W582-589.

[66] Paramo, T.; East, A.; Garzon, D.; Ulmschneider, M.B.; Bond, P.J. Efficient Characterization of Protein Cavities within Molecular Simulation Trajectories: trj_cavity. *J. Chem. Theory. Comput*., **2014**, *10*, 2151-2164.

[67] Goodford, P.J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem*., **1985**, *28*, 849-857.

[68] Stubbs, M.T.; Oschkinat, H.; Mayr, I.; Huber, R.; Angliker, H.; Stone, S.R.; Bode, W. The interaction of thrombin with fibrinogen. A structural basis for its specificity. *Eur. J. Biochem*., **1992**, *206*, 187-195.

[69] Molecular Operating Environment (MOE), 2013.0801: Chemical Computing Group Inc., Montreal, QC, Canada, **2013**.

[70] Frauenfelder, H.; Sligar, S.G.; Wolynes, P.G. The energy landscapes and motions of proteins. *Science*, **1991**, *254*, 1598-1603.

[71] Boehr, D.D.; Nussinov, R.; Wright, P.E. The role of dynamics conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.*, **2009**, *5*, 789-796.

[72] Frederick, K.K.; Marlow, M.S.; Valentine, K.G.; Wand, J.A. Conformational entropy in molecular recognition by proteins. *Nature*, **2007**, *448*, 325-330.

[73] Dror, R.O.; Dirks, R.M.; Grossman, J.P.; Xu, H.; Shaw, D.E. Biomolecular simulation: a computational microscope for molecular biology. *Annu. Rev. Biophys*., **2012**, *41*, 429-452.

[74] Shaw, D.E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R.O.; Eastwood, M.P.; Bank, J.A.; Jumper, J.M.; Salmon, J.K.; Shan, Y.; Wriggers, W. Atomic-level charactierzation of the structural dynamics of proteins. *Science*, **2010**, *330*, 341-346.

[75] Piana, S.; Klepeis, J.L.; Shaw, D.E. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Curr. Opin. Struct. Biol*., **2014**, *21*, 579-584.

[76] Fuchs, J.E.; Waldner, B.J.; Huber, R.G.; von Grafenstein, S.; Kramer, C.; Liedl, K.R. Independent metrics for protein backbone and side-chain flexibility: time scales and effects of ligand binding. *J. Chem. Theory Comput*., **2015**, *11*, 851-860.

[77] Fuchs, J.E.; Huber, R.G.; von Grafeinstein, S.; Wallnoefer, H.G.; Spitzer, G.M.; Fuchs, D.; Liedl, K.R. Dynamics regulation of phenylalanine hydroxylase by simulated redox manipulation. *PLoS One*, **2012**, *7*, e53005.

[78] Ai, R.; Qaiser Fatmi, M.; Chang, C.E. T-Analyst: a program for efficient analysis of protein conformational changes by torsional angles. *J. Comput. Aided Mol. Des*., **2010**, *24*, 819-827.

[79] Huber, R.G.; Eibl, C.; Fuchs, J.E. Intrinsic flexibility of NLRP pyrin domains is a key factor in their conformational dynamics, fold stability and dimerization. *Protein Sci*., **2015**, *24*, 174-181.

[80] Haider, K.; Huggins, D.J. Combining solvent thermodynamic profiles with functionality maps of the Hsp90 binding site to predict the displacement of water molecules. *J. Chem. Inf. Model*., **2013**, *53*, 2571-2586.

[81] Breiten, B.; Lockett, M.R.; Sherman, W.; Fujita, S.; Al-Sayah, M.; Lange, H.; Bowers, C.M.; Heroux, A.; Krilov, G.; Whitesides, G.M. Water networks contribute to enthalpy/entropy compensation in protein-ligand binding. *J. Am. Chem. Soc*., **2013**, *135*, 15579-15584.

[82] Huber, R.G.; Fuchs, J.E.; von Grafenstein, S.; Laner, M.; Wallnoefer, H.G.; Abdelkader, N.; Kroemer, R. T.; Liedl, K.R.J. Entropy from states probabilities: hydration entropy of cations. *Phys. Chem. B*, **2013**, *117*, 6466-6472.

[83] Nguyen, C.N.; Cruz, A.; Gilson, M.K.; Kurtzman, T. Thermodynamics of water in an enzyme active site: grid-based hydration analysis of coagulation factor Xa. *J. Chem. Theory Comput*., **2014**, *10*, 2769-2780.

[84] Wallnoefer, H.G.; Handschuh, S.; Liedl, K.R.; Fox, T. Stabilizing of a globular protein by a highly complex water network: a molecular dynamics simulation study on factor Xa. *J. Phys. Chem. B*., **2010**, *114*, 7405-7412.

[85] Wallnoefer, H.G.; Liedl, K.R.; Fox, T. A GRID-derived water network stabilizes molecular dynamics simulations of a protease. *J. Chem. Inf. Model*., **2011**, *51*, 2860-2867.

[86] Guvench, O.; MacKerell, A.D. Jr. Computational fragment-based binding site identification by ligand competitive saturation. *PLoS Comput. Biol*., **2009**, *5*, e1000435.

[87] Alvarez-Garcia, D.; Barril, X. Molecular simulations with solvent competition quantify water displaceability and provide accurate interaction maps of protein binding sites. *J. Med. Chem*., **2014**, *57*, 8530-8539.

[88] Mulder, F.A.; Schipper, D.; Bott, R.; Boelens, R. Altered flexibility in the substrate-binding site of related native and engineered high-alkaline *Bacillus subtilisins*. *J. Mol. Biol*., **1999**, *292*, 111-123.

[89] Lechtenberg, B.C.; Johnson, D.J.; Freund, S.M.; Huntington, J.A. NMR resonance assignments of thrombin reveal the conformational and dynamic effects of ligation. *Proc. Natl. Acad. Sci. USA,* **2010**, *107*, 14087-14092.

[90] Nicholson, L.K.; Yamazaki, T.; Torchia, D.A.; Grzesiek, S.; Bax, A.; Stahl, S.J.; Kaufman, J.D.; Wingfield, P.T.; Lam, P.Y.; Jadhav, P.K.; Hodge, C.N.; Domaille, P.J.; Chang, C.H. Flexibility and function in HIV-1 protease. *Nat. Struct. Biol*., **1995**, *2*, 274-280.

[91] Forneris, F.; Burnley, B.T.; Gros, P. Ensemble refinement shows conformational flexibility in crystal structures of human complement factor D. *Acta Crystogr. D Biol. Crystallogr*., **2014**, *70*, 733-743.

[92] Lang, P.T.; Holton, J.M.; Fraser, J.S.; Alber, T. Protein structural ensembles are revealed by redefining X-ray electron density noise. *Proc. Natl. Acad. Sci. USA,* **2014**, *111*, 237-242.

[93] Steitz, T.A.; Shulman, R.G. Crystallographic and NMR studies of the serine proteases. *Annu. Rev. Biophys. Bioeng*., **1982**, *11*, 419-444.

[94] Perona, J.J.; Craik, C.S. Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold. *J. Biol. Chem*., **1997**, *272*, 29987-29990.

[95] Steitz, T.A.; Henderson, R.; Blow, D.M. Structure of crystalline alpha-chymotrypsin. 3. Crystallographic studies of substrates and inhibitors bound to the active site of alpha-chymotrypsin. *J. Mol. Biol*., **1969**, *46*, 337-348.

[96] Hedstrom, L. Trypsin: a case study in the structural determinants of enzyme specificity. *Biol. Chem*., **1996**, *377*, 465-470.

[97] Hedstrom, L.; Szilagyi, L.; Rutter, W.J. Converting trypsin to chymotrypsin: the role of surface loops. *Science*, **1992**, *255*, 1249-1253.

[98] Hung, S.H.; Hedstrom, L. Converting trypsin to elastase: substitution of the S1 site and adjacent loops reconstitutes esterase specificity but not amidase activity. *Protein Eng*., **1998**, *11*, 669-673.

[99] Bode, W.; Mezer, E.Jr.; Powers, J.C. Human leukocyte and porcine pancreatic elastase: X-ray crystal structures, mechanism, substrate specificity, and mechanism-based inhibitors. *Biochemistry*, **1989**, *28*, 1951-1963.

[100] Mace, J.E.; Wilk, B.J.; Agard, D.A. Functional linkage between the active site of alpha-lytic protease and distant regions of structure: scanning alanine mutagenesis of a surface loop affects activity and substrate specificity. *J. Mol. Biol*., **1995**, *251*, 116-134.

[101] Ozen, A.; Haliloglu, T.; Schiffer, C. Dynamics of preferential substrate recognition in HIV-1 protease: redefining the substrate envelope. *J. Mol. Biol*., **2011**, *410*, 726-744.

[102] Wallnoefer, H.G.; Lingott, T.; Gutierrez, J.M.; Merfort, I.; Liedl, K.R. Backbone flexibility controls the activity and specificity of a protein-protein interface: specificity in snake venom metalloproteases. *J. Am. Chem. Soc*., **2010**, *132*, 10330-10337.

[103] Fuchs, J.E.; Huber, R.G.; Waldner, B.J.; Kahler, U.; von Grafenstein, S.; Kramer, C.; Liedl, K.R. Dynamics govern specificity of a protein-protein interface: substrate recognition by thrombin. *PLoS One,* **2015**, *10,* e0140713.

[104] Tsai, C.J.; Kumar, S.; Ma, B.; Nussinov, R. Folding funnels, binding funnels, and protein function. *Protein Sci*., **1999**, *8*, 1181-1190.

[105] Vogt, A.D.; Di Cera, E. Conformational selection is a dominant mechanism of ligand binding. *Biochemistry*, **2013**, *52*, 5723-5729.

[106] Nooren, I.M.; Thornton, J.M. Structural characterization and functional significance of transient protein-protein interactions. *J. Mol. Biol*., **2003**, *325*, 991-1018.

[107] Wells, J.A.; McClendon, C.L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature*, **2007**, *450*, 1001-1009.

[108] Wang, L.; Chen, R.X.; Kallenbach, N.R. Proteolysis as a probe of thermal unfolding of cytochrome C. *Proteins*, **1998**, *30*, 435-441.

[109] Sides, C.R.; Liyanage, R.; Lay, J.O. Jr.; Philominathan, S.T.L.; Matsushita, O.; Sakon, J. Probing the 3-D structure, dybnamics, and stability of bacterial collagenase collagen binding domain (apo-versus holo-) by limited proteolysis MALDI-TOF MS. *J. Am. Soc. Mass Spectrom*., **2012**, *23*, 505-519.

[110] Novotny, J.; Bruccoleri, R.E. Correlation among sites of limited proteolysis, enzyme accessibility and segmental mobility. *FEBS Lett*., **1987**, *211*, 185-189.

[111] Pey, A.L.; Maggi, M.; Valentini, G. Insights into human phosphoglycerate kinase 1 deficiency as a conformational disease from biochemical, biophysical, and *in vitro* expression analyses. *J. Inherit. Metab. Dis*., **2014**, *37*, 909-916.

[112] Perez, B.; Desviat, L.R.; Gomez-Puertas, P.; Martinez, A.; Stevens, R.C.; Ugarte, M. Kinetic and stability analysis of PKU mutations identified in BH4-responsive patients. *Mol. Genet. Metab*., **2005**, *86*, S43-S53.

[113] Keller, M.A.; Zander, U.; Fuchs, J.E.; Kreutz, C.; Watschinger, K.; Mueller, T.; Golderer, G.; Liedl, K.R.; Ralser, M.; Kraeutler, B.; Wener, E.R.; Marquez, J.A. A gatekeeper helix determines the substrate specificity of Sjoegren-Larsson Syndrome enzyme fatty aldehyde dehydrogenase. *Nat. Commun*., **2014**, *5*, 4439.

[114] Timmer, J.C.; Zhu, W.; Pop, C.; Regan, T.; Snipas, S.J.; Eroshkin, A.M.; Riedl, S.J.; Salvesen, G.S. Structural and kinetic determi-

nants of protease substrates. *Nat. Struct. Mol. Biol*., **2009**, *16*, 1101-1108.

[115] Kazanov, M.D.; Igarashi, Y.; Eroshkin, A.M.; Cieplak, P.; Ratnikov, B.; Zhang, Y.; Li, Z.; Godzik, A.; Osterman, A.L.; Smith, J.W. Structural determinants of limited proteolysis. *J. Proteome Res*., **2011**, *10*, 3642-3651.

[116] Feng, Y.; De Francesci, G.; Kahraman, A.; Sose, M.; Melnik, A.; Boersema, P.J.; de Laureto, P.P.; Nikolaev, Y.; Oliveira, A.P.; Picotti, P. Global analysis of protein structural changes in complex proteosomes. *Nat. Biotechnol*., **2014**, *32*, 1036-1044.

[117] Hubbard, S.J. The structural aspects of limited proteolysis of native proteins. *Biochim. Biophys. Acta*, **1998**, *1382*, 191-206.

[118] Tsai, C.J.; Polverino de Laureto, P.; Fontana, A.; Nussinov, R. Comparison of protein fragments identified by limited proteolysis and by computational cutting of proteins. *Protein Sci*., **2002**, *11*, 1753-1770.

[119] Kettenbach, A.N.; Schweppe, D.K.; Faherty, B.K.; Pechenick, D.; Pletnev, A.A.; Gerber, S.A. Quantitative phosphoproteomics identifies substrates and functional modules of aurora and polo-like kinase activities in mitotic cells. *Sci. Signal*., **2011**, *4*, rs5.

[120] Xue, L.; Geahlen, R.L.; Tao, W.A. Identification of direct tyrosine kinase substrates based on protein kinase assay-linked phosphoproteomics. *Mol. Cell. Proteomics*, **2013**, *12*, 2969-2980.

[121] Munz, M.; Hein, J.; Biggins, P.C. The role of flexibility and conformational selection in the binding promiscuity of PDZ domains. *PLoS Comput. Biol*., **2012**, *8*, e1002749.

[122] Dai, D.; Huang, Q.; Nussinov, R.; Ma, B. Promiscuous and specific recognition among ephrins and Eph receptors. *Biochim. Biophys. Acta*., **2014**, *1844*, 1729-1740.

[123] Michielssens, S.; Peters, J.H.; Ban, D.; Pratihar, S.; Seeliger, D.; Sharma, M.; Giller, K.; Sabo, T.M.; Becker, S.; Lee, D.; Griesinger, C.; de Groot, B.L. A designed conformational shift to control binding specificity. *Angew. Chem. Int. Ed. Engl*., **2014**, *53*, 10367-10371.

[124] Levit, A.; Beuming, T.; Krilov, G.; Sherman, W.; Niv, M.Y. Predicting GPCR promiscuity using binding site features. *J. Chem. Inf. Model*., **2014**, *54*, 184-194.

[125] Manivel, V.; Sahoo, N.C.; Salunke, D.M.; Rao, K.V.S. Maturation of an antibody response is governed by modulations in flexibility of the antigen-combining site. *Immunity*, **2000**, *13*, 611-620.

[126] Sydow, J.F.; Lipsmeier, F.; Larraillet, V.; Hilger, M.; Mautz, B.; Molhoj, M.; Kuentzer, J.; Klostermann, S.; Schoch, J.; Voelger, H.R.; Regula, J.T.; Cramer, P.; Papadimitriou, A.; Kettenberger, H. Structure-based prediction of asparagine and aspartate degradation sites in antibody variable regions. *PLoS One*, **2014**, *9*, e100736.

[127] Huang, Y.; Liu, Z. Do intrinsically disordered proteins possess high specificity in protein-protein interactions? *Chem. Eur. J*., **2013**, *19*, 4462-4467.

[128] Suskiewicz, M.J.; Sussman, J.L.; Silman, I.; Shaul, Y. Context-dependent resistance to proteolysis of intrinsically disordered proteins. *Protein Sci*., **2011**, *20*, 1285-1297.

[129] Cumberworth, A.; Lamour, G.; Babu, M.M.; Gsponer, J. Promiscuity as a functional trait: intrinsically disordered regions as central players of interactomes. *Biochem. J*., **2013**, *454*, 361-369.

[130] Fernandez-Suarez, X.M.; Rigden, D.J.; Galperin, M.Y. The 2014 nucleic acids research database issue and an updated NAR online molecular biology database collection. *Nucleic Acids Res*., **2014**, *42*, D1-D6.

[131] UniProt Consortium. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res*., **2014**, *42*, D191-D198.

[132] Vizcaino, J.A.; Cote, R.G.; Csordas, A.; Dianes, J.A.; Fabregat, A.; Foster, J.M.; Griss, J.; Alpi, E.; Birim, M.; Contell, J.; O'Kelly, G.; Schoenegger, A.; Ovelleiro, D.; Perez-Riverol, Y.; Reisinger, F.; Rios, D.; Wang, R.; Hermjakob, H. The proteomics identifications (PRIDE) database and associated tools. *Nucleic Acids Res*., **2013**, *41*, D1063- D1069.

[133] Barsner, H.; Vizcaino, J.A.; Eidhammer, I.; Martens, L. PRIDE Converter: making proteomics data-sharing easy. *Nat. Biotechnol*., **2009**, *27*, 598-599.

[134] Vizcaino, J.A.; Deutsch, E.W.; Wang, R.; Csordas, A.; Reisinger, F.; Rios, D.; Dianes, J.A.; Sun, Z.; Farrah, T.; Bandeira, N.; Binz, P.A.; Xenarios, I.; Eisenacher, M.; Mayer, G.; Gatto, L.; Campos, A.; Chalkey, R.J.; Kraus, H.J.; Albar, J.P.; Martinez-Bartolome, S.; Apweiler, R.; Omenn, G.S.; Martens, L.; Jones, A.R.; Hermjakob, H. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol*., **2014**, *32*, 223-226.

[135] Schlage, P.; Egli, F.E.; Nanni, P.; Wang, L.W.; Kizhakkedathu, J.N.; Apte, S.S.; auf dem Keller, U. Time-resolved analysis of the matrix metalloproteinase 10 substrate degradome. *Mol. Cell. Proteomics*, **2014**, *13*, 580-593.