Monika Łęgowska
Katedra Biochemii
Nr indeksu: 2541

SUMMARY OF DOCTORAL DISSARTATION

"In search of substrates and inhibitors of human cathepsin C. Design, chemical synthesis and biological studies"

Cathepsin C (Cat C) is a wildly expressed lysosmal cysteine dipeptidyl peptidase that is mostly recognized for the activation of the neutrophil serine proteinases and related granule-associated enzymes essential for host defense in humans. Cat C is also involved in many other important physiological processes; however, in some cases its exact role has not been fully understood. Although it has been shown that the enzyme can be secreted extracellularly, its occurrence in human bodily fluids has not been thoroughly studied.

Cat C is a unique member of the family of cysteine cathepsins due to its elaborate activation mechanism, oligomeric structure and exopeptidase activity. The presence of exclusion domain within Cat C's structure is responsible for its aminodipeptidase activity, as it blocks the enzyme active site beyond the S_2 pocket (nomenclature according to Schechter-Berger model), making it only accessible for a *N*-terminus of its substrates. Due to the dipeptidase activity of Cat C, most of the reported synthetic Cat C substrates are dipeptides bearing reporter groups such as chromophores and fluorophores. Synthetic dipeptide substrates of Cat C enabled the detailed specificity study of non-prime sites of this cathepsin; however, preferences for amino acid residues in P′ positions of Cat C substrates have not been thoroughly investigated.

Because of the Cat C's participation in the activation of neutrophil serine proteinases (NSPs), its inhibition is a promising therapeutic strategy in the treatment of many inflammatory diseases, in which the increased activity of NSPs results in the degradation of extracellular matrix and chronic inflammation. Bearing it in mind, scientists from many research centers have been working on the development of Cat C inhibitors. Dipeptidyl nitriles seem to be the most promising inhibitors among the investigated compounds due to their high inhibitory activity towards Cat C and low reactivity against cellular nucleophiles.

The first aim of my dissertation was the synthesis of novel, highly selective and sensitive human Cat C substrates, which would be suitable for the detection of Cat C activity in complex biological samples and human bodily fluids. I expected that the new substrates would give a new insight into Cat C activity and would allow the further characterization of this enzyme. Based on the structure of the active site of Cat C and its substrate specificity, I designed and synthesized two different classes of substrates.

Firstly, I focused on the fluorogenic dipeptide substrates, which are commonly used for the measurement of Cat C activity. The designed substrates were based on two dipeptide sequences: Gly-Phe and Thi-Phe (where Thi is *β*-(2-thienyl)-L-alanine), that are known to be recognized by Cat C. I wanted to optimize the fluorescent reporter group occupying P_1' position, thus I synthesized two series of substrates with the general formulas **Gly-Phe-Y** and **Thi-Phe-Y**, where Y is a coumarin or a quinolin-2(1*H*)-one derivative. The synthesis of the majority of the compounds was carried out in solution using Boc strategy and coupling reagent HATU. Compounds bearing 7-amino-coumarin-4-acetic acid amide (ACC) were obtained by the solid phase method using Fmoc strategy. The designed compounds were purified using liquid chromatography on silica gel. They were identified with MALDI-TOF MS and their homogeneity was analyzed by RP-HPLC. My goal was to investigate whether the compounds with Thi in P_2 position, a non-proteinogenic amino acid residue previously used in the potent and selective Cat C inhibitors, could be efficiently hydrolyzed in the presence of Cat C. The obtained compounds were studied in enzymatic assays with human recombinant Cat C and Cat L, that was used to activate the proform of Cat C. Substrates with the general formula Thi-Phe-Y were efficiently hydrolyzed in the presence of Cat C and displayed higher affinity towards Cat C than the corresponding compounds from the Gly-Phe-Y series. My other aim was to study the influence of the fluorophore structure on the substrate affinity toward Cat C and to find the optimal reporter group for Cat C. Firstly, I observed that four substrates with methyl and propyl groups in the position 4 of either coumarin or quinolinone ring were proteolyzed the most efficiently. Then, kinetic parameters (Michaelis constant, catalytic constant and specificity constant) were determined for the selected substrates in MES buffer at pH 6. Thi-Phe-APC (where APC is 7-amino-4-propylcoumarin) was found to be the most specific Cat C substrate within the series with the specificity constant of $(1.65\pm0.02)\times10^6$ $M^{-1}\times s^{-1}$. In order to evaluate the selectivity of the compound, it was incubated with the following cathepsins: Cat C, L, B, V, K, S and X. Unfortunately, it was proteolyzed not only by Cat C, but also Cat L, B, K and X.

The second class of the substrates designed for the measurement of Cat C were internally-quenched fluorescent substrates, that had been successfully applied for the measurement of the activity of different endopeptidases and had been proven to be superior over chromogenic and fluorogenic substrates in terms of specificity and sensitivity. The great advantage of FRET substrates is that they enable the investigation of S′ specificity of proteases. My objective was to design and synthesize FRET substrates of Cat C, that would not only allow for the sensitive and selective detection of Cat C activity, but also the detailed study of the binding preference of the protease S′ subsites, that had never been done before.

In the course of the study I synthesized two groups of FRET substrates, differing by the donor/acceptor pair for FRET. The first proposed substrates had the general formula **ABZ-Bip-X-Tyr(3-NO_2)-NH_2**, where ABZ was 2-aminobenzoic acid, Bip stands for biphenylalanine, whereas Tyr(3-NO_2) was 3-nitro-L-tyrosine and X was a set of proteinogenic amino acid residues except for Cys. Unfortunately, they were not hydrolyzed in the presence of Cat C, but turned out to be substrates of Cat L, a related lysosomal cathepsin, that is used for the activation of the proform of Cat C *in vitro.* Based on the sequence of the substrate that was very efficiently proteolyzed by Cat L, ABZ-Bip-Arg-Tyr(3-NO_2)-NH_2, I decided to develop a selective and ultrasensitive Cat L substrate. My goal was also to study in detail the substrate specificity of Cat L and use the obtained knowledge in the design of selective Cat C substrate. In order to achieve this goal, I synthesized the next two series of FRET peptides: **ABZ-Bip-Arg-X_1′-Tyr(3-NO_2)-NH_2** and **ABZ-Bip-Arg-X_1′-X_2′-Tyr(3-NO_2)-NH_2**, where **X_1′** and **X_2′** are a set of proteinogenic amino acid residues except for Cys. The FRET peptides were synthesized manually on the solid support using Fmoc strategy. I performed enzymatic assays with recombinant Cat L to study in detail the prime site specificity of Cat L, which allowed for the identification of the specific and sensitive Cat L substrates. I determined kinetic parameters for the selected hexapeptide substrates and evaluated their sensitivity. The selectivity of the most specific substrates was assessed in the assay with several cysteine cathepsins. Finally, the selected Cat L substrate ABZ-Bip-Arg-Ala-Gln-Tyr(3-NO_2)-NH_2 was applied for the measurement of Cat L activity in cell lysates of healthy and cancer cell lines. The results showed that Cat L activity was approximately 40% higher in cancer cells than in healthy cell lines.

The second group of the FRET peptides that were considered for the development of Cat C substrates were the peptidomimetics employing the FRET pair Ala(Mca)/Tyr(3-NO_2), where Ala(Mca) is β-(7-methoxy-coumarin-4-yl)-L-alanine. Firstly, I investigated the optimal

position within the peptide chain for Ala(Mca) and based on the obtained results I designed two series of Cat C FRET substrates with the general formula **Thi-Ala(Mca)-X_1'-Tyr(3-NO_2)-NH_2** and **Thi-Ala(Mca)-X_1'-X_2'-Tyr(3-NO_2)-NH_2**, where **X_1'** and **X_2'** stand for a set of proteinogenic amino acid residues except for Cys. The designed compounds were synthesized manually by the solid phase method using Fmoc strategy. Owing to the performed enzymatic assays with both recombinant Cat C and Cat L, I selected amino acid residues for the specific and selective FRET substrates of Cat C; Ser was introduced in P_1' position, whereas Gly and Gln were the most optimal residues in P_2'. In order to gain a deeper insight into the Cat C activity, I determined kinetic parameters for the selected substrates from both series of FRET peptides at pH 5 and 6. The results indicated that the extension of the peptide chain up to five amino acid residues resulted in 2-3 fold increase in Cat C turnover number, whereas it did not affect the affinity towards Cat C. The best substrate **Thi-Ala(Mca)-Ser-Gly-Tyr(3-NO_2)-NH_2** was applied for the measurement of Cat C activity in biological samples of different origin such as neutrophil cell lysate, lysates of selected cell lines and concentrated urine of healthy donors. The developed compound was also employed for the detection of Cat C activity in bronchoalveolar lavage fluids from patients suffering from chronic inflammatory respiratory diseases characterized by the predominance of macrophages: histiocytosis X and desquamative interstitial pneumopathy. The obtained compounds were also used in the detailed study of the prime site specificity of Cat C. FRET peptides were incubated with Cat C in buffers differing in pH: 5.0, 6.0 and 7.4. I observed that the specificity of Cat C was pH-dependent and it was the broadest at pH 5.0, in the environment resembling the one of lysosomes, where Cat C is responsible for protein degradation. The differences in substrates affinity towards Cat C were more pronounced at pH 6.0 and 7.4. The results revealed the preference for polar, basic and small aliphatic amino acid residues in P_1' and P_2' positions in Cat C substrates.

In the second part of my dissertation, the design and synthesis of potent and proteolysis-resistant inhibitors of Cat C became my main objective. Dipeptidyl nitriles have been the subject of much interest for the last ten years, as they are among the most potent low-molecular inhibitors of Cat C. Having it in mind, as a part of my thesis I planned the synthesis of two series of dipeptidyl nitriles. The first series consisted of all four configurational isomers of nitriles of *β*-(2-thienyl)-alanylphenylalanine, the LL isomer of which, **Thi-Phe-CN**, has been already proven to be a potent Cat C inhibitor with IC_{50}=0.9±0.3 nM. The main disadvantage of this compound was its lack of proteolytic stability when incubated with

hepatocytes and rat whole blood. Hence, I wanted to study the importance of the configuration of both chiral centers of the dipeptidyl nitrile, not only for the inhibitory activity, but also for the inhibitor's stability in human bodily fluids. Diastereomers were synthesized by the coupling of Boc-*β*-2-(thienyl)-L- or D-alanine to either Phe-CN (nitrile of L-phenylalanine) or D-Phe-CN. They were purified using RP-HPLC and identified by MALDI-TOF MS. The inhibitory activity of all isomers was evaluated in two systems: with the native enzyme from the neutrophil cell lysate and urine. The results showed that all the isomers of Thi-Phe-CN inhibit Cat C. Among the compounds bearing D-amino acid residues Thi-D-Phe-CN displayed the highest inhibitory activity, both in lysate and urine. These results suggest that the configuration of the amino acid residue in P_2 of the inhibitor is crucial for the high inhibitory activity against Cat C. Next, the stability of the obtained nitriles was assessed in human plasma. While Thi-Phe-CN was completely degraded during 3 hours of incubation, the isomers bearing D-amino acid residues were proteolysis-resistant. As a result, Thi-D-Phe-CN was applied for the control of the assay developed for the measurement of Cat C activity in sputa supernatants from cystic fibrosis patients.

The second series of the dipeptidyl nitriles was designed based on the electrostatic potential of human Cat C. The general formula of the compounds from the series was **Thi-X-CN**, where X stands for a basic amino acid residue such as 2,3-diaminopropionic acid (Dap), 2,4-diaminobutyric acid (Dab), Orn and Lys. The aim of these studies was to assess the influence of the basic amino acid residue in P_1 position on the affinity of the inhibitors towards Cat C. The starting point of the syntheses were the following amino acid residues: Z-Dap(Boc), Z-Dab(Boc), Z-Orn(Boc) and Boc-Lys(Fmoc), that were firstly used to obtain Boc-protected dipeptides (Boc-Thi-X(Boc)) and then transformed to the corresponding amides. Nitriles were obtained in the dehydration reaction of the amides with trifluoroacetic anhydride. The final products were purified by RP-HPLC after Boc deprotection. The inhibitory activity of the obtained nitriles was studied in an assay with recombinant Cat C. I was interested to see whether the designed compounds were more potent inhibitors than Thi-Phe-CN. I observed that the extension of the length of the side chain of the basic amino acid residue resulted in the increase in the inhibitory potency towards Cat C and that Thi-Lys-CN was the most potent inhibitor within the series; however, it was less potent than Thi-Phe-CN.