choosing the most relevant enzyme-ligand complex for their analyses in structure-based drug design applications but will also help the structural bioinformatics community in selecting suitable structural representatives for protein family-based studies on ligand diversity (Najmanovich, 2017), function evolution (Das et al., 2015), and structural modeling (Lam et al., 2017), among others. For example, the carbonic anhydrase superfamily in CATH (Dawson et al., 2017) contains many carbonic anhydrase enzyme structures with different ligands bound in different parts of the structure (Figure 2A). Using the ranking of PDB structures for each EC by Tyzack et al. (2018), it is possible to easily find the enzyme structure with the cognate ligand bound (exact match in this case, Figure 2B). Another research area to which this study will contribute massively is the structure-guided interpretation of how genetic variants may impact the structure and function of proteins.

In summary, the study presents a very valuable strategy for selecting structures of enzyme-cognate ligand complexes for structural biologists, structural bioinformaticians, and biomedical researchers alike, with helpful data on PDB structures provided for these users.

REFERENCES

Ahmed, A., Smith, R.D., Clark, J.J., Dunbar, J.B., Jr., and Carlson, H.A. (2015). Recent improvements to Binding MOAD: a resource for protein-ligand binding affinities and structures. Nucleic Acids Res. *43*, D465–D469.

Bashton, M., Nobeli, I., and Thornton, J.M. (2006). Cognate ligand domain mapping for enzymes. J. Mol. Biol. *364*, 836–852.

Das, S., Dawson, N.L., and Orengo, C.A. (2015). Diversity in protein domain superfamilies. Curr. Opin. Genet. Dev. *35*, 40–49.

Dawson, N.L., Lewis, T.E., Das, S., Lees, J.G., Lee, D., Ashford, P., Orengo, C.A., and Sillitoe, I. (2017). CATH: an expanded resource to predict protein function through structure and sequence. Nucleic Acids Res. *45* (D1), D289–D295.

Lam, S.D., Das, S., Sillitoe, I., and Orengo, C. (2017). An overview of comparative modelling and resources dedicated to large-scale modelling of genome sequences. Acta Crystallogr. D Struct. Biol. *73*, 628–640.

Najmanovich, R.J. (2017). Evolutionary studies of ligand binding sites in proteins. Curr. Opin. Struct. Biol. *45*, 85–90.

Śledź, P., and Caflisch, A. (2018). Protein structure-based drug design: from docking to molecular dynamics. Curr. Opin. Struct. Biol. *48*, 93–102.

Tyzack, J.D., Fernando, L., Ribeiro, A.J.M., Borkakoti, N., and Thornton, J.M. (2018). Ranking enzyme structures in the PDB by bound ligand similarity to biological substrates. Structure *26*, this issue, 565–571.

Yang, J., Roy, A., and Zhang, Y. (2013). BioLiP: a semi-manually curated database for biologically relevant ligand-protein interactions. Nucleic Acids Res. *41*, D1096–D1103.

Unexpected Specificity of a Trypsin-like Enzyme

Zbigniew Dauter[1,*] and Alexander Wlodawer[2,*]
[1]Synchrotron Radiation Research Section, Macromolecular Crystallography Laboratory, NCI, Argonne National Laboratory, Argonne, IL 60439, USA
[2]Protein Structure Section, Macromolecular Crystallography Laboratory, National Cancer Institute, Frederick, MD 21702, USA
*Correspondence: dauter@anl.gov (Z.D.), wlodawer@nih.gov (A.W.)
https://doi.org/10.1016/j.str.2018.03.007

In this issue of *Structure*, Stach et al. (2018) describe the properties of SplE, a trypsin-like enzyme from *Staphylococcus aureus*. They report, on the basis of biochemical and structural studies, the unusual specificity requiring the presence of histidine in the P1 subsite of its substrate.

A search of the Protein Data Bank (PDB; Berman et al., 2000) with the term "trypsin" identifies over 1,000 entries, with a further 500 or so found with the term "chymotrypsin." The oldest deposited structures of these enzymes are 40 years old (PDB: 1tgb; Fehlhammer et al., 1977), so it should be a safe bet that by now all should be known about the mode of activity and about specificity of this family of enzymes. Nevertheless, a paper by Stach et al. (2018) in this issue of *Structure* describes unusual properties of another trypsin-like enzyme and its specificity for the presence of histidine in the P1 subsite (Schechter and Berger, 1967) of its substrate.

This trypsin-like enzyme SplE from *Staphylococcus aureus* is a product of an operon that encodes, in total, six closely related proteases and might constitute a first example of a proteolytic system in bacteria. This operon is located on a pathogenicity island, suggesting that the produced enzymes might be involved in bacterial virulence. All six enzymes share a similar fold, whereas their specificity differs very significantly, ranging from chymotrypsin-like specificity of SplA (P1 tyrosine or phenylalanine) and glutamine, asparagine, and aspartic acid for SplB, to elastase-like specificity of SplD (small residues in P1). SplE, however, differs in its specificity not only from the other enzymes encoded by this operon but from practically all other members of the trypsin family, which rarely if ever are capable of cleaving substrates containing a histidine residue at the P1 site.

Determination of the specificity of SplE was a relatively straightforward part of this project. A tetrapeptide library of the type Ac-X4-X3-X2-P1-Acc (where Acc denotes 7-amino-4-carbamoylmethylcoumarin) containing 18 sub-libraries

was utilized. In each sub-library, the P1 position contained a particular natural amino acid, while the X4, X3, and X2 positions contained equimolar mixtures of tested residues. Hydrolysis of peptides in the library resulted in an increase of fluorescence, and the results were completely clear—high specificity for a histidine in P1, with much lower activity for a glutamine, and practically no activity with other amino acids. Further elaboration of the specificity of SplE was performed with a cellular library of peptide substrates, in which a large number of potential substrates (8-mer peptides) were displayed on the surface of *E. coli* cells. The cleavage of these peptide substrates would lead to a change in the fluorescence properties of the cells, amenable to fluorescence-activated cell sorting (FACS). The results of the latter studies have shown that the specificity of SplE was not limited to the P1 site alone, but was also influenced by other sites more distant from the cleavage point. Thus, the P2 site would preferentially contain leucine, whereas the preference at the P3 site was for a large aromatic residue (tyrosine, phenylanine, or tryptophan).

The authors determined a high-resolution (1.75 Å) crystal structure of SplE to gain insight into the structural basis of the observed specificity of the enzyme. Whereas the structure is that of the apoenzyme, the conformation of the substrate could be ascertained by modeling based on the data available for other trypsin-like enzymes with ligands present in their active sites. Thus, modeling of the consensus tetrapeptide Trp-Leu-His-Gly in the active site of SplE provides convincing clues about structural features around the S1 pocket of the enzyme responsible for its preference of His in this place. The peptide in the active site of SplE with the surrounding enzyme residues interacting with it are shown in Figure 1 (as well as Figure 4 of Stach et al., 2018). The optimal positioning of the peptide requires only very small rearrangement of the P1 pocket, as observed in the crystal structure of apo-SplE presented by Stach et al. (2018). The imidazole ring of P1 His is hydrogen bonded through its ND1 atom with the enzyme histidine and through the NE2 atom with the carbonyl oxygen atom of glycine at the bottom of the S1 pocket. These interactions, as well as the presence of Val residue lining the S1 pocket, provide an effective arrangement encompassing the His side chain in the pocket. It is also possible that a Glu side chain, similar in length to His, may also create similar interactions with the same His residue, which would explain the partial specificity of SplE to this amino acid in the P1 site.

All members of the Spl family of proteases differ to some extent in the structural details around their active sites from the canonical arrangement observed in other trypsin-like proteases. Stach et al. (2018) suggest a possibility that only some specific substrates induce certain conformational changes leading to the catalytically effective arrangement of the active site. This—currently speculative—idea could explain the exceptional substrate specificity of Spl proteases, which is stricter than that of other related enzymes from different sources. However, as these hypotheses are solely based on the structure of the apoenzyme, and since it is necessary to make some rearrangements to the coordinates to accommodate modeled substrates, these suggestions need to be considered with some caution. It may be expected that the work of Stach et al. (2018) will be pursued, particularly by investigating SplE structures in the presence of non-cleavable substrates or inhibitors, which will shed light on the structural and functional characteristics of this unusual and interesting group of proteases.

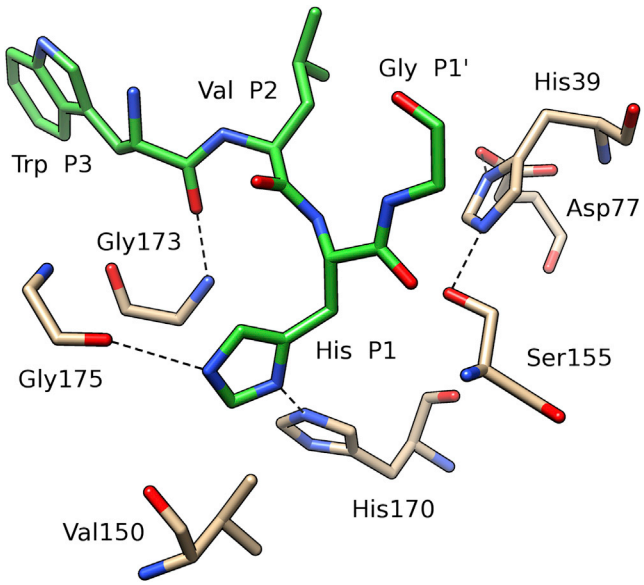

Figure 1. A Model of a Fragment of a Substrate Including the Subsites P3-P1′ Interacting with SplE

This model was created independently of the model shown by Stach et al. (2018) by superimposing on SplE a structure of trypsin complexed with a Kunitz inhibitor. The reactive loop of the inhibitor was computationally mutated, and the side chain of His39, part of the catalytic triad (Dodson and Wlodawer, 1998) but not observed due to disorder, was added.

REFERENCES

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The protein data bank. Nucleic Acids Res. *28*, 235–242.

Dodson, G., and Wlodawer, A. (1998). Catalytic triads and their relatives. Trends Biochem. Sci. *23*, 347–352.

Fehlhammer, H., Bode, W., and Huber, R. (1977). Crystal structure of bovine trypsinogen at 1-8 A resolution. II. Crystallographic refinement, refined crystal structure and comparison with bovine trypsin. J. Mol. Biol. *111*, 415–438.

Schechter, I., and Berger, A. (1967). On the size of the active site in proteases. I. Papain. Biochem. Biophys. Res. Commun. *27*, 157–162.

Stach, N., Kalinska, M., Zdzalik, M., Kitel, R., Karim, A., Serwin, K., Rut, W., Larsen, K., Jabaiah, A., Firlej, M., et al. (2018). Unique substrate specificity of SpiE serine protease from *Staphylococcus aureus*. Structure *26*, this issue, 572–579.