CHARLES UNIVERSITY IN PRAGUE

FACULTY OF SCIENCE

STUDY PROGRAMME: BIOCHEMISTRY

Mgr. ALAN KÁDEK

STUDY OF CONFORMATIONS AND CONFORMATIONAL CHANGES OF PROTEINS USING MASS SPECTROMETRIC METHODS

STUDIUM KONFORMACÍ A KONFORMAČNÍCH ZMĚN PROTEINŮ METODAMI HMOTNOSTNÍ SPEKTROMETRIE

DOCTORAL THESIS

Supervisor: RNDr. Petr Man, Ph.D.

PRAGUE 2016

To my dearest parents Zdeňka and Florián,
and to my beloved wife Anna.

DECLARATION

I declare that I have worked on this thesis under the guidance of my supervisor and that all sources of the previous knowledge are properly cited. No part of this work was used and will not be used for obtaining any other academic degree than Ph.D. from Charles University in Prague.

Prague ……,

……

Mgr. Alan Kádek

DECLARATION OF AUTHORSHIP

I declare that Mgr. Alan Kádek contributed significantly to the experiments and to all 4 scientific publications contained in this Ph.D. thesis. He performed most of the experiments, substantially contributed to their planning, and took a significant part in the primary data interpretation and their preparation for publication.

Prague ……,

……

RNDr. Petr Man, Ph.D.

ACKNOWLEDGMENTS

At this place my thanks are due to many people, who helped me throughout my studies and supported me in every possible way.

First and foremost, my most sincere thanks go to my parents, Zdeňka and Florián, for their endless love and support throughout the years.

Next, I would like to express my gratitude to my supervisor Petr Man, who guided me during my studies and kept me busy, but happy. Moreover, he allowed me to take part in all the tasks of scientific life and also let me travel abroad to many wonderful conference destinations and study stays, which is very much appreciated.

Thanks are also due to all the people in the Laboratory of Structural Biology and Cell Signaling and the Laboratory of Molecular Structure Characterization at the Institute of Microbiology CAS both in Prague and in Vestec. Above all, I would like to name the other three Peters - Petr Novák, Petr Halada, Petr Pompach, Zdenek Kukacka and also all the other colleagues – either postdocs, graduate or diploma students, who helped me and made the lab such a good and friendly place to work in.

Finally, a huge thank you is also meant for all the collaborators, above all the guys from Austria – Roland Ludwig, Alfons Felice and Daniel Kracher, who supplied me with lots of enzymes to play with. Also all the people who I met abroad during my study stays in Oxford and Strasbourg are fondly remembered – namely Julien Marcoux, who spent a lot of time with me teaching me new tricks.

Last but definitely not least I must thank a lot my beloved wife Anna, for all her care, love and patience.

ABSTRACT (in English)

Mass spectrometry (MS) techniques have, over the last twenty years, found their stable place in the structural biology toolkit. They are not only employed to provide information on the protein primary sequence, but are increasingly used to probe higher orders of protein structure as well. They may not boast the atomic resolution and the ability to directly provide structural coordinates, but on the other hand suffer from very few experimental limitations as they are able to work under native conditions in solution, provide data fast, with low sample consumption and for proteins and complexes of vastly differing sizes. Perhaps most importantly, they may often be employed to study conformational dynamics of proteins and can thus complement other methods with higher spatial resolution in integrative structural biology approaches.

The main focus of this Ph.D. thesis was hydrogen / deuterium exchange coupled to MS (HXMS), which is one of the most widespread structural MS methods. Recombinantly produced aspartic protease nepenthesin-1 from *Nepenthes* pitcher plants was characterized, immobilized and extensively tested with the intention to expand the portfolio of aspartic proteases in HXMS workflow and to improve the spatial resolution of the technique.

Following successful implementation of nepenthesin-1 into the HXMS protocol, it was used in combination with rhizopuspepsin in analyses of a challenging highly-flexible cellulolytic enzyme cellobiose dehydrogenase (CDH). This biotechnologically interesting protein was studied in order to obtain structural explanation for the regulation of its activity by pH and divalent cations. For this, HXMS was complemented by native MS with ion mobility and protein surface electrostatics calculations.

Together, the structural MS and computational techniques brought some interesting observations concerning the analyses of transient protein complexes. Moreover, they provided a direct experimental proof that repulsion of negative charge patches close to interdomain interface in CDH is the key mechanism governing its functioning in solution.

Keywords: Structural mass spectrometry, nepenthesin-1, hydrogen / deuterium exchange mass spectrometry (HXMS), native mass spectrometry with ion mobility (IMMS), aspartic protease, protein immobilization, cellobiose dehydrogenase (CDH), protein surface electrostatics, flavocytochrome, direct electron transfer.

ABSTRAKT (in Czech)

Hmotnostně spektrometrické (MS) techniky si v průběhu posledních dvou desetiletí našly své trvalé místo mezi nástroji strukturní biologie. Kromě získání informace o primární sekvenci proteinů jsou stále častěji využívány i pro studium vyšší strukturní organizace bílkovin. Nedosahují sice atomárního prostorového rozlišení, jsou ale naopak prosty řady experimentálních omezení. MS strukturní techniky tak jsou schopny studovat molekuly za nativních podmínek v roztoku, jsou rychlé, mají nízkou spotřebu vzorku a jsou použitelné pro molekuly a jejich komplexy s velmi širokým rozsahem velikostí. Možná nejdůležitější je však jejich schopnost poskytnout informace o konformační dynamice proteinů, které tak mohou doplnit data získaná jinou strukturní technikou s vyšším prostorovým rozlišením v rámci integrativní strukturní biologie.

V této disertační práci byla hlavní pozornost věnována technice vodík / deuteriové výměny v kombinaci s hmotnostní spektrometrií (HXMS), která je jednou z nejvíc rozšířených strukturních MS metod. Rekombinantně připravená aspartátová proteasa nepenthesin-1 z láčkovek rodu *Nepenthes* byla charakterizována, imobilizována a podrobně testována s cílem rozšířit portfolio proteas dostupných pro HXMS experimenty a zvýšit prostorové rozlišení této techniky.

Po úspěšné implementaci do HXMS experimentálního protokolu byl nepenthesin-1 využit v kombinaci s rhizopuspepsinem při analyticky náročném studiu vysoce flexibilního proteinu celobiosadehydrogenasy (CDH). Tento biotechnologicky zajímavý protein byl zkoumán s cílem strukturně objasnit mechanismy regulace jeho aktivity pomocí pH a divalentních kationtů. Za tímto účelem byla metoda HXMS doplněna nativní hmotnostní spektrometrií s iontovou mobilitou a výpočetním modelováním distribuce elektrostatického náboje na povrchu proteinů.

Společně tyto strukturní MS a výpočetní techniky přinesly zajímavé poznatky o analýze transientních proteinových komplexů. Zejména ale poskytly přímý experimentální důkaz, že repulse oblastí negativního náboje v blízkosti mezidoménového rozhraní CDH je klíčovým mechanismem řídícím fungování tohoto enzymu.

Klíčová slova: Strukturní hmotnostní spektrometrie, nepenthesin-1, vodík / deuteriová výměna s MS (HXMS), nativní hmotnostní spektrometrie s iontovou mobilitou (IMMS), aspartátové proteasy, imobilizace proteinů, celobiosadehydrogenasa (CDH), elektrostatika povrchu proteinu, flavocytochrom, přímý přenos elektronů.

TABLE OF CONTENTS

ABBREVIATIONS

CDH	cellobiose dehydrogenase
CNBr	cyanogen bromide
cryo-EM	cryogenic electron microscopy
CYT	cytochrome domain of CDH
Da	Dalton – SI-recognized mass unit equivalent to atomic mass unit
DH	dehydrogenase domain of CDH
ESI	electrospray ionization
ECD	electron capture dissociation
EDTA	ethylenediaminetetraacetate
Endo Hf	endoglycosidase H fusion protein
ETD	electron transfer dissociation
EX1 / EX2	kinetic regimes of hydrogen / deuterium exchange
FAD	flavin adenine dinucleotide
FRET	Förster resonance energy transfer
FTICR	Fourier transform ion cyclotron resonance mass spectrometer
HX	hydrogen exchange
HXMS	hydrogen / deuterium exchange coupled to mass spectrometry
IET	interdomain electron transfer
IMMS	ion mobility mass spectrometry
irNep-1	immobilized recombinant nepenthesin-1
irRpn	immobilized recombinant rhizopuspepsin
LC	liquid chromatography
LPMO	lytic polysaccharide monooxygenase
MALDI	matrix-assisted laser desorption / ionization
*Mt*CDH	cellobiose dehydrogenase from *Myriococcum thermophilum*
MS	mass spectrometry
MS/MS	tandem mass spectrometry
*Nc*CDH	cellobiose dehydrogenase from *Neurospora crassa*
Nep-1	nepenthesin-1
NMR	nuclear magnetic resonance
Pep	porcine pepsin A
QTOF	quadrupole and time-of-flight combined mass analyzer
Rpn	rhizopuspepsin from *Rhizopus chinensis;* protease XVIII
SAXS	small-angle X-ray scattering
SDS-PAGE	polyacrylamide gel electrophoresis in the presence of sodium dodecylsulphate
TCEP	tris(2-carboxyethyl)phosphine
TOF	time-of-flight mass analyzer
UV	ultraviolet radiation
wt	wild-type

1. INTRODUCTION

1.1. Proteins and their structure

Proteins are probably the most versatile molecules in nature. They evolved to function almost everywhere in cells and their biological roles range widely from purely structural, through biocatalytic and regulative to providing interactions and communication with the cells' environment. This impressive variability is allowed by their polymeric structure and the vast number of possibilities how to combine their twenty (or twenty two) amino acid building blocks.

Furthermore, the order of amino acids (protein primary structure) is not the only level of protein organization (**Figure 1**). Directly determined by the order of amino acids and their properties is so-called secondary structure, where parts of polypeptide chains form into sub-structures. These are alpha-helices or individual beta-strands, which in turn interact to form larger beta-sheets. In both alpha-helices and beta-sheets extensive hydrogen bond networks exist between C=O carbonyl and N-H secondary amine groups in peptide backbone. These bonds are weak, but in large enough numbers greatly stabilize the protein secondary structural elements and make them less polar. This is important for folded proteins, which most often have a non-polar core, where ordered elements of secondary structure are usually buried to minimize their contact with surrounding water. The relative three-dimensional arrangement of secondary structure elements is then termed protein tertiary structure and is mainly stabilized by interactions of amino acid side chains. Lastly, multiple protein chains can form supramolecular assemblies (quaternary

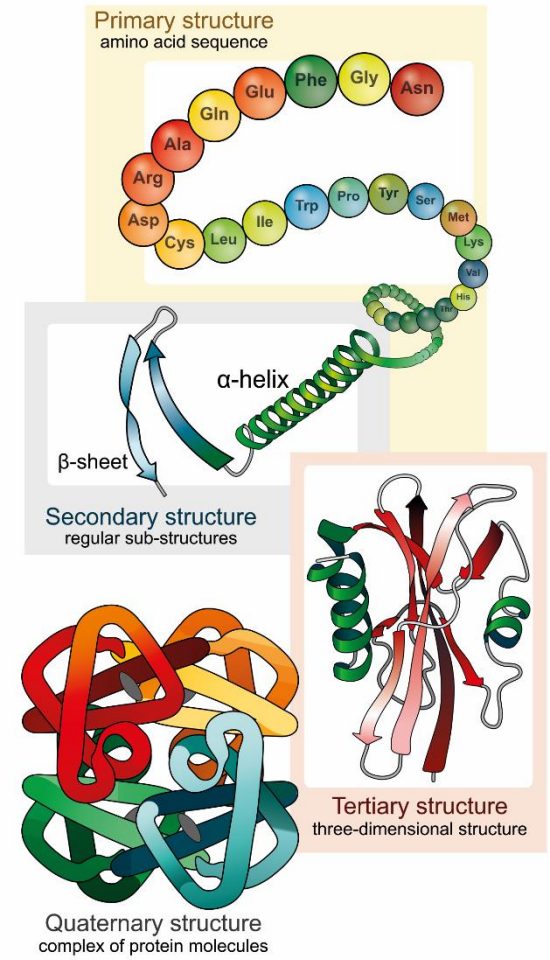


Figure 1: Four levels of protein structure. *Scheme representing protein structures from a chain of amino acids to supramolecular non-covalent assemblies. Source: wikimedia.org.*

structure), which are key for many biological processes involving concerted actions of several proteins.[1]

As the structure of protein is crucial to determine its function, it is conversely true that problems in the structural organization of a protein can affect its functional properties as well, which can lead to a range of diseases. Such diseases, generally termed proteopathies or proteinopathies, may involve point mutations in protein amino acid sequence resulting in the loss of function of enzymes or receptors – e.g. glucocerebrosidase in Gaucher's disease[2] or lipoprotein receptor in familial hypercholesterolemia[3]. Mutated proteins may also be misfolded and aggregate as insoluble plaque deposits in cells – such is the case for Aβ and τ proteins in Alzheimer's disease[4,5], α-synuclein in Parkinson's disease[6] or prion proteins in Creutzfeld-Jacob's disease[7]. Alternatively, conformation changes may result in a range of other less common effects in proteins, such as improper localization, incorrect degradation or gain of toxic function.[8]

However, even correct static protein structure is often not enough for a protein to properly play its role. Most biological processes require movement within proteins or between their subunits. Protein movements can be of varying nature and range from localized conformational changes in enzyme active site during catalysis[9], through allosteric effects modulating proteins' affinities to ligands such as in hemoglobin[10] or regulating their activity as observed in dimeric protein kinases[11]. Some protein movements even encompass large scale rearrangements of whole protein assemblies - with the most complex being proteins participating in active membrane translocation[12] or so-called molecular motors[13]. Protein conformational changes can also occur on very different time scales often ranging from microseconds to milliseconds[14], but sometimes even considerably longer such as in the case of slow biological processes[15].

Therefore, to properly understand proteins, it is necessary to study them in terms of their atomic structures as well as in the context of their conformational dynamics.[16]

1.2. Experimental methods to study protein structure and dynamics

In order to study both protein structures and dynamics a range of experimental methods was developed. These various biophysical properties of molecules, differ in sample and experimental condition requirements and provide data with varying levels of detail.

1.2.1. Protein crystallography with X-ray diffraction

Since the first three-dimensional structure of protein determined in 1958 for sperm-whale myoglobin by Sir John Kendrew[17], X-ray diffraction analysis of protein crystals has been the most successful method for determining high-resolution structures of proteins.[18] It utilizes the scattering of powerful X-ray radiation beams by electrons in sample. The diffraction reflections produced by individual molecules are very weak, therefore it is often necessary to work with protein monocrystals, where the molecules are arranged in a periodical lattice. This hugely increases the intensities of reflections in particular directions by constructive interferences of photons diffracting from repeating individual molecules in the crystal grid. From the observed reflections it is possible to deduce the three-dimensional map of electron density in the crystal, after experimentally or computationally finding complex phases of individual signals, which are necessary to deconvolute contributions of individual atoms in molecule by Fourier transform. Finally, atomistic protein structure is modelled into the electron density to provide experimental protein structure.[18,19]

One major advantage of X-ray crystallography is the fact that under favorable conditions the technique is able to provide structure with single atom resolution (in best cases close to or even less than 1Å)[18] and can work with a broad range of molecule sizes. Moreover, thanks to the long time the technique has been used, the theory underlying the experiments is well described, software is readily available and also the measurements and data interpretation are largely automated.[19]

On the other hand, common bottlenecks in X-ray crystallography are the processes of signal phase determination and the crystal preparation itself.[19] In many cases, proteins will not easily crystallize, which is often due to their heterogeneity or structural flexibility. And even when crystallization conditions are found, it is not common for the user to freely choose the solution conditions the protein crystallizes from. These are often far from native with very high protein concentrations, various additives and precipitants or non-optimal pH and ionic strengths. All this can in some cases lead to structures containing experimental artifacts when compared with structures in solution.[20,21] Last but not least, crystallography almost always provides static pictures of protein systems with strong preference for energetically favorable and thus stable conformers, largely ignoring the dynamic nature of proteins.

1.2.2. Nuclear magnetic resonance spectroscopy

Another technique, which supports determination of atomic resolution of structures is nuclear magnetic resonance (NMR). Contrary to X-ray crystallography, this technique most often works in solution (albeit highly concentrated) and utilizes a completely different biophysical principle. In NMR, atomic nuclei having uneven spin numbers (in biological samples namely ^{1}H, ^{13}C, ^{15}N and ^{31}P) may be studied, because energy levels of nuclei with opposite spin angular momentums can be separated in high external magnetic field. It is then possible to excite all the nuclei to the higher energy state through radio frequency pulses and then measure signal as energy is released by nuclear relaxation. By utilizing ingenious sequences of excitation pulses[22] it is then possible to deduce from the measured signals information about the chemical surroundings of individual "visible" atoms, their connectivities and interatomic distances. These are provided as spatial distance constraints, which are then in high enough numbers used to model the three-dimensional structure of the studied molecule.[19,22]

Major advantage of NMR for studying proteins lies in its ability to determine atomic-level structures of molecules in solution and most importantly to study their dynamics. It can also sample lower occupancy states of molecules and not only the most energetically favorable conformations. This leads to protein structures reported as ensembles of multiple conformations with structural factors describing conformational flexibility of their regions.[19]

Despite the ability to study protein dynamics and structure in solution there are still some limiting factors in protein NMR. Similar to protein crystallography, sample consumption and high protein concentrations during analyses can be problematic. Additionally, the acquisition of data necessary for generating three-dimensional structures takes considerable amounts of time, which may be an issue with less stable samples. Moreover, routine NMR experiments have usually been limited to proteins smaller than 30-40 kDa due to broadening of spectral peaks and overlapping of signals.[23] This has partially been overcome with recent developments of specialized experimental setups such as methyl-TROSY (transverse relaxation optimized spectroscopy)[24], but this techniques is still far from routine. Also, there is the problem of preparing multiply isotopically labelled samples as the natural abundance of NMR-compatible isotopes of biogenic elements (except for ^{1}H) is very low. The isotope enrichment is usually achieved with protein recombinant expression using only ^{13}C-glucose and ^{15}N-ammonium chloride as sources of

carbon and nitrogen, respectively. Moreover, large proteins often require perdeuteration to limit the signal complexity. Therefore, production of such isotopically modified proteins can easily become financially prohibitive.

1.2.3. Electron microscopy (EM) techniques

The last of common techniques, which aspire to provide atomic resolution of molecular structures, are methods utilizing electron beams. These can basically be differentiated into electron diffraction and electron microscopy approaches. Electron diffraction is very similar to X-ray diffraction, only the sample damage caused by electrons is much higher than that caused by photons. Therefore, the technique is mainly limited to two-dimensional crystals and the provided resolution is lower.[19] Microscopic approaches utilizing electrons are nowadays much more promising. In transmission electron microscopy samples can be fixed onto carbon grids and stained by thin layers of heavy atoms such as gold or platinum, which have many electrons and thus interact more strongly with electron beams and increase resolution. Alternatively, unstained biological samples in thin solution films may be cryogenically vitrified.[25] Electron beams, which are passed through such films then interact with electrons in studied proteins, are scattered and carry the image information. In contrast to X-rays, electron beams can be manipulated using electromagnetic lenses and therefore the image is reconstructed and visualized directly, thus eliminating problems with phase determination.[19,26] As vitrified samples in cryo-EM are not arranged periodically in identical orientations such as proteins are in crystal, large amounts of individual particles need to be imaged. The individual 2D images are then computationally classified according to their rotation and following extensive computational processing 3D electron densities of molecules can be reconstructed by tomographic approaches.[26] In favorable conditions and when enough single particles are imaged, the resolution provided by cryo-EM can be near-atomic (below 4-5 Å)[27] even though most commonly the resolution obtained tends to be around 10 Å.[19] Such medium-resolution data are however usually complemented by X-ray structures of fragments of the studied structure. These can then be fitted into the cryo-EM electron density envelope and lead to so-called pseudo-atomic structural model.[19]

Apart from approaching single-atomic resolution, the ability of cryo-EM to study hydrated macromolecules in low concentration solutions is very beneficial. The image classification approach can even enable the distinguishing of several conformers

concurrently present in solution, even though this increases computational demands and requires large enough sets of single particle images.[26] On the other hand, as single-particle images are very weak, cryo-EM is most effective on large structures ideally with high levels of symmetry, such as viral capsids or large protein assemblies typically bigger than several hundred kDa.[19] Consequently, oligomeric states of protein with largely different sizes are not monitored with equal efficiency – large amounts of small monomer may not be visible at all in EM images, while a minor population of large oligomers may be the only resolvable cryo-EM structure. Finally, when directly competing with crystallography, the resolution obtainable with cryo-EM is significantly lower. Nevertheless, it is not limited by crystallization issues and is therefore very powerful emerging technique for structural biology.

1.2.4. Lower resolution approaches

In addition to high-resolution techniques, a range of other approaches to study protein structures and dynamics exists. These are sometimes slightly overlooked in favor of their aforementioned high-resolution cousins as they either do not provide direct molecular coordinates or generate just vague structural information with low resolution. Among these, small-angle scattering of X-ray radiation (SAXS) or spectroscopic methods such as Förster resonance energy transfer (FRET) are probably the most commonly known and used. These methods have benefits such as the fact that they can study heterogeneous and highly dynamic systems. SAXS allows the users to determine ratios of various conformers occurring in the system together and their shapes, albeit with resolution of only 50-100Å.[28,29] FRET on the other hand can help to study distances between molecules in solution and their dynamics in time-resolved manner.[30] However, this depends on the presence of suitable fluorophores in the molecule or on the possibility to introduce these without affecting protein structure and behavior.

Although these and other (usually spectral) low-resolution techniques are not powerful enough to describe protein structure on their own, they can nicely complement high-resolution techniques to allow better understanding of protein structure and functioning. Thus, by combining data from different structural techniques an approach called integrative or hybrid structural biology has emerged.

1.3. Mass spectrometric structural techniques

Among experimental approaches, which contribute significantly to integrative structural biology, are methods based on mass spectrometry (MS).[31] Originally a purely analytical chemistry method identifying atoms (and later molecules) based on their molecular weight[32], mass spectrometry has in the last three decades evolved significantly in the structural analysis of biologically-relevant systems.[33] Nowadays it can provide information on all levels of protein structure ranging from amino acid sequence to dynamics of huge non-covalent protein assemblies.[33] Although not able to provide atomic coordinates *per se*, structural MS can provide distance constraints, which may be used in computer modelling to derive structures. Furthermore, it can quickly generate data with minute sample amounts, analyze very heterogeneous samples and even study dynamics of macromolecular systems in solution. Also, in contrast to many other low resolution techniques, structural MS can often provide localized structural information.

1.3.1. Soft ionization for protein analyses

No overview of structural MS techniques would be complete without at least a brief mention of ionization techniques used. Since its discovery in 1913 by J.J. Thomson[32], mass spectrometry has been used for analyses of small molecules. Factor preventing the analyses of biomacromolecules was the ionization process necessary to transfer proteins from solution into vacuum inside mass spectrometer and to charge them so they could be manipulated in the gas phase. The early methods used for producing ions such as electric glow discharge or bombardment of samples by electrons proved too energetic for fragile molecules such as biopolymers. This limitation was overcome in 1980s when two “soft” ionization techniques were discovered simultaneously. First was matrix-assisted laser desorption ionization (MALDI), which employs energy transfer between UV-laser irradiated organic acid matrix molecules and sample to ionize it and aids in its transfer to the gas phase.[34–36] The other was electrospray ionization (ESI), which is nowadays the most commonly used technique to generate ions of biopolymers in structural studies. It disperses small charged solvent droplets with analyte into the gas phase through a Taylor cone formed at the tip of a highly charged emitter. The droplets gradually desolvate in the vacuum inside mass spectrometer until only charged analyte molecules remain.[37,38] Utilizing MALDI and ESI, it is currently possible to gently ionize a vast range of biopolymeric analytes including peptides, proteins, polysaccharides or nucleic acids.

1.3.2. Native mass spectrometry with ion mobility

Through properly tuned ESI process it is even possible to ionize protein assemblies without breaking non-covalent interactions between protein and its ligand or those holding together non-covalent protein complexes.[39,40] Despite some initial controversies, it has now become widely accepted due to mounting evidence in literature, that proteins may retain significant amounts of their secondary, tertiary and quaternary structure in the vacuum inside mass spectrometer, at least on the time scale of milliseconds.[41–44] Assuming, of course, that instrument parameters and ionization process are carefully optimized to be as gentle as possible. This is to prevent extensive "ion activation" by increased internal energy of analyte ions, which would lead to their denaturation and unfolding.[45] Therefore, under such experimental conditions the technique has been frequently referred to as "native" mass spectrometry.

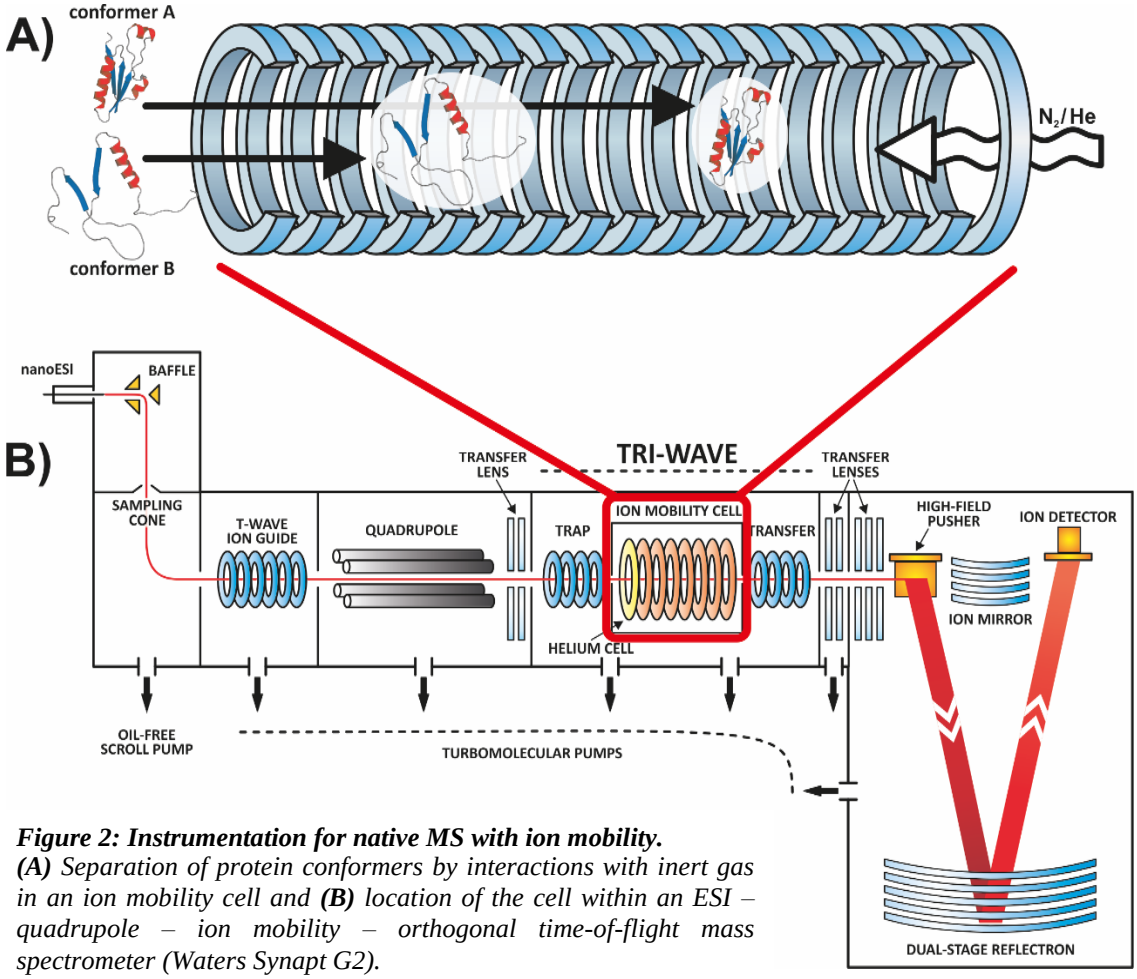


Figure 2: Instrumentation for native MS with ion mobility.
(A) *Separation of protein conformers by interactions with inert gas in an ion mobility cell and* ***(B)*** *location of the cell within an ESI – quadrupole – ion mobility – orthogonal time-of-flight mass spectrometer (Waters Synapt G2).*

Native MS often employs miniaturized ESI interface (so-called nanoelectrospray) to generate analyte ions.[46] Even though it has recently been shown that carefully tuned classical electrospray source can provide similarly gentle ionization process as nanospray[47], the latter is usually preferred. That is because it produces smaller droplets, which are easier to evaporate, and the sample consumption as well as the amount of salt adducts present is lower. Native MS is also commonly combined with ion mobility spectrometry (IMMS). (**Figure 2**, p. 17) In this technique ions travel ("drift") through a cell with low pressure of inert gas molecules (usually helium or nitrogen). There, analyte ions separate not according to their mass to charge ratio (m/z), but based on their overall shape. Therefore, ions of the same m/z but with extended conformation interact more with the gas molecules, are more slowed down and arrive later than compact ions.[42,45,48] The time the ions take to travel through the cell may then be converted into collisional cross sections, which reflect the size and shape of ions in a rotationally-averaged value. These values may be compared with candidate structure models or experimental high resolution structures.[49] Overall, the IMMS process is sometimes likened to electrophoretic or analytic ultracentrifugation separation, but occuring in the gas phase, on millisecond timescales and with significantly less sample.

By combining native MS and ion mobility it is possible to obtain information about conformations of proteins and their conformational flexibility. The techniques can as well report on proteins' oligomeric states, strength of their interactions with ligands or subunit composition of complexes. Through destabilization of complexes either by chaotropic agents in solution[50] or by fragmentation methods in the gas phase[51] it is even possible to deduce complex topologies and connectivity between individual complex subunits. Furthermore, probably the most interesting aspect of native MS analyses of whole proteins is the fact that analyte size as well as sample heterogeneity are not big limitations. Using the separation power of both IM and MS, it is possible to analyze various protein forms or ligand-bound states in parallel[52,53] as well as protein assemblies of very different sizes.[54,55]

1.3.3. Chemical cross-linking

Apart from analyses of intact proteins, mass spectrometry can provide structural information of more localized nature such as experimental spatial restraints. These limit distances between amino acid side chains in the molecule or between interacting protein partners, and can be used to derive structural models by computational modelling

approaches. To obtain the restraints, proteins are reacted with bifunctional chemical reagents, which covalently link functional groups (most commonly primary amines or carboxyl groups) in proteins.[56] Primary amines are targeted by N-hydroxysuccinimide esters[57] or imidoesters[58], whereas dihydrazides can be used to react carboxy groups pre-activated by carbodiimides.[59] (**Figure 3**) Alternatively, specific amino acids in protein sequence may be substituted with their structural analogs bearing photo-activatable labile groups. These upon UV irradiation form highly reactive carbene intermediates which attack any residue in their vicinity.[60,61] After the cross-linking reaction, studied proteins are enzymatically digested and individual cross-linked amino acids are identified in the resulting peptides by LC-MS analyses. Restraints determining the maximum distance between the identified residues in protein structure are then derived from the length of inflexible linker arm between the reactive groups of the used cross-linker.

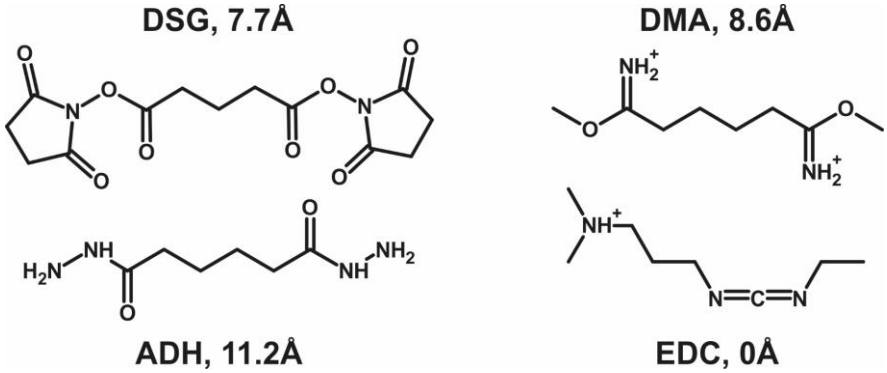


Figure 3: Some examples of chemical cross-linking reagents with their linker lengths. *Amino-reactive agents – N-hydroxysuccinimide ester DSG (disuccinimidyl glutarate) and imidoester DMA (dimethyl adipimidate). Carboxyl-reactive cross-linkers – dihydrazide ADH (adipic acid dihydrazide) and carbodiimide EDC (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide.*

Information obtained from chemical cross-linking can be used to study 3D structures of proteins[62] or to refine structures provided by other high-resolution techniques[20]. It can also define structural contacts between proteins in large assemblies[63] or even study interaction networks in living cells[64]. Additionally, through the use of isotopically labelled cross-linking reagents, quantitative information can be obtained describing the ratios of protein conformers in solution.[65]

1.3.4. Surface labeling approaches

The last category of MS-based structural biology techniques studies the solvent accessibility of protein surfaces by covalent labeling. This includes two main methods –

hydroxyl radical protein footprinting and hydrogen exchange. The latter will, as a main research tool employed in this thesis, be covered in a dedicated chapter.

Protein footprinting is a rather young technique, which probes surface solvent accessibility of proteins by exposing them to short pulses of hydroxyl radicals $OH^{\bullet}$. These are extremely reactive and unstable species, which quickly react with any amino acid in their vicinity (even though some amino acids react faster – namely those containing sulphur).[66] The reaction with label generates stable oxidation modification, which can then be detected and localized through the use of LC-MS/MS approaches.

The main premise of the method is, that as hydroxyl radicals have similar size as water molecules, they constitute a good probe to study what parts of protein are exposed to solvent. In order for the method to function, it is necessary to generate significant concentrations of hydroxyl radicals on very fast time scales to prevent unfolding of the studied molecule during labeling. This is achieved by short synchrotron X-ray pulses which directly fragment water molecules to radicals.[67] Alternative experimental setup is based on fast UV laser pulses which decompose minute amounts of hydrogen peroxide mixed with protein solution in a fluidic device just prior to labeling.[68] Both approaches claimed that with the use of capillary fluidics and addition of radical scavengers, the labeling reaction could be limited to low microsecond time scales.[67,68] This claim has however just recently been challenged by work of Prof. Konermann's group, which showed the reaction to generate secondary radicals that may last much longer in solution than expected (up to tens of milliseconds).[69] Nevertheless, despite this argument about the radical life times, the technique has successfully been used to probe protein organization in membrane as well as fast protein transitions during folding/unfolding experiments.[70–72]

1.4. Hydrogen/deuterium exchange

The other structural MS technique utilizing water-like probes to monitor which parts of the studied molecule are exposed to solvent, is hydrogen/deuterium exchange coupled to mass spectrometry (HXMS). Since its introduction in 1991 by Katta and Chait[73], HXMS has rapidly evolved into one of the most powerful and widely used structural MS techniques. Owing to its ability to work under virtually any solution conditions, it has successfully been applied to study a vast range of biological problems. These for example included adding solution dynamics information to static X-ray protein structures[21,74], studying mechanisms of intramembrane transporters[75,76] and membrane-associated

G-proteins[77], mapping binding epitopes of antibodies and monitoring conformations of biopharmaceuticals[78,79] or probing amyloid aggregation[80]. Moreover, as HXMS is able to work with heterogeneous samples and proteins of vastly differing sizes, it enabled studies of structure and dynamics of big protein complexes[81,82] or even viral capsid assemblies[83,84].

1.4.1. Basic principles of protein hydrogen exchange

HXMS is based upon principles first observed in 1950s by Kaj Linderstrøm-Lang and his students Aase Hvidt and Sigurd Nielsen.[85,86] They noticed that hydrogen atoms in a protein can spontaneously exchange with surrounding water molecules and also, that not all hydrogen atoms in protein behave equally.[85,86] (**Figure 4**) Hydrogens attached directly to carbon atoms are bonded too strongly and do not exchange at all. On the other hand, very labile protons in carboxyl groups, primary amines as well as in hydroxyl and sulfhydryl groups exchange too fast to be of any analytical use in standard experimental setups. Finally, owing to the partial double nature of peptide bond, protein backbone amide hydrogens exchange on intermediate time scales (seconds to days) compatible with biophysical analyses. The analyses exploits the fact that the exchange can be monitored when heavy water (D_2O) is used to prepare the buffer proteins are incubated in. Thanks to the different nuclear spin or increased atomic mass, exchange of hydrogen for deuterium can be detected by analytical methods such as NMR[87,88] or mass spectrometry[73,89], respectively. Using these techniques, it is possible to monitor the exchange kinetics of individual amide hydrogens or protein backbone regions.

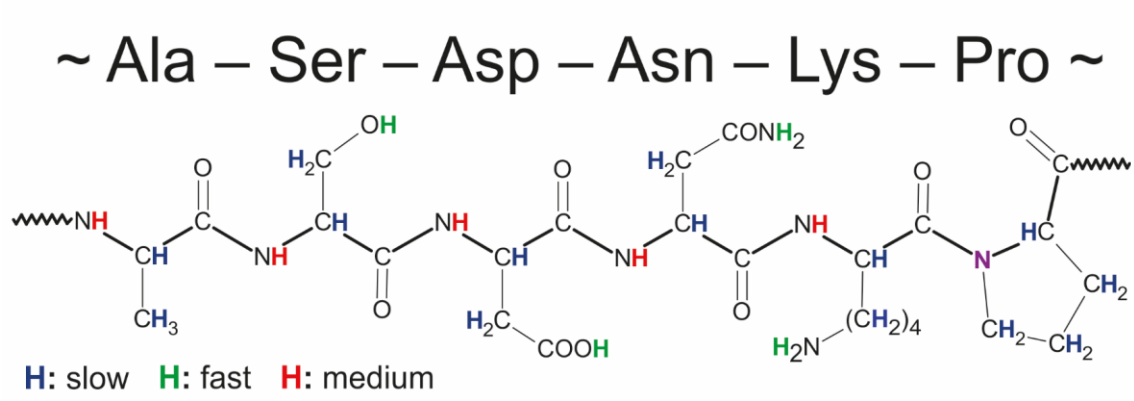


Figure 4: Types of hydrogen atoms in proteins. *Hydrogens are divided based on the rate of their exchange for hydrogen in solvent. Very slow are C-H bonded atoms (blue), fast are atoms at carboxyl groups, hydroxyls and sulfhydryls (green). Peptide backbone hydrogens exchange with intermediate speeds (red) and can be monitored analytically. Proline has no peptide amide hydrogen (purple).*

It is also known, that amide hydrogens in unstructured protein regions exchange much more rapidly than those buried inside the structure and those involved in intramolecular hydrogen bonds.[86,90] This is because regions with bonded amides first need to get into a transient unfolded state to exchange. Such transitions occur on local level in protein naturally through protein breathing and fast localized structural fluctuations.[86,90] After the temporary breaking of hydrogen bonds, exchange can occur and then regions refold again. Whole mechanism can formally be described by the following equation[86,91]:

$$\text{N-H}_{\text{closed}} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \text{N-H}_{\text{open}} \xrightarrow[\text{D}_2\text{O}]{k_{HD}} \text{N-D}_{\text{open}} \underset{k_{op}}{\overset{k_{cl}}{\rightleftharpoons}} \text{N-D}_{\text{closed}}$$

where N-H and N-D represent a hydrogenated and deuterated backbone amide, respectively. Kinetic constants k_{op} and k_{cl} define the rate of local structural unfolding and refolding, respectively, and k_{HD} is a constant reflecting the speed of deuterium exchange itself. Finally, the deuteration reaction is showed as irreversible in the above equation. This is not caused by the chemistry behind the process as, of course, the amide deuteration is a fully-reversible reaction. Rather it reflects the fact that during the labeling protein is placed in a buffer composed of vast excess of D_2O. Therefore, even though some protons are introduced from the protein itself, the overwhelming amount of buffer in the system makes the probability of deuterium back exchange during the labeling negligible.

Depending on the ratio between kinetic constants k_{HD} and k_{cl}, two situations can occur during the labeling. (**Figure 5**) Most commonly, transient unfolding fluctuations in proteins are only localized and very short-lived ($k_{cl} >> k_{HD}$). In such cases overall deuteration kinetics is governed by k_{HD}, which results in gradually incrementing deuterium content during the course of deuteration – such deuteration regime is called EX2.[86,90,92] In relatively rare cases, where slow, large scale concerted rearrangements of protein occur, some regions in protein can remain longer in temporarily unfolded state ($k_{cl} << k_{HD}$).

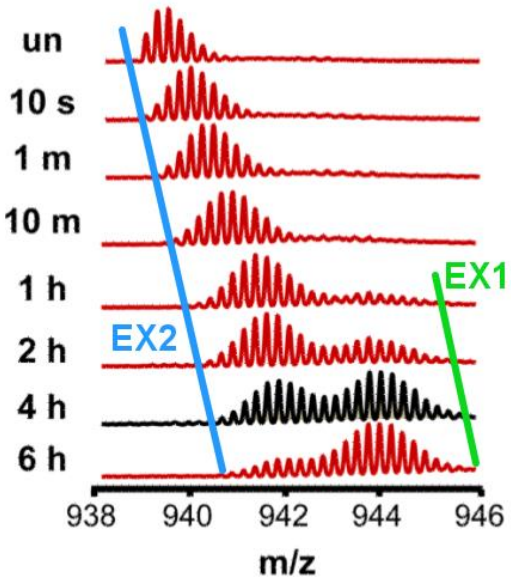


Figure 5: HXMS exchange kinetics. *Real data[81] showing mixed exchange with clearly distinguishable EX2 and EX1 components as discussed in text. "un" – undeuterated control.*

Deuteration rate is then defined by the rate of structural rearrangement k_{cl}, which results in deuteration regime EX1 and typical two-state data profiles.[86,90,92] Even though the observation of pure EX1 kinetics is rather uncommon, mixed EX1/EX2 behavior can sometimes be detected in raw NMR or HXMS data and in such cases it can provide valuable additional information about the conformational dynamics of protein and/or about the presence of populations of different conformers in the system.[92]

1.4.2. Factors influencing hydrogen/deuterium exchange

The task of hydrogen exchange analyses of proteins is to monitor changes in k_{HD} for an analyte under various conditions. This rate constant is primarily determined by the accessibility of amide hydrogens to solvent molecules and by their involvement in hydrogen bonding. Therefore, analyses can provide information about secondary structures in proteins, their conformational changes, interactions with various ligands and interaction partners and about overall dynamics of protein in solution.[79] Usually, if protein interacts with a partner or ligand, this leads to deuteration decrease in protein regions, where the interaction takes place. This is commonly explained by solvent exclusion from the interface and by formation of new stabilizing hydrogen bonds between interacting partners. Despite this being the most common outcome of HX analyses, it is not the only possibility as demonstrated by recent works of Prof. Konermann's group.[91,93] In the case when unfavorable structural changes in protein are necessary for ligand binding or if the interaction with ligand does not perturb free energy levels of protein molecule, much more uncommon scenarios may occur leading to backbone deprotection or no changes observable by hydrogen exchange.[91,93] Sometimes, the scenarios may even occur together in distinct parts of the protein – such as in the case of allosteric effects in hemoglobin.[93]

Apart from solvent exposure and hydrogen bonding, which are directly probed by HX analyses, there are other factors, which influence the deuteration behavior of proteins and k_{HD}. These are mainly pH, temperature and the effects of neighboring amino acids.[94] To a smaller extent the rates also depend on isotopic effects caused by deuterium, but these are commonly considered to be negligible.[95] Neighboring amino acids influence backbone amide exchange by steric side chain shielding and by inductive effects caused by polar residues.[94] However, primary structure of a protein does not change during the analyses. It basically may come into consideration only in the case of studying mutated proteins or when predicting the deuteration ratios of peptides computationally.[94] On the other hand,

pH and temperature are key parameters, which are set and controlled during analyses. Alternatively, if the effects of these factors on proteins are to be studied, it is necessary to appropriately compensate for their intrinsic effect on H/D exchange.[96]

The effect of temperature can be determined by the classical Arrhenius equation:

$$\boldsymbol{k(T) = k(293) \times e^{\left[\frac{-E_a}{R}\left(\frac{1}{T}-\frac{1}{293}\right)\right]}}$$

where k is deuteration rate constant, T is the experimental temperature, R is universal gas constant and E_a denotes activation energy, which depends on the mode of catalysis according to reaction pH. For base catalysis prevailing at the most commonly used physiological pH: $E_a(k_B)$ = 17 kcal/mol.[94] Overall, H/D exchange kinetics is approximately 10× slower at 0°C than at 22°C.[97]

The influence of pH on k_{HD} is even stronger and depends on the catalysis mode of H/D exchange reaction at the experimental pH. Under physiological pH, the reaction is mainly mediated by OH^-, whereas at extremely acidic conditions H^+ catalysis is predominant. Also, pH-independent catalysis by undissociated water molecules takes place as well, but is very slow and only significant at the pH of minimal exchange.[86,94,97] The rate can be determined by the following equation[94,97,98]:

$$\boldsymbol{k_{HD} = k_{int,A}[H^+] + k_{int,B}[OH^-] + }$$
$$\boldsymbol{+k_{int,w}[H_2O]}$$

where k_{int} is intrinsic rate constant for acid (A), base (B) and water (w) catalysis.[94,97] Research has shown that $k_{int,B}$ is much higher than both $k_{int,A}$ and $k_{int,w}$.[94] Even though pH where k_{HD} is minimal depends on the sequence of peptide, for an average random unfolded peptide it is considered to be around pH 2.5 (**Figure 6**). Increase of one pH unit then results in 10× faster H/D exchange rate.[94,97]

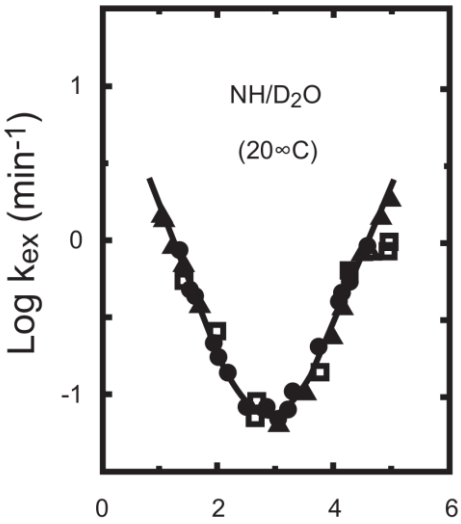


Figure 6: Influence of pH on the H/D exchange kinetics ($k_{ex} = k_{HD}$). *Conditions of minimal exchange are around pH 2.5 for an average unstructured polypeptide.*[94,97]

1.4.3. Experimental approaches to HXMS analyses

Historically, many methods were used for deuterium incorporation determination, including H_2O/D_2O density measurements, infrared spectroscopy or radiography when tritium was used in place of deuterium.[85,86] However, nowadays protein NMR and most

importantly MS are exclusively used. NMR has the advantage of directly determining kinetics for each backbone amide, but suffers from the usual protein size, sample concentration and data interpretation issues as discussed in chapter 1.2.2. Coupling MS to hydrogen/deuterium exchange (HXMS) on the other hand offers fast and high-throughput analyses with small sample consumption, easy data interpretation and virtually no protein size limitations.[73,89] On the other hand, HXMS has lower structural resolution than NMR, providing information on a short peptide instead of single residue level. Nevertheless, this detail is usually sufficient for structural conclusions. Alternatively, gas-phase fragmentation MS techniques can be employed to enhance spatial resolution to near single residue level.[99,100]

Although some experiments to study protein stability and folding utilize pulsed deuterium labeling[80,101–103], most often used is continuous labeling approach. (**Figure 7**)

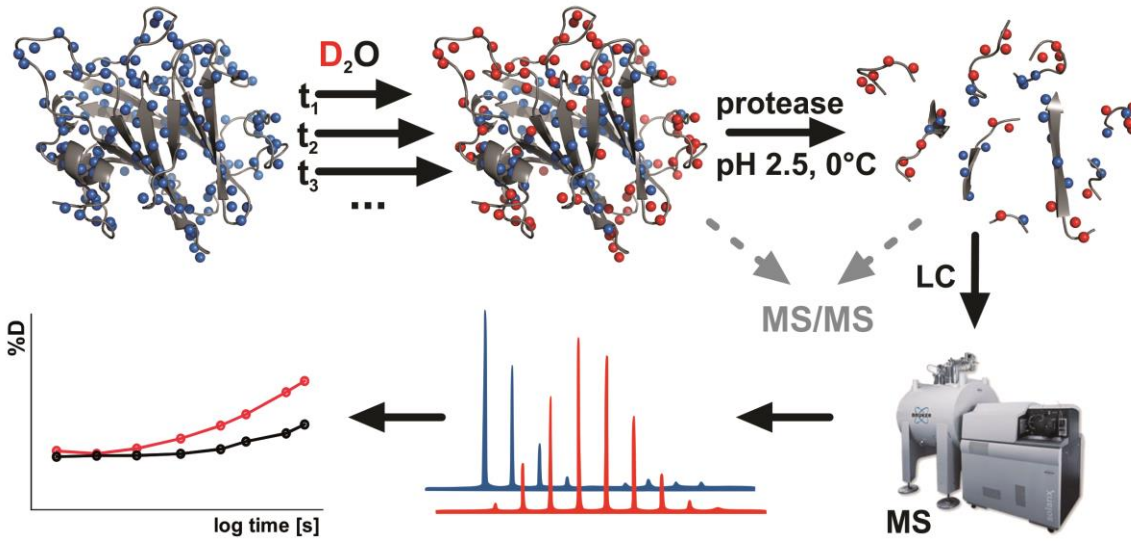


Figure 7: Experimental setup of a bottom-up H/D exchange analysis by mass spectrometry. *Protein amide backbone hydrogens (blue spheres) exchange for deuterium from solution (red spheres) depending on their solvent accessibility and hydrogen bonding. Following labeling, the deuteration is "quenched" by rapid pH and temperature decrease and protein is digested to peptides. These are separated by LC, analyzed by MS and the observed mass increase is converted to deuterium uptake curves. Steps where gas-phase fragmentation may be employed for top-down or middle-down HXMS are shown in grey.*

In this setup protein deuteration is initiated by its dilution into an excess of D_2O-based buffer. The exchange is then left to proceed and at predetermined time points, aliquots are removed from the labeling reaction. From this moment, speed of the sample analysis is of essence. This is because deuterium label is not stable and as further analytical step are performed in normal H_2O-based solvents, the deuterium introduced into protein would exchange back and the structural information would be lost. Therefore, the effects of pH

and temperature on the H/D exchange rate (discussed in the previous chapter) are used as tools to prolong the time window for analysis. This is achieved by a process called "quenching", during which the pH in sample is quickly dropped to around 2.5, where the H/D exchange is minimal.[90,94] Additionally, the sample is cooled to 0°C to further slow down the deuterium back-exchange. By doing this the half-life of deuterium label increases dramatically (approximately by a factor of 10^6 compared with the usual labeling conditions of pH 7.4 and 22°C). The low back-exchange conditions (pH 2.5, 0°C) are kept during the rest of sample analysis. Moreover, to compensate for the amount of back-exchange encountered during analyses, two types of experimental controls can be performed. First is done by analyzing a non-deuterated sample and is used to provide initial mass measurements before the deuterium incorporation analysis. Second, which compensates for the loss of deuterium during sample handling and LC-MS, analyzes the studied protein in which all backbone amide hydrogens have been exchanged for deuterium.[89] From this the maximal deuteration level practically achievable by HXMS can be determined for each peptide. This is often crucial when directly comparing the data against theoretically calculated exchange values, especially when studying, which parts of a molecule are involved in secondary structures.

Following the quenching, further sample analysis usually involves fast protein digestion into peptides – often employing a small flow-through column with immobilized protease to increase the effectivity of proteolysis and limit the time necessary for digestion.[104] Following the digestion, sample is online desalted and separated by a liquid chromatography (LC) coupled to an ESI source of a mass spectrometer. Mass increase caused by deuteration is then monitored for individual peptides by MS analysis. Measured deuterium content in the peptides is in the end plotted against labeling time to construct deuterium uptake plots, based on which the behavior of proteins under different conditions (e.g. active / inactive, alone / with ligand) is compared.

As dictated by the low pH quench conditions necessary to minimize deuterium back-exchange, sample proteolysis cannot be carried out by specific proteases such as trypsin, which are commonly used in MS-based proteomic experiments. Therefore, aspartic proteases which have their pH optima in strongly acidic pH are employed instead. The most commonly used protease is porcine pepsin. However, pepsin and pepsin-like proteases do not have well defined cleavage specificities. On one hand this is beneficial, as the protease cleaves the studied protein in many places and shorter resulting peptides

enable monitoring of H/D exchange with higher spatial resolution. On the other hand this means that it is impossible to predict the cleavage sites in the analyzed protein. It is therefore necessary to first subject its digest by the utilized protease to an LC-MS/MS analysis. This is done in order to identify what peptides the protease produces and what signals to search for in LC-MS analyses of deuterium uptake.

1.4.4. Increasing spatial resolution of HXMS

From the overall scheme of HXMS experimental setup it is obvious that the length of peptides produced by proteolysis is the key factor, which determines the spatial resolution of HXMS. The shorter the analyzed peptides, the bigger the precision in localizing structural changes in a protein detected through H/D exchange behavior. Some analyte proteins, however, do not provide sufficiently short peptides in digestion or cannot be completely digested by porcine pepsin at all. In such cases there are two approaches to increasing the spatial resolution of HXMS.

First, gas-phase fragmentation MS/MS techniques may be used to (a) fragment the intact deuterated protein or (b) to fragment the peptides produced by proteolysis. Direct top-down HDX-MS/MS has been successful for some small proteins[105,106], even though for large disulfide bond stabilized proteins its usefulness is only limited[107]. So-called middle-down approach, where only protease-produced peptides get fragmented in the gas phase has shown great promise, providing almost single-residue monitoring of deuterium uptake.[100,108–110] The most important analytical consideration of gas phase fragmentation HXMS approaches is the use of electron-based fragmentation techniques such as ETD (electron transfer dissociation) or ECD (electron capture dissociation) to prevent deuterium atoms from scrambling (changing positions in a peptide) during the fragmentation.[111] However, not all mass spectrometers support these fragmentation techniques. Moreover, both ETD and ECD require at least triply charged peptides for optimal fragmentation and some peptides also do not fragment well. Therefore, despite their undisputed powers, it is necessary to not only depend on fragmentation techniques and to actively seek in parallel other means of increasing HXMS spatial resolution.

Second, aspartic proteases other than the gold standard porcine pepsin can be used to optimize the digestion and provide shorter peptides for analysis. Short overlapping peptides are beneficial for HXMS as computational approaches can then be used to localize deuteration changes to very short regions, even approaching single-residue

resolution.[112,113] Although all pepsin-like aspartic proteases have somewhat similar broad cleavage preferences, the exact cleavage sites in a specific protein often vary between individual proteases. Therefore, several alternative proteases have so far been tried for HXMS-compatible digestion in the place of the most commonly used porcine pepsin. These included aspergillopepsin (protease XIII from *Aspergillus saitoi*) and rhizopuspepsin (protease XVIII) from *Rhizopus chinensis*.[114,115] Both were shown to be complementary to pepsin as they showed higher preference for cleavage after basic amino acid residues.[116,117] Other, more exotic sources of acidic proteases were investigated as well – including plasmepsin[118], pepsins from Antarctic rock cod[119] and pepsin from rice field eel[120]. However, these did not find much use in HXMS, probably due to their inconvenient sourcing.

Most recent addition to the HXMS protease portfolio is represented by acidic digestive fluid from carnivorous plants of genus *Nepenthes*[121] (**Figure 8**) These fluids exhibit interesting behavior, namely their high activity and unique cleavage preferences. In addition to typical pepsin-like nonspecific cleavage, proteolytic activity of pitcher fluids also exhibited strong preference to cleave after basic amino acids and uniquely after proline.[121] Since the discovery of proteolytic activity of *Nepenthes* pitcher fluids back in 1874, much controversy remained to the origin of this activity – whether it originated form microbial contamination or from proteases produced by the plants themselves.[122,123] It has since been discovered that pitcher plants indeed produce into their traps at least two proteases nepenthesins-1 and 2.[124,125] These were identified to be aspartic proteases, but interestingly showed unique tolerance to high pH values and in the case of nepenthesin-1 low stability in denaturing agents, completely unlike porcine pepsin.[126,127] Due to the unique cleavage preferences of pitcher fluids and interesting enzymatic properties of nepenthesins, we set out to study these in more detail with the aim to produce them recombinantly, characterize them and ultimately to test them for the use in HXMS digestion setup.

Figure 8: Nepenthes sp. pitcher plant. *One of plants in a greenhouse of Prague Botany Garden, where pitcher fluids were collected.*

1.5. Cellobiose dehydrogenase enzymatic system

One protein system, which is due to its heterogeneity, high flexibility and conformational dynamics challenging to study using structural biology approaches is cellobiose dehydrogenase (CDH). In this thesis, CDH is used to demonstrate how integrative structural MS can be used to complement X-ray crystallography and study this enzyme's dynamics in solution.

CDH - cellobiose:(acceptor) 1-oxidoreductase, E.C. 1.1.99.18 - is the only known extracellular flavocytochrome. It is produced by a number of wood-degrading and phytopathogenic fungi from the *Basidiomycota* and *Ascomycota* phyla, which utilize CDH for lignocellulose degradation.[128] The enzyme is also highly relevant in the field of biotechnology as it forms a component of promising third generation biosensors and biofuel cells based on direct electron transfer.[129,130] Additionally, CDH is a crucial enzyme in the process of cellulose depolymerization to shorter saccharides that are foreseen as a potential source of renewable energy in place of fossil-based fuels.[131]

From a structural point of view, CDH is a monomeric enzyme consisting of two domains joined by a flexible linker region. The larger dehydrogenase (DH) domain carries FAD as a cofactor while the smaller cytochrome (CYT) domain contains a heme *b*. During its catalytic cycle (**Figure 9**) cellobiose dehydrogenase oxidizes β-D-cellobiose to δ-cellobionolactone.[132] Electrons obtained at FAD by the reaction are then due to redox

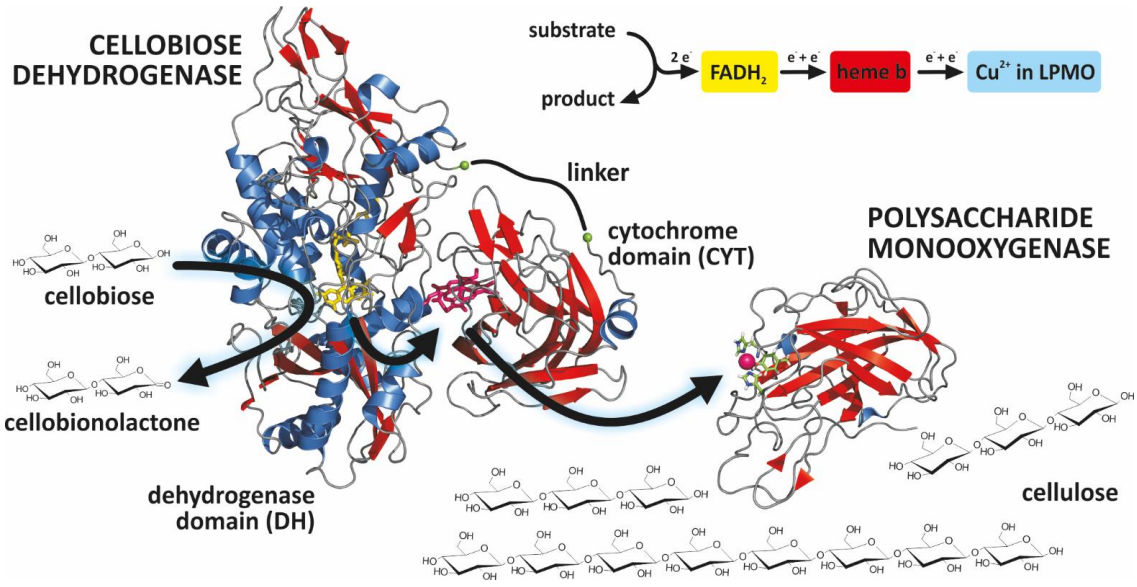


Figure 9: CDH / LPMO cellulolytic system. *Electrons obtained by the oxidation of cellobiose are transferred between FAD (yellow) and heme (magenta) cofactors in CDH via a direct electron transfer mechanism. They subsequently end up on terminal acceptor - polysaccharide monooxygenase, which fragments cellulose chains. For the electron transfer a temporary close contact between the cofactors is enabled by cytochrome domain movements.*

potential difference delivered to heme *b* in the CYT domain. From the heme cofactor, the electrons are finally channeled to a terminal electron acceptor, which can be either a small molecule or a protein partner such as lytic polysaccharide monooxygenase (LPMO).[133] For such electron flow to proceed, the key reaction is the interdomain electron transfer (IET) between FAD and heme within the CDH molecule. This transfer occurs in CDHs spontaneously only under acidic solution conditions. Dependence of the IET on the pH has been observed since early studies.[134] While both FAD and heme are reduced rapidly at slightly acidic pH in the presence of cellobiose, at neutral pH only FAD reduction is fast, while the IET and heme reduction is extremely slow.[134,135] Moreover, it has also recently been found that the presence of divalent cations in mM concentrations can induce enzyme activity even at pH 7.5, where it is otherwise completely inactive.[136,137] Based on these observations a theory of CDH functioning was proposed, in which the close contact of the two domains is necessary for the electrons to be transferred directly between the cofactors. Physical separation of CDH domains on the other hand prevents the IET and stops the enzyme's functioning.[138]

Detailed knowledge of the structure and dynamics of CDH is a prerequisite to improve its biotechnological potential and to tailor its enzymatic properties. Most likely due to its dynamic nature and high degree of post-translational modification, the only known high resolution structures for CDH were until recently derived from isolated dehydrogenase[139] and cytochrome[140] domains from *Phanerochaete chrysosporium*. Until recently, these structures thus served as a modelling template for the full length protein. However, in 2015 a breakthrough paper describing high-resolution X-ray structures of two conformations of CDH was published after more than a decade of attempts to crystallize the enzyme. The study characterized two different full-length CDH molecules (proteins originating from *Myriococcum thermophilum* - *Mt*CDH and *Neurospora crassa* - *Nc*CDH).[141] Crystals were obtained from pH~6.5, which is for *Nc*CDH close to its pH optimum while for *Mt*CDH this is an intermediate value between its most active and inactive form.[137] Interestingly, the enzymes crystallized in different conformations, each representing one conformational state predicted to occur during pH mediated inter-domain cross-talk in CDH. The static pictures of X-ray structures were further complemented by small angle X-ray scattering data, which allowed estimating the representation of different conformers occurring in the system. This analysis showed that both the closed and open forms of CDH (resembling the crystal forms of *Mt*CDH and *Nc*CDH, respectively) are

simultaneously present among other conformers.[141] Finally, molecular modelling and domain docking approaches suggested that the molecular mechanism preventing the domain interaction and IET at neutral pH is the repulsion of negative charge patches at the domain-domain interface.[137]

In order to structurally validate the theory of domain charge repulsion and to better understand the mechanisms involved in controlling the IET, we focused our structural MS studies presented in this thesis on determining what effects the pH and divalent cations have on the conformation and conformational dynamics of CDH in solution. For this, we selected *Mt*CDH, which was successfully crystallized in its closed (IET-capable) form[141], as a suitable model system.

2. AIMS OF THE THESIS

The overall aims of this Ph.D. dissertation were to (1) contribute to ongoing development of novel tools for hydrogen / deuterium exchange mass spectrometry and (2) to apply structural MS techniques to study conformational dynamics of cellobiose dehydrogenase in solution.

The specific goals then were:

- To characterize the enzymatic behavior of recombinantly expressed aspartic protease nepenthesin-1 in solution.

- To immobilize nepenthesin-1 onto perfusion resin and test its usability in an HXMS-compatible online digestion setup.

- To study the conformational dynamics of cellobiose dehydrogenase in solution using structural mass spectrometry techniques with main focus on elucidating the mechanism of its ion- and pH-mediated regulation.

- To contribute to the development of novel computational tools for processing and evaluation of HXMS data.

3. METHODS

The research papers included in this Ph.D. thesis provide a detailed description of all methods and experimental procedures used together with details necessary for the reproduction of the presented results. Therefore, this chapter only lists experimental techniques used throughout the thesis.

List of used research methods:

- Protein sample preparation for MS analyses
- Non-denaturing enzymatic deglycosylation
- UV/VIS spectrophotometric determination of aspartic protease activity
- Enzyme immobilization
- MS-based analysis of protein primary structure and post-translational modifications (MALDI-TOF, ESI-FTICR directly or in LC-MS and LC-MS/MS setup)
- Hydrogen / deuterium exchange coupled to mass spectrometry
- Native mass spectrometry with ion mobility (ESI-qTOF)
- Protein homology modeling
- Protein surface electrostatics computational simulations

4. RESULTS AND DISCUSSION

The aims of this thesis involved both HXMS method development as well as the application of this structural MS technique in an integrative structural biology approach to address questions involving protein conformational dynamics in solution. Therefore the first part of the results section will discuss the development of a novel aspartic protease nepenthesin-1 as a potential addition to the portfolio of HXMS proteolytic enzymes. In the second part, nepenthesin-1 will be used during the analyses of cellobiose dehydrogenase cellulolytic enzyme, which were performed in order to explain the mechanisms underlying its regulation in solution.

4.1. Developing aspartic protease nepenthesin-1 as a tool for HXMS

One of the factors limiting the spatial resolution of HXMS structural studies of proteins is the efficiency of protein digestion (see section 1.4.4). If short peptides with many overlaps can be generated, it is ultimately possible to achieve almost single residue resolution in localizing the deuterium uptake and structural changes in the studied protein molecule.[112,113] For this reason, proteases complementary to the most commonly used porcine pepsin may be utilized.[114] Recent work described the use of pitcher fluid from carnivorous plants of genus *Nepenthes* as a viable approach to achieve better spatial resolution of HXMS.[121] In spite of this, pitcher fluids are not very convenient for routine uses and also their enzymatic composition cannot be guaranteed to be stable. Moreover, the low amounts of proteases present in the pitcher fluid together with their glycosylation prevented their efficient immobilization for online HXMS-compatible digestion using a protease column. (M. Rey – personal communication)

Therefore, the interesting properties of pitcher fluids prompted us to recombinantly prepare and study nepenthesin-1 – the major protease present in *Nepenthes* pitcher fluid.

4.1.1. Recombinant expression and characterization of nepenthesin-1

*Results in this section were included in the attached **paper I** – Kadek, A. et al. Expression and characterization of plant aspartic protease nepenthesin-1 from Nepenthes gracilis. Protein Expr. Purif. 95, 121–128 (2014).*

Recombinant expression protocol for nepenthesin-1 (Nep-1) from *Nepenthes gracilis* was developed in our laboratory by Dr. Hynek Mrázek and consisted of gene expression in *E. coli* cells followed by *in vitro* renaturation of Nep-1 from inclusion bodies according to

a modified protocol of Flentke.[142] It consisted of gradual lowering of pH and denaturant concentration by dialysis. After the final acidification to pH 2.5 the solution contained nearly pure Nep-1 as contaminants either precipitated upon acidification or were digested by the active protease. Yield of active rNep-1 was usually within 30-60 mg per liter of the production culture, which was much more than the reported 1.8 mg of naturally occurring Nep-1 isolated from approximately 30 litres of pitcher fluid.[143] Protease used in the following work for enzymatic characterization and immobilization was produced by undergraduate students Vyacheslav Tretyachenko and Jan Kukla.

Following the protein production, sequential identity of Nep-1 was confirmed by enzymatic and CNBr digestion followed by LC-MS and LC-MS/MS analyses. SDS-PAGE and MALDI-TOF analyses were used to monitor its autoactivation upon solution acidification as well as to confirm its ability to digest other proteins. The mechanism of autoactivation was shown by MALDI MS to involve the cleavage of Nep-1 propeptide, which proceeded sequentially from protein N-terminus. First, a slightly longer form (75-437) is formed quickly, which is further processed to the theoretically expected mature polypeptide (79-437) in the range of hours. The results were in agreement with the previously published partial N-terminal sequence of nepenthesin-1 obtained by Edman sequencing of protein purified from pitchers of *N. distillatoria.*[143]

Upon confirmation of Nep-1 identity and functionality, enzymatic properties of the mature protease were then studied in solution by spectrophotometric activity assays. The behavior of rNep-1 was compared with porcine pepsin and with results published for nepenthesin-1 isolated from pitcher fluid.[127,143] For this a modified version of the activity assay developed in 1938 by Anson[144] was used. It was based on incubating bovine hemoglobin with the tested protease at pH 2.5 and 37°C, followed by precipitation of undigested protein by trichloroacetic acid. Enzymatic activity was then derived from UV spectrophotometric absorbance measurements at 280 nm, which measured the amount of TCA-soluble peptides released into solution by proteolysis.

Recombinantly produced nepenthesin-1 showed characteristics very similar to its naturally occurring form isolated from plants. The only difference was lower stability of recombinant protein at high temperatures. This was probably caused by missing glycosylation, which is not present in protein expressed in bacteria, but seems to stabilize naturally occurring enzyme in plants.[143] However, this did not compromise the proteolytic activity of the enzyme. Comparison of enzymatic properties between rNep-1 and porcine

pepsin provided more interesting results. Optimal pH for the activity of both proteases was found in strongly acidic solutions around pH 2.0, whereas at pH > 4 both enzymes were virtually inactive. Interestingly, rNep-1 was found to be very stable in basic pH during long term stability assays. Even after 30 days at pH 8.0 it retained close to 75% of its activity. Such behavior is in striking contrast with porcine pepsin, which becomes rapidly irreversibly denatured even after short exposure to pH above 6.

Finally, recombinant Nep-1 demonstrated its typical feature[127] – high sensitivity to denaturing agents in solution. (**Figure 10**) In the presence of guanidine, Nep-1 rapidly lost its activity. This became even more pronounced when reducing agents such as tris-(2-carboxyethyl)phosphine (TCEP) were added into solution. This pointed at the presence of disulfide bonds in the Nep-1 molecule and proved their necessity for the stabilization and activity of the protease. Such behavior was in contrast to porcine pepsin, which could withstand even high concentrations of both the denaturing agents and also the addition of TCEP did not seem to have any significant effect on its stability.

This observation might be discussed in the context of numbers and locations of disulfide bonds involved in the stabilization of each of the two proteases. Nepenthesins are

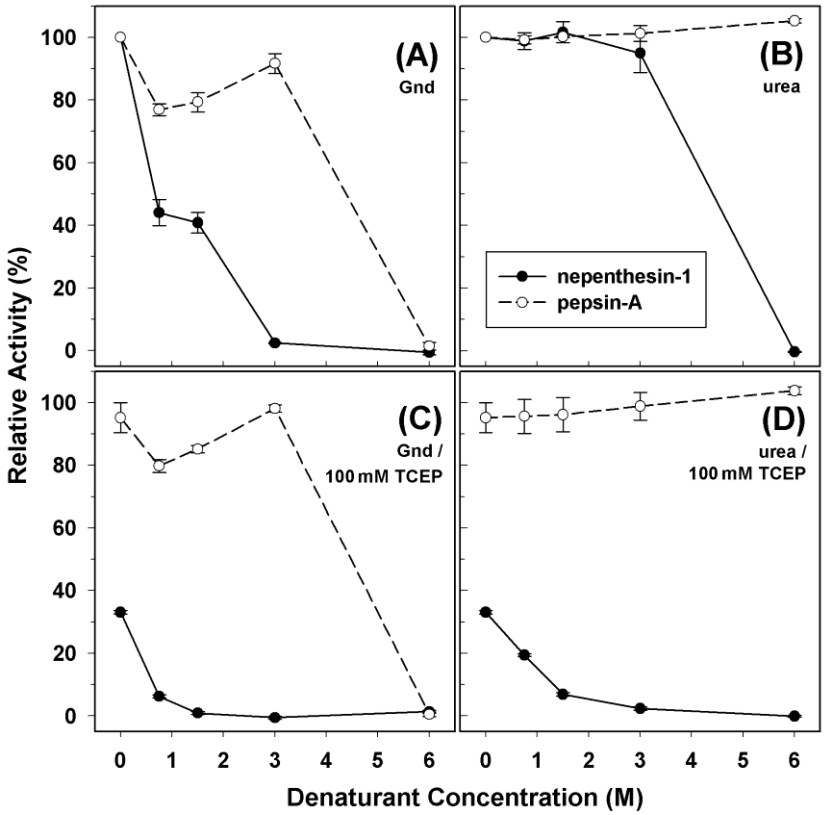


Figure 10: High sensitivity of rNep-1 to denaturing and reducing agents in solution. *Proteolytic activity towards hemoglobin measured at 37 °C and pH 2.4 in the presence of guanidine (A) and urea (B) alone or in combination with 100 mM TCEP (C, D). Values are means of independent replicates with standard deviations. For comparison, porcine pepsin data are plotted as well, while relative activities (normalized individually for each of the enzymes) are shown.*

supposed to contain more disulfide bonds in the N-terminal lobe of the protein (five disulfides versus one for nepenthesin and porcine pepsin respectively).[127] It has also been shown that the thermal denaturation of pepsin starts by melting of this N-terminal lobe, followed by structural rearrangements of the rest of the molecule.[145] Provided that nepenthesin unfolds along a similar pathway starting with the N-terminal domain, its higher dependence on the disulfide stabilization can explain the more severe effects of reduction on its function.[127]

4.1.2. Immobilization of recombinant nepenthesin-1 for HXMS online digestion

*Results in this section were included in the attached **paper II** – Kadek, A. et al. Aspartic protease nepenthesin-1 as a tool for digestion in hydrogen/deuterium exchange mass spectrometry. Anal. Chem. 86, 4287–94 (2014).*

Large amounts of active Nep-1 protease available through recombinant expression together with the interesting reported cleavage properties of *Nepenthes* pitcher fluids[121] prompted us to test the suitability of recombinant Nep-1 for HXMS workflow. During bottom-up HXMS analyses it is beneficial to use protease immobilized onto resin and filled into small protease columns.[104] This "online" digestion has many advantages over the digestion by protease added into solution. First and foremost, it increases digestion efficiency by enhancing the local protease/substrate ratio. As the protease is immobilized, it is less liable to autolysis and does not produce autolytic peptides, which would otherwise elevate analytical background in LC-MS.[104] Also, online digestion is better suited for robotics-based automated HXMS analyses by eliminating additional pipetting and mixing steps. Finally, as the immobilization of an enzyme to resin usually increases its stability, the prepared immobilized protease column can be used repeatedly without deterioration of its enzymatic activity.

Therefore, we followed a protocol previously used for porcine pepsin to immobilize recombinantly prepared Nep-1 onto perfusion resin.[104] We used commercially available POROS-20AL resin, which has 20 μm porous particles functionalized on their surface by aldehyde groups. Onto these groups activated Nep-1 molecules were attached via one of their two primary amines. The resulting Schiff base was then softly reduced by cyanoborohydride to form a stable –CH–NH– linkage. In the protease coupling step previously characterized pH stability of Nep-1 was utilized as it allowed performing the coupling reaction in pH 7.0. At this pH the reaction proceeds much more effectively than

at pH 5.0, which must be used for porcine pepsin. Resin with the attached protease was then filled into a 2×20 mm column, which could then be coupled to an online HXMS experimental setup.

We tested the activity of the Nep-1 column and showed that it digests proteins efficiently under HXMS-compatible conditions (pH 2.5, 0°C). Nevertheless, as some protein structures are very compact and not accessible to protease digestion without denaturation and/or reduction, digestion under such harsh conditions is essential to render these proteins amenable to H/D exchange studies. However, Nep-1 has been shown to not function in solution under denaturing conditions.[127,146] Therefore, we needed to test whether the immobilization influenced this behavior of nepenthesin.

To study in detail the cleavage behavior of Nep-1 under HXMS conditions, we employed a test based on LC-MS/MS mass spectrometric monitoring of horse myoglobin digestion by pepsin or nepenthesin in a pH 2.5 buffer alone or in the presence of 2 M guanidine with 0.4 M TCEP. To clearly identify the effect of these conditions on the protease we selected a model substrate protein – myoglobin, which does not require denaturation or reduction for proper digestion. Thus, it allowed us to observe the effect of the denaturing conditions on the protease action. To obtain a reference point, myoglobin was first digested under the tested conditions by non-immobilized nepenthesin-1 and porcine pepsin in solution. Resulting LC-MS/MS chromatographic traces clearly corroborated our previous findings that pepsin digestion is not affected much by the denaturing and reducing conditions. (**Figure 11 - A, B**) On the other hand the activity loss observed for nepenthesin-1 was dramatic (**Figure 11 - C, D**).

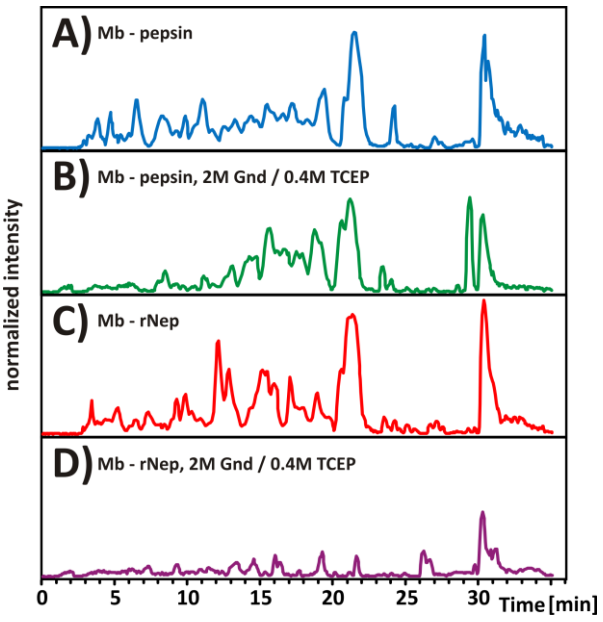


Figure 11: Effect of denaturants on in-solution digestion of myoglobin by porcine pepsin and rNep-1. *Horse myoglobin was digested in solution by pepsin A (A, B) or rNep-1 (C, D) either in glycine buffer pH 2.3 (A, C) or in the same buffer supplemented with 2 M Gnd and 0.4 M TCEP (B, D). Samples were analyzed by LC-MS/MS – normalized base-peak chromatograms were plotted to visualize the impact of denaturing and reducing agents on each protease.*

We then proceeded with analyses of the online digestion using the columns filled with immobilized proteases. Here, the immobilization resulted in a striking difference as the digestion by nepenthesin was maintained even in the presence of denaturants, albeit with some differences apparent in the chromatogram. (**Figure 12 - E, F**) These were

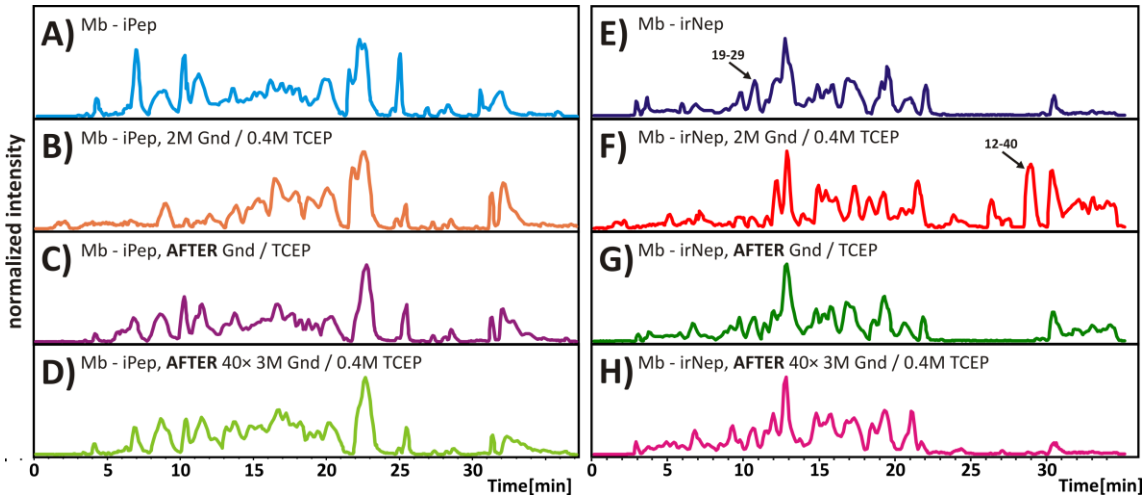


Figure 12: Effect of denaturing and reducing conditions on on-line digestion on columns with immobilized pepsin (A-D) and immobilized rNep-1 (E-H). *Myoglobin was analyzed by LC-MS/MS after online digestion in glycine buffer (A, E), in the same buffer with 2 M Gnd, 0.4 M TCEP (B, F), in glycine buffer after denaturing/reducing run (C, G) and in glycine buffer after 40 injections of 3 M Gnd, 0.4 M TCEP (D, H). The arrows point on example peptide which is present in normal digest (19-29) only and on longer peptide covering this region (12-40) in denaturing/reducing digest.*

apparent as a chromatogram shift towards higher retention times meaning that larger peptides were produced (e.g. peptide 19-29 vs. 12-40). Porcine pepsin was affected as well, although to a smaller extent. (**Figure 12 - A, B**) Importantly, the protein sequence coverage was preserved under all conditions tested, albeit under denaturing conditions it was accomplished with slightly longer peptides. With this in mind, it is important to note that the changes to the cleavage pattern of immobilized enzymes were shown to be reversible (original activity was restored upon removal of denaturing conditions). When, after analyzing myoglobin under denaturing conditions, we injected the next sample, consisting of myoglobin in the glycine buffer alone, the resulting digests for both proteases (**Figure 12 - C, G**) were similar to digests before denaturant injection (**Figure 12 - A, E**). Moreover, we tested the stability of Nep-1 column by 40 injections of buffer containing 3M guanidine and 0.4M TCEP, which was once again followed by the analysis of myoglobin alone. The chromatogram showed (**Figure 12 - H**) that the column easily withstood this treatment and its activity was virtually unperturbed. A precise mechanism for the observed protease stabilization by immobilization is not yet definitely known but is supposed to be the result of multiple effects including limited conformational flexibility of

the immobilized protein.[147–149] We also assume that one of the main factors involved in our case is, among other effects, the short time during which the protease is in contact with the quenched sample buffer. In the case of immobilized proteases the online digestion typically only takes around 40 seconds, whereas the in solution digestion normally lasts several minutes.

Further, we used both pepsin and Nep-1 columns to digest a panel of "real-world" protein substrates, which had been previously studied in our laboratory. For comparison, we digested the same proteins in solution by concentrated pitcher fluids isolated from *Nepenthes* plants as described by Rey *et al.*[121] Generated peptides (more than 1300 for each protease) were identified by LC-MS/MS and processed by an in-house developed script to identify all the unique cleavage sites for the proteases. The data were statistically normalized according to Keil[150] for the occurrence of individual amino acids in the sequences of the substrates and were used to construct plots summarizing the cleavage preferences of the used proteases. (**Figure 13**)

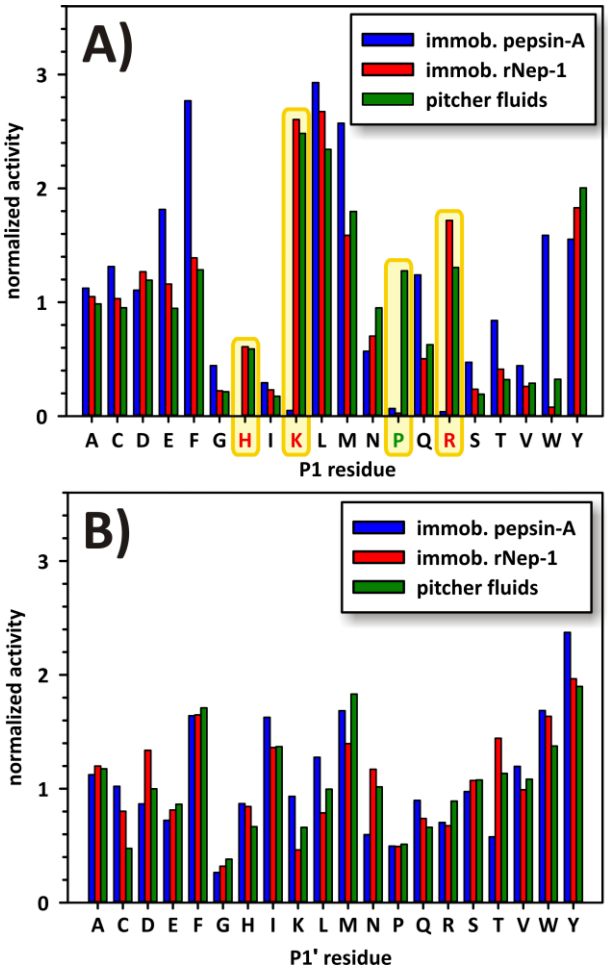


Figure 13: Plots summarizing cleavage preferences after (A) – position P1, and before (B) – position P1', individual amino acids. *Digestion of 8 model proteins was done using immobilized pepsin A (blue), immobilized rNep-1 (red) and pitcher fluid (green). Significant differences in P1 preferences (K, R, H, P) are highlighted.*

All the proteases or protease columns exhibited somewhat similar preferences for the substrate in P1′ position with no apparent differences present (panel B). Although, this was not the case for P1 position (panel A). Here, although generally similar in behavior to porcine pepsin, immobilized rNep-1 showed lower preferences to cleave after phenylalanine and glutamic acid. Moreover, a huge drop in cleavage C-terminal to tryptophan was observed, which contrasted nicely with pepsin. On the other hand, both immobilized rNep-1 and pitcher

fluid cleaved extremely efficiently after all three basic residues, lysine, arginine and histidine. This cleavage pattern was consistent with behavior reported previously for the pitcher fluid.[121] The only striking difference, which we have observed when comparing cleavage preferences of rNep-1 with those of the pitcher fluid obtained by both us and by Rey *et al.*[121], was the missing cleavage activity C-terminal to proline residues. The exact reason for this behavior remained elusive at the time, although we later in collaboration with the group of Prof. David Schriemer ruled out nepenthesin-2 as the source of this activity[151] and attributed it instead to a newly discovered prolyl endopeptidase, which was named neprosin[152].

Overall, the immobilization protocol for nepenthesin-1 enabled us to use it in an HXMS online digestion. It also dramatically increased the tolerance of the enzyme to denaturing and reducing agents. Most importantly, by having cleavage preferences different from porcine pepsin, immobilized Nep-1 has shown promise as a tool to increase the digestion efficiency of proteins in HXMS and thus to increase the spatial resolution of the method.

4.2. Structural MS studies of CDH conformational dynamics

*Results in this section were included in the attached **paper III** – Kadek, A. et al. Structural insight into the calcium ion modulated interdomain electron transfer in cellobiose dehydrogenase. FEBS Lett. 589, 1194–1199 (2015) as well as in the manuscript of **paper IV** – Kadek, A. et al. Interdomain electron transfer in cellobiose dehydrogenase is governed by surface electrostatics. - submitted.*

Cellobiose dehydrogenase is a key enzymatic system involved in cellulose depolymerization in nature.[133] Additionally, it is being studied extensively by biotechnologists with the aim to harness its enzymatic activity for the construction of biosensors and fuel cells.[129,130] During its enzymatic cycle, CDH oxidizes mono- and disaccharides to lactones. Electrons obtained from this reaction are then channeled by CDH to a terminal electron acceptor.[133] However, for this process to occur, electrons first need to be internally transferred between FAD and heme cofactors in its two domains. This is supposed to occur through a direct interdomain electron transfer, for which the CDH domains need to get into close contact.[138] The IET has been described to be modulated by both pH and divalent cations in solution[134,136,137] and the process has been supposed to involve dynamic protein motions. As the CDH enzyme is quite large and at the same time

very flexible, the studies of its function in solution proved to be challenging. Therefore, this motivated us to employ structural MS techniques to study CDH's conformational dynamics with the aim to experimentally explain the mechanisms underlying its functioning and modulation of its activity.

4.2.1. Optimization of HXMS conditions for CDH analyses

In order to provide insight into the conformational dynamics of CDH in solution, HXMS was selected as a suitable structural approach. However, experimental conditions of the method needed to be optimized to provide us with as much detail in structural data as possible. We started by tuning the process of *Mt*CDH enzymatic digestion since proteolysis is the key factor determining the spatial resolution of HXMS.

For this, *Mt*CDH was digested under various HXMS-compatible conditions (pH 2.5, 0°C), while the resulting peptides were analyzed using LC-MS/MS. (**Figure 14**) In **paper III** we showed that when it was digested by immobilized porcine pepsin in

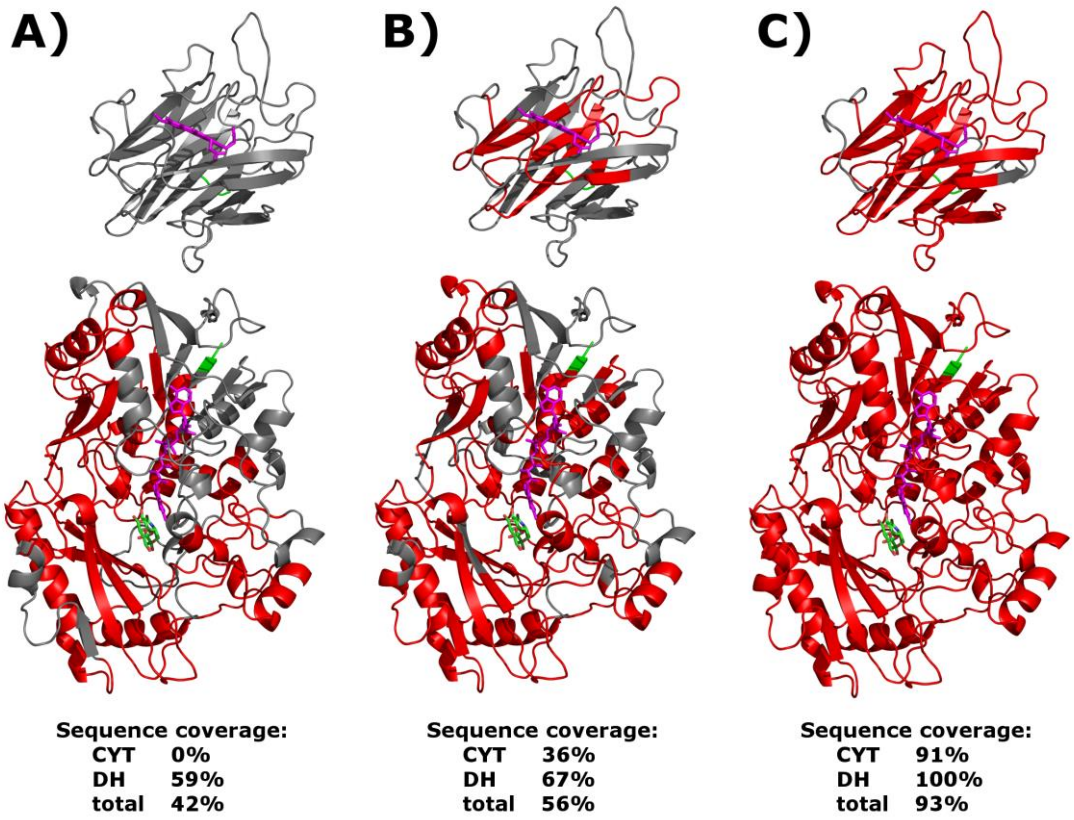


Figure 14: Optimization of HXMS-compatible digestion conditions of MtCDH. *Three different digestion conditions are compared:* ***(A)*** *- native MtCDH;* ***(B)*** *– MtCDH after reduction and deglycosylation;* ***(C)*** *– deglycosylated protein subjected to reduction and digestion in 3 M guanidine. Regions covered under individual conditions are shown in red. Green highlights sites where a flexible linker connects the two domains. The orientation of the domains is purely schematic and does not reflect their natural position. Sequence coverage for the individual conditions for each domain separately as well as for the whole protein is shown below.*

glycine buffer pH 2.5 alone, the identified peptides covered only about 42% of the whole protein and no coverage was obtained for the CYT domain whatsoever. (**Figure 14a**, p. 42) It then became obvious that post-translational modifications are present in the molecule and hamper its analysis. Therefore, we used Endo Hf endoglycosidase to remove N-linked glycans from the *Mt*CDH molecule, while preserving the proximal N-acetylhexosamine attached to the molecule. This was deliberately chosen in order to not perturb the amino acid structure of *Mt*CDH and the distribution of charge on its surface. This would happen if the glycans were removed completely and asparagines were thus converted to aspartates (e.g. by the action of PNGase F glycosidase). For the Endo Hf deglycosylation, we utilized non-denaturing conditions, so that the enzymatic activity of CDH was not compromised.[153] The non-denaturing nature of the deglycosylation process was later confirmed experimentally, as our HXMS analyses comparing the N-deglycosylated and fully glycosylated enzyme showed no significant structural changes in *Mt*CDH molecule. (**paper IV**) Following the deglycosylation we once again digested *Mt*CDH under HXMS conditions, but this time also in the presence of reducing agent TCEP, as we suspected disulfide bonds to be present in the CDH molecule. This led to partial increase in its sequence coverage, but still, the digestion efficiency of the tightly folded CYT domain was insufficient. (**Figure 14b**, p. 42) In the end, quenching buffer resulting in a final concentration of 3 M guanidine with 0.45 M TCEP was needed to achieve almost complete peptide coverage of *Mt*CDH. (**Figure 14c**, p. 42) The only region we missed was the flexible interdomain linker, where we identified O-glycosylation at multiple sites. This then prevented the identification of peptides produced by non-specific proteolysis. Regardless, the overall sequence coverage of 93% already provided a very solid base for HXMS structural studies.

Later, as we prepared immobilized recombinant nepenthesin-1 and rhizopuspepsin for the use in HXMS, we tested their efficiency in CDH digestion under the denaturing conditions identified with porcine pepsin. In **paper IV** we demonstrated that the best results were obtained with combined digestion utilizing sequentially coupled nepenthesin and rhizopuspepsin (Rpn) protease columns. Cleavage sites partially complementary to those provided by porcine pepsin resulted in more efficient digestion of *Mt*CDH with more peptide overlaps and generally shorter produced fragments. The practical benefit of this can be demonstrated on the example of the ultimate N-terminus of CYT domain. There

deuteration changes identified by HXMS could only be localized into two rather long peptides (orange and red color) when porcine pepsin was used (**Figure 15b**). However, the complementary cleavage sites introduced by Nep-1 and Rpn allowed us to localize the deuteration differences more precisely to the two loops in the structure (**Figure 15c**).

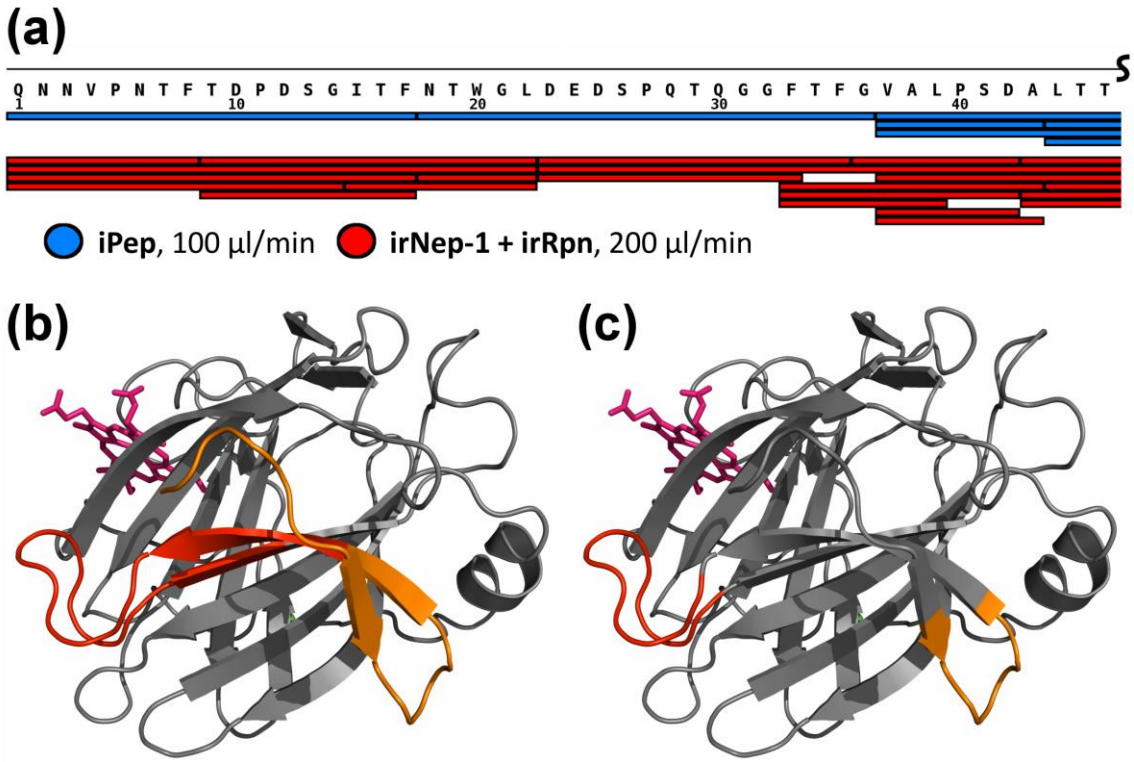


Figure 15: Alternative aspartic proteases increase the spatial resolution of HXMS. (a) *Peptide mapping of the ultimate N-terminus of MtCDH. Individual peptides are drawn as blue bars (immobilized pepsin - Pep) and red bars (combined digestion by immobilized nepenthesin-1 – Nep-1 and rhizopuspepsin - Rpn). The additional cleavage sites enabled more precise localization of deuteration changes at the N-terminus of CYT domain (red and orange regions) with the combined irNep-1+irRpn digestion* ***(c)*** *compared to porcine pepsin* ***(b)****.*

In conclusion of this section, HXMS-compatible proteolysis conditions were optimized for *Mt*CDH, which provided reasonable spatial resolution and enabled analyzing the behavior of almost all the regions in the studied protein. The immobilization of Nep-1 here proved to be critical. It enabled the use of Nep-1 for the digestion of CDH, which probably due to its very stable β-fold and seven disulfide bonds in its molecule necessitated the use of harsh denaturing conditions during the quenching of H/D exchange reaction. Furthermore, together with Dr. Petr Man and Dr. Daniel Kavan, who wrote the computer code, we designed, beta-tested and have been actively developing a novel software tool DeutEx for dealing with raw HXMS data. (as yet unpublished, but with a demonstration video available at http://ms.biomed.cas.cz/SWD/DeutExCDH.wmv) It enabled rapid processing of large dataset with the possibility to manually validate all peak assignments in

an interactive graphical user interface. Also it provided means to easily correct for the influence of pH or other factors. Overall, the software led to much faster and less ambiguous HXMS workflow than previously utilized script-based or manual approach[154]. This development later proved invaluable for dealing with complex CDH datasets.

4.2.2. Mechanism of CDH modulation by pH and cations studied by HXMS

New software tools and optimized experimental procedures for HXMS analyses of *Mt*CDH allowed us to study the mechanisms of its modulation in solution. In **paper III** we focused on the recently described activation of *Mt*CDH in slightly alkaline pH by the presence of divalent cations.[137] We approached this by analyzing the protein in solution at pH 7.4 alone or in the presence of 30 mM calcium cations. For comparison, we also tested calcium-free conditions, where the exactly same ionic strength was instead set by potassium ions or chelating agent EDTA anions. Structural changes observed on both CYT (**Figure 16**) and DH (**Figure 17**, p. 46) domains were localized around the proposed interdomain interface in the enzyme and could be distinguished into two categories.

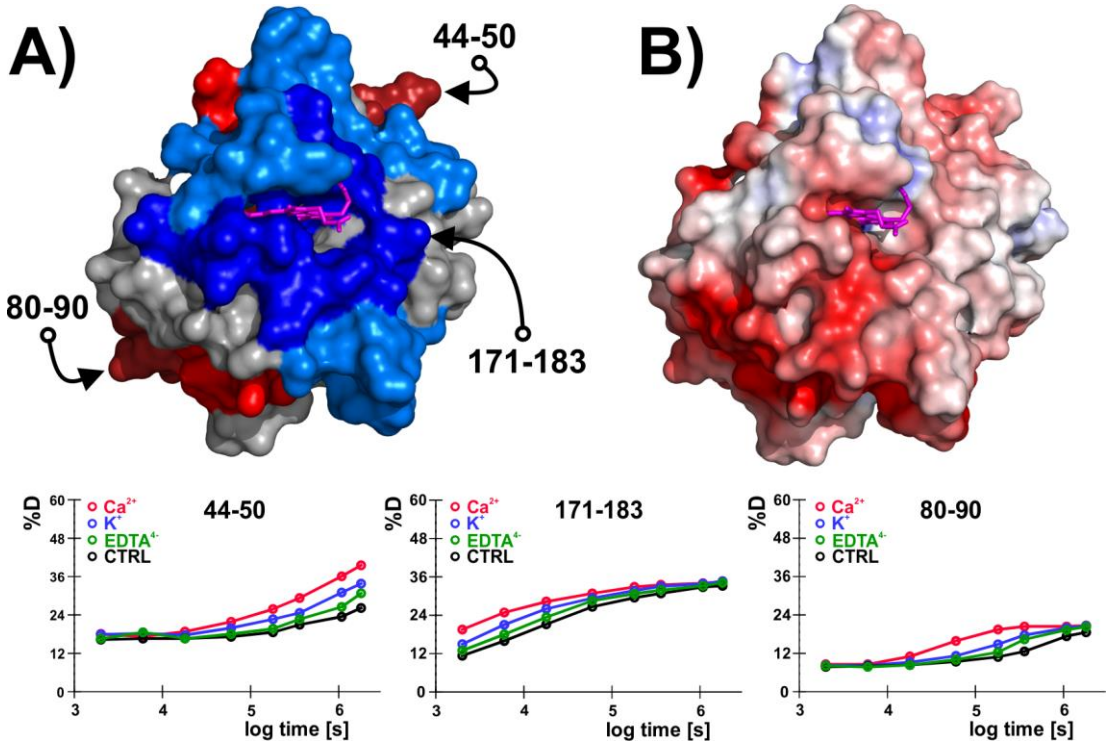


Figure 16: Influence of calcium ions on the MtCDH cytochrome domain at pH 7.4. A) *HXMS results (below) visualized as colored regions on the homology model of CYT domain. The heme b cofactor is shown in magenta. Observed changes can be attributed mainly to the ionic strength itself (blue colors), to the presence of divalent calcium ions (red colors) or to the combination of the two factors (44-50, 80-90 or 171-183; coloring according to the more prominent component).* ***B)*** *Calculated surface electrostatics at pH 7.4 - colored as a gradient from red to blue (-4 to +4 kT/e, respectively). Patches of negative charge correlate with regions of calcium interaction identified by the HXMS. Domain size is not to scale with DH domain in* ***Figure 17, p. 46****.*

In some parts of the protein, changes were mostly caused by ionic strength (blue regions), while in others calcium ions caused even stronger effect than the ionic strength alone (red).

Comparing these regions with protein surface electrostatics simulations calculated by Adaptive Poisson-Boltzmann Solver algorithm[155,156], we noticed that regions of increased calcium influence collocalized well with patches of negative charge on the surfaces of *Mt*CDH domains. This confirmed a role of divalent cations in shielding these to enable domain contacting and thus experimentally validated previous mechanism predictions based on molecular docking approaches.[137]

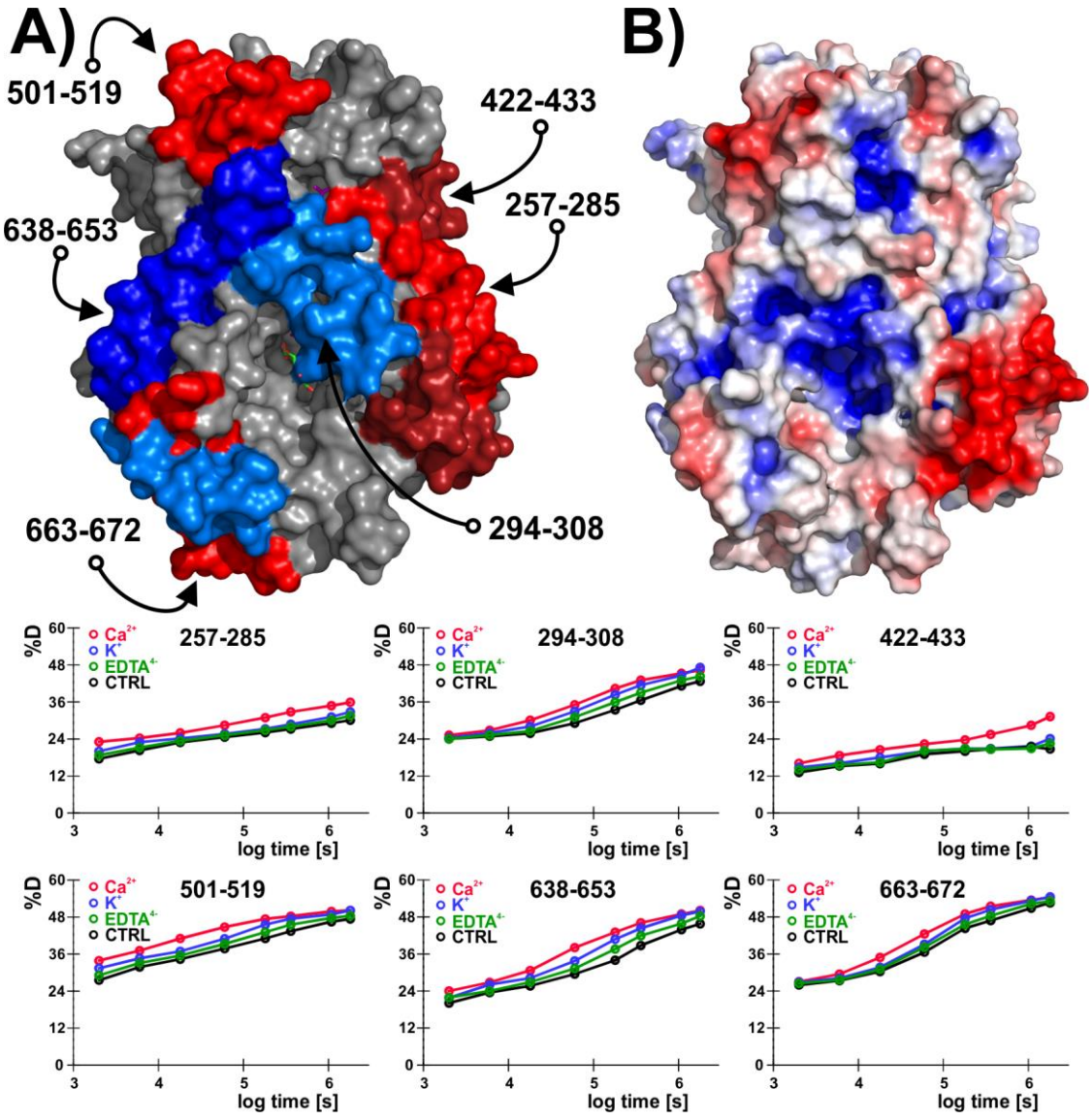


Figure 17: Influence of calcium ions on the MtCDH dehydrogenase domain at pH 7.4. A) *HXMS results (below) visualized as colored regions on the homology model of DH domain. Part of the cellobiose molecule is visible in the substrate entry channel close to the buried FAD cofactor. Observed changes can be attributed mainly to the ionic strength itself (blue colors; region 294-308), to the presence of divalent calcium ions (red colors; 257-285 or 422-433) or to the combination of the two factors (e.g. 638-653; coloring according to the more prominent component).* ***B)*** *Calculated surface electrostatics at pH 7.4 - colored as a gradient from red to blue (-4 to +4 kT/e, respectively). Patches of negative charge correlate with regions of calcium interaction identified by the HXMS.*

In **paper IV** we observed similar high importance of negatively charged regions when studying the modulation of CDH activity by pH. Using HXMS we compared *Mt*CDH at pH 5.4, where the enzyme is close to its pH optimum and pH 7.4, where it is completely inactive.[137,157,158] In this analysis, we again observed that the regions perturbed by the change in pH were mostly localized close around the domain interface. (**Figure 18,** p. 48) The difference in deuteration behavior can for these experiments be observed in the overlapping regions of HXMS curves, as in this case the curves for pH 5.4 and 7.4 conditions are offset. This is in order to correct for the intrinsic effect of pH on the deuteration rate. As at higher pH the exchange is faster by a factor of $10^{\Delta pD}$, the difference of 2 pH units results in one curve being offset by a factor of 100 on the logarithmic time axis. Similarly as with the calcium ion modulation, we correlated the HXMS data with surface electrostatics simulations of the protein in solution at the two studied pH values. This led to the observation that the most pronounced HXMS changes are in regions, where negative charge on protein surface gets neutralized when the solution is acidified. Furthermore, we repeated the experiment working with the two separated *Mt*CDH domains and it was shown that the individual CYT and DH domains behaved virtually identically as the full-length protein. Therefore, we concluded that the changes observed by HXMS at the two different pH values were not in fact caused by the interaction of the two CDH domains at the lower pH, but were rather caused by the neutralization of charge due to protonation.

Together, our results from both the ion modulation (**paper III**) and pH regulation (**paper IV**) experiments thus provided direct structurally localized experimental support for the theory of charge repulsion in CDH. According to this theory the electrostatic repulsion between domains at higher pH prevents their contacting and successful IET.[137,138] According to HXMS data, divalent cations can shield and/or bridge the patches of negative charge and enable the domain contacting even at slightly alkaline pH. Alternatively, when the pH of solution is lowered, the charge around the edge of domain interface on DH gets neutralized by protonation. Both these mechanisms ultimately enable the IET between domains and CDH functioning.

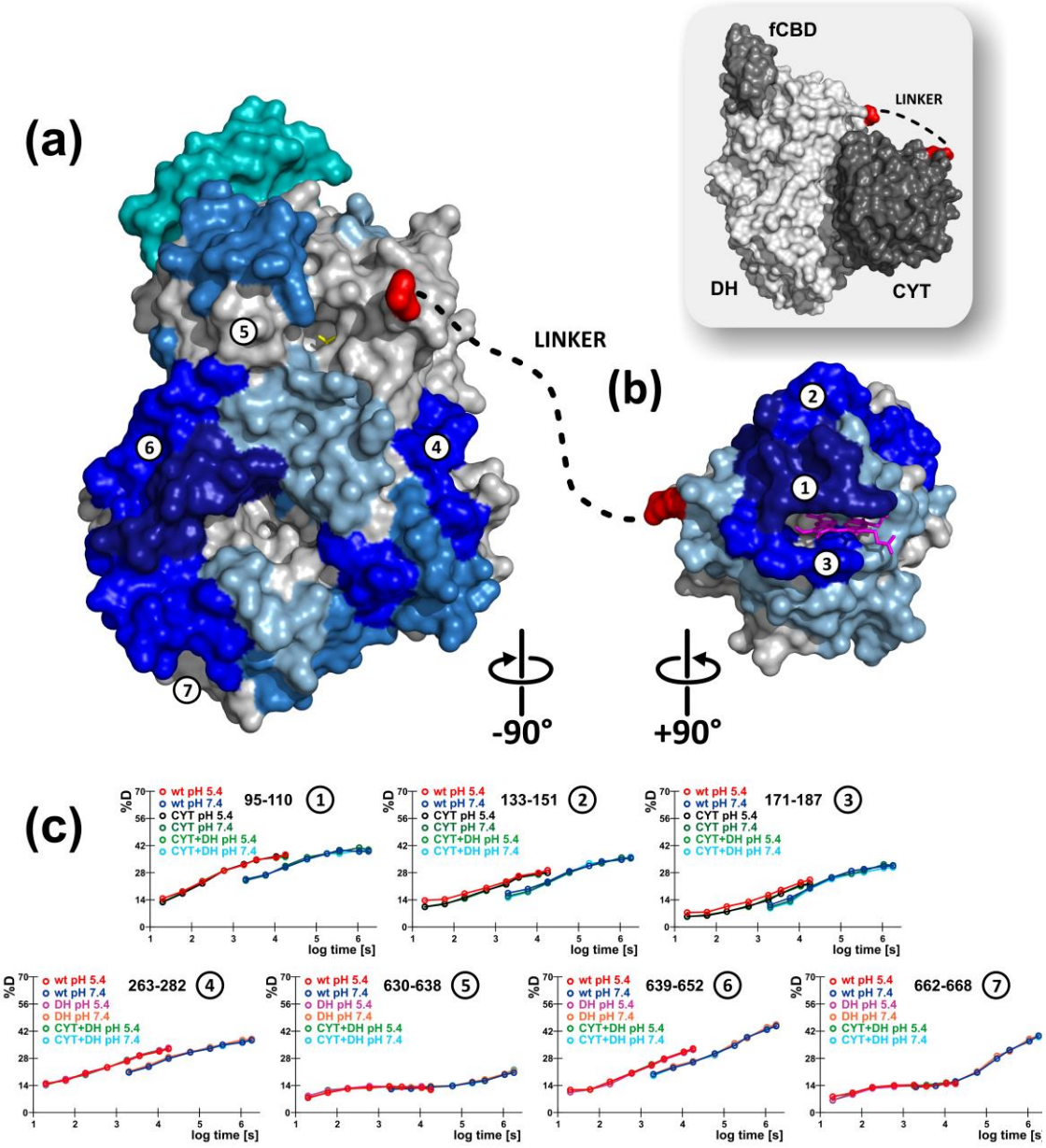


Figure 18: HXMS detected changes at the MtCDH interdomain interface between pH 5.4 (active) and 7.4 (inactive). ***(a)*** *DH domain with the most intense changes around the cusp of FAD cavity. Some perturbance was also seen for the fungal carbohydrate binding module (shown in teal color).* ***(b)*** *CYT domain with strongest pH effects detected around the heme cofactor at the DH-binding surface. For the sake of clarity the domains were rotated by 90° in opposite directions from their natural orientation (inset). MtCDH structure (PDB ID: 4QI6) was colored according to the detected changes in deuteration. Grey means no difference while different shades of blue mark the deprotected regions with dark blue showing the biggest difference. Red regions show the connecting points of interdomain linker.* ***(c)*** *HDX-MS curves for the individual regions of full-length MtCDH at pH 5.4 and 7.4 (red and dark blue curves, respectively). Separated domains (black, dark green, magenta and orange) and their mixture (light green and cyan) behave similarly as full-length protein. MtCDH regions unaffected by pH show no difference (5, 7), while in the perturbed regions there is significant deprotection at the lower pH.*

Interestingly, we did not observe any "classical" decrease of deuteration at the interdomain interface when the two domains interacted in the presence of calcium ions or at the lower pH. The interaction should theoretically have led to solvent exclusion from the

interface, to formation of stabilizing interprotein hydrogen bonds and to decrease in the deuteration level of affected peptides. However, in HXMS we exclusively saw increased levels of deuterium incorporation in all the affected regions. This can be explained by two contributing factors: (1) The CDH is highly flexible and its IET competent closed complex is only very short-lived, transient and not very stable in solution. This is in agreement with previously published SAXS data[141], which showed CDH to be present in a range of semi-open states in solution with only a small amount of the conformers actually being in the closed active state. (2) The increased deuteration itself is then the result of protonation- or ion-mediated reorganization of hydrogen bonding and charge networks present in the molecule, which are however necessary for successful transient domain interaction. This for some protein backbone amides led to their increased solvent accessibility. Alternatively, the affected regions around the *Mt*CDH interdomain interface might have slightly loosened structurally to better accommodate the transient domain-domain contact.

The results proved to be rather non-canonical in the field of HXMS. Nevertheless, taken into account the highly dynamic nature of CDH enzymatic action, they fit well into the current concept of different possible outcomes observable in HXMS.[91,93,159] Our data thus emphasize the need to look beyond the most common deuteration scenarios when interpreting dynamic behavior of complex protein systems – especially of those that involve side chain-mediated transient interactions with no distinct transition between active and inactive states.

4.2.3. Electrostatic stability of CDH ions probed in the gas phase

Finally, we validated our HXMS findings about the identified electrostatic repulsion in CDH by a complementary structural MS method – native MS coupled with ion mobility. (**paper IV**) Although this technique studies molecules in the gas phase, it has been shown that when properly tuned, it can gently transfer proteins into vacuum while preserving significant features of their tertiary and quaternary structures.[44] Ion mobility then provides information about the size and shape of studied ions by measuring the time ions take to traverse through a cell filled with inert gas.

*Mt*CDH was ionized by nanoelectrospray from solutions of pH 5.4 as well as 7.4. In the raw MS data collected, identical protein charge state was studied for ions from both studied pH solutions. Comparison of ion mobility “drift time” profiles extracted for these ions (**Figure 19**, p. 50) under carefully tuned experimental conditions initially did not

show any difference in the behavior of ions. However, intentional gradual "activation" of the studied ions in the gas phase (collisional increasing of their internal energy) showed that they collapse to more compact structures. Most importantly, we observed that the ions produced from pH 5.4 required less activation energy and collapsed easier than those from pH 7.4.

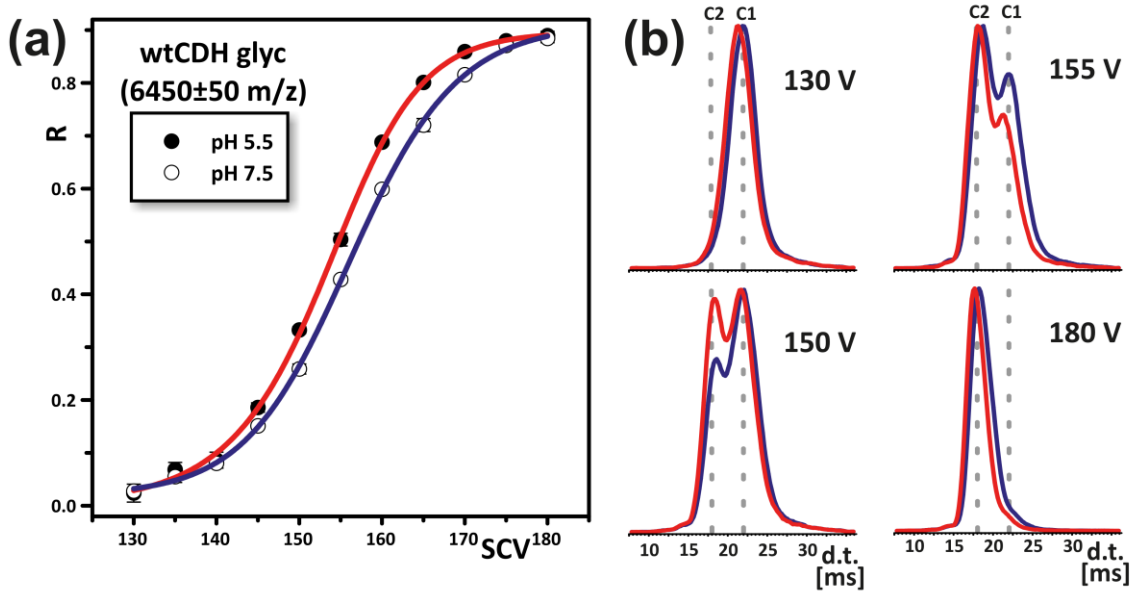


Figure 19: Native IM-MS shows different electrostatic stability of MtCDH ionized from pH 5.4 (red) and pH 7.4 (blue). (a) *Ions of fully-glycosylated MtCDH generated from lower pH collapse more easily with increasing sample cone voltage (SCV) than ions of the same charge state from higher pH. Conformer ratio R = C2 / (C1+C2), where C1 and C2 are areas-under-curve in drift time distributions for individual activation voltages* ***(b)****. d.t. - "drift" time through the ion mobility cell.*

Such collapse of ions is not the most common scenario upon collisional activation of proteins, which normally tend to unfold and dissociate.[160,161] Nevertheless, it has been previously observed for molecules containing internal cavities.[161] Therefore, this led us to propose an explanation that as CDH is very flexible, it gets transferred into the gas phase in similarly-sized partially open conformations from solutions of both pH values. Only when the energy of ions is increased through collision with gas, the molecules are forced into a more closed conformation, where CYT domain of CDH might be protruding into the cavity of DH domain. The observed differences in how easily this occurs then reflect for ions the amount of repulsion between negative charges at their domain-domain interface. Our observation that the energy necessary to compact *Mt*CDH molecules ionized from pH 5.4 is less than for those from pH 7.4 is then consistent with both our surface electrostatics calculations and enzymatic activity pH profile of *Mt*CDH.[137] Although the strength of electrostatic interactions is slightly overestimated *in vacuo* and cannot be directly compared with the strength of forces acting in solution[162], the clear difference in

electrostatic stability of *Mt*CDH ions observed by native IMMS corroborated the results obtained by HXMS and computational surface electrostatics modeling.

In conclusion of this thesis, novel recombinantly produced aspartic protease nepenthesin-1 was characterized in solution. Following its successful immobilization onto perfusion resin, it was shown to be active even in the presence of high concentrations of denaturing and reducing agents, which are otherwise incompatible with its use in solution. Immobilized Nep-1 was then used for online digestion of substrates in an HXMS bottom-up workflow and it was shown to significantly differ in its cleavage preferences from classical porcine pepsin. Specifically, combined online digestion by immobilized Nep-1 column sequentially coupled to rhizopuspepsin column proved beneficial for increasing the digestion efficiency of *Mt*CDH and the spatial resolution of its HXMS analyses. These analyses provided structurally localized information about the effects of calcium ions as well as changes of solution pH on the structure of *Mt*CDH. Together with IMMS, which probed CDH ion electrostatics in the gas phase, and computational protein surface electrostatics calculations, HXMS provided an insight into the dynamics of *Mt*CDH. In the context of existing SAXS data and a static high resolution structure of this enzyme obtained by crystallography, structural MS techniques experimentally confirmed that surface electrostatic repulsion at the interdomain interface is the key factor governing its functioning in solution. This showcased the power of integrative structural biology approaches combining data from different experimental techniques to structurally explain challenging biological questions.

5. SUMMARY

The aims of this doctoral thesis were (1) to contribute to the development of structural mass spectrometric techniques for studies of protein conformational dynamics and (2) to apply these techniques to study enzymatic system of cellobiose dehydrogenase. Following results were obtained and included in the four scientific publications attached to directly support this thesis:

- Recombinant aspartic protease nepenthesin-1 was shown to behave similarly as the enzyme isolated directly from plants.
- Nep-1 was shown to be highly susceptible to the presence of denaturing and reducing agents in solution.
- The denaturation and reduction sensitivity of Nep-1 was overcome by its immobilization onto perfusion resin.
- Immobilized Nep-1 was successfully tested for HXMS-compatible online digestion of proteins.
- Proteolytic preferences of Nep-1 differed significantly from porcine pepsin by cleavages after basic residues.
- The combined use of immobilized Nep-1 with rhizopuspepsin increased spatial resolution of HXMS in studies of *Mt*CDH structure and dynamics.
- The effect of divalent cations on the activity of *Mt*CDH was shown to involve bridging and shielding of opposing patches of negative charge at the interdomain interface.
- The pH regulation of *Mt*CDH activity was showed to similarly be based on the negative charge neutralization by protonation at slightly acidic pH.
- Native MS with ion mobility showed different gas phase electrostatic stability of *Mt*CDH ions produced from acidic or neutral pH.
- Integration of structural MS data provided direct experimental support for the charge repulsion theory of CDH functioning in solution.
- Additionally, powerful software for handling HXMS datasets has been co-developed during the work on this thesis to streamline data evaluation and processing for *Mt*CDH and other projects.

LIST OF PUBLICATIONS

Publications directly supporting this doctoral thesis:

1) Kadek A, Tretyachenko V, Mrazek H, Ivanova L, Halada P, Rey M, Schriemer DC & Man P (2014) Expression and characterization of plant aspartic protease nepenthesin-1 from Nepenthes gracilis. Protein Expr. Purif. 95, 121–128.

2) Kadek A, Mrazek H, Halada P, Rey M, Schriemer DC & Man P (2014) Aspartic protease nepenthesin-1 as a tool for digestion in hydrogen/deuterium exchange mass spectrometry. Anal. Chem. 86, 4287–94.

3) Kadek A, Kavan D, Felice AKG, Ludwig R, Halada P & Man P (2015) Structural insight into the calcium ion modulated interdomain electron transfer in cellobiose dehydrogenase. FEBS Lett. 589, 1194–1199.

4) Kadek A, Kavan D, Marcoux J, Stojko J, Felice AKG, Cianférani S, Ludwig R, Halada P & Man P (2016) Interdomain electron transfer in cellobiose dehydrogenase is governed by surface electrostatics. *submitted*

Other publications by the author:

5) Yang M, Hoeppner M, Rey M, **Kadek A**, Man P & Schriemer DC (2015) Recombinant Nepenthesin II for Hydrogen/Deuterium Exchange Mass Spectrometry. Anal. Chem. 87, 6681–7.

6) Kacirova M, Kosek D, **Kadek A**, Man P, Vecer J, Herman P, Obsilova V & Obsil T (2015) Structural Characterization of Phosducin and Its Complex with the 14-3-3 Protein. J. Biol. Chem. 290, 16246–60.

7) Fejfarová K, **Kádek A**, Mrázek H, Hausner J, Tretyachenko V, Koval' T, Man P, Hašek J & Dohnálek J (2016) Crystallization of nepenthesin I using a low-pH crystallization screen. Acta Crystallogr. Sect. F Struct. Biol. Commun. 72, 24–28.

8) Vit O, Man P, **Kadek A**, Hausner J, Sklenar J, Harant K, Novak P, Scigelova M, Woffendin G & Petrak J (2016) Large-scale identification of membrane proteins based on analysis of trypsin-protected transmembrane segments. J. Proteomics *in press*.

BIBLIOGRAPHY

(1) Berg, J. M.; Tymoczko, J. L.; Stryer, L.; Gatto, G. J. *Biochemistry - 7th edition*; W.H.Freeman and Company: New York, 2012.
(2) Tsuji, S.; Martin, B. M.; Barranger, J. A.; Stubblefield, B. K.; LaMarca, M. E.; Ginns, E. I. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85* (7), 2349–2352.
(3) Angarica, V. E.; Orozco, M.; Sancho, J. *Hum. Mol. Genet.* **2015**, *25* (6), 1233–1246.
(4) Glenner, G. G.; Wong, C. W. *Biochem. Biophys. Res. Commun.* **1984**, *122* (3), 1131–1135.
(5) Grundke-Iqbal, I.; Iqbal, K.; Quinlan, M.; Tung, Y. C.; Zaidi, M. S.; Wisniewski, H. M. *J. Biol. Chem.* **1986**, *261* (13), 6084–6089.
(6) Masliah, E.; Iwai, A.; Mallory, M.; Uéda, K.; Saitoh, T. *Am. J. Pathol.* **1996**, *148* (1), 201–210.
(7) Silvestrini, M. C.; Cardone, F.; Maras, B.; Pucci, P.; Barra, D.; Brunori, M.; Pocchiari, M. *Nat. Med.* **1997**, *3* (5), 521–525.
(8) Valastyan, J. S.; Lindquist, S. *Dis. Model. Mech.* **2014**, *7* (1), 9–14.
(9) McGowan, L. C.; Hamelberg, D. *Biophys. J.* **2013**, *104* (1), 216–226.
(10) Jayaraman, V.; Rodgers, K. R.; Mukerji, I.; Spiro, T. G. *Science* **1995**, *269* (5232), 1843–1848.
(11) Mendieta, J.; Gago, F. *J. Mol. Graph. Model.* **2004**, *23* (2), 189–198.
(12) Morth, J. P.; Pedersen, B. P.; Toustrup-Jensen, M. S.; Sørensen, T. L.-M.; Petersen, J.; Andersen, J. P.; Vilsen, B.; Nissen, P. *Nature* **2007**, *450* (7172), 1043–1049.
(13) Murata, T.; Yamato, I.; Kakinuma, Y.; Leslie, A.; Walker, J. *Science (80-.).* **2005**, *308* (5722), 654–659.
(14) Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D. *Nature* **2007**, *450* (7171), 913–916.
(15) Wu, Z.; Xing, J. *Biophys. J.* **2012**, *103* (5), 1052–1059.
(16) Henzler-Wildman, K.; Kern, D. *Nature* **2007**, *450* (7172), 964–972.
(17) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. *Nature* **1958**, *181* (4610), 662–666.
(18) Wlodawer, A.; Minor, W.; Dauter, Z.; Jaskolski, M. *FEBS J.* **2008**, *275* (1), 1–21.
(19) Egli, M. In *Current Protocols in Nucleic Acid Chemistry*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2010; Vol. 2009, pp 7.13.1–7.13.35.
(20) Rozbesky, D.; Man, P.; Kavan, D.; Chmelik, J.; Cerny, J.; Bezouska, K.; Novak, P. *Anal. Chem.* **2012**, *84* (2), 867–870.
(21) Rozbesky, D.; Sovova, Z.; Marcoux, J.; Man, P.; Ettrich, R.; Robinson, C. V; Novak, P. *Anal. Chem.* **2013**, *85* (3), 1597–1604.
(22) Nagayama, K.; Wüthrich, K. *Naturwissenschaften* **1977**, *64* (11), 581–583.
(23) Wang, G.; Zhang, Z.-T.; Jiang, B.; Zhang, X.; Li, C.; Liu, M. *Anal. Bioanal. Chem.* **2014**, *406*, 2279–2288.
(24) Tugarinov, V.; Sprangers, R.; Kay, L. E. *J. Am. Chem. Soc.* **2004**, *126* (15), 4921–4925.
(25) Dubochet, J.; Adrian, M.; Chang, J. J.; Homo, J. C.; Lepault, J.; McDowall, A. W.; Schultz, P. *Q. Rev. Biophys.* **1988**, *21* (2), 129–228.
(26) Saibil, H. R. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2000**, *56* (10), 1215–1222.
(27) Zhou, Z. H. *Curr. Opin. Struct. Biol.* **2008**, *18* (2), 218–228.
(28) Putnam, C. D.; Hammel, M.; Hura, G. L.; Tainer, J. A. *Q. Rev. Biophys.* **2007**, *40* (3), 191–285.
(29) Petoukhov, M. V.; Svergun, D. I. *Int. J. Biochem. Cell Biol.* **2013**, *45* (2), 429–437.
(30) Schuler, B.; Eaton, W. A. *Curr. Opin. Struct. Biol.* **2008**, *18* (1), 16–26.
(31) Politis, A.; Borysik, A. J. *Proteomics* **2015**, *15* (16), 2792–2803.
(32) Thomson, J. J. *Proc. R. Soc.* **1913**, *A 89*, 1–20.
(33) Benesch, J. L.; Ruotolo, B. T. *Curr. Opin. Struct. Biol.* **2011**, *21* (5), 641–649.
(34) Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T. *Rapid Commun. Mass Spectrom.* **1988**, *2* (8), 151–153.
(35) Karas, M.; Bachmann, D.; Bahr, U.; Hillenkamp, F. *Int. J. Mass Spectrom. Ion Process.*

1987, *78*, 53–68.
(36) Karas, M.; Bachmann, D.; Hillenkamp, F. *Anal. Chem.* **1985**, *57* (14), 2935–2939.
(37) Yamashita, M.; Fenn, J. B. *J Phys.Chem.* **1984**, *88* (20), 4451–4459.
(38) Whitehouse, C. M.; Dreyer, R. N.; Yamashita, M.; Fenn, J. B. *Anal. Chem.* **1985**, *57* (3), 675–679.
(39) Loo, J. A.; Giordani, A. B.; Muenster, H. *Rapid Commun. Mass Spectrom.* **1993**, *7* (3), 186–189.
(40) Loo, J. a. *Mass Spectrom. Rev.* **1997**, *16* (1), 1–23.
(41) Ruotolo, B. T.; Giles, K.; Campuzano, I.; Sandercock, A. M.; Bateman, R. H.; Robinson, C. V. *Science (80-.).* **2005**, *310* (5754), 1658–1661.
(42) Uetrecht, C.; Rose, R. J.; van Duijn, E.; Lorenzen, K.; Heck, A. J. *Chem. Soc. Rev.* **2010**, *39* (5), 1633–1655.
(43) Chen, S.-H.; Russell, D. H. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (9), 1433–1443.
(44) Ruotolo, B. T.; Robinson, C. V. *Curr. Opin. Chem. Biol.* **2006**, *10* (5), 402–408.
(45) Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S.-J.; Robinson, C. V. *Nat. Protoc.* **2008**, *3* (7), 1139–1152.
(46) Wilm, M. S.; Mann, M. *Int. J. Mass Spectrom. Ion Process.* **1994**, *136*, 167–180.
(47) Sun, Y.; Vahidi, S.; Sowole, M. A.; Konermann, L. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (1), 31–40.
(48) Clemmer, D. E.; Jarrold, M. F. *J. mass Spectrom.* **1997**, *32* (April), 577–592.
(49) Politis, A.; Stengel, F.; Hall, Z.; Hernández, H.; Leitner, A.; Walzthoeni, T.; Robinson, C. V; Aebersold, R. *Nat. Methods* **2014**, *11* (4), 403–406.
(50) Hernández, H.; Robinson, C. V. *Nat. Protoc.* **2007**, *2* (3), 715–726.
(51) Blackwell, A. E.; Dodds, E. D.; Bandarian, V.; Wysocki, V. H. *Anal. Chem.* **2011**, *83* (8), 2862–2865.
(52) Stojko, J.; Fieulaine, S.; Petiot-Bécard, S.; Van Dorsselaer, A.; Meinnel, T.; Giglione, C.; Cianférani, S. *Analyst* **2015**, *140* (21), 7234–7245.
(53) Dyachenko, A.; Gruber, R.; Shimon, L.; Horovitz, A.; Sharon, M. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (18), 7235–7239.
(54) Debaene, F.; Boeuf, A.; Wagner-Rousset, E.; Colas, O.; Ayoub, D.; Corvaïa, N.; Van Dorsselaer, A.; Beck, A.; Cianférani, S. *Anal. Chem.* **2014**.
(55) Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V; Ruotolo, B. T. *Anal. Chem.* **2010**, *82* (22), 9557–9565.
(56) Sinz, A. *Mass Spectrom. Rev.* **2006**, *25* (4), 663–682.
(57) Lomant, A. J.; Fairbanks, G. *J. Mol. Biol.* **1976**, *104* (1), 243–261.
(58) Wang, D.; Wilson, G.; Moore, S. *Biochemistry* **1976**, *15* (3), 660–665.
(59) Novak, P.; Kruppa, G. H. *Eur. J. Mass Spectrom. (Chichester, Eng).* **2008**, *14* (6), 355–365.
(60) Schwarz, R.; Tänzler, D.; Ihling, C. H.; Müller, M. Q.; Kölbel, K.; Sinz, A. *J. Med. Chem.* **2013**, *56* (11), 4252–4263.
(61) Ptackova, R.; Jecmen, T.; Novak, P.; Hudecek, J.; Stiborova, M.; Sulc, M. *Int. J. Mol. Sci.* **2014**, *15* (6), 9224–9241.
(62) Lübben, M.; Portmann, R.; Kock, G.; Stoll, R.; Young, M. M.; Solioz, M. *Biometals* **2009**, *22* (2), 363–375.
(63) Morgner, N.; Schmidt, C.; Beilsten-Edmands, V.; Ebong, I. obong; Patel, N. A.; Clerico, E. M.; Kirschke, E.; Daturpalli, S.; Jackson, S. E.; Agard, D.; Robinson, C. V. *Cell Rep.* **2015**, *11* (5), 759–769.
(64) Navare, A. T.; Chavez, J. D.; Zheng, C.; Weisbrod, C. R.; Eng, J. K.; Siehnel, R.; Singh, P. K.; Manoil, C.; Bruce, J. E. *Structure* **2015**, *23* (4), 762–773.
(65) Schmidt, C.; Zhou, M.; Marriott, H.; Morgner, N.; Politis, A.; Robinson, C. V. *Nat. Commun.* **2013**, *4* (May), 1985.
(66) Xu, G.; Chance, M. R. *Anal. Chem.* **2005**, *77* (14), 4549–4555.
(67) Hambly, D. M.; Gross, M. L. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (12), 2057–2063.
(68) Gau, B. C.; Sharp, J. S.; Rempel, D. L.; Gross, M. L. *Anal. Chem.* **2009**, *81* (16), 6563–6571.

(69) Vahidi, S.; Konermann, L. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (7), 1156–1164.
(70) Stocks, B. B.; Konermann, L. *J. Mol. Biol.* **2010**, *398* (2), 362–373.
(71) Gau, B. C.; Chen, J.; Gross, M. L. *Biochim. Biophys. Acta* **2013**, *1834* (6), 1230–1238.
(72) Pan, Y.; Brown, L.; Konermann, L. *J. Mol. Biol.* **2011**, *410* (1), 146–158.
(73) Katta, V.; Chait, B. T. *Rapid Commun. Mass Spectrom.* **1991**, *5* (4), 214–217.
(74) Lewis, H. A.; Wang, C.; Zhao, X.; Hamuro, Y.; Conners, K.; Kearins, M. C.; Lu, F.; Sauder, J. M.; Molnar, K. S.; Coales, S. J.; Maloney, P. C.; Guggino, W. B.; Wetmore, D. R.; Weber, P. C.; Hunt, J. F. *J. Mol. Biol.* **2010**, *396* (2), 406–430.
(75) Rey, M.; Forest, E.; Pelosi, L. *Biochemistry* **2012**, *51* (48), 9727–9735.
(76) Pan, Y.; Piyadasa, H.; O'Neil, J. D.; Konermann, L. *J. Mol. Biol.* **2012**, *416* (3), 400–413.
(77) Chung, K. Y.; Rasmussen, S. G. F.; Liu, T.; Li, S.; DeVree, B. T.; Chae, P. S.; Calinski, D.; Kobilka, B. K.; Woods, V. L.; Sunahara, R. K. *Nature* **2011**, *477* (7366), 611–615.
(78) Zhang, Q.; Willison, L. N.; Tripathi, P.; Sathe, S. K.; Roux, K. H.; Emmett, M. R.; Blakney, G. T.; Zhang, H. M.; Marshall, A. G. *Anal. Chem.* **2011**, *83*, 7129–7136.
(79) Houde, D.; Berkowitz, S. A.; Engen, J. R. *J. Pharm. Sci.* **2011**, *100* (6), 2071–2086.
(80) Zhang, Y.; Rempel, D. L.; Zhang, J.; Sharma, A. K.; Mirica, L. M.; Gross, M. L. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (36), 14604–14609.
(81) Fang, J.; Engen, J. R.; Beuning, P. J. *Biochemistry* **2011**, *50* (26), 5958–5968.
(82) Trcka, F.; Durech, M.; Man, P.; Hernychova, L.; Muller, P.; Vojtesek, B. *J. Biol. Chem.* **2014**, *289* (14), 9887–9901.
(83) Tuma, R.; Coward, L. U.; Kirk, M. C.; Barnes, S.; Prevelige, P. E. *J. Mol. Biol.* **2001**, *306*, 389–396.
(84) Monroe, E. B.; Kang, S.; Kyere, S. K.; Li, R.; Prevelige, P. E. *Structure* **2010**, *18* (11), 1483–1491.
(85) Hvidt, A.; Linderstrøm-Lang, K. *Biochim. Biophys. Acta* **1954**, *14* (4), 574–575.
(86) Hvidt, A.; Nielsen, S. O. *Adv. Protein Chem.* **1966**, *21*, 287–386.
(87) Roder, H.; Wüthrich, K. *Proteins Struct. Funct. Genet.* **1986**, *1* (1), 34–42.
(88) Roder, H.; Elöve, G. A.; Englander, S. W. *Nature* **1988**, *335* (6192), 700–704.
(89) Zhang, Z.; Smith, D. L. *Protein Sci.* **1993**, *2* (4), 522–531.
(90) Wales, T. E.; Engen, J. R. *Mass Spectrom. Rev.* **2006**, *25* (1), 158–170.
(91) Konermann, L.; Rodriguez, A.; Sowole, M. A. *Analyst* **2014**, *139* (23), 6078–6087.
(92) Weis, D. D.; Wales, T. E.; Engen, J. R.; Hotchko, M.; Ten Eyck, L. F. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1498–1509.
(93) Sowole, M. A.; Konermann, L. *Anal. Chem.* **2014**, *86* (13), 6715–6722.
(94) Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins* **1993**, *17* (1), 75–86.
(95) Connelly, G. P.; Bai, Y.; Jeng, M. F.; Englander, S. W. *Proteins* **1993**, *17*, 87–92.
(96) Man, P.; Montagner, C.; Vitrac, H.; Kavan, D.; Pichard, S.; Gillet, D.; Forest, E.; Forge, V. *J. Mol. Biol.* **2011**, *414* (1), 123–134.
(97) Englander, S. W. *J. Am. Soc. Mass Spectrom.* **2006**, *17* (11), 1481–1489.
(98) Hvidt, A. *C. R. Trav. Lab. Carlsberg* **1964**, *34*, 299–317.
(99) Rand, K. D.; Zehl, M.; Jørgensen, T. J. D. *Acc. Chem. Res.* **2014**, *47* (10), 3018–3027.
(100) Abzalimov, R. R.; Bobst, C. E.; Kaltashov, I. A. *Anal. Chem.* **2013**, *85* (19), 9173–9180.
(101) Ghaemmaghami, S.; Fitzgerald, M. C.; Oas, T. G. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 8296–8301.
(102) Roulhac, P. L.; Powell, K. D.; Dhungana, S.; Weaver, K. D.; Mietzner, T. a.; Crumbliss, A. L.; Fitzgerald, M. C. *Biochemistry* **2004**, *43*, 15767–15774.
(103) Khanal, A.; Pan, Y.; Brown, L. S.; Konermann, L. *J. Mass Spectrom.* **2012**, *47* (12), 1620–1626.
(104) Wang, L.; Pan, H.; Smith, D. L. *Mol. Cell. Proteomics* **2002**, *1* (2), 132–138.
(105) Pan, J.; Han, J.; Borchers, C. H.; Konermann, L. *J. Am. Chem. Soc.* **2009**, *131* (35), 12801–12808.
(106) Abzalimov, R. R.; Kaplan, D. A.; Easterling, M. L.; Kaltashov, I. A. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (8), 1514–1517.
(107) Pan, J.; Zhang, S.; Chou, A.; Hardie, D. B.; Borchers, C. H. *Anal. Chem.* **2015**, *87* (12),

5884–5890.
(108) Rand, K. D.; Zehl, M.; Jensen, O. N.; Jørgensen, T. J. D. *Anal. Chem.* **2009**, *81* (14), 5577–5584.
(109) Huang, R. Y. C.; Garai, K.; Frieden, C.; Gross, M. L. *Biochemistry* **2011**, *50* (43), 9273–9282.
(110) Landgraf, R. R.; Chalmers, M. J.; Griffin, P. R. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (2), 301–309.
(111) Zehl, M.; Rand, K. D.; Jensen, O. N.; Jørgensen, T. J. D. *J. Am. Chem. Soc.* **2008**, *130* (51), 17453–17459.
(112) Mayne, L.; Kan, Z.-Y.; Chetty, P. S.; Ricciuti, A.; Walters, B. T.; Englander, S. W. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (11), 1898–1905.
(113) Althaus, E.; Canzar, S.; Ehrler, C.; Emmett, M. R.; Karrenbauer, A.; Marshall, A. G.; Meyer-Bäse, A.; Tipton, J. D.; Zhang, H.-M. *BMC Bioinformatics* **2010**, *11*, 424.
(114) Cravello, L.; Lascoux, D.; Forest, E. *Rapid Commun. Mass Spectrom.* **2003**, *17* (21), 2387–2393.
(115) Rey, M.; Man, P.; Brandolin, G.; Forest, E.; Pelosi, L. *Rapid Commun. Mass Spectrom.* **2009**, *23* (21), 3431–3438.
(116) Zhang, H.-M.; Kazazic, S.; Schaub, T. M.; Tipton, J. D.; Emmett, M. R.; Marshall, A. G. *Anal. Chem.* **2008**, *80* (23), 9034–9041.
(117) Mazon, H.; Marcillat, O.; Forest, E.; Vial, C. *Biochimie* **2005**, *87* (12), 1101–1110.
(118) Marcoux, J.; Thierry, E.; Vives, C.; Signor, L.; Fieschi, F.; Forest, E. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (1), 76–79.
(119) Brier, S.; Maria, G.; Carginale, V.; Capasso, A.; Wu, Y.; Taylor, R. M.; Borotto, N. B.; Capasso, C.; Engen, J. R. *FEBS J.* **2007**, *274* (23), 6152–6166.
(120) Ahn, J.; Cao, M.-J.; Yu, Y. Q.; Engen, J. R. *Biochim. Biophys. Acta* **2013**, *1834* (6), 1222–1229.
(121) Rey, M.; Yang, M.; Burns, K. M.; Yu, Y.; Lees-Miller, S. P.; Schriemer, D. C. *Mol. Cell. Proteomics* **2013**, *12* (2), 464–472.
(122) Hooker, J. D. *Nature* **1874**, *10* (253), 366–372.
(123) Frazier, C. K. *Carniv. Plant Newsl.* **2000**, *29*, 56–61.
(124) Vines, S. H. *Ann. Bot.* **1897**, *os-11* (4), 563–584.
(125) Nakayama, S.; Amagase, S. *Proc. Japanese Acad.* **1968**, *44* (5), 358–362.
(126) Takahashi, K.; Athauda, S. B. P.; Matsumoto, K.; Rajapakshe, S.; Kuribayashi, M.; Kojima, M.; Kubomura-Yoshida, N.; Iwamatsu, A.; Shibata, C.; Inoue, H. *Curr. Protein Pept. Sci.* **2005**, *6* (6), 513–525.
(127) Kubota, K.; Metoki, Y.; Athauda, S. B. P.; Shibata, C.; Takahashi, K. *Biosci Biotechnol Biochem* **2010**, *74* (11), 2323–2326.
(128) Zamocky, M.; Ludwig, R.; Peterbauer, C.; Hallberg, B. M.; Divne, C.; Nicholls, P.; Haltrich, D. *Curr. Protein Pept. Sci.* **2006**, *7* (3), 255–280.
(129) Ludwig, R.; Harreither, W.; Tasca, F.; Gorton, L. *ChemPhysChem* **2010**, *11* (13), 2674–2697.
(130) Felice, A. K. G.; Sygmund, C.; Harreither, W.; Kittl, R.; Gorton, L.; Ludwig, R. *J. Diabetes Sci. Technol.* **2013**, *7* (3), 669–677.
(131) Fan, Z.; Wu, W.; Hildebrand, A.; Kasuga, T.; Zhang, R.; Xiong, X. *PLoS One* **2012**, *7* (2), 1–8.
(132) Higham, W. C.; Gordon-Smith, D.; Dempsey, E.; Wood, P. M. *Fed. Eur. Biochem. Soc.* **1994**, *351*, 128–132.
(133) Kracher, D.; Scheiblbrandner, S.; Felice, A. K. G.; Breslmayr, E.; Preims, M.; Ludwicka, K.; Haltrich, D.; Eijsink, V. G. H.; Ludwig, R. *Science (80-.).* **2016**, *3165* (April), 1–13.
(134) Samejima, M.; Eriksson, K. E. L. *Eur. J. Biochem.* **1992**, *207* (1), 103–107.
(135) Igarashi, K.; Momohara, I.; Nishino, T.; Samejima, M. *Biochem. J.* **2002**, *365* (Pt 2), 521–526.
(136) Schulz, C.; Ludwig, R.; Micheelsen, P. O.; Silow, M.; Toscano, M. D.; Gorton, L. *Electrochem. commun.* **2012**, *17*, 71–74.

(137) Kracher, D.; Zahma, K.; Schulz, C.; Sygmund, C.; Gorton, L.; Ludwig, R. *FEBS J.* **2015**, *282* (16), 3136–3148.
(138) Igarashi, K.; Yoshida, M.; Matsumura, H.; Nakamura, N.; Ohno, H.; Samejima, M.; Nishino, T. *FEBS J.* **2005**, *272* (11), 2869–2877.
(139) Hallberg, B. M.; Henriksson, G.; Pettersson, G.; Divne, C. *J. Mol. Biol.* **2002**, *315* (3), 421–434.
(140) Hallberg, B. M.; Bergfors, T.; Bäckbro, K.; Pettersson, G.; Henriksson, G.; Divne, C. *Structure* **2000**, *8* (1), 79–88.
(141) Tan, T.-C.; Kracher, D.; Gandini, R.; Sygmund, C.; Kittl, R.; Haltrich, D.; Hällberg, B. M.; Ludwig, R.; Divne, C. *Nat. Commun.* **2015**, *6* (May), 7542.
(142) Flentke, G. R.; Glinski, J.; Satyshur, K.; Rich, D. H. *Protein Expr Purif* **1999**, *16* (2), 213–220.
(143) Athauda, S. B. P.; Matsumoto, K.; Rajapakshe, S.; Kuribayashi, M.; Kojima, M.; Kubomura-Yoshida, N.; Iwamatsu, A.; Shibata, C.; Inoue, H.; Takahashi, K. *Biochem. J.* **2004**, *381* (Pt 1), 295–306.
(144) Anson, M. L. *J. Gen. Physiol.* **1938**, *22* (1), 79–89.
(145) Privalov, P. L.; Mateo, P. L.; Khechinashvili, N. N.; Stepanov, V. M.; Revina, L. P. *J. Mol. Biol.* **1981**, *152* (2), 445–464.
(146) Kadek, A.; Tretyachenko, V.; Mrazek, H.; Ivanova, L.; Halada, P.; Rey, M.; Schriemer, D. C.; Man, P. *Protein Expr. Purif.* **2014**, *95*, 121–128.
(147) Mansfeld, J.; Vriend, G.; Van den Burg, B.; Eijsink, V. G.; Ulbrich-Hofmann, R. *Biochemistry* **1999**, *38* (26), 8240–8245.
(148) Grazú, V.; Abian, O.; Mateo, C.; Batista-Viera, F.; Fernández-Lafuente, R.; Guisán, J. M. *Biotechnol. Bioeng.* **2005**, *90* (5), 597–605.
(149) Singh, R. K.; Tiwari, M. K.; Singh, R.; Lee, J.-K. *Int. J. Mol. Sci.* **2013**, *14* (1), 1232–1277.
(150) Keil, B. *Specificity of Proteolysis*; Springer-Verlag: New York, 1992.
(151) Yang, M.; Hoeppner, M.; Rey, M.; Kadek, A.; Man, P.; Schriemer, D. C. *Anal. Chem.* **2015**, *87* (13), 6681–6687.
(152) Rey, M.; Yang, M.; Lee, L.; Zhang, Y.; Sheff, J. G.; Sensen, C. W.; Mrazek, H.; Halada, P.; Man, P.; McCarville, J. L.; Verdu, E. F.; Schriemer, D. C. *Sci. Rep.* **2016**, *in press*.
(153) Ortiz, R.; Matsumura, H.; Tasca, F.; Zahma, K.; Samejima, M.; Igarashi, K.; Ludwig, R.; Gorton, L. *Anal. Chem.* **2012**, *84* (23), 10315–10323.
(154) Kavan, D.; Man, P. *Int. J. Mass Spectrom.* **2011**, *302* (1-3), 53–58.
(155) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (18), 10037–10041.
(156) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. *Nucleic Acids Res.* **2007**, *35* (Web Server issue), W522–W525.
(157) Harreither, W.; Nicholls, P.; Sygmund, C.; Gorton, L.; Ludwig, R. *Langmuir* **2012**, *28* (16), 6714–6723.
(158) Harreither, W.; Coman, V.; Ludwig, R.; Haltrich, D.; Gorton, L. *Electroanalysis* **2007**, *19* (2-3), 172–180.
(159) Engen, J. R. *Analyst* **2003**, *128* (6), 623–628.
(160) Hopper, J. T. S.; Oldham, N. J. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (10), 1851–1858.
(161) Hall, Z.; Politis, A.; Bush, M. F.; Smith, L. J.; Robinson, C. V. *J. Am. Chem. Soc.* **2012**, *134* (7), 3429–3438.
(162) Yin, S.; Xie, Y.; Loo, J. A. *J. Am. Soc. Mass Spectrom.* **2008**, *19* (8), 1199–1208.

PAPER I

Kadek A, Tretyachenko V, Mrazek H, Ivanova L, Halada P, Rey M, Schriemer DC & Man P

Expression and characterization of plant aspartic protease nepenthesin-1 from *Nepenthes gracilis*.

Protein Expr. Purif. 95, 121–128 (2014)

My contribution: *research performing (spectrophotometric assays – pH and temperature optima and stabilities, stability in denaturants), data collection, data analysis & interpretation, manuscript writing*

Protein Expression and Purification 95 (2014) 121–128

Contents lists available at ScienceDirect

Protein Expression and Purification

journal homepage: www.elsevier.com/locate/yprep

Expression and characterization of plant aspartic protease nepenthesin-1 from *Nepenthes gracilis*

Alan Kadek [a,b], Vyacheslav Tretyachenko [a,b], Hynek Mrazek [a], Ljubina Ivanova [a,b], Petr Halada [a], Martial Rey [c], David C. Schriemer [c], Petr Man [a,b,*]

[a] Institute of Microbiology, v.v.i., Academy of Sciences of the Czech Republic, Prague, Czech Republic
[b] Faculty of Science, Charles University in Prague, Prague, Czech Republic
[c] Department of Biochemistry & Molecular Biology, University of Calgary, Calgary, Alberta, Canada

ARTICLE INFO

Article history:
Received 7 November 2013
and in revised form 6 December 2013
Available online 21 December 2013

Keywords:
Plant aspartic protease
Nepenthesin
Protease characterization
Recombinant protein
Protein stability
Carnivorous plant

ABSTRACT

Carnivorous plants of the genus *Nepenthes* produce their own aspartic proteases, nepenthesins, to digest prey trapped in their pitchers. Nepenthesins differ significantly in sequence from other aspartic proteases in the animal or even plant kingdoms. This difference, which also brings more cysteine residues into the structure of these proteases, can be a cause of uniquely high temperature and pH stabilities of nepenthesins. Their detailed structure characterization, however, has not previously been possible due to low amounts of protease present in the pitcher fluid and also due to limited accessibility of *Nepenthes* plants. In the present study we describe a convenient way for obtaining high amounts of nepenthesin-1 from *Nepenthes gracilis* using heterologous production in *Escherichia coli*. The protein can be easily refolded *in vitro* and its characteristics are very close to those described for a natural enzyme isolated from the pitcher fluid. Similarly to the natural enzyme, recombinant nepenthesin-1 is sensitive to denaturing and reducing agents. It also has maximal activity around pH 2.5, shows unusual stability at high pH and its activity is not irreversibly inhibited even after prolonged incubation in the basic pH range. On the other hand, temperature stability of the recombinant enzyme is lower in comparison with the natural enzyme, which can be attributed to missing *N*-glycosylation in the recombinant protein.

Introduction

Aspartic proteases (APs, EC 3.4.23)[1] are a widespread group of proteolytic enzymes that can be found in viruses, bacteria, yeast, plants, fungi and animals [1,2]. According to the MEROPS database they are divided into several families, e.g., A1 – pepsin-like or A2 – retroviral [3]. The common features of pepsin-like enzymes are activity at low pH and sensitivity towards inhibition by pepstatin A. From a structural point of view, they have a single polypeptide chain with molecular weight between 32 and 38 kDa and their three-dimensional structure is composed of two lobes of a very similar fold. The active site is located in the cleft between the two lobes and is characterized by the presence of two aspartate residues, Asp32 and Asp215 (in pepsin numbering) occurring in the -Asp-Thr/Ser-Gly- triad. Aspartic proteases are expressed as zymogens capable of auto-activation in an acidic environment by cleavage of the pro-sequence [4,5].

A specific group of APs is represented by plant enzymes. Their general characteristics are very similar to their analogues in the animal kingdom, but they have some specific structural and sequence differences. The biggest subgroup of plant aspartic proteases is formed by vacuolar APs. These enzymes have a typical plant-specific insert (PSI) that is inserted into the C-terminal domain and bears similarity to saposins [6,7]. Another class of plant APs is represented by the group of nepenthesins. Nepenthesins are produced by secretory glands in the lower parts of pitchers of carnivorous plants of the genus *Nepenthes*, although they are also found elsewhere, for instance in *Drosera* and *Arabidopsis* [8]. The proteolytic activity of the pitcher fluid was discovered a long time ago, but uncertainty lingered as to the source of the activity, whether it arose from a protease produced by the plant itself or from microbial contamination of the fluid [9,10]. It has been experimentally proven that *Nepenthes* produce their own aspartic proteases, and they provide the main proteolytic activity of the fluid [11–13]. Nepenthesins are distinct from other plant APs. Their amino acid sequences differ significantly and in addition, they lack the PSI. Rather, they have a short insertion called the nepenthesin-type

* Corresponding author at: Institute of Microbiology, Academy of Sciences of the Czech Republic, Videnska 1083, Praha 4 142 20, Czech Republic. Tel.: +420 241062631.
E-mail address: pman@biomed.cas.cz (P. Man).

[1] *Abbreviations used:* AP, aspartic protease; PSI, plant-specific insert; rNep-1, recombinant nepenthesin-1 from *Nepenthes gracilis*; TCEP, tris-(2-carboxyethyl)phosphine; FT-ICR, fourier transform ion cyclotron resonance; TCA, trichloroacetic acid.

1046-5928/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.pep.2013.12.005

AP-specific insert, which contributes additional cysteines to the primary structure [14,15]. It is most probably higher number of disulfide bonds, in conjunction with putative *N*-glycosylation, which is responsible for remarkable features of nepenthesins, such as their stability over a wide range of pH and temperature. In contrast to these properties, nepenthesins are rather sensitive to denaturing and reducing agents [16]. Interesting are also cleavage properties of nepenthesins, which differ from those typical for aspartic proteases [17].

In this report we describe the expression of a recombinant form of nepenthesin-1 (rNep-1) from *Nepenthes gracilis*, as well as its biochemical and enzymatic characterization. The enzyme can be easily produced in *Escherichia coli* and a straightforward *in vitro* refolding protocol provides large amounts of folded active protease. The enzyme has enzymatic characteristics very similar to those described for the protease isolated from the pitcher fluid, although some specific differences are observed [14]. This includes a lower stability to auto-digestion under certain conditions, which is in contrast to the observations described for the glycosylated natural enzyme.

Materials and methods

All chemicals were from Sigma–Aldrich unless otherwise stated. Porcine pepsin A, used for comparison and mentioned throughout the text, was from Sigma–Aldrich (#P6887: 3200–4500 units/mg).

Preparation of nepenthesin-1 expression vector

The expression plasmid pET21a (Invitrogen, USA) encoding nepenthesin-1 from *N. gracilis* (residues 25-437, NCBI GI number 41016420, UniProt Q766C3) was prepared from pET21d vector containing full gene of nepenthesin-1 (NCBI GI number 41016420, UniProt Q766C3) kindly provided by Prof. Hideshi Inoue (Tokyo University of Pharmacy and Life Science). A set of primers was designed to amplify the nepenthesin-1 gene without the plant signal sequence. Forward primer (5′-CAT ATG ACG TCA AGA ACA GCT C-3′) contained *NdeI* restriction site (underlined), whereas reverse primer (5′-AAG CTT TCA CGA CGC ACC ACA TTG-3′) contained *HindIII* restriction site (underlined) and a stop codon (italics). The gene fragment was amplified using DeepVent DNA polymerase (New England Biolabs, USA). The PCR product (fragment of 1251 bp) was cloned into the pBSSK$^-$ vector (Invitrogen, USA) and the resulting product was amplified in *E. coli* NovaBlue (DE3) (Strategene, USA). The presence of nepenthesin-1 gene without the plant signal sequence in pBSSK$^-$ was confirmed by restriction digestion analysis and DNA sequencing. The fragment coding the nepenthesin-1 gene was subcloned into pET21a (Invitrogen, USA) using *NdeI* and *HindIII*. The resulting production vector (pET21a/Nep1) was verified by DNA sequencing.

Protein expression and renaturation

E. coli C41 (DE3) cells (Lucigen) were transformed by pET21a/Nep1. A starter culture was established overnight in 1 ml of LB medium containing ampicillin (100 μg/ml). Large-scale protein production was carried out by adding 0.5 ml of the starter culture to 500 ml of ampicillin-supplemented LB medium. Cells were grown at 37 °C with shaking at 220 rpm for 4 h. Expression of the recombinant protein was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM. After induction, cells were grown at 37 °C for 4 h and then harvested by centrifugation (6000×*g*, 10 min, 21 °C). Cells were resuspended in 20 ml of TSE buffer (50 mM Tris–HCl pH 7.4, 25% (w/v) sucrose, 1 mM EDTA, 1 mM NaN_3) supplemented with 1 mM phenyl-methylsulphonyl fluoride (PMSF) and 1 μM leupeptin. The cell suspension was subjected to four cycles of freezing (−80 °C) and thawing (40 °C). Next, 400 μl of 1 M $MgCl_2$, 3 μl of DNAse I (10 mg/ml), and 5 μl of RNAse I (10 U/ml) were added and the mixture was incubated at 21 °C for 30 min. Following the incubation, 5 short (30 s) sonication/incubation cycles performed on ice were used to lyse the cells. Soluble material was separated by centrifugation (10,000×*g*, 15 min, 4 °C) and the pellet was resuspended in 20 ml of Tris buffer with detergent (50 mM Tris–HCl pH 7.4, 100 mM NaCl, 0.5% Triton X-100 (v/v), 1 mM 2-mercaptoethanol, 1 mM NaN_3, 1 mM PMSF, 1 μM leupeptin). After incubation on ice with occasional short (30 s) sonication, the soluble content was removed by centrifugation (10,000×*g*, 15 min, 4 °C) and the pellet was washed with the same Tris buffer as in the previous step, only without the detergent. The pellet obtained after the final centrifugation (10,000×*g*, 15 min, 4 °C) represented enriched inclusion bodies and was stored at −80 °C.

Protein refolding followed a modified protocol of Flentke et al. [18]. Briefly, the inclusion bodies were solubilized in a denaturing buffer containing 8 M urea, 50 mM CAPS (*N*-cyclohexyl-3-aminopropanesulfonic acid) pH 10.5, 1 mM EDTA, 1 mM glycine, 500 mM NaCl, 300 mM 2-mercaptoethanol at a concentration of 1 mg of inclusion bodies (wet weight) per ml of buffer. After 1 h of incubation at 21 °C the solution was cleared of insoluble material by centrifugation (50,000×*g*, 30 min, 4 °C). The denatured protein solution was dialyzed against 5 volumes of 50 mM Tris–HCl (pH 11) for 1 h, the buffer was exchanged and the dialysis was repeated under the same conditions. Next, the dialysis buffer was exchanged for 50 mM Tris–HCl pH 7.5 and the dialysis proceeded for 12 h at 4 °C. The final dialysis was carried out for 24 h against PBS buffer (pH 7.5) at 4 °C. The dialyzed protein sample was cleared by centrifugation (24,000×*g*, 30 min, 4 °C) and the protein was concentrated using pressure ultrafiltration with a PLTK cellulosic ultrafiltration disk with 30 kDa cut-off (Millipore, USA). Finally, the solution was cleared by centrifugation (24,000×*g*, 30 min, 4 °C) and the protein solution was stored at 4 °C.

Protein digestion

Verification of the expressed protein and assignment of disulfide bonds was performed on peptides generated from rNep-1 by enzymatic or chemical cleavage. Nepenthesin was fragmented either in solution by cyanogen bromide (CNBr) or in-gel by enzymatic digestion. For in-gel digestion the protein was separated on 12% SDS–PAGE and stained by Coomassie Brilliant Blue G 250. Protein bands containing 10 μg of rNep-1 were excised, cut into small pieces and destained. Protein in the gel was subjected to reduction by 10 mM tris-(2-carboxyethyl)phosphine (TCEP) for 30 min at 75 °C and alkylation by 20 mM iodoacetamide for 60 min at 21 °C. Alkylation reactions were performed in 100 mM Tris-HCl buffer (pH 8.2) in the dark. Alternatively (for disulfide bond mapping), the reduction and alkylation were omitted and the procedure was modified as previously described [19]. Briefly, the gel pieces were washed with water and acetonitrile, dried and reconstituted in 50 mM 4-ethylmorpholine acetate buffer pH 8.2 with a protease (trypsin or chymotrypsin – enzyme:protein ratio 1:20 (w/w)) and, alternatively, with 300 μM cystamine (for disulfide bond mapping). For in-solution digestion, aliquots (10 μg) of nepenthesin were lyophilized and reconstituted in 50 μl of 70% trifluoroacetic acid (TFA) with 20 mg/ml CNBr. Proteolysis was carried out for 8 h in the dark. Next, the reaction mixture was dried down in a SpeedVac concentrator, reconstituted in 100 μl of methanol/water (1:1) and dried down again. Finally, the sample was reconstituted in 20% acetonitrile and split into two halves. One part was left unmodified and the other was subjected to reduction by TCEP (10 min, 75 °C). Finally all samples

were desalted on a peptide microtrap (Michrom Bioresources, USA) and analyzed by mass spectrometry.

Mass spectrometry

Matrix-assisted laser desorption/ionization combined with time of flight (MALDI-TOF) was used to measure intact protein mass. 1 µl of protein solution was mixed on a MALDI target with 1 µl of a saturated solution of sinapinic acid in 50% acetonitrile/50% 50 mM triethylammonium hydrogen carbonate buffer pH 8.0. After complete drying, the target was introduced into a MALDI-TOF mass spectrometer (Ultraflex III, Bruker Daltonics, Bremen, Germany). Positive-ion protein mass spectra were acquired in linear mode. Protein calibration mixtures I and II (Bruker Daltonics) were used to externally calibrate the spectrometer. The data were processed in mMass [20].

LC-MS and LC-MS/MS analyses were done using a capillary HPLC system (Agilent Technologies, Germany) connected directly to an ESI source of the Apex-ULTRA Qe Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer (Bruker Daltonics), equipped with a 9.4 *T* superconducting magnet. Samples were loaded on a reverse phase column (MAGIC C18 AQ, 0.2 × 150 mm, Michrom Bioresources) and separated by the following gradient: 1–10% *B* in 1 min, 10–90% *B* in 50 min, where solvent *A* was 0.2% formic acid, 2.5 % acetonitrile and 2.5% isopropanol in water and solvent *B* was 0.16% formic acid in 90% acetonitrile and 5% isopropanol in water. The flow rate was 4 µl/min. ESI-FT-ICR MS was calibrated externally using arginine clusters resulting in a mass accuracy below 2 ppm. In LC-MS/MS mode, the top three ions in each MS scan were selected for fragmentation using collision-induced dissociation in a quadrupole. LC-MS data were processed using MS Links software and LC-MS/MS data were searched by MASCOT algorithm (MatrixScience) against a Uniprot database and also a single protein database containing sequence of nepenthesin-1.

Spectrophotometric assays

All spectrophotometric measurements were done using the pedestal mode of a Nanodrop 2000c UV/VIS spectrophotometer (Thermo Scientific, USA). Protein content was determined by a commercial BCA assay kit.

Determination of proteolytic activity of rNep-1 or porcine pepsin A followed a modified protocol of Anson [21]. Briefly, in the standard assay 5 µl of 1 mg/ml protease solution were added to 25 µl of 20 mg/ml hemoglobin in 0.8 M glycine–HCl buffer pH 2.4. Reaction was allowed to proceed for 15 min at 37 °C and then stopped by adding 50 µl of 5 % trichloroacetic acid (TCA). After 5 min incubation at 37 °C the mixture was centrifuged for 10 min at 16,000×*g*. The amount of released peptides non-precipitable by TCA was used as a measure of proteolytic activity, as determined by measuring the absorbance of the supernatant at 280 nm against a blank sample.

For determination of pH effects on the activity of enzymes the standard activity assay was modified using 20 mg/ml BSA as a substrate in different buffers. Glycine–HCl buffer (100 mM) and standard McIlvaine's citrate–phosphate buffers (obtained by mixing 0.1 M citric acid and 0.2 M disodium phosphate) were used for pH 1.0 and pH 2.2–8.0 range, respectively.

To investigate the effect of pH and temperature on the stability of nepenthesin and porcine pepsin, the enzymes (1 mg/ml) were incubated at 37 °C in the different buffers or in the McIlvaine's buffer pH 3.0 at different temperatures (4, 22, 37, 55 and 75 °C). After 7 and 30 days, the residual proteolytic activity of the solutions was determined at pH 2.4 and 37 °C.

To investigate the effect of denaturing (guanidine, urea) and/or reducing (TCEP) agents on the enzyme stability, the standard activity assay was modified by pre-incubation of nepenthesin or porcine pepsin for 15 min at 37 °C in the solutions containing desired concentrations of these agents. Afterwards, equally pre-incubated solutions containing 20 mg/ml hemoglobin in 0.5 M glycine–HCl buffer with the desired concentrations of denaturing/reducing agents were used as a substrate.

Results

In order to prepare nepenthesin using recombinant expression in *E. coli*, a plasmid bearing the gene encoding nepenthesin-1 from *N. gracilis* was constructed. Expression trials with different *E. coli* production strains showed that the best results are obtained with C41 or C43 strains. We further continued with large scale expression in C41 cells. The production in LB media and the refolding were done according to the protocols described in the experimental section. During the solubilization of inclusion bodies and subsequent refolding steps, low amounts of the protein were lost either due to problems with resolubilization, or precipitation occurring during the dialysis and final concentration. This precipitate was removed by centrifugation after the last dialysis step and after the concentration to a final volume of 50 ml.

Each step of protein expression, purification of inclusion bodies and refolding was monitored by SDS–PAGE (Fig. 1A) and closing of disulfide bonds during the renaturation was followed as well. Here, samples were taken after each dialysis step and run on SDS-PAGE under reducing (addition of DTT) and non-reducing conditions (Fig. 1B). Clear differences in electrophoretic mobility point to a more compact protein structure under non-reducing conditions, indicating that disulfide bonds were formed already during the first dialysis step. Since the loading on the SDS–PAGE was not corrected for protein amount, it is evident how the protein becomes diluted during the course of dialysis. This impairs detection of

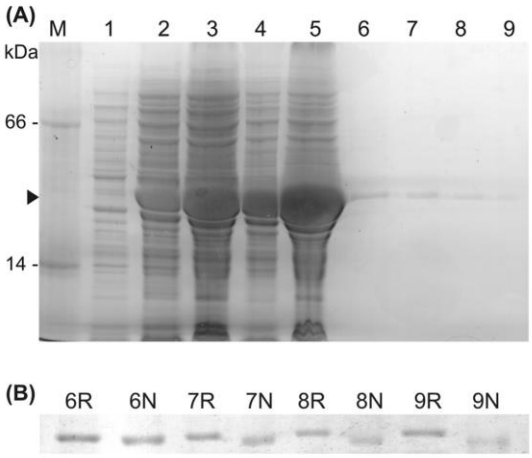


Fig. 1. SDS–PAGE analysis of nepenthesin-1 expression and refolding. (A) – *E. coli* cell culture before induction with IPTG (1) and after production (2). Process of inclusion bodies purification by saccharose buffer (3), Tris buffer with detergent (4) and detergent free Tris buffer (5). Refolding of nepenthesin-1 by dialysis against Tris–HCl pH 11 for 1 h (6), and 2 h with buffer exchange (7) and against Tris–HCl pH 7.5 for 12 h (8) and finally against PBS buffer for one day (9). (B) – Formation of disulfide bonds during refolding monitored under non-reducing (N) and reducing (R) SDS–PAGE conditions. Numbers correspond to lanes in panel A. Arrowhead indicates position of nepenthesin-1. *M* is simple molecular weight marker (66 and 14 kDa).

possible contaminations. However when the final solution was concentrated and the precipitate centrifuged, the solution contained nearly pure nepenthesin-1 (as observed by MALDI-MS). Some smaller proteins and peptides were detected but these either precipitated upon acidification (activation of the protease) or were digested by the active protease. Therefore no chromatographic steps were required and pure protease was obtained just by dialysis and activation of the enzyme. This is similar to the protocol described for another aspartic protease, rhizopuspepsin [22].

Basic characterization of the expressed protein, auto-activation and proteolytic activity

The refolded, concentrated protein was tested by SDS–PAGE for the presence of disulfide bonds, its autoactivation upon acidification and proteolytic activity (Fig. 2). SDS–PAGE of the proenzyme as well as the enzyme incubated for one day in 100 mM glycine–HCl buffer (pH 2.5) was done under reducing and non-reducing conditions (Fig. 2, lanes 1–4).

A clear shift in electrophoretic mobility upon acidification of the enzyme solution is consistent with an autolytic removal of the propeptide. In addition, proenzyme as well as mature enzyme shows distinct differences in electrophoretic mobility after treatment with a reducing agent (DTT). This confirms the presence of disulfide bonds in both forms of the enzyme. The proteolytic activity of both forms was tested using horse heart myoglobin as a substrate. Two micrograms of the proenzyme were incubated with 10 μg of myoglobin for 10 min at 37 °C either in 100 mM Tris–HCl buffer (pH 8.5) or 100 mM glycin–HCl buffer (pH 2.5). The proteolysis was stopped by raising the sample pH to 8.5 through the addition of SDS-PAGE buffer, and boiling. The result presented in Fig. 2, lane 6 showed that there was no activity at basic pH as the myoglobin remained intact. On the other hand, acidification of the proenzyme (Fig. 2, lane 7) leads to auto-activation (shift of nepenthesin band to lower molecular weight) and complete digestion of myoglobin. Mature enzyme, which was activated at pH 2.5 for one day, was also tested for its proteolytic activity at pH 8.5 and 2.5 (Fig. 2, lanes 8 and 9). While basic pH prevented any cleavage of myoglobin, the acidic conditions resulted once again in complete digestion of the myoglobin. Such behavior is typical for acid proteases and is in agreement with the known enzymatic mechanism of aspartic proteases.

Enzyme was also subjected to MALDI-TOF MS analysis to verify the mass of the protein product. Due to the ability of the protease to auto-activate upon acidification, we used a slightly basic solution of sinapinic acid as a MALDI matrix. The MALDI-TOF measurement in linear mode provided an average molecular mass of 43.73 kDa (Fig. 3), which fits to sequence 25–437 (numbering of pre-pro-nepenthesin is used throughout this paper) with predicted average mass of 43.74 kDa.

Upon acidification of the protease solution by trifluoroacetic acid (to pH 2.5) we observed very rapid auto-cleavage of the protein accompanied by disappearance of the signal at 43.73 kDa and occurrence of different signals within the range of 37–40 kDa with a dominant signal at 37.91 kDa consistent with a protein sequence 75–437 (theoretical average mass 37.92 kDa). This peak was accompanied by a smaller signal at 37.46 kDa (consistent with residues 79–437, theoretical average mass 37.47 kDa). With longer incubation times at acidic pH, the intensity ratio between these two signals shifted fully to the lower mass component after approximately one day. These observations show the mechanism of auto-activation to be a very rapid process initially, proceeding in a sequential order from protein N-terminus, but producing a slightly longer form (75–437) than expected. This form is further processed to the theoretically expected mature polypeptide (79–437) in the range of hours. These results are also in agreement with the previously published N-terminal sequence of nepenthesin-1 obtained by Edman sequencing of protein purified from pitchers of *Nepenthes distillatoria* [14].

Verification of the protein sequence and assignment of disulfide bonds

Expressed nepenthesin-1 was further characterized by mass spectrometry. In the first experiments, both proenzyme as well as mature enzyme were analyzed after enzymatic digestion (trypsin or chymotrypsin) and/or CNBr cleavage. Both protein forms fitted to nepenthesin-1 sequence and no contamination from *E. coli* was observed. In addition, N-terminal peptides were also found which justified the proposed processing mechanism. Proenzyme started with Thr25 and the mature enzyme with Asn79. Further

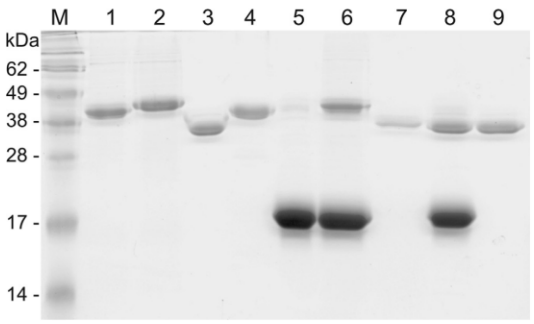


Fig. 2. SDS–PAGE analysis of disulfide bonds in nepenthesin-1, its auto-activation and activity. Proenzyme under non- (1) and reducing conditions (2). Mature enzyme under non- (3) and reducing (4) conditions. Myoglobin 10 μg (5). Digestion of myoglobin by the proenzyme at pH 8.5 (6) and pH 2.5 (7) and by the mature enzyme at pH 8.5 (8) and 2.5 (9). *M* – molecular weight marker.

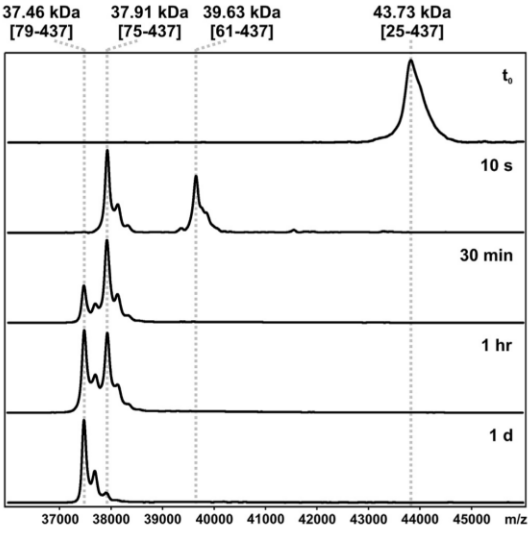


Fig. 3. Process of nepenthesin-1 activation. Solution of the proenzyme was analyzed by MALDI-TOF MS before (top, t_0) and after acidification. Aliquots were taken at different time points (indicated on the right) and spotted on MALDI target with basic solution of SA matrix to stop the reaction. Measured masses of the major peaks are indicated at the top together with amino acid limits corresponding to individual forms of nepenthesin-1.

analyses aimed at full sequence coverage and identification of disulfide bonds were done on the mature form only. Protein digested by proteases or by CNBr was analyzed by LC-MS or LC-MS/MS. Full sequence coverage was achieved with CNBr digestion. By comparing reduced and non-reduced peptides generated by CNBr, we were able to partially assign the disulfide bonds. Cysteine 240 is bonded to Cys434 and Cys354 is bonded to Cys395. The remaining cysteines (Cys123, 126, 129, 150, 155, 163, 168, 203) are all involved in disulfide bonds, but the exact linkage could not be deciphered from the CNBr digest. Further details were obtained from chymotrypsin digest of nepenthesin-1. Here peptides 202–226 (Cys203) and 121–130 (Cys123, 126, 129) were found to be linked. In addition, peptide 146–170 confirms links between Cys150, 155, 163 and 168. As we were not able to map all disulfides, we do not know if Cys203 is linked to Cys123, Cys126 or Cys129. The theoretically proposed links based on a homology model are 123–126 and 129–203 [14]. Similarly, we were not able to confirm the nested disulfide pattern Cys150–Cys168 and Cys155–Cys163, proposed again based on the model. However, we did experimentally prove that all cysteines are involved in disulfide bonds and that the disulfide pattern most probably fits to the scheme proposed by Athauda based on homology modeling.

Temperature and pH optimum for proteolytic activity

In order to compare optimal digestion conditions for rNep-1 and the prototypical aspartic protease porcine pepsin A, we tested the proteolytic activity of both mature enzymes at different pH and temperatures. When compared, our preparation of rNep-1 showed absolute activity similar to that of commercial pepsin A (Sigma–Aldrich #P6887).

Fig. 4 shows relative activities (normalized individually for each of the enzymes) as a function of pH for both the proteases towards bovine serum albumin.

As expected, both aspartic proteases are most active in the highly acidic environment. However, the pH optimum for rNep-1 is slightly higher than for pepsin, with a maximum in the pH range 2.2–3.1. In slightly acidic to neutral pH, both the proteases are effectively inactive.

Temperature-activity profiles (Fig. 5) demonstrate proteolytic activities of the enzymes towards acid-denatured hemoglobin at different temperatures.

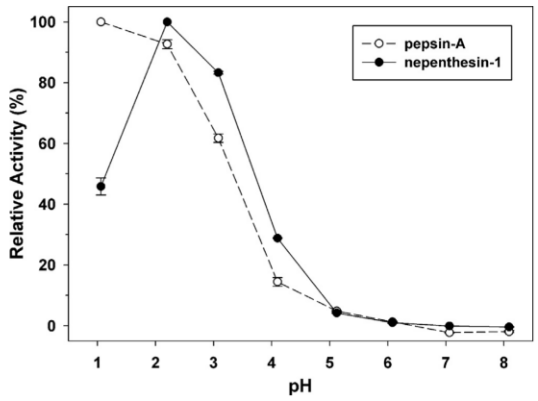


Fig. 4. Effect of pH on the activity of rNEP-1 and pepsin. Proteolytic activity towards BSA measured at 37 °C. Values are means of independent replicates with standard deviations. Relative activities (normalized individually for each of the enzymes) are shown.

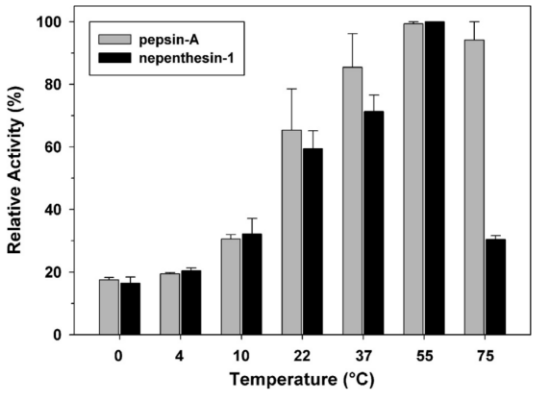


Fig. 5. Effect of temperature on the activity of rNEP-1 and pepsin. Proteolytic activity towards hemoglobin measured in pH 2.4. Values are means of independent replicates with standard deviations. Relative activities (normalized individually for each of the enzymes) are shown.

Relative activities of both the proteases are similar with a maximum at 55 °C. However, there is a marked difference in the proteolytic activity between rNep-1 and porcine pepsin at 75 °C, which may be due to a lower thermal stability of rNep-1.

Temperature and pH stability

To test for stability, we next compared resistances of porcine pepsin-A and rNep-1 in long-term pH and temperature stability assays. Fig. 6 shows residual activities of the proteases upon their incubation in solution at 37 °C and different pH values for 7 days (panel A) and 30 days (panel B).

After this incubation the activities were assayed at 37 °C and pH 2.4 against hemoglobin as a substrate. Porcine pepsin was very stable at acidic pH, with most of its activity retained up to pH 5.0 (over 80% after 30 days), where its activity is known to be low but not irreversibly inhibited [23]. At more basic pH pepsin completely lost its activity even after 7 days, whereas rNep-1 easily tolerated basic pH, but instead demonstrated a significant loss of activity at low pH and 37 °C.

The temperature stability assay (Fig. 7) shows the residual activities of the enzymes after 7 days (panel A) and 30 days (panel B) of incubation at pH 3.0 and different temperatures.

After the incubation period, the activities were once again assayed at 37 °C and pH 2.4 with the hemoglobin substrate. At lower incubation temperatures, the stability of recombinant nepenthesin clearly increased and approached that of pepsin (having retained almost 90% of activity after 30 days at 4 °C and pH 3.0). At 55 °C and higher, nepenthesin lost its activity completely, even after 7 days of incubation. This loss of activity was most probably connected with protein precipitation observed during the first few hours of incubation in nepenthesin samples at 55 and 75 °C.

Stability in denaturing and reducing agents

Nepenthesins isolated from natural sources have recently been shown to be markedly unstable in the presence of denaturing and/or reducing agents [16]. We observed similar behavior for the recombinant protease as well. Fig. 8 shows relative enzymatic activities of rNep-1 and porcine pepsin-A towards hemoglobin at pH 2.4 and 37 °C in different concentrations of guanidinium hydrochloride (A) and urea (B) alone, or in combination with 100 mM TCEP as a reducing agent (C, D).

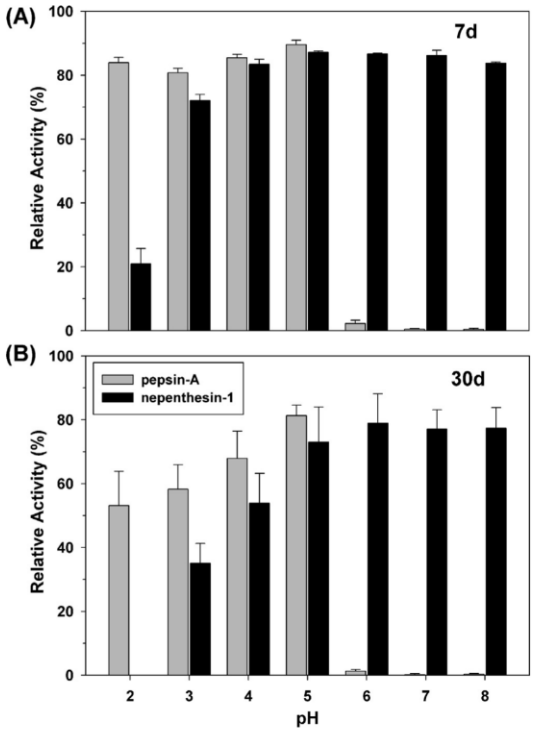


Fig. 6. Long-term pH stability of rNEP-1 and pepsin. Proteolytic activity towards hemoglobin measured at 37 °C and pH 2.4 after incubation for 7 days (A) and 30 days (B) in different pH. Values are means of independent replicates with standard deviations shown.

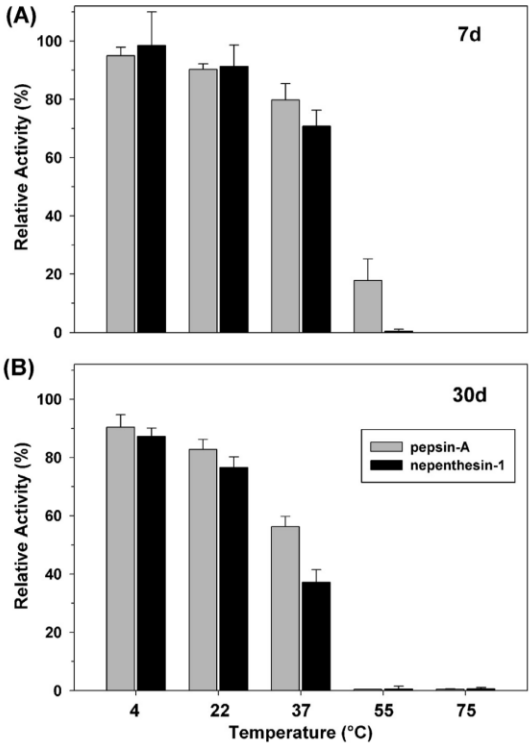


Fig. 7. Long-term temperature stability of rNEP-1 and pepsin. Proteolytic activity towards hemoglobin measured at 37 °C and pH 2.4 after incubation for 7 days (A) and 30 days (B) at different temperatures. Values are means of independent replicates with standard deviations. Relative activities (normalized individually for each of the enzymes) are shown.

In all the conditions rNep-1 was more affected by the denaturing and reducing agents than pepsin. In particular, nepenthesin rapidly lost activity with increasing concentration of guanidine, although it tolerated urea quite well up to 3 M concentration. The addition of the reducing agent TCEP greatly decreased nepenthesin's activity *per se* and also enhanced the effects of denaturants, thus demonstrating the stabilizing effect of closed disulfide bonds on the structure of nepenthesin. On the other hand, porcine pepsin tolerated both the denaturing agents well and lost its activity only at elevated levels of guanidine (>3 M). Moreover, the activity of pepsin was neither affected by 100 mM TCEP alone nor in combination with denaturing reagents.

The activity drop visible for pepsin in 0.75 and 1.5 M guanidine is probably due to effects of this denaturant on hemoglobin, which was used as a substrate for both the proteases. Indeed, we observed some hemoglobin precipitation in the course of its incubation at 37 °C with lower concentrations of guanidine. However, in 3 and 6 M guanidine hemoglobin formed a stable clear brown solution. Although the definite mechanism of guanidine protein denaturation is still not completely understood, it is expected to influence the solvation of the protein [24]. We believe it may have affected hemoglobin in a kind of salting out/salting in process. Therefore, the activity data in the low guanidine concentration region represent a convolution of protease activity and substrate accessibility, but they support a comparison of the two proteases nevertheless.

Discussions

This report is the first describing large-scale preparation of nepenthesin-1, a representative member of A1B subfamily in the MEROPS database covering nepenthesin-like aspartic proteases. Protein expression was tested in various bacterial strains but the best results were obtained with OverExpress C41 and C43 strains [25]. Protein produced into inclusion bodies can be easily purified just by washing the insoluble subcellular fraction. Surprisingly, the refolding protocol is straightforward despite the high number of disulfide bonds in the structure of the protein. No redox system (e.g., cysteamine/cystamine or reduced/oxidized glutathione) was needed and the final product has all cysteines involved in disulfide bonds. Their linkages were partially confirmed and the map fits to that described previously on the basis of homology modeling. Final purification of the refolded protein can be achieved by acidification, which leads to the cleavage of the pro-sequence, activation of the protease and digestion of all the remaining contaminants and incorrectly folded protein. The final yield is within 30–60 mg of active rNep-1 per liter of the production culture. The process of enzyme activation proceeds, under the conditions we used, in several sequential steps, but requires a long incubation before it is fully cleaved to the expected mature form (79–437).

Comparison of the enzymatic properties was done against the typical aspartic protease – porcine pepsin A and it was also possible to compare our results with the study done by Athauda et al. on

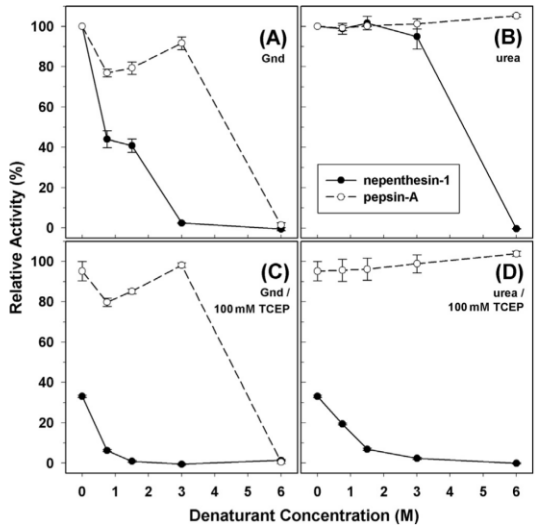


Fig. 8. Effects of denaturing and reducing agents on rNEP-1 and pepsin. Proteolytic activity towards hemoglobin measured at 37 °C and pH 2.4 in the presence of guanidine (A) and urea (B) alone or in combination with 100 mM TCEP (C, D). Values are means of independent replicates with standard deviations. Relative activities (normalized individually for each of the enzymes) are shown.

natural enzyme purified from the pitcher fluid [14]. In contrast to pepsin, rNep-1 has a pH optimum shifted slightly to higher values and this pH optimum value is similar to the ones obtained for isolated nepenthesin-1 or even the crude pitcher fluid. Quite similar are also the temperature profiles (T optimum), where the maximal activity is reached around 55 °C for both the recombinant and naturally occurring enzyme.

In terms of long-term stability, we observed some similarities as well as some differences when compared to nepenthesin-1isolated directly from plants [14]. Stability after incubation in solution of different pH (at 37 °C) showed that rNep-1 is, in contrast to pepsin, markedly stable and is not irreversibly inhibited in neutral to basic pH range. However in comparison with the isolated enzyme, recombinant nepenthesin is prone to autodigestion/degradation in the more acidic pH range. Moreover, our findings on thermal stability are in clear contrast to high temperature resistance reported earlier by Athauda et al. for plant isolates [14]. The rNep-1 was very stable at low temperatures, but in contrast to plant isolates it rapidly lost the activity at higher temperatures. At 37 °C there was a visible drop even after 7 days and this was more pronounced after 30 days. Striking differences were observed at higher temperatures (above 55 °C) where rNep-1 lost activity completely, likely due to massive protein precipitation. We attribute this thermal and low pH instability to the effect of glycosylation, which is known to be present in the nepenthesin-1 isolated directly from plants [14] and which is missing in the recombinant protein produced in *E. coli*. Glycosylation may play a crucial role in the protection of nepenthesin-1 against auto-digestion and in its stabilization similarly as was described for other proteins [26,27]. However, our temperature and pH optimum characterization results demonstrate that it is not necessary for nepenthesin's function.

Another interesting feature we observed was a much higher stability of porcine pepsin A in both the pH and temperature long-term stability assays compared with previously reported results by Athauda et al. [14]. Their pepsin was stable only at pH 5.0 for 7 days, but in all other cases its activity was nearly or completely lost. Our pepsin, on the other hand, was quite stable up to pH 5.0 and the activity was present even after one month. Although the definite cause of this discrepancy remains unclear, we speculate that the pepsin-A used in our study (Sigma–Aldrich #P6887) may be different from the one which Athauda et al. used (Sigma–Aldrich with no additional information available).

Finally, we focused on another interesting feature of nepenthesins, which is their sensitivity to denaturing and reducing agents. This phenomenon was described earlier [16] and it was shown that in contrast to pepsin, nepenthesins are much more sensitive to denaturation and reduction. Indeed, rNep-1 showed high sensitivity to guanidium hydrochloride and was more tolerant to urea, which agrees well with the results observed for isolated natural nepenthesin-1 [16]. Addition of reducing agent (TCEP) changed these profiles and caused nearly complete loss of activity under all denaturant concentrations tested. This again confirmed the presence of disulfide bonds in the molecule and proved their necessity for the stabilization and activity of the protease. However, it is in contrast to pepsin-A, which could withstand even high concentrations of both the denaturing agents and addition of TCEP did not seem to have any significant effect on its stability. This behavior might be explained by the numbers and locations of disulfide bonds involved in the stabilization of each of the two proteases. Nepenthesins are supposed to contain more disulfide bonds in the N-terminal lobe of the protein (five disulfides versus one for nepenthesin and porcine pepsin respectively) [16]. It was also shown that the thermal denaturation of pepsin starts by a melting of this N-terminal lobe, followed by structural rearrangements of the rest of the molecule [28]. Provided that nepenthesin unfolds along a similar pathway starting with an N-terminal domain, its higher dependence on the disulfide stabilization can explain the more pronounced effects of reduction on its structure.

In summary, we describe here a convenient protocol for the preparation of large amounts of active nepenthesin-1. The recombinant enzyme has enzymatic and physico-chemical properties expected for this protease [14,16]. Some of the observed differences can be attributed to the lack of glycosylation, which seem to have a protective effect but does not affect the activity of the enzyme. Considering that the amount of nepenthesin secreted by the plant into the pitcher is minute, and the access to larger amounts of plants is rather restricted, it is important that we are now able to prepare larger amounts of homogeneous nepenthesin-1. This opens the way for nepenthesin structural studies as well as basic protein folding studies, and also for usage of nepenthesin as a tool in structural proteomics research.

Acknowledgments

The authors thank prof. Hideshi Inoue and prof. Kenji Takahashi from Tokyo University of Pharmacy and Life Science for their kind gift of plasmid coding nepenthesin-1. This project was supported by Grant Agency of the Czech Republic (Grant P206/12/0503), Charles University (Project UNCE_204025/2012), Ministry of Education Youth and Sports of the Czech Republic (OPVK30 CZ.1.07/2.3.00/30.0003), Operational Program Prague – Competitiveness project CZ.2.16/3.1.00/24023 and by the Institutional Concept of the Institute of Microbiology (RVO61388971). DCS acknowledges the support of the National Sciences and Engineering Research Council of Canada (NSERC, grant 298351).

References

[1] D.R. Davies, The structure and function of the aspartic proteinases, Annu. Rev. Biophys. Biophys. Chem. 19 (1990) 189–215.
[2] B.M. Dunn, Structure and mechanism of the pepsin-like family of aspartic peptidases, Chem. Rev. 102 (2002) 4431–4458.

[3] N.D. Rawlings, A.J. Barrett, A. Bateman, MEROPS: the database of proteolytic enzymes, their substrates and inhibitors, Nucleic Acids Res. 40 (2012) D343–D350.
[4] B.M. Dunn, Splitting image, Nat. Struct. Biol. 4 (1997) 969–972.
[5] A.R. Khan, M.N. James, Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes, Protein Sci. 7 (1998) 815–836.
[6] I. Simoes, C. Faro, Structure and function of plant aspartic proteinases, Eur. J. Biochem. 271 (2004) 2067–2075.
[7] P. Runeberg-Roos, K. Törmäkangas, A. Ostman, Primary structure of a barley-grain aspartic proteinase. A plant aspartic proteinase resembling mammalian cathepsin D, Eur. J. Biochem. 202 (1991) 1021–1027.
[8] K. Takahashi, H. Niwa, N. Yokota, K. Kubota, H. Inoue, Widespread tissue expression of nepenthesin-like aspartic protease genes in *Arabidopsis thaliana*, Plant Physiol. Biochem. 46 (2008) 724–729.
[9] J.D. Hooker, The carnivorous habits of plants, Nature 10 (1874) 366–372.
[10] C.K. Frazier, The enduring controversies concerning the process of protein digestion in *Nepenthes* (Nepenthaceae), Carniv. Plant Newslett. 29 (2000) 56–61.
[11] S.H. Vines, The proteolytic enzyme of Nepenthes, Ann. Bot. os-11 (1897) 563–584.
[12] S. Nakayama, S. Amagase, Acid protease in *Nepenthes* partial purification and properties of enzyme, Proc. Jpn. Acad. 44 (1968) 358–362.
[13] S. Amagase, S. Nakayama, A. Tsugita, Acid protease in *Nepenthes*. II. Study on the specificity of nepenthesin, J. Biochem. 66 (1969) 431–439.
[14] S.B.P. Athauda, K. Matsumoto, S. Rajapakshe, M. Kuribayashi, M. Kojima, N. Kubomura-Yoshida, et al., Enzymic and structural characterization of nepenthesin, a unique member of a novel subfamily of aspartic proteinases, Biochem. J. 381 (2004) 295–306.
[15] K. Takahashi, S.B.P. Athauda, K. Matsumoto, S. Rajapakshe, M. Kuribayashi, M. Kojima, et al., Nepenthesin, a unique member of a novel subfamily of aspartic proteinases: enzymatic and structural characteristics, Curr. Protein Pept. Sci. 6 (2005) 513–525.
[16] K. Kubota, Y. Metoki, S.B.P. Athauda, C. Shibata, K. Takahashi, Stability profiles of nepenthesin in urea and guanidine hydrochloride: comparison with porcine pepsin A, Biosci. Biotechnol. Biochem. 74 (2010) 2323–2326.
[17] M. Rey, M. Yang, K.M. Burns, Y. Yu, S.P. Lees-Miller, D.C. Schriemer, Nepenthesin from monkey cups for hydrogen/deuterium exchange mass spectrometry, Mol. Cell. Proteomics 12 (2013) 464–472.
[18] G.R. Flentke, J. Glinski, K. Satyshur, D.H. Rich, Purification and crystallization of rhizopuspepsin: the use of nickel chelation chromatography to select for catalytically active species, Protein Expr. Purif. 16 (1999) 213–220.
[19] P. Pompach, P. Man, D. Kavan, K. Hofbauerova, V. Kumar, K. Bezouska, et al., Modified electrophoretic and digestion conditions allow a simplified mass spectrometric evaluation of disulfide bonds, J. Mass Spectrom. 44 (2009) 1571–1578.
[20] M. Strohalm, D. Kavan, P. Novak, M. Volny, V. Havlicek, mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data, Anal. Chem. 82 (2010) 4648–4651.
[21] M.L. Anson, The estimation of pepsin, trypsin, papain, and cathepsin with hemoglobin, J. Gen. Physiol. 22 (1938) 79–89.
[22] M. Rey, P. Man, G. Brandolin, E. Forest, L. Pelosi, Recombinant immobilized rhizopuspepsin as a new tool for protein digestion in hydrogen/deuterium exchange mass spectrometry, Rapid Commun. Mass Spectrom. 23 (2009) 3431–3438.
[23] J.H. Northrop, The influence of hydrogen ion concentration on the inactivation of pepsin solutions, J. Gen. Physiol. 2 (1920) 465–470.
[24] J.L. England, G. Haran, Role of solvation effects in protein denaturation: from thermodynamics to single molecules and back, Annu. Rev. Phys. Chem. 62 (2011) 257–277.
[25] B. Miroux, J.E. Walker, Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels, J. Mol. Biol. 260 (1996) 289–298.
[26] C. Wang, M. Eufemi, C. Turano, A. Giartosio, Influence of the carbohydrate moiety on the stability of glycoproteins, Biochemistry 35 (1996) 7299–7307.
[27] M.A. Yoshimasu, T. Tanaka, J. Ahn, R.Y. Yada, Effect of N-linked glycosylation on the aspartic proteinase porcine pepsin expressed from *Pichia pastoris*, Glycobiology 14 (2004) 417–429.
[28] P.L. Privalov, P.L. Mateo, N.N. Khechinashvili, V.M. Stepanov, L.P. Revina, Comparative thermodynamic study of pepsinogen and pepsin structure, J. Mol. Biol. 152 (1981) 445–464.

PAPER II

Kadek A, Mrazek H, Halada P, Rey M, Schriemer DC & Man P

Aspartic protease nepenthesin-1 as a tool for digestion in hydrogen / deuterium exchange mass spectrometry.

Anal. Chem. 86, 4287–94 (2014)

My contribution: *research design, research performing (protein verification by MALDI-TOF, protein immobilization, LC-MS/MS cleavage preference determination, LC-MS stability in denaturants determination), data collection, data analysis & interpretation, manuscript writing*

analytical chemistry

Article

pubs.acs.org/ac

Aspartic Protease Nepenthesin-1 as a Tool for Digestion in Hydrogen/Deuterium Exchange Mass Spectrometry

Alan Kadek,[†,‡] Hynek Mrazek,[†] Petr Halada,[†] Martial Rey,[§] David C. Schriemer,[§] and Petr Man*[,†,‡]

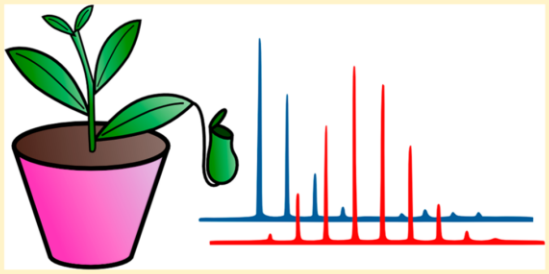
[†]Institute of Microbiology, Academy of Sciences of the Czech Republic, Prague, Czech Republic

[‡]Department of Biochemistry, Faculty of Science, Charles University in Prague, Prague, Czech Republic

[§]Department of Biochemistry & Molecular Biology, University of Calgary, Calgary, Alberta, Canada

Supporting Information

ABSTRACT: Hydrogen/deuterium exchange coupled to mass spectrometry (HXMS) utilizes enzymatic digestion of proteins to localize the information about altered exchange patterns in protein structure. The ability of the protease to produce small peptides and overlapping fragments and provide sufficient coverage of the protein sequence is essential for localizing regions of interest. Recently, it was shown that there is an interesting group of proteolytic enzymes from carnivorous pitcher plants of the genus *Nepenthes*. In this report, we describe successful immobilization and the use of one of these enzymes, nepenthesin-1, in HXMS workflow. In contrast to pepsin, it has different cleavage specificities, and despite its high inherent susceptibility to reducing and denaturing agents, it is very stable upon immobilization and withstands even high concentration of guanidine hydrochloride and reducing agents. We show that denaturing agents can alter digestion by reducing protease activity and/or substrate solubility, and additionally, they influence the trapping of proteolytic peptides onto the reversed phase resin.



Protein hydrogen/deuterium exchange coupled to mass spectrometry (HXMS) is an integral part of structural biology and is routinely used in the biopharmaceutical industry.[1−7] It can address various questions about protein structure changes under various conditions (pH, temperature, ionic strength, etc.) or interactions with different ligands (e.g., protein, DNA, small molecule).[8−11] A typical workflow involves a continuous isotope exchange reaction applied to different states of a protein and sampled at select time points. The protein is then digested, and the deuterium uptake is measured by mass spectrometry for each proteolytic fragment generated. Finally, the deuteration profiles of the individual states of the protein are compared. Despite its many advantages, HXMS suffers from low spatial resolution, which is dictated by the length of the peptides generated by the protease of choice.

Due to restrictions imposed by the quenched exchange conditions that must be maintained during the analysis (pH close to 2.5, temperature close to 0 °C), the choice of the protease is quite limited. The one most frequently used is porcine pepsin, which cleaves preferentially at the C-terminus of large hydrophobic amino acids.[12−16] Pepsin is readily available at high purity and retains functionality after immobilization, which leads to increased digestion efficiency by enhancing the local protease/protein ratio and which imparts higher resistance to denaturing and reducing agents. Therefore, in the immobilized form, pepsin can be used to digest proteins of highly compact structure and/or proteins stabilized by disulfide bonds.[17−20] However, there are many proteins for which pepsin does not provide sufficient spatial resolution and sequence coverage.[11,21]

One approach to increasing the spatial resolution of HXMS involves tandem mass spectrometry. Here, techniques based on electron-induced fragmentation are promising but currently limited to small proteins or a subset of digest peptides.[22−25] An alternative strategy is based on a computational approach requiring a large number of overlapping peptides,[26−28] provided certain conditions are met.[29] In all cases, novel proteases with different cleavage preferences would increase the number of cleavage sites, improve sequence coverage and spatial resolution, and/or process proteins not easily digested by pepsin. The first report on this topic appeared ten years ago and described the use of two aspartic proteases: aspergillo- and rhizopuspepsin (known also as protease type XIII and XVIII, respectively).[30] These proteases were shown to be complementary to pepsin, based on a higher preference for cleavage after basic amino acid residues.[31,32] Later reports described the preparation and immobilization of recombinant rhizopuspepsin and the suitability of protease type XIII for simultaneous reduction and digestion of proteins in HXMS workflow.[33,34] Several papers also identified new proteases from various

Received: December 16, 2013
Accepted: March 24, 2014
Published: March 24, 2014

sources. Among these, plasmepsin 2 and pepsin from Antarctic rock cod did not find much use in HXMS.[35,36] Two recent additions to the protease portfolio include pepsin from the stomach of the rice field eel and protease-enriched extracts from pitchers of the carnivorous plants genus *Nepenthes*.[37,38] While rice-field eel pepsin brings another possibility for protein digestion in HXMS, it does not seem to have cleavage preferences and efficiency as extraordinary as was shown for the *Nepenthes* extract. This extract appears to offer promise in extending the spatial resolution of the HXMS method, which we explore in the current study.

The nepenthesin aspartic proteases, produced by specialized cells in the lower part of the pitchers, are primarily aimed at the digestion of prey trapped by the plant. They are the only aspartic proteases described so far in the pitcher fluid and can be enriched from the crude fluid.[38] Reasonable levels of enzyme activity can be obtained despite the low amount of proteases secreted by the plant. Earlier reports described nepenthesins as remarkable aspartic proteases having slightly different sequence and structural organization than their animal (e.g., porcine pepsin A) or typical plant (vacuolar) counterparts.[39] Nepenthesins differ from the classical pepsin-like proteases in terms of a wide range pH and temperature stability. Also their long-term stability was shown to be significantly higher when compared to porcine pepsin A. These properties are thought to be enhanced by a higher number of disulfide bonds stabilizing the structure and by N-glycosylation.[39−41] The extract shows a strong preference for cleavage after basic amino acids (histidine, lysine, arginine) but, interestingly, also after proline.[38] Recently, we used recombinant protein expression to prepare large amounts of nepenthesin-1 from *N. gracilis*.[41] The biochemical and enzymatic properties of this recombinant protease nearly matched those described for the plant extract.[39,41] In this report, we focus on immobilizing the enzyme for online digestion and the use of recombinant nepenthesin-1 in the HXMS workflow. Testing of the immobilized enzyme also points to some general aspects of online protein digestion in HXMS and the use of denaturing agents, which we discuss in detail.

■ EXPERIMENTAL SECTION

Protein Expression and Immobilization. Expression and renaturation of recombinant nepenthesin-1 (rNep-1) was done as described elsewhere.[41,42] Briefly, protein was produced in *E. coli* C41 (DE3) cells (Lucigen) transformed by pET21a plasmid coding sequence 25-437 of nepenthesin-1 from *N. gracilis*. Following production, inclusion bodies were isolated, and protein was solubilized at 1 mg/mL in denaturing buffer containing 300 mM 2-mercaptoethanol. Renaturation of the protein was started by dialysis against 50 mM Tris-HCl (pH 11). In the next steps, the buffer was exchanged and pH was gradually decreased to final value of 7.5. Following renaturation, the protein solution was concentrated and acidified with 1 M glycine-HCl buffer (pH 2.3). After overnight incubation at 4 °C, the solution was clarified by centrifugation (24 000*g*, 30 min, 4 °C) and the soluble mature nepenthesin-1 was concentrated to 7 mg/mL. Purity and identity of the protein was verified by SDS-PAGE, MALDI-TOF MS, and LC-MS/MS analysis of cyanogen bromide generated peptides.[41]

Recombinant nepenthesin-1 was immobilized on an aldehyde-functionalized resin in a similar way as described earlier for porcine pepsin and rhizopuspepsin.[14,33] Briefly, activated rNep-1 was desalted and transferred into 50 mM 3-(*N*-morpholino)propanesulfonic acid (MOPS) buffer (pH 7.0) using a PD-10 gel size-exclusion chromatography column (GE Healthcare, USA). Protein solution after buffer exchange was concentrated using an Amicon Ultra-15 concentrator (10 kDa molecular weight cutoff) to reach concentration of 37 mg/mL (as determined by a bicinchoninic acid (BCA) assay kit). Six hundred microliters of the concentrated protease solution was mixed in a 15 mL Falcon tube with 90 mg of POROS-20AL resin (Applied Biosystems, USA), and in the course of 30 min, 40 μL increments of 2 M Na_2SO_4 were added to the solution to reach a final salt concentration of 1 M. Finally, 75 μL of 2 M $NaBH_3CN$ solution (highly toxic reagent) was added to the reaction to reduce the Schiff's base formed between aldehyde functions of the resin and primary amines of the protein to an amide bond. The reaction was allowed to proceed overnight at room temperature with a gentle end-over-end mixing. Finally, the reaction was quenched by the addition of 600 μL of solution containing 2 M ethanolamine/100 mM $NaBH_3CN$/400 mM Tris-Cl (pH 7.3). After 2 h of gentle mixing at room temperature, the resin with immobilized protease was centrifuged for 5 min at 1000*g* and extensively washed with 50 mM Tris-Cl buffer pH 7.3. The immobilization of porcine pepsin A (catalogue number P6887: 3200−4500 units/mg, Sigma-Aldrich, USA) followed the procedure published previously.[14] Both the resins with immobilized proteases were packed into empty stainless steel columns (2 × 20 mm, IDEX Health & Science, USA) and extensively washed by 0.4% formic acid using an LC-20AD high-performance liquid chromatography (HPLC) pump (Shimadzu, Japan).

Protein Digestion. The following model proteins (UniProt accession numbers are shown) were used in this study: myoglobin (*E. caballus*), P68082; carbonic anhydrase I (*H. sapiens*), P00915; haptoglobin (*H. sapiens*), P00738; Endo H deglycosylated recombinant cellobiose dehydrogenase (*M. thermophilum*), A9XK88, residues 22-828; recombinant diphtheria toxin T domain (*C. diphtheriae*), P00588, residues 201-386 with mutation C201S; cytochrome C (*E. caballus*), P00004; recombinant hydrogenosomal processing peptidase (*T. vaginalis*), A2D7B7 and A2ES04; and rabbit glycogen phosphorylase B (*O. cuniculus*), P00489. Protein digestion was done by porcine pepsin A (catalogue number P6887: 3200−4500 units/mg, Sigma-Aldrich, USA), recombinantly produced nepenthesin-1, or concentrated pitcher fluid from *Nepenthes* plants.

Concentrated acidified rNep-1 was screened for proteolytic activity by MALDI-TOF using myoglobin as substrate. One microliter of mixture containing 1 μg of horse myoglobin was spotted on a MALDI target with 1 μL of saturated solution of sinapinic acid (SA) in 50% acetonitrile or with 1 μL of 4-hydroxy-α-cyanocinnamic acid (HCCA). In parallel, 10 μg of horse myoglobin in 10 μL of 50 mM glycine-HCl buffer pH 2.3 was subjected to digestion by 1 μg of rNep-1 protease for 10 min at 20 °C. The digestion was stopped by increasing the pH to 8, and samples were spotted similarly to the undigested myoglobin. Except for this initial test, solution digestions were always performed under HXMS compatible conditions (0 °C, 2 min).

Preparation of protease-enriched digestive fluid from the pitcher of *Nepenthes* plants was done as described earlier.[38] Protein substrates were mixed with an amount of pitcher fluid equal to 1 μg of total protein (as assayed by the BCA method). Digestions with the pitcher fluid were done in solution at 10 °C

for 10 min and without the use of denaturing or reducing agents.

To test the in-solution activity of rNep-1 under denaturing and reducing conditions, digestions were carried out for 2 min in an ice−water bath in 0.5 M glycine-HCl buffer (pH 2.3) or in the same buffer supplemented with 2 M guanidine hydrochloride (Gnd) and 0.4 M tris-(2-carboxyethyl)-phosphine (TCEP). The substrate/enzyme ratio was 1:1 (w/w) in each case. After the incubation period, individual samples containing 300−500 pmol of a substrate protein (depending on an individual substrate protein size and the resulting peptide mixture complexity) were immediately subjected to desalting and LC-MS/MS analyses in an H/D exchange system (described below) without a protease column.

For the online digestion with immobilized protease columns, 300−500 pmol of a substrate protein was injected into the H/D exchange system with a protease column mounted before a peptide desalting trap. Again, the proteins were injected into the system either in 0.5 M glycine-HCl buffer (pH 2.3) or in the same buffer supplemented with 2 M Gnd and 0.4 M TCEP. For stability testing, myoglobin in pure 0.5 M glycine-HCl buffer (pH 2.3) was analyzed before and after 40 repeated injections of a denaturant solution containing 3 M Gnd and 0.4 M TCEP in 0.5 M glycine-HCl buffer (pH 2.3).

Liquid Chromatography Coupled to Mass Spectrometry. LC-MS/MS analyses were done under H/D exchange compatible conditions. The system used for analyses was home-built, similar to the one described by Wang et al.[14] It consisted of injection and switching valves mounted with a desalting trap column (peptide MicroTrap, Bruker-Michrom Bioresources), an analytical column (Jupiter C18, 0.5 × 50 mm, 5 μm, 300 Å, Phenomenex), and alternatively an immobilized protease column, with all the components immersed in an ice−water bath. Digestion and desalting (3 min for glycine buffer or 4 min when denaturing/reducing agents were used) were driven by a Shimadzu LC20-AD pump isocratically delivering 0.4% formic acid in water at a flow rate of 100 μL/min. Gradient separation on the analytical column was done by an HPLC system (Agilent Technologies 1200) running at a flow rate of 15 μL/min. Gradient elution from 5% B to 35% B in 30 min, followed by 1 min gradient to 95% B, was used for separation. Solvents used were A: 0.4% formic acid and 2% acetonitrile in water; and solvent B: 0.4% formic acid in 95% acetonitrile. The outlet of the analytical column was directly connected to an electrospray ionization (ESI) source of an Apex-ULTRA Qe Fourier transform ion cyclotron resonance (FTICR) mass spectrometer (Bruker Daltonics) equipped with a 9.4 T superconducting magnet. The ESI-FTICR MS was calibrated externally using arginine clusters resulting in a mass accuracy below 2 ppm. For LC-MS/MS, the instrument was operated in data-dependent mode selecting the six most intense ions in each MS scan for an MS/MS analysis using collisionally induced dissociation in a quadrupole. The data were searched by MASCOT (MatrixScience) against single protein databases containing each of the model proteins and against a database containing selected aspartic proteases (validation of possible autolysis).

The identified peptides were plotted against the substrate protein sequence using a web-based software package MSTools[43] to visualize the sequence coverage. Alternatively, to extract cleavage preferences of the proteases, a Microsoft Excel script (Supporting Information) was employed to nonredundantly extract the locations of cleavage sites for individual proteases on each of the model substrate proteins. Cleavage preferences were monitored for residues in P1 and P1′ positions of the peptides resulting from the cleavage of all the substrate proteins. Results were expressed as a probability of cleavage, normalized according to Keil[15] for the occurrence of individual amino acids in the sequences of the digested proteins.

RESULTS AND DISCUSSION

Nepenthesin-1 from *N. gracilis* (rNep-1) was recombinantly expressed in *E. coli* and refolded *in vitro* as described elsewhere.[41] The protease was activated *in vitro* by autoproteolytic cleavage upon acidification of the solution. Proteolytic activity of the resulting active rNep-1 was tested by in-solution digestion of equine myoglobin at pH 2.3 followed by a MALDI-TOF analysis (Figure 1). Panels A and B show

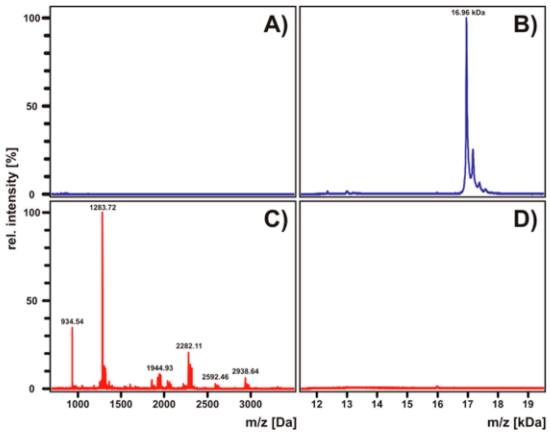


Figure 1. In-solution digestion of horse myoglobin by rNep-1 demonstrating complete digestion to peptides. Protein was spotted directly in HCCA (A) and SA (B) and after 10 min incubation with rNep-1 again in both MALDI matrices, HCCA (C) and SA (D).

peptide and protein mass spectra, respectively, of myoglobin before the digestion. After the digestion, myoglobin was completely degraded to short peptide fragments (panel C) with no intact protein remaining (panel D). This confirms that, although expressed in bacteria and thus devoid of glycosylation, rNep-1 possesses characteristics of an acidic protease and is functional under H/D exchange compatible conditions.

Next, we tested the enzyme under HXMS compatible conditions (0 °C, 2 min) and varied the substrate/enzyme ratios (w/w) ranging from 1:5 to 50:1. However, we did not observe any significant effect on the cleavage pattern, apart from the decrease in overall cleavage efficiency and in the resulting intensities of peptide MS signals as the amount of the protease decreased. Also, no improvement was observed with higher protease amounts. Therefore, we used a 1:1 substrate/enzyme ratio throughout the solution digestion experiments. Such an observation is in contrast to the substrate/enzyme ratios described earlier, where much lower amount of the protease-enriched digestive fluid was required for complete protein digestion.[38]

As some protein structures are very compact and not accessible to protease digestion without denaturation and/or reduction, digestion under such conditions is essential to render

these proteins amenable to H/D exchange studies.[14] To clearly identify the effect of these conditions on the enzyme alone, we selected myoglobin as substrate, which does not require denaturation or reduction for successful digestion. Figure 2

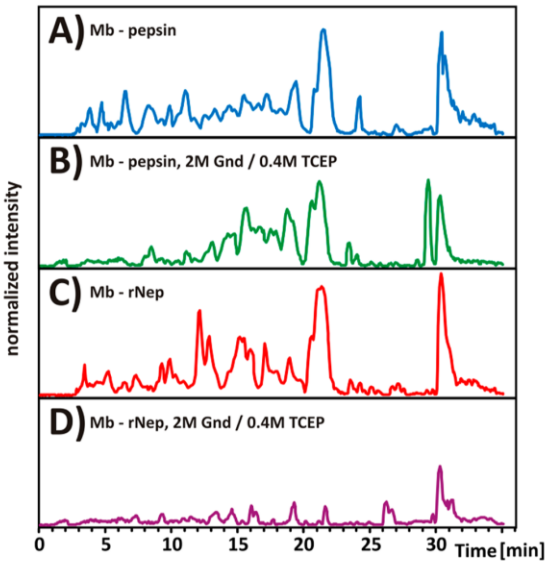


Figure 2. Effect of denaturing and reducing conditions on in-solution digestions of horse myoglobin by commercial pepsin A and recombinant nepenthesin-1. Myoglobin was digested in solution by pepsin A (A, B) or rNep-1 (C, D) either in glycine buffer (pH 2.3) (A, C) or in the same buffer supplemented with 2 M Gnd and 0.4 M TCEP (B, D). Samples were analyzed by LC-MS/MS, and normalized base-peak chromatograms were plotted to visualize the impact of denaturing and reducing agents on each protease.

shows smoothed base-peak chromatograms of peptide mixtures obtained by solution digestion of myoglobin using either pepsin or rNep-1 and normalized to a common scale (maximal absolute intensity is the same for all the graphs). Digestions performed in glycine buffer resulted in a complex peptide mixture for both pepsin and rNep-1 (panels A and C, respectively). Digestion in the presence of denaturing and reducing agents did not affect the activity of pepsin dramatically (B), although a trend toward less complete digestion accompanied by generation of longer and more hydrophobic peptides can be seen from the appearance of new chromatographic peaks at higher retention times accompanied by the disappearance of peaks at lower retention times. On the other hand, the proteolytic ability of rNep-1 was almost completely lost when denaturants were present in the solution (D). This sensitivity is in agreement with the enzymatic characteristics of nepenthesins described earlier for both our recombinantly produced protein[41] and naturally occurring enzyme isolated from plants.[40]

To overcome this intolerance to denaturants, to limit the amount of peptide background caused by autolysis, and to increase efficiency, we immobilized rNep-1 on an aldehyde-functionalized resin to produce a protease column as done with pepsin.[14] We used this column to digest myoglobin online in an H/D exchange system and extracted base-peak chromatograms from these analyses. Myoglobin was injected onto the immobilized pepsin and rNep-1 columns in a glycine buffer (Figure 3A, E) or in a buffer containing 2 M guanidine/0.4 M TCEP (Figure 3B,F). In contrast to the digestions performed in solution (see Figure 2), digestion under denaturing conditions was maintained with the immobilized rNep-1, although some changes in chromatograms were apparent for both proteases. We initially attributed these differences to lower enzymatic activity of the proteases in the presence of guanidine, but we explored them in more detail. For this purpose, we compared the digestion and chromatographic separation of myoglobin using both protease columns in the presence or absence of denaturing agents (Supplementary methods and Figure S1, Supporting Information). Denaturants were added either before or after the digestion. On

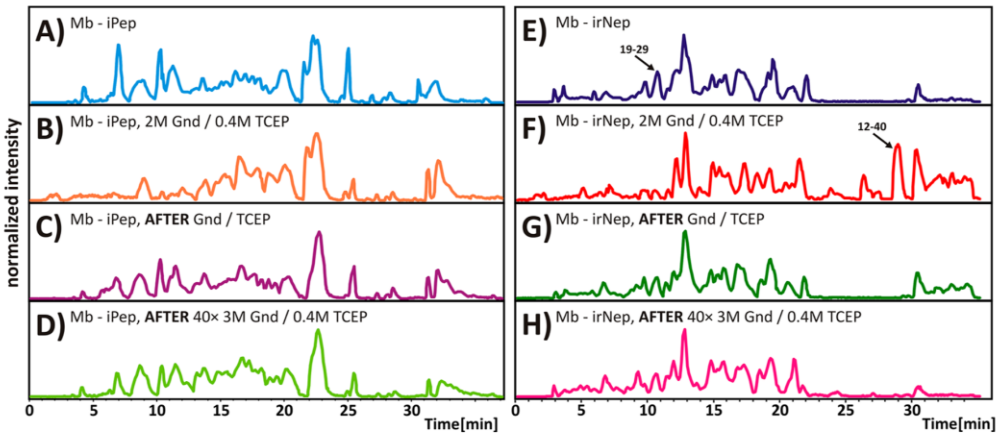


Figure 3. Effect of denaturing and reducing conditions on online digestion on columns with immobilized pepsin (A–D) and immobilized rNep-1 (E-H). Myoglobin was analyzed by LC-MS/MS after online digestion in glycine buffer (A, E), in the same buffer with 2 M Gnd, 0.4 M TCEP (B, F), in glycine buffer after denaturing/reducing run (C, G), and in glycine buffer after 40 injections of 3 M Gnd, 0.4 M TCEP (D, H). The arrows point to the example peptide which is present in normal digest (19-29) only and on the longer peptide covering this region (12-40) in denaturing/reducing digest.

the basis of these analyses, it is clear that the presence of guanidine affects not only the protease but also the chromatography, by modifying the trapping efficiency/selectivity of the reversed phase trap column. This can be demonstrated by poorer retention of small hydrophilic peptides (e.g., Figure S1A,B, Supporting Information, peptides 1-6, 8-11) and better trapping of larger hydrophobic peptides (e.g., Figure S1A,B, Supporting Information, peptide 1-11). The effect of the denaturant on protease activity can be seen through poorer digestion: lower intensity of shorter peptides and a concomitant increase in the intensity of larger peptides (e.g., Figure S1A,C, Supporting Information, peptides 33-40 and 33-69). In some cases, we observe a combination of effects: altered trapping and digestion (e.g., Figure S1A–D, Supporting Information, peptide 1-29). We observed such behavior for both protease columns tested: pepsin (Figure S1A–D, Supporting Information) and rNep-1 (Figure S1E–H, Supporting Information). However, it is clear that the digestion by nepenthesin is negatively influenced more than the digestion by pepsin (Figure S1A and C versus E and G, Supporting Information), which is in good agreement with nepenthesin's inherent susceptibility to denaturation. To further monitor the effects of denaturants on the protease columns, we also searched all the MS/MS data for possible autoproteolytic products of porcine pepsin and rNep-1. No autolysis products were observed for pepsin column regardless of the digestion conditions. On the other hand, rNep-1 column exhibited minor bleeding after exposure to 2 M guanidine/0.4 M TCEP. After a single analysis under denaturing conditions, one peptide originating from rNep-1 was identified, but after a long-term stability test consisting of 40 repeated injections of the denaturing solution (discussed below), 8 unique rNep-1 peptides were found (covering 20% of the rNep-1 protein sequence).

With this in mind, it is important to note that changes in the digestion pattern were shown to be largely reversible in nature when using the immobilized enzyme (original activity was restored upon removal of denaturing conditions). After analyzing myoglobin under denaturing conditions, we injected the next sample in the glycine buffer alone. The resulting digests (Figure 3, panels C, G) were similar to the digests before denaturant injection (A, E). In contrast to the mostly detrimental effects of denaturants on chromatography and proteolysis, we also show a different example where the denaturation of the substrate protein is crucial to achieve good sequence coverage. This is illustrated by digestion of cellobiose dehydrogenase (CDH) which is a rather large protein (807 AAs) with multiple N- and O-glycosylation sites, five disulfide bonds, and very compact fold, especially in its N-terminal domain. The data are shown in the Supporting Information (Figures S2 and S8).

Finally, we tested the long-term stability of the immobilized rNep-1 column by 40 injections of a solution containing even higher concentration of guanidine (3 M guanidine/0.4 M TCEP in glycine buffer, pH 2.3), followed by an analysis of myoglobin in the glycine buffer alone. The chromatogram traces (Figure 3D,H) showed that the activity of both the immobilized proteases was mostly unimpaired. A precise mechanism for protease stabilization by immobilization is not yet definitely known but is supposed to be the result of multiple effects including limited conformational flexibility.[44−47] We assume that one of the main factors involved in our case is, among other effects, the short time during which the protease is in contact with the quenched sample buffer. An offline solution digestion usually takes minutes to complete, but in the case of the online proteolysis using a column, the process of protein digestion is finished in tens of seconds, depending on the flow rate. We hypothesize that during this short time frame the immobilized protease does not have enough time to completely unfold, even though some structure perturbation probably occurs as seen from the impaired digestion efficiencies for immobilized proteases under denaturing and reducing conditions. However, this unfolding effect seems to be mostly reversible as demonstrated by the recovery of the activity for both immobilized rNep-1 and pepsin, when the denaturants are washed away. This may suggest some partial unfolding of the protease structure, probably due to the effects of guanidine, but we believe the disulfide bonds in the molecule remain intact, as they are vital for the stabilization of rNep-1[40,41] and their reformation in an acidic pH is not likely. The salting-out effect of guanidinium ions must also be taken into account, as at some concentration they may cause precipitation of the proteins in the reaction. This effect we observed in our previous experiments with nepenthesin-1 and porcine pepsin A, where we studied the effect of guanidine hydrochloride on the digestion using spectrophotometry.[41] It is clear that the addition of denaturing and reducing agents influences the conformational state of the substrates, and its effects on the protease itself are limited by surface-induced stabilization of the protease fold and/or by the time frame of reaction.

The slight differences in chromatograms between freshly prepared protease columns (Figure 3, panels A for pepsin and E for rNep-1) and columns after some digestion cycles (panels C and G) stress the importance of running several digestions on a column first to obtain reproducible peptides as described by Ahn et al.[37] In keeping with their results for pepsin, we have also observed that both our columns are stable for many digestions after equilibration, even in denaturing/reducing conditions (panels D and H). No significant alteration of the digestion pattern and chromatography was observed after 40 injections of cleaning solution, as was also described by Majumdar et al.[48] for pepsin digestion. This stability of immobilized rNep-1 supports regenerating the protease column with denaturing and reducing cocktails between analyses. Such behavior is beneficial because it reduces sample carry-over and increases the precision of H/D exchange results.[48] On the basis of all these findings, we conclude that the addition of denaturing and reducing agents, generally highly recommended in HXMS workflow as a way for more complete digestion, may lead to poor results if the above-mentioned effects override the benefit of denaturation and unfolding of the protein substrate.

Finally, we investigated the cleavage preferences and efficiencies of immobilized rNep-1 and compared them with immobilized porcine pepsin A. We used both protease columns to digest a panel of 7 different proteins under H/D exchange compatible conditions and analyzed the resulting peptide mixtures by LC-MS/MS. For comparison, we also used extracts directly from *Nepenthes* plants by concentrating pitcher fluids as described earlier by Rey et al.[38] However, because the amounts of enzyme obtained by this preparation method are low and the proteins are present as an impure mixture, we did not immobilize the extract.

Substrate protein sequence coverage and cleavage preferences obtained with immobilized pepsin and rNep-1 columns as well as with the pitcher fluid in solution are shown here on a model protein, equine myoglobin (Figure 4), with additional

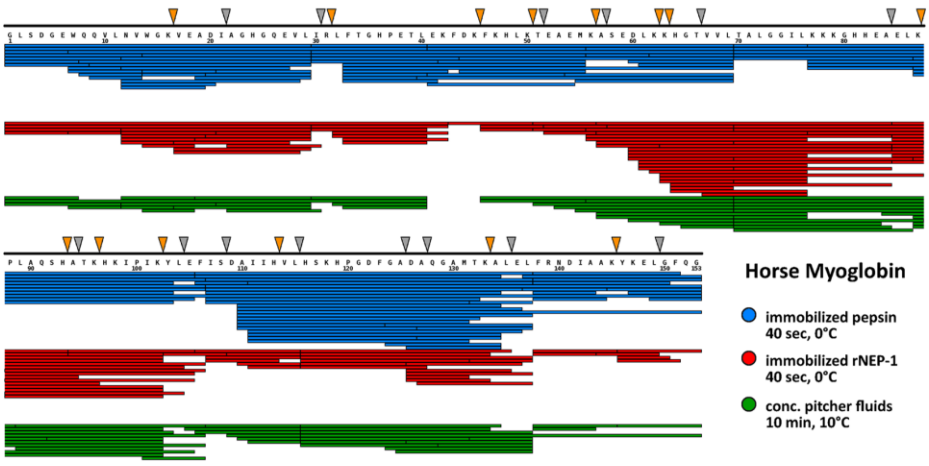


Figure 4. Myoglobin peptide map showing sequence coverage and peptides generated by immobilized pepsin (blue bars), immobilized rNep-1 (red bars), and in-solution digestion by pitcher fluid (green bars). Digestion conditions are indicated in the right lower corner. Arrowheads are pointing at pepsin-like (gray symbols) and nonpepsin (orange symbols, cleavage after basic AAs) new cleavage sites in rNep-1.

protein substrate data available in Supporting Information (Figures S3−S9). Peptides identified by LC-MS/MS analyses in individual online myoglobin digests are represented by colored bars underneath the myoglobin primary sequence. The overall cleavage pattern of the rNep-1 column is similar to that of immobilized porcine pepsin. However, rNep-1 generated several new cleavage sites that we label "pepsin-like" (marked by gray vertical arrows above the sequence). Furthermore, it also introduced a considerable number of novel cleavages (marked by orange arrows), located C-terminal to basic residues. These additional cleavage sites can help increase the spatial resolution of an H/D exchange analysis, especially in cases where they are located in regions not readily processed by pepsin (e.g., myoglobin cleavages after K16, K50, H93, or L115) or when they produce many overlapping peptides (e.g., myoglobin region 56-64). The digestion pattern generated by immobilized rNep-1 also did not differ significantly from that produced by the pitcher fluid, despite the different time and temperature of digestion.

The LC-MS/MS analyses of all selected model proteins yielded 1394 nonredundant peptides for immobilized pepsin, 1310 peptides for immobilized rNep-1, and 1319 peptides for the pitcher fluid digest. Average peptide lengths were 21.1, 14.0, and 16.2 AAs for immobilized pepsin, rNep-1, and pitcher fluid, respectively. From these data, cleavage preferences of individual preparations were extracted. Here, we took all the non-redundant cleavage sites found among the identified peptides and calculated the frequencies of occurrence of individual amino acids in P1 or P1′ positions. To obtain meaningful values, we also normalized the results for the occurrence of individual amino acids in the sequences of the substrate proteins according to a formula derived from the work of Keil:[15]

$$CS_{norm} = (\Sigma(P1_{AA})/\Sigma(P1_{tot}))/(\Sigma(seq_{AA})/\Sigma(seq_{tot}))$$

where CS_{norm} is a normalized cleavage specificity; $P1_{AA}$ is the number of occurrences of an individual amino acid in P1 position; $P1_{tot}$ is the total number of identified cleavage sites; seq_{AA} is the number of occurrences of an individual amino acid in all protein substrate sequences, and seq_{tot} is the total number of residues in all protein substrate sequences. The obtained normalized values expressing the cleavage preferences of the tested proteases are shown in Figure 5. All the proteases or protease columns exhibit somewhat similar preferences for the substrate in P1′ position with no apparent differences present (panel B). However, this is not the case for the P1 position (panel A). Here, although generally similar in behavior to porcine pepsin, immobilized rNep-1 shows lower preferences to cleave after phenylalanine and glutamic acid. Moreover, we have observed a huge drop in cleavage C-terminal to tryptophan, which contrasts nicely with pepsin. On the other hand, both immobilized rNep-1 and pitcher fluid cleave extremely efficiently after all three basic residues, lysine, arginine, and histidine. This cleavage pattern is consistent with behavior reported previously for the pitcher fluid.[38] The only striking difference, which we have observed when comparing cleavage preferences of rNep-1 with those of the pitcher fluid obtained by both us and by Rey et al.,[38] was the missing cleavage activity C-terminal to proline residues. The exact reason for this behavior remains elusive so far, although there are several possible explanations, which are currently under investigation. In our opinion, the lack of postproline activity can be either attributed to lack of glycosylation influencing the different molecular fold of native versus recombinant nepenthesin-1 protein or due to some interspecies variability of its structure. Other possible reasons may include postproline activity of nepenthesin-2 protease or the presence of some other protease(s) not described to date, in the *Nepenthes* pitcher fluid.

In summary, we have shown that novel recombinant aspartic protease nepenthesin-1 can be used in an HXMS workflow and can be efficiently immobilized. Immobilization leads to significant protease stabilization and increases its tolerance to denaturing and reducing agents. Even though this protease shows some minor differences when compared to the *Nepenthes* pitcher fluid, its ability to cleave after different residues than that of pepsin may be used to increase spatial resolution in HXMS experiments. Moreover, our experiments show that the addition of denaturing agents to HXMS compatible digestion may lead to better digestion of the substrate protein, but other

dx.doi.org/10.1021/ac404076j | *Anal. Chem.* 2014, 86, 4287−4294

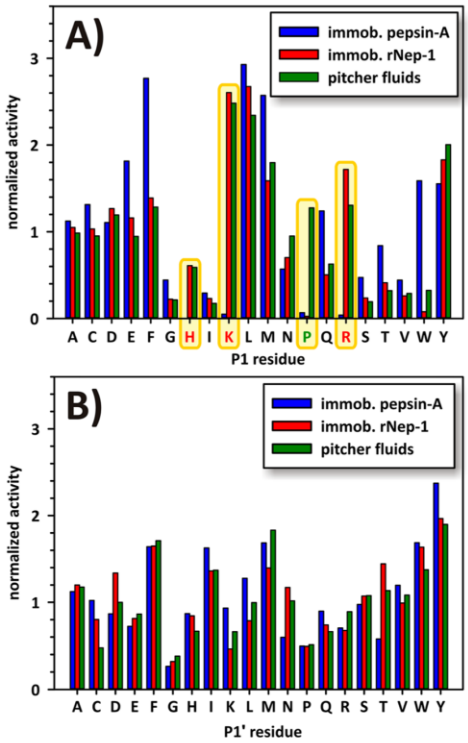


Figure 5. Plots summarizing cleavage preferences after (A) position P1 and before (B) position P1′, individual amino acids. Digestion was done using immobilized pepsin A (blue), immobilized rNep-1 (red), and pitcher fluid (green). Significant differences in P1 preferences (K, R, H, P) are highlighted. Data obtained on several model proteins were normalized according to Keil.

effects, including salting-out by the guanidinium ions, lowering protease activity, and changes to reversed phase trapping and separation, should not be overlooked. We show that these effects on the immobilized proteases are largely reversible.

■ ASSOCIATED CONTENT

Supporting Information

Additional methods and chromatogram traces describing the effects of denaturing/reducing agents on the proteolysis efficiency and chromatographic peptide trapping. Peptide coverage maps of cytochrome C, diphtheria toxin T domain, carbonic anhydrase, hydrogenosomal processing peptidase, cellobiose dehydrogenase, and glycogen phosphorylase. Microsoft Excel-based script for extraction of cleavage preferences from MS/MS-derived lists of peptides. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: pman@biomed.cas.cz.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Tomas Kucera from the Institute of Microbiology in Prague for recombinant hydrogenosomal processing peptidase and Daniel Gillet from CEA Saclay for recombinant diphtheria toxin T domain. We also thank Petra Darebna and Jan Kukla for assistance with pitcher fluid collection and preparation and Vlastimil Rybka from Prague Botanic Garden for access to the *Nepenthes* collection. This project was supported by Grant Agency of the Czech Republic (Grant P206/12/0503), Charles University (Project UNCE_204025/2012), Ministry of Education Youth and Sports of the Czech Republic (OPVK30 CZ.1.07/2.3.00/30.0003), the Institutional Concept of the Institute of Microbiology (RVO61388971), and Operational Program Prague - Competitiveness project CZ.2.16/3.1.00/24023. D.C.S. acknowledges the support of the National Sciences and Engineering Research Council of Canada (NSERC, Grant 298351).

■ REFERENCES

(1) Wei, H.; Mo, J.; Tao, L.; Russell, R. J.; Tymiak, A. A.; Chen, G.; Iacob, R. E.; Engen, J. R. *Drug Discovery Today* **2014**, *19*, 95−102.
(2) Engen, J. R. *Anal. Chem.* **2009**, *81*, 7870−7875.
(3) Marcsisin, S. R.; Engen, J. R. *Anal. Bioanal. Chem.* **2010**, *397*, 967−972.
(4) Houde, D.; Berkowitz, S. A.; Engen, J. R. *J. Pharm. Sci.* **2011**, *100*, 2071−2086.
(5) Iacob, R. E.; Engen, J. R. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1003−1010.
(6) Berkowitz, S. A.; Engen, J. R.; Mazzeo, J. R.; Jones, G. B. *Nat. Rev. Drug Discovery* **2012**, *11*, 527−540.
(7) Kaltashov, I. A.; Bobst, C. E.; Abzalimov, R. R. *Protein Sci.* **2013**, *22*, 530−544.
(8) Sperry, J. B.; Shi, X.; Rempel, D. L.; Nishimura, Y.; Akashi, S.; Gross, M. L. *Biochemistry* **2008**, *47*, 1797−1807.
(9) Hernychova, L.; Man, P.; Verma, C.; Nicholson, J.; Sharma, C.-A.; Ruckova, E.; Teo, J. Y.; Ball, K.; Vojtesek, B.; Hupp, T. R. *Proteomics* **2013**, *13*, 2512−2525.
(10) Macakova, E.; Kopecka, M.; Kukacka, Z.; Veisova, D.; Novak, P.; Man, P.; Obsil, T.; Obsilova, V. *Biochim. Biophys. Acta* **2013**, *1830*, 4491−4499.
(11) Man, P.; Montagner, C.; Vitrac, H.; Kavan, D.; Pichard, S.; Gillet, D.; Forest, E.; Forge, V. *J. Mol. Biol.* **2011**, *414*, 123−134.
(12) Zhang, Z.; Smith, D. L. *Protein Sci.* **1993**, *2*, 522−531.
(13) Ehring, H. *Anal. Biochem.* **1999**, *267*, 252−259.
(14) Wang, L.; Pan, H.; Smith, D. L. *Mol. Cell. Proteomics* **2002**, *1*, 132−138.
(15) Keil, B. *Specificity of Proteolysis*; Springer-Verlag: New York, 1992.
(16) Hamuro, Y.; Coales, S. J.; Molnar, K. S.; Tuske, S. J.; Morrow, J. A. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 1041−1046.
(17) Woods, V. L., Jr. Methods for the high-resolution identification of solvent-accessible amide hydrogens in polypeptides or proteins and for characterization of the fine structure of protein binding sites. United States Patent 6291189, 2001.
(18) Woods, V. L.; Hamuro, Y. *J. Cell. Biochem. Suppl.* **2001**, *37* (Suppl), 89−98.
(19) Rozbesky, D.; Man, P.; Kavan, D.; Chmelik, J.; Cerny, J.; Bezouska, K.; Novak, P. *Anal. Chem.* **2012**, *84*, 867−870.
(20) Chung, K. Y.; Rasmussen, S. G. F.; Liu, T.; Li, S.; DeVree, B. T.; Chae, P. S.; Calinski, D.; Kobilka, B. K.; Woods, V. L.; Sunahara, R. K. *Nature* **2011**, *477*, 611−615.
(21) Marcoux, J.; Man, P.; Petit-Haertlein, I.; Vivès, C.; Forest, E.; Fieschi, F. *J. Biol. Chem.* **2010**, *285*, 28980−28990.
(22) Kaltashov, I. A.; Eyles, S. J. *J. Mass Spectrom.* **2002**, *37*, 557−565.
(23) Abzalimov, R. R.; Kaplan, D. A.; Easterling, M. L.; Kaltashov, I. A. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1514−1517.

(24) Pan, J.; Han, J.; Borchers, C. H.; Konermann, L. *J. Am. Chem. Soc.* **2009**, *131*, 12801−12808.
(25) Zehl, M.; Rand, K. D.; Jensen, O. N.; Jørgensen, T. J. D. *J. Am. Chem. Soc.* **2008**, *130*, 17453−17459.
(26) Mayne, L.; Kan, Z.-Y.; Chetty, P. S.; Ricciuti, A.; Walters, B. T.; Englander, S. W. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1898−1905.
(27) Kan, Z.-Y.; Walters, B. T.; Mayne, L.; Englander, S. W. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 16438−16443.
(28) Althaus, E.; Canzar, S.; Ehrler, C.; Emmett, M. R.; Karrenbauer, A.; Marshall, A. G.; Meyer-Bäse, A.; Tipton, J. D.; Zhang, H.-M. *BMC Bioinf.* **2010**, *11*, 424.
(29) Sheff, J. G.; Rey, M.; Schriemer, D. C. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 1006−1015.
(30) Cravello, L.; Lascoux, D.; Forest, E. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2387−2393.
(31) Zhang, H.-M.; Kazazic, S.; Schaub, T. M.; Tipton, J. D.; Emmett, M. R.; Marshall, A. G. *Anal. Chem.* **2008**, *80*, 9034−9041.
(32) Mazon, H.; Marcillat, O.; Forest, E.; Vial, C. *Biochimie* **2005**, *87*, 1101−1110.
(33) Rey, M.; Man, P.; Brandolin, G.; Forest, E.; Pelosi, L. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 3431−3438.
(34) Zhang, H.-M.; McLoughlin, S. M.; Frausto, S. D.; Tang, H.; Emmett, M. R.; Marshall, A. G. *Anal. Chem.* **2010**, *82*, 1450−1454.
(35) Marcoux, J.; Thierry, E.; Vives, C.; Signor, L.; Fieschi, F.; Forest, E. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 76−79.
(36) Brier, S.; Maria, G.; Carginale, V.; Capasso, A.; Wu, Y.; Taylor, R. M.; Borotto, N. B.; Capasso, C.; Engen, J. R. *FEBS J.* **2007**, *274*, 6152−6166.
(37) Ahn, J.; Cao, M.-J.; Yu, Y. Q.; Engen, J. R. *Biochim. Biophys. Acta* **2013**, *1834*, 1222−1229.
(38) Rey, M.; Yang, M.; Burns, K. M.; Yu, Y.; Lees-Miller, S. P.; Schriemer, D. C. *Mol. Cell. Proteomics* **2013**, *12*, 464−472.
(39) Athauda, S. B. P.; Matsumoto, K.; Rajapakshe, S.; Kuribayashi, M.; Kojima, M.; Kubomura-Yoshida, N.; Iwamatsu, A.; Shibata, C.; Inoue, H.; Takahashi, K. *Biochem. J.* **2004**, *381*, 295−306.
(40) Kubota, K.; Metoki, Y.; Athauda, S. B. P.; Shibata, C.; Takahashi, K. *Biosci. Biotechnol. Biochem.* **2010**, *74*, 2323−2326.
(41) Kadek, A.; Tretyachenko, V.; Mrazek, H.; Ivanova, L.; Halada, P.; Rey, M.; Schriemer, D. C.; Man, P. *Protein Expression Purif.* **2014**, *95*, 121−128.
(42) Flentke, G. R.; Glinski, J.; Satyshur, K.; Rich, D. H. *Protein Expression Purif.* **1999**, *16*, 213−220.
(43) Kavan, D.; Man, P. *Int. J. Mass Spectrom.* **2011**, *302*, 53−58.
(44) Mansfeld, J.; Vriend, G.; Van den Burg, B.; Eijsink, V. G.; Ulbrich-Hofmann, R. *Biochemistry* **1999**, *38*, 8240−8245.
(45) Grazú, V.; Abian, O.; Mateo, C.; Batista-Viera, F.; Fernández-Lafuente, R.; Guisán, J. M. *Biotechnol. Bioeng.* **2005**, *90*, 597−605.
(46) Singh, R. K.; Tiwari, M. K.; Singh, R.; Lee, J.-K. *Int. J. Mol. Sci.* **2013**, *14*, 1232−1277.
(47) Cao, L. *Carrier-bound Immobilized Enzymes*; Wiley-VCH Verlag: Weinheim, Germany, 2005; pp 25−27.
(48) Majumdar, R.; Manikwar, P.; Hickey, J. M.; Arora, J.; Middaugh, C. R.; Volkin, D. B.; Weis, D. D. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 2140−2148.

PAPER III

Kadek A, Kavan D, Felice AKG, Ludwig R, Halada P & Man P

Structural insight into the calcium ion modulated interdomain electron transfer in cellobiose dehydrogenase.

FEBS Lett. 589, 1194–1199 (2015)

My contribution: *research design, research performing (enzymatic deglycosylation, optimization of digestion conditions, HXMS analyses, electrostatics computations), data collection, data analysis & interpretation, manuscript writing*

FEBS Letters 589 (2015) 1194–1199

FEBS Letters

journal homepage: www.FEBSLetters.org

Structural insight into the calcium ion modulated interdomain electron transfer in cellobiose dehydrogenase

Alan Kadek [a,b], Daniel Kavan [a,b], Alfons K.G. Felice [c], Roland Ludwig [c], Petr Halada [a], Petr Man [a,b,*]

[a] Institute of Microbiology, The Czech Academy of Sciences, Videnska 1083, Prague, Czech Republic
[b] Department of Biochemistry, Faculty of Science, Charles University in Prague, Hlavova 8, Prague, Czech Republic
[c] Department of Food Sciences and Technology, Food Biotechnology Laboratory, BOKU – University of Natural Resources and Life Sciences, Muthgasse 18, 1190 Vienna, Austria

ARTICLE INFO

Article history:
Received 29 January 2015
Revised 10 March 2015
Accepted 29 March 2015
Available online 8 April 2015

Edited by Miguel De la Rosa

Keywords:
Hydrogen/deuterium exchange
Cellobiose dehydrogenase
Calcium effect
Interdomain electron transfer
Flavocytochrome
Electrostatic interaction

ABSTRACT

Cellobiose dehydrogenase (CDH) from wood degrading fungi represents a subclass of oxidoreductases with unique properties. Consisting of two domains exhibiting interdomain electron transfer, this is the only known flavocytochrome involved in wood degradation. High resolution structures of the separated domains were solved, but the overall architecture of the intact protein and the exact interface of the two domains is unknown. Recently, it was shown that divalent cations modulate the activity of CDH and its pH optimum and a possible mechanism involving bridging of negative charges by calcium ions was proposed. Here we provide a structural explanation of this phenomenon confirming the interaction between negatively charged surface patches and calcium ions at the domain interface.

1. Introduction

Cellobiose dehydrogenase (EC 1.1.99.18, CDH) is an intriguing oxidoreductase produced and secreted by several wood degrading and phytopathogenic fungi. CDH is composed of two domains which are covalently linked by a flexible linker. At the C-terminus a flavin adenin dinucleotide (FAD) bearing dehydrogenase domain (DH) performs the oxidation of carbohydrates, e.g. cellobiose or cello-oligosaccharides. The electrons obtained during this reaction are stored on the FAD and can be either transferred to soluble electron acceptors or by interdomain electron transfer (IET) to the second domain of the enzyme – a heme *b* containing cytochrome domain (CYT). CYT can further transport electrons to cytochrome *c* (cyt *c*, an artificial substrate) or reduce lytic polysaccharide monooxygenase (LPMO, the proposed natural substrate) which in turn depolymerizes cellulose [1].

Despite the high interest in CDH, the exact structure and organization of the whole protein remains elusive. High-resolution structure of individual domains of CDH from *Phanerochaete chrysosporium* were already solved a decade ago [2,3] and based on them a possible assembly of the full length protein and the mechanisms underlying IET were drawn [4]. In this model the two domains face each other in a way which allows contact between the DH domain and the CYT domain. The crystal structure of the isolated CYT domain showed that the heme *b* propionates are facing outwards and thus are available for a close contact with FAD in the DH domain. In such orientation, the interacting surfaces are complementary and the buried surface area between the domains is rather large [4]. However, it is also known, that the IET in CDHs is pH dependent and that the electron transfer is blocked at a pH above 6 for most CDHs. Based on these facts the generally accepted view of the domain interaction is such, that at higher pH, the surfaces of both domains are negatively charged due to the deprotonation of amino acid side chains, which in turn causes

Abbreviations: CDH, cellobiose dehydrogenase; DH, dehydrogenase domain; CYT, cytochrome domain; HDX-MS, hydrogen/deuterium exchange coupled to mass spectrometry; IET, interdomain electron transfer; TCEP, tris(2-carboxyethyl)phosphine; EDTA, ethylenediaminetetraacetic acid

Author contributions: A.K. and P.M. performed HDX-MS analyses. D.K. designed and wrote the software for HDX-MS data interpretation and together with A.K. and P.M. they interpreted the data. A.K.G.F. prepared the protein. R.L., P.H., P.M. and A.K. designed the study and wrote the manuscript. D.K. and A.K.G.F. contributed to manuscript drafting.

* Corresponding author at: Institute of Microbiology, Videnska 1083, Prague 4, 142 20, Czech Republic.
E-mail address: pman@biomed.cas.cz (P. Man).

http://dx.doi.org/10.1016/j.febslet.2015.03.029
0014-5793/

electrostatic repulsion and results in the separation of the DH and CYT domain. As pH decreases, amino acid side chains become protonated and this renders the surfaces neutral. In this situation the domains can get into a close contact and IET may occur. Recently, it was found that divalent cations influence this pH dependent behavior by enhancing the enzymatic activity of CDH [5]. Interestingly, the effect of calcium differed between individual cellobiose dehydrogenases from various sources which pointed on a relation to the individual protein sequence/structure rather than to the binding of calcium by a specific site in CDH structure. A recent study explored this phenomenon in detail and screened CDHs from twelve different fungi [6]. It was shown, that any divalent alkali metal cation (regardless its atomic radius or electronegativity) at concentrations above 3 mM increases the IET of most cellobiose dehydrogenases. The divalent cation-dependent IET enhancing effect can be seen at pH 5.5 and 7.5, but is more pronounced at pH 7.5, with the most striking increase in activity observed for cellobiose dehydrogenase from *Myriococcum thermophilum* (*Mt*CDH). In addition to these observations, monovalent cations and anions were shown to have no impact, which rules out the effect of ionic strength alone. The lack of divalent cation selectivity together with the requirement for their rather high concentration (mM) and with the differences between the levels of high pH IET activation among individual CDHs pointed again on the possible elimination of negative charges by calcium cations *via* the cation bridging effect. This hypothesis was further explored by molecular modeling and domain docking and suggested an explanation by highlighting much higher number of possible divalent cation interacting residues in *Mt*CDH in contrast to other CDHs from *Phanerochaete sordida* or *Corynascus thermophilus* [6].

Hydrogen/deuterium exchange coupled to mass spectrometry (HDX-MS) is nowadays a well-established technique for fast and straightforward monitoring of protein dynamics and protein interactions [7]. It has virtually no limitation in terms of size or flexibility of the studied proteins and thus even quite complex and dynamic systems can be investigated [8–10]. Based on the measured time-resolved kinetics of backbone amide hydrogen exchange for two or more states of the protein we can identify the regions of a protein that are influenced by e.g. ligand binding. Here we used HDX-MS to provide structurally localized answer to the question: how calcium ions bind to the *Mt*CDH?

2. Materials and Methods

2.1. Materials

All chemicals were from Sigma–Aldrich unless otherwise stated. Endoglycosidase Endo Hf (1.000.000 U/mL) was purchased from New England Biolabs. The immobilization of porcine pepsin A followed the procedure described previously [11].

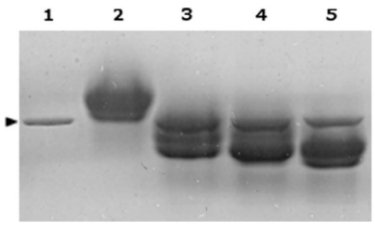


Fig. 1. Deglycosylation of *Mt*CDH by Endo Hf prior to the HDX-MS. 1 – Endo Hf; 2 – native *Mt*CDH; 3 – 5 *Mt*CDH – deglycosylation of *Mt*CDH by Endo Hf after 1 h, 4 h and overnight treatment. Position of Endo Hf is indicated by an arrowhead.

2.2. Protein preparation

Recombinant full length cellobiose dehydrogenase from *M. thermophilum* (Uniprot accession number A9XK88) was expressed in *Pichia pastoris* and purified as described previously [12]. Prior to the analyses, *Mt*CDH was deglycosylated overnight by Endo Hf under non-denaturing conditions (15 U Endo Hf/1 μg *Mt*CDH, at 37 °C in 50 mM sodium acetate buffer pH 5.5).

2.3. Hydrogen/deuterium exchange

Deglycosylated *Mt*CDH was pre-incubated for 30 min in an H_2O-based 50 mM 4-morpholinepropanesulfonic acid (MOPS) buffer pH 7.4, alone or in the presence of the studied ions. The buffers contained either 30 mM $CaCl_2$, 90 mM KCl or 9 mM EDTA-Na_2 (disodium ethylenediaminetetraacetate) to reach identical ionic strength under all the added ion conditions tested. The deuterium labeling was initiated by a 10-fold dilution of the protein into a deuterated buffer (50 mM MOPS, pD 7.4) alone or including the ions. The final *Mt*CDH concentration during the labeling was 5 μM. The exchange was left to proceed at 21 °C and aliquots (50 μL) were removed after 0.33, 1, 3, 10, 30, 60, 180 and 300 min. In these aliquots the exchange was quenched by the addition of 50 μL of a buffer containing 6 M guanidine, 0.9 M tris-(2-carboxyethyl)phosphine (TCEP) and 1 M glycine pH 2.4. The quenched mixture was incubated for 10 min on ice before being rapidly frozen in liquid nitrogen.

2.4. Digestion and liquid chromatography

Each sample was quickly thawed and injected onto an immobilized pepsin column (bed volume 66 μL). Digestion was driven by a flow of 0.4% formic acid in water at a flow rate of 100 μL/min (LC-20AD pump, Shimadzu). The resulting peptides were trapped and desalted online on a peptide microtrap (Michrom Bioresources). After a desalting step (4 min), the peptides were eluted onto a Jupiter C18 analytical column (0.5 × 5 mm, 5 μm, 300 Å, Phenomenex) and separated by a linear gradient of 10–35% B in 12 min, followed by a quick jump to 99% B, where A was 0.2% formic acid/2% acetonitrile in water and B was 95% acetonitrile/0.2% formic acid in water. The solvent was delivered at a constant flow rate of 15 μL/min (Agilent Technologies 1200). For peptide mapping of non-deuterated samples the same conditions were used. All the valves, capillaries as well as protease, desalting and analytical columns were kept at 0 °C to minimize the deuterium back-exchange.

2.5. Mass spectrometry and data analysis

The outlet of the LC system was interfaced to an electrospray ionization source of a Fourier transform ion cyclotron resonance mass spectrometer (9.4 T Apex-Qe, Bruker Daltonics). For peptide mapping (LC–MS/MS) the instrument was operated in data-dependent mode, where each MS scan was followed by up to six MS/MS collision-induced fragmentations of the most intense ions. Data were searched using MASCOT against a single protein database containing the sequence of *Mt*CDH. Identified peptides were plotted using the DrawMap script (MSTools) [13].

To determine the amount of deuterium incorporated into the peptides after the HDX, the instrument was operated in an LC–MS mode and the acquired data were processed using an in-house developed program DeutEx. The deuterium content of each peptide was reported as a percentage of maximal achievable deuteration based on the number of exchangeable amide hydrogens in each peptide.

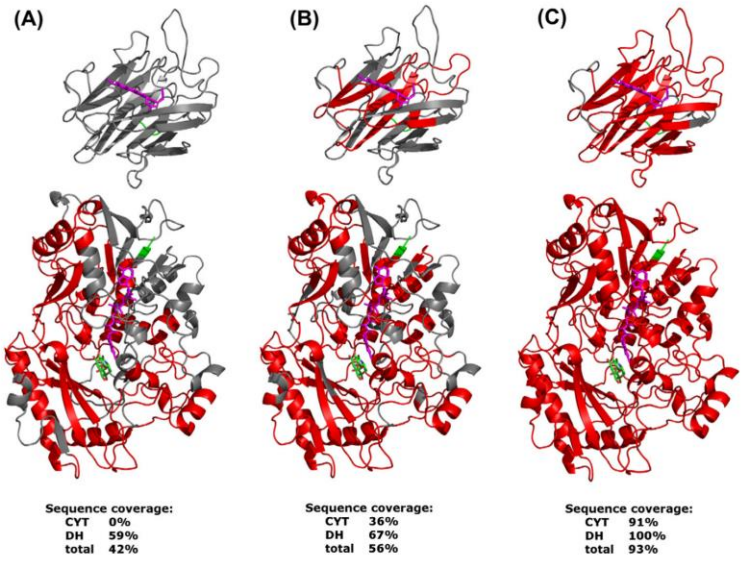


Fig. 2. Visualization of *Mt*CDH protein sequence coverage on the models of cytochrome and flavin domains. Three different digestion conditions are compared: (A) native *Mt*CDH; (B) *Mt*CDH after reduction and deglycosylation; (C) deglycosylated protein subjected to reduction and digestion in 3 M guanidine. Regions covered under individual conditions are shown in red. Green highlights sites where a flexible linker connects the two domains. The orientation of the domains is purely schematic and does not reflect their natural position. Sequence coverage for the individual conditions for each domain separately as well as for the whole protein is shown below.

2.6. Bioinformatics

Structure models of the individual domains of *Mt*CDH were prepared as described previously [6] by homology modeling using the automated Swiss Modeler [14] based on the X-ray structures of separated domains of *P. chrysosporium* CDH as templates. The surface electrostatics calculations were performed on these models using the Adaptive Poisson–Boltzmann Software (APBS) [15] with the use of PDB2PQR web server for structure preparation [16,17] and PROPKA web server for protonation assignment at chosen pH values [18,19]. All structures were visualized using PyMOL package 1.7.2 (Schrödinger) with APBS Tools 2.1 plugin.

3. Results and discussion

3.1. Optimization of HDX-MS

Cellobiose dehydrogenase from *M. thermophilum* is a large protein (86.6 kDa) and similarly to other CDHs it is *N*-glycosylated. Additionally, *O*-glycosylation is supposed to be present in its interdomain linker region and the cysteine residues in the sequence are likely involved in disulfide bonds [12,20,21]. Our experiments started with the optimization of proteolytic digestion which dictates the spatial resolution and coverage of the protein sequence. To simplify the analysis we first removed the *N*-glycans. This step was reasonable since the deglycosylation does not affect the activity of *Mt*CDH [22]. In order to remove the glycans without affecting the charge of the molecule, we used endoglycosidase Endo Hf which leaves the anchoring N-acetylglucosamine of the glycan moiety attached to the protein backbone (Fig. 1).

In the next step, we optimized protein digestion under HDX-MS conditions. First, we injected the native protein in a glycine buffer. Here we reached sequence coverage of 42% (Fig. 2A) and it became obvious that the posttranslational modifications are hampering complete sequence coverage (Fig. S1, peptides shown in red). Deglycosylation using Endo Hf improved the coverage (Fig. S1, peptides shown in yellow) but clearly, the reduction of disulfide bonds was also required. Using the Endo Hf-deglycosylated protein we optimized the reduction of the disulfide bonds using TCEP by testing various combinations of reducing agent concentration (0.1–0.5 M), temperature (0, 4 and 10 °C) and incubation time (1–10 min). The final protocol, providing the best results under HDX-MS compatible conditions, required the addition of TCEP to the final concentration of 0.45 M followed by a 10 min reduction of the protein on ice prior to the injection on the pepsin column. The result of this setup can be seen in Fig. 2B and in Fig. S1 (green dataset). Even though we captured some peptides bearing disulfides and N-glycosylation sites, the sequence coverage was still insufficient. Taking into account the stability of CDHs we also implemented denaturing agents. Out of the conditions tested (guanidine hydrochloride or urea at concentrations between 0.5 to 4 M), 3 M guanidine provided the highest sequence coverage as well as the most reproducible results. Finally, we were able to almost completely cover the *Mt*CDH sequence with the only missing part being the linker region. The flexible linker most likely bears heterogeneous *O*-glycosylation and with one exception we failed to find any peptide from this part of the protein. However, we ended the optimization reaching 93% sequence coverage (Fig. 2C and S1, blue dataset) which already constitutes a very solid basis for HDX-MS experiments.

3.2. HDX-MS analyses of the ion effects

Having optimized the digestion conditions, we proceeded to the monitoring of the structural changes of *Mt*CDH upon its activation by calcium ions at slightly alkaline pH. It has been recently reported that *Mt*CDH is activated at this pH by divalent cations [5,6]. Therefore, we monitored the protein alone (in 50 mM MOPS pH 7.4) or in the presence of 30 mM calcium ions. Calcium was chosen to be a representative of divalent alkali earth metal

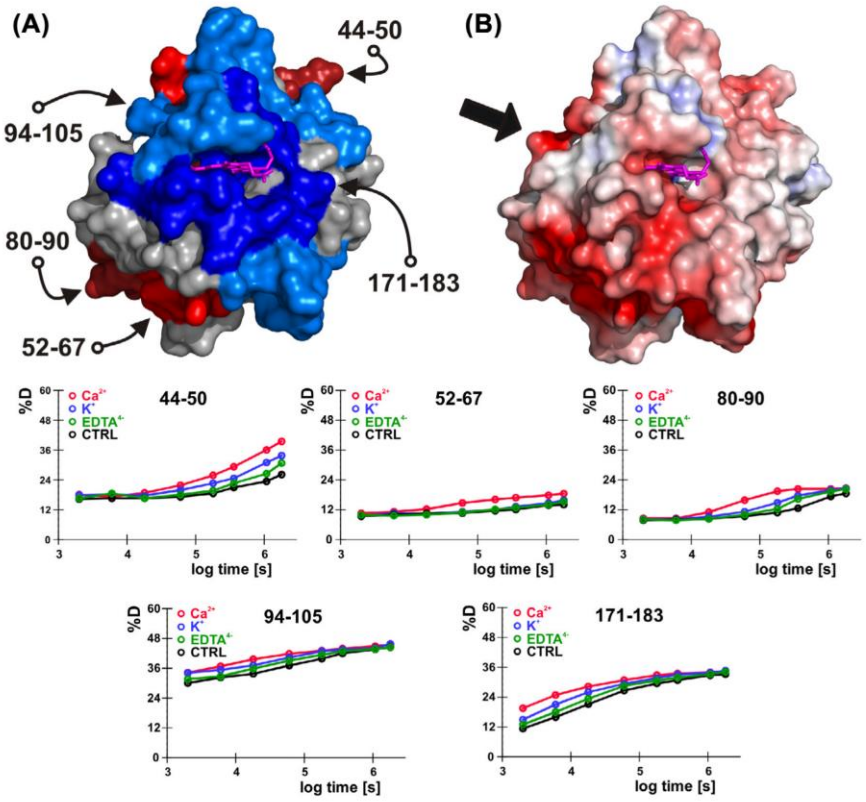


Fig. 3. Influence of calcium ions on the *Mt*CDH cytochrome domain at pH 7.4. (A) HDX-MS results (below) visualized as colored regions on the structure model of CYT domain. The heme *b* cofactor is shown in magenta. Observed changes can be attributed mainly to the ionic strength itself (blue colours; region 94–105), to the presence of divalent calcium ions (red colours; 52–67) or to the combination of the two factors (44–50, 80–90 or 171–183; coloring according to the more prominent component). (B) Calculated surface electrostatics at pH 7.4 – colored as a gradient from red to blue (−4 to +4 kT/e, respectively). Patches of negative charge correlate with regions of calcium interaction identified by the HDX-MS. Bold arrow highlights the patch of unaccounted for negative charge (partly on the opposite side of the domain) as discussed in the text.

cations as the other ions tested showed essentially the same behavior [6]. To distinguish between the role of divalent ions and the increase in ionic strength itself, a control utilizing monovalent potassium ions was also performed. Another control included the chelating agent EDTA to completely remove any residual divalent cations present in the protein preparation and buffers.

Upon the incubation with the ions, we observed several regions with modified deuteration kinetics for both parts of the *Mt*CDH – CYT domain (Fig. 3) as well as the DH domain (Fig. 4). For all these regions the presence of additional ions led to increased deuteration demonstrating a deprotection of protein amide hydrogens. Interestingly, we saw no protein backbone protection by the interaction of the two domains. We found this surprising as we were initially expecting closer interdomain contact upon addition of calcium ions and thus also protection from the exchange. However, higher protection upon interaction is not the only possible scenario in HDX-MS when two proteins interact [23]. In this case, it points toward a transient, sidechain-mediated interaction between the domains. A closed conformation of CDH featuring both domains in close contact is the prerequisite of interdomain electron transfer. The high concentration of Ca^{2+} needed to achieve IET at neutral pH suggests that the closed conformation is not strongly stabilized. It points more toward the shielding of electrostatic charges than strong binding to specific sequence motifs.

Taking a closer look at the deuteration changes, we see that the effects of ions were basically of two different types. First, we observed some parts of the protein, where only the calcium ions had an effect (CYT residues 52–67 and DH residues 257–285 or 422–433) hinting at more specific interactions requiring divalent cations. Second, we also saw regions, where solely the ionic strength was the major cause of changes (CYT 94–105 or DH 294–308), suggesting a more generic surface charge shielding effect by counter ions. Most often, however, we found both factors acting together meaning that while some changes were introduced by ionic strength and unspecific charge shielding alone, these were even more pronounced when the divalent cations were present (e.g. CYT 44–50 and 80–90 or DH 501–519).

When visualizing these results on the homology models of the two domains (Figs. 3A and 4A), it is apparent that the influenced protein regions are located around or close to the proposed domain interaction interface [4,6]. Indeed, we did not observe any important changes on the opposite sides of the two domains (Figs. S2 and S3). The protein regions where the effect of calcium is most prominent (red colored) are on both domains located to the side of the interaction interface, precisely matching to where patches of strong negative surface charge at pH 7.4 were detected by computational electrostatics simulations (Figs. 3B and 4B). These regions contain clusters of aspartic and glutamic acid residues as

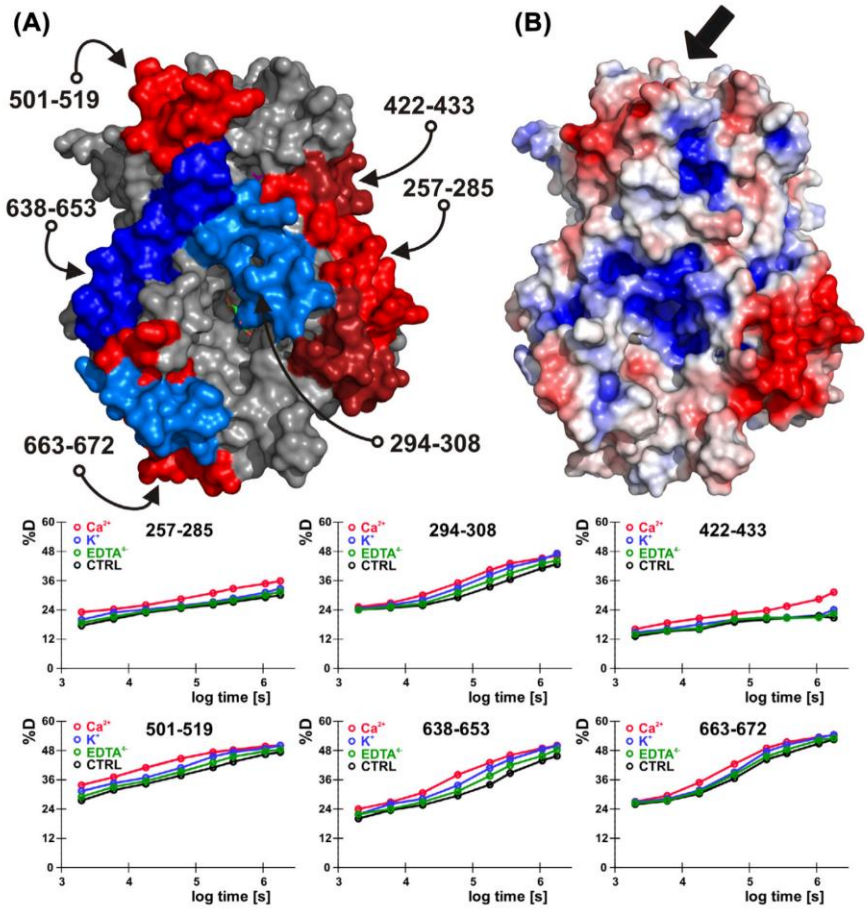


Fig. 4. Influence of calcium ions on the *Mt*CDH dehydrogenase domain at pH 7.4. (A) HDX-MS results (below) visualized as colored regions on the structure model of DH domain. Part of the cellobiose molecule is visible in the substrate entry channel close to the buried FAD cofactor. Observed changes can be attributed mainly to the ionic strength itself (blue colours; region 294–308), to the presence of divalent calcium ions (red colours; 257–285 or 422–433) or to the combination of the two factors (e.g. 638–653; coloring according to the more prominent component). (B) Calculated surface electrostatics at pH 7.4 – colored as a gradient from red to blue (−4 to +4 kT/e, respectively). Patches of negative charge correlate with regions of calcium interaction identified by the HDX-MS. Bold arrow highlights the patch of unaccounted for negative charge (completely on the opposite side of the domain) as discussed in the text.

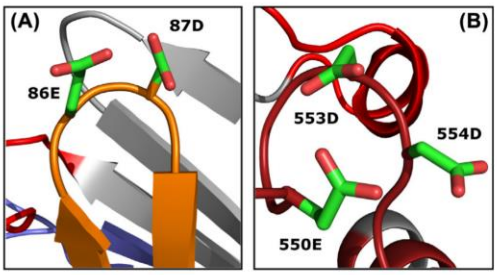


Fig. 5. Structural details of the clustered acidic amino acids. Patches of negative surface charge on both cytochrome domain (panel A, region 80–90) and dehydrogenase domain (panel B, region 549–563) of *Mt*CDH contain clustered acidic amino acids whose sterically acceptable sidechain rotamers form potential sites for calcium interaction. Coloring of the protein backbone regions remained the same as in Figs. 3 and 4.

well as their amides (Fig. 5), thus potentially forming binding pockets for the divalent ions – although probably ill-defined, resulting in the high concentrations of calcium needed to induce the effect – as also suggested previously [6].

When comparing our simulations with experimental data, we observed HDX-MS changes for all significant pockets of negative charge detected on the domain models, except for two cases. One, fairly large, unaccounted for area was located on the backside of DH and the other, smaller one, was on the side of CYT (positions indicated by bold arrows in Figs. 3B, 4B and Figs. S2, S3). We explain the first one by a shielding caused by the fungal carbohydrate binding domain (CBM1, amino acids 772–807), which is missing in the homology model, but whose connection to the rest of the domain can be predicted in the aforementioned area. The other, on the cytochrome domain, can in our opinion be explained by a probable interaction with the flexible interdomain linker, whose spatial position is as yet precisely unknown as it

has not been elucidated in any structural data, but which is most likely passing through this area.

Overall, our HDX-MS experimental data provided structural insight into the phenomenon of calcium-induced gain of function of *Mt*CDH at high pH values and support the theory of electrostatics-driven interactions between the two domains of CDH. The negative charges of the two domains, which are believed to be preventing the interdomain electron transfer in *Mt*CDH at a pH > 6.0 seem to be shielded by both increased ionic strength and in various places more specifically by divalent cations, which can bridge the two surfaces of negative charge and enable the close contacting of domains. It also seems that although some of the acidic residues forming pockets of negative charge at high pH are not located directly at the proposed interdomain interface, they are present in clusters and in high numbers, increasing the strength with which they contribute their long-distance electrostatic interactions to the overall electrostatic repulsion similarly as described for other proteins previously [24].

Acknowledgements

This project was supported by the Czech Science Foundation (Grant P206/12/0503). MS instrumentation was purchased from OPPK (CZ.2.16/3.1.00/24023). Additional funding: Charles University project UNCE_204025/2012 and European funds CZ.1.07/2.3.00/20.0055 and CZ.1.05/1.1.00/02.0109. A.K.G.F. is a recipient of a DOC-fellowship from the Austrian Academy of Sciences.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.febslet.2015.03.029.

References

[1] Agger, J.W., Isaksen, T., Várnai, A., Vidal-Melgosa, S., Willats, W.G.T., Ludwig, R., Horn, S.J., Eijsink, V.G.H. and Westereng, B. (2014) Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation. Proc. Natl. Acad. Sci. USA 111, 6287–6292.
[2] Hallberg, B.M., Henriksson, G., Pettersson, G. and Divne, C. (2002) Crystal structure of the flavoprotein domain of the extracellular flavocytochrome cellobiose dehydrogenase. J. Mol. Biol. 315, 421–434.
[3] Hallberg, B.M., Bergfors, T., Bäckbro, K., Pettersson, G., Henriksson, G. and Divne, C. (2000) A new scaffold for binding haem in the cytochrome domain of the extracellular flavocytochrome cellobiose dehydrogenase. Structure 8, 79–88.
[4] Zamocky, M., Ludwig, R., Peterbauer, C., Hallberg, B.M., Divne, C., Nicholls, P. and Haltrich, D. (2006) Cellobiose dehydrogenase–a flavocytochrome from wood-degrading, phytopathogenic and saprotropic fungi. Curr. Protein Pept. Sci. 7, 255–280.
[5] Schulz, C., Ludwig, R., Micheelsen, P.O., Silow, M., Toscano, M.D. and Gorton, L. (2012) Enhancement of enzymatic activity and catalytic current of cellobiose dehydrogenase by calcium ions. Electrochem. Commun. 17, 71–74.
[6] Kracher, D., Zahma, K., Schulz, C., Sygmund, C., Gorton, L. and Ludwig, R. (2015) Interdomain electron transfer in cellobiose dehydrogenase: modulation by pH and divalent cations. FEBS J.. in revision.
[7] Engen, J.R. (2009) Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS. Anal. Chem. 81, 7870–7875.
[8] Trcka, F., Durech, M., Man, P., Hernychova, L., Muller, P. and Vojtesek, B. (2014) The assembly and intermolecular properties of the Hsp70-Tomm34-Hsp90 molecular chaperone complex. J. Biol. Chem. 289, 9887–9901.
[9] Macakova, E., Kopecka, M., Kukacka, Z., Veisova, D., Novak, P., Man, P., Obsil, T. and Obsilova, V. (2013) Structural basis of the 14-3-3 protein-dependent activation of yeast neutral trehalase Nth1. Biochim. Biophys. Acta 1830, 4491–4499.
[10] Marcoux, J., Man, P., Petit-Haertlein, I., Vivès, C., Forest, E. and Fieschi, F. (2010) P47phox molecular activation for assembly of the neutrophil NADPH oxidase complex. J. Biol. Chem. 285, 28980–28990.
[11] Wang, L., Pan, H. and Smith, D.L. (2002) Hydrogen exchange-mass spectrometry: optimization of digestion conditions. Mol. Cell. Proteomics 1, 132–138.
[12] Zamocky, M., Schumann, C., Sygmund, C., O'Callaghan, J., Dobson, A.D., Ludwig, R., Haltrich, D. and Peterbauer, C.K. (2008) Cloning, sequence analysis and heterologous expression in *Pichia pastoris* of a gene encoding a thermostable cellobiose dehydrogenase from *Myriococcum thermophilum*. Protein Expr. Purif. 59, 258–265.
[13] Kavan, D. and Man, P. (2011) MSTools—Web based application for visualization and presentation of HXMS data. Int. J. Mass Spectrom. 302, 53–58.
[14] Arnold, K., Bordoli, L., Kopp, J. and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22, 195–201.
[15] Baker, N.A., Sept, D., Joseph, S., Holst, M.J. and McCammon, J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. USA 98, 10037–10041.
[16] Dolinsky, T.J., Nielsen, J.E., McCammon, J.A. and Baker, N.A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res. 32, W665–W667.
[17] Dolinsky, T.J., Czodrowski, P., Li, H., Nielsen, J.E., Jensen, J.H., Klebe, G. and Baker, N.A. (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res. 35, W522–W525.
[18] Olsson, M.H.M., Søndergaard, C.R., Rostkowski, M. and Jensen, J.H. (2011) PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. J. Chem. Theory Comput. 7, 525–537.
[19] Søndergaard, C.R., Olsson, M.H.M., Rostkowski, M. and Jensen, J.H. (2011) Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. J. Chem. Theory Comput. 7, 2284–2295.
[20] Górka-Nieć, W., Bańkowska, R., Palamarczyk, G., Krotkiewski, H. and Kruszewska, J.S. (2007) Protein glycosylation in pmt mutants of *Saccharomyces cerevisiae*. Influence of heterologously expressed cellobiohydrolase II of *Trichoderma reesei* and elevated levels of GDP-mannose and cis-prenyltransferase activity. Biochim. Biophys. Acta Gen. Subj. 1770, 774–780.
[21] Stapleton, P.C., O'Brien, M.M., O'Callaghan, J. and Dobson, A.D.W. (2004) Molecular cloning of the cellobiose dehydrogenase gene from *Trametes versicolor* and expression in *Pichia pastoris*. Enzyme Microb. Technol. 34, 55–63.
[22] Ortiz, R., Matsumura, H., Tasca, F., Zahma, K., Samejima, M., Igarashi, K., Ludwig, R. and Gorton, L. (2012) Effect of deglycosylation of cellobiose dehydrogenases on the enhancement of direct electron transfer with electrodes. Anal. Chem. 84, 10315–10323.
[23] Konermann, L., Rodriguez, A. and Sowole, M.A. (2014) Type 1 and type 2 scenarios in hydrogen exchange mass spectrometry studies on protein-ligand complexes. Analyst 139, 6078–6087.
[24] Joughin, B.A., Green, D.F. and Tidor, B. (2005) Action-at-a-distance interactions enhance protein binding affinity. Protein Sci. 14, 1363–1369.

PAPER IV

Kadek A, Kavan D, Marcoux J, Stojko J, Felice AKG, Cianférani S, Ludwig R, Halada P & Man P

Interdomain electron transfer in cellobiose dehydrogenase is governed by surface electrostatics.

submitted manuscript

My contribution: *research design, research performing (enzymatic deglycosylation, identification of post-translational modifications, HXMS analyses, electrostatics computations, native mass spectrometry with ion mobility), data collection, data analysis & interpretation, manuscript writing*

Interdomain electron transfer in cellobiose dehydrogenase is governed by surface electrostatics

Alan Kadek[a,b], Daniel Kavan[a,b], Julien Marcoux[c,d,#], Johann Stojko[c,d], Alfons K.G. Felice[e], Sarah Cianférani[c,d], Roland Ludwig[e], Petr Halada[a], Petr Man[a,b,*]

[a] *BioCeV - Institute of Microbiology, The Czech Academy of Sciences, Prumyslova 595, Vestec, Czech Republic*
[b] *Department of Biochemistry, Faculty of Science, Charles University in Prague, Hlavova 8, Prague, Czech Republic*
[c] *BioOrganic Mass Spectrometry Laboratory (LSMBO), IPHC, Université de Strasbourg, 25 rue Becquerel, 67087 Strasbourg, France*
[d] *IPHC, CNRS, UMR7178, 67087 Strasbourg, France*
[e] *Department of Food Sciences and Technology, BOKU - University of Natural Resources and Life Sciences, Muthgasse 18, 1190 Vienna, Austria*
[#] *Current Address: IPBS, CNRS, UMR 5089, 205 Route de Narbonne, 31077 Toulouse, France*

[*]*Corresponding author: BioCeV - Institute of Microbiology, Prumyslova 595, Vestec, 252 42, Czech Republic. E-mail: pman@biomed.cas.cz*

Keywords

flavocytochrome; direct electron transfer; electrostatic interactions; hydrogen/deuterium exchange; mass spectrometry;

Abbreviations

CDH, cellobiose dehydrogenase; DH, dehydrogenase domain; CYT, cytochrome domain; HDX-MS, hydrogen/deuterium exchange mass spectrometry; IET, interdomain electron transfer; IM-MS, ion mobility mass spectrometry

RESEARCH HIGHLIGHTS

- Cellobiose dehydrogenase and its interdomain interaction was studied in solution
- CDH is extensively O-glycosylated in the linker region
- Hydrogen/Deuterium exchange was used to describe structure changes in CDH at different pH values
- Ion mobility points at higher electrostatic repulsion between domains at neutral pH
- Charge neutralization at acidic pH enables the domain interaction in CDH

ABSTRACT
Cellobiose dehydrogenase (CDH) is a fungal extracellular oxidoreductase involved in the degradation of cellulose and hemicelluloses together with its partner enzyme lytic polysaccharide monooxygenase (LPMO). CDH provides electrons to LPMO, which are collected from CDH's active site FAD by its cytochrome domain via interdomain electron transfer. This electron transfer can be modulated by both pH and ions. In this study, we investigate recently postulated electrostatic domain repulsion model by structural mass-spectrometric methods using full-length CDH from *Myriococcum thermophilum* as well as its papain-separated individual domains. Primary structure analysis answered the long-standing questions regarding O-glycosylation in the linker domain of CDH as well as the papain cleavage site. Hydrogen/deuterium exchange mass spectrometry is then used to probe the domain-domain interaction in CDH under two different pH values, while utilizing novel combined digestion by acidic proteases nepenthesin-1 and rhizopuspepsin to increase spatial resolution. The results are further complemented with ion mobility measurements as well as with protein electrostatics simulations. Together, these techniques probe the interaction between CDH domains in its transient electrocompetent closed state and the stability of its open state conformation, which prevents interdomain electron transfer. Transition between these states is shown to be governed by changes in the protein surface electrostatics at the interdomain interface of CDH. Our study thus confirms that electrostatic repulsion between the domains is indeed the key factor modulating the functioning of CDH.

INTRODUCTION
Cellobiose dehydrogenase (CDH, cellobiose:(acceptor) 1-oxidoreductase, E.C. 1.1.99.18) is the only known extracellular flavocytochrome. It is produced by wood-degrading and phytopathogenic fungi from the *Basidiomycota* and *Ascomycota* phyla. [1] CDH participates in the early events of lignocellulose degradation, as found in several fungi at the transcriptional and translational levels. [2,3] The enzyme is also highly relevant from the biotechnological point of view as CDH forms a promising component for third generation biosensors and biofuel cells based on direct electron transfer. [4–6] Additionally, CDH is a crucial enzyme in the process of cellulose depolymerization to shorter saccharides [7,8], that are envisaged as a potential source of renewable energy replacing fossil-based fuels. [9]
From a structural point of view, CDH is a monomeric enzyme consisting of two domains joined by a flexible linker region. The larger dehydrogenase (DH) domain carries FAD as a cofactor while the smaller cytochrome (CYT) domain contains a heme *b*. The catalytic cycle of cellobiose dehydrogenase is divided into a reductive and an oxidative half-reaction. During the reductive half-reaction, β-D-cellobiose is oxidized at the anomeric C1 carbon to yield δ-lactone, which is further hydrolyzed in bulk water to the corresponding carboxylic acid. [10] Due to the higher redox potential of the heme *b* an electron from the catalytic center (FAD) can be delivered to the CYT domain during the oxidative half-reaction. The electrons are finally channeled from the heme cofactor to a terminal electron acceptor, which can be either a small molecule or a protein partner such as lytic polysaccharide monooxygenase (LPMO). [11] For such an electron flow to proceed, the key reaction is the interdomain electron transfer (IET) between FAD and heme within the CDH molecule. This transfer occurs only under acidic solution conditions as has been observed since early studies. [12] While the specific pH optima for this reaction differ between CDHs from different organisms [4,13,14], both FAD and heme are in general reduced rapidly in the presence of cellobiose at slightly acidic pH. At neutral pH only FAD reduction is fast, while the IET and the heme reduction are extremely slow. [12,15] Based on these findings a theory of CDH functioning was proposed, in which the close contact of the two domains is necessary for the electrons to be transferred directly between the cofactors. Physical separation of CDH domains on the other hand prevents the IET to occur and stops the electron flow to the interaction partner. [16]
Detailed knowledge of the structure and dynamics of CDH is a prerequisite for improving of its biotechnological potential and tailoring its enzymatic properties. Most likely due to its dynamic nature and high degree of modification, the only high resolution structures available for CDH were derived from isolated dehydrogenase [17] and cytochrome [18] domains from *Phanerochaete chrysosporium*. Until recently, these structures thus served as a modelling template for the full-length protein. However, paper describing high-resolution X-ray structures of two conformations of CDH was published in 2015.

[19] This study characterized two different full-length CDH molecules (proteins originating from *Myriococcum thermophilum* - *Mt*CDH and *Neurospora crassa* - *Nc*CDH). Crystals were obtained at pH ~6.5, which is for *Nc*CDH close to its pH optimum while for *Mt*CDH this is an intermediate value between its active and inactive form. [14] Interestingly, the enzymes crystallized in different conformations, each representing one conformational state predicted to occur during pH-mediated interdomain cross-talk in CDH. The static pictures of X-ray structures were further complemented by small angle X-ray scattering (SAXS) data, which allowed estimating the occurrence of different conformers in the system. This analysis showed that both forms (closed and open) of CDH (resembling the crystal forms of *Mt*CDH and *Nc*CDH, respectively) are simultaneously present among other conformers. [19] Also, molecular modelling and domain docking approaches suggest that the molecular mechanism preventing the domain interaction and IET at neutral pH is the repulsion of negative charge patches at the domain-domain interface. [14] The role of charged residues around the interdomain interface was further highlighted by our recent studies where the effect of calcium ions on the IET at neutral pH was described. We showed that calcium enables IET even at higher pH values and this effect can be explained by shielding of the negative charge patches, mostly localized on the interface of DH and CYT domains. [14,20]

In order to structurally validate the theory of pH-regulated domain charge repulsion and to better understand the mechanisms involved in controlling the IET, we focused our present study on determining what effects the pH has on the conformation and dynamics of CDH in solution. For this, we selected CDH from *M. thermophilum*, which was successfully crystallized in its closed (IET-capable) form. [19] As the study of proteins under different pH values can be quite challenging we approached this task using a combination of techniques based on structural mass spectrometry. Here we utilized hydrogen/deuterium exchange coupled with mass spectrometry (HDX-MS) and native ion mobility mass spectrometry (IM-MS). Although these techniques usually have lower structural resolution and do not directly provide 3D structures, they are well suited for monitoring conformational changes of molecules in their various functional states and are theoretically unlimited by the size and complexity of the studied systems. [21–23] In addition, HDX-MS can work under native-like conditions, with the studied protein directly in aqueous environment and at low concentrations, which rules out possible artifacts created by concentrated protein solutions. [24] If interpreted in the context of high-resolution structures, structural MS data add another dimension to our understanding of the studied biological system.

Here we provide experimental evidence for the “charge repulsion model” of the pH-dependent interdomain electron transfer in CDH. We show that at lower pH values, the areas on the interface of CYT and DH domains get protonated and change their organization with respect to hydrogen bonding as evidenced by increased deuteration kinetics. We also demonstrate that this property is an intrinsic behavior of each domain and does not depend on their physical contact.

MATERIALS AND METHODS

Materials

All chemicals were obtained from Sigma-Aldrich unless stated otherwise. Endoglycosidase Endo Hf was supplied by New England Biolabs. Porcine pepsin (3.200-4.500 units/mg) was bought from Sigma-Aldrich, whereas nepenthesin-1 and rhizopuspepsin were recombinantly prepared and purified as described previously. [25,26] All the three proteases were immobilized in-house onto POROS-20AL perfusion resin (Applied Biosystems) using a published procedure. [26–28]

Protein production and deglycosylation

Full-length cellobiose dehydrogenase from *Myriococcum thermophilum* (Uniprot ID: A9XK88, without the 21-aminoacid N-terminal signal sequence) was recombinantly expressed in *Pichia pastoris* and purified as described previously. [29] In order to prepare the two individual *Mt*CDH domains, part of the full-length sample was treated by papain and separated using ion exchange chromatography. [19,30] Enzymatic activity of the *Mt*CDH preparations was tested using the standard electron transfer assays based on monitoring the reduction of 2,6-dichloroindophenol and cytochrome *c* [31,32], while protein concentration was determined by Pierce BCA protein assay kit (Thermo Fisher).

Non-denaturing deglycosylation

For analyses requiring N-deglycosylated protein, *Mt*CDH was treated by glycosidase Endo Hf (New England Biolabs) as described elsewhere.[20] Briefly, *Mt*CDH was incubated overnight at 37°C in 50 mM sodium acetate buffer pH 5.4 with 15 U Endo Hf per 1µg of protein.

Post-translational modification characterization

Analysis of the disulfide bond pattern was performed utilizing a method developed previously. [33] Proteases used for digestion of the protein were Trypsin Gold MS grade (Promega), chymotrypsin, Asp-N and Glu-C (all three sequencing grade from Roche) at 5 ng/µl concentration.

To characterize other post-translational modifications and identify papain cleavage site, native or Endo Hf treated samples were resolved on SDS electrophoresis. Proteins were further in-gel reduced by tris(2-carboxyethyl)phosphine (TCEP), alkylated with iodoacetamide and digested by specific proteases (Trypsin, Asp-N and Glu-C).

Samples were subjected to an LC-MS/MS analysis using collision-induced fragmentation on an Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer equipped with 15T superconducting magnet (SolariX XR, Bruker Daltonics). Tandem mass spectra were searched against a single-protein database containing the *Mt*CDH sequence using Mascot algorithm with carbamidomethylation (Cys, +57.0215), N-acetylhexosamine (Asn, +203.0794), hexose (Ser, Thr, +162.0528) and pyroglutamic acid (N-terminal Gln, -17.0265) set as variable modifications. N- and O-glycosylation sites and glycan structures were identified by manual searching of the data and with the help of the GlycoPeptideSearch engine. [34]

Intact mass measurements of CYT domain were performed using sample desalted by benchtop gel filtration (0.5 ml Zeba Spin column with 7 kDa cut-off, Thermo Fisher Scientific) and diluted to 10 µM into 50% ACN / 0.1% TFA. Protein was ionized by ESI and analyzed using 15T FT-ICR mass spectrometer operated in positive ion mode with continuous accumulation of selected ions (1370±130 m/z window).

Hydrogen/deuterium exchange

*Mt*CDH, either fully glycosylated or N-deglycosylated under native conditions, was transferred into either 50 mM 4-morpholineethanesulfonate (MES) pH 5.4 or 50 mM 4-morpholinepropanesulfonate (MOPS) pH 7.4. After 30 min preincubation, the H/D exchange workflow proceeded as described before. [20] Briefly, after 10-fold dilution by D_2O-based MES or MOPS buffer (pD 5.4 and pD 7.4, respectively) the deuteration was left to proceed at 21°C for 0.33, 1, 3, 10, 30, 60, 180 and 300 min. Protein concentration during the exchange was 5 µM. Following the labelling, 50 µl aliquots were quenched by the addition of 50 µl 6M guanidine, 0.9 M TCEP and 1 M glycine pH 2.4, and incubated for 10 min at 0 °C before being rapidly frozen in liquid nitrogen as optimized previously for proper unfolding and digestion efficiency. [20]

Upon quick thawing, the quenched samples were injected into an LC system described in detail before [20] and consisting of immobilized protease column, peptide desalting microtrap (Michrom Bioresources) and analytical chromatographic column Jupiter C18 (0.5 × 5 mm, 5 µm, 300 Å; Phenomenex). Proteases used for the digestion were either porcine pepsin, recombinant rhizopuspepsin [26] or recombinant nepenthesin-1 [28]. Digestion was carried out in 0.4% formic acid in water with constant flow of 100 µl/min for pepsin column and 200 µl/min when nepenthesin-1 and rhizopuspepsin columns were used together, joined serially by a PEEK column coupler (IDEX Health&Science). After digestion, peptides were desalted (3 min) before being eluted onto the analytical column operated at flow rate 15 µl/min. Peptides were separated by a linear gradient of 10-35% B in 12 min, followed by a jump to 99% B, where solvent A was 0.2% formic acid/2% acetonitrile in water and B was 95% acetonitrile/0.2% formic acid in water. All the fluidic pathway, protease and analytical columns were kept at 0°C to keep the deuterium back-exchange to a minimum.

The LC system was directly interfaced to an electrospray ionization source of a 15T FT-ICR mass spectrometer. The instrument was operated either in LC-MS/MS mode (peptide identification) or LC-MS mode (analysis of deuterated samples). Tandem mass spectrum data were searched against a MASCOT database containing the sequence of *Mt*CDH. HDX-MS data were processed using DeutEx, an in-house developed program.

Native mass spectrometry and ion mobility

Fully glycosylated *Mt*CDH was transferred into 200 mM ammonium acetate pH 5.4 or pH 7.4 by two cycles of benchtop size exclusion chromatography (0.5 ml Zeba Spin column with 7 kDa cut-off). Desalted sample was then introduced into a Synapt G2 HDMS mass spectrometer (Waters) through an automated nanoelectrospray microfluidic injection system (Triversa Nanomate, Advion). Voltage on the electrospray tip was kept at 1.75 kV, while the mass spectrometer parameters were tuned to favor transmission of intact *Mt*CDH molecule with minimal ion activation in the gas phase. Specifically, the instrument acquired data in the m/z range 1,000 – 10,000, with backing pressure kept at 6 mbar while the extraction cone voltage, trap collision energy and trap bias voltage were set to 4 V, 10 V and 40 V, respectively. For collision induced activation experiments, the sample cone voltage was systematically increased from 130 V to 180 V in 5 V increments to gradually activate the ions in a controlled manner. The ion mobility settings were optimized to provide good separation with the drift time of *Mt*CDH ions kept in the second third of the accessible drift time range. Travelling wave height and wave velocity were set to 40 V and 910 m/s, while the gas flows in the helium and ion mobility cells were kept at 100 and 25 ml/min, respectively.

Calculation of protein surface charge

For the simulation of protein surface electrostatics, structures of cytochrome and flavin domain were individually extracted from the crystallographic structure of full-length *Mt*CDH (PDB ID: 4QI6). [19] Hydrogen atoms were assigned to the structures at pH 5.4 and pH 7.4 by PROPKA algorithm [35] used by PDB2PQR web server [36,37] during the preparation of molecules for calculations. The simulations of molecular electrostatics were then performed using the Adaptive Poisson-Boltzmann Software package (APBS) [38] and the results were visualized in PyMOL 1.8.0 (Schrödinger) with APBS Tools 2.1 plugin.

RESULTS AND DISCUSSION

***Mt*CDH is extensively post-translationally modified**

Detailed knowledge of protein primary structure is a prerequisite for successful protein characterization using HDX-MS. Therefore, it was necessary to characterize post-translational modifications present in the *Mt*CDH. This proved especially important as *Mt*CDH expressed in *Pichia pastoris* has been shown to differ from *Mt*CDH produced by the fungus itself. Variations in the molecular mass are due to different N-glycosylation pattern and possible O-glycosylation in the linker region, which was predicted by bioinformatic tools. [29]

Using LC-MS/MS analyses, we confirmed the primary amino acid sequence of *Mt*CDH along with the N-terminal processing occurring during protein maturation (first 21 amino acids in Uniprot ID: A9XK88) and the N-terminal pyroglutamylation. Furthermore, we determined the disulfide bond network for the full-length *Mt*CDH (Figure 1a). This proved to be in agreement with the recent crystal structure of *Mt*CDH [19], but additionally confirmed the existence of the Cys167-Cys211 disulfide bridge, which is not directly observable in the X-ray density.

Concerning the glycosylation status, all the six potential N-glycosylation sites (Asn-119, 400, 437, 516, 671 and 678, present in Asn-Xxx-Ser/Thr consensus sequence) were shown to be modified with high-mannose type glycans (Figure 1b). The only exception was the Asn671, which was only partially glycosylated. Moreover, we also confirmed that the predicted O-glycosylation is indeed present in the

interdomain linker region. In the 42-amino acid long region (188-230) twelve potential O-glycosites (Ser and Thr residues) with up to twenty hexose units attached (Figure 1c) were found. Based on the numbers of hexoses in overlapping peptides produced by trypsin, chymotrypsin, Asp-N and Glu-C proteases, it seems that a kind of clustered O-glycosylation sites is present in the *Mt*CDH linker region. This situation resembles the hinge region in IgA antibodies, where a similar pattern of multiply glycosylated clustered serine and threonine residues was described. [39] The findings are corroborated by the crystal structure of *Mt*CDH, where some mannose units were found in the vicinity of the potential O-glycosylation sites (Ser195, Thr197, Thr204, Thr206 and Thr226). [19]

Papain cleaves *Mt*CDH before the flavin domain

It has long been known that it is possible to use papain to cleave the intact *Mt*CDH molecule into its two constituent domains – CYT and DH. [30] However, the precise cleavage site remained unknown.

After papain cleavage of full-length *Mt*CDH, the two domains were separated, further digested and analyzed as described in the Materials and methods section. Spectra of peptides generated by Asp-N and trypsin digests of the separated DH (Figure 2a) and CYT domains (Figures 2b, 2c) show that the papain cleavage results at the separation of the two domains at either Val220 or Gly221. The presence of these two residues at the peptide termini in the specific protease digests proves that these are indeed the terminal peptides produced by the papain action. Otherwise, this would be in contradiction with the cleavage rules of the specific proteases used.

ESI-FT-ICR spectrum of the intact CYT domain corroborates these findings by showing peak distributions attributable to three types of mass differences (Figure 2d). First, there is a mass difference of 162 Da, which arises from the heterogeneity in the number of hexose units in the O- and N-linked glycans present in the CYT domain. Second, the 57 Da difference between peaks of the same glycosylation state signifies the presence (or absence) of the Gly221 residue at the C-terminus of the domain. Judging by the similar intensity of these two proteolytic products, we can conclude that the papain cleaves N- and C-terminally of this residue with roughly the same probability. Finally, peaks with the 18 Da difference show partial hydrolysis of the N-terminal pyroglutamic acid. The high resolution of FT-ICR mass spectrometer also allowed us to determine the precise monoisotopic masses of intact CYT domain proteoforms. For the most abundant form in Figure 2d the determined mass 25459.387 Da (mass error ±0.3 ppm) corresponds to *Mt*CDH amino acids 1-220 with N-terminal glutamine converted to pyroglutamic acid, two N-acetylhexosamines and eleven hexoses. It also confirms that in the CYT domain three are disulfide bonds formed.

Increasing HDX-MS spatial resolution by alternative aspartic proteases

Having characterized the modifications present on the *Mt*CDH molecule, we focused on optimizing the HDX-MS method for *Mt*CDH structural studies. In previous work, we described HDX-MS compatible conditions for the pepsin digestion of *Mt*CDH. [20] By incubating the deuterium labeled N-deglycosylated *Mt*CDH in a quenching buffer containing denaturing and reducing agents on ice prior to the online pepsinolysis, we achieved almost complete (93%) coverage of the protein missing basically only the O-glycosylated linker region.

However, some of the peptides produced by this immobilized pepsin digestion of *Mt*CDH were quite long, leading to less precise localization of the deuteration changes. (Figure S1, blue) Therefore, we selected alternative aspartic proteases that are complementary in their cleavage preferences to the classical porcine pepsin, in order to increase the spatial resolution and other digestion parameters (peptide redundancy, sequence coverage, etc.). The two proteases, recombinant nepenthesin-1 [28] and rhizopuspepsin [26], were immobilized onto POROS-20AL resin and filled into individual columns. These were coupled serially (with nepenthesin preceding the rhizopuspepsin) in an online LC setup. Initially, we observed that the efficiency of the combined digestion was too high, leading to an overdigestion and a lower coverage of several protein regions due to the production of very short peptides. We limited the digestion time by doubling the sample flow rate through the protease columns

to 200 µl/min. The combined digestion resulted in a slight increase in total sequence coverage (95%) but most importantly, it also led to an increase in the spatial resolution and other HDX-MS protease efficiency metrics (Table 1). This occurred by introducing new cleavage sites different from those obtained by pepsin, at both "pepsin-like" (Trp, Phe, Leu) as well as "non pepsin-like" cleavage sites (basic residues Lys, Arg, His). [26,28] The structural resolution increase is apparent when looking at the sequence coverage map in several parts of the *Mt*CDH molecule – e.g. in regions 1-46, 439-478, 586-616 or 690-720 (Figure S1). Taking a closer look at one of these regions – the ultimate N-terminus of the cytochrome domain – we can clearly see increase in spatial resolution (Figure 3). Originally, with immobilized pepsin (Figure 3a - blue bars) the changes in deuteration could only be mapped onto two rather long peptides (Figure 3b). The shorter and more overlapping peptides produced by nepenthesin in combination with rhizopuspepsin (Figure 3a - red bars) made it possible to narrow down the structurally perturbed regions to the two loops in this part of the domain (Figure 3c).

From a practical perspective, it is important to mention that even though nepenthesin-1 was shown to be very sensitive to reducing and denaturing agents in solution [28,40,41], the immobilization and the flow-through setup of the digestion greatly stabilizes this enzyme [28]. Indeed, during the course of this study including hundreds of injection no adverse effect on the digestion efficiency was observed.

HDX-MS study of pH-dependent domain interaction

In order to study the structural dynamics of *Mt*CDH in solution, we utilized the HDX-MS method as described above. First of all, we checked whether the removal of N-linked glycans by the Endo Hf treatment had any influence on the structure of *Mt*CDH. Although the glycosylation might be important for the long-term as well as thermal stability of *Mt*CDH, HDX-MS data showed that the N-deglycosylation performed under non-denaturing conditions had negligible effect on the overall structure of *Mt*CDH. (Figure S2) This was corroborated by spectrophotometric assays proving that the removal of N-linked glycans did not change *Mt*CDH's enzymatic activity as determined by the cytochrome c assay and as observed in previous work performed with deglycosylated *Mt*CDH immobilized on electrodes.[32,42]

Moreover, as the deglycosylation was performed by Endo Hf glycosidase rather than PNGase F, the proximal N-acetylhexosamine of each glycan remained covalently attached to an asparagine residue. This was desirable as it prevented possible changes in the hydrogen bonding networks of the *Mt*CDH molecule through the conversion of neutral Asn to acidic Asp residues by the action of PNGase F. As a result, we performed further HDX-MS analyses with the more homogeneous N-deglycosylated form.

To gain insight into the mechanism of *Mt*CDH regulation by pH, we used HDX-MS to monitor the enzyme under two different pH values. Based on previous results, we selected pH 5.4, which is close to the pH optimum of IET for class II cellobiose dehydrogenases, as an "active" state with the two domains presumably coming into close contact. In contrast, pH 7.4, with negligible IET, was chosen to represent an "inactive" state. [14,43,44]

Since hydrogen/deuterium exchange strongly depends on solution pH, the deuteration kinetics at the two selected pH values are not directly comparable. [45–48] Such a comparison is however possible after simple correction which, in this particular case (pH 5.4 and 7.4), can be done by offsetting the deuteration curve for the higher pH by a factor of 100 ($10^{\Delta pD}$) on the logarithmic time scale.

The differences in deuteration between the *Mt*CDH at the two pH values mapped on the surface of DH and CYT domains are shown in Figure 4 (complete dataset is shown in Figure S3). The structures were taken from the crystallographic model of the intact *Mt*CDH molecule (PDB ID: 4QI6; Figure 4 inset) and the domains were rotated, so the interdomain interface can be clearly seen. Regions with detected pH-dependent changes in deuteration are color coded in different shades of blue, with the biggest differences represented in the darkest blue.

For both domains, most of the regions affected by the pH-change are localized around their expected interaction interface with some changes on the negatively charged fungal carbohydrate binding module

as well. On the cytochrome domain, these regions covered the majority of its DH-facing side with the most affected parts mapped to loops forming the active center around the heme cofactor (Figure 4b). Similarly, for the flavin domain, the majority of pH-perturbed regions was found around the edge of the cavity leading to the FAD cofactor in the enzymatic active site (Figure 4a). These identified regions fit well into the context of the solved crystal structure of the whole *Mt*CDH in its closed conformation, where the cytochrome domain fits into the cavity on the flavin domain enabling a propionate group of the heme to contact the buried flavin resulting in the transfer of electrons. [19]

Surprisingly, in all the protein regions influenced by the change of pH, we observed increased deuteration at the lower pH at which the two domains interact (Figure 4c, red and dark blue curves). This was unexpected, since in most cases the interaction of proteins or protein domains leads to a decrease in deuteration, resulting from the exclusion of solvent and formation of interprotein hydrogen bonds. However, it has been pointed out by other groups [49–51] that a protein/ligand interaction can result in a variety of observable outcomes in HDX-MS ranging from the canonical protection to deprotection due to conformational rearrangements. Interestingly, highly dynamic or amino acid side chain mediated interactions in rigid systems may also fail to provide any change in deuteration. In order to prove whether the observed changes are caused by the interdomain contact or are instead caused by the changes in pH, we repeated the HDX-MS experiments with individual separated domains of *Mt*CDH. The changes in deuteration around the interdomain interacting surface observed on the individual domains proved to be identical as those detected for the full-length protein (Figure 4c). The only exception was slightly increased pH independent protection of some regions in the separated CYT domain. This is likely associated with the physical separation of DH and CYT domains, which leads to changes of the overall flexibility and/or accessibility of CYT domain (Figure 4c, regions 2 and 3). Despite these changes in the CYT domain we can conclude that in the case of *Mt*CDH pH-modulation, the observed protein backbone deprotection is in fact not directly caused by the domain contacting, but is instead a result of the pH change.
Furthermore, we correlated the HDX-MS data with protein surface electrostatic charge distributions modelled computationally for both *Mt*CDH domains at pH 5.4 and 7.4. In protein regions where the most significant deuteration changes were discovered by HDX-MS, we also observed neutralization of negative charge on protein surface due to protonation at the more acidic pH. (Figure 5a and c) The neutralization is most prominent on the DH domain and especially in its part around and on the sides of the cavity leading to the FAD cofactor in the active site. Thus, the domain surface effectively repolarizes in this region from negative charge at pH 7.4 to neutral or slightly positive values at pH 5.4, suppressing the electrostatic repulsion between the two domains. The involved protonation of amino acid side chains, which in turn leads to the rearrangement of hydrogen bonding networks in the affected regions of *Mt*CDH, is then offering the explanation for the unusual deuteration behavior of *Mt*CDH. As the hydrogen bonds are perturbed, the protein backbone amide hydrogens get more accessible for the exchange with deuterium from solution and/or the protein regions around the *Mt*CDH interdomain interface slightly loosen structurally to better accommodate the domain-domain contact.

Of interest is also the lack of any detectable “classical” backbone amide protection, which is the most commonly observed deuteration scenario in HDX-MS, and normally leads to a deuteration decrease in the regions where two proteins interact. One reason for this unusual behavior may be the fact, that based on the crystal structure, the contact between the two domains seems to be mainly mediated by protruding heme propionate moieties on CYT domain and amino acid side chains on DH domain. [19] Also, the deuteration behavior suggests an inherent high flexibility and rapid dynamics of *Mt*CDH in solution. It seems that the system does not go through a well-defined transition between its active (closed) and inactive (open) states, but rather occupies a plethora of semi-open transitional conformations. The probability of the system to move into its closed IET-competent conformation is then determined by the interdomain electrostatics, which is in turn governed by the pH of the solution. This explanation of the

*Mt*CDH dynamics are in accordance with the data obtained by small-angle X-ray scattering measurements reported together with the crystal structure of full-length CDHs. [19] Here, the SAXS measurements revealed that the open and closed conformations represent only small subset of the populations present in the solution at pH 4.5 with a range of other conformer clusters present concomitantly. [19] Therefore, as the IET-competent state seems to be only occupied for short periods of time and, as no more than 20-30% of the molecules [19] seem to be present in the closed conformation even for the *Mt*CDH active state, it is hardly surprising that only the deuteration changes induced by protonation are directly observable by HDX-MS.

Protein stability probed by IM-MS

Finally, we attempted to validate the HDX-MS findings using native mass spectrometry coupled to ion mobility. This technique, when tuned carefully, can gently transfer macromolecules into the gas phase inside the mass spectrometer, while largely keeping the features of their tertiary and quaternary structure unperturbed. [52] Ion mobility then separates ions based on their size and shape, adding another dimension of information to molecular mass values obtained by MS.

The fully-glycosylated *Mt*CDH was introduced into the mass spectrometer via a nanoelectrospray ionization interface from ammonium acetate solutions of pH 5.4 and 7.4. Even though the presence of both N- and O-glycosylation and their heterogeneity resulted in broad mass peaks, the drift time distributions extracted at both pH for the lowest charge state present (16+, mass window 6450 ± 50 Da) resulted in fairly narrow Gaussian curves (Figure 6b). By comparing the drift time distributions for ions generated from solutions of pH 5.4 and 7.4 (red and blue curves in Figure 6b, respectively) no difference could be detected directly. This means that the ions produced from both pH conditions adopt conformation(s) of similar size in the gas phase. However, differences between the ions appeared when we started to increase their internal energy. This was achieved by gradually increasing the sample cone voltage in the ion source of the mass spectrometer, leading to faster acceleration and more collisions of the ions with residual gas molecules. As the accelerating voltage was incrementally increased from 130 V to 180 V, the ions transitioned distinctly from the original conformation to a more compact one, marked C1 and C2 respectively (Figure 6b). Most importantly, this transition proceeded differently depending on the pH of the solution *Mt*CDH was electrosprayed from. The intermediate accelerating voltages clearly showed that the *Mt*CDH ions formed from the more acidic solution collapsed more easily than those from solution of pH 7.4. This can be seen even more clearly in Figure 6a, where a ratio of conformers “R” was plotted against the accelerating sample cone voltage increased at 5 V increments. The curves were constructed by fitting the drift time distributions at individual voltages with two Gaussian curves, leading to the ratio R of the area under the curve for the collapsed conformer (C2) divided by the sum of areas under the curves of both conformers (C2 + C1). The resulting sigmoidal curves clearly show that *Mt*CDH molecules ionized from pH 5.4 solution require less activation energy to collapse. It should be also noted that exactly the same experiments were performed with the N-deglycosylated form of *Mt*CDH and provided identical result. Thus we can conclude that this behavior is not influenced by the N-glycosylation.

The observed collapse of *Mt*CDH molecules upon gas phase activation is an interesting phenomenon, as most proteins tend to unfold and adopt more extended conformations during such collision-induced unfolding analyses as described above. [53,54] However, for some proteins with internal cavities, ionic collapse was observed preceding their unfolding/fragmentation. [54] Therefore, this could suggest that under both pH the *Mt*CDH molecules are transferred into the gas phase in (at least partially) open conformation and as their internal energy increases, the molecules are forced into their closed state where the cytochrome domain protrudes into the cavity of the flavin domain. As the distribution of negative charges at the domain interface (especially on the DH domain) changes between the two pH states (see Figure 5), the energy barrier of domain contacting stemming from the repulsion of patches of negative charge changes as well. The observed lower energy needed for the transition of ions at pH 5.4 then means that the energy barrier is smaller for the closing of the domains at slightly acidic pH,

which agrees with both the surface electrostatics calculations and the *Mt*CDH enzymatic pH profile data. It is important to point out here that at both pH values ions of identical charge state were compared. Therefore, the difference in ion stability could indeed be explained by a different repartitioning/localization of the same number of protons.
It is also necessary to mention at this point, that the strength of electrostatic interactions determined in the gas phase cannot be directly compared with the strength of forces acting in solution [55] as bulk water, that surrounds molecules in solution, attenuates the effects of electrostatic interactions by partially shielding the charges. The desolvation also leads to the complete absence of hydrophobic interactions in the gas phase. However, despite the fact that the strength of electrostatic interactions between the two domains may be slightly overestimated *in vacuo*, the difference in the electrostatic stability of *Mt*CDH was clearly dependent on its protonation arrangement state.
Most importantly, together with HDX-MS experiments and computational electrostatics calculations, which studied the system directly in solution or took the presence of solvent into account through calculation with implicit solvation [38], the experimental data show that the electrostatic repulsion is indeed the key factor in pH-dependent regulation of the transient interaction of *Mt*CDH domains in solution. This finding is also consistent with and experimentally supports the theory of domain charge repulsion proposed in previous studies, which described how the divalent cations bridge the patches of negative charge and enable the *Mt*CDH domains to come into close contact necessary for IET, even at neutral pH. [14,20]

CONCLUSIONS

In this work mass spectrometry was used to study *Mt*CDH in solution. Primary structure analysis confirmed extensive O-glycosylation in the flexible linker region and also answered the long-standing question of the exact site of papain cleavage, which has long been used to produce separate *Mt*CDH domains. This enzymatic action was found to take place at both N- and C- terminal sides of Gly221 in the linker region.
HDX-MS was used to probe the conformational dynamics and the pH-regulated electron transfer function of *Mt*CDH. Here we benefited from improved proteolytic setup combining immobilized nepenthesin-1 [28] with rhizopuspepsin [26], which increased the digestion efficiency and spatial resolution. Together with computational electrostatics simulations HDX-MS provided a direct experimental evidence that the contacting of *Mt*CDH domains in solution is regulated by the repulsion of patches of negative charge at the interdomain interface. This supports the charge repulsion theory of CDH functioning as proposed in previous works and in our study of the effects of divalent cations on the activity of *Mt*CDH at neutral to alkaline pH. [14,17,20] According to our current results, at acidic pH, regions of negative charge mainly on the DH domain are neutralized by protonation, allowing the interaction with the CYT domain and enabling efficient interdomain electron transfer. This observation was also linked with results provided by native IM-MS, which corroborated the conclusion that electrostatic repulsion between domains is significantly stronger at higher pH.
Also, our results indicate that the interdomain interaction in *Mt*CDH is transient with many protein conformations existing concomitantly in the solution, as suggested by previous SAXS measurements. [19] This manifested itself in HDX-MS by the absence of “classical” protection by interaction. In the case of *Mt*CDH, only deprotection resulting from the protonation of amino acid sidechains was observed. Therefore, this enzymatic system reiterated the necessity of looking beyond the most common scenarios and outcomes when interpreting HDX-MS results. [49,50]

Acknowledgements

This project was supported by the Czech Science Foundation (Grant P206/12/0503) and the Austrian Science Foundation (Grant P12069). S.C. was supported by the CNRS and the University of Strasbourg and also thanks the GIS IBiSA, the Région Alsace and the Communauté Urbaine de

Strasbourg for the financial support in purchasing the Synapt G2 HDMS system. J.S. acknowledges the Institut de Recherches Servier for supporting his doctoral fellowship.

References

[1] M. Zamocky, R. Ludwig, C. Peterbauer, B.M. Hallberg, C. Divne, P. Nicholls, D. Haltrich, Cellobiose dehydrogenase--a flavocytochrome from wood-degrading, phytopathogenic and saprotropic fungi., Curr. Protein Pept. Sci. 7 (2006) 255–280. doi:10.2174/138920306777452367.

[2] C.M. Phillips, A.T. Iavarone, M.A. Marletta, Quantitative proteomic approach for cellulose degradation by Neurospora crassa, J. Proteome Res. 10 (2011) 4177–4185. doi:10.1021/pr200329b.

[3] J. Sun, C. Tian, S. Diamond, N. Louise Glassa, Deciphering transcriptional regulatory mechanisms associated with hemicellulose degradation in Neurospora crassa, Eukaryot. Cell. 11 (2012) 482–493. doi:10.1128/EC.05327-11.

[4] R. Ludwig, W. Harreither, F. Tasca, L. Gorton, Cellobiose dehydrogenase: A versatile catalyst for electrochemical applications, ChemPhysChem. 11 (2010) 2674–2697. doi:10.1002/cphc.201000216.

[5] R. Ludwig, R. Ortiz, C. Schulz, W. Harreither, C. Sygmund, L. Gorton, Cellobiose dehydrogenase modified electrodes: advances by materials science and biochemical engineering, Anal. Bioanal. Chem. (2013). doi:10.1007/s00216-012-6627-x.

[6] A.K.G. Felice, C. Sygmund, W. Harreither, R. Kittl, L. Gorton, R. Ludwig, Substrate Specificity and Interferences of a Direct-Electron-Transfer-Based Glucose Biosensor, J. Diabetes Sci. Technol. 7 (2013) 669–677. doi:10.1177/193229681300700312.

[7] C.M. Phillips, W.T. Beeson, J.H. Cate, M. a Marletta, Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by Neurospora crassa., ACS Chem. Biol. 6 (2011) 1399–406. doi:10.1021/cb200351y.

[8] R.J. Quinlan, M.D. Sweeney, L. Lo Leggio, H. Otten, J.-C.N. Poulsen, K.S. Johansen, K.B.R.M. Krogh, C.I. Jørgensen, M. Tovborg, A. Anthonsen, T. Tryfona, C.P. Walter, P. Dupree, F. Xu, G.J. Davies, P.H. Walton, Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components., Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 15079–84. doi:10.1073/pnas.1105776108.

[9] Z. Fan, W. Wu, A. Hildebrand, T. Kasuga, R. Zhang, X. Xiong, A novel biochemical route for fuels and chemicals production from cellulosic biomass, PLoS One. 7 (2012) 1–8. doi:10.1371/journal.pone.0031693.

[10] W.C. Higham, D. Gordon-Smith, E. Dempsey, P.M. Wood, Direct ‘H NMR evidence for conversion of b-D-cellobiose to cellobionolactone by cellobiose dehydrogenase from Phanerochaete chrysosporium, Fed. Eur. Biochem. Soc. 351 (1994) 128–132.

[11] D. Kracher, S. Scheiblbrandner, A.K.G. Felice, E. Breslmayr, M. Preims, K. Ludwicka, D. Haltrich, V.G.H. Eijsink, R. Ludwig, Extracellular electron transfer systems fuel cellulose oxidative degradation, Science (80-.). 3165 (2016) 1–13. doi:10.1126/science.aaf3165.

[12] M. Samejima, K.E.L. Eriksson, A comparison of the catalytic properties of cellobiose:quinone oxidoreductase and cellobiose oxidase from Phanerochaete chrysosporium, Eur. J. Biochem. 207 (1992) 103–107. doi:10.1111/j.1432-1033.1992.tb17026.x.

[13] W. Harreither, C. Sygmund, M. Augustin, M. Narciso, M.L. Rabinovich, L. Gorton, D.

Haltrich, R. Ludwig, Catalytic properties and classification of cellobiose dehydrogenases from ascomycetes, Appl. Environ. Microbiol. 77 (2011) 1804–1815. doi:10.1128/AEM.02052-10.

[14] D. Kracher, K. Zahma, C. Schulz, C. Sygmund, L. Gorton, R. Ludwig, Inter-domain electron transfer in cellobiose dehydrogenase: modulation by pH and divalent cations, FEBS J. 282 (2015) 3136–3148. doi:10.1111/febs.13310.

[15] K. Igarashi, I. Momohara, T. Nishino, M. Samejima, Kinetics of inter-domain electron transfer in flavocytochrome cellobiose dehydrogenase from the white-rot fungus Phanerochaete chrysosporium., Biochem. J. 365 (2002) 521–6. doi:10.1042/BJ20011809.

[16] K. Igarashi, M. Yoshida, H. Matsumura, N. Nakamura, H. Ohno, M. Samejima, T. Nishino, Electron transfer chain reaction of the extracellular flavocytochrome cellobiose dehydrogenase from the basidiomycete Phanerochaete chrysosporium, FEBS J. 272 (2005) 2869–2877. doi:10.1111/j.1742-4658.2005.04707.x.

[17] B.M. Hallberg, G. Henriksson, G. Pettersson, C. Divne, Crystal structure of the flavoprotein domain of the extracellular flavocytochrome cellobiose dehydrogenase., J. Mol. Biol. 315 (2002) 421–434. doi:10.1006/jmbi.2001.5246.

[18] B.M. Hallberg, T. Bergfors, K. Bäckbro, G. Pettersson, G. Henriksson, C. Divne, A new scaffold for binding haem in the cytochrome domain of the extracellular flavocytochrome cellobiose dehydrogenase., Structure. 8 (2000) 79–88.

[19] T.-C. Tan, D. Kracher, R. Gandini, C. Sygmund, R. Kittl, D. Haltrich, B.M. Hällberg, R. Ludwig, C. Divne, Structural basis for cellobiose dehydrogenase action during oxidative cellulose degradation., Nat. Commun. 6 (2015) 7542. doi:10.1038/ncomms8542.

[20] A. Kadek, D. Kavan, A.K.G. Felice, R. Ludwig, P. Halada, P. Man, Structural insight into the calcium ion modulated interdomain electron transfer in cellobiose dehydrogenase, FEBS Lett. 589 (2015) 1194–1199. doi:10.1016/j.febslet.2015.03.029.

[21] J.R. Engen, Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS, Anal. Chem. 81 (2009) 7870–7875. doi:10.1021/ac901154s.

[22] P. Man, C. Montagner, H. Vitrac, D. Kavan, S. Pichard, D. Gillet, E. Forest, V. Forge, Accessibility changes within diphtheria toxin T domain upon membrane penetration probed by hydrogen exchange and mass spectrometry, J. Mol. Biol. 414 (2011) 123–134. doi:10.1016/j.jmb.2011.09.042.

[23] E. Trabjerg, R.U. Jakobsen, S. Mysling, S. Christensen, T.J.D. Jørgensen, K.D. Rand, Conformational Analysis of Large and Highly Disulfide-Stabilized Proteins by Integrating Online Electrochemical Reduction into an Optimized H/D Exchange Mass Spectrometry Workflow, Anal. Chem. 87 (2015) 8880–8888. doi:10.1021/acs.analchem.5b01996.

[24] D. Rozbesky, P. Man, D. Kavan, J. Chmelik, J. Cerny, K. Bezouska, P. Novak, Chemical cross-linking and H/D exchange for fast refinement of protein crystal structure., Anal. Chem. 84 (2012) 867–70. doi:10.1021/ac202818m.

[25] A. Kadek, V. Tretyachenko, H. Mrazek, L. Ivanova, P. Halada, M. Rey, D.C. Schriemer, P. Man, Expression and characterization of plant aspartic protease nepenthesin-1 from Nepenthes gracilis., Protein Expr. Purif. 95 (2014) 121–128. doi:10.1016/j.pep.2013.12.005.

[26] M. Rey, P. Man, G. Brandolin, E. Forest, L. Pelosi, Recombinant immobilized rhizopuspepsin as a new tool for protein digestion in hydrogen/deuterium exchange mass spectrometry, Rapid Commun. Mass Spectrom. 23 (2009) 3431–3438. doi:10.1002/rcm.4260.

[27] L. Wang, H. Pan, D.L. Smith, Hydrogen exchange-mass spectrometry: optimization of

digestion conditions., Mol. Cell. Proteomics. 1 (2002) 132–138.

[28] A. Kadek, H. Mrazek, P. Halada, M. Rey, D.C. Schriemer, P. Man, Aspartic protease nepenthesin-1 as a tool for digestion in hydrogen/deuterium exchange mass spectrometry., Anal. Chem. 86 (2014) 4287–94. doi:10.1021/ac404076j.

[29] M. Zamocky, C. Schumann, C. Sygmund, J. O'Callaghan, A.D. Dobson, R. Ludwig, D. Haltrich, C.K. Peterbauer, Cloning, sequence analysis and heterologous expression in Pichia pastoris of a gene encoding a thermostable cellobiose dehydrogenase from Myriococcum thermophilum, Protein Expr Purif. 59 (2008) 258–265. doi:10.1016/j.pep.2008.02.007.

[30] G. Henriksson, G. Pettersson, G. Johansson, A. Ruiz, E. Uzcategui, Cellobiose oxidase from Phanerochaete chrysosporium can be cleaved by papain into two domains, Eur. J. Biochem. 196 (1991) 101–106.

[31] G. Canevascini, P. Borer, J.L. Dreyer, Cellobiose dehydrogenases of Sporotrichum (Chrysosporium) thermophile., Eur. J. Biochem. 198 (1991) 43–52.

[32] U. Baminger, S.S. Subramaniam, V. Renganathan, D. Haltrich, Purification and characterization of cellobiose dehydrogenase from the plant pathogen Sclerotium (Athelia) rolfsii., Appl. Environ. Microbiol. 67 (2001) 1766–74. doi:10.1128/AEM.67.4.1766-1774.2001.

[33] P. Pompach, P. Man, D. Kavan, K. Hofbauerova, V. Kumar, K. Bezouska, V. Havlícek, P. Novák, Modified electrophoretic and digestion conditions allow a simplified mass spectrometric evaluation of disulfide bonds., J. Mass Spectrom. 44 (2009) 1571–8. doi:10.1002/jms.1609.

[34] P. Pompach, K.B. Chandler, R. Lan, N. Edwards, R. Goldman, Semi-automated identification of N-Glycopeptides by hydrophilic interaction chromatography, nano-reverse-phase LC-MS/MS, and glycan database search., J. Proteome Res. 11 (2012) 1728–40. doi:10.1021/pr201183w.

[35] M.H.M. Olsson, C.R. Søndergaard, M. Rostkowski, J.H. Jensen, PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions, J. Chem. Theory Comput. 7 (2011) 525–537. doi:10.1021/ct100578z.

[36] T.J. Dolinsky, J.E. Nielsen, J.A. McCammon, N.A. Baker, PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations., Nucleic Acids Res. 32 (2004) W665–667. doi:10.1093/nar/gkh381.

[37] T.J. Dolinsky, P. Czodrowski, H. Li, J.E. Nielsen, J.H. Jensen, G. Klebe, N.A. Baker, PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations., Nucleic Acids Res. 35 (2007) W522–525. doi:10.1093/nar/gkm276.

[38] N.A. Baker, D. Sept, S. Joseph, M.J. Holst, J.A. McCammon, Electrostatics of nanosystems: application to microtubules and the ribosome., Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 10037–10041. doi:10.1073/pnas.181342398.

[39] K. Takahashi, S.B. Wall, H. Suzuki, A.D. Smith, S. Hall, K. Poulsen, M. Kilian, J.A. Mobley, B.A. Julian, J. Mestecky, J. Novak, M.B. Renfrow, Clustered O-Glycans of IgA1: Defining Macro- and Microheterogeneity by Use of Electron Capture/Transfer Dissociation, Mol. Cell. Proteomics. 9 (2010) 2545–2557. doi:10.1074/mcp.M110.001834.

[40] K. Kubota, Y. Metoki, S.B.P. Athauda, C. Shibata, K. Takahashi, Stability profiles of nepenthesin in urea and guanidine hydrochloride: comparison with porcine pepsin A, Biosci Biotechnol Biochem. 74 (2010) 2323–2326. doi:10.1271/bbb.100391.

[41] M. Yang, M. Hoeppner, M. Rey, A. Kadek, P. Man, D.C. Schriemer, Recombinant Nepenthesin II for Hydrogen/Deuterium Exchange Mass Spectrometry., Anal. Chem. 87 (2015) 6681–7. doi:10.1021/acs.analchem.5b00831.

[42] R. Ortiz, H. Matsumura, F. Tasca, K. Zahma, M. Samejima, K. Igarashi, R. Ludwig, L. Gorton, Effect of deglycosylation of cellobiose dehydrogenases on the enhancement of direct electron transfer with electrodes, Anal. Chem. 84 (2012) 10315–10323. doi:10.1021/ac3022899.

[43] W. Harreither, P. Nicholls, C. Sygmund, L. Gorton, R. Ludwig, Investigation of the pH-dependent electron transfer mechanism of ascomycetous class II cellobiose dehydrogenases on electrodes, Langmuir. 28 (2012) 6714–6723. doi:10.1021/la3005486.

[44] W. Harreither, V. Coman, R. Ludwig, D. Haltrich, L. Gorton, Investigation of graphite electrodes modified with cellobiose dehydrogenase from the ascomycete Myriococcum thermophilum, Electroanalysis. 19 (2007) 172–180. doi:10.1002/elan.200603688.

[45] A. Hvidt, A Discussion of the pH Dependene of the Hydrogen-Deuterium Exchange of Proteins., C. R. Trav. Lab. Carlsberg. 34 (1964) 299–317.

[46] Y. Bai, J.S. Milne, L. Mayne, S.W. Englander, Primary structure effects on peptide group hydrogen exchange., Proteins. 17 (1993) 75–86. doi:10.1002/prot.340170110.

[47] Z. Zhang, D.L. Smith, Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation., Protein Sci. 2 (1993) 522–31. doi:10.1002/pro.5560020404.

[48] P. Man, C. Montagner, G. Vernier, B. Dublet, A. Chenal, E. Forest, V. Forge, Defining the Interacting Regions between Apomyoglobin and Lipid Membrane by Hydrogen/Deuterium Exchange Coupled to Mass Spectrometry, J. Mol. Biol. 368 (2007) 464–472. doi:10.1016/j.jmb.2007.02.014.

[49] M.A. Sowole, L. Konermann, Effects of protein-ligand interactions on hydrogen/deuterium exchange kinetics: canonical and noncanonical scenarios., Anal. Chem. 86 (2014) 6715–22. doi:10.1021/ac501849n.

[50] L. Konermann, A. Rodriguez, M.A. Sowole, Type 1 and type 2 scenarios in hydrogen exchange mass spectrometry studies on protein-ligand complexes, Analyst. 139 (2014) 6078–6087. doi:10.1039/c4an01307g.

[51] J.R. Engen, Analysis of protein complexes with hydrogen exchange and mass spectrometry., Analyst. 128 (2003) 623–8.

[52] B.T. Ruotolo, C. V Robinson, Aspects of native proteins are retained in vacuum., Curr. Opin. Chem. Biol. 10 (2006) 402–8. doi:10.1016/j.cbpa.2006.08.020.

[53] J.T.S. Hopper, N.J. Oldham, Collision Induced Unfolding of Protein Ions in the Gas Phase Studied by Ion Mobility-Mass Spectrometry: The Effect of Ligand Binding on Conformational Stability, J. Am. Soc. Mass Spectrom. 20 (2009) 1851–1858. doi:10.1016/j.jasms.2009.06.010.

[54] Z. Hall, A. Politis, M.F. Bush, L.J. Smith, C. V Robinson, Charge-state dependent compaction and dissociation of protein complexes: insights from ion mobility and molecular dynamics, J. Am. Chem. Soc. 134 (2012) 3429–3438. doi:10.1021/ja2096859.

[55] S. Yin, Y. Xie, J.A. Loo, Mass Spectrometry of Protein-Ligand Complexes: Enhanced Gas-Phase Stability of Ribonuclease-Nucleotide Complexes, J. Am. Soc. Mass Spectrom. 19 (2008) 1199–1208. doi:10.1016/j.jasms.2008.05.012.

Figure legends

Figure 1. Post-translational modifications of *Mt*CDH. **(a)** MS-derived pattern of disulfide bond linkages in the full-length protein. **(b)** All six N-glycosylation sites have a broad distribution of high-mannose-type glycans similar as the Asn119 site shown here. Blue square – N-acetylhexosamine, green circle – hexose. **(c)** Overlapping Asp-N/trypsin-generated peptides show clustered O-glycosylation in the linker region of *Mt*CDH. The spectrum illustrates the O-glycan heterogeneity for the 206-247 protein region.

Figure 2. Papain separates *Mt*CDH domains by cleaving in the linker region. Asp-N/trypsin-generated peptides of DH **(a)** and CYT **(b, c)** domains show the cleavage to occur both N- and C-terminally to Gly221. **(d)** High-resolution ESI-FT-ICR spectrum of CYT domain shows a range of proteoforms differing by glycine residue (Δ57 Da), number of hexoses (Δ162 Da) and partial hydrolysis of N-terminal pyroglutamic acid (Δ18 Da). Inset shows baseline isotopic separation of the 1427.5890 (18+) peak.

Figure 3. Alternative aspartic protease cleavage increases the spatial resolution of HDX-MS. **(a)** Peptide mapping of the ultimate N-terminus of *Mt*CDH. Individual peptides are drawn as blue bars (immobilized pepsin) and red bars (combined digestion by immobilized nepenthesin-1 and rhizopuspepsin). The additional cleavage sites enabled more precise localization of deuteration changes at the N-terminus of CYT domain (red and orange regions) with the combined irNep-1+irRpn digestion **(c)** compared to porcine pepsin **(b)**.

Figure 4. HDX-MS detected changes at the *Mt*CDH interdomain interface between pH 5.4 (active) and 7.4 (inactive). **(a)** DH domain with the most intense changes around the cusp of FAD cavity. Some perturbance was also seen for the fungal carbohydrate binding module (shown in teal color). **(b)** CYT domain with strongest pH effects detected around the heme cofactor at the DH-binding surface. For the sake of clarity the domains were rotated by 90° in opposite directions from their natural orientation (inset). *Mt*CDH structure (PDB ID: 4QI6) was colored according to the detected changes in deuteration. Grey means no difference while different shades of blue mark the deprotected regions with dark blue showing the biggest difference. Red regions show the connecting points of interdomain linker. **(c)** HDX-MS curves for the individual regions of full-length *Mt*CDH at pH 5.4 and 7.4 (red and dark blue curves, respectively). Separated domains (black, dark green, magenta and orange) and their mixture (light green and cyan) behave similarly as full-length protein. *Mt*CDH regions unaffected by pH show no difference (5, 7), while in the perturbed regions there is significant deprotection at the lower pH.

Figure 5. Protein surface electrostatics calculated for *Mt*CDH domains – DH (top) and CYT (bottom) – at pH 5.4 (a, c) and pH 7.4 (b, d). Electrostatic potential is shown as color gradient from red to blue (-12 to +12 kT/e). Green lines mark areas of negative charge neutralization, which correlate with areas of the most pronounced differences detected by HDX-MS.

Figure 6. Native IM-MS shows different electrostatic stability of *Mt*CDH ionized from pH 5.4 (red) and pH 7.4 (blue). **(a)** Ions of fully-glycosylated *Mt*CDH generated from lower pH collapse more easily with increasing sample cone voltage (SCV) than ions of the same charge state from higher pH. Conformer ratio R = C2 / (C1+C2), where C1 and C2 are areas-under-curve in drift time distributions for individual activation voltages **(b)**.

Table legends

Table 1. *Mt*CDH digestion metrics - Comparison of protein digestion by immobilized pepsin and the combination of nepenthesin-1 with rhizopuspepsin. iPep – immobilized porcine pepsin at 100 µl/min; irNep-1 + irRpn – sequentially coupled immobilized recombinant nepenthesin-1 and recombinant rhizopuspepsin columns at 200 µl/min; Cleavage efficiency – percentage of cleaved peptide bonds; redundancy score – average times a residue is covered by unique peptides. AAs – amino acids.

Figure 1

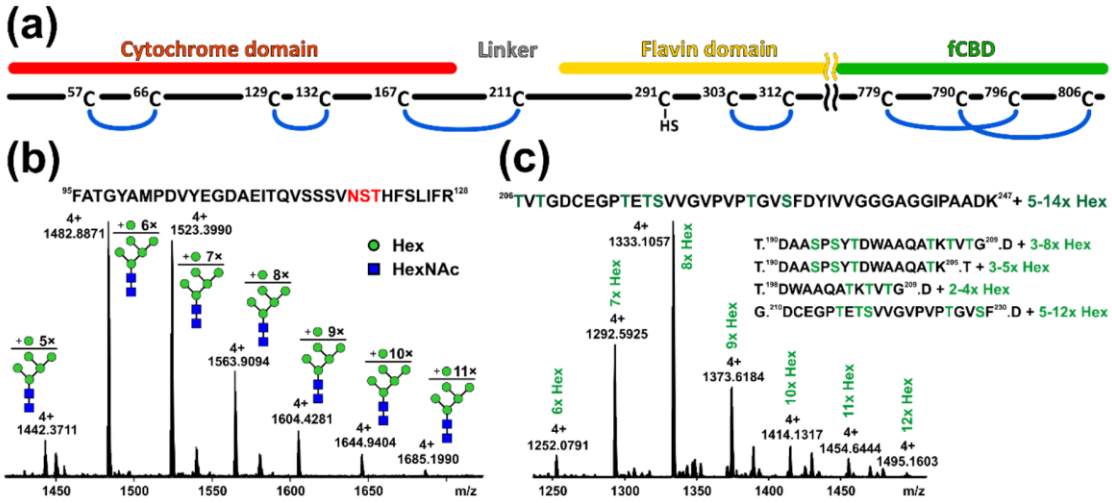

(a)
Cytochrome domain
Linker
Flavin domain
fCBD
57C 66C 129C 132C 167C 211C 291C 303C 312C 779C 790C 796C 806C
HS
(b)
95FATGYAMPDVYEGDAEITQVSSSVNSTHFSLIFR128
Hex
HexNAc
4+ 1482.8871
4+ 1523.3990
4+ 1563.9094
4+ 1604.4281
4+ 1644.9404
4+ 1685.1990
4+ 1442.3711
5×
6×
7×
8×
9×
10×
11×
1450
1500
1550
1600
1650
m/z
(c)
206TVTGDCEGPTETSVVGVPVPTGVSFDYIVVGGGAGGIPAADK247+ 5-14x Hex
T.190DAASPSYTDWAAQATKTVTG209.D + 3-8x Hex
T.190DAASPSYTDWAAQATK205.T + 3-5x Hex
T.198DWAAQATKTVTG209.D + 2-4x Hex
G.210DCEGPTETSVVGVPVPTGVSF230.D + 5-12x Hex
4+ 1333.1057
4+ 1292.5925
4+ 1373.6184
4+ 1414.1317
4+ 1454.6444
4+ 1495.1603
4+ 1252.0791
8x Hex
7x Hex
9x Hex
10x Hex
11x Hex
12x Hex
6x Hex
1250
1300
1350
1400
1450
m/z

Figure 2

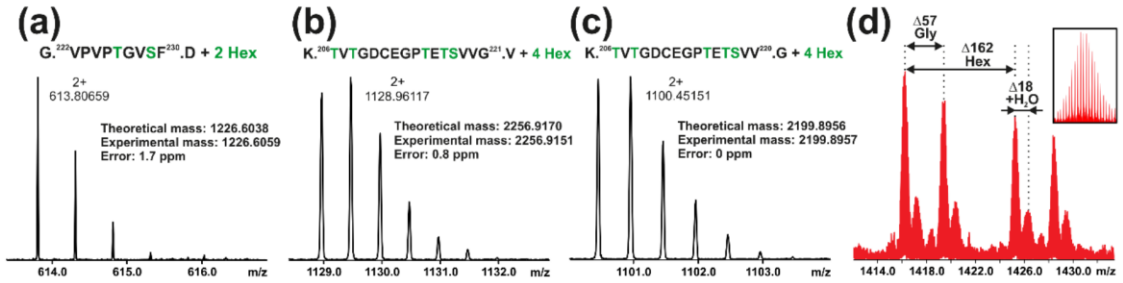

(a)
G.222VPVPTGVSF230.D + 2 Hex
2+
613.80659
Theoretical mass: 1226.6038
Experimental mass: 1226.6059
Error: 1.7 ppm
614.0
615.0
616.0
m/z
(b)
K.206TVTGDCEGPTETSVVG221.V + 4 Hex
2+
1128.96117
Theoretical mass: 2256.9170
Experimental mass: 2256.9151
Error: 0.8 ppm
1129.0
1130.0
1131.0
1132.0
m/z
(c)
K.206TVTGDCEGPTETSVV220.G + 4 Hex
2+
1100.45151
Theoretical mass: 2199.8956
Experimental mass: 2199.8957
Error: 0 ppm
1101.0
1102.0
1103.0
m/z
(d)
Δ57
Gly
Δ162
Hex
Δ18
+H2O
1414.0
1418.0
1422.0
1426.0
1430.0 m/z

Figure 3

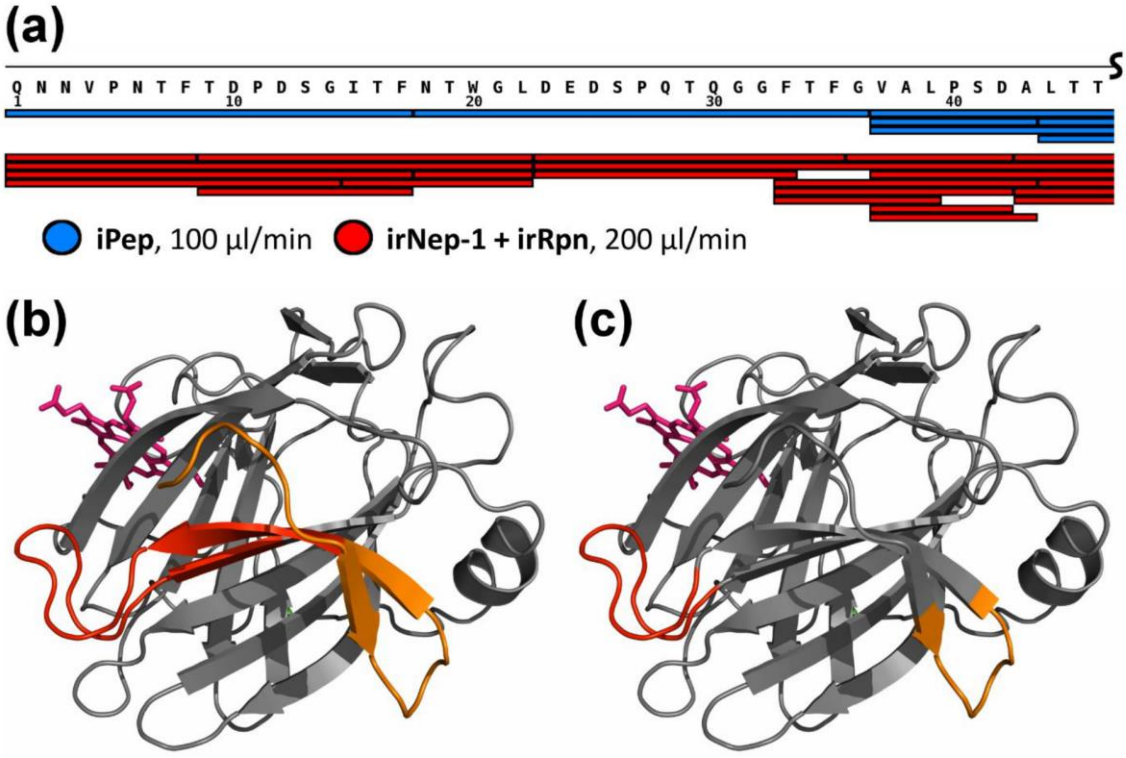

(a)
Q N N V P N T F T D P D S G I T F N T W G L D E D S P Q T Q G G F T F G V A L P S D A L T T
1
10
20
30
40
S
iPep, 100 µl/min
irNep-1 + irRpn, 200 µl/min
(b)
(c)

Figure 4

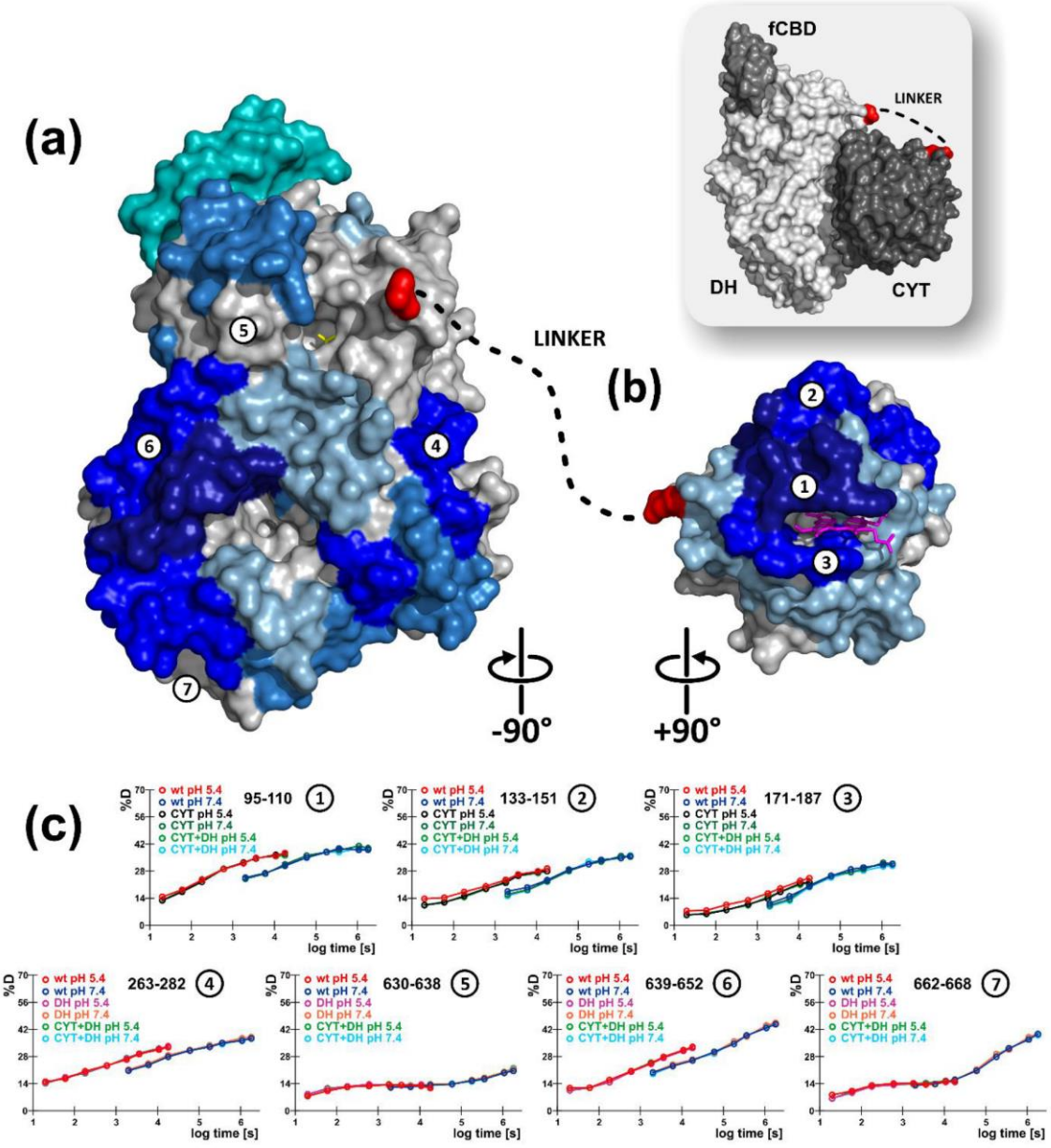

fCBD
LINKER
DH
CYT
(a)
LINKER
(b)
-90°
+90°
(c)
wt pH 5.4
wt pH 7.4
CYT pH 5.4
CYT pH 7.4
CYT+DH pH 5.4
CYT+DH pH 7.4
DH pH 5.4
DH pH 7.4
95-110
133-151
171-187
263-282
630-638
639-652
662-668
%D
log time [s]

Figure 5

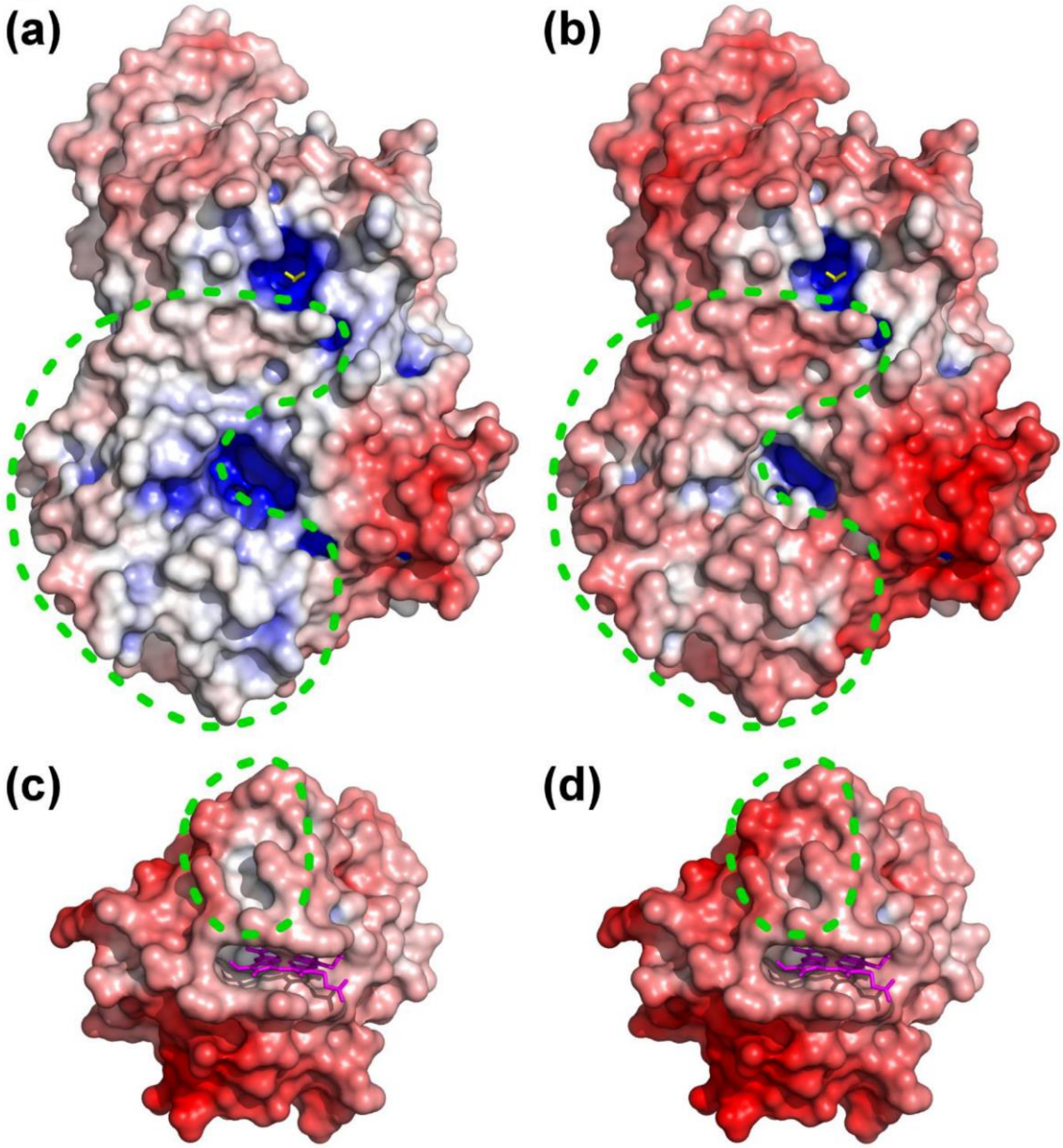


Figure 6

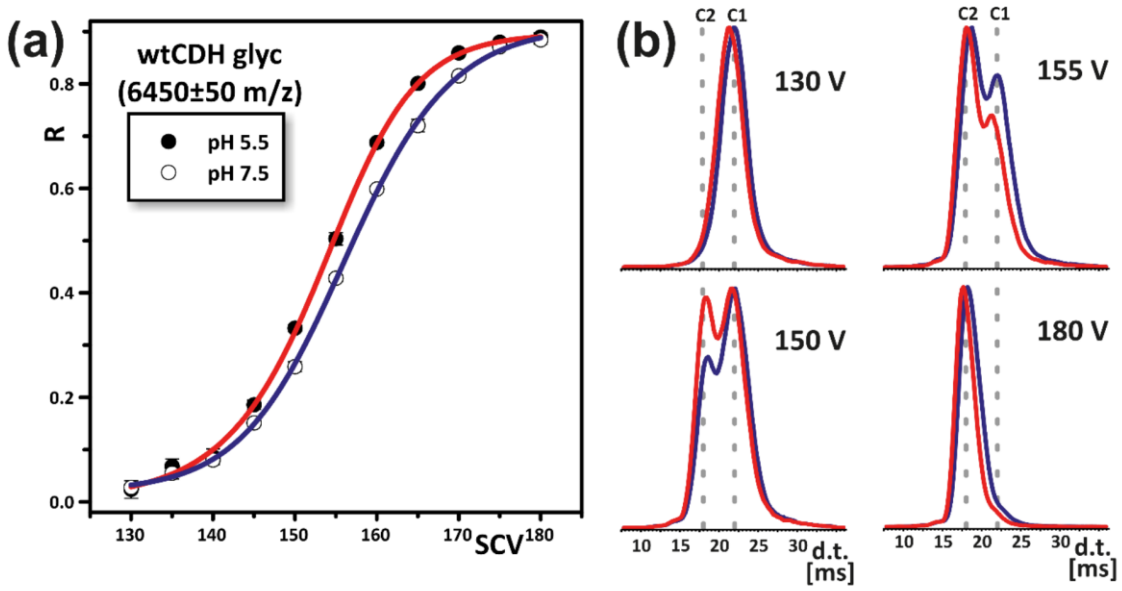

(a)
wtCDH glyc
(6450±50 m/z)
pH 5.5
pH 7.5
R
SCV
(b)
C2 C1
130 V
155 V
150 V
180 V
d.t. [ms]

Table 1

Table 1

	iPep	irNep-1 + irRpn
Number of peptides	229	370
Average peptide length (AAs)	25.5	20.5
Sequence coverage (%)	92.7	94.7
Cleavage efficiency (%)	14.1	23.3
Redundancy score	7.23	9.41