Chemical tools for the study of proteolytic activities associated with antigen presentation

Swieten, Paul Franciscus van

General introduction

This thesis describes the development of strategies to study the proteasome and cysteine proteases of the cathepsin family. Key to all strategies is the use of an electrophilic trap, which allows covalent labeling of the proteolytic activities with a number of desirable properties, depending on the question at hand. The development of such labels, often based on irreversible inhibitors, is not unique to this thesis. Indeed, many research efforts on the development of irreversible protease inhibitors have appeared in the literature over the past decades.[1] These probes are widely regarded to be very useful in medicinal and biological research, both for the purpose of drug development and, perhaps more importantly, for the assessment of protease activities in complex biological samples in what is now called chemical, or functional, proteomics.[2]

An ideal protease label in the field of proteomics should have a number of functionalities embedded in its structure. These include comparable and high binding to all members of a given class of proteases, but no reactivity towards other components in the biological sample. Other useful attributes are cell permeability and the means to track (for instance via a fluorescent tag) and/or retrieve the modified proteases. A further attractive feature would be the ability to monitor the relative abundance or proteolytic activities stemming from different, but closely related biological samples. Obviously, such ideal tools

are not within reach for each class of proteolytic activities present in nature. The first requirement is the ability to covalently and irreversibly modify proteolytic activities. This can be achieved with relative ease in case a side chain heteroatom of the protease(s) at hand participates in proteolysis, exerting a nucleophilic attack on the peptide bond of the substrate. Fortunately, this is the case for many protease families,[1] for instance the threonine secondary alcohol (the proteasome), the serine primary alcohol (serine proteases such as trypsin) and the cysteine thiol (many cathepsins, caspases, ubiquitin isopeptidases). The highly reactive nucleophiles can be used as a site for labeling of the protease via selective alkylation. However, protease families that recruit a water molecule for the initial nucleophilic attack (matrix metalloproteases, aspartic proteases) have proven to be more resistant towards chemical proteomics strategies employing protease inhibitor tools.[3] A broad spectrum irreversible protease inhibitor is an appropriate starting point for designing protease labels. Some restraint is required, though, because covalent attachment of an identification or isolation tag to a broad spectrum irreversible protease inhibitor may abolish cell permeability or may result in a drop in activity.[4]

The aim of this thesis is to address some of the issues raised in the previous paragraph. In case it is not possible to include all desirable properties of broad spectrum activity, cell permeability and identification potential in one chemical probe, an alternative would be to modify proteases in a two-step procedure: first labeling with a cell-permeable probe and second selectively functionalizing the labeled protease with an identification moiety in a cell-free environment. One solution, presented in this thesis for the two-step labeling of the proteasome, relies on the use of the Bertozzi-Staudinger ligation,[5] one of the major recent breakthroughs in the development of selective reactions in complex mixtures of biological origin. A second advance in protease probe design is the inclusion of an isotope encoded spacer, described in this thesis as a set of 'light' (H8) and 'heavy' (D8) probes that inhibit a broad array of cathepsin cysteine proteases. This strategy extends the isotope coded affinity tagging (ICAT) strategy as developed by Aebersold and coworkers,[6] enabling the monitoring of relative levels of protease activity, rather than levels of expression. The ICAT strategy is based on the alkylation of cysteine residues in two samples with either a heavy or light reagent, the mass difference of the products stemming from isotopes incorporated. The underlying idea is that the incorporated stable isotopes neither influence the reactivity of the alkylating agent towards cysteine thiols nor influence its chromatography properties. The two samples are combined and prepared for mass

spectrometric analysis, where fragments containing alkylated cysteine residues serve as mutual internal standard. The relative abundance of proteins in the two samples becomes apparent through comparing the magnitudes of the signal pairs originating from cysteine containing fragments.

Although the probes described in the research chapters in this thesis can be used to study either the proteasome or the cathepsin family of cysteine proteases, the general strategy upon which their design is based should be easy to transfer to probes for studying other classes of biomolecules. As such, the contents of this thesis fall within the broad area of research that can be described loosely as 'Chemical Biology'.[7] Although not new, especially the last decade has witnessed an explosion of research activities that take place at the interface of chemistry and biology. Next to the two examples briefly outlined above (bioorthogonal labeling, ICAT based proteomics), diverse areas of research such as chemical genetics (disabling gene products through inhibition with a chemical compound, rather than silencing the gene),[8] but also engineering the actual genetic code[9] are considered to belong to the Chemical Biology area. A specific area of Chemical Biology research with high relevance for part of the contents of this thesis is highlighted in **Chapter 1** and encompasses the current status in the two-step labeling of selected biomolecules in live cells.

Chapters 2 and 3 describe a new strategy to decorate the proteolytic activities of the 26S proteasome with a biotin moiety in a two-step strategy. The 26S proteasome is the business end of the ubiquitin-proteasome machinery that is responsible for the degradation of the majority of cytosolic and nuclear proteins in eukaryotes.[10] The degradation and turnover of proteins is indispensable for living cells, both to remove malfunctioning or redundant proteins, and to allow reuse of amino acids in the synthesis of other proteins. Attachment of a polyubiquitin chain (ubiquitin is a 76 residue signal protein) to lysine residues in a given protein marks it for degradation. The proteasome processes ubiquitinylated polypeptides to produce oligopeptides of a length varying from three to about 20 amino acid residues. The bulk of these are then further processed through the action of downstream aminopeptidases. A small portion, however, will (after potential further trimming by aminopeptidases) escape to the endoplasmic reticulum and complex to major histocompatibility complex (MHC) class I assemblies for presentation to the immune system.[11] As such, the proteasome is indispensable to mammalian immunity in monitoring

the cytosolic and nuclear protein content which may, next to self proteins, also encompass viral encoded proteins.

The 26S proteasome is assembled from a 20S core proteolytic particle to which two 19S cap protein assemblies are associated. Substrate recognition takes place at the 19S caps, where also protein unfolding and ubiquitin removal activities are located. The actual protein degradation takes place within the 20S core particle, which consists of 28 protein subunits. Of these, 14 α-subunits make up the outer parts as two rings of 7 subunits each, and 14 β-subunits make up the inner parts, again as two rings of 7 subunits. The 7 α-subunits in each ring, although highly homologues, have an unique sequence and the same holds true for the 7 individual β-subunits in the inner rings. Three of the β-subunits in higher organisms possess catalytic activity, each with a distinct (although as yet poorly understood) substrate preference. To complicate matters further, higher organisms may express a second set of proteasome proteolytic activities, which also assemble into a 20S particle. This particle, induced by specific cytokines (for instance interferon gamma) that are the result of initial immune response processes, is therefore called the immunoproteasome,[12] and it also harbors three distinct catalytic activities. In **Chapter 2** a two step labeling strategy is described that enables the visualization of the combined six proteolytic activities of both the constitutively expressed proteasome and the immunoproteasome with almost equal efficiency. The strategy hinges on the use of two synthetic tools, one being a peptide vinyl sulfone broad spectrum proteasome inhibitor equipped with an azide, and the other a biotinylated phosphine. After addition of the vinyl sulfone compound to growing cells, labeling of all proteasome catalytic subunits with azido groups occurs. Cell lysis is followed by denaturation of the protein content. In the next step the azide functionalities are addressed through Staudinger ligation, resulting in the attachment of a biotin to each catalytic subunit. These are now visualized through SDS-PAGE gel electrophoresis followed by streptavidin blotting. **Chapter 3** describes the application of this two-step labeling strategy in the identification of a peptide vinyl sulfone based inhibitor that specifically inhibits one of the three proteasome proteolytic activities, namely the one with preference to cleave after acidic residues (β1 or the caspase-like activity). Further, this inhibitor was found to specifically inhibit the corresponding immunoproteasome subunit (β1i), making this compound potentially useful in the study of the involvement of the immunoproteasome in MHC class I antigen presentation.

Chapters 4, 5 and 6 describe the development of tools to modify and analyse lysosomal cysteine proteases of the cathepsin class.[13] Lysosomal proteases, of which the cathepsins constitute the major class, are responsible for the degradation of proteins that are collected into the lysosomal compartment through phagocytosis, pinocytosis or autophagy. Proteins targeted for lysosomal degradation can be either cell surface membrane associated proteins or proteins from the extracellular matrix, and may be either self proteins or proteins originating from host pathogens, typically bacteria. As such, the body of lysosomal proteases complement the ubiquitin-proteasome pathway, in that the latter degrades most intracellular proteins and the former takes care of the extracellular protein content. Another parallel is their role in the adaptive immune system: a small portion of the oligopeptides resulting from lysosomal proteolysis escape further degradation and are presented on the cell surface by MHC class II molecules.[14] In this way, also the protein content of the extracellular matrix, including that arising from bacterial infection, is monitored by the immune system. This thesis describes the development of a number of advanced chemical probes for the study of cathepsin activity in a number of different settings. All probes are based on the leucine epoxysuccinate motif as present in the parent compound known in the literature as DCG-04,[15] which was developed by Bogyo and coworkers as a broad-spectrum cysteine protease label. **Chapter 4** presents the development of a set of isotope-encoded activity based probes for the assessment of the relative abundance of cysteine proteases. The probes are based on the aforementioned DCG-04, but contain either eight protons or eight deuterium nuclei. The first inroads towards the use of these compounds in biological studies and the difficulties in these are discussed. In **Chapter 5** a strategy is presented through which the progress of an extracellular protein through the endocytic pathway and the cysteine proteases it encounters along that way can be monitored.[16] A broad-spectrum cysteine protease label equipped with a biotin moiety is linked to ovalbumin as a model protein through a photocleavable linker. Initial studies in which the conjugate is incubated with recombinant cathepsin S followed by photocleavage are presented. **Chapter 6** makes use of a related strategy. A cysteine protease probe is linked to nona-arginine, a peptide known for its ability to deliver cargo to the cytosol.[17] Endocytosis studies demonstrate that this conjugate labels at least some of the pool of cathepsin cysteine proteases, thereby indicating that cell penetrating peptides of this type do make use of endocytic processes in reaching the cytoplasm. Finally, some concluding remarks and future prospects are presented.

References

1 J. C. Powers, J. L. Asgian, O. D. Ekici and K. E. James, *Chem. Rev.* 2002, **102**, 4639-4750.

2 D. A. Jeffery and M. Bogyo, *Curr. Op. Chem. Biol.*, 2003, **14**, 87-95.

3 A. Saghatelian, N. Jessani, A. Joseph, M. Humphrey and B. F. Cravatt, *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 10000-10005.

4 B. M. Kessler, D. Tortorella, M. Altun, A. F. Kisselev, E. Fiebinger, B. G. Hekking, H. L. Ploegh and H. S. Overkleeft, *Chem. Biol.*, 2001, **8**, 913-929.

5 M. Kohn and R. Breinbauer, *Angew. Chem., Int. Ed.*, 2004, **43**, 2-12.

6 S. P. Gygi, B. Rist, S. A. Gerber, F. Turecek, M. H. Gelb and R. Aebersold, *Nat. Biotechnol.*, 1999, **17**, 994-999.

7 A perspective on the field of chemical biology is given: S. L. Schreiber, *Nat. Chem. Biol.*, 2005, **1**, 64-66.

8 D. R. Spring, *Chem. Soc. Rev.*, 2005, **34**, 472-482.

9 L. Wang and P. G. Schultz, *Angew. Chem., Int. Ed.*, 2005, **44**, 34-66.

10 K. L. Rock and A. L. Goldberg, *Annu. Rev. Immunol.*, 1999, **17**, 739-797.

11 J. W. Yewdell, E. Reits and J. Neefjes, *Nat. Rev. Immunol.*, 2003, **3**, 952-961.

12 B. J. van den Eynde and S. Morel, *Curr. Opin. Immunol.*, 2001, **13**, 147-153.

13 K. Honey and A. Y. Rudensky, *Nat. Rev. Immunol.*, 2003, **3**, 472-482.

14 C. Watts, *Nat. Immunol.*, 2004, **5**, 685-692.

15 D. Greenbaum, K. F. Medzihradszky, A, Burlingame and M. Bogyo, *Chem. Biol.*, 2000, **7**, 569-581.

16 A.-M. Lennon-Duménil, A. H. Bakker, R. Maehr, E. Fiebinger, H. S. Overkleeft, M. Rosemblatt, H. L. Ploegh and C. Lagaudrière-Gesbert, *J. Exp. Med.*, 2002, **196**, 529-539.

17 S. Futaki, *Int. J. Pharm.*, 2002, **245**, 1-7.