Popular science summary of:

"Production of recombinant TEV protease in *Escherichia coli* and analysis of enzymatic reactions"

Recent experiments, investigated in this thesis, may have shown that the TEV protease sequence recognition specificity is not as stringent as previously thought. This due to the possible observed hydrolysis of the substrate β-casein when digested with recombinant TEV protease. This new discovery may lead to new opportunities when using TEV protease as a tool in the lab as well as unexpected challenges.

The stringent sequence specificity of TEV protease has been studied by many and it is well-known by now that this recognition sequence consists of the amino acids ENLYFQ*G/S (where the asterisk marks the point of hydrolysation). The G, at the P1' position, can be substituted by several other amino acids, including S (Serine), without compromising the recognition pattern for the enzyme. The TEV protease has an auto-lysis tendency which is important to counter and can be done by fusing the fusion protein MBP (maltose-binding protein) to the protease, which prevents this detrimental function. The typical activity for the enzyme is to hydrolyse peptide bonds between Q (Glutamine) and G (Glycine), making the protease usable as a tool for precise removal of affinity tags. Recent experiments however have possibly shown that this specificity is not as strict as previously considered. The investigation in this thesis was conducted on the protein β-casein from bovine milk (UniProt accession no. P02666), which has an amino acid sequence similar to the recognition site for TEV protease between amino acids 131-142 (ENLHLPLPLLQS for β-casein compared to ENLYFQG for classic TEV protease recognition). The investigation started by docking the 12 amino acid long ligand to TEV protease which acts as the receptor in the UCSF Chimera program, which is a software allowing in-depth visualisation of molecular structures. The results showed a possible interaction between the thiol part of the cysteine (Cys151) in the active site of TEV protease to the carbonyl atom between amino acid residues Q and S on the β-casein ligand. This discovery led to the need for more experimental assays in the lab to further strengthen the results.

After successfully producing the recombinant his-tagged TEV protease in the host organism *E. coli*, the first assay to confirm a digestion of β-casein was the colorimetric assay using Folin-Ciocalteu´s phenol reagent. The reagent consists of the two acids phosphomolybdate and phosphotungstate, which are yellow in their oxidized form but turn blue after reacting with reducing phenolic compounds. The phenolic compounds are usually released upon digestion of substrates by enzymes. The blue colour is directly proportional to the concentration of the released reducing phenolic compounds and can be measured using a spectrophotometer to measure absorbance. The quantification can be done by comparing to a phenolic compound

used as a standard with known concentrations. In this case, tyrosine was used as a standard to measure the proteolytic activity. The assay showed that a phenolic compound was released upon interaction between β-casein and TEV protease, however further investigation was needed to confirm a digestion by the enzyme. Another assay used was SDS-PAGE technology, which showed slightly more bands on the β-casein control (undigested protein) compared to digested β-casein. To see if these bands corresponded to a peptide obtained after cleavage with TEV protease, a more detailed evaluation had to be done using MALDI-TOF mass spectrometry. At first glance, the results from these showed very little difference between the β-casein control (undigested) and the digested β-casein, however a peak at approximately 17.5 kDa could be observed on the digested sample of β-casein and not (or at least much less) on the unhydrolyzed control. Another small peak was observed at 7649.959 Da in the digested sample. Both polypeptide fragments obtained after digestion correspond well to the weights (when checked on Expasy´s ProtParam tool on the internet) of these peptides after a possible digestion at the hypothesized location of the β-casein. The first weight includes the N-terminal of the polypeptide to the Q (Glutamine) while the second polypeptide fragment includes the S (Serine) after digestion and downstream to the C-terminal. If TEV protease did, in fact, digest the β-casein at this position, there could be problems when using this protease in the future as a reliable tool for precise protein processing. However, to confirm these obtained results more extensive analysis would be needed.

The activity of the recombinant TEV protease was later investigated using a synthetic polypeptide constructed specifically as a target for the protease. This synthetic polypeptide had an 8-time repeated amino acid sequence where each repetition ended in the recognition site for TEV protease and a histidine-tag at the very end. When running the SDS-PAGE and mass spectrometry assay with both an undigested polypeptide and a digested sample, a difference could be observed. A peak at roughly 2403 kDa could be observed in the digested polypeptide but not in the control. This peak corresponded well to the weight of one of the 8-time repeated peptides, indicating a successful cleavage of this part of the polypeptide. The peptide products were later analysed for bioactive properties and showed a dual-inhibiting function (antihypertension and antidiabetic). This makes for a possible production of a multi-functional therapeutic agent in the future.

Fisnik Nerjovaj

Lund, June 2024