Beilstein-Institut

ESCEC, March 19th – 23rd, 2006, Rüdesheim/Rhein, Germany

Investigation of Proteases – Suggestions

Hartmut Schlüter

Core Facility Protein Analytik, Charité,
Tucholskystr. 2, 10117 Berlin, Germany

E-Mail: hartmut.schlueter@charite.de

Received: 6th November 2006 / Published: 31st August 2007

Proteases are enzymes catalysing the hydrolysis of peptides or proteins. They are key players in a wide range of biological processes such as the release of peptide hormones, nutrient acquisition, cell growth, differentiation, antigen processing and protein turnover, in all living organisms. Furthermore it is becoming more and more obvious that the abnormal functioning of some proteases may lie behind several types of diseases, including inflammation, cancer and Alzheimer's disease. Therefore proteases are attracting an increasing interest.

The MEROPS database, which is specialized in proteases, lists 555 known and putative genes encoding proteases in *Homo sapiens* (31st of August 2006). The number of proteins acting as proteases in the human organism may even be much higher, since proteins can develop proteolytic activities although they are not assigned as proteases. The protein disulfide isomerase A3 (PDIA3, primary accession number: P30101), which main function is protein folding, is an example for the latter case [1,2], since Kito, Urade and coworkers published convincing data about a protease activity of PDIA3 [1].

From many of these protease-encoding genes the endogenous substrates and as a result the physiological roles are as yet unknown. One strategy for deciphering the physiological roles of proteases is to start with known reaction products of the proteolytic action of unknown proteases (Fig. 1). For example, the peptide urotensin-II, a potent vasoconstrictor, is cleaved from its inactive urotensin-precursor by the proteolytic action of an unknown protease. The knowledge of the sequence of both, the peptide urotensin and its precursor, allows a reaction specific probe (substrate) to be developed, which can be used in an assay like the MES (mass spectrometry assisted enzyme screening system) [1] for detecting urotensin-II-generating activity. After having developed the enzyme assay it can be used for screening for the presence of the target enzyme in protein fractions and guiding the purification of the target enzyme to near homogeneity (Fig. 1). For the identification of the

target enzyme the purified active fraction can be subjected to enzymatic cleavage and mass spectrometric analysis of the enzymatic peptide fragments followed by database research. After protein identification the results must be validated by either expressing recombinantly the identified protein candidate or by simply purchasing it, if possible and by demonstrating that the enzymatic activity and properties are identical to that of the purified enzyme and its properties.

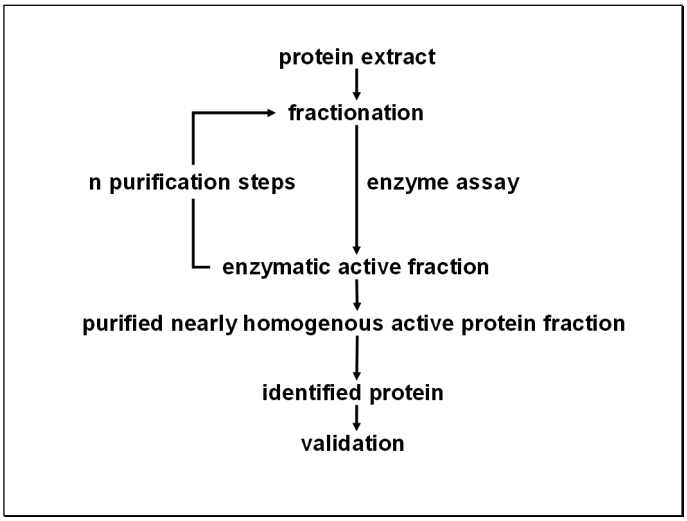

Figure 1 Scheme showing the strategy for detection, purification, identification and verification of proteases from protein extracts.

The MES-system is one of the core instruments within the protease-deciphering strategy. Figure 2 demonstrates a typical read out of the MES system. In this case 2 different protein fractions were monitored for angiotensin-II-generating activity by incubating the proteins with the reaction specific probe angiotensin-I. Since the signal intensity of the reaction product angiotensin-II in the incubate of fraction A increases faster with increasing incubation time than in fraction B, the angiotensin-II-generating activity of fraction A is higher.

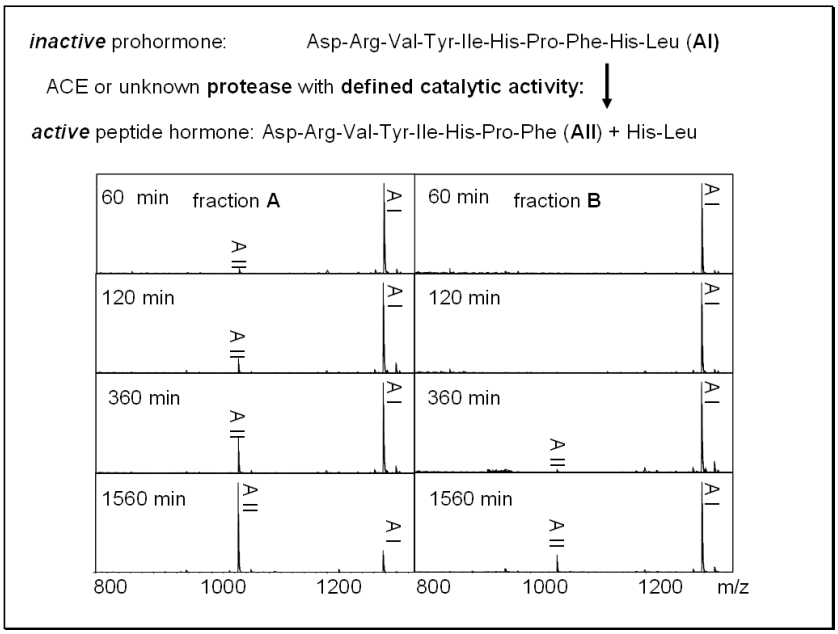

Figure 2. MES results of a MES assay monitoring angiotensin-II-generating activity of 2 chromatographic fractions A and B. MALDI-MS spectra of the reaction products of the incubation of immobilized proteins of fraction A and fraction B derived from porcine renal tissue. AI: Angiotensin I; AII: Angiotensin II; A(1 – 7): Angiotensin (1 – 7).

Mass spectrometry based enzyme assays are advantageous compared to UV- or fluorescence based enzyme assays because they give information about the identity of the reaction products and about the fate of the substrate. The control of the identity of the reaction products reduces the risk of false positive results. Being able to monitor the fate of the substrate gives the opportunity to notice the presence of additional proteolytic activities accompanying the target protease. An example for this latter case is given in Fig. 3A. The MES mass spectrum was obtained after incubating a crude protein fraction obtained from porcine renal tissue with angiotensin-I. In the spectrum in Fig. 3A, beside the signal of angiotensin-II, additional signals point to the presence of several other peptidases. The peptide des-Asp-A-I may be generated by the enzyme ACE-II, which is known to be present in renal tissue and A(1 – 7) by the renal peptidase neprilysin. With increasing purity of the angiotensin-II-generating activity the number of the additional peptides decreases. Nearly homogenous fractions yield the signal of the reaction product of the target protease only (Fig. 3B).

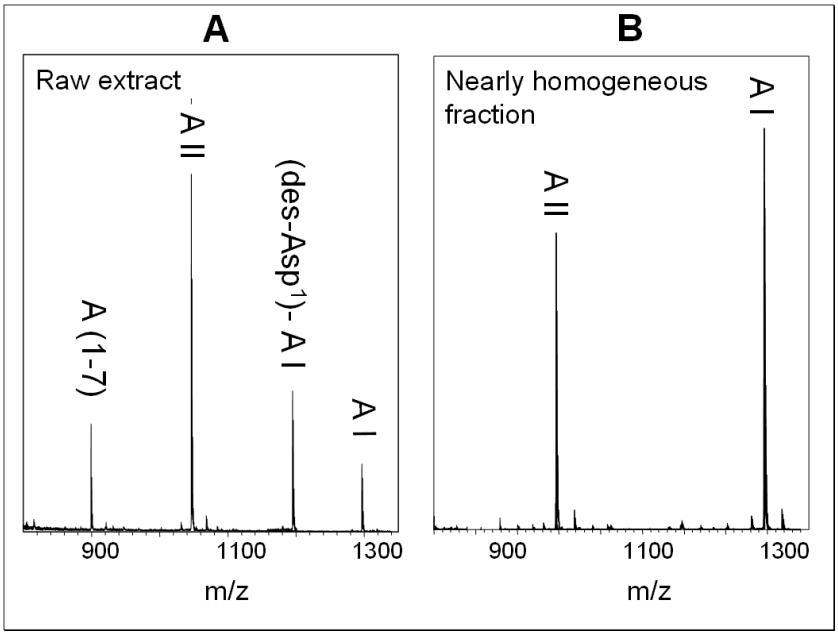

Figure 3. Typical MES results of a MES assay monitoring angiotensin-II-generating activity: MALDI-MS spectra of the reaction products of the incubation of immobilized proteins of a raw extract of porcine renal tissue (A) and of proteins from a fraction purified to near homogeneity. AI: Angiotensin I; AII: Angiotensin II; A(1 – 7): Angiotensin (1 – 7); (des-Asp^{1})-A-I: des-asparaginc acid angiotensin I.

The experience with the mass-spectrometry based assay system MES results in the 1st suggestion:
Control experiments should include a mass spectrometric analysis of the enzymatic reaction products thus minimizing false positive results, verifying the chemistry of the catalytic conversion and being able to detect other accompanying enzymatic activities, which may interfere with the target enzyme.

After purifying the active fraction to near homogeneity the protein will be identified (Fig. 1). Usually within the purified fractions not only one but several proteins are identified. Therefore the question arises as to which of the identified proteases may have proteolytical activities. Comparison of the own experimental data describing the properties of the protease with those described in the literature helps to verify the identification data. The verification procedure may be easy, if the identified protein is known as a protease, the proteolytic activity is its main function and the enzymatic properties are well described. However proteins may be identified, which have several different functions. An example for the latter case is the protein disulfide isomerase A3 (PDIA3_human, P30101), which major catalytic property comprises the disulfide isomerase activity. In such cases it has to

be proven, if the proteolytic activity is physiologically relevant. Therefore a comprehensive database analysis and analysis of the original papers is necessary. If this work will give more confidence about the proteolytic activity of the candidate, the protein should be recombinantly expressed to prove its proteolytic activity experimentally.

Performing the database analysis about the candidate usually is accompanied with some trouble, arising from the many synonyms often used for a protein encoded by one single gene and the still missing standardization of the nomenclature of proteins. Here an example will be given. The database SwissProt summarizes the following molecular functions of PDIA3: Cysteine-type endopeptidase activity, phospholipase C activity, protein disulfide isomerase activity, protein retention in ER, protein import into nucleus and signal transduction. In the next step the original papers have to be searched for. Using the synonyms shown in SwissProt for PubMed database searches yielded the results given in Table 1. Because of the confusion concerning nomenclature, some authors used several synonyms within the title of their papers: "Association of the chaperone **glucose-regulated protein 58** (**GRP58/ER-60/ERp57**) with Stat3 in cytosol and plasma membrane complexes" [1, 2].

Table 1. The numbers indicate the hits in the database search of PubMed performed with the synonyms of PDIA3 without and with additional keywords.

Keywords **Synonyms**	**Σ**	**Isomerase** **+ human**	**Isomerase** **+ human + protease**
ERp57	119	73	**5**
p58	**480**	**8**	1
58 kDa microsomal protein	0		
58 kDa glucose regulated protein	0		
ER60	23	10	**4**
ER-60	32	10	**5**
ERp60	**833**	73	**8**
PDIA3	98	97	**10**

Table 1 lists only a few of many synonyms known for proteins encoded by the gene PDIA3. A PubMed data base search was performed with each of the synonyms, with the combination of the synonyms with the key words "isomerase" and "human" and with the keywords "isomerase", "human" and "protease". The synonyms "p58", "58 kDa microsomal protein", "58 kDa glucose regulated protein" yielded the worst results, especially "p58", but also "ERp60" yielded a huge number of false positive results. More helpful are the synonyms "ER-60", "ER60" and "PDIA3". However, only in those papers, where the gene sequence of the amino acid sequence or the complete amino acid sequence is given, one can be sure, that the protein described is identical with that described under PDIA3 in SwissProt. Therefore in many cases there remains some doubt, whether the properties described in the publication really belong to PDIA3.

Suggestion 2:

A standardized nomenclature for enzymes (and all other proteins) is needed, e. g. the accession number, which gives an unambiguous hint towards its origin and which should be used by all data bases and all journals!

As soon as the database research has yielded positive results towards the proteolytic function of the candidate protein a validation by an appropriate experiment is necessary. In some cases a recombinant expression of the candidate protein may be circumvented because it can be purchased. Independently from the source of the candidate protein it is strongly recommended that both the identity and the purity of the candidate protein preparation are checked. The following example demonstrates the importance of this recommendation: A protease preparation was purchased, here named protease 1. A size exclusion chromatography (SEC) of the protease 1 containing fraction was performed (Fig. 4). The UV-absorption profile of the SEC chromatogram (Fig 4A) monitored at 280 nm already shows that the protease 1 containing fraction is not pure. The protease 1 activity of the fractions of the SEC was measured with an appropriate substrate. The protease 1 activity co-eluted with the fraction with the highest UV absorption. An LC–MS analysis of the tryptic peptides of the active fraction confirmed the identity of protease 1. Furthermore the fractions were tested for the target enzyme activity with the MES assay described above. Surprisingly the target enzyme activity eluted in front of the protease 1 fraction. The purchased protease 1 fraction was chromatographed with an affinity chromatography and an aliquot of the eluate was applied to the size exclusion chromatography again (Fig. 4B). In the resulting fractions no protease 1 activity was detected any more beside the target protease activity. An LC–MS analysis of the fraction with the target protease activity confirmed that the target protease is not identical to protease 1.

The example of the impurified protease 1 demonstrates the need for controlling the identity of the protease responsible for the activity as well as the control of the purity. Without controlling the purity the false protease would have been assigned with the defined proteolytic reaction.

The problem of impurities is also given with recombinantly expressed proteins, even after passing affinity chromatography. Figure 5 shows the SDS–PAGE analysis of the lysate of the host cells expressing a recombinant protein with a His-tag and the eluate of a immobilized-Ni affinity chromatography (Ni-IMAC). Therefore suggestion 3 is recommended.

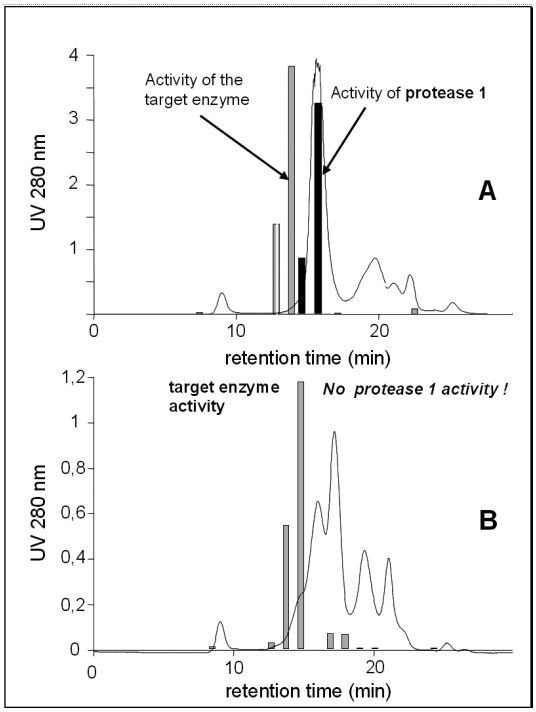

Figure 4. Chromatograms of size exclusion chromatographies of a purchased protein fraction before (A) and after purification (B) with an affinity chromatography. Black bars: Activity of protease 1. Grey bars: Activity of the target protease.

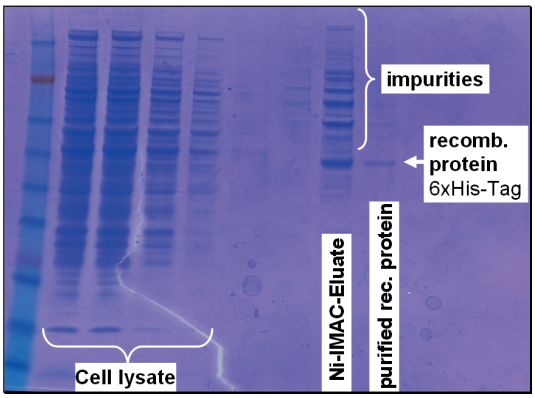

Figure 5.

Suggestion 3:

Publishing data about the properties of an enzyme should also include 1, proof of the purity of the enzyme fraction and 2, proof of the identity of the enzyme. The proof of the purity should not be performed by *SDS–PAGE* but 2DE including identification of all proteins visible in the 2DE-gel *or* tryptic digest followed by LC–MS/MS analysis.

References

[1] Spiess, C., Ehrmann, M. (1999) A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. *Cell* **97:**339–347.

[2] Krojer, T., Garrido-Franco, M., Huber, R., Ehrmann, M., Clausen, T. (2002) Crystal structure of DegP (HtrA) reveals a new protease chaperone machine. *Nature* **416:**455–459.

[3] Urade, R., Oda, T., Ito, H., Moriyama, T., Utsumi, S., Kito, M., (1997) Functions of characteristic Cys-Gly-His-Cys (CGHC) and Gln-Glu-Asp-Leu (QEDL) motifs of microsomal ER-60 protease. *J. Biochem.* **122:**834–842.

[4] Schlüter, H., Jankowski, J., Rykl, J., Thiemann, J., Belgardt, S., Zidek, W., Wittmann, B., Pohl, T. (2003) Detection of protease activities with the mass-spectrometry-assisted enzyme-screening (MES) system. *Analyt. Bioanalyt. Chem.* **377:**1102–1107.

[5] Guo, G.G., Patel, K., Kumar, V., Shah, M., Fried, V.A., Etlinger, J.D., Sehgal, P.B. (2002) Association of the chaperone glucose-regulated protein 58 (GRP58/ER-60/ERp57) withStat3 in cytosol and plasma membrane complexes. *J. Interferon Cytokine Res.* **22:**555–563.

[6] Kita, K., Okumura, N., Takao, T., Watanabe, M., Matsubara, T., Nishimura, O., Nagai, K.(2006) Evidence for phosphorylation of rat liver glucose-regulated protein 58,GRP58/ERp57/ER-60, induced by fasting and leptin. *FEBS Lett.* **580:**199–205.