Cathepsin B: Active Site Mapping with Peptidic Substrates and Inhibitors

Schmitz, J.; Gilberg, E.; Löser, R.; Bajorath, J.; Bartz, U.; Gütschow, M.;

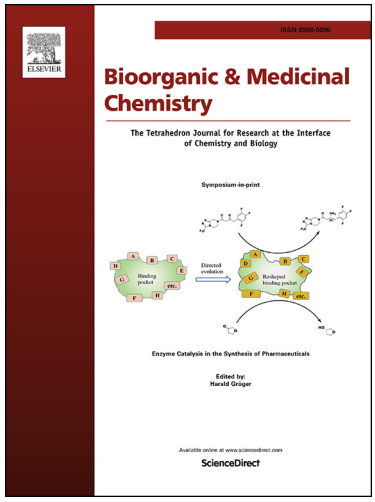
Originally published:

January 2019

Bioorganic & Medicinal Chemistry 27(2019)1, 1-15

DOI: https://doi.org/10.1016/j.bmc.2018.10.017

Accepted Manuscript



Cathepsin B: Active Site Mapping with Peptidic Substrates and Inhibitors

Janina Schmitz, Erik Gilberg, Reik Löser, Jürgen Bajorath, Ulrike Bartz, Michael Gütschow

PII: S0968-0896(18)31561-X
DOI: https://doi.org/10.1016/j.bmc.2018.10.017
Reference: BMC 14576

To appear in: *Bioorganic & Medicinal Chemistry*

Received Date: 4 September 2018
Revised Date: 16 October 2018
Accepted Date: 18 October 2018

Please cite this article as: Schmitz, J., Gilberg, E., Löser, R., Bajorath, J., Bartz, U., Gütschow, M., Cathepsin B: Active Site Mapping with Peptidic Substrates and Inhibitors, *Bioorganic & Medicinal Chemistry* (2018), doi: https://doi.org/10.1016/j.bmc.2018.10.017

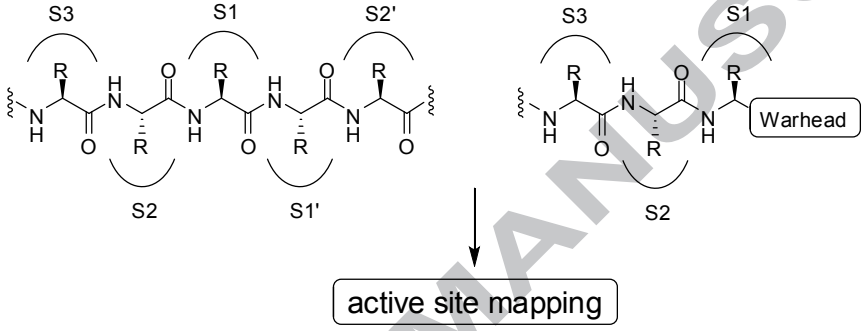
Graphical Abstract

Cathepsin B: Active Site Mapping with Peptidic Substrates and Inhibitors

Leave this area blank for abstract info.

Janina Schmitz [a,b,e], Erik Gilberg [a,c,e], Reik Löser [d], Jürgen Bajorath [c], Ulrike Bartz [b], Michael Gütschow [a]

[a] *Pharmaceutical Institute, Pharmaceutical Chemistry I, University of Bonn, An der Immenburg 4, D-53121 Bonn, Germany*
[b] *Department of Natural Sciences, University of Applied Sciences Bonn-Rhein-Sieg, von-Liebig-Str. 20, D-53359 Rheinbach, Germany*
[c] *Department of Life Science Informatics, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, University of Bonn, Dahlmannstr. 2, D-53113 Bonn, Germany*
[d] *Department of Radiopharmaceutical and Chemical Biology, Institute of Radiopharmaceutical Cancer Research, Helmholtz-Zentrum Dresden-Rossendorf, Bautzner Landstraße 400, D-01328 Dresden, Germany*



Bioorganic & Medicinal Chemistry
journal homepage: www.elsevier.com

Cathepsin B: Active Site Mapping with Peptidic Substrates and Inhibitors

Janina Schmitz [a,b,e], Erik Gilberg [a,c,e], Reik Löser [d], Jürgen Bajorath [c], Ulrike Bartz [b], Michael Gütschow [a]

[a] *Pharmaceutical Institute, Pharmaceutical Chemistry I, University of Bonn, An der Immenburg 4, D-53121 Bonn, Germany*
[b] *Department of Natural Sciences, University of Applied Sciences Bonn-Rhein-Sieg, von-Liebig-Str. 20, D-53359 Rheinbach, Germany*
[c] *Department of Life Science Informatics, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, University of Bonn, Dahlmannstr. 2, D-53113 Bonn, Germany*
[d] *Department of Radiopharmaceutical and Chemical Biology, Institute of Radiopharmaceutical Cancer Research, Helmholtz-Zentrum Dresden-Rossendorf, Bautzner Landstraße 400, D-01328 Dresden, Germany*

ARTICLE INFO

Article history:
Received
Received in revised form
Accepted
Available online

Keywords:
Active site mapping
Cathepsin B
Fluorescence-quenched substrates
Peptidomimetic inhibitors
Substrate specificity

ABSTRACT

The potential of papain-like cysteine proteases, such as cathepsin B, as drug discovery targets for systemic human diseases has prevailed over the past years. The development of potent and selective low-molecular cathepsin B inhibitors relies on the detailed expertise on preferred amino acid and inhibitor residues interacting with the corresponding specificity pockets of cathepsin B. Such knowledge might be obtained by mapping the active site of the protease with combinatorial libraries of peptidic substrates and peptidomimetic inhibitors. This review, for the first time, summarizes a wide spectrum of active site mapping approaches. It considers relevant X-ray crystallographic data and discloses propensities towards favorable protein-ligand interactions in case of the therapeutically relevant protease cathepsin B.

1. Introduction

Cathepsin B is a ubiquitously expressed human lysosomal cysteine protease which is involved in several physiological processes. Cathepsin B has been identified as an important tumor promoting factor, involved in the proteolytic cascade for the breakdown of the extracellular matrix (ECM). This degradation can occur either by extracellular proteolysis of the ECM components, such as laminin, fibronectin, collagen type I and IV and proteoglycans, or indirectly *via* activation of other proteases. Moreover, ECM components can be internalized to intracellular vesicles for a continued proteolysis. ECM degradation remodels the tumor environment and enables angiogenesis, tumor migration, invasion and even metastasis.[1-3] Based on several *in vitro* and *in vivo* models, including cathepsin B knock-down or over-expression mouse models, as well as pharmacological tools, such as specific inhibitors or siRNAs, the crucial role of cathepsin B at multiple points of the tumor development has been established.[4] Cathepsin B has now been recognized as a prognostic marker in tumor patients,[1,5] and has been validated as a promising target for drug development.[1,2,6,7]

* Corresponding author. *E-mail address:* guetschow@uni-bonn.de.
[c] Contributed equally.

Cathepsin B, a 30-kDa protein, is folded into two distinct domains, which interact through an extended polar interface that opens to a solvent-accessible, V-shaped active site cleft. The protein possesses six disulfide bridges and two unpaired cysteine residues, i.e. Cys29, the active site cysteine, and Cys240. The active site cysteine that is part of the left domain and a histidine residue (His199) from the right domain interact as an ion pair. The specificity pockets S3, S2, S1, S1' and S2' are formed by loops of the left (S3, S1, S2') or right (S2, S1') domain.[5,6,8]

Cathepsin B shows both endopeptidase and exopeptidase activity (Fig. 1). In the open conformation, large endopeptidase substrates can be bound and cleaved. However, cathepsin B possesses a unique structural feature, the occluding loop, which covers the primed binding region in the closed conformation of the enzyme (Fig. 2). Thus, the occluding loop confers a dipeptidyl carboxypeptidase activity to cathepsin B. The polar interaction of two protonated histidine residues (His110 and His111) of the occluding loop with the substrate's C-terminal carboxyl group contributes to the exopeptidase activity of cathepsin B in the closed conformation.[8,9] A mutant, mature form of cathepsin B, where twelve residues of the occluding were deleted, was proteolytically active as endopeptidase, but lost its exopeptidase activity.[9]

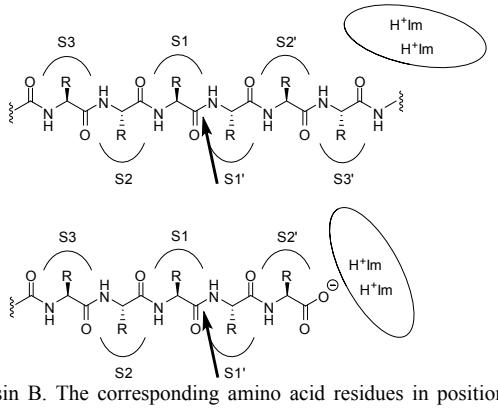


Figure 1. Binding of a peptidic substrate to cathepsin B. The corresponding amino acid residues in position P3, P2, P1, P1', P2', P3' occupy the subsites (specificity pockets) S3, S2, S1, S1', S2', S3'. The arrow indicates the scissile bond. H^+Im refers to the protonated histidine residues of the occluding loop. Top*:* Cathepsin B in its open conformation exhibits an endopeptidase activity. Bottom*:* Cathepsin B in its closed conformation with an exopeptidase (*i.e.* a dipeptidyl carboxypeptidase) activity.

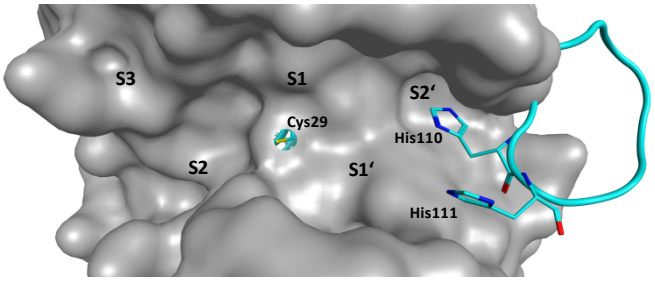


Figure 2. Crystallographic view on the active site cleft of human cathepsin B (PDB-ID: 1HUC). Surface representation is shown in grey. The occluding loop as ribbon diagram and selected residues are colored cyan.[8]

The exopeptidase activity is preferred in the acidic lysosomal compartments, whereas the endopeptidase activity is thought to predominate in the extracellular degradation of ECM components under less acidic conditions. Both activities of cathepsin B are considered to participate in tumor progression.[1,7,8,10]

Cysteine proteases cleave peptide bonds in the course of an acyl-transfer mechanism, initiated by the nucleophilic attack of the deprotonated active site cysteine residue, leading to an acyl-enzyme intermediate, which subsequently undergoes hydrolysis. Accordingly, the majority of peptidic inhibitors for cysteine cathepsins bear an electrophilic warhead, prone to covalently react with the sulfur of the active site cysteine, and a peptide portion for non-covalent interactions with the protease's binding pockets.

The design of new inhibitors for the therapeutically relevant protease cathepsin B relies on the detailed knowledge of the preferred interactions of the amino acid residues with the corresponding specificity pockets. Thus, mapping approaches for the active site provide substantial support to receive optimized low-molecular weight inhibitors. In case of chromogenic or fluorogenic peptidic substrates, the kinetic parameters of substrate consumption have been used to perform active site mappings of cathepsin B. Following this approach, an accelerated cleavage of defined peptides enables the identification of advantageous interactions between the amino acid residues and the specificity pockets. Although a variety of studies on the substrate specificity has been published, these data have not been reviewed so far. In our review, we summarize and discuss recent developments on the active site mapping of cathepsin B.

Moreover, inhibition data of small-molecule inhibitor series can provide insight into the interaction of residues with the corresponding subsites. In this case, a pronounced inhibition of the protease-catalyzed reaction arises from particular favorable interactions. Thus, we have also covered a selection of studies on cathepsin B inhibition by combinatorial inhibitor libraries, whose members possess the same warhead, but differ in distinct substructures. Herein, such data has been reviewed with respect to the active site mapping of cathepsin B.

2. Epoxide-derived Inhibitors of Cathepsin B

Hanada *et al*. isolated the epoxysuccinyl peptide E-64 (Fig. 3) from cultures of *Aspergillus japonicas* and determined first activities of this new inhibitor type against serine and cysteine proteases.[11] Later, inhibition of cysteine cathepsins B, H and L has been reported and the specificity of E-64-type epoxides for cysteine *versus* serine proteases elucidated.[12] Structure-activity relationships have been drawn from modifications at the N-terminal moieties (Fig. 3).

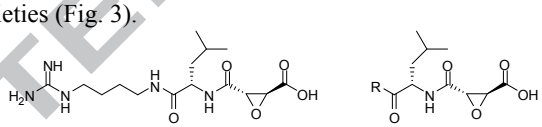


Figure 3. E-64 (left) and derived inhibitors (right). The cathepsin B assay was performed at pH 6.0 with Cbz-Phe-Arg-AMC as fluorogenic substrate. k_{obs}/[I] values in descending order for compounds with different residues R: 7-aminoheptylamino > 4-(benzyloxycarbonylamino)butylamino > 4-guanidinobutylamino > 7-aminoheptylamino > (*S*)-1-carboxy-3-methylbutylamino > OH. The two first noted derivatives showed a stronger cathepsin B inhibition than E-64.[12]

To achieve selectivity for cathepsin B, both termini of E-64 have been systematically modified.[13,14] Via alteration of the epoxysuccinyl carboxyl groups to various ester and amide moieties and exchange of the guanidinobutylamino group to a proline with a free carboxylic acid function, a series of selective inhibitors has been obtained. The most prominent of these carboxy terminal dipeptide-based compounds were CA-030 (*i.e. N*-((2*S*,3*S*)-3-(ethoxycarbonyl)oxirane-2-carbonyl)-isoleucyl-proline, Fig. 4, R = CO_2H, IC_{50} = 2.28 nM) and CA-074 (*i.e. N*-((2*S*,3*S*)-3-(propylcarbamoyl)oxirane-2-carbonyl)-isoleucyl-proline, Fig. 6; R = Pr, IC_{50} = 2.24 nM).

For cathepsin B inhibition, the free carboxylic group of CA-030 was mandatory (Fig. 4). This finding indicates a polar interaction with the occluding loop, as it has been later established by means of X-ray crystal structure analyses. The inhibitory activity of CA-030 towards cathepsin B could be somewhat improved by extending the alkyl part of the ester substructure (Fig. 5). Only minor changes in the potency were obtained with variation in the amide part of CA-074 (Fig. 6).

E-64 bound to papain revealed the covalent binding mode of epoxysuccinyl-based cysteine protease inhibitors.[15,16] The cocrystallization of human cathepsin B with CA-030 by Turk *et al.*[17] and that of bovine cathepsin B with E-64c and with CA-074 by Yamamoto *et al.*[18,19,20] confirmed the oxirane ring acting as the thiol-alkylating entity of the E-64-type inhibitors.[17,18] These structures provided first clues that the occluding loop can successfully be addressed with a terminal free carboxylic acid functionality to receive cathepsin B specificity.[21,22]

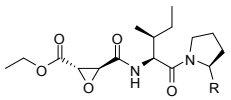


Figure 4. CA-030 type derivatives with modified proline moieties. The assay was performed at pH 5.5 with Cbz-Phe-Arg-AMC as fluorogenic substrate. pIC_{50} values in descending order for compounds with different residues R: carboxy >> H > methyloxycarbonyl > carbamoyl > hydroxymethyl.[13]

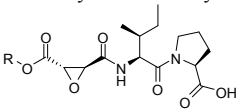


Figure 5. CA-030 type derivatives with modified ester moieties. The assay was performed at pH 5.5 with Cbz-Phe-Arg-AMC as fluorogenic substrate. pIC_{50} values in descending order for compounds with different residues R: cyclohexyl > isobutyl > isopropyl > ethyl > methyl > H.[13]

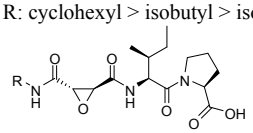


Figure 6. CA-074 type derivatives with modified amide moieties. The assay was performed at pH 5.5 with Cbz-Phe-Arg-AMC as fluorogenic substrate. pIC_{50} values in descending order for compounds with different residues R: isobutyl > cyclohexyl > *n*-propyl > *n*-butyl > isoamyl > *n*-amyl > *n*-hexyl > isopropyl > benzyl > ethyl > phenyl.[13]

After nucleophilic attack of the thiol group, a covalent thioether bond is formed between the active site Cys29 sulfur and one carbon of the oxirane moiety of CA-030 and CA-074, respectively. The nucleophilic attack occurred at the ring carbon next to the carbonyl carbon which is attached to the nitrogen of isoleucine. Instead of cleaving the scissile bond, the active site cysteine undergoes alkylation, thus converting a potential dipeptidyl carboxypeptidase substrate into a suicide inhibitor. At the same time, the ring opening is accompanied by the formation of a solvent exposed hydroxy group at the other carbon atom, resulting in a inversion of the configuration from (*S,S*) to (*R,R*) at the affected carbons (Fig. 7).

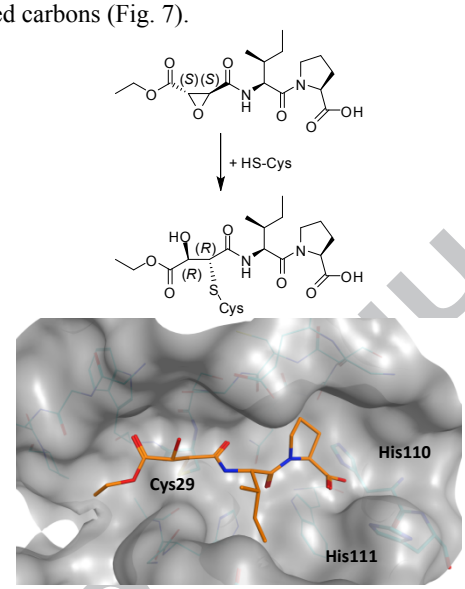


Figure 7. Reaction of CA-030 with cathepsin B. Formation of the thioether bond (top). Crystal structure of human cathepsin B with the cocrystallized covalent inhibitor shown in orange (bottom, PDB-ID: 1CSB).[17]

The binding modes of CA-030 (Fig. 7) and of CA-074 (Fig. 8) are similar. Most notably, the proline residues are consistently oriented towards the positively charged histidines of the occluding loop and their alkylidene chains occupy the S2' pocket. The isoleucine residue in P1' position forms van-der-Waals interactions with the hydrophobic amino acids Val176, Leu181 and Met196, comprising the S1' pocket.[19,20]

In contrast to the aforementioned binding modes, the backbone orientation of E-64c is reversed and the primed subsites remain unoccupied in the crystal structure of E-64c in complex with cathepsin B.[18] The terminal carboxyl group cannot display direct interactions with the occluding loop (Fig. 9). Instead, the closed conformation is stabilized by hydrogen bonds of cocrystallized water molecules to the occluding loop.

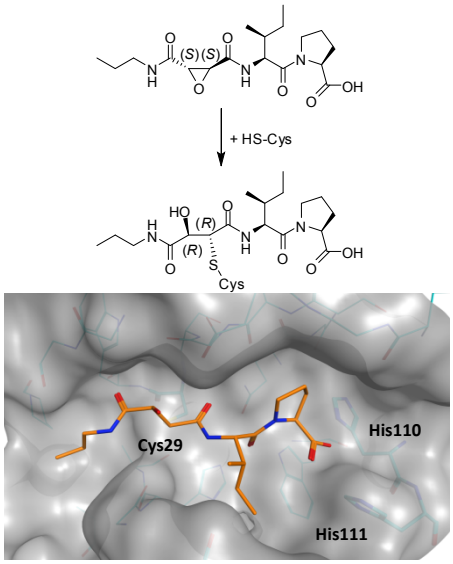


Figure 8. Reaction of CA-074 with cathepsin B (top). Crystal structure of bovine cathepsin B with the cocrystallized covalent inhibitor CA-074 (bottom, PDB-ID: 1QDQ).[19,20]

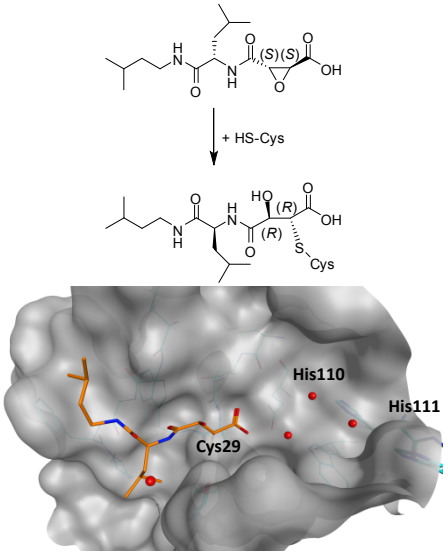


Figure 9. Reaction of E-64c with cathepsin B (top). Crystal structure of bovine cathepsin B with the cocrystallized covalent inhibitor (bottom, PDB-ID: 1ITO). Water molecules are illustrated as red spheres.[18]

The hydrogen bond network also includes the bound inhibitor whose carboxy moiety forms charge assisted hydrogen bonds with His110 and Trp221, as well as with Cys29 and Gln23. The conserved Gly74 residue addresses the N-terminal carbonyl oxygen of E-64c via a hydrogen bond. The S2-P2 interaction of E-64c is constituted by the accommodation of the leucine residue in the deep, hydrophobic S2 pocket, while the methylbutyl moiety performs hydrophobic interactions with Phe75 in the S3 subsite.[18]

Concerning the S2-P2 interaction, Giordano *et al.* synthesized a series of E-64c analogues bearing phenylalanine-derived residues in P2 position, which included mono- and diiodinated tyrosine (Fig. 10).[23] Kinetic evaluation at bovine cathepsin B has shown that iodination results in significantly enhanced potency, as reflected by k_{inac}/K_i values of 105,000 and 12,300 $M^{-1}s^{-1}$ for the compounds containing 3,5-diiodotyrosine and unmodified phenylalanine, respectively. The beneficial influence of tyrosine iodination in P2 on the inactivation of cathepsins B was also found for tyrosine-containing diazomethylketones in an earlier study.[24]

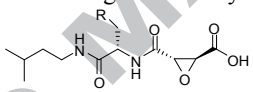


Figure 10. Influence of iodination of E-64c analogues containing phenylalanine derivatives in P2 on inactivation of cathepsin B. The assay was performed at pH 6.8 with Z-Gly-O-*para*-nitrophenyl as chromogenic substrate. k_{inac}/K_i values in descending order for compounds with different residues R: 4-hydroxy-3,5-diiodophenyl > 4-hydroxy-3-iodophenyl > 4-hydroxyphenyl > phenyl.[23]

The stereochemical requirements of the epoxysuccinyl moiety for optimal inactivation of human cathepsin B have been assessed by Schaschke *et al.*[25] Notably, E-64 derivatives with (*2R*,*3R*) configuration exhibited stronger inhibitory activity than the corresponding (*2S*,*3S*) diastereomers. In case of the Leu-Pro containing derivatives, the (*2S*,*3S*)-configured CA-030 analogue had a ten-fold lower second-order rate constant of inactivation (R = ethoxy, k_{inac}/K_i = 44,400 $M^{-1}s^{-1}$) than the corresponding (*2R*,*3R*)-configured compound (k_{inac}/K_i = 567,000 $M^{-1}s^{-1}$, Fig. 11). The introduction of the ornithyl-agmatine pseudo-dipeptide substructure provided the same result. Only the (*2S*,*3S*)-configured CA-074 analogue (R = *n*-propylamino, k_{inac}/K_i = 152,000 $M^{-1}s^{-1}$) was more potent than the (*2R*,*3R*)-configured diastereomer (k_{inac}/K_i = 29,400 $M^{-1}s^{-1}$).

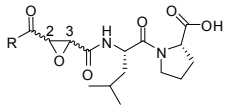
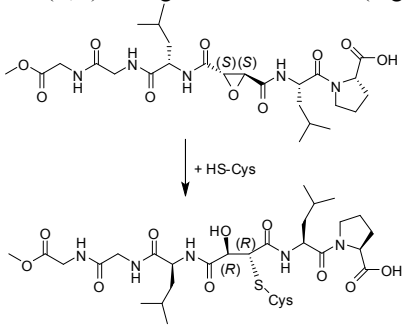


Figure 11. (*2R*,*3R*) and (*2S*,*3S*) epoxysuccinyl-peptides. The assay was performed at pH 5.5 with Cbz-Phe-Arg-AMC as fluorogenic substrate. Second-order rate constants of inactivation, k_{inac}/K_i, in descending order for compounds with different residues R and different configurations at the carbons C2 and C3: ethoxy (*2R*,*3R*) > HN=C(NH_2)NH(CH_2)$_4$NHCOCH((CH_2)$_3$$NH_2$)NH (*2S*,*3S*) > *n*-propylamino (*2S*,*3S*) > HN=C(NH_2)NH(CH_2)$_4$NHCOCH((CH_2)$_3$$NH_2$)NH (*2R*,*3R*) > ethoxy (*2S*,*3S*) > *n*-propylamino (*2R*,*3R*) > hydroxy (*2R*,*3R*).[25]

The high similarity of bovine and human cathepsin B was emphasized suggesting that binding of low-molecular weight inhibitors will lead to similar complexes.[26] The crystal structure of the complex of human cathepsin B with the two-headed epoxysuccinyl inhibitor NS-134 is shown in Figure 12. Compared to previous inhibitors, NS-134 extends to the whole active site cleft from the S4 to the S2' specificity pocket. The Leu-Gly-Gly-OMe pattern contributes to binding to the non-primed subsites while the Leu-Pro-OH pattern was maintained and occupies the primed specificity pockets. This extraordinary potent cathepsin B inhibitor had a second-order rate constant of inactivation of 1,520,000 $M^{-1}s^{-1}$; the (*R*,*R*)-configured diastereomer (Fig. 13) was 7-fold less active.[26,27]



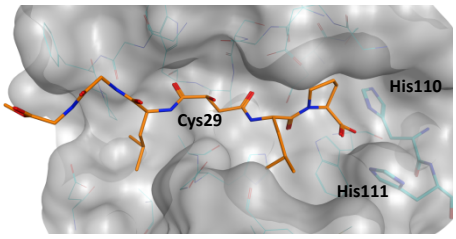


Figure 12. Reaction of NS-134 with cathepsin B (top). Crystal structure of human cathepsin B with the cocrystallized covalent inhibitor NS-134 (PDB-ID: 1SP4).[26]

To enable studies of cathepsin B activity in pathophysiological contexts, Schaschke *et al.* expanded the concept of double-headed epoxide inhibitors to *endo*-epoxysuccinyl peptides which bear a rhodamine-B or a biotin label (Fig. 13). These activity-based probes exhibited second-order rate constants of inactivation of about 1,600,000 $M^{-1}s^{-1}$, indicating that the functionalization of NS-134 did not affect the inhibitory properties.[28]

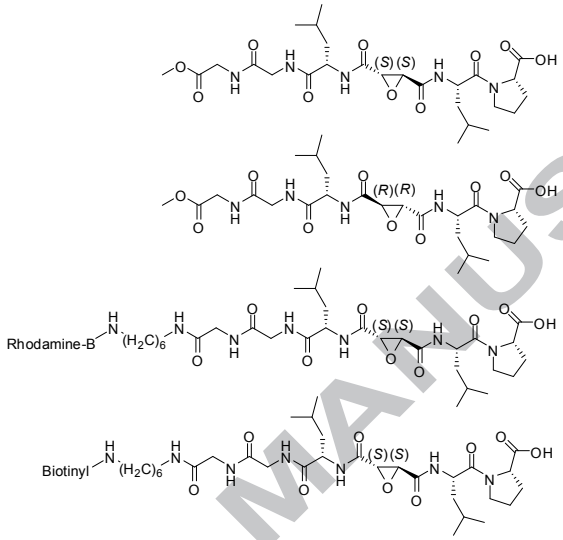


Figure 13. Two-headed epoxysuccinyl cathepsin B inhibitors and affinity labels. The assay was performed at pH 5.5 with Cbz-Phe-Arg-AMC as fluorogenic substrate.[27,28]

The effects of binding kinetics have been estimated by correlating crystallographic data of epoxysuccinyl derivatives with inhibitory activities. The isoleucine residue of the structures shown in Figure 14 is capable of interacting with the S1' subsite of cathepsin B. Adressing the S2' binding pocket with a second hydrophobic amino acid such as isoleucine resulted in a 100-fold higher activity compared to the derivative lacking the P2' amino acid (153,000 $M^{-1}s^{-1}$ *versus* 1,250 $M^{-1}s^{-1}$).[29]

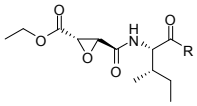


Figure 14. Epoxysuccinyl derivatives with modifications in P2' position. The assay was performed at pH 6.8 with Cbz-Phe-Arg-AMC as fluorogenic substrate. pIC_{50} values in descending order for compounds with different residues R: Ile-OH, Ala-OH > hydroxy.[29]

Peptidyl epoxides which contain the oxirane ring in place of the C-terminal carboxyl group (Fig. 15, top) have been described as cathepsin B inhibitors already in 1996.[30] Endo-peptidyl epoxides, on the other hand, facilitate interactions with both S and S' subsites of cathepsin B (Fig. 15, bottom). These structurally extended compounds did not show a mechanism-based irreversible inactivation, but behaved as reversible competitive inhibitors with improved inhibitory effects. The *cis*- exhibited a lower K_i value than the *trans*-derivative.[31]

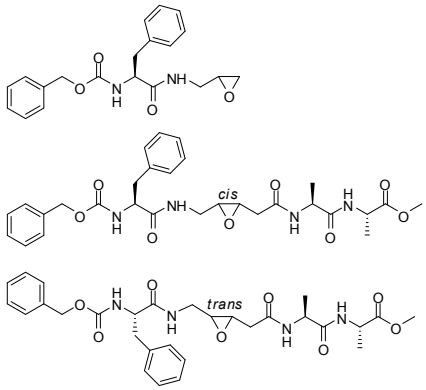


Figure 15. Non-succinyl epoxy-peptides.[30,31]

3. Active Site Mapping with Peptidic Substrates

Several approaches have been carried out to investigate substrate specificity of proteases.[32-35] In positional scanning synthetic combinatorial libraries (PS-SCL), one amino acid at the position under investigation has been continuously modified in the peptidic

structure and the kinetics of substrate cleavage has been determined. Hence, in such a positional profiling, each position is separately screened. PS-SCL allow for the real-time monitoring of proteolysis *in vitro* and *in vivo*. The panel of substrate sequences was successfully expanded by introducing unnatural amino acids, leading to hybride combinatorial substrate libraries (HyCoSuL) as has recently been applied to human cathepsin L.[36]

Fluorophore- and chromophore-based PS-SCL can be utilized to quantitatively define the non-prime specificity of the protease of interest. The inherent limitation to the analysis of the non-prime specificity pockets is attributed to the structure of such fluorogenic and chromogenic substrates that bear a C-terminal reporter group in place of a P1' amino acid. Internally quenched fluorescent (IQF) peptide substrates originate from Förster Resonance Energy Transfer (FRET) overcome this limitation and represent excellent tools for the active site mapping of both non-primed and primed specificity pockets. IQF substrates are assembled with a fluorophore and a quencher separated by several amino acids. Fluorescence is generated following the proteolytic cleavage between any peptide bond within the substrate. Several donor/acceptor pairs have been employed for the design of individual substrates and combinatorial libraries. Mass spectrometry is required to accurately validate the site of cleavage and to reconstruct the specificity profile.[32-34]

Several of such approaches based on PS-SCL and IQF substrates have been performed for the active site mapping of human cathepsin B. These studies on human enzymes are reviewed herein in chronological order. Preferred substrates will exhibit high values for the specificity constant, k_{cat}/K_m. In an early study of Ménard *et al*. applying an IQF library, Dns-Phe-Arg-X-Trp-Ala (Fig. 16, Dns = dansyl) was used as general substrate structure to investigate the S1' specificity of cysteine cathepsins.[37] This implies that the substrate is cleaved at the Arg-X peptide bond. For X, twelve different amino acids were introduced. In the unaffected substrate, the dansyl group with an absorption wavelength of 330 nm quenches the fluorescence of tryptophan with an emission wavelength range of 330-340 nm after excitation at 290 nm. When the substrate is cleaved between P1 and P1', the quenching of tryptophan is relieved resulting in a decreased emission of the dansyl fluorescence at 550 nm and an increase of the tryptophan fluorescence at 330-340 nm. The design of the substrates was undertaken by considering the general features of the active site of papain-like cysteine proteases. Accordingly, phenylalanine was introduced in P2 position. Arginine was chosen for the P1 position because it is poorly accepted in the S2 subsite of most cathepsins and, therefore, the binding mode of the substrate was secured. Alanine was placed in P3' position. The authors identified an acidic residue in P1' to be especially unfavorable for cathepsin B. Furthermore, a preference for large hydrophobic residues in the P1' position of a ligand was determined, *e.g*. phenylalanine, leucine, tyrosine, or tryptophan. Nevertheless, the S1' subsite of most cysteine proteases is not a primary site of specificity when compared to the S2 site.[37]

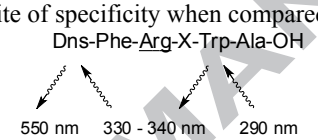


Figure 16. Design of substrates to determine the S1' subsite specificity of cathepsin B. Arg = P1 position. X = diverse amino acids introduced. The assays were performed at pH 6.0. k_{cat}/K_m in descending order: Leu, Phe, Tyr, Trp > Ala, Val, Ser > Glu, Asn, Gln, Asp, Lys.[37]

Brömme *et al*. discussed the S2 subsite specificity of cathepsin B.[38] In this study, an AMC-based PS-SCL was used, which, after hydrolytical amide cleavage, produce the fluorescent compound 7-amino-4-methylcoumarin (AMC, excitation wavelength 355-380 nm, emission wavelength 440-460 nm). In the S2 pocket of cathepsin B, a glutamic acid residue is present; therefore, a preference for basic residues was predicted.[8] The results in Figure 17 obtained with four different dipeptides showed, however, that the substrate with the non-basic phenylalanine in P2 position was kinetically most efficient. Its k_{cat}/K_m value was 460 $mM^{-1}s^{-1}$.[38] Semashko *et al.* obtained similar results.[39]

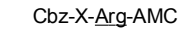


Figure 17. Substrates with different P2 amino acids. Arg = P1 position. X = diverse amino acids introduced. The assays were performed at pH 6.0. k_{cat}/K_m in descending order: Phe > Arg, Leu > Val.[38]

In a similar study by Taralp *et al*., tyrosine was identified as a preferred amino acid in P3 position, probably because of favorable aromatic-aromatic and hydrogen-bond interactions (Fig. 18).[40]

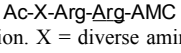


Figure 18. Substrates with different P3 amino acids. Arg = P1 position. X = diverse amino acids introduced. The assays were performed at pH 6.0. k_{cat}/K_m in descending order: Tyr > Val > Arg > Gly > Glu.[40]

By applying anthranilic acid and 2,4-dinitrobenzene as fluorophore-quencher pair, Portaro *et al*., used IQF peptide substrates derived from Abz-Ala-Ala-Phe-Arg-Ser-Ala-Gln-EDDnp (Fig. 19, Abz = 2-aminobenzoyl, EDDnp = *N*-(2,4-dinitrophenyl)ethylenediamine) and Abz-Ala-Phe-Arg-Ser-Ala-Ala-Gln-EDDnp, the former to study the specificities of S4 and S3, and the latter to investigate those of S2, S2' and S3'. The cleavage site of both peptides was expected to be between arginine and serine. In the employed peptide sequences, diverse amino acids were introduced for X to study cathepsin B subsite specificities.[41] The results are shown in Figure 20.

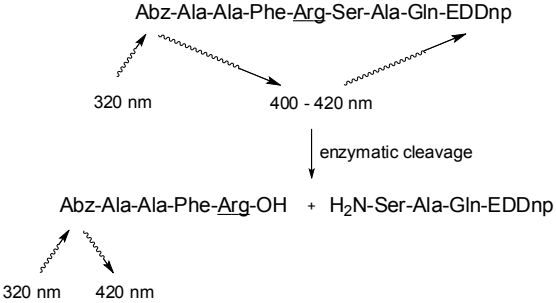


Figure 19. Principle of measurement with Abz-Ala-Ala-Phe-Arg-Ser-Ala-Gln-EDDnp.[41]

For investigations of the S4 subsite of cathepsin B, the substrate Abz-X-Ala-Phe-Arg-Ser-Ala-Gln-EDDnp was used. In the profile of cathepsin B, certain hydrophobic residues present in the P4 position of the substrate produced the highest specificity constants (k_{cat}/K_m ~ 40 $mM^{-1}s^{-1}$), but no significant differences were detected with basic, small neutral or acidic amino acid residues. It was assumed that this subsite is not involved in the substrate binding process.[42,43] However, the results of Portaro *et al*. did indicate that

substrate interactions with the S4 subsite were deducible and hydrophobic residues are accepted in this position.[41] The S3 subsite was elucidated with Abz-Ala-X-Phe-Arg-Ser-Ala-Gln-DDnp. A preference for hydrophobic residues was monitored. Substrate specificity for the S2 subsite, which is particularly important for selectivity,[44] was measured with Abz-Ala-X-Arg-Ser-Ala-Ala-Gln-EDDnp, where X was substituted by phenylalanine, leucine, alanine, proline or arginine. Cathepsin B accepted all amino acid residues with similar kinetic behavior. However, with proline in this position, the kinetic efficiency was weakened. Abz-Ala-Phe-Arg-Ser-X-Ala-Gln-EDDnp was used as substrate to determine the S2' specificity of cathepsin B. With asparagine in place of X, the highest k_{cat}/K_m value was figured out for cathepsin B (k_{cat}/K_m = 38 $mM^{-1}s^{-1}$); tryptophan, tyrosine and alanine were also well accepted in the S2' subsite. The lowest specificity constants were observed with histidine, lysine, aspartic acid, leucine and proline (for proline: k_{cat}/K_m < 5 $mM^{-1}s^{-1}$). S3' specificity was measured with Abz-Ala-Phe-Arg-Ser-Ala-X-Gln-EDDnp, but cathepsin B did not show a preference for a specific amino acid, indicating the P3' position not to be important for substrate recognition.[41]

Abz-X-Ala-Phe-Arg-Ser-Ala-Gln-EDDnp — to map the P4 position

k_{cat}/K_m: Leu, Phe, Trp > Ser, Asn, Asp, Ala > Tyr, His, Lys

Abz-Ala-X-Phe-Arg-Ser-Ala-Gln-EDDnp — to map the P3 position

k_{cat}/K_m: Leu, Tyr > Phe, Trp, Ala, Ser, Pro > His, Lys > Asn, Asp

Abz-Ala-X-Arg-Ser-Ala-Ala-Gln-EDDnp — to map the P2 position

k_{cat}/K_m: Phe, Leu > Arg > Ala > Pro

Abz-Ala-Phe-Arg-Ser-X-Ala-Gln-EDDnp — to map the P2' position

k_{cat}/K_m: Asn > Trp, Tyr, Ala > Ser, Phe > His, Lys > Asp, Leu, Pro

Abz-Ala-Phe-Arg-Ser-Ala-X-Gln-EDDnp — to map the P3' position

k_{cat}/K_m: Asn > Tyr, His, Leu, Trp > Ser, Phe, Pro, Lys > Asp, Ala

Figure 20. Design of substrates to determine the subsite specificities of the active site of cathepsin B. Arg = P1 position. X = diverse amino acids introduced. The assays were performed at pH 6.0. k_{cat}/K_m in descending order.[41]

To dissect the substrate specificity of cathepsin B with respect to the P1 positions, Del Nery *et al*. developed peptidyl-coumarin amides and quenched fluorogenic peptides with the general structures ε-H_2N-Cap-Leu-X-AMC (Cap = caproic acid) and Abz-Lys-Leu-X-Phe-Ser-Lys-Gln-EDDnp, respectively (Fig. 21 and 22). For X in P1 position, six amino acids were introduced. Peptidyl-coumarin amides with X being glycine, *O*-methyl-tyrosine or phenylalanine showed low k_{cat}/K_m values < 20 $mM^{-1}s^{-1}$, whereas *O*-benzyl-threonine exhibited the highest k_{cat}/K_m value of 1069 $mM^{-1}s^{-1}$. The IQF substrates (Fig. 22) were in general susceptible to cathepsin B, but with low k_{cat}/K_m values. Even arginine and *S*-benzyl-cysteine, the most appropriate amino acids in P1 position, showed k_{cat}/K_m values of only 13 $mM^{-1}s^{-1}$ and 10 $mM^{-1}s^{-1}$, respectively.[45]

ε-H_2N-Cap-Leu-X-AMC

Figure 21. Substrates to determine the P1 specificity. X = P1 position. X = diverse amino acids introduced. The assays were performed at pH 6.0. k_{cat}/K_m in descending order: Thr(*O*Bn) > Cys(*S*Bn) > Ser(*O*Bn) > Phe > Tyr(*O*Me) > Gly.[45]

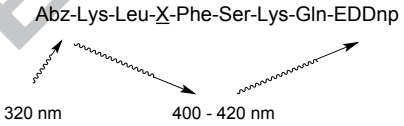


Figure 22. Design of quenched substrates to determine the P1 subsite specificity. X = P1 position. X = diverse amino acids introduced. The assays were performed at pH 6.0. k_{cat}/K_m in descending order: Arg > Cys(*S*Bn) > Ser(*O*Bn), Gly, Phe > Thr(*O*Bn).[45]

Melo *et al*. developed sensitive substrates for cysteine cathepsins with the sequence Abz-Phe-Arg-X-EDDnp to map the corresponding S1' subsites (Fig. 23).[46] For X, nine different amino acids were introduced. A clear preference of cathepsin B was found for hydrophobic amino acids in S1' position as reflected by the k_{cat}/K_m value of 828 $mM^{-1}s^{-1}$ for phenylalanine (for glutamine: k_{cat}/K_m = 20 $mM^{-1}s^{-1}$).

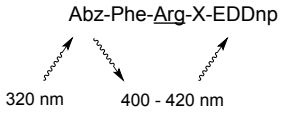


Figure 23. Substrates to map the S1' site. Arg = P1 position. X = diverse amino acids introduced. The assays were performed at pH 6.0. k_{cat}/K_m in descending order: Phe, Leu > Ile, Ala > Ser > Gln; Pro, Lys and Arg containing peptides were not hydrolyzed.[46]

Krupa *et al*. determined k_{cat}/K_m values for the cleavage of the IQF substrates Abz-Phe-Arg-Phe-(4-NO_2)-X-OH catalyzed by cathepsin B (Fig. 24).[47] For X, nine different amino acids were introduced to determine S2' subsite specificities. As control, the substrate lacking the P2' amino acid was also examined, showing that the k_{cat}/K_m value was two orders of magnitude lower than that of the most effective compound with tryptophan in P2'. The poorest substrate in this series was Abz-Phe-Arg-Phe-(4-NO_2)-Glu. In general, it was concluded that the S2' subsite of cathepsin B prefers large aromatic residues.

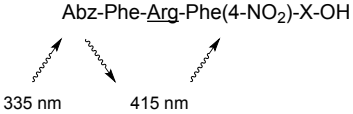


Figure 24. Substrates with different P2' amino acids. Arg = P1 position. X = diverse amino acids introduced. The assays were performed at pH 6.0. k_{cat}/K_m in descending order: Trp, Phe > Ser, Ala, Leu > Gln > Arg > Pro > Glu.[47]

The S2-S2' carboxydipeptidase subsite specificities were investigated by Cezari *et al*. with four series of internally quenched substrates with the parent structure Dnp-Gly-Phe-Arg-Phe-Trp-OH.[48] Its exemplary cleavage is shown in Figure 25 (Dnp = 2,4-dinitrophenyl). The substrates were designed in a way that each position was occupied by 15 different amino acids (Fig. 26). Aromatic amino acids were well accommodated in the S2 subsite. For phenylalanine, k_{cat}/K_m was 1725 $mM^{-1}s^{-1}$ and the substrate with tyrosine in

P2 revealed a k_{cat}/K_m of 1336 $mM^{-1}s^{-1}$. The S1 site accepted preferentially basic amino acids for hydrolysis. Arginine and lysine caused the highest k_{cat}/K_m values for this subsite, 1725 $mM^{-1}s^{-1}$ and 1285 $mM^{-1}s^{-1}$, respectively. The S1' site showed preferences for phenylalanine and serine, while the substrate with phenylalanine exhibited a k_{cat}/K_m value of 1725 $mM^{-1}s^{-1}$. When examining the S2' subsite, phenylalanine in P2' position showed the best k_{cat}/K_m value of 1254 $mM^{-1}s^{-1}$. All other second-order rate constants for the S2' substrates were lower than 180 $mM^{-1}s^{-1}$.

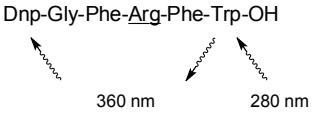


Figure 25. Principle of measurement with a quenched pentapeptide.[48]

Dnp-Gly-X-Arg-Phe-Trp-OH to map the P2 position

k_{cat}/K_m:
Phe > Tyr > Lys > Val > Ala > Ile > Ser, Pro > His > Leu > Arg > Asp, Glu
Substrates with Gly and Gln were not hydrolyzed.

Dnp-Gly-Phe-X-Phe-Trp-OH to map the P1 position

k_{cat}/K_m:
Arg > Lys > Phe > Leu > Val > His, Gln > Ser > Gly > Ala > Glu > Tyr > Ile, Asp
The substrate with Pro was not hydrolyzed.

Dnp-Gly-Phe-Arg-X-Trp-OH to map the P1' position

k_{cat}/K_m:
Phe > Ser > Leu > Ala > Ile > Tyr, Val > Gln > Lys > Gly > His, Asp > Arg > Glu
The substrate with Pro was not hydrolyzed.

Dnp-Gly-Phe-Arg-Trp-X-OH to map the P2' position

k_{cat}/K_m:
Phe > Tyr, Ile > Leu > Val > Arg > Lys, Pro > Gln, Ser > Asp, Glu, Ala
Substrates with His and Gly were not hydrolyzed.

Figure 26. Design of substrates to determine the subsite specificities of the active site of cathepsin B. Arg or X = P1 position. X = diverse amino acids introduced. The assays were performed at pH 6.0. k_{cat}/K_m in descending order.[48]

Further substrate mapping studies were performed by Cotrin *et al*. with IQF peptides of the general structure Abz-Gly-Mix-Mix-Arg-Mix-Lys(Dnp)-OH (Fig. 27).[49] The assay was performed at pH 4.5 and all of the peptides contained the free carboxylate of the P2' amino acid. Thus, this study was aimed at mapping the active site of cathepsin B in its closed conformation. In different sublibraries, positions P3, P2, P1 and P1' were occupied with one of 19 amino acids (except of cysteine) (Fig. 28). The Mix-positions represented randomly incorporated amino acid residues introduced by means of isokinetic mixtures.[50] Sublibraries with Abz-Gly-X-Mix-Arg-Mix-Lys(Dnp)-OH, Abz-Gly-Mix-X-Arg-Mix-Lys(Dnp)-OH, Abz-Gly-Mix-Mix-X-Mix-Lys(Dnp)-OH and Abz-Gly-Mix-Mix-Arg-X-Lys(Dnp)-OH were consulted to identify preferred residues for the S3-S1' subsites, respectively. For the P3 and P2 position, phenylalanine and valine were the privileged amino acids. The mapping approach confirmed arginine to be preferred at P1 position. The S1' subsite favored hydrophobic residues with either aromatic or aliphatic side chains. The peptide Abz-Gly-Ile-Val-Arg-Ala-Lys(Dnp)-OH was synthesized on the basis of the aforementioned results. It contained the favorable residues in P3-P1' positions and was the most efficient substrate described for cathepsin B in the course of the study (k_{cat}/K_m = 7288 $mM^{-1}s^{-1}$).[49]

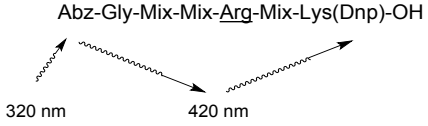


Figure 27. Principle of measurement with quenched hexapeptides.[49]

Abz-Gly-X-Mix-Arg-Mix-Lys(Dnp)-OH to map the P3 position

k_{cat}/Km:
Phe, Val > Ile, Ala, Lys, Tyr > Met, Leu > His, Arg, Ser
> Gln, Pro, Thr, Trp > Gly, Asp, Asn, Glu

Abz-Gly-Mix-X-Arg-Mix-Lys(Dnp)-OH to map the P2 position

k_{cat}/K_m:
Val, Phe > Tyr, Ala, Arg > Lys, Ser, Leu, His > Ile, Thr
> Asn, Met, Pro > Glu, Asp, Trp > Gln, Gly

Abz-Gly-Mix-Mix-X-Mix-Lys(Dnp)-OH to map the P1 position

k_{cat}/K_m:
Arg > Ser, Leu, Ala > Gly, Lys > His, Glu, Asp > Ile, Phe, Tyr, Gln
> Met, Val > Trp, Thr, Asn, Pro

Abz-Gly-Mix-Mix-Arg-X-Lys(Dnp)-OH to map the P1' position

k_{cat}/K_m:
Ala > Ile, Ser, Thr, Trp, Leu > Val, Arg, Tyr, Phe > Met, Asn, Lys
> Gly, Gln > His, Glu, Asp > Pro

Figure 28. Design of substrates to determine the subsite specificities of the active site of cathepsin B. Arg or X = P1 position. X = one fixed amino acid out of 19 proteinogenic amino acids, without Cys. Mix = isokinetic mixture of amino acids. The assays were performed at pH 4.5. k_{cat}/K_m in descending order.[49]

Stachowiak *et al*. synthesized series of peptide substrates specifically created for the carboxydipeptidase activity of cathepsin B.[51] The substrates were designed on the basis of the P5-P1 binding fragments of cystatin C and cystatin SA, which are natural, reversible inhibitors of papain-like cysteine proteases. In P1' and P2', amino acid residues were chosen according to the composition of the cathepsin B primed-binding sites. As fluorescence donor and acceptor groups, EDANS (5-((2-aminoethyl)amino)naphthalene-1-sulfonic acid) and DABCYL (4-(dimethylaminoazo)benzene-4-carboxylic acid) were introduced, the former via an amide bond with the side chain of an acidic amino acid, the latter via an amide bond with the N-terminal amino group of the peptide. Substrates were assayed fluorimetrically at pH 6.0 and 5.0. The results are shown in Figures 29 and 30. Position X-Y in the substrates in Figure 29 was occupied by five different dipeptides. For X in the substrates in Figure 30, six different di-, tri- or tetrapeptides were introduced.

Concerning the P1' and P2' positions, an efficient substrate identified to be used at both pH values was DABCYL-Arg-Leu-Val-Gly-Phe-Glu(EDANS)-OH (Fig. 29). When phenylalanine was replaced by β-(2-naphthyl)-alanine (Nal), k_{cat}/K_m slightly increased at pH 5.0 (k_{cat}/K_m = 122.9 $mM^{-1}s^{-1}$ *versus* k_{cat}/K_m = 140.0 $mM^{-1}s^{-1}$) and unexpectedly decreased at pH 6.0 (k_{cat}/K_m = 72.7 $mM^{-1}s^{-1}$ *versus* k_{cat}/K_m = 8.7 $mM^{-1}s^{-1}$).[51]

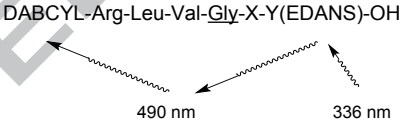


Figure 29. Quenched substrates to map the S1' and S2' sites. Gly = P1 position. X-Y = diverse dipeptides introduced. The assays were performed at pH 5.0 and 6.0. k_{cat}/K_m in descending order, pH 5.0: Nal-Glu > Phe-Glu > Phe-L-α-aminoadipic acid, Phe-Asp > Trp-Glu. k_{cat}/K_m in descending order, pH 6.0: Phe-L-α-aminoadipic acid > Phe-Glu > Phe-Asp > Nal-Glu > Trp-Glu.[51]

To analyze the P3-P2, P4-P2 and P5-P2 positions, substrates derived from DABCYL-Arg-Ile-Ile-Glu-Gly-Ile-Glu(EDANS)-OH, based on the amino acid sequence of cystatin SA, were synthesized. Isoleucine was placed in P1' position, deduced from the structure of inhibitor CA-074. The specificity constants of the substrates depicted in Figure 30 were generally smaller than 10 $mM^{-1}s^{-1}$; an exchange of the tetrapeptide motif Arg-Ile-Ile-Glu by Leu-Arg was advantageous. In the resulting substrate, DABCYL-Leu-Arg-Gly-Ile-Glu(EDANS)-OH, the replacement of isoleucine in P1' by phenylalanine increased the k_{cat}/K_m values as follows; at pH 6.0, k_{cat}/K_m for isoleucine was 7.1 $mM^{-1}s^{-1}$, for phenylalanine 24.9 $mM^{-1}s^{-1}$. At more acidic pH values, the k_{cat}/K_m value for isoleucine was 5.7 $mM^{-1}s^{-1}$ at pH 4.5 and for phenylalanine 51.2 $mM^{-1}s^{-1}$ at pH 5.0.[51]

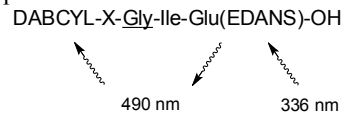


Figure 30. IQF substrates to map the S3-S2, S4-S2 and S5-S2 sites. sites. Gly = P1 position. X = diverse di-, tri- or tetrapeptides introduced. The assays were performed at pH 4.5 and 6.0. k_{cat}/K_m in descending order, pH 4.5: Leu-Arg > Leu-Glu > Arg-Leu-Glu > Ile-Glu > Arg-Ile-Ile-Glu > Arg-Ile-Glu. k_{cat}/K_m in descending order, pH 6.0: Leu-Arg > Arg-Leu-Glu > Arg-Ile-Glu. Arg-Ile-Ile-Glu, Ile-Glu and Leu-Glu were not hydrolyzed at pH 6.0.[51]

Ruzza *et al*. dissected the P2 substrate specificity of cathepsin B with seven quenched fluorescent decapeptides (Fig. 31).[52] The peptides differ in position X of the common structure Nma-Orn-Ala-Gly-Arg-Arg-X-Ala-Lys(Dnp)-D-Ala-Ala-OH (Nma = *N*-methylanthranoyl, Orn = ornithine) and were examined at pH 5.0 and pH 7.4 to investigate the exo- and endoproteolytic activity of cathepsin B, respectively. However, with alanine in P1 position, such substrates would rather address a carboxytri- than a carboxydipeptidase activity. At pH 7.4, only the substrates with alanine and phenylalanine in P2 were hydrolyzed by the enzyme, the highest k_{cat}/K_m value for the alanine containing substrate was 327.2 $mM^{-1}s^{-1}$. At pH 5.0, all substrates were hydrolyzed, the one with alanine in P2 position exhibited a k_{cat}/K_m value of 635.8 $mM^{-1}s^{-1}$, while the other specificity constants were reported to be less than 10 $mM^{-1}s^{-1}$.[52]

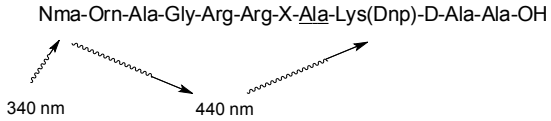


Figure 31. Substrates to determine the P2 subsite specificity of cathepsin B. Ala = P1 position. X = diverse amino acids introduced. The assays were performed at pH 5.0 and 7.4. k_{cat}/K_m in descending order, pH 5.0: Ala > Phe > Lys > Glu, Val > Asn; β-Ala was not hydrolyzed. k_{cat}/K_m in descending order, pH 7.4: Ala > Phe; Lys, Glu, Val, β-Ala and Asn were not hydrolyzed.[52]

A report on the preparation and analysis of 7-amino-4-carbamoylmethylcoumarin (ACC) containing tetrapeptides was provided by Choe *et al*.[53] The PS-SCL approach was applied to dissect the P4-P1 substrate specificity of cathepsin B. In these peptides, the P4, P3, P2 or P1 position was fixed with one of 20 amino acids, where cysteine was replaced by norleucine (Nle). The other positions contained an isokinetic amino acid mixture. Each library had 20 sublibraries, all together, 160,000 tetrapeptidic substrates were investigated. The ACC-substrates had an absorption maximum of 325 nm and an emission maximum of 400 nm. When the substrate was cleaved by the enzyme, the maxima of ACC shifted to 350 nm and 450 nm, respectively. The assays were monitored fluorimetrically with excitation at 380 nm and emission at 460 nm. Cathepsin B displayed a preference for hydrophobic amino acids at the P2 position but showed a rather broad P2 specificity. Arginine and lysine were strongly preferred in P1 and broad specificities were identified for P3 and P4 (Fig. 32). Specificity constants for two individual tetrapeptide substrates, Ac-His-Arg-Leu-Arg-ACC and Ac-His-Arg-Tyr-Arg-ACC, were reported to be 15 $mM^{-1}s^{-1}$ and 7.0 $mM^{-1}s^{-1}$, respectively. Noteworthy, among dipeptide substrates, the N-terminal capping groups influenced the substrate properties and Cbz-Phe-Arg-AMC (k_{cat}/K_m = 190 $mM^{-1}s^{-1}$) was much more appropriate than Ac-Phe-Arg-ACC (k_{cat}/K_m = 4.3 $mM^{-1}s^{-1}$).

Ac-X-Mix-Mix-Mix-ACC — to map the P4 position

k_{cat}/K_m:
Arg > His > Lys > Met, Thr, Ser, Gly, Gln, Nle, Asn, Ala, Val, Ile, Leu, Glu, Trp, Tyr, Phe, Pro > Asp

Ac-Mix-X-Mix-Mix-ACC — to map the P3 position

k_{cat}/K_m:
Nle > Leu > Lys, Met > Arg > Tyr, Pro, Phe, Trp, Ala, Val, Ile > His, Glu, Gln, Ser, Thr, Gly > Asn, Asp

Ac-Mix-Mix-X-Mix-ACC — to map the P2 position

k_{cat}/K_m:
Val > Met, Nle > Lys, Leu, Arg, Ser, Gln, Ala > Ile, Asp, Phe, His, Tyr > Glu, Asn, Thr, Gly, Trp, Pro

Ac-Mix-Mix-Mix-X-ACC — to map the P1 position

k_{cat}/K_m:
Arg, Lys > Gln, Nle, Met > Leu, Thr > Trp, Glu > Ser, Ala, His > Tyr, Gly, Phe > Asn, Pro, Asp, Val, Ile

Figure 32. Sublibraries of cathepsin B substrates. X or Mix = P1 position. X = diverse amino acids introduced. Mix = isokinetic mixture of amino acids. The assays were performed at pH 5.5. k_{cat}/K_m in descending order.[53]

Kim *et al*. developed sensitive dityrosine-based substrates for the selective assay of cathepsin B (Fig. 33).[54] Dityrosine is an unusual bulky amino acid formed from two L-tyrosine molecules. Its excitation wavelength is 320 nm at alkaline conditions and 284 nm at acidic conditions. As quencher, isoniazid was introduced in the molecules. Two ligands with the sequence DBDY-(X-INH)$_2$ (Fig. 33, DBDY = N,N'-diBoc-dityrosine, INH = isoniazid, isonicotinic acid hydrazide) were synthesized to map the S1 site and tested them for selective hydrolysis by cathepsins B. The substrates were assembled by connecting two equal amino acids (Gly or Lys) through two peptide bonds to both carboxylic groups of the N,N'-protected dityrosine. The carboxylic groups of the amino acids are linked to the terminal NH_2 group of INH. The enzyme hydrolyzed DBDY-(Gly-INH)$_2$ with a k_{cat}/K_m value of 3.87 $mM^{-1}s^{-1}$ and DBDY-(Lys-INH)$_2$ with a k_{cat}/K_m value of 1.80 $mM^{-1}s^{-1}$.[54]

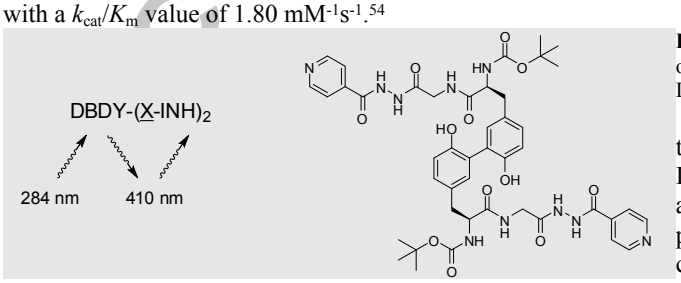


Figure 33. Dityrosine-derived substrates for cathepsin B. X = P1 position. X = Gly or Lys. The assays were performed at pH 6.0. k_{cat}/K_m in descending order: Gly > Lys (left). Structure of the glycine-containing substrate (right).[54]

As exemplified below, cathepsin B has extensively been used in targeted drug release because of its overexpression in various tumors. In an effort to develop an enzyme-triggered, tumor-specific release of an antibody-drug conjugate payload, Wei *et al.* identified peptidomimetic substrates and determined kinetic parameters for the cathepsin B-catalyzed cleavage, leading to the formation of norfloxacin.[55] A series of compounds possessing a citrulline residue in P1 and a cyclobutyl ring in P2 position were optimized by variation of the P3 moieties. K_m values for these substrates between 10 and 100 μM were obtained, with a remarkable impact of the stereochemistry of the thien-2-yl residue (Fig. 34).

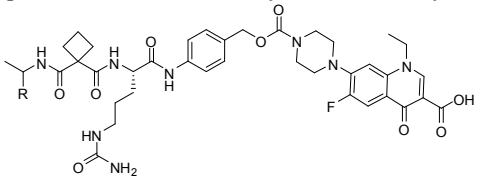


Figure 34. Cathepsin B catalyzed cleavage of the anilide bond in peptide-drug conjugates leading to the sequential release of norfloxacin. The assay was performed at pH 6.0. Norfloxacin formation was quantified by a LC-MS multiple reaction monitoring method. V_{max}/K_m in descending order for substrates with different residues R: thien-2-yl (the active diastereomer of unknown stereochemistry) > (*R*)-phenyl > thien-2-yl (the other diastereomer), (*R*)-isopropyl, H.[55]

As discussed in detail in the next chapter, peptidic nitrile-based inhibitors can be designed from an appropriate substrate. One of the malondiamide-type substrates (Fig. 34) was employed for such an approach and the corresponding nitrile inhibitor was subjected to crystallization experiments (Fig. 35). The electron density was not sufficient to determine the coordinates of the citrulline moiety and the terminal methyl group, but confirmed the covalent thioimidate linkage between the active site cysteine and the inhibitor. The malondiamide backbone contributes to an orientation of the cyclobutyl ring towards the S2 pocket.[55]

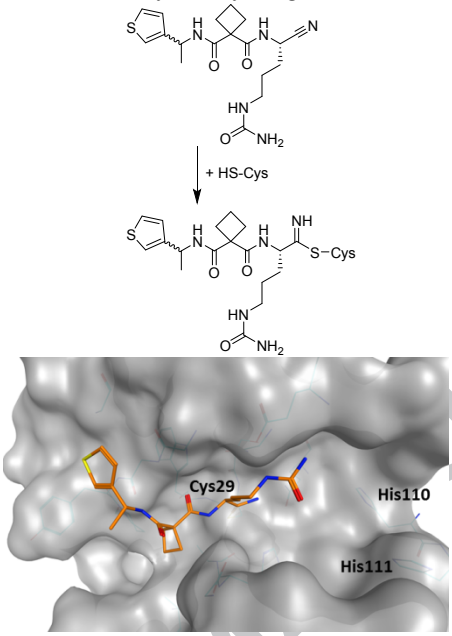


Figure 35. Formation of the thioimidate (top). Crystal structure of human cathepsin B with a covalently bound peptidomimetic nitrile inhibitor. (bottom, PDB-ID: 6AY2).[55]

4. Active Site Mapping with Inhibitors

Peptide substrates can be converted into protease inhibitors through coupling to an electrophilic warhead. Thus, besides the analysis of the consumption of different substrates, peptidic or peptidomimetic inhibitors can be used for an active site mapping. Typically, such approaches with electrophile-based libraries are limited to the non-prime specificity profiling.[34] Greenbaum *et al.* used libraries of peptidic epoxides and combined chemical- and computational-based methods to generate and analyze affinity fingerprints for the papain family of cysteine proteases.[56] Based on the epoxide electrophile scaffold of E-64, three positional scanning libraries were synthesized by fixing each of the P2, P3, and P4 positions with each of the 20 possible natural amino acids, while variable positions were composed of a mixture of all amino acids. These libraries were screened in a competition assay for cysteine proteases and subsequent computational cluster analysis of inhibition data revealed patterns for binding pocket specificity. For cathepsin B, a strong preference in P2 position for leucine, isoleucine and valine with competition values greater than 60% were observed.

Besides epoxides, several other warhead types have been employed for the cathepsin inhibitor development. The following data determined with human cathepsin B will be presented by using general structures with only one variable position in order to allow for conclusions with respect to the interaction of the inhibitors with the corresponding subsites.

Peptide nitriles are prototypic inhibitors for cysteine cathepsins.[57-59] The interaction of cathepsins with peptide nitriles involves a nucleophilic attack of the active site cysteine at the nitrile carbon leading to the reversible formation of a covalent thioimidate adduct. To perform mapping studies, the enzymatic cleavage of an appropriate substrate in the presence of different inhibitor concentrations has to be recorded and the rates to be plotted versus the inhibitor concentrations. From the so obtained IC_{50} values, K_i values can be calculated. Typically, peptide nitriles show a kinetic behavior of fast-binding inhibitors and linear progress curves have been obtained. In the following, we summarize cathepsin B inhibition data generated with series of peptide nitriles. All equipped with the same warhead, the different K_i values result from differently pronounced non-covalent interactions of their residues with the cathepsin's specificity pockets. Hence, an active site mapping can be carried out on the basis of the structure-activity relationships drawn from such libraries of peptidic inhibitors.

Cathepsin B inhibition of a first series of dipeptide nitriles with aminoacetonitrile at P1, different amino acids at P2 position and a fixed Cbz capping group is shown in Figure 36. The introduction of β-(2-naphthyl)alanine furnished the most potent inhibitor with a K_i value of 3 µM, while the other compounds had double-digit micromolar inhibition constants. Fluorogenic and chromogenic substrates producing AMC and *para*-nitroaniline (pNA), respectively, were used.[60,61]

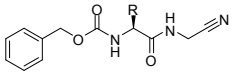


Figure 36. Inhibition of cathepsin B by Cbc-protected dipeptide nitriles. The assays were performed at pH 6.0 with Boc-Leu-Lys-Arg-AMC and Cbz-Arg-Arg-pNA as fluorogenic and chromogenic substrate, respectively. pK_i values in descending order for compounds with different residues R: 2-naphthylmethyl > 4-hydroxybenzyl, benzyl > 4-methoxybenzyl > 2,2-dimethylpropyl, (*S*)-2-butyl > H.[60,61]

An extended series of Boc-protected dipeptide nitriles with variations within the P2 position (Fig. 37) revealed (again) β-(2-naphthyl)alanine (K_i = 3.46 µM) and 3-bromo-phenylalanine (K_i = 5.17 µM) to be the most advantageous amino acids in this position.[61-64] The K_i value of the former inhibitor increased to 240 µM, when the isomeric amino acid β-(1-naphthyl)alanine was introduced. The

position of the bromo substituent at the benzene ring of phenylalanine derivatives was also critical, K_i = 5.17 µM (3-bromo-), K_i = 52 µM (4-bromo-), K_i > 500 µM (2-bromophenylalanine containing inhibitors). As in the aforementioned series of Cbz-protected compounds, the dipeptide nitrile with glycine at P2 position lacked affinity for cathepsin B, indicating the strong contribution of the P2-S2 interaction for the inhibitor's affinity.

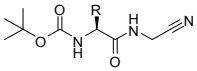


Figure 37. Inhibition of cathepsin B by Boc-protected dipeptide nitriles. The assay was performed at pH 6.0 with Cbz-Arg-Arg-pNA as chromogenic substrate. pK_i values in descending order for compounds with different residues R: 2-naphthylmethyl > 3-bromobenzyl > 3-phenylbenzyl > 4-hydroxybenzyl > (1*H*-indol-3-yl)methyl > *n*-propyl > (furan-2-yl)methyl > 2-thienylmethyl > benzyl > ethyl > cyclohexylmethyl > cyclopropylmethyl > *n*-butyl, 4-bromobenzyl, phenethyl > 2-(methylthio)ethyl > isopropyl > 3-thienylmethyl > (*S*)-2-butyl > 4-fluorobenzyl > isobutyl > benzyloxymethyl > 1-naphthylmethyl > methyl > (biphenyl-4-yl)methyl > phenyl, 2-bromobenzyl, H.[61]

One of the structural variations realized by Greenspan *et al*.[65] concerned the P2 position, while the diphenylacetyl capping group was maintained. The racemic compounds with this substitution pattern are included in Figure 38. Such compounds are generally more potent than analogs with a Cbz or Boc capping group. The combination of a diphenylacetyl moiety which occupies the S3 subsite with the 3,5-dimethylphenylalanine residue to be accommodated in the S2 pocket led to an IC_{50} value of 11.9 nM. Although the bioactivity of the other listed compounds was gradually attenuated, they all possessed nanomolar IC_{50} values. A comparison of IC_{50} values of enantiopure dipeptide nitriles with (*S*)-phenylalanine in P2 and aminoacetonitrile in P1 position showed the advantage of the diphenylacetyl over the Cbz capping group (496 nM versus 62,000 nM).[56]

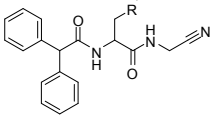


Figure 38. Inhibitors of cathepsin B with an N-terminal diphenylacetyl group. The assay was performed at pH 5.8 with Cbz-Arg-Arg-AMC as fluorogenic substrate. pIC_{50} in descending order for compounds with different residues R: 3,5-dimethylphenyl > 3-methylphenyl > 3-iodophenyl > 3-chlorophenyl > 3-ethylphenyl > 3-(5-methylpyridyl).[56]

The binding mode of this chemotype of dipeptide nitriles was also confirmed by X-ray crystallographic data. A complex of a diphenylacetylated inhibitor with 3-methylphenylalanine in P2 and aminoacetonitrile in P1 position is exemplarily shown in Figure 39. The attack of the active site cysteine sulfur at the electrophilic nitrile carbon produced a covalent thioimidate bond between the protease and the inhibitor

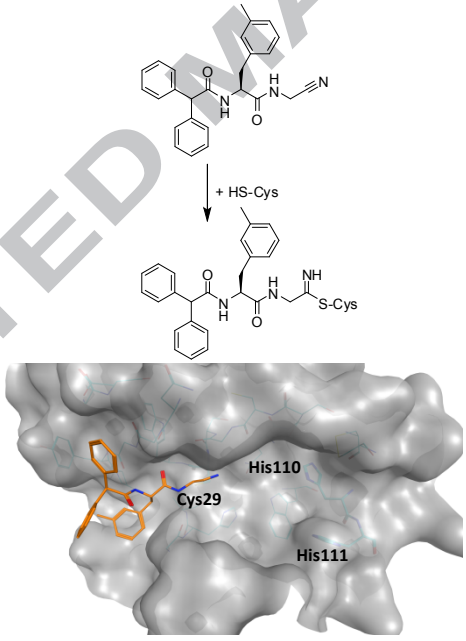


Figure 39. Thioether bond formation (top). Crystal structure of human cathepsin B with a dipeptide nitrile (bottom, PDB-ID: 1GMY).[65]

Palmer *et al.* and Falgueyret *et al*. reported on the cathepsin inhibiting activity of dipeptide nitriles comprising the amino acid homocycloleucine in P2 position (Fig. 40).[60,66] This building block was shown to be favorable for an inhibition of human cathepsin K and was incorporated into the cathepsin K inhibitor balicatib (R = 4-(4-propylpiperazin-1-yl)phenyl). Nevertheless, as kinetic data for the inhibition of cathepsin B have also been obtained, they can be used for a cathepsin B active site mapping. Large tri-ring or two-ring capping groups were found to be best suited for cathepsin B inhibition. For example, the first listed compound whose structure was terminated with the 4-(2-(4-methylpiperazin-1-yl)thiazol-4-yl)benzoic acid building block had a K_i value of ca 600 nM.[66] Starting from compounds shown in Figure 40, two hydrogen atoms at the 2-position of the cyclohexane ring were exchanged for two geminal fluorine substituents. The enantiomers of the resulting chiral difluoro derivatives were separated and their biological activity towards four cathepsins was determined. Noteworthy, the fluorinated eutomers displayed lower K_i values against cathepsin B than the corresponding non-fluorinated analogs. Accordingly, a fluorophilic property of the S2 pocket of cathepsin B could be concluded.[67]

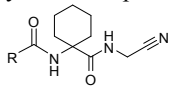


Figure 40. Homocycloleucine-derived inhibitors. The assays were performed at pH 6.0 with Boc-Leu-Lys-Arg-AMC as fluorogenic substrate. pK_i values in descending order for compounds with different residues R: (4-(2-(4-methylpiperazin-1-yl)thiazol-4-yl)phenyl), 4'-(dimethylamino)biphenyl-4-yl, biphenyl-4-yl >

4-bromophenyl, 4-(2-morpholinothiazole-4-yl)phenyl, 4-(4-methylpiperazin-1-yl)phenyl, 4-(4-propylpiperazin-1-yl)phenyl, 2-bromothienyl-5-yl > 4-(dimethylamino)phenyl > 3-bromophenyl, biphenyl-3-yl, 4-morpholinophenyl > 4-(2-(pyridin-4-yl)thiazol)-yl, 4-aminophenyl > phenylethyl.[60,66]

A mapping of the P3 specificity pocket could also be performed based on the data of dipeptide nitriles with leucine in P2 position (Fig. 41). This series included the biphenyl derivative (K_i = 1.79 µM) which was somewhat more potent than the next listed tri-ring compound (K_i = 2.89 µM). A replacement of the amide- by a urea linker (R = 4-[5-(2-thienyl)-1,2,4-oxadiazol-3-yl]phenylamino) was disadvantageous for cathepsin B inhibition.[61,64]

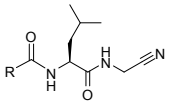


Figure 41. Leucine-derived inhibitors. The assay was performed at pH 6.0 with Cbz-Arg-Arg-pNA as chromogenic substrate. pK_i values in descending order for compounds with different residues R: biphenyl-4-yl > 4-[5-(2-thienyl)-1,2,4-oxadiazol-3-yl]phenyl > 4-[5-(2-thienyl)-1,2,4-oxadiazol-3-yl]phenylamino, benzyloxy > *tert*-butyloxy.[61]

Greenspan *et al.* expanded the dipeptide nitrile structure by a *meta*- (Fig. 42) or *para*-benzoic acid motif, which were linker-connected to the α-carbon of the P1 aminonitrile.[65,68] The carboxylate moiety is capable of interacting with the positively charged histidine residues of the occluding loop. This design successfully provided highly potent and selective cathepsin B inhibitors. At P2 position, 3-methylphenylalanine was introduced.

Noteworthy, the assay to determine the kinetic parameters was performed with Cbz-Arg-Arg-AMC as fluorogenic substrate. Cbz-Arg-Arg-AMC is a 'small' substrate which binds exclusively to the S3-S1 region of the active site of cathepsin B. It has been referred that the use of 'small' substrates caused a preferred endopeptidase activity executed by the protease in the so called open conformation.[13,61,69-71] However, such 'small' substrates are not able to prevent the formation of the closed conformation as the occluding loop is a highly flexible element that can easily rearrange. The inhibitors synthesized by Greenspan *et al*. possess a free carboxylate which is assumed to interact with the histidines of the occluding loop in the closed conformation. To determine the exopeptidase activity of cathepsin B, it is advisable to select true exopeptidase substrates *e.g.* Abz-Gly-Ile-Val-Arg-Ala-Lys(Dnp)-OH,[69] that occupy the S1' and S2' binding region and possess a C-terminal carboxylate to form and stabilize the closed conformation. Accordingly, Cbz-Arg-Arg-AMC was also mentioned not to be an appropriate model substrate to exclusively study the endo- or exoproteolysis of cathepsin B.[70,71]

The different aromatic residues in the listed compounds (Fig. 42) caused IC_{50} values of less than 10 nM. The most potent derivative again (see also Fig. 38) contained the diphenylacetic acid building block and showed an IC_{50} value of 1.8 nM. The exchange of the linker oxygen by a methylene group reduced the inhibitory activity 10 to 15-fold (R = diphenylmethyl, 2,4-difluorophenyl).

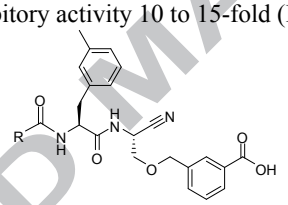


Figure 42. Dipeptidyl nitrile inhibitors**.** The assays were performed at pH 5.8 with Cbz-Arg-Arg-AMC as a fluorogenic substrate. pIC_{50} values in descending order for compounds with different residues R: diphenylmethyl > 2-fluoro-4-chlorophenyl > 2,4-difluorophenyl > phenyl > *n*-butyl.[65]

Inhibitors presented in Figure 43 were prepared to study the endo- and exopeptidase activity of cathepsin B.[61] A mapping approach of cathepsin B was performed to optimize the inhibitor structure. Finally, biphenyl-4-carboxylic acid in P3 and 3-bromophenylalanine in P2 were introduced in the peptidic scaffold. In P1 position, a linker structure assembled by click-chemistry with a C-terminal free carboxylate was placed to form salt bridges with the occluding loop's histidines. The compound with the oxygen-containing linker structure and the acetic acid-substituted triazole ring (m = 1, n = 0) was the most potent inhibitor in this study with a K_i value of 41.3 nM. When shortening the linker by omitting the oxygen bridge, the inhibitors clearly lost affinity (m = 0, n = 1 and m = n = 0). The pH dependency of cathepsin B activity in the presence of the depicted inhibitors was investigated at pH 4.5 with Cbz-Arg-Arg-AMC as substrate (m = 1, n = 0; K_i = 27.3 nM). When maintaining pH 4.5 and using the true exopeptidase substrate Abz-Gly-Ile-Val-Arg-Ala-Lys(Dnp)-OH, further reduced K_i values were obtained (m = 1, n = 0; K_i = 19.2 nM). These data illustrate that the inhibitory potency of such inhibitors toward cathepsin B increased when assays conditions were used which favor the exopeptidase activity of the targeted enzyme. Then, the closed conformation of the protease is stabilized and the inhibitor's carboxylate groups can beneficially interact with the protonated histidines of the occluding loop.

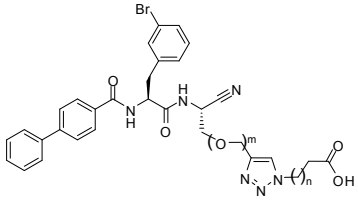


Figure 43. Dipeptidyl nitrile inhibitors with varying linker length**.** The assays were performed at pH 6.0 with Cbz-Arg-Arg-pNA as a chromogenic substrate. pK_i values in descending order for compounds with varying linker lengths m and n: m = 1, n = 0 > m = 0, n = 1 > m = n = 0.[61]

The chemotype of azadipeptide nitriles was developed by exchanging the α-CH by nitrogen.[72] Azadipeptide nitriles show time-dependent cathepsin inhibition. Progress curves can be analyzed by means of the slow-binding equation. Some representatives with a Cbz-capping group have been investigated towards cathepsin B. The comparison of those compounds, which are thought to occupy the S2 pocket with the residue R, is outlined in Figure 44.[64,73,74] The corresponding tyrosine and leucine derivatives exhibited K_i values of around 0.4 nM. The former inhibitor was *O*-fluoroethylated to obtain a non-radioactive reference compound for an ^{18}F-labeled analog.[74]

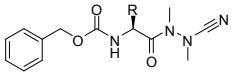


Figure 44. Azadipeptide nitriles with Cbz-capping group. The assay was performed at pH 6.0 with Cbz-Arg-Arg-pNA as a chromogenic substrate. pK_i values in descending order for compounds with different residues R: 4-hydroxybenzyl, isobutyl > 4-(2-fluoroethoxy)benzyl > isobutylsulfonylmethyl.[64,73,74]

The introduction of homocycloleucine at P2 position of three azadipeptide nitriles resulted in inhibitors with similar, relatively high K_i values between 140 and 170 nM (Figure 45). One of these derivatives bears a fluorescent coumarin as P3 capping group.[63,75]

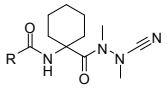


Figure 45. Azadipeptide nitriles with homocycloleucine. The assay was performed at pH 6.0 with Cbz-Arg-Arg-pNA as a chromogenic substrate. pK_i values in descending order for compounds with different residues R: 7-(diethylamino)coumarin-3-yl > 4-[5-(2-thienyl)-1,2,4-oxadiazol-3-yl]phenyl > benzyloxy.[63,75]

A larger series of structurally analogous azadipeptide nitriles which share leucine at P2 position and bear different P3 capping groups has been investigated. The comparison of those groups, which are thought to occupy the S3 pocket, is outlined in Figure 46. Although the affinity for cathepsin B was weaker than for other human cysteine cathepsins, a mapping of the S3 pocket was possible. The first compound mentioned in Figure 46 had a K_i value for cathepsin B inhibition of 0.36 nM, followed by the Cbz-protected compound with a K_i value of 0.43 nM, whereas the urea derivatives were less active with K_i values up to 2.8 nM.[64]

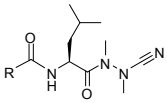


Figure 46. Azadipeptide nitriles with with different P3 capping groups. The assay was performed at pH 6.0 with Cbz-Arg-Arg-pNA as a chromogenic substrate. pK_i values in descending order for compounds with different residues R: 4-[5-(2-thienyl)-1,2,4-oxadiazol-3-yl]phenyl > benzyloxy > phenylamino > benzylamino, 4-[5-(2-thienyl)-1,2,4-oxadiazol-3-yl]benzylamino > 4-[5-methyl-1,2,4-oxadiazol-3-yl]phenylamino, 4-[5-(2-thienyl)-1,2,4-oxadiazol-3-yl]phenylamino > 4-[5-methyl-1,2,4-oxadiazol-3-yl]benzylamino.[64]

Ren *et al.* reported on racemic cathepsin K inhibitors of the same chemotype,[76] which have been investigated for their cathepsin B inhibitory capability (Fig. 47). The introduction of *p*-terphenyl-4-carboxylic acid as N-terminal capping group resulted in the strongest cathepsin B inhibition (K_i = 0.69 nM). The other aromatic acyl groups caused azadipeptide nitriles with K_i values between 2 and 5 nM.

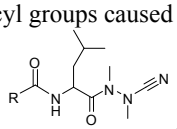


Figure 47. Azadipeptide nitriles with with different P3 capping groups. The assay was performed at pH 6.0 with Cbz-Arg-Arg-AMC as a fluorogenic substrate. pK_i values in descending order for compounds with different residues R: *p*-terphenyl-4-yl > 4-bromophenyl > biphenyl-4-yl > 4-(2-thienyl)phenyl, 3-(2-thienyl)phenyl, *p*-terphenyl-3-yl, 4-trifluoromethylphenyl > 4-methoxyphenyl, biphenyl-3-yl > 4-nitrophenyl, 4-(4-pyridyl)phenyl > 3-bromphenyl, 3-(4-pyridyl)phenyl.[76]

Yuan *et al*. have excised the amide moiety in the aforementioned type of compounds, leading to a generally diminished activity as cathepsin inhibitors (Fig. 48). In particular the 'straight' *p*-terphenyl-4-yl substituent was less advantageous than the 'bent' substituents, *i.e.* 3-(2-thienyl)phenyl, 3-(4-pyridyl)phenyl and biphenyl-3-yl.[77]

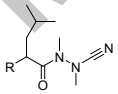


Figure 48. Azadipeptide nitriles with with different P3 capping groups. The assay was performed at pH 6.0 with Cbz-Arg-Arg-AMC as a fluorogenic substrate. pK_i values in descending order for compounds with different residues R: 3-(2-thienyl)phenyl, 3-(4-pyridyl)phenyl, biphenyl-3-yl > 4-(4-pyridyl)phenyl > biphenyl-4-yl, 4-(2-thienyl)phenyl > *p*-terphenyl-3-yl > *p*-terphenyl-4-yl.[77]

Krantz *et al.* discovered peptidyl acyloxymethyl ketones as affinity labels for cathepsin B.[78] Equipped with the acyloxymethyl unit containing a leaving group in form of the carboxylate nucleofuge, such inhibitors are capable of an irreversible inhibition of the target protease by alkylating the active site cysteine.[78,79] This chemotype of cysteine protease inactivators has widely been applied for activity-based probing.[80-83] Second-order rate constants of inactivation, k_{inac}/K_i, have been determined to assess the inhibitory potency of the compounds shown in Figure 49.[78] Electron-withdrawing groups lower the pK_a of the benzoyloxy leaving group and thus were found to increase the reactivity of the compound because of electronic effects rather than specific non-covalent interactions within the binding site.

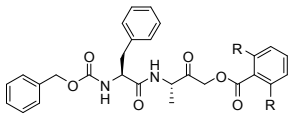


Figure 49. Acyloxymethyl ketone inactivators with different leaving groups. The assay was performed at pH 6.0 with Cbz-Phe-Arg-AMC or Bz-Val-Lys-Lys-Arg-4-trifluoromethylcoumarinamide as fluorogenic substrates. k_{inac}/K_i values in descending order for compounds with different residues R: CF_3 > Cl > F > Me > OMe > H.[78]

Different acyl groups R were connected with the fixed Ac-Phe-Gly motif resulting in a further series of acyloxymethyl ketones (Fig. 50).[79] The amino acid leaving groups are thought to interact with the primed binding region of cathepsin B and different inactivation parameters might reflect the magnitude of this interaction. Accordingly, the amino acid residue and its stereochemistry had an impact on the potency. Replacement of D-Leu with D-Asn led to a 100-fold decreased k_{inac}/K_i value and inversion of the Leu-Cbz configuration from L to D resulted in a 10-fold reduction (k_{inac}/K_i = 2,700 $M^{-1}s^{-1}$, 300 $M^{-1}s^{-1}$).

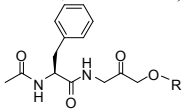


Figure 50. Acyloxymethyl ketone inactivators with different leaving groups. The assay was performed at pH 6.8 with Cbz-Phe-Arg-pNA as chromogenic substrate. k_{inac}/K_i values in descending order for compounds with different acyl groups R: L-Leu-Cbz > D-Trp-Ac, D-Leu-Cbz > D-Phe-Ac > Bz, D-Asn-Cbz. Note that the amino acids are connected to the unaltered substructure via an ester bond.[79]

Variations in the leaving group of a further acyloxymethyl ketone (Fig. 51) revealed the introduction of a terminal carboxylate (R = OCH_2CO_2H) not to be superior to methyl.[84]

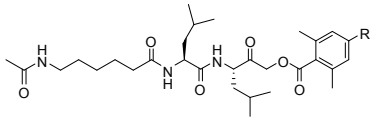


Figure 51. 4-Functionalized 2,6-dimethylbenzoate acyloxymethyl ketones. The assays were performed at pH 5.0 with Cbz-Arg-Arg-AMC as fluorogenic substrate. pIC_{50} values in descending order for compounds with different residues R: carboxymethoxy, methyl > H.[84]

A combination of different peptide portions with a fixed acyloxy methyl ketone warhead, which is relevant with respect to the active site mapping of cathepsin B, is shown in Figure 52.[85] Lysine at P1 position was superior to alanine and glycine as could be concluded by comparison of corresponding Cbz-protected derivatives with phenylalanine at P2 position (k_{inac}/K_i = 230,000 $M^{-1}s^{-1}$, 14,000 $M^{-1}s^{-1}$, 9,900 $M^{-1}s^{-1}$, respectively). Moreover, among the three capping groups, carbobenzoxy was favored.

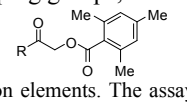


Figure 52. Acyloxymethyl ketones with different peptidic recognition elements. The assay was performed at pH 6.0 with Cbz-Phe-Arg-AMC or Bz-Val-Lys-Lys-Arg-4-trifluoromethylcoumarinamide as fluorogenic substrates. k_{inac}/K_i values in descending order for compounds with different residues R: Cbz-Phe-Lys > Cbz-Tyr(OMe)-Lys > Cbz-Phe-Ala > Cbz-Phe-Gly > MeO-Suc-Phe-Ala > HO-Suc-Phe-Ala, HO-Suc-Phe-Gly.[85]

Radioiodinated acyloxymethyl ketones have been developed as activity-based probes for cathepsin B.[86] The general structure (Fig. 53) has only one point of diversity. The most potent compound was the non-iodinated one with a terminal primary amine moiety (R = H) exhibiting a k_{inac}/K_i value of 40,000 $M^{-1}s^{-1}$. By extending the P1 residue with a PEG linker, terminated by *para*-iodobenzamide, the most potent iodinated inhibitor was obtained with a k_{inac}/K_i value of 21,100 $M^{-1}s^{-1}$. The other derivatives listed in Figure 53 exhibited lower second-order rate constants of inactivation, obviously because of unfavorable interactions of R with the active site of cathepsin B.[86]

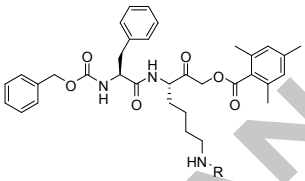


Figure 53. Acyloxymethyl ketone inactivators. The assay was performed at pH 6.0 with Cbz-Arg-Arg-pNA as chromogenic substrate. k_{inac}/K_i values in descending order for compounds with different residues R: H > $CO(CH_2CH_2O)_3(CH_2)_2NHCOC_6H_4$-*p*-I > COC_6H_4-*p*-I > $CO(CH_2)_5NHCOC_6H_4$-*p*-I > $CONHCOC_6H_4$-*p*-I > $CO(CH_2)_5NHCONHC_6H_4$-*p*-I.[86]

A library of 400 compounds with different cyclohexanone-derived reversible inhibitors was synthesized by Abato *et al*. using solid-phase chemistry and was evaluated toward three proteases, including cathepsin B.[87] The proteinogenic amino acids were incorporated at positions P2 and P2' and cysteine and methionine were replaced by hydroxyproline and ornithine. In Figure 54, the structure with phenylalanine in P2 position is shown. For cathepsin B, rather small differences in the binding profile were observed with this amino acid (or isoleucine, tryptophan, tyrosine) in P2 position. Phenylalanine and tryptophan appeared to be preferred in P2' position, although the general affinity of this type of compounds for cathepsin B was weak (R = Bn: K_i = 1.1 mM).[87] Probably, the free carboxylate of the P2' amino acid did not prove to be advantageous because the assay conditions used here did not specifically address the exopeptidase activity of cathepsin B.

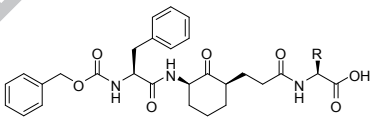


Figure 54. Cyclohexanone derived inhibitors. The assays were performed at pH 7.4 with Cbz-Arg-Arg-pNA as chromogenic substrate. pIC_{50} values in descending order for compounds with different residues R: benzyl > (1*H*-indol-3-yl)methyl.[87]

Vinyl sulfones have received much attention as covalent, irreversible inhibitors for cysteine cathepsins.[88-94] Their mode of action is based upon the nucleophilic addition of the active site cysteine residue at the β-carbon of the α,β-unsaturated Michael acceptor moiety. The electrophilic site is located at the position of the carbonyl carbon of the scissile bond of a peptidic substrate. Once the preceding noncovalent enzyme-inhibitor complex is formed and the peptide portion of the inhibitor is oriented toward the non-primed binding region, the catalytic thiolate attacks the prochiral electrophilic β-carbon from the *si* face. This kind of attachment was confirmed, for example, by an X-ray crystal structure of a vinyl sulfon-cathepsin S complex.[95,96]

Palmer *et al.* investigated morpholino urea inhibitors with the homophenylalanine residue and different substituents in P2 position. Introduction of the large 3,5-diiodotyrosine moiety gave rise to a particularly potent inactivator (Fig. 55).[97]

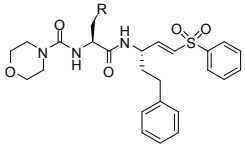


Figure 55. Vinyl sulfone inhibitors. The assays were performed at pH 6.0 with Cbz-Arg-Arg-AMC as fluorogenic substrate. Second-order rate constants of inactivation in descending order for compounds with different residues R: 3,5-I_2-4-HO-Ph > Ph > *i*-Pr.[97]

The glycine-homophenylalanine portion was incorporated into peptidomimetic vinyl sulfones with the terminal nitrogen as part of the benzodiazepine nucleus. The smallest substituent at the sulfur atom resulted in the weakest activity (Fig. 56).[98]

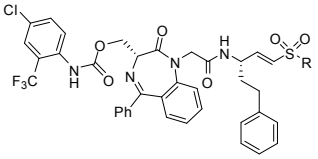


Figure 56. Peptidomimetic vinyl sulfone inhibitors. The assays were performed at pH 6.0 with Cbz-Phe-Arg-AMC as fluorogenic substrate. k_{inac}/K_i values in descending order for compounds with different residues R: 4-CH_3O-Ph, Ph, Et > Me.[98]

In the study of Mendieta *et al*.,[99] the series of vinyl sulfones (see Fig. 55) was extended and different aromatic substituents were placed at the γ-carbon (Fig. 57). The compounds with a biphenyl-4-yl or 2-naphthyl residue were most potent (IC_{50} = 300 nM) against cathepsin B. When maintaining these P1 residues and replacing phenylalanine by leucine or 4-methylpiperazine by morpholine, a further improved inhibitory capacity was not achieved.

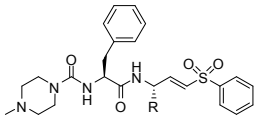


Figure 57. Vinyl sulfone inhibitors. The assays were performed at pH 6.2 with Cbz-Arg-Arg-AMC as fluorogenic substrate. pIC_{50} values in descending order for compounds with different residues R: biphenyl-4-yl, 2-naphthyl > phenyl > 1-naphthyl > mesityl.[99]

Several classes of cathepsin B inhibitors have been investigated, and some of them are mentioned in this review with respect to the active site mapping of cathepsin B. However, limited preclinical data are available and only one compound of undisclosed structure (VBY-376) targeting cathepsin B has entered clinical trials.[100] It can be expected that active site mapping information will be useful for the future development of low-molecular weight inhibitors for this protease.

5. Conclusions

Summarizing the data of the cathepsin B active site mapping presented in this review deciphers tendencies for favorable residues and amino acids in low-molecular weight inhibitors and peptidic substrates, respectively. By taking crystallographic data into account, these trends can be explained on the basis of specific protein-ligand interactions and are in agreement with characteristic features of the cathepsin B binding pockets.[8,17-20,26,55,65] A propensity for aromatic residues at the P3 position has been reported in various, aforementioned studies. These aromatic moieties may likely address Tyr75 of the S3 specificity pocket. Lipophilic and aromatic substituents and amino acid residues were frequently found to be feasible for an accommodation in the hydrophobic environment of the S2 subsite, characterized by Ala173, Gly172 and Ala200. A trend was identified for substrates and inhibitors with basic moieties in P1 position, able to perform hydrogen bond and charged interactions with the residues Asn72 and Glu122 of the S1 pocket. Hydrophobic substituents in ligands, such as isoleucine or phenylalanine residues, showed a preference for occupation of the lipophilic S1' binding site of cathepsin B, assembled by Val176, Leu181 and Met196.

This survey has considered a variety of reports on the ligand specificities of cathepsin B. It reveals a to some extent heterogeneous picture. Subsite cooperativity might be one reason for such inconsistencies. Owing to conformational plasticity and shared determinants among binding pockets, the accommodation of a particular residue at one subsite can have either a positive or negative influence on the binding of another residue at a different subsite of the protease.[34,101] In several of the aforementioned mapping studies that rely on the use of limited substrate or inhibitor pools, cooperativity information has not been considered.

Different assay conditions will interfere with the equilibrium between the open and closed conformation of cathepsin B, influence the balance between exo- and endoproteolytic activity, and thus limit the comparability of different profiling studies.

Moreover, in case of IQF substrates, a cleavage event at any peptide bond will produce the fluorescent outcome and the precise deconvolution requires MS analysis.

The active site mapping studies reviewed herein followed conventional procedures. However, a variety of novel technologies have been developed in order to map small molecule-target interactions. These include proteomic approaches, such as the affinity-based biochemical analysis by means of small-molecule probes. Novel genetic approaches comprise, for example, the transcriptional profiling of cellular states or the identification of mutations in the target encoding genes. New computational approaches can be used for *in silico* target prediction. Such attempts have recently been summarized in an instructive review article.[102]

Multiplex substrate profiling by mass spectrometry (MSP-MS) was initially developed to assess the extended prime and non-prime substrate specificity of proteases. This degradome analysis utilizes liquid chromatography combined with tandem mass spectrometry (LC-MS/MS) for the deconvolution of proteolytic signatures derived from a library of *e.g.* 228 peptides composed of *e.g.* 14 amino acids, covering a maximal physico-chemical diversity.[34]

An important N-terminomic method for the proteome-wide analysis of proteolytic processing is termed 'proteomic identification of protease cleavage sites' (PICS). It utilizes natural sequence diversity together with LC-MS/MS to determine the peptides which are generated from cellular proteomes by endoproteolytic digestion. For example, a peptide library is prepared by digesting a cell lysate with trypsin. Free amino groups are blocked by reductive methylation and the library is then treated with the protease of interest to create *neo* N-termini. These newly formed NH_2 groups are then chemically or enzymatically biotinylated, enriched and identified by MS. The PICS technique is applicable for the biochemical profiling of protease prime and non-prime specificity and, accordingly, for an active site mapping of the protease of interest.[103-105]

A pioneer PICS study for cathepsin B was performed by Biniossek *et al*.[106] The authors surveyed hundreds of cleavage products from proteome-derived peptide libraries. The LC/MS-based assignment required peptides of at least seven residues which qualifies this study for the endopeptidase activity of cathepsin B. The P3-S3 interaction was assessed to be an only minor specificity determinant. A canonical specificity of cysteine proteases for alanine, valine, tryptophan and phenylalanine in P2 position and a notable selectivity for glycine in P1 position was detected. Cathepsin B exhibited a slight preference for aromatic amino acids in P1' and for valine and isoleucine in P2' position. Noteworthy, the major specificity determinant for endopeptidase activity was determined for glycine in P3' position,[106] as also observed for the proteolysis of rabbit muscle fructose 1,6-bisphosphate aldolase by cathepsin B.[70] Thus, these PICS data are, by and large, in accordance with the results obtained with chemically defined peptidic substrates or combinatorial series of low-molecular weight inhibitors. PICS studies for substrate mapping are limited by the specificity of the protease used to generate the peptide library prior to analysis. For example, Biniossek *et al.* used tryptic peptide libraries, lacking internal arginine and lysine. The obtained tryptic PICS profile does, therefore, not allow for statements with regard to cathepsin B endoproteolytic specificity for basic residues. Notably, a positive cooperativity between aromatic residues in P2 and P1' was discovered.[106] Future cell-contextual proteomic approaches will be beneficial for the substrate profiling of cathepsin B.

References and notes

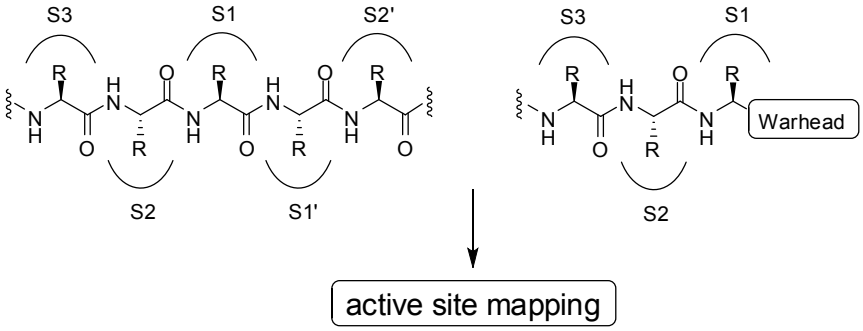
1. Kos, J.; Mitrović, A.; Mirković, B. *Future Med. Chem*. **2014**, *6*, 1355.
2 Löser, R.; Pietzsch, J. *Front. Chem*. **2015**, 3:37.
3. Sloane, B. F.; Yan, S.; Podgorski, I.; Linebaugh, B. E.; Cher, M. L.; Mai, J.; Cavallo-Medved, D.; Sameni, M.; Dosescu, J.; Moin, K. *Semin. Cancer Biol*. **2005**, *15*, 149.
4. Joyce, J. A.; Hanahan, D. *Cell Cycle* **2004**, *3*, 1516.
5. Reiser, J.; Adair, B.; Reinheckel, T. *J. Clin. Invest.* **2010**, *120*, 3421.
6. Frlan, R.; Gobec, S. *Curr. Med. Chem.* **2006**, *13*, 2309.
7. Mirković, B.; Markelc, B.; Butinar, M.; Mitrović, A.; Sosič, I.; Gobec, S.; Vasiljeva, O.; Turk, B.; Čemažar, M.; Serša, G.; Kos, J. *Oncotarget* **2015** 6, 19027.
8. Musil, D.; Zucic, D.; Turk, D.; Engh, R. A.; Mayr, I.; Huber, R.; Popovic, T.; Turk, V.; Towatari, T.; Katunuma, N.; Bode, W. *EMBO J.* **1991**, *10*, 2321.
9. Illy, C.; Quraishi, O.; Wang, J., Purisima, E., Vernet, T., Mort, J. S. *J. Biol. Chem.* **1997**, *272*, 1197.
10. Mitrović, A.; Mirković, B.; Sosič, I.; Gobec, S.; Kos, J. *Biol. Chem.* **2016**, *397*, 165.
11. Hanada, K.; Tamai, M.; Yamagishi, M.; Ohmura, S. Sawada, I.; Tanaka I. *Agric. Biol. Chem.* **1978**, *42*, 523.
12. Barrett, A. J.; Kembhavi, A. A.; Brown, M. A.; Kirschke H.; Knight, C. G.; Tamai, M.; Hanada, K. *Biochem. J.* **1982**, *201*, 189.
13. Murata, M.; Miyashita, S.; Yokoo, C.; Tamai, M.; Handa, K.; Hatayama, K.; Towatari, T.; Nikawa, T.; Katunuma, N. *FEBS Lett.* **1991**, *280*, 307.
14. Towatari, T.; Nikawa, T.; Murata, M.; Yokoo, C.; Tamai, M.; Handa, K.; Katunuma, N. *FEBS Lett.* **1991**, *280*, 311.
15. Matsumoto, K.; Yamamoto, D.; Ohishi, H.; Tomoo, K.; Ishida, T.; Inoue, M.; Sadatome, T.; Kitamura, K.; Mizuno, H. *FEBS Lett.* **1989**, *245*, 177.
16. Varughese, K. I.; Ahmed, F. R.; Carey, P. R.; Hasnain, S.; Huber, C. P.; Storer, A. C. *Biochemistry* **1989**, *28*, 1330.
17. Turk, D.; Podobnik, M.; Popovic, T.; Katunuma, N.; Bode, W.; Huber, R.; Turk, V. *Biochemistry* **1995**, *34*, 4791.
18. Yamamoto, A.; Tomoo, K.; Matsugi, K.; Hara, T.; In, Y.; Murata, M.; Kitamura, K.; Ishida, T. *Biochim. Biophys. Acta* **2002**, *1597*, 244.
19. Yamamoto, A.; Hara, T.; Tomoo, K.; Ishida, T.; Fujii, T.; Hata, Y.; Murata, M.; Kitamura, K. *J. Biochem*. **1997**, *121*, 974.
20. Yamamoto, A.; Tomoo, K.; Hara, T.; Murata, M.; Kitamura, K.; Ishida, T. *J. Biochem*. **2000**, *127*, 635.
21. Gour-Salin, J. B.; Lachance, P.; Plouffe, C.; Storer, A.C.; Ménard, R. *J. Med. Chem*. **1993**, *36*, 720.
22. Buttle, D. J.; Murata, M.; Knight, C. G.; Barret, A. J. *Arch. Biochem. Biophys*. **1992**, *299*, 377.
23. Giordano, C.; Calabretta, R.; Gallina, C.; Consalvi, V.; Scandurra, R.; Chiaia Noya, F.; Franchini, C. *Eur. J. Med. Chem.* **1993**, 28, 917.
24. Crawford, C.; Mason, R. W.; Wikstrom, P.; Shaw, E. *Biochem. J.* **1988**, *253*, 751.
25. Schaschke, N.; Assfalg-Machleidt, I.; Machleidt, W.; Turk, D.; Moroder, L. *Bioorg. Med. Chem*. **1997**, *5*, 1789.
26. Stern, I.; Schaschke, N.; Moroder, L.; Turk, D. *Biochem. J*. **2004**, *381*, 511.
27. Schaschke, N.; Assfalg-Machleidt, I.; Machleidt, W.; Moroder, L. *FEBS Lett.* **1998**, *421*, 80.
28. Schaschke, N.; Assfalg-Machleidt, I.; Laßleben, T.; Sommerhoff, C. P.; Moroder, L.; Machleidt, W. *FEBS Lett*. **2000**, *482*, 91.
29. Watanabe, D.; Yamamoto, A.; Tomoo, K.; Matsumoto, K.; Murata, M.; Kitamura, K.; Ishida, T. *J. Mol. Biol.* **2006**, *362*, 979.
30. Albeck, A.; Fluss, S.; Persky, R. *J. Am. Chem. Soc.* **1996**, *118*, 3591.
31. Perlman, N.; Hazan, M.; Shokhen, M.; Albeck, A. *Bioorg. Med. Chem.* **2008**, *16*, 9032.
32. Poreba, M.; Szalek, A.; Rut, W.; Kasperkiewicz, P.; Rutkowska-Wlodarczyk I.; Snipas, S. J.; Itoh, Y.; Turk, D.; Turk, B.; Overall, C. M.; Kaczmarek, L.; Salvesen, G. S.; Drag, M. *Sci. Rep.* **2017**, *7*, 43135.
33. Rut, W.; Kasperkiewicz, P.; Byzia, A.; Poreba, M.; Groborz, K.; Drag M. *Biol. Chem.* **2015**, *396*, 329.
34. Ivry, S. L.; Meyer, N. O.; Winter, M. B.; Bohn, M. F.; Knudsen, G. M., O'Donoghue, A. J.; Craik, C. S. *Protein Sci.* **2018**, *27*, 584.
35. Kasperkiewicz, P.; Poreba, M.; Groborz, K.; Drag, M. *FEBS J.* **2017**, *284*, 1518.
36. Poreba, M.; Rut, W.; Vizovisek, M.; Groborz, K.; Kasperkiewicz, P.; Finlay, D.; Vuori, K.; Turk, D.; Turk, B.; Salvesen, G. S.; Drag, M. *Chem. Sci.* **2018**, *9*, 2113.
37. Ménard, R.; Carmona, E.; Plouffe, C.; Brömme, D.; Konishi, Y.; Lefebvre, J.; Storer, A. C. *FEBS Lett*. **1993**, *328*, 107.
38. Brömme, D.; Bonneau, P. R.; Lachance, P.; Storer, A. C. *J. Biol. Chem*. **1994**, *269*, 30238.
39. Semashko, T. A.; Vorotnikova, E. A.; Sharikova, V. F.; Vinokurov, K. S.; Smirnova, Y. A.; Dunaevsky, Y. E.; Belozersky, M. A.; Oppert, B.; Elpidina, E. N.; Filippova, I. Y. *Anal. Biochem.* **2014**, *449*, 179.
40. Taralp, A.; Kaplan, H.; Sytwu, I. I.; Vlattas, I.; Bohacek, R.; Knap, A. K.; Hirama, T.; Huber, C. P.; Hasnain, S. *J. Biol. Chem.* **1995**, *270*, 18036.
41. Portaro, F. C.; Santos, A. B.; Cezari, M. H.; Juliano, M. A.; Juliano, L.; Carmona, E. *Biochem. J.* **2000**, *347*, 123.
42. Jia, Z.; Hasnain, S.; Hirama, T.; Lee, X.; Mort, J. S.; To, R.; Huber, C. P. *J. Biol. Chem.* **1995**, *270*, 5527.
43. Turk, D.; Podobnik, M.; Kuhelj, R.; Dolinar, M.; Turk, V. *FEBS Lett.* **1996**, *384*, 211.
44. Nägler, D. K.; Tam, W.; Storer, A. C.; Krupa, J. C.; Mort, J. S.; Ménard, R. *Biochemistry* **1999**, *38*, 4868.
45. Del Nery, E.; Alves, L. C.; Melo, R. L.; Cesari, M. H.; Juliano, L.; Juliano, M. A. *J. Protein Chem*. **2000**, *19*, 33.
46. Melo, R. L.; Alves, L. C.; Del Nery, E.; Juliano, L.; Juliano, M. A. *Anal. Biochem.* **2001**, *293*, 71.
47. Krupa, J. C.; Hasnain, S.; Nägler, D. K.; Ménard, R.; Mort, J. S. *Biochem. J*. **2002**, *361*, 613.
48. Cezari, M. H.; Puzer, L.; Juliano, M. A.; Carmona, A. K.; Juliano, L. *Biochem. J.* **2002**, *368*, 365.
49. Cotrin, S. S.; Puzer, L.; de Souza Judice, W. A.; Juliano, L.; Carmona, A. K.; Juliano, M. A. *Anal. Biochem.* **2004**, *335*, 244.
50. Giulianotti, M. A.; Debevec, G.; Santos, R. G.; Maida, L. E.; Chen, W.; Ou, L.; Yu, Y.; Dooley, C. T.; Houghten, R. A. *ACS Comb. Sci.* **2012**, *14*, 503.
51. Stachowiak, K.; Tokmina, M.; Karpińska, A.; Sosnowska, R.; Wiczk, W. *Acta Biochim. Pol.* **2004**, *51*, 81.
52. Ruzza, P.; Quintieri, L.; Osler, A.; Calderan, A.; Biondi, B.; Floreani, M.; Guiotto, A.; Borin G. *J. Pept. Sci.* **2006**, *12*, 455.
53. Choe, Y.; Leonetti, F.; Greenbaum, D. C.; Lecaille, F.; Bogyo, M.; Brömme, D.; Ellman, J. A.; Craik, C. S. *J. Biol. Chem.* **2006**, *281*, 12824.
54. Kim, C. J.; Lee, D. I.; Zhang, D.; Lee, C. H.; Ahn, I. S. *Anal. Biochem.* **2013**, *435*, 166.
55. Wei, B.; Gunzner-Toste, J.; Yao, H.; Wang, T.; Wang, J.; Xu, Z.; Chen, J.; Wai, J.; Nonomiya, J.; Tsai, S. P.; Chuh, J.; Kozak, K.R.; Liu, Y.; Yu, S. F.; Lau, J.; Li, G.; Phillips, G. D.; Leipold, D.; Kamath, A.; Su, D.; Xu, K.; Eigenbrot, C.; Steinbacher, S.; Ohri, R.; Raab, H.; Staben, L. R.; Zhao, G.; Flygare, J. A.; Pillow, T. H.; Verma, V.; Masterson, L. A.; Howard, P. W.; Safina, B. *J. Med. Chem.* **2018**, *61*, 989.
56. Greenbaum, D. C.; Arnold, W. D.; Lu, F.; Hayrapetian, L.; Baruch, A.; Krumrine, J.; Toba, S.; Chehade, K.; Brömme, D.; Kuntz, I. D.; Bogyo, M. *Chem Biol.* **2002**, *10*, 1085.
57. Frizler, M.; Stirnberg, M.; Sisay, M. T.; Gütschow M. *Curr. Top. Med. Chem.* **2010**, *10*, 294.
58. Silva, D. G.; Ribeiro, J. F.; De Vita, D.; Cianni, L.; Franco, C. H.; Freitas-Junior, L. H.; Moraes, C. B.; Rocha, J. R.; Burtoloso, A. C.; Kenny, P. W.; Leitão, A.; Montanari, C. A. *Bioorg. Med. Chem. Lett.* **2017**, *27*, 5031.
59. Avelar, L. A.; Camilo, C. D.; de Albuquerque, S.; Fernandes, W. B.; Gonçalez, C.; Kenny, P. W.; Leitão, A.; McKerrow, J. H.; Montanari, C. A.; Orozco, E. V.; Ribeiro, J. F.; Rocha, J. R.; Rosini, F.; Saidel, M. E. *PLoS Negl. Trop. Dis.* **2015**, *9*, e0003916.
60. Palmer, J. T.; Bryant, C.; Wang, D. X.; Davis, D. E.; Setti, E. L.; Rydzewski, R. M.; Venkatraman, S.; Tian, Z. Q.; Burrill, L. C.; Mendonca, R. V.; Springman, E.; McCarter, J.; Chung, T.; Cheung, H.; Janc, J. W.; McGrath, M.; Somoza, J. R.; Enriquez, P.; Yu, Z. W.; Strickley, R. M.; Liu, L; Venuti, M. C.; Percival, M. D.; Falgueyret, J. P.; Prasit, P.; Oballa, R.; Riendeau, D.; Young, R. N.; Wesolowski, G.; Rodan, S. B.; Johnson, C.; Kimmel, D. B.; Rodan, G. *J. Med. Chem*. **2005**, *48*, 7520.
61. Schmitz, J.; Li, T.; Bartz, U.; Gütschow, M. *ACS Med. Chem. Lett.* **2015**, *7*, 211.
62. Löser, R.; Schilling, K.; Dimmig, E.; Gütschow, M. *J. Med. Chem.* **2005**, *48*, 7688.
63. Frizler, M.; Lohr, F.; Lülsdorff, M.; Gütschow, M. *Chem. Eur. J.* **2011**, *17*, 11419.
64. Frizler, M.; Lohr, F.; Furtmann, N.; Kläs, J.; Gütschow, M. *J. Med. Chem.* **2011**, *54*, 396.
65. Greenspan, P. D.; Clark, K. L.; Tommasi, R. A.; Cowen, S. D.; McQuire, L. W.; Farley, D. L.; van Duzer, J. H.; Goldberg, R. L.; Zhou, H.; Du, Z.; Fitt, J. J.; Coppa, D. E.; Fang, Z.; Macchia, W.; Zhu, L.; Capparelli, M. P.; Goldstein, R.; Wigg, A. M.; Doughty, J. R.; Bohacek, R. S.; Knap, A. K. *J. Med. Chem.* **2001**, *44*, 4524.

66. Falgueyret, J. P.; Desmarais, S.; Oballa, R.; Black, W. C.; Cromlish, W.; Khougaz, K.; Lamontagne, S.; Massé, F.; Riendeau, D.; Toulmond, S.; Percival, M. D. *J. Med. Chem*. **2005**, *48*, 7535.
67. Fustero, S.; Rodrigo, V.; Sánchez-Roselló, M.; del Pozo, C.; Timoneda, J.; Frizler, M.; Sisay, M. T.; Bajorath, J.; Calle, L. P.; Cañada, F. J.; Jiménez-Barbero, J.; Gütschow M. *Chem. Eur. J.* **2011**, *17*, 5256.
68. Greenspan, P. D.; Clark, K. L.; Cowen, S. D.; McQuire, L.W.; Tommasi, R. A.; Farley, D. L.; Quadros, E.; Coppa, D. E.; Du, Z.; Fang, Z.; Zhou, H.; Doughty, J.; Toscano, K. T.; Wigg, A. M.; Zhou, S. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 4121.
69. Sosič, I.; Mirković, B.; Arenz, K.; Stefane, B.; Kos, J.; Gobec S. *J. Med. Chem*. **2013**, *56*, 521.
70. Schenker, P.; Alfarano, P.; Kolb, P.; Caflisch, A.; Baici A. *Protein Sci.* **2008**, *17*, 2145.
71. Nägler, D. K.; Storer, A. C.; Portaro, F. C.; Carmona, E.; Juliano, L.; Ménard, R. *Biochemistry* **1997**, *36*, 12608.
72. Löser, R.; Frizler, M.; Schilling, K.; Gütschow, M. *Angew. Chem. Int. Ed*. **2008**, *47*, 4331.
73. Frizler, M.; Schmitz, J.; Schulz-Fincke, A. C.; Gütschow, M. *J. Med. Chem*. **2012**, *55*, 5982.
74. Löser, R.; Bergmann, R.; Frizler, M.; Mosch, B.; Dombrowski, L.; Kuchar, M.; Steinbach, J.; Gütschow, M.; Pietzsch, J. *ChemMedChem* **2013**, *8*, 1330.
75. Frizler, M.; Mertens, M. D.; Gütschow, M. *Bioorg. Med. Chem. Lett*. **2012**, *22*, 7715.
76. Ren, X. F.; Li, H. W.; Fang, X.; Wu, Y.; Wang, L.; Zou, S. *Org. Biomol. Chem.* **2013**, *11*, 1143.
77. Yuan, X. Y.; Fu, D. Y.; Ren, X. F.; Fang, X.; Wang, L.; Zou, S.; Wu, Y. *Org. Biomol. Chem.* **2013**, *11*, 5847.
78. Krantz, A.; Copp, L. J.; Coles, P. J.; Smith, R. A.; Heard, S. B. *Biochemistry* **1991**, *30*, 4678.
79. Dai, Y.; Hedstrom, L.; Abeles, R. H. *Biochemistry* **2000**, *39*, 6498.
80. Edgington, L. E.; Verdoes, M.; Bogyo, M. *Curr. Opin. Chem. Biol.* **2011**, *15*, 798.
81. Blum, G.; Mullins, S. R.; Keren, K.; Fonovic, M.; Jedeszko, C.; Rice, M. J.; Sloane, B. F.; Bogyo, M. *Nat. Chem. Biol.* **2005**, *1*, 203.
82. Blum, G.; von Degenfeld, G.; Merchant, M. J.; Blau, H. M.; Bogyo, *M. Nat. Chem. Biol.* **2007**, *3*, 668.
83. Oresic Bender, K.; Ofori, L.; van der Linden, W. A.; Mock, E. D.; Datta, G. K.; Chowdhury, S.; Li, H.; Segal, E.; Sanchez Lopez, M.; Ellman, J. A.; Figdor, C. G.; Bogyo, M.; Verdoes, M. *J. Am. Chem. Soc.* **2015**, *137*, 4771.
84. Coman, A. G.; Paraschivescu, C. C.; Hadade, N. D.; Juncu, A.; Vlaicu, O.; Popescu, C.; Matache, M. *Synthesis* **2016**, *48*, 3917.
85. Wagner, B. M.; Smith, R. A.; Coles, P. J.; Copp, L. J.; Ernest, M. J.; Krantz, A. *J. Med. Chem.* **1994**, *37*, 1833.
86. Edem, P. E.; Czorny, S.; Valliant, J. F. *J. Med. Chem*. **2014**, *57*, 9564.
87. Abato, P.; Conroy, J. L.; Seto, C. T.; *J. Med. Chem*. **1999**, *42*, 4001.
88. Santos, M. M.; Moreira, R. *Mini Rev. Med. Chem.* **2007**, *7*, 1040.
89. Jílková, A.; Rezácová, P.; Lepsík, M.; Horn, M.; Váchová, J.; Fanfrlík, J.; Brynda, J.; McKerrow, J. H.; Caffrey, C. R.; Mares M. *J. Biol. Chem.* **2011**, *286*, 35770.
90. Wiggers, H. J.; Rocha, J. R.; Fernandes, W. B.; Sesti-Costa, R.; Carneiro, Z. A.; Cheleski, J.; da Silva, A. B.; Juliano, L.; Cezari, M. H.; Silva, J. S.; McKerrow, J. H.; Montanari, C. A. *PLoS Negl. Trop. Dis.* **2013**, *7*, e2370.
91. Ettari, R.; Tamborini, L.; Angelo, I. C.; Micale, N.; Pinto, A.; De Micheli, C.; Conti, P. *J. Med. Chem*. **2013**, *56*, 5637.
92. Frizler, M.; Yampolsky, I. V.; Baranov, M. S.; Stirnberg, M.; Gütschow, M. *Org. Biomol. Chem.* **2013**, *11*, 5913.
93. Previti, S.; Ettari, R.; Cosconati, S.; Amendola, G.; Chouchene, K.; Wagner, A.; Hellmich, U.; Ulrich, K.; Krauth-Siegel, R. L.; Wich, P. R.; Schmid, I.; Schirmeister, T.; Gut, J.; Rosenthal, P. J.; Grasso, S.; Zappalà, M. *J. Med. Chem*. **2017**, *60*, 6911.
94. Breuer, C.; Lemke, C.; Schmitz, J.; Bartz, U.; Gütschow, M. *Bioorg. Med. Chem. Lett.* **2018**, *28*, 2008.
95. Pauly, T.A.; Sulea, T.; Ammirati, M.; Sivaraman, J.; Danley, D. E.; Griffor, M. C.; Kamath, A. V.; Wang, I. K.; Laird, E. R.; Seddon, A. P.; Ménard, R.; Cygler, M.; Rath, V. L. *Biochemistry* **2003**, *42*, 3203.
96. Mertens, M. D.; Schmitz, J.; Horn, M.; Furtmann, N.; Bajorath, J.; Mareš, M.; Gütschow, M. *ChemBioChem* **2014**, *15*, 955.
97. Palmer, J. T.; Rasnick, D.; Klaus, J. L.; Brömme, D. *J. Med. Chem.* **1995**, *38*, 3193.
98. Ettari, R.; Nizi, E.; Di Francesco, M. E.; Dude, M. A.; Pradel, G.; Vicík, R.; Schirmeister, T.; Micale, N.; Grasso, S.; Zappalá, M. *J. Med. Chem.* **2008**, *51*, 988.
99. Mendieta, L.; Picó, A.; Tarragó, T.; Teixidó, M.; Castillo, M.; Rafecas, L.; Moyano, A.; Giralt, E. *ChemMedChem* **2010**, *5*, 1556.
100. Kramer, L.; Turk, D.; Turk, B. *Trends Pharmacol. Sci*. **2017**, *38*, 873.
101. Ng, N. M.; Pike, R. N.; Boyd, S. E. *Biol. Chem.* **2009**, *390*, 401
102. Kapoor, S.; Waldmann, H.; Ziegler, S. *Bioorg Med. Chem.* **2016**, *24*, 3232.
103. Shahinian, H.; Tholen, S.; Schilling, O. *Expert Rev. Proteomics* **2013**, *10*, 421.
104. Vizovišek, M.; Vidmar, R.; Fonović, M.; Turk, B. *Biochimie* **2016**, *122*, 77.
105. Schilling, O.; Huesgen, P. F.; Barré, O.; Auf dem Keller, U.; Overall, C. M. *Nat. Protoc.* **2011**, *6*, 111.
106. Biniossek, M. L.; Nägler, D. K.; Becker-Pauly, C.; Schilling, O. *J. Proteome Res.* **2011**, *10*, 5363.

Cathepsin B: Active Site Mapping with Peptidic Substrates and Inhibitors

Janina Schmitz [a,e], Erik Gilberg [a,b,e], Reik Löser [c], Jürgen Bajorath [b], Ulrike Bartz [d], Michael Gütschow [a]

[a] *Pharmaceutical Institute, Pharmaceutical Chemistry I, University of Bonn, An der Immenburg 4, D-53121 Bonn, Germany*
[b] *Department of Life Science Informatics, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, University of Bonn, Dahlmannstr. 2, D-53113 Bonn, Germany*
[c] *Department of Radiopharmaceutical and Chemical Biology, Institute of Radiopharmaceutical Cancer Research, Helmholtz-Zentrum Dresden-Rossendorf, Bautzner Landstraße 400, D-01328 Dresden, Germany*
[d] *Department of Natural Sciences, University of Applied Sciences Bonn-Rhein-Sieg, von-Liebig-Str. 20, D-53359 Rheinbach, Germany*



active site mapping