

Structural Study of eIF Complexes by H/D Exchange FT-ICR Mass Spectrometry

Jianqing Wu

▶ **To cite this version:**

Jianqing Wu. Structural Study of eIF Complexes by H/D Exchange FT-ICR Mass Spectrometry. Analytical chemistry. Ecole Polytechnique X, 2013. English. NNT : . pastel-00914013

HAL Id: pastel-00914013
https://pastel.hal.science/pastel-00914013

Submitted on 4 Dec 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE

présentée pour obtenir le grade de

DOCTEUR DE L'ÉCOLE POLYTECHNIQUE

Spécialité : Chimie

Par

Jianqing WU

Structures de complexes d'initiation de la traduction étudiée par échanges isotopiques couplés à la spectrométrie de masse FT-ICR

Directeur de thèse : Guillaume van der Rest

Soutenue le 27 septembre 2013 devant la commission d'Examen composée de**:**

Pr.	Thomas J.D. Jørgensen	Rapporteur
Dr.	Martine Cadène	Rapporteur
Dr.	Delphine Pflieger	Examinateur
Dr.	Emmanuelle Schmitt	Co-directrice de thèse
Pr.	Guillaume van der Rest	Directeur de thèse

工欲善其事，必先利其器

孔子，《论语·卫灵公》

He that would perfect his work must first sharpen his tools.

Confucious, *Analects*

Acknowledgement

I would first like to thank *Dr. Gilles Ohanessian*, director of the laboratoire des Mécanismes Réactionnels, for the opportunity to work in this laboratory and complete the study of this thesis.

My deepest gratitude goes to *Prof. Guillaume van der Rest*, my thesis supervisor. What is important that I learned from you is not only the chemistry and biology written in this thesis, but the philosophy of scientific research from and beyond our day-to-day discussions which I will always benefit.

In addition, I shall truly thank *Dr. Emmanuelle Schmitt*, my co-supervisor, whose instructions and inspirations from the biochemistry aspect are essential for this work. Thanks *Pierre-Damien* for instructing me the biochemistry experiments as well as helping me with the protein 3D models. I indeed appreciate your knowledge, kindness and patience. My gratitude also goes to *Dr. Yves Méchulam*, director of the laboratoire Biochemie for your supports during the experiments and fruitful discussions.

I would also like to thank *Edith*, who has always been patient and optimistic helping me with the difficulties of the instruments and experiments. Without you, my experiments would never be finished within three years.

A lot of thanks to those with whom we shared the same office: *Ophélie, Yasmine, Jingjing, Jana, Christian, Joe, Yannick, Aziz* and *Florian*. It has always been a pleasure working and talking with you, during which I did improve my French a lot! Besides I shall specially thank Renjie, who was like my elder sister in DCMR taking warm care of me when she was a PhD student in the same laboratory.

My heartful thanks also go to all the staff, researchers (or former researchers) and students in DCMR, to *Thérèse, Julia, Gilles, Carine, Guy, Sophie, Stephane, Michel, Christophe, Julien, Vanessa, Manjeet*… You have made this three years in DCMR a memory that I will cherish forever in my life. I wish you all the best for the future.

I appreciate, as well, the helps from the Proteomics Platform in Pasteur Institute on the MS/MS experiments, which were critical to a major part of this work.

My gratitude also goes to Prof. *Thomas J.D. Jørgensen* and Dr. *Martine Cadène* for reviewing my thesis, and to Dr. *Delphine Pflieger* for evaluating my work and being my thesis committee member.

I shall also express my gratitude to my family and friends --

My parents, no matter where I am or what decisions I make, you are there for me. There are never enough words to describe how much you support me and how grateful I am. I am the luckiest girl in the world because I am your daughter.

My family, your love is the harbour of my heart. Even though many of you know nothing about my work, you have always been encouraging and supporting me to pursue the achievement of this thesis.

My boyfriend, Fanji, I have always expected a relationship with love, respect, fun and full supports, and we are having it now.

My dear Friend, Xin, whom I respect as a scientist and as a person. The time in Ecole Polytechnique with you will remain a page of joy in my life. My friends Lili, Zheng, Zhibo, Zixian, Xue, Jie, Shiguang, Songzhe, Ling, Xianzhou, Yizhou…You are my family here. I wish you all a bright future.

同样也要感谢我的家人和朋友们——

我的父母，无论我人在何方，所往何处，你们都在我的身边。父母之恩，言语不可表其万一。只能说生为你们的女儿，是我的最大幸运。

我的家人，你们的爱就如同我心的港湾。虽然你们并不一定了解我的工作，但从未停止过对我学业的支持与鼓励。

我的男友，孟凡济，我期望的是充满爱，尊重，快乐与互相支持的感情，正如我们的感情。

我可爱的朋友，钱欣，无论作为一名科研工作者还是作为一个独立的人，你都让我敬佩。与你在巴黎综合理工度过的这段日子，在我人生中留下了美好的一页。我的朋友们，鲁黎黎，瞿铮，刘志博，蒋子贤，陈雪，杨杰，李士广，韩嵩哲，覃岭，白显洲，刘一周，你们就如同我在法国的家人。岂曰无衣，与子同袍。愿大家都有如意的未来。

谢谢大家。

List of abbrevations

3D:	three -dimensional
AA:	amino acid
aMb:	apo-myoglobin
AP-MS:	affinity purification mass spectrometry
ATP:	adenosine-5'-triphosphate
CID:	collision-induced dissociation
CsmA:	chlorosome protein A
CTD:	C-terminal domain
CX-MS:	chemical cross-linking mass spectrometry
DMF:	dimethylformamide
ECD:	electron capture dissociation
eIF:	eukaryotic initiation factor
EM:	electromicroscopy
ESI:	electrospray ionization
ETD:	electron transfer dissociation
FAB:	fast atom bombardment
FMO protein:	Fenna–Matthews–Olson protein
FT-ICR:	fourier transform ion cyclotron resonance
GDP:	guanosine diphosphate
GTP:	guanosine-5'-triphosphate
HDX-MS:	hydrogen/deuterium exchange mass spectrometry
HPLC:	high performance liquid chromatography
IM:	ion mobility
IRES:	internal ribosome entry site
MALDI:	matrix-assisted laser desorption ionization
Met-tRNA$_i^{Met}$:	initiator methionyl tRNA charged with methionine
MFC:	multi-initiation factor complex
mRNA:	messenger ribonucleic acid
MS:	mass spectrometry
MSP1D1:	membrane scaffold protein 1D1
mTOR:	mammalian target of rapamycin

NH:	amide hydrogen
NMP:	N-methylpyrrolidone
NMR:	nuclear magnetic resonance
NTD:	N-terminal domain
PABP:	poly(A)-binding protein
PIC:	pre-initiation complex
PTM:	post-translational modification
RNA:	ribonucleic acid
SAXS:	Small-angle X-ray scattering
SFC:	supercritical fluid chromatography
TAP:	tandem affinity purification
TC:	ternary complex

Content

General Introduction **1**

Part I Introduction **7**

Chapter 1 Structural analysis of protein complexes by MS **9**

1 The structure of proteins 12

1.1 The secondary structure: α-helix, β-sheet and motif 12

1.2 The tertiary structure: protein folding process 13

1.3 The quaternary structure: protein complex and assembly 14

2 Probing protein structure by mass spectrometry 14

2.1 AP-MS 14

2.2 Native MS 15

2.3 Chemical cross-linking MS 17

2.4 Labeling techniques 17

2.5 Perspectives of combining the structural MS strategies 18

3 Hydrogen/deuterium exchange 18

3.1 Brief history of HDX studies 18

3.2 General principle of hydrogen exchange 19

3.2.1 The exchange of free-exposed hydrogens 19

3.2.2 The exchange of amide hydrogen 22

3.3 Incorporation of deuterons: labeling strategies 24

3.3.1 Continuous labeling 24

3.3.2 Pulse labeling 25

3.4 Measuring the HDX 26

3.4.1 Bottom-up approach 28

3.4.1.1 Back exchange 28

3.4.1.2 Choice of protease 29

3.4.1.3 Chromatography 30

3.4.2 Top-down approach 31

3.4.3 Mass measurement and data processing 32

3.5 Application of HDX-MS in protein structural studies 33

3.5.1 Protein folding 33

3.5.2 Protein interactions 34

3.5.3 Aid in structure elucidation and crystal production 36

3.5.4 Quality control 36

3.6 The on-going improvements and perspectives for HDX-MS 36

3.6.1 Automated experimental work-flow and data processing 37

3.6.2 Combining the bottom-up and top-down approach 38

3.6.3 High performance mass spectrometer 39
4 HDX vs. other structural tools 41
References 43

Chapter 2 The eIF3 complexes: functions, structure, production and purification 49

1 Pathway of eukaryotic translation initiation.....51
2 The role of eIF3 in translation initiation.....52
3 The structure of eIF3 53
3.1 MS revealed the overall interaction network of human eIF3.....54
3.2 The atomic structure of yeast eIF3i:b(654-700).....56
4 Synthesis and purification of proteins of interest 59
References 62

Part II Results.....65

Chapter 3 Method development in the bottom-up HDX approach 67

1 Initial workflow of HDX experiments and data processing steps 69
1.1 The HDX-MS experiments 69
1.2 Initial data processing workflow.....70
1.2.1 Retention time correction 71
1.2.2 Establishment of the peptide reference list 72
1.2.3 Peak assignment for deuterated peptides 74
1.2.4 Calculation of deuteration level and kinetics plotting 77
2 Method improvements in the establishment of peptide reference list 77
2.1 Insufficiency of mass accuracy for peptic peptide identification 78
2.2 Pepsin specificity as another approach for peptic peptide identification.....80
2.2.1 Probability of occurrence of a peptide 80
2.2.2 POc filter and mass error filter 83
2.2.3 Resolving ambiguous assignments.....85
2.2.4 Reproducibility control 87
2.2.5 Method validation by MS/MS and evaluation 87
3 Method improvements in the processing of deuterated sample data 92
3.1 Mass recalibration 92
3.2 Program upgrades for peak assignments of deuterated samples 92
References 94

Chapter 4 HDX analysis of the eIF3 complexes 95

1 Peptide selection, centroid mass verification and deuteration level calculation for deuterated samples 97

2 The eIF3i monomer and complexes of eIF3i:bC3 and eIF3i:gC1ΔC 102
2.1 HDX of eIF3i monomer 102
2.2 HDX of eIF3i in complex with eIF3bC3 106
2.3 HDX of eIF3i in complex with eIF3gC1ΔC 118
2.4 Discussion 127
References 129

Part III Methods 131

Chapter 5 Materials and methods 133

1 NanoLC system in HDX-MS 135
2 FT-ICR MS 137
2.1 The principles of FT-ICR MS 138
2.2 Factors ensuring the high performance of FT-ICR MS 140
2.3 The evolution of instruments during this work 141
2.3.1 APEX III 141
2.3.2 APEX Q 142
2.3.3 SolariX 142
3 Experimental protocols 144
3.1 Protein production and purification 144
3.1.1 Production of eIF3i, eIF3bC3 and eIF3gC1ΔC 144
3.1.2 Protein purification 145
3.1.3 Complex purification 146
3.1.4 Protein storage 146
3.2 Hydrogen/deuterium exchange reaction 146
3.3 NanoLC-MS 146
3.4 LC-MS/MS 148
References 150

Conclusion and perspectives 151

General introduction

Mass spectrometry has entered the biology and biochemistry laboratories since the end of the 1980's with the introduction of MALDI and ESI as powerful ionization sources. These ion sources provided instruments with ability to produce ions of biological molecules with high efficiency, and thus allowed the analysis of high molecular weight biomolecules. The most routinely used technique relies on the measurement of the mass of peptides within a peptide digest, allowing efficient identification of proteins, provided a protein or genomic database is available. The peptide sequencing capabilities provided by MS/MS instruments allows the gain of sequence information on these peptides which can then be used to confirm identification, even in complex situations such as in the presence of protein post-translational modifications (PTM analysis). But these "soft" ionization techniques have also been shown to allow resolving large intact macromolecular assemblies, including large protein complexes. Thus mass spectrometry can be used as a tool not only for protein identification and characterization but also for protein structure analysis. The major advantage of mass spectrometry in this area relies on the extreme sensitivity of the mass measurements: although it is s destructive technique, only minute quantities (on the order of 10 atomoles) of biological material are required to perform an analysis on the most sensitive instruments currently available.

Another major advance in mass spectrometry has been the huge increase in mass resolving power and mass accuracy. Prior to the 2000, only FT-ICR instruments could provide routine resolutions over 100 000 and mass accuracy on the order of 1 ppm. The development of the Orbitrap™ mass analyser by Thermo has been a major breakthrough, allowing high mass accuracy and high resolution for much less cumbersome instruments. Currently, ESI-TOF analyzers are also in constant progress and mass accuracies on the order of 1 ppm are routinely achieved on a range of instruments. It is thus of interest to develop techniques that make use of established protocols but also include increased mass accuracy and high resolution in the protocol.

The development of improved techniques is not a goal in itself, and a mass spectrometry technique has to demonstrate its problem-solving capabilities to be widely used. Thus this thesis has also relied on attempts to gain insights in a major pathway in biology, protein synthesis. The initial aim was to focus on eIF3, one of the protein complexes involved in translation initiation: translation initiation involves a number (on the order of 10 for simple eukaryotic systems) protein complexes which interact together in a well-orchestrated molecular machinery, opening the way for efficient and accurate expression of proteins from

the mRNA. The eIF3 complex is itself a heteromultimeric protein complex, comprised of 5 stoechiometric units in the budding yeast *Saccharomyces cerevisiae* and on the order of 13 subunits in *Homo sapiens*. The crystal structure of this large (around 300 kDa) complex is not known yet, and only partial information is available on the assembly of subunits together. Even the contact regions between some subunits remains to be solved, and this seemed a good system on which to apply the mass spectrometry methods based on the measurement of amide H/D exchange (HDX) as developed in the DCMR group at the Ecole Polytechnique in the past years.

Thus this manuscript is organized to reflect the major advances undergone in the course of this three year work. In a first introductive part, a summary the key problems tackled in this thesis are presented. In a first chapter, the current status on the uses of mass spectrometry for protein complex analysis is reviewed. A focus is put on the HDX methods, which represent the core of the work presented in this thesis. The standard protocol is presented and discussed and recent developments, including the uses fragmentation techniques to gather improved resolution on site-specific exchange rate constants will be surveyed. In a second chapter, a review of the literature concerning the eukaryotic translation initiation factor is done. This review puts in perspective the current work with both prior knowledge, and also advances that occurred in the course of the thesis, including a crystal structure of partial domains which are very close to one of the systems that has been studied in the course of this work. Finally, the chapter will focus on the experimental procedures for the production of the protein complexes used in this work.

In a second part, the experimental results gathered throughout this work will be presented. This part is itself separated in two. Chapter 3 deals with the methodological developments that have been required to increase the size of complexes that could be addressed. As will be shown, the increase in size of the protein complexes made that mass accuracy alone was not sufficient to correctly assign each isotope distribution to a single peptide. In these conditions, the standard procedure would have been inclusion of an MS/MS step to validate individually each peptide. But after a change in instrumentation, we noticed that on the newer mass spectrometers, sensitivity gains led to the detection of a vastly increased number of peptides. Individual validation of each of these peptides was a tough task, and we sought alternatives to the MS/MS validation step. The major limitation arises from the lack of specificity of pepsin in conditions used for HDX mass spectrometry. But using prior work of another group, we devised a predictive tool that can assign a statistical score to each putative peptide, which we

termed “Probability of occurrence” (POc). Combining this score with high mass accuracy, it seemed possible to correctly assign peptides without requiring MS/MS validation. Of course, in order to demonstrate the validity of this POc scoring system, a thorough comparison of datasets with the MS/MS reference method has been performed, and showed very few instances for which the POc / accurate mass combination led to falsely assigned peptides. As a consequence of this method, it should also be noted that not requiring MS/MS validation allows inclusion of a much larger number of peptides within the reference dataset: as noticed in this work as well as in earlier work, pepsin digest peptides do not necessarily fragment as well in collision induced dissociation conditions as the more common tryptic peptides usually present in proteomics studies. Thus a number of peptides are present in the sample but are not taken into account when an MS/MS validation step is required because their fragmentation results are not sufficient to meet the validation criteria.

In addition to this central technical development, this chapter will also deal with other improvements that have been performed on the processing of the data. The huge datasets that are obtained from nano-LC/FT-ICR MS cannot be manually processed. An in-house developed software has been used to automate the task of spectrum analysis as much as possible. The whole workflow is described and the critical steps which still require manual intervention of the scientist are identified and the limitation of automation for these steps will be discussed in this chapter.

Chapter 4 is devoted to the biological results achieved in the course of this thesis on the eIF3 subunits interaction domain. The first system which will be discussed is the eIF3i:bC3 interaction region. In 2012, a crystal structure of this binding region was obtained while HDX MS data was acquired in our group. Although a bit disappointing since we expected to be able to determine the interaction region based on HDX MS measurements, this comparison structure allows a good critical discussion on the HDX results for this system. Furthermore, this crystal structure also helped a lot in providing a first proposed structure for the eIF3i subunit. The existence of this structure helped a lot in the discussion of the results for the other system which was explored in this thesis, the eIF3i:gC1ΔC binding region, which will constitute the second part of this chapter. Unfortunately, at the time of writing this manuscript, only preliminary results are available on the eIF3i:b:g ternary complex: thus only a sketch of the global interaction region between the three subunits will be presented in this manuscript.

In a final part, the manuscript will present a complete material and methods section. It must be pointed out that in the course of this work, the auther had the opportunity to work in close collaboration with two groups: in the biochemistry laboratory (Laboratoire de Biochimie), all the steps required for obtaining the biological samples were performed, which included protein expression, purification, production and purification of the protein complexes. In the mass spectrometry laboratory (Labratoire des Mécanismes Réactionnels), the HDX kinetics as well as the mass spectrometry analysis and data processing were performed. This leads to a very complete work, and all the relevant protocols as well as the technical details are provided in this manuscript.

Part I

Introduction

Chapter 1

Structural analysis of protein complexes by MS

1 The structure of proteins 12
1.1 The secondary structure: α-helix, β-sheet and motif 12
1.2 The tertiary structure: protein folding process 13
1.3 The quaternary structure: protein complex and assembly 14
2 Probing protein structure by mass spectrometry 14
2.1 AP-MS 14
2.2 Native MS 15
2.3 Chemical cross-linking MS 17
2.4 Labeling techniques 17
2.5 Perspectives of combining the structural MS strategies 18
3 Hydrogen/deuterium exchange 18
3.1 Brief history of HDX studies 18
3.2 General principle of hydrogen exchange 19
3.2.1 The exchange of free-exposed hydrogens 19
3.2.2 The exchange of amide hydrogen 22
3.3 Incorporation of deuterons: labeling strategies 24
3.3.1 Continuous labeling 24
3.3.2 Pulse labeling 25
3.4 Measuring the HDX 26
3.4.1 Bottom-up approach 28
3.4.1.1 Back exchange 28
3.4.1.2 Choice of protease 29
3.4.1.3 Chromatography 30
3.4.2 Top-down approach 31
3.4.3 Mass measurement and data processing 32
3.5 Application of HDX-MS in protein structural studies 33
3.5.1 Protein folding 33
3.5.2 Protein interactions 34
3.5.3 Aid in structure elucidation and crystal production 36
3.5.4 Quality control 36
3.6 The on-going improvements and perspectives for HDX-MS 36
3.6.1 Automated experimental work-flow and data processing 37
3.6.2 Combining the bottom-up and top-down approach 38
3.6.3 High performance mass spectrometer 39
4 HDX vs. other structural tools 41
References 43

Before the development of the so-called “soft-ionization” techniques, mass spectrometry (MS) was used exclusively in the field of small molecules for the elucidation of unknown molecular composition. When it comes to the late 1980s, the development of “soft” electrospray ionization (ESI) and matrix assisted laser desorption/ionization (MALDI) techniques opened the door of biological MS, as bio-macromolecules can be transferred into the gas phase as intact ions, thereby making them amenable to detection by different types of mass analyzers. [1-4] Since then, mass spectrometry has been an indispensible tool in the development of proteomics study. According to the introductory book of *Proteomics,* [5] five categories of experiments can be considered as applied MS in proteomics: (1) protein identification, (2) protein quantitation or differential analysis, (3) post-translational modifications (PTMs), (4) protein-protein interactions, and (5) structural proteomics.

Among all the above aspects, MS is most commonly known for its capacity to identify proteins and monitor post-translational modifications (PTMs). Thanks to the large amount of protein sequence data available, the identification of several thousand proteins in a single study can nowadays be achieved upon a database search using the MS and MS/MS data from tryptic (or other protease) peptides. When this technology is applied carefully and correctly, the rapid, accurate, sensitive and reproducible characters that modern MS has shown in protein identification make it accepted as the routine method for such topics, and bring the broad expectations to generate high-coverage to complete proteome maps. [5-7] In the PTMs strategies, types and sites of modifications are identifiable for single or multiple protein(s). While typical bottom-up approach has been the most widespread method for standard PTMs study, complementary methods are also introduced with interesting performances. Top-down approach, though they require more advanced instrumentation and data processing software, can provide a more comprehensive overview of the protein sequence and the PTM pattern, where as middle-down approach that uses proteases cleaving less frequently-occurring amino acid residues to generate long peptides (5-6 kDa) has attracted increasing interests as it is particularly suitable to focus on regions in which major modifications reside. [8-9] With the still undergoing developments of these methods, the proteomic-scale study on PTMs and PTMs changes in biological processes has started to grow.

This work, on the other hand, focuses on another domain of proteomics studies, namely the structural proteomics. This involves the fundamental structure determination of proteins, as well as the dynamics and structural changes that proteins may undergo due to different

factors, e.g. the interaction with ligands, other proteins, the conformational changes during various physiological processes, etc.

1 The structure of proteins

Proteins are extremely complex molecules. Biochemists define the protein structure into several levels (Figure 1): primary structure refers to the amino acid (AA) sequence that forms the polypeptide. As the peptide chain folds, it forms certain localized arrangements of adjacent (but not necessarily contiguous) AAs that constitute secondary structure. The three-dimensional (3D) shape that the peptide chain assumes is termed the tertiary structure. When consisting of more than one polypeptide chains the protein has a quaternary structure. [10]

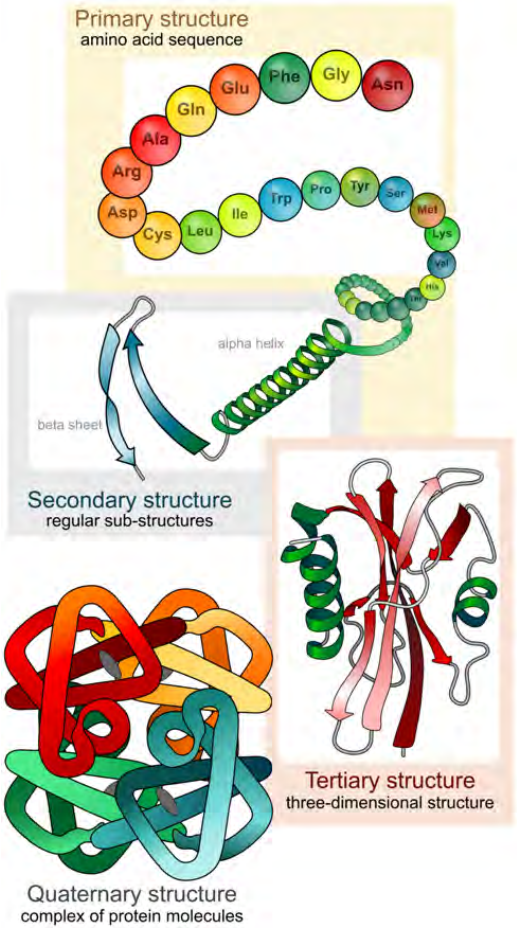


Figure 1 Brief illustration of 4-level protein structures. [11]

1.1 The secondary structure: α-helix, β-sheet and motif

In the secondary structure of proteins, peptide chains can fold into several repeating patterns. The most widely observed patterns are the α-helix and the β-pleated sheet. Both of them are

stabilized by localized hydrogen bonding between the carbonyl and N-H groups in the peptide backbone.

The α-helix is a rigid, rodlike structure that forms when a polypeptide chain twists into a right-handed helical conformation. Hydrogen bonds form between each N-H of amide group and the carbonyl group of the amino acid four residues away, while side chains extend outward of the helix. Each turn in the helix consists of 3.6 residues and the distance between the same positions of two turns is 0.54 nm.

β-sheets are formed when two or more peptide chains line up side by side, with each chain segment termed as β-strand. The stabilizing hydrogen bonds form between the backbone amide N-Hs and the carbonyl groups in the adjacent chains. β-sheets can be either parallel, with both chains N-C oriented in the same direction, or anti-parallel with the two chains oriented heads to tail.

Combinations of several α-helix(es) and the β-pleated sheet(s) structures form the so-called supersecondary structures or motifs, e.g. the βαβ unit with two parallel β-sheets connected by an α-helix, or the β-meander unit with two or more consecutive antiparallel β-sheets linked together by direction-reversed hairpin loops. These structures are usually stabilized by hydrophobic interaction involving the side chains.

1.2 The tertiary structure: protein folding process

As globular proteins fold into their unique native structures and prosthetic groups, if any, are inserted, they assume the tertiary structure. The formation process, protein folding, makes highly organized structures out of nascent polypeptides. This process occurs as a consequence of the interactions between the AAs side chains. Main interactions that stabilize tertiary structure include: 1) hydrophobic interactions, when hydrophobic side chains are brought closer as they are excluded from water, while highly ordered water molecules are released from the interior of the tertiary structure. Yet a few water molecules may remain within the core of the folded protein and form up to 4 hydrogen bonds with the peptide backbones and increase the flexibility of some of the internal interactions, which may be crucial in the binding of ligands. 2) electrostatic interactions, involving the strong interactions between ionic groups of opposite charge (salt bridges) in water-free regions, and weak interactions between ion-dipole, dipole-dipole, etc. These interactions can be significant in holding the interiors of the tertiary structure, as well as contributing to the interactions of the

subunits/ligands in the quaternary structure, as will be discussed later. 3) hydrogen bonds, which is not only the stabilization factor in the secondary structure, but also important and often observed in the interior and surface of the globular proteins. 4) covalent bonds, most prominently the disulfide bridges in many extracellular proteins, are significant in protecting proteins from adverse changes in pH or salt concentrations. 5) hydration, i.e. the dynamic hydration shells formed around the tertiary structures.

1.3 The quaternary structure: protein complex and assembly

A lot of proteins, especially those with high molecular weights, contain more than one polypeptide chains (subunits), and these subunits can be either identical or different. The quaternary structures, similar to the tertiary structures, are stabilized either by noncovalent interactions such as the hydrophobic effect, electrostatic interactions and hydrogen bonds, or by covalent cross-links such as disulfide bridges. It is worth noticing that the assembling of subunits is often triggered by a conformational change that is affected by the binding of ligands, and such conformational change is called allosteric transition.

2 Probing protein structure by mass spectrometry

Protein structures, as well as the formation and transition of protein structures, are essential to the understanding of various protein functions and metabolic processes. Proteins expressed in living organisms mostly function in the form of protein complexes, which are further organized in protein interaction networks in the living cell. [12] In order to understand the function of a protein, it is of great importance to obtain comprehensive knowledge of its structure and dynamics on the molecular level and preferably in its native state, i.e. in the solution phase, in a medium close to the cellular content. Structural MS has drawn increasing attentions and shown great successes on these topics since the introduction of ion sources adapted for biomolecules. Typical structural MS techniques includes affinity purification mass spectrometry (AP-MS), chemical cross-linking mass spectrometry (CX-MS), MS of intact assemblies (native MS), and labeling techniques such as hydroxyl radical labeling and hydrogen/deuterium exchange mass spectrometry (HDX-MS).

2.1 AP-MS

AP-MS uses the affinity purification technique, one that has been commonly applied by biochemists. In AP, the protein of interest and its associated interactors are selectively

purified using either antibodies targeted towards this protein or an affinity tag introduced in the protein of interest. After pull down the proteins present are then analyzed and identified by proteomics-based MS (Figure 2 D). [13-15] Using AP-MS, it is therefore possible to investigate the connectivity within a complex or establish whole protein interaction networks. AP-MS is nowadays probably the method of choice when looking at protein-protein interactions on a global scale. With the application of the powerful tandem affinity purification (TAP) protocols and the integration of other strategies such as quantitative interaction proteomics, elegant works have been reported on s olving large-scale protein network architecture, for example the stoichiometry and architecture of the cullin-RING ubiquitin ligase network was revealed by a study which combined AP-MS with multiplex absolute quantification technology. [16]

2.2 Native MS

In native MS, intact complexes of up t o several MDa are no l onger hydrolyzed as in conventional MS-based proteomics studies, but directly ionized and analyzed by mass spectrometry (Figure 2 E). By careful manipulation of the experimental conditions of pressure and accelerating voltages, the target protein assembly can be ionized in the intact form as a whole without dissociation in the travel through the instrument. [12, 17-18] The overall stoichiometry is the initial information obtainable from the native MS, yet the connectivity of the different subunits are also important information achievable when partial dissociation is performed in controlled conditions. Native MS is often combined with proteomics MS to identify each of the subunits in the assembly, ion mobility to understand the topological structure of the global complex, and modeling approaches to stitch together information obtained through various techniques, and has shown powerful ability in a number of complicated topics, including studies of notoriously difficult targets such as membrane proteins, [19-20] and investigations of ligand binding processes such as the report of bound lipids and the effects of nucleotide binding of ATPases. [21]

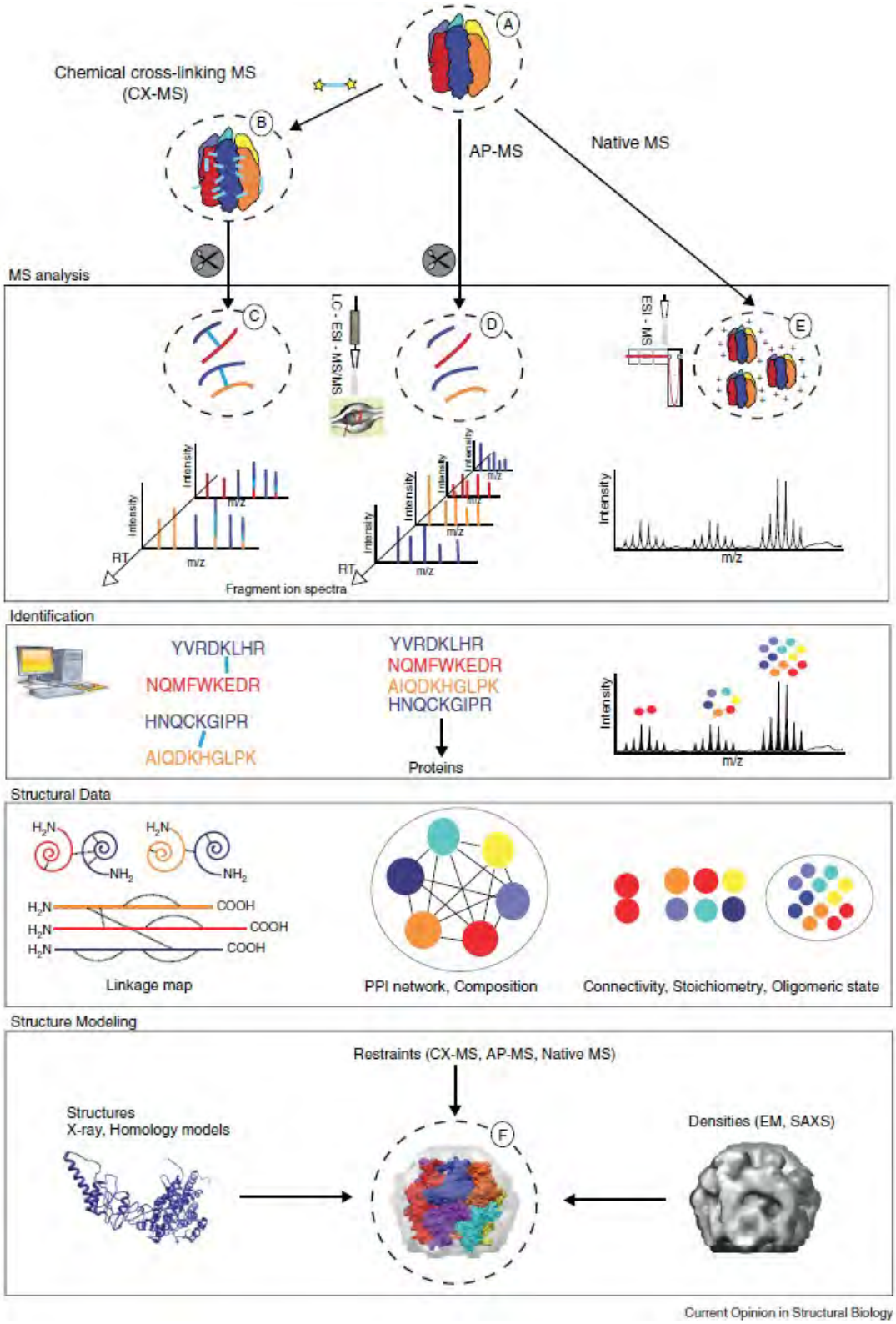


Figure 2 Selected mass spectrometry based methods for structural biology. (A) Purification of a protein complex by a biochemical approach. (B-C) Chemical cross-linking MS (CX-MS) workflow. (D) Affinity purification coupled to MS (AP-MS) workflow. (E) Native MS workflow. (F) Integration of diverse structural data by computational modeling. Thereby a model (or an ensemble of models) is searched that satisfies the structural data. [12]

2.3 Chemical cross-linking MS

Chemical cross-linking MS (Figure 2 B, C) is based on the fact that in proteins or protein complexes, amino acid residues in spatial proximity (20-30 Å) can be physically connected by covalent bonds via a cross-linking reagent. Target proteins or protein complexes are chemically cross-linked in solution under near-physiological conditions, enzymatically hydrolyzed to peptides, and the resulting peptides are analyzed by mass spectrometry. The cross-linkers can be either cleavable, so that the linked peptides are uncoupled before or during the mass spectrometric fragmentation process, or in more common cases non-cleavable, where ions of the cross-linked peptides are fragmented as single chemical entities. [12, 22] Various analytical enrichment strategies have been developed to increase the detectability of the cross-linked peptides in the complex peptide mixture, which is crucial to the efficiency of the technique because both the amount and the number of cross-linked peptides are vastly lower as compared to the abundance and number of non cross-linked peptides, and facilitate the latter data processing steps to the MS data. [23-25] With the help of these efforts, successful applications of CX-MS integrated with other techniques to challenging large macromolecular assemblies have been reported: recent examples include the determination of the architecture of the 2.5-MDa 26S proteasome using CX-MS together with cryo-electron microscopy (EM) and X-ray crystallography [26], and the deciphering the subunit topology of the TRiC/CCT chaperonin using CX-MS combined with molecular modeling. [27-28]

2.4 Labeling techniques

The local structures of the proteins are studied conventionally by X-ray crystallography and NMR. Yet different labeling techniques combined with MS measurement provide alterative pathway towards the structural information as well as the changes and dynamics of protein structures. Hydroxyl radical labeling uses the hydroxyl radicals (•OH) to induce covalent modifications at solvent-exposed amino acid side chains, which is detected as a +16 Da mass shift. The difference in the reactivity of individual side chains, keeping notion that it varies according to their chemical nature, reflects the solvent accessibility of their specific location. [29-32] The HDX technique, on the other hand, probes the protein structure with the information of the backbone amide hydrogens. The details of this technique will be discussed in the later part of this chapter.

2.5 Perspectives of combining the structural MS strategies

Increasing attention has been paid to the combination of the above approaches as each of them has unique strengths and weaknesses. AP-MS is able to detect subtle changes in interaction partners and dynamics and is as well able to deal with lowly expressed complexes. MS of intact assemblies is able to assess the absolute stoichiometry of the intact complex and multiple subcomplexes, yet both of them lack the ability to provide finer structural details on the amino acid or pseudo-amino acid scale resolution. On the other hand, CX-MS provides peptide resolution on t he interaction site of two subunits, and HDX-MS (which will be presented in the next section) reveals protein conformations and dynamics in seconds to hours time scale and down to single-amide resolution, but both methods cannot probe the overall stoichiometry or higher oligomeric state. Therefore it is believed that these approaches are complementary and their combination could generate complete interaction maps on global protein assemblies (Figure 2 F), and furthermore, enable subsequent merging of such datasets with topological information from IM-MS and cryo-EM results. [12, 15, 17]

3 Hydrogen/deuterium exchange

Protein molecules contain a huge amount of hydrogen atoms. Part of these hydrogen atoms are bound to carbon atoms through stable covalent bonding. Whereas other hydrogen atoms, those in O–H, N–H, and S–H groups, are labile due to the higher eletronegativities of the binding atoms, and they can exchange with the surrounding water. HDX targets all labile hydrogen atoms.

3.1 Brief history of HDX studies

The study of protein hydrogen exchange dates back to the pioneering work of Linderstrøm-Lang and his coworkers at the Carlsberg Laboratories in Copenhagen in the 1950s. [33-34] Although a lot of the early data from these days were incorrect, Linderstrøm-Lang inferred the basic dynamic mechanisms of the hydrogen exchange processes and suggested the equations that govern measurable HDX, which are still accepted today.

The extraordinary potential that hydrogen exchange offers was, however, challenged for a long time by the difficulty of accurately measuring the incorporated deuterium. The development of multidimensional NMR methods first provided an efficient way to measure HX in rather small protein molecules on single amide level, as the amide proton peaks (from

hydrogen) disappear after deuteration because ^{2}H has a spin of 1and ^{1}H has a spin of ½. [34-36].

Alternatively, due to the mass shift from 1.0078 Da of ^{1}H to 2.0141 Da of ^{2}H, the exchange of deuterium can be measured by mass spectrometry. In 1991, Katta and Chait reported the first application of MS-based HDX [37], where they probed different conformers of ubiquitin by HDX-ESI-MS. The first HDX MS experiments to include protease (pepsin) digestion and HPLC separation were reported 1993. [38] In this work, exchange reaction was quenched at different time points by acidification and temperature decrease, followed by pepsin proteolysis, HPLC separation and FAB-MS analysis. This prototype develops into the standard bottom-up approach without major changes in the basic experimental design, but it becomes increasingly powerful thanks to the vast advances of instrumentation and software. The development of fragmentation methods, electron capture dissociation (ECD) [39] and electron transfer dissociation (ETD) [40], leads to the alternative top-down approach.

3.2 General principle of hydrogen exchange

3.2.1 The exchange of free-exposed hydrogens

The mechanism of such reaction has been studies ever since the first reports of the HDX method. [33, 41-43] Starting from the fundamental of the proton transfer reaction, now the widely accepted pathway can be described in three steps (Equation 1): formation of a hydrogen-bonded complex via diffusional collisions between a proton donor (X-H) and an acceptor (Y); rapid equilibrium redistribution of the proton between the proton donor and acceptor within the complex, and dissociation of the hydrogen-bonded complex. [42]

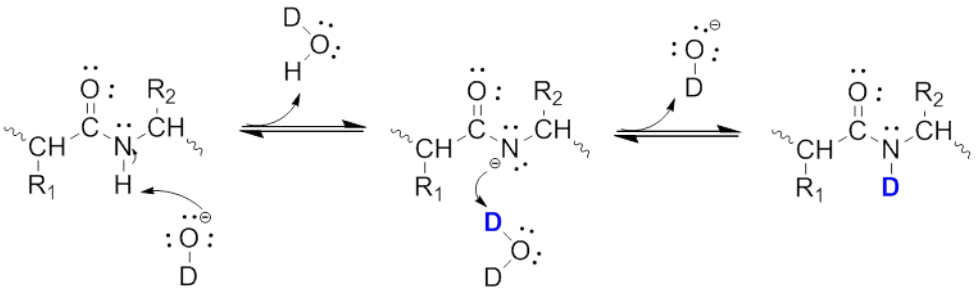
$$X{-}H + Y \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} [X{-}H\text{--}Y \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} X\text{--}H{-}Y] \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} X+Y{-}H \qquad (1)$$

$$k = k_1\left(\frac{10^{\Delta pK}}{1+10^{\Delta pK}}\right) \qquad (2)$$

The overall rate constant can be approximated as Equation 2, where ΔpK represents the difference between the pK value of the proton acceptor and donor. [44-45] Therefore, if the acceptor is a much weaker acid than the donor, i.e. $pK_{acceptor} \gg pK_{donor}$, the rate constant ($k$) of the proton transfer will be equal to that of the diffusional collision k_1, meaning that every collision between X-H and Y leads to a successful transfer. On the contrary, if the transfer proceeds from a weaker to a stronger acid, the transfer will be much slower than the diffusion.

This explains the different behavior of the hydrogen atoms in different groups in proteins and allows delineating three groups of hydrogen atoms. (1) Hydrogens bonded to carbon show a very high pKa and thus these hydrogens hardly exchange with surrounding hydrogens. (2) Hydrogens bonded to heteroatoms of -N, -O and -S on polar side chains or the N/C termini display much a lower pKa values for deprotonation (lower than 13 at pH 7, 25 °C) than the pKa (15.7 at pH 7, 25 °C) for OH^- protonation, and therefore they are easily exchanged with OH^-. (3) Hydrogens at backbone amide groups, however, has a much higher p*K* value (~18), in other words, the exchange of such hydrogens proceeds from a weaker to a stronger acid, and thus is much slower. This explains why the hydrogen exchange that is most often measured in proteins is that of the backbone amide hydrogen (NH).

(a) Base catalysis



(b) Acid catalysis

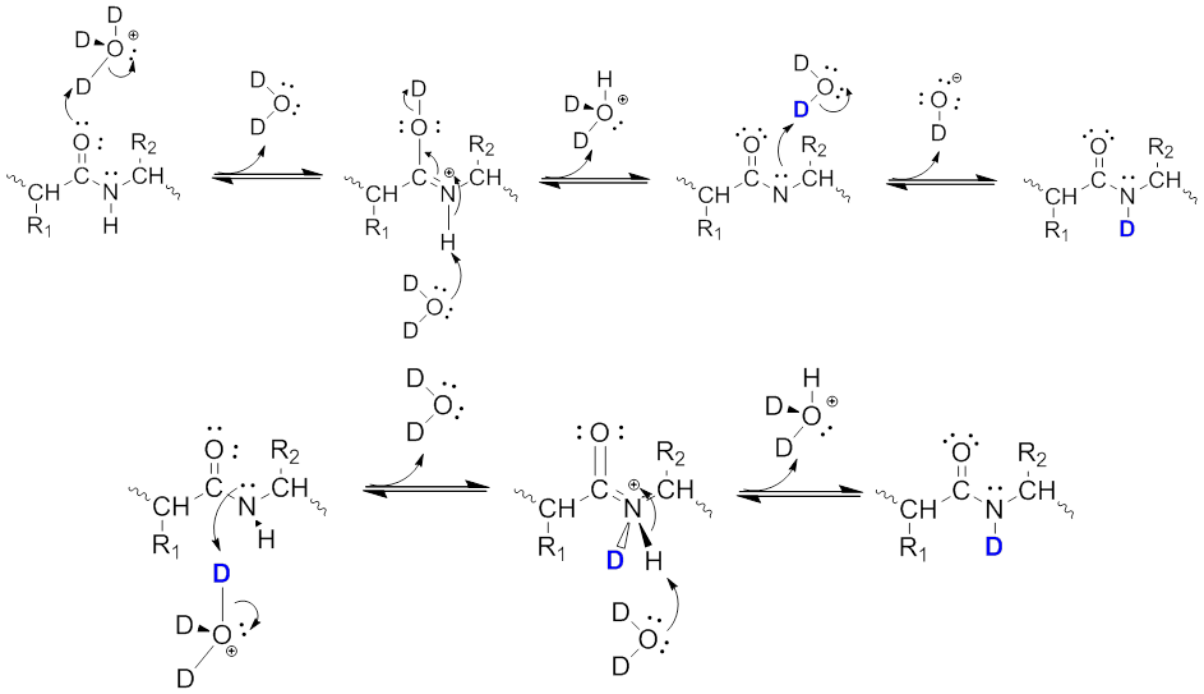


Figure 3 mechanisms of hydrogen/deuterium exchange in peptide/protein backbone amides.

The H/D exchange on peptide/protein backbone amide groups may proceed via base- or acid-catalyzed mechanisms. [46-47] The base-catalyzed pathway, first proposed by Berger et al. in 1959, has been quickly accepted without much debate. Exchange begins from abstraction of the amide proton by a hydroxide ion and the formation of an imidate anion, which is subsequently re-protonated to complete the transfer (Figure 3 (a)). This is the more important mechanism for the physiological conditions in which proteins exist, yet it is also possible that the exchange happens via the acid-catalyzed pathway. The latter, however, has gone through a long way of discussions between two distinct mechanisms. [46, 48-49] Nowadays the more commonly accepted pathway is the so-called O-protonation (Figure 3 (b)): the protonation of the carbonyl oxygen happens first, lowering the p*K* for deprotonation of the amide group. The amide hydrogen is therefore removed by a water molecule to produce the imidic acid intermediate that extracts a deuterium and returns to an amide. This mechanism is more reasonable due to the fact that on the backbone amide, the more basic group is the oxygen rather than the nitrogen. Whereas an earlier-raised pathway, N-protonation (Figure 3 (c)), starts directly from the protonation of the amide nitrogen. This mechanism is mostly inhibited but still not strictly impossible. In sum, the base-catalyzed exchange occurs via the imidate anion, and is related to the structural information of the amide nitrogen. The acid-catalyzed exchange, on the other hand, is predominantly occurring via O-protonation pathway and is influenced by the amide carbonyl. Thus exchange rate constant of freely exposed backbone amide hydrogens can expressed as:

$$k_{ex} = k_{acid}[D_3O^+] + k_{base}[OD^-] + k_{water} \quad (3)$$

According to reference studies using small peptides [33, 48, 50], at low salt conditions and 20 °C k_{base} can be as much as eight orders larger than k_{acid} and eleven orders larger than k_{water}. This implies that the base-catalyzed exchange is predominant at physiological conditions.

Plotting log(k_{ex}) versus pH leads to a V-shaped curve (Figure 4), with a minimum exchange rate at pH 2.5 - 3.0. This can be deduced by differentiating Equation 3 as

$$\frac{\partial k_{ex}}{\partial[\mathrm{D_3O^+}]} = k_{acid} - k_{base}\frac{K_w}{[\mathrm{D_3O^+}]^2} \quad (4)$$

and setting to zero, where K_w represents the ionization constant of water. On either side of the minimum pH, the slope of $\Delta\log(k_{ex})/\Delta\mathrm{pH}$ is approximately equal to 1 for higher pH and to -1 for lower pH, indicating a close to first-order pH dependence of k_{ex}. This reflects that each

shift of one pH unit from the pH_{min} changes the exchange rate by 10-fold, and that the exchange is nearly four orders slower at pH 2.5 than at pH 7.

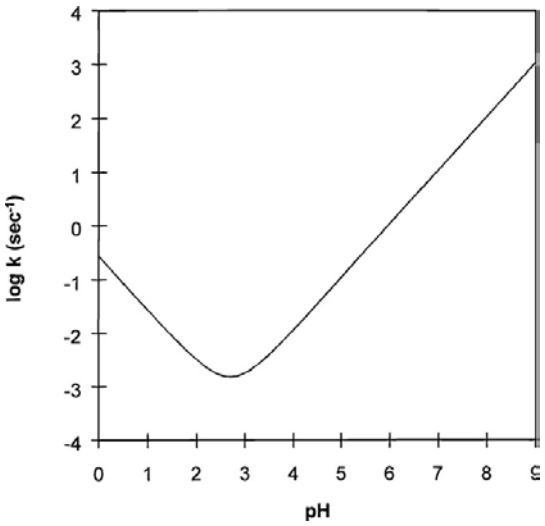


Figure 4 Dependency of the exchange rate k_{ex} on solution pH for amide hydrogens within a polyalanine peptide. [51]

Another feature that affects significantly the exchange rate is the temperature. At fixed pH, reducing the temperature alters the concentration of OD^- through modification of the K_w. In addition, the diffusional collision rate k_1 also changes in accordance to temperature. The exchange rate reduces approximately one order of magnitude when the temperature lowers from 20 to 0 °C.

3.2.2 The exchange of amide hydrogen

In native (folded) proteins, a few hydrogen atoms are fully solvent-accessible that can readily exchange at a fast and often similar rate. In contrast, some of the NHs exchange quickly and others exchange much more slowly with half-lives up to months. [9, 41, 45, 52-54] This difference in the exchange rate can be demonstrated through the following kinetic model:

$$NH\ (protected) \underset{k_{cl}}{\overset{k_{op}}{\Longleftrightarrow}} NH\ (unprotected) \overset{k_{ex}}{\Longleftrightarrow} ND\ (unprotected) \overset{k_{op}}{\Longleftrightarrow} ND\ (protected) \qquad (5)$$

where k_{op} and k_{cl} are the rate constants for the unfolding (opening) and refolding (closing) events that expose amide hydrogen to the solvent. This assumes that amide hydrogen requite to be in a open state in order to exchange with the surrounding water. Although there is also experimental evidences that the exchange reaction can occur when the protein is folded [55-56], most studies have been based on the above model and assume that the exchange-

competent state for an amide hydrogen is either a fully or partially unfolded state. In these conditions, the overall exchange rate constant k_{HDX} is deduced as:

$$k_{HDX} = \frac{k_{op}}{k_{cl} + k_{ex}} k_{ex} \quad (6)$$

Two extreme regimes have typically been demonstrated: ordinarily proteins are stable in their native state so the refolding process is much faster than the exchange, $k_{cl} \gg k_{ex}$, and the k_{HDX} is therefore approximated as:

$$k_{HDX} = K_{op} k_{ex} \quad (7)$$

where $K_{op} = (k_{op}/k_{cl})$ is the equilibrium constant of the unfolding process, implying the probability of the exchange reaction occurring in a single unfolding event. This regime is referred to as EX2, which is more prevalent in physiological conditions than the other regime, EX1. The latter refers to the case when $k_{ex} \gg k_{cl}$, meaning that the unfolding event is rate limiting, i.e.

$$k_{HDX} = k_{op} \quad (8)$$

The EX1 exchange is often observed under protein denaturing conditions or by increasing the intrinsic exchange rate. The two kinetics are easily distinguishable by mass spectrometry [57]: EX1 regime gives rise to two distinct and separated mass envelope (the undeuterated and deuterated forms) while EX2 regime gives a gradual shift of the mass envelope with time.

An example of the EX1 and EX2 scheme is given in Figure 5, from a work on l ipid dependent changes of the membrane scaffold protein MSP1D1. [58] The HDX of the lipid-associated states (Figure 5, spectra on the right and red plots) shows a typical EX2 pattern, i.e. a single isotopic distribution. In the HDX of the lipid-free form, the peptide displays EX1 kinetics which can be distinguished by the presence of two isotopic distributions, e.g. in the spectrum of 5m. For the peptide with EX1 unfolding, each time point is fitted with two Gaussian distributions. The center of these distributions is used to plot the deuterium uptake for both the low (solid line, filled symbols) and high (dashed line, empty symbols) mass populations as shown in blue in Figure 5.

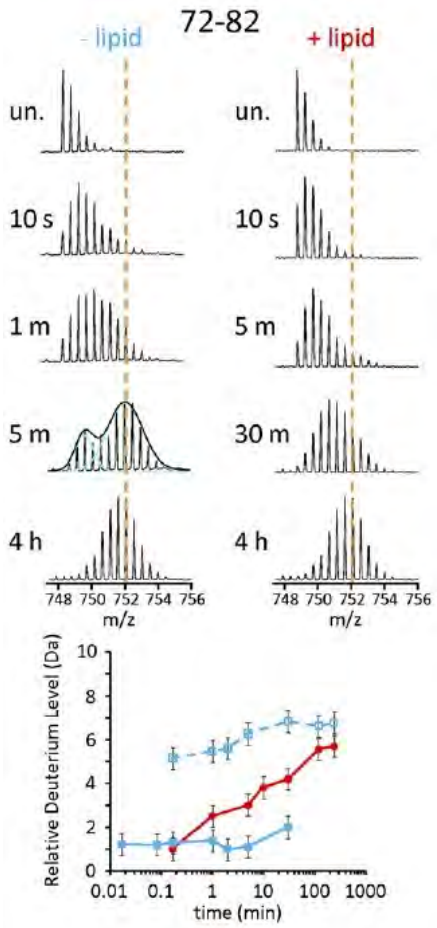


Figure 5 Mass spectra and deuterium uptake plots for peptide MSP1D1(72-82) in both the lipid-free (blue) and lipid-associated (red) states.[58]

3.3 Incorporation of deuterons: labeling strategies

Exposing a protein to a D_2O-containing environment leads to H-D replacements that increase the mass of the protein by one unit per exchange event, this is referred to as an ‘‘exchange-in’’ scenario, which will be mostly used throughout this thesis. Although less commonly used, it is also possible to conduct measurements in an ‘‘exchange-out’’ protocol. In the latter case the protein is first fully deuterated before placing it in H_2O, so that labeling occurs in the D-H direction. [54, 59] The protocols to induce the deuterium incorporation into the protein can be generally grouped into two types: continuous labeling and pulse labeling.

3.3.1 Continuous labeling

In the continuous labeling strategy (Figure 6), the exchange is initiated by placing the unlabeled protein into a D_2O-based solvent system (e.g., by a rapid dilution of 15-20 fold).

The final deuterium concentration after the dilution exceeds 95%, so that the exchange reaction is forced in only one direction, H→D. Unstructured and highly dynamic protein segments undergo fast exchange (blue and red colors represent protons and deuterons, respectively), while less dynamic segments exchange more slowly (Equations 5-8). At various time points from minutes to hours, aliquots are removed from the deuteration solution, and the exchange is quenched by rapid acidification.

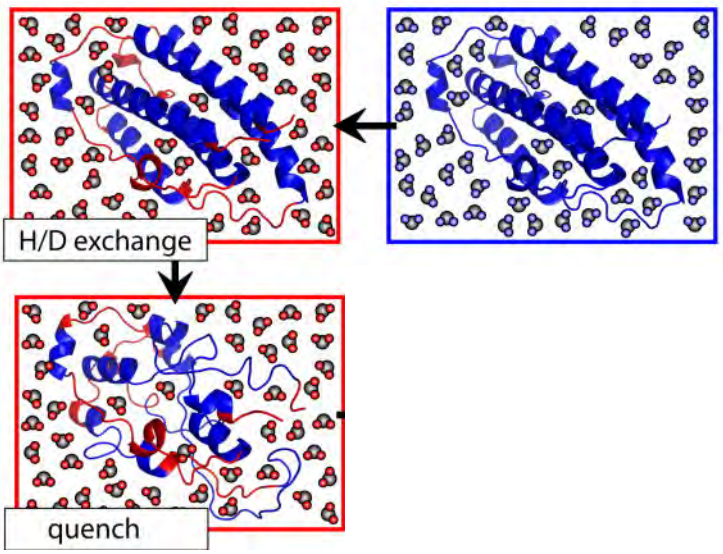


Figure 6 Schematic representation of typical continuous labeling work flow. [60]

The nature of the protein properties that are being monitored with continuous labeling has gone through different debates. It has always been demonstrated that HDX provides direct information on the solvent accessibility of amide groups. [52, 61-62] Yet according to detailed studies of the exchange patterns [54, 63], it seems more appropriate to state that HDX rates reflect the intactness of the hydrogen bonding network, i.e. solvent exposure and breakage of hydrogen bonds are both required for the exchange to occur, which should be met simultaneously in the "unprotected" conformation in Equation 5. It is also claimed that that HDX primarily reports on the stability and dynamics of secondary structure elements, because most intramolecular hydrogen bonds are associated with α-helices and β-sheets. [54, 64]

3.3.2 Pulse labeling

Pulse labeling strategy, although less commonly used, brings unique type of information: it detects and characterizes protein folding intermediates, short-lived species that exhibit structural features in-between the native and the unfolded states. [58, 65-66] In these studies

(Figure 7) an initially denatured protein is rapidly transferred to a suitable solvent environment, thereby triggering the refolding process. Subsequently, the protein is exposed to a brief (millisecond to several seconds) HDX pulse at a well defined time point during the reaction. A series of measurements is conducted by varying the time interval between folding trigger and pulse. In this way it is possible to obtain detailed insights into the temporal sequence of events that lead from the unfolded state to the native conformation. Pulsed HDX is typically handled under basic conditions (pD 8–10) to ensure that extensive isotope exchange can occur during the short labeling step, taking advantage of the pD-tunability of k_{ex} according to Equation 3. The labeling pulse may be quenched by rapid acidification to pH 2.5 or solvent evaporation during ESI in the case of on-line HDX-MS approaches. Often the folding intermediates contain regions that are hydrogen bonded to some extent, whereas other segments remain disordered. When the pulse conditions are well-controlled, the free amides become completely deuterated, whereas those that are even marginally protected remain in their N–H form. The ‘‘all-or-nothing’’ (protected or not) readout under such conditions allows folded and unfolded regions to be readily differentiated.

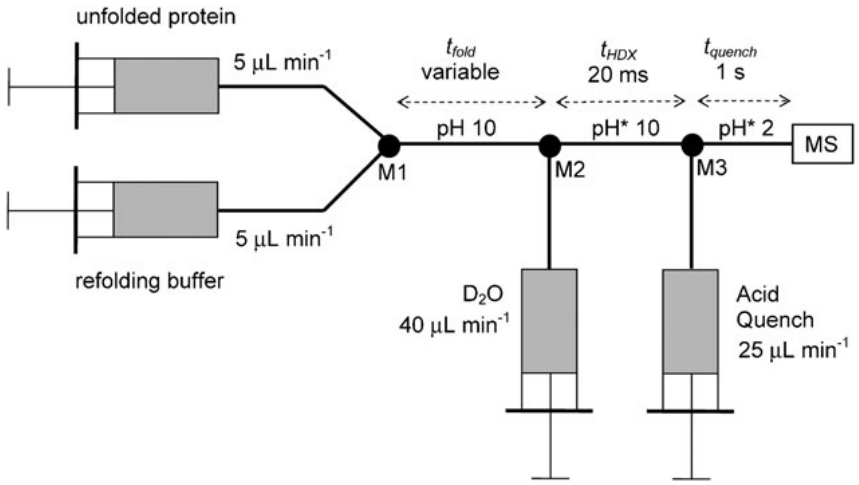


Figure 7 Example of typical pulsed HDX work flow. [67] Mixing of acid-denatured protein with ammonium hydroxide triggers refolding at mixer M1 in a H_2O environment. After a variable folding time (t_{fold}, ranging between 10 ms and 1 s), addition of D_2O at mixer M2 initiates a 20 ms labeling pulse. HDX is quenched by formic acid solution after M3 for 1 s, before the solution is infused into the ESI source of a mass spectrometer. pH-meter readings for D_2O-containing solutions are referred to as pH.*

3.4 Measuring the HDX

Any method that is sensitive to the properties of the isotopes of hydrogen can be used to monitor the HD exchange. The prior applications were the measurements of increasing

radioactivity of the labeled proteins after incorporation of tritium (T_2O) [44], and later NMR was introduced because deuterium is NMR silent and replacement of amide hydrogens by deuterons causes the disappearance of the corresponding ^1H NMR peaks. [68] The first report of protein HDX-MS appeared in 1991[37], thanks to the application of ESI.

Mass analyses of deuterated proteins can be conducted at the global protein level, which represent a straightforward approach for characterizing global changes in protein structure and stability, either when the protein is alone to monitor natural motions under equilibrium conditions or in response to ligand binding, complex formation, activation, presence of denaturants, etc.

To achieve spatial resolution, the deuterium levels of short peptide fragments have to be measured. This is achieved, before the mass measurement, either by proteolysis of the deuterated protein, or by fragmentation of the protein in gas-phase with ECD or ETD, referred to as the bottom-up and top-down approach, respectively (Figure 8).

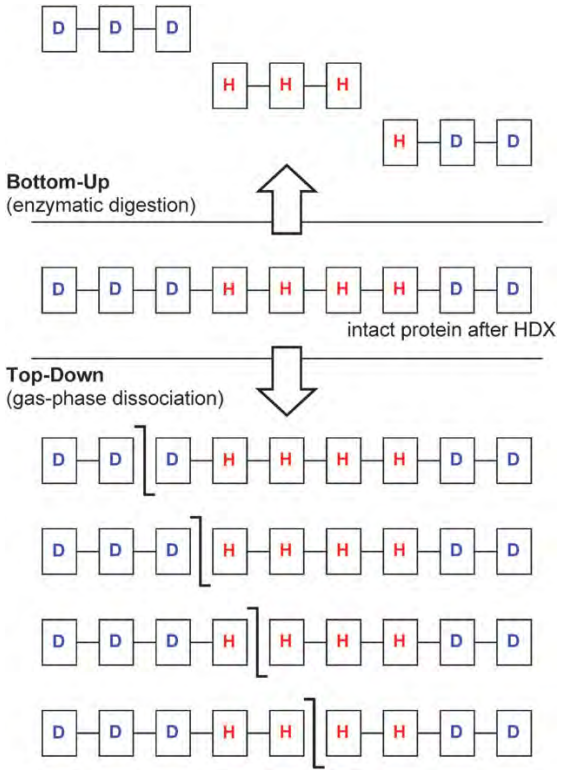


Figure 8 Brief illustration of the difference between bottom-up and top-down HDX-MS protocols. [54] In bottom-up experiments the protein is subjected to limited proteolysis after labeling. Each protein molecule is cleaved multiple times, generating solution-phase peptides. Top-down experiments involve the dissociation of intact proteins in the gas phase. Individual protein molecules usually experience only one cleavage event, resulting in complementary N- and C-terminal ion-pairs (c/z-ions in the case of ECD and ETD).

3.4.1 Bottom-up approach

The classical proteolysis/LC-MS approach, or the bottom-up approach of HDX-MS, was first reported in 1993. [38] It was originally developed as an extension of the NMR and radioactivity strategies, yet through the past decades this approach has been gradually improved thanks to the great instrumental improvements of the chromatography and mass spectrometry as well as the variations of the experimental protocols. Typically, continuous-labeling HD exchange is followed by acid quenching. [54] Subsequently, an acid-tolerant protease such as pepsin is added to induce limited proteolysis. The resulting peptides are desalted and separated by reversed phase LC, and analyzed by ESI-MS. The solution temperature is lowered to ~0 °C during quenching and all subsequent steps, and proteolysis and LC separation are conducted in a minimum of time (~30 min. or less), in order to minimize the back-exchange effect that some deuterons incorporated in backbone amides exchange back with surrounding hydrogen. In this process, no information is obtainable for side chains due to their very high back exchange rates.

3.4.1.1 Back exchange

It is critical for HDX-MS to retain the label as much as possible during the experimental process. However, as stated above, after deuteration the proteins have to undergo proteolysis, desalting, concentration, separation and mass measurement, all of which are carried out in protiated solution. Therefore, deuterated positions on amide groups as well as side chains may all revert to hydrogen. These back-exchange processes cannot be completely avoided. On the side-chains, due to the high k_{ex} value, all deuterons are back-exchanged to hydrogen. On the amide backbones, contrarily, the back-exchange can be minimized if careful control is taken during the experiment.

Primarily, a pH at which exchange is at the minimum state and a temperature as low as possible are maintained during the whole procedure after the HD exchange. According to Equations 3 and 4, the exchange rate reaches the minimum at pH ~ 2.5, whereas the quenching temperature is usually set 0-4 °C, the temperature limit at which the LC experiment may be carried out. Comparing to the condition at pH 7 and 25 °C, the exchange rate constant of the unprotected hydrogen/deuteron is five orders of magnitude lower in this quenching condition. The half-life for back-exchange, depending on the sequence of the proteins and proteolyzed peptides, varies from 30 to 120 min [45, 50-51], and therefore the

whole analysis process after the exchange reaction has to be completed as quickly as possible. It has been proven that if proteolysis and separation steps are kept under 12-15 min total, deuterium recovery can be as much as 85%, while the losses in the ESI interface during spraying can be kept to less than 5%. [66] Notably, although the measured increases in mass correspond predominantly to the exchange on the backbone amides, rich content of arginines can complicate the analysis because arginine has a side chain amide hydrogen with a minimum exchange rate close to that of the backbone NHs, so that the deuterons on these groups may not be completely washed away before the mass analysis.

Various methods have been reported to adjust or reduce the back-exchange. For example, the deuteration level can be corrected by compensating the extent of labeling loss for completely deuterated proteins. [38] The use of supercritical fluid chromatography (SFC) is also introduced as an alternative to HPLC, taking advantage of separating the peptides using CO_2 rather than protiated solvents as mobile phase, which partly reduces the possibility of back-exchange although the proteolysis step cannot be accomplished in supercritical CO_2. [69]

3.4.1.2 Choice of protease

The quenching condition, i.e. pH 2.5 and ~0 °C has to be maintained strictly during the whole analytical process. The protease to hydrolyse the proteins, therefore, has to be active at this chosen pH and temperature. Pepsin has always been the enzyme of choice. [38, 51] Pepsin is most active at pH 1.5-2.0. [70] Its optimal temperature is 37-42 °C, but is still active at 0 °C. Unlike trypsin and other strictly specific enzymes, pepsin shows only a preference of cleavage towards some amino acid residues. The weak specificity of pepsin results in both advantages and hindrances for the HDX MS methodology. The peptic peptides detected by mass spectrometry can achieve sequence coverage of 100% or higher, i.e. overlaps of peptide fragments are frequently observed. However, the identification of peptides from such a complicated mixture is not straightforward via cleavage position alone. Therefore, tandem mass spectrometry (MS/MS) has been considered as an indispensable step for the peptic peptide identification.

However, incomplete cleavage of large proteins and limited number of peptic peptides sometimes restrict the range of applicability when using pepsin alone. Additional proteases (protease type XIII from *Aspergillus saitoi* and protease type XVIII from *Rhizhopus*) were introduced in the HDX MS method, and have been demonstrated to increase the sequence coverage of the peptide mixture. [65, 71-72] These proteases are, similarly as pepsin, not

specific in the cleavage site, but they are not as efficient as pepsin under quench conditions. Yet combining the different proteases yields significantly higher sequence coverage and better spatial resolution as they each have slight preferences of the cleavage sites.

3.4.1.3 Chromatography

Typically, for a protein of about 50 kDa, more than one hundred peptides are generated after pepsin hydrolysis. [57] In addition, the peptide mixture is usually in a salt solution which cannot be directly analyzed by MS. Thus prior to mass analysis, the peptides are desalted and separated using HPLC to minimize mass overlap and ion suppression in the mass spectrometer. The length in time of the HPLC run and the detailed gradient needs to be appropriately designed, so that the loss of deuterium through back exchange with solvent is minimized, while the resolution can be achieved as high as possible. As stated above, 85% of the total exchange can be retained if the proteolysis and separating steps are finished within 15 min. IN order to compensate the loss of labeling and the gain in final sequence coverage, a 30-40-min time scale for proteolysis and separation is typically applied nowadays.

The complexity of the samples that can be analyzed in such a fast gradient is generally limited comparing to standard LC separation gradient that sometimes last up to several hours, especially when using slower scanning or lower-resolution mass spectrometers. In addition, the short time allocated to the separation makes it difficult to implement capillary chromatographic systems that have typically higher elution delays due to larger relative in system dead volumes. [73] Therefore different approaches have been reported to improve the LC efficiency and reduce the back-exchange during the chromatography.

Besides the above-mentioned use of SFC, the replacement of H_2O in the aqueous mobile phase with mixed solvent of H_2O and polar aprotic modifiers [dimethylformamide (DMF) or N-methylpyrrolidone (NMP)] during the HPLC step [74] was another option to decrease the back-exchange derived from the idea of avoid the use of protiated solvents.

Other strategies involve the decrease of the temperature below zero point. The half time of back-exchange extends to 18 hours at -30 °C. [73] To achieve successful LC at subzero conditions, buffer modifiers need to be carefully selected that can reduce the freezing point of commonly used aqueous solvent systems, are mass spectrometry compatible, and do not preclude the use of liquid chromatography. Once such a system is well designed, it allows

much longer time to perform extensive chromatographic separation or subsequent sample processing.

As discussed in 2.4, ion mobility separation (IMS) has been widely applied to the analysis of proteins after ESI and MALDI. Based on the drift behaviors through an inert gas under the influence of an electric field, ions are separated due to the differences in the sizes, shapes and charges. These features are different from that of reversed-phase (RP)-HPLC in bottom-up HDX-MS workflow, thereby making IMS independent to HPLC and adding an additional dimension of separation. [75-76] Commercial IMMS instrument uses IM to separate ions between the ion source and mass analyzer, which occurs on a millisecond timescale; many mobility separations can occur during each chromatographic peak (which are typically on more of a seconds timescale). It is also proven from a demonstrative study of a small protein [75] that no additional losses of deuterium were created during the IMS process, and the drift time does not change upon solution-based deuterium incorporation, thus the addition of IMS increases the LC peak capacity of the experiment and should aid in HDX studies of very large proteins and protein complexes.

3.4.2 Top-down approach

Unlike the proteolysis/LC-MS approach that cleaves the deuterated proteins before introducing them into the mass spectrometer, in the "top-down" approach, fragmentation of the proteins takes place in the gas phase. Every individual chain typically experiences only a single cleavage event, resulting in a complementary N- and C-terminal ion pair. Spatial resolution can be obtained in top-down experiments because not every protein is cleaved at the same backbone site, such that a ladder of ion pairs is produced (Figure 8, lower part).

Early attempts to apply top-down strategies in HDX/MS were based on the use of collision-induced dissociation (CID). Yet the collisional heating of CID leads to a "scrambling" process, the intramolecular migration of hydrogen upon activation and can completely randomize the H/D exchange pattern that was imprinted onto the protein backbone during solution-phase HDX. The uses of ECD [39] and ETD [40] have basically solved the scrambling problem. Both techniques involve the transfer of a low energy electron to a gas-phase protein cation, generating the c/z- ion pairs. And the low energy induced in such fragmentations is critical as scrambling is negligible during the process, provided that collisional activation in the ion sampling interface is kept at a minimum. Potential advantages of these electron-based fragmentation techniques include an improved spatial resolution and

the opportunity to minimize back exchange. [62] Both ECD and ETD have claimed high sequence coverage, allowing enhanced spatial resolution to be achieved, in some cases even down to the single amide level. [67] As chromatography is no longer a limit for the solvent selection and the analysis can be much faster because the proteolysis and separation steps are replaced by the gas-phase fragmentation, more flexible strategies can be applied to reduce the back-exchange. The temperature in the quenching condition can be subzero, e.g. Amen et al. [77] have reported a chip-based nanoelectrospray approach based on cooling to subzero temperature (-15 °C), offering additional advantages over nanoelectrospray needles such as high spray reproducibility and spray stability and the possibility for automation [78], and the deuteration level remains ~95% after 10 min of back-exchange. Yet when the protein size increases, i.e. when the protein is larger than 30 kDa, the top-down approach has not conquered the limitation of incomplete sequence coverage, which is a hindrance from the gas phase fragmentation efficiency. [79] Combining the bottom-up and top-down approach is therefore another option to take advantages of both methods, which is going to be discussed later.

3.4.3 Mass measurement and data processing

Relative HDX levels of individual protein fragments are determined as a function of deuteration time *t* according to

$$deuteration\ level\ (t) = \frac{m(t) - m_0}{m_{max} - m_0} \qquad (9)$$

where $m(t)$ is the centroid mass of the peptide of interest, whereas m_{max} and m_0 are the corresponding reference masses for completely labeled and unlabeled peptide, respectively. Each amino acid residues in the protein corresponds to a unique kinetic, yet this single site resolution are difficult to achieve with common experimental condition. Therefore the exchange rate constants on the peptide level are always determined rather than the single amide group, as:

$$deuteration\ level\ (t) = A_0 + A_1(1 - exp[-k_1 t]) + A_2(1 - exp[-k_2 t]) \qquad (10)$$

where A_0 is the fraction of amide sites that undergo burst phase exchange, and A_1, A_2 are the fractions that are labeled with average rate constants k_1 and k_2, respectively. [54] In many cases, a single exponential term is sufficient for the fitting of the observed kinetics, and it is not very common to observe kinetics that requires more than two exponentials.

3.5 Application of HDX-MS in protein structural studies

3.5.1 Protein folding

HDX-MS is powerful in the study of protein folding mechanisms. This arises from the nature of the exchange reaction: transient intermediates in a folding pathway will exhibit different exchange rate constants according to their conformation, especially in the EX1 regime. [9] Thus different protein populations become labeled to different degrees and are separated in the masses, and the respective exchange rates can be measured.

Such topics are usually studied using the pulse-labeling strategy, as described in xx. The mass of the labeled proteins can be measured directly with MS, which provides global information on folding process. Recently researches have been carried out on the more detailed level of the different conformers, by analyzing the different deuteration level on protein fragments either by bottom-up or top-down approaches. For example, Pan et al. reported a study of structural characterizing of an apo-myoglobin (aMb) kinetic folding intermediate with pulsed top-down HDX. [67]

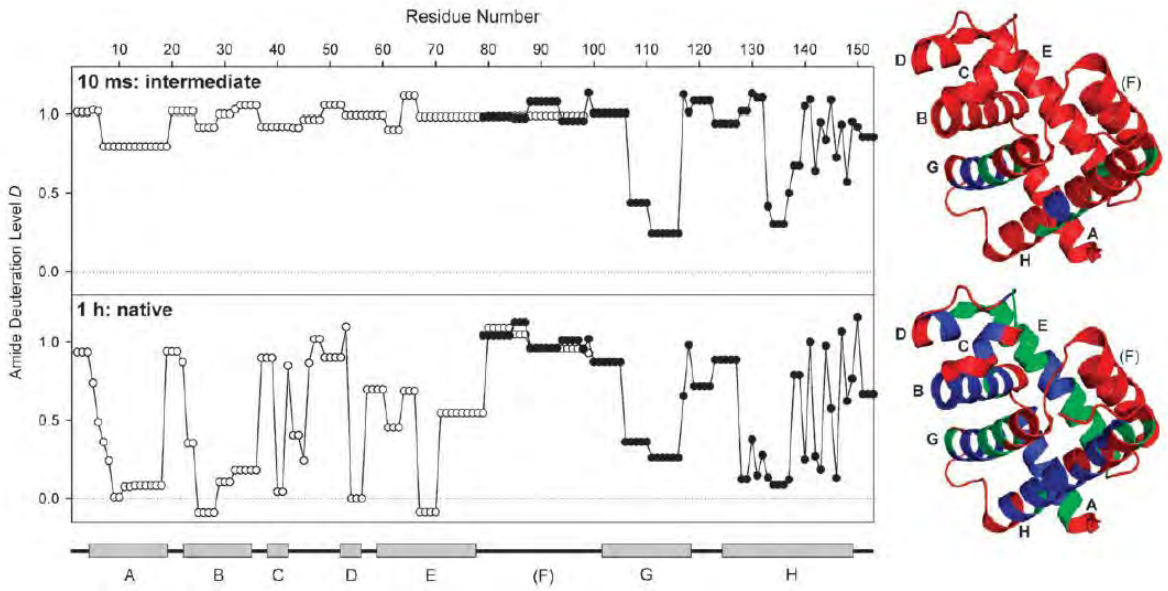


Figure 9 Deuteration level D of backbone amides for the apo-myoglobin (aMb) folding intermediate (top, refolding time t_{fold} = 10 ms), and for the native control (bottom, t_{fold} = 1 h) after pulsed HDX. The data were obtained by top-down ECD. Open and filled symbols represent deuteration levels deducted from c and z- ion data, respectively. Also shown is the α-helical secondary structure of native aMb. Helices are denoted as A–E, G, and H. On the right hand side of this figure the D values are mapped to the crystal structure of holo-myoglobin, using the following color scheme: blue, D < 0.33; green, 0.33 < D > 0.66; red D > 0.66. [67]

The work flow is demonstrated in Figure 9. Top-down ECD fragmentation of electrosprayed aMb allowed mass shifts to be determined for 36 c- and 34 z-ions that covered the entire protein sequence (Figure 9). Since deuteration level $D = 1$ corresponds to amides that are unprotected whereas fully protected sites are characterized by $D = 0$, it can be revealed that the aMb folding intermediate (Figure 9, top) remains mostly unprotected throughout the first 105 residues, marginal protection is seen only for helix A and only helices G and H are well developed at this time point. The deuteration behavior in folded conformer (Figure 9, bottom), on the other hand, shows that completely deuterated regions and virtually unlabeled regions are well-correlated with the the secondary structure of native aMb, i.e. the loop regions in the native protein are heavily deuterated, whereas helices are mostly protected. The partial protection observed in the C–D loop is consistent with hydrogen bonding data obtained by crystallography. These results reveal that the initial step of aMb folding involves formation of helices G and H. An additional marginal stabilization of helix A seems likely. Formation of the remaining helices B, C, D, and E occurs at a later stage.

3.5.2 Protein interactions

Equally important, HDX-MS has been applied in the studies of protein interactions. Upon interaction with metal ions, ligands, other proteins or large biomolecules, the amide hydrogens at the binding interface may be protected from exchange. Therefore a decrease in the deuteration level should be observed. Such a behavior is used to identify the binding sites of the above situations by comparing the deuteration level and the exchange kinetics of the non-complexed proteins with those of the complexed ones.

The change in the deuteration level upon complex formation, however, does not only involve the direct binding site, which is a precaution that should be carefully taken in the studies of protein interactions, especially in the case of small molecules. The alterations of the deuteration behavior can be caused by the protection of the binding agents, or in many cases, by the change in the unfolding/refolding motion, which can happen in regions far from the binding sites. In addition, not all binding regions may be probed under the standard HDX-MS condition, e.g. information of interactions involving the side chains can be lost.

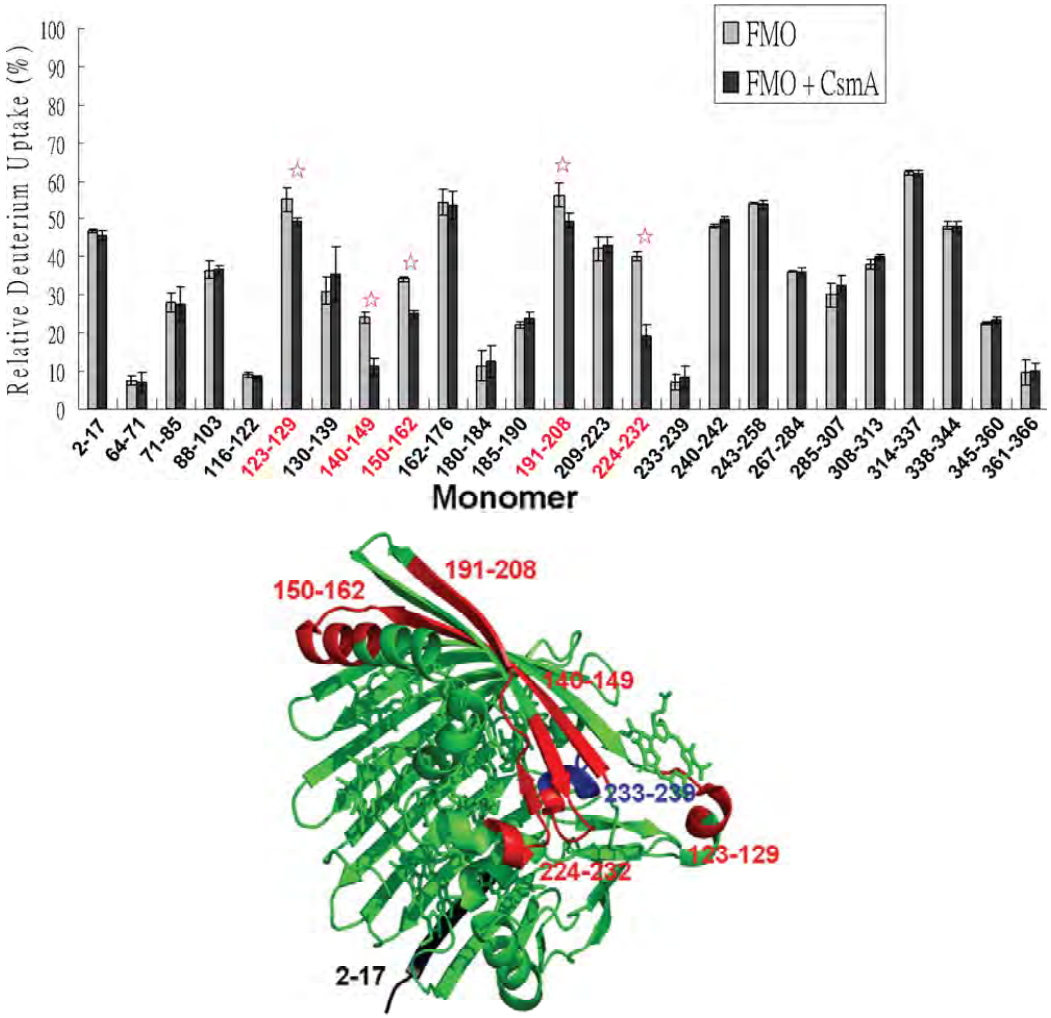


Figure 10 Top: Relative deuterium uptakes for peptic peptides from N- to C-terminus of FMO after 10 min of exchange. Peptide regions showing significant differences in deuterium uptake (>5% difference) in the presence of CsmA are labeled in red font and marked with red stars. Bottom: FMO monomer structure mapping with the deuterium uptake patterns. Peptide regions of FMO undergoing significant deuterium uptake differences between apo and holo states are in red. Peptide 233−239 located in the center of the trimer underwent low deuterium uptake (blue), whereas peptide 2−17 located on the outer shell of the protein underwent relatively extensive deuterium uptake (black). [80]

Nonetheless, when the deuteration behavior is carefully interpreted, it can be powerful to resolve the binding site and conformation changes, especially when coupling with computational modeling and docking studies. A recent work by the Gross group represents a delicate illustration of typical bottom-up HDX-MS strategy to resolve the interaction question. [80] In this work, the binding interface of the FMO antenna protein and the CsmA baseplate protein was revealed: global exchange kinetics were first compared for FMO and FMO+CsmA without proteolysis, showing clear decrease in the deuteration level upon CsmA binding. Then the proteins were proteolyzed by pepsin after a 10-min deuteration, and deuterated peptides were analyzed that generate information on the binding sites on FMO.

Sequence coverage of ~80% was achieved, and peptides consisting of 123−129, 140−149, 150−162, 191−208, and 224−232 showed significant decreases of deuteration levels after CsmA binding (Figure 10, top). All of these sites were located on one side of the FMO protein when they were mapped on the previously generated X-ray crystal structure, which is referred to as the Bchl *a* #1 side of the protein (Figure 10, bottom). The results indicated that the CsmA protein interacts with the Bchl *a* #1 side of the FMO protein, and were coincident with results from a previous study [81] on the membrane orientation of FMO in green-sulfur bacteria showing the Bchl *a* #1 side of the FMO interacts with the chlorosome.

3.5.3 Aid in structure elucidation and crystal production

As a major tool to determine protein structure, crystallography always meets the obstacles of successfully obtaining the protein crystals due to unstructured regions in some proteins. HDX MS can provide valuable information to precisely locate the disordered regions. After a brief D_2O pulse, unstructured regions are intensively deuterated and can be probed by MS, whereas folded regions stay with low deuteration level.

3.5.4 Quality control

A less commonly mentioned domain where HDX MS may be used is the quality control of protein folding. As the different folding behavior controls the deuteration behavior that can be monitored either by continuous-labeling or pulse-labeling HDX, by comparing the deuteration level of certain regions in the native conformer and refolded conformer, one can investigate various factors regarding the protein folding, e.g. the effects of mutations on protein structure, the correctness of the folding of a recombinant protein, the protein preparation protocols to produce correctly folded proteins, etc.

3.6 The on-going improvements and perspectives for HDX-MS

As the method of HDX-MS continues to develop as a mature tool for structural proteomics, the use of it has extend from small proteins (~50 kDa) to larger proteins and protein complexes (up to 300 kDa), and more complicated problems are subjected to be answered, often in combination with other structural analysis tools. Meanwhile, HDX MS has been questioned for the following limitations, namely the control of back-exchange, the reproducibility, the spatial resolution, the protein size and the consumption of time. Based on these issues, various approaches have been reported.

3.6.1 Automated experimental work-flow and data processing

Since the first prototype of fully automated H/D exchange system was reported in 2006 [82], different groups have designed the automated work-flow of bottom-up HDX-MS [83-84], and commercial products have now become available. Such robotic systems can conduct replicate runs in temperature-controlled autosamplers for multiple HDX times, followed by on-line proteolysis using immobilized enzyme columns, LC separation, and MS analysis. [54]

Considering the varieties that have large effects on the isotope pattern measured by the mass spectrometer as a result of the isotope incorporation, one will realize that the experimental handling during the labeling, quenching and proteolysis steps, the pH and the temperature are the most crucial factors, and a slight difference of these parameters leads to large error bar in the final measured mass. [85] This is why the automated systems are highly demanding: although such systems takes much longer to install, once well operated, they can help control the many variables and make HX MS experiments significantly easier to perform reproducibly. Yet the development of the robotics for HDX-MS still needs to overcome certain problems. For instance, as immobilized pepsin columns are commonly used in on-line systems, they increase digestion efficiency and limit enzyme autolysis. However, they add an extra element to a peptide carryover problem in the flow path. Carryover depletes deuteration and generates an isotopic distribution that mimics an EX1 signature in the next run. Therefore careful attention to system regeneration between runs is therefore required. [79]

Automated data processing software is another topic that is intensively discussed. With only manually data processing, routine HX MS is generally out of reach for many, especially in an industrial setting. Therefore, even at present it is the processing capability rather than the instrumental capability that limits the protein size that HDX-MS can work with. In the past decade, different software platforms have been introduced for the HDX-MS data analysis. For example, Kan et al. reported the ExMS [85-86] platform that processes high-resolution MS data, verifies peptide assignment, extracts deuterium and incorporation information, and the checkpoints during processing allow for manual intervention in processing routines. The HeXicon workflow was introduced in 2010 [87] for automatic deuteration distribution estimation with increased sequence coverage, and it allows insight into possible bimodal exchange rather than just average deuteration for each time point. Waters developed their commercial software, DynamX, with simultaneous visualization of raw MS data, identified

peptides, deuterium uptake graphs and comparability plots, and is also integrated with ion mobility separation. [88]

3.6.2 Combining the bottom-up and top-down approach

The pursuing of higher spatial resolution is critical for the competence of the HDX-MS, and although single amide-bond resolution has been claimed in several works to be achievable, such instances remain very rare. In the top-down approach, the limiting factor is usually the ability to transfer protein ions into the gas phase and fragment the parent ions into a sufficient number of product ions, especially when it comes to larger proteins above 40 kDa. [76] In the bottom-up approach, the spatial resolution mainly limited by the extent of peptide overlaps by proteolysis, which has to be performed under rather extreme conditions. Approaches aimed on the goal of site-specific resolution are therefore typically derived from two strategies: the combination of different proteases to create larger number and more overlaps of peptides as demonstrated in 3.4.1.2, or the incorporation of ETD or ECD into the classical bottom-up approach to further fragment the peptides into smaller pieces. The use of proteolysis in the typical bottom-up research allows large proteins in excess of 100 kDa to be analyzed routinely [76], whereas the use of ECD/ETD within peptide fragments generated via proteolysis allows sublocalizations of single amide exchange kinetics, providing that the instrumental conditions are carefully tuned to avoid the hydrogen scrambling. The work of Rand et al. [89] (Figure 11) illustrated the use of ETD fragmentation in bottom-up HDX to measure the exchange of individual amide linkages in the amyloidogenic protein β_2-microglobulin. A comparison of the deuterium levels of 60 individual backbone amides of β_2-microglobulin measured by HX-ETD-MS analysis to the corresponding values measured by NMR spectroscopy showed an excellent correlation, given that hydrogen scrambling of the uptaken deuterons in the gas phase was minimized as the instrumental conditions were tuned to ensure minimal vibrational excitation of deuterium labeled peptide ions.

This combined approach is superior to top-down as it does not have similar limitations regarding protein size, and to bottom-up approach as it provides further information on individual sites. However, the ability to record MS/MS spectra of individual peptides that coelute into the mass spectrometer depends on the times required during each acquisition to accumulate sufficient ion statistics to accurately determine the average masses of MS/MS fragment ions, and this is the main hindrance for the combined approach to reach high spatial resolution for extremely large proteins.

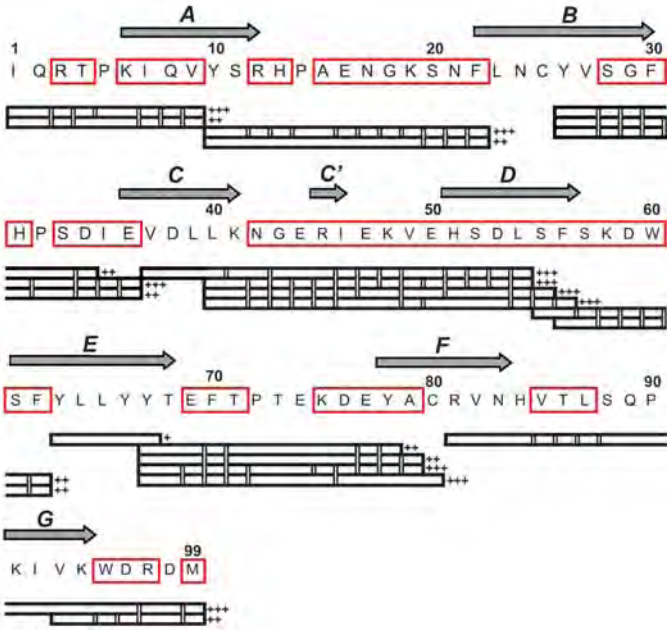


Figure 11 Sequence coverage of β_2-microglobulin obtained from peptic peptides and their ETD fragment ions. Peptides are shown as horizontal bars where double lines denote ETD cleavage sites producing abundant fragment ions suitable for measuring the deuterium levels of individual residues (bounded by red squares). The charge states of peptide precursor ions are indicated by plus signs, and secondary structure elements of β_2-microglobulin are shown above the sequence. [89]

3.6.3 High performance mass spectrometer

As a mass spectrometry-based method, HDX-MS strongly depends on the development of the mass spectrometers. High performance MS instruments such as Orbitrap and FT-ICR mass spectrometers are equipped nowadays in many laboratories, and they provide significantly high resolution and high sensitivity. Higher sequence overlap is therefore achievable due to the increasing number of peptides that can be identified, and this is crucial to the increase of the structural resolution of the HDX-MS method.

In a recent work [90], data acquired by an Orbitrap instrument and a QTOF instrument were compared. Three proteins with different sizes were proteolyzed by pepsin and separated by LC following a typical bottom-up workflow, and analyzed by the different mass spectrometers. The resulting MS/MS data were identified with Mascot search, giving a comparison of identified peptides in Table 1. Clearly both instruments result in rich data sets with high sequence coverage, indicating that the current state of instrumentation and data processing methods have generally overcome the problem of incomplete sequence coverage. However, on average the Orbitrap identified 64% more peptides than the QTOF, with a 42% overlap between the two systems, independent of protein size.

Table 1 Comparison of peptide identification statistics for the Orbitrap and QTOF for three different protein substrates. [90]

	XRCC4 (20 kDa)			PNK (55 kDa)			*α/β*-TUB (110 kDa)		
	QTOF	Orbitrap	Overlap	QTOF	Orbitrap	Overlap	QTOF	Orbitrap	Overlap
Count[a]	241	409	200	294	450	223	466	750	341
Seq. Cov. (%)	100	100	100	97	98.1	94.5	93.3	95.8	92.1
Redundancy[b]	13	19	10	7	12	5	6	9	4

[a] Unique pairs of retention time and m/z.

[b] Average number of times a given location is represented by a peptide.

The structure elucidation of nucleotide-binding GroEL protein [91] represents an example of the state-of-the-art bottom-up HDX-MS with advanced instrumentation and data processing. GroEL is composed of 14 identical subunits of 57 kDa each (~800 kDa total), arranged in two homoheptameric rings stacked "back to back" to form a cylinder with a central folding channel for substrate proteins. The HDX experiments for apo GroEL and GroEL-ATPγS were optimized and automated with an autosampler interfaced with an HPLC instrument, and an LTQ 14.5 T FT-ICR MS was used for data acquisition with mass resolving power of $m/\Delta m_{50\%}$ = 100,000 at m/z 400, sufficient to resolve virtually all of the overlapped peptide isotopic distributions in such experiments. After data processing, more than 500 peptides were identified for both apo GroEL and GroEL-ATPγS. 134 were selected for data analysis which provided ~99% sequence coverage, including the C-terminus unstructured region and ensured access to structural changes throughout the protein assembly.

The overall enhanced solvent accessibility and/or structure flexibility were also observed upon ATPγS binding, as indicated by higher deuteration levels throughout the sequence. For some of the peptides, the bimodal isotopic distributions were observed, and this could be ascribed to structural asymmetry between the two heptameric rings caused by ATP binding to one ring, which was coincident with previous cryo-EM studies. [92] Down to more details, upon ATP binding, significant weakening of the interactions between the GroEL subunits within one ring as well as between the two GroEL rings was simultaneously observed: in Figure 12, several interaction regions showed higher deuteration level in GroEL-ATPγS. Notably, such interpretations were also revealed by cryo-EM study [92] but within domain or sub-domain level resolution, while the HDX-MS results pinpoint the detailed sequence composition for such conformational changes.

Regions	I	II	III	IV	V
Representative Peptides	6-13 16-23	459-466	46-52	422-441 425-440	515-530 522-528

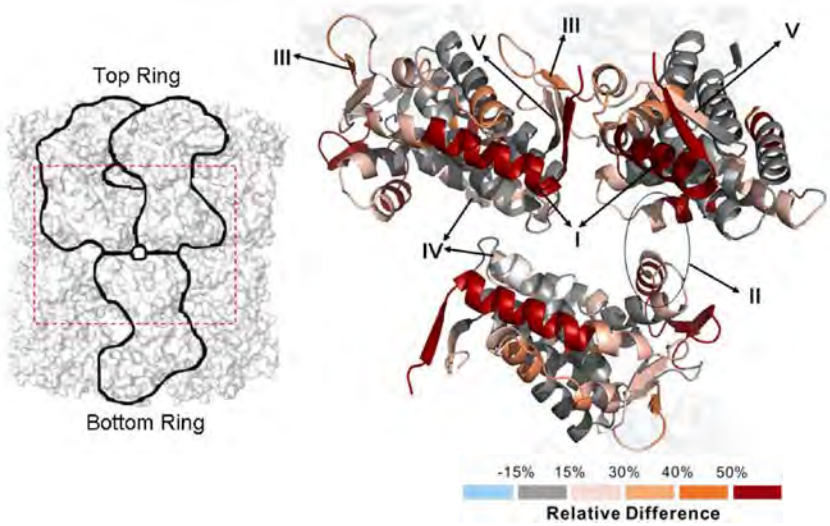


Figure 12 The equatorial domains in three interacting GroEL subunits. Five regions (I-V) from the HDX comparison of GroEL- ATPγS vs. apo GroEL are colored as the color coding at bottom right. Two subunits in the top GroEL ring and one in the bottom yring are chosen to represent the inter-subunit and inter-ring interactions (viewed from the outside of the cylinder). Representative peptides for the regions are shown in the inset table. [91]

4 HDX *vs.* other structural tools

Conventional biophysical tools used for protein structural studies include spectroscopic and chromatographic methods, X-ray crystallography, NMR, small angle X-ray scattering (SAXS), cryo-EM, etc. [54, 93] Optical spectroscopy and chromatographic methods are often employed for interrogating global structural aspects. X-Ray crystallography remains the gold standard for obtaining atomically-resolved conformational information on na tive proteins. NMR techniques provide single-proton level resolution and can be applied to partially unfolded proteins.

The HDX-MS method has proven itself a powerful complement or even alternative to these traditional tools. Unlike X-ray crystallography, HDX-MS is carried out in solution phase and gas phase, and therefore crystallization is not required. It can reveal changes to conformation and dynamics on a wide timescale, whereas X-ray crystallography only provides static crystal structures. Comparing to NMR, HDX-MS is not restricted by the protein size and can routinely analyze large proteins of more than 100 kDa, while NMR experiments on proteins larger than 40 kD a remain challenging. Latest HDX-MS experiments have reached the single-amide resolution. In addition, although considering the total amount of sample

required for a complete kinetics experiments (500 - 1000 pmol) [93] it is not totally convincing that HDX-MS consumes significantly less samples, this method is amenable to studying proteins that are hard to purify or that can only be handled at low concentrations (as low as 0.1 μM).

However, due to the nature of the HDX method, it can provide information only related to the amide backbone, and cannot reveal the structural information in three-dimension. One cannot therefore rely on HDX-MS alone to resolve a protein structure issue; rather, it should be used in combination with complementary structural information from other technologies (both experimental and computational) to generate models of large protein complexes.

Reference

1. Griffiths, J., A brief history of mass spectrometry. *Anal. Chem.* **2008,** *80*, 5678-5683.
2. Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T.; Matsuo, T., Protein and polymer analyses up to m/z 100 000 by laser ionization time? of? flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **1988,** *2*, 151-153.
3. Karas, M.; Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* **1988,** *60*, 2299-2301.
4. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989,** *246*, 64-71.
5. John T. Stults, D. A., Proteomics. In *Biological Mass Spectrometry*, Burlingame, A. L., Ed. Elsevier Academic Press: USA, 2005.
6. Nilsson, T.; Mann, M.; Aebersold, R.; Yates, J. R.; Bairoch, A.; Bergeron, J. J., Mass spectrometry in high-throughput proteomics: ready for the big time. *Nat. Methods* **2010,** *7*, 681.
7. Ahrens, C. H.; Brunner, E.; Qeli, E.; Basler, K.; Aebersold, R., Generating and navigating proteome maps using mass spectrometry. *Nat. Rev. Mol. Cell. Biol.* **2010,** *11*, 789-801.
8. Sidoli, S.; Cheng, L.; Jensen, O. N., Proteomics in chromatin biology and epigenetics: Elucidation of post-translational modifications of histone proteins by mass spectrometry. *J Proteomics* **2012,** *75*, 3419-3433.
9. Banks, G. C.; Deterding, L. J.; Tomer, K. B.; Archer, T. K., Hormone-mediated dephosphorylation of specific histone H1 isoforms. *J. Biol. Chem.* **2001,** *276*, 36467-36473.
10. Trudy McKee, J. R. M., *Biochemistry: the molecular basis of life*. 4 ed.; Oxford University Press: USA, 2008; p 880.
11. http://en.wikipedia.org/wiki/Protein_structure.
12. Walzthoeni, T.; Leitner, A.; Stengel, F.; Aebersold, R., Mass spectrometry supported determination of protein complex structure. *Curr. Opin. Struct. Biol.* **2013**.
13. Pardo, M.; Choudhary, J. S., Assignment of Protein Interactions from Affinity Purification/Mass Spectrometry Data. *J Proteome Res* **2012,** *11*, 1462-1474.
14. Kaake, R. M.; Wang, X.; Huang, L., Profiling of protein interaction networks of protein complexes using affinity purification and quantitative mass spectrometry. *Mol. Cell Proteomics* **2010,** *9*, 1650-1665.
15. Gavin, A.-C.; Maeda, K.; Kühner, S., Recent advances in charting protein–protein interaction: mass spectrometry-based approaches. *Curr. Opin. Biotechnol.* **2011,** *22*, 42-49.
16. Low, T. Y.; Magliozzi, R.; Guardavaccaro, D.; Heck, A. J., Unraveling the ubiquitin-regulated signaling networks by mass spectrometry-based proteomics. *Proteomics* **2012**.
17. Stengel, F.; Aebersold, R.; Robinson, C. V., Joining forces: integrating proteomics and cross-linking with the mass spectrometry of intact complexes. *Mol. Cell Proteomics* **2012,** *11*.
18. Van Duijn, E., Current limitations in native mass spectrometry based structural biology. *J. Am. Soc. Mass. Spectrom.* **2010,** *21*, 971-978.
19. Barrera, N. P.; Isaacson, S. C.; Zhou, M.; Bavro, V. N.; Welch, A.; Schaedler, T. A.; Seeger, M. A.; Miguel, R. N.; Korkhov, V. M.; van Veen, H. W., Mass spectrometry of membrane transporters reveals subunit stoichiometry and interactions. *Nat. Methods* **2009,** *6*, 585-587.
20. Barrera, N. P.; Di Bartolo, N.; Booth, P. J.; Robinson, C. V., Micelles protect membrane complexes from solution to vacuum. *Science* **2008,** *321*, 243-246.
21. Zhou, M.; Morgner, N.; Barrera, N. P.; Politis, A.; Isaacson, S. C.; Matak-Vinković, D.; Murata, T.; Bernal, R. A.; Stock, D.; Robinson, C. V., Mass spectrometry of intact V-type ATPases reveals bound lipids and the effects of nucleotide binding. *Science* **2011,** *334*, 380-385.

22. Paramelle, D.; Miralles, G.; Subra, G.; Martinez, J., Chemical cross-linkers for protein structure studies by mass spectrometry. *Proteomics* **2013**.
23. Leitner, A.; Reischl, R.; Walzthoeni, T.; Herzog, F.; Bohn, S.; Förster, F.; Aebersold, R., Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Mol. Cell Proteomics* **2012,** *11*.
24. Petrotchenko, E. V.; Borchers, C. H., Crosslinking combined with mass spectrometry for structural proteomics. *Mass Spectrom. Rev.* **2010,** *29*, 862-876.
25. Walzthoeni, T.; Claassen, M.; Leitner, A.; Herzog, F.; Bohn, S.; Förster, F.; Beck, M.; Aebersold, R., False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nat. Methods* **2012,** *9*, 901-903.
26. Lasker, K.; Förster, F.; Bohn, S.; Walzthoeni, T.; Villa, E.; Unverdorben, P.; Beck, F.; Aebersold, R.; Sali, A.; Baumeister, W., Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *PNAS* **2012,** *109*, 1380-1387.
27. Leitner, A.; Joachimiak, L. A.; Bracher, A.; Mönkemeyer, L.; Walzthoeni, T.; Chen, B.; Pechmann, S.; Holmes, S.; Cong, Y.; Ma, B., The molecular architecture of the eukaryotic chaperonin TRiC/CCT. *Structure* **2012**.
28. Kalisman, N.; Adams, C. M.; Levitt, M., Subunit order of eukaryotic TRiC/CCT chaperonin by cross-linking, mass spectrometry, and combinatorial homology modeling. *PNAS* **2012,** *109*, 2884-2889.
29. Xu, G.; Chance, M. R., Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chem. Rev.* **2007,** *107*, 3514-3543.
30. Konermann, L.; Tong, X.; Pan, Y., Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches. *J. Mass Spectrom.* **2008,** *43*, 1021-1036.
31. Maleknia, S. D.; Downard, K. M., Radical approaches to probe protein structure, folding, and interactions by mass spectrometry. *Mass Spectrom. Rev.* **2001,** *20*, 388-401.
32. Sharp, J. S.; Becker, J. M.; Hettich, R. L., Protein surface mapping by chemical oxidation: structural analysis by mass spectrometry. *Anal. Biochem.* **2003,** *313*, 216-225.
33. Berger, A.; Linderstrøm-Lang, K., Deuterium exchange of poly-DL-alanine in aqueous solution. *Arch. Biochem. Biophys.* **1957,** *69*, 106-118.
34. Englander, S. W., Hydrogen exchange and mass spectrometry: A historical perspective. *J. Am. Soc. Mass. Spectrom.* **2006,** *17*, 1481-9.
35. Dyson, H. J.; Wright, P. E., Unfolded proteins and protein folding studied by NMR. *Chemical Reviews-Columbus* **2004,** *104*, 3607-3622.
36. Wales, T. E.; Engen, J. R., Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrom. Rev.* **2006,** *25*, 158-170.
37. Katta, V.; Chait, B. T.; Carr, S., Conformational changes in proteins probed by hydrogen-exchange electrospray-ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **1991,** *5*, 214-217.
38. Zhang, Z.; Smith, D. L., Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Sci.* **1993,** *2*, 522-531.
39. Pan, J.; Han, J.; Borchers, C. H.; Konermann, L., Electron capture dissociation of electrosprayed protein ions for spatially resolved hydrogen exchange measurements. *J. Am. Chem. Soc.* **2008,** *130*, 11574-11575.
40. Abzalimov, R. R.; Kaplan, D. A.; Easterling, M. L.; Kaltashov, I. A., Protein conformations can be probed in top-down HDX MS experiments utilizing electron transfer dissociation of protein ions without hydrogen scrambling. *J. Am. Soc. Mass. Spectrom.* **2009,** *20*, 1514-1517.

41. Hvidt, A.; Nielsen, S. O., Hydrogen exchange in proteins. *Adv. Protein Chem.* **1966,** *21*, 287-386.
42. Eigen, M., Proton Transfer, Acid-Base Catalysis, and Enzymatic Hydrolysis. Part I: ELEMENTARY PROCESSES. *Angew Chem Int Ed Engl* **1964,** *3*, 1-19.
43. Woodward, C.; Simon, I.; Tüchsen, E., Hydrogen exchange and the dynamic structure of proteins. *Mol. Cell. Biochem.* **1982,** *48*, 135-160.
44. Englander, S.; Englander, J., [19] Hydrogen-tritium exchange. *Methods Enzymol.* **1972,** *26*, 406-413.
45. Englander, S. W.; Kallenbach, N. R., Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q. Rev. Biophys.* **1983,** *16*, 521-655.
46. Perrin, C. L., Proton exchange in amides: surprises from simple systems. *Acc. Chem. Res.* **1989,** *22*, 268-275.
47. Berger, A.; Loewenstein, A.; Meiboom, S., Nuclear Magnetic Resonance Study of the Protolysis and Ionization of N-Methylacetamide1. *J. Am. Chem. Soc.* **1959,** *81*, 62-67.
48. Molday, R. S.; Kallen, R. G., Substituent effects on amide hydrogen exchange rates in aqueous solution. *J. Am. Chem. Soc.* **1972,** *94*, 6739-6745.
49. Eriksson, M.; Härd, T.; Nilsson, L., On the pH dependence of amide proton exchange rates in proteins. *Biophys. J.* **1995,** *69*, 329-339.
50. Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W., Primary structure effects on peptide group hydrogen exchange. *Proteins: Structure, Function, and Bioinformatics* **1993,** *17*, 75-86.
51. Smith, D. L.; Deng, Y.; Zhang, Z., Probing the non-covalent structure of proteins by amide hydrogen exchange and mass spectrometry. *J. Mass Spectrom.* **1997,** *32*, 135-146.
52. Engen, J. R.; Smith, D. L., Peer Reviewed: Investigating Protein Structure and Dynamics by Hydrogen Exchange MS. *Anal. Chem.* **2001,** *73*, 256 A-265 A.
53. Kaltashov, I. A.; Bobst, C. E.; Abzalimov, R. R., Mass spectrometry-based methods to study protein architecture and dynamics. *Protein Sci.* **2013**.
54. Konermann, L.; Pan, J.; Liu, Y. H., Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.* **2011,** *40*, 1224-34.
55. Kim, K. S.; Fuchs, J. A.; Woodward, C. K., Hydrogen exchange identifies native-state motional domains important in protein folding. *Biochemistry* **1993,** *32*, 9600-9608.
56. Woodward, C.; Hilton, B., Hydrogen isotope exchange kinetics of single protons in bovine pancreatic trypsin inhibitor. *Biophys. J.* **1980,** *32*, 561-575.
57. Miranker, A.; Robinson, C. V.; Radford, S. E.; Aplin, R. T.; Dobson, C. M., Detection of transient protein folding populations by mass spectrometry. *Science* **1993,** *262*, 896-900.
58. Morgan, C. R.; Hebling, C. M.; Rand, K. D.; Stafford, D. W.; Jorgenson, J. W.; Engen, J. R., Conformational transitions in the membrane scaffold protein of phospholipid bilayer nanodiscs. *Mol Cell Proteomics* **2011,** *10*.
59. Xiao, H.; Hoerner, J. K.; Eyles, S. J.; Dobo, A.; Voigtman, E.; Mel'čuk, A. I.; Kaltashov, I. A., Mapping protein energy landscapes with amide hydrogen exchange and mass spectrometry: I. A generalized model for a two-state protein and comparison with experiment. *Protein Sci.* **2005,** *14*, 543-557.
60. Bobst, C. E.; Kaltashov, I. A., Advanced mass spectrometry-based methods for the analysis of conformational integrity of biopharmaceutical products. *Curr. Pharm. Biotechnol.* **2011,** *12*, 1517.
61. Zhang, H.-M.; McLoughlin, S. M.; Frausto, S. D.; Tang, H.; Emmett, M. R.; Marshall, A. G., Simultaneous reduction and digestion of proteins with disulfide bonds for hydrogen/deuterium exchange monitored by mass spectrometry. *Anal. Chem.* **2010,** *82*, 1450-1454.

62. Pan, J.; Han, J.; Borchers, C. H.; Konermann, L., Structure and dynamics of small soluble Aβ (1–40) oligomers studied by top-down hydrogen exchange mass spectrometry. *Biochemistry* **2012,** *51*, 3694-3703.
63. Chetty, P. S.; Mayne, L.; Lund-Katz, S.; Stranz, D.; Englander, S. W.; Phillips, M. C., Helical structure and stability in human apolipoprotein AI by hydrogen exchange and mass spectrometry. *PNAS* **2009,** *106*, 19005-19010.
64. Pan, J.; Han, J.; Borchers, C. H.; Konermann, L., Hydrogen/deuterium exchange mass spectrometry with top-down electron capture dissociation for characterizing structural transitions of a 17 kDa protein. *J. Am. Chem. Soc.* **2009,** *131*, 12801-12808.
65. Zhang, H.-M.; Kazazic, S. a.; Schaub, T. M.; Tipton, J. D.; Emmett, M. R.; Marshall, A. G., Enhanced digestion efficiency, peptide ionization efficiency, and sequence resolution for protein hydrogen/deuterium exchange monitored by Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **2008,** *80*, 9034-9041.
66. Chance, M., *Mass Spectrometry Analysis for Protein-Protein Interactions and Dynamics*. John Wiley & Sons, Inc.: 2008.
67. Pan, J.; Han, J.; Borchers, C. H.; Konermann, L., Characterizing short-lived protein folding intermediates by top-down hydrogen exchange mass spectrometry. *Anal. Chem.* **2010,** *82*, 8591-8597.
68. Englander, S. W.; Mayne, L., Protein folding studied using hydrogen-exchange labeling and two-dimensional NMR. *Annu. Rev. Biophys. Biomol. Struct.* **1992,** *21*, 243-265.
69. Emmett, M. R.; Kazazic, S.; Marshall, A. G.; Chen, W.; Shi, S. D.-H.; Bolaños, B.; Greig, M. J., Supercritical fluid chromatography reduction of hydrogen/deuterium back exchange in solution-phase hydrogen/deuterium exchange with mass spectrometric analysis. *Anal. Chem.* **2006,** *78*, 7058-7060.
70. Corporation, W. B.; Diagnostics, W., *Manual of Clinical Enzyme Measurements*. Worthington Diagnostics: 1972.
71. Cravello, L.; Lascoux, D.; Forest, E., Use of different proteases working in acidic conditions to improve sequence coverage and resolution in hydrogen/deuterium exchange of large proteins. *Rapid Commun. Mass Spectrom.* **2003,** *17*, 2387-2393.
72. Man, P.; Montagner, C.; Vernier, G.; Dublet, B.; Chenal, A.; Forest, E.; Forge, V., Defining the interacting regions between apomyoglobin and lipid membrane by hydrogen/deuterium exchange coupled to mass spectrometry. *J. Mol. Biol.* **2007,** *368*, 464-472.
73. Venable, J. D.; Okach, L.; Agarwalla, S.; Brock, A., Subzero temperature chromatography for reduced back-exchange and improved dynamic range in amide hydrogen/deuterium exchange mass spectrometry. *Anal. Chem.* **2012,** *84*, 9601-8.
74. Valeja, S. G.; Emmett, M. R.; Marshall, A. G., Polar Aprotic Modifiers for Chromatographic Separation and Back-Exchange Reduction for Protein Hydrogen/Deuterium Exchange Monitored by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* **2012,** *23*, 699-707.
75. Iacob, R. E.; Murphy, J. P.; Engen, J. R., Ion mobility adds an additional dimension to mass spectrometric analysis of solution-phase hydrogen/deuterium exchange. *Rapid Commun. Mass Spectrom.* **2008,** *22*, 2898-2904.
76. Chalmers, M. J.; Busby, S. A.; Pascal, B. D.; West, G. M.; Griffin, P. R., Differential hydrogen/deuterium exchange mass spectrometry analysis of protein–ligand interactions. *Expert Rev Proteomics* **2011,** *8*, 43-59.
77. Amon, S.; Trelle, M. B.; Jensen, O. N.; Jørgensen, T. J., Spatially resolved protein hydrogen exchange measured by subzero-cooled chip-based nanoelectrospray ionization tandem mass spectrometry. *Anal. Chem.* **2012,** *84*, 4467-4473.

78. Zhang, S.; Van Pelt, C. K.; Henion, J. D., Automated chip-based nanoelectrospray-mass spectrometry for rapid identification of proteins separated by two-dimensional gel electrophoresis. *Electrophoresis* **2003,** *24*, 3620-3632.
79. Percy, A. J.; Rey, M.; Burns, K. M.; Schriemer, D. C., Probing protein interactions with hydrogen/deuterium exchange and mass spectrometry—a review. *Anal. Chim. Acta* **2012,** *721*, 7-21.
80. Huang, R. Y.; Wen, J.; Blankenship, R. E.; Gross, M. L., Hydrogen–deuterium exchange mass spectrometry reveals the interaction of Fenna–Matthews–Olson protein and chlorosome CsmA protein. *Biochemistry* **2011,** *51*, 187-193.
81. Wen, J.; Zhang, H.; Gross, M. L.; Blankenship, R. E., Membrane orientation of the FMO antenna protein from Chlorobaculum tepidum as determined by mass spectrometry-based footprinting. *PNAS* **2009,** *106*, 6134-6139.
82. Chalmers, M. J.; Busby, S. A.; Pascal, B. D.; He, Y.; Hendrickson, C. L.; Marshall, A. G.; Griffin, P. R., Probing protein ligand interactions by automated hydrogen/deuterium exchange mass spectrometry. *Anal. Chem.* **2006,** *78*, 1005-1014.
83. Burkitt, W.; O'Connor, G., Assessment of the repeatability and reproducibility of hydrogen/deuterium exchange mass spectrometry measurements. *Rapid Commun. Mass Spectrom.* **2008,** *22*, 3893-3901.
84. Houde, D.; Berkowitz, S. A.; Engen, J. R., The utility of hydrogen/deuterium exchange mass spectrometry in biopharmaceutical comparability studies. *J. Pharm. Sci.* **2011,** *100*, 2071-2086.
85. Iacob, R. E.; Engen, J. R., Hydrogen exchange mass spectrometry: are we out of the quicksand? *J. Am. Soc. Mass. Spectrom.* **2012,** *23*, 1003-1010.
86. Kan, Z.-Y.; Mayne, L.; Chetty, P. S.; Englander, S. W., ExMS: Data analysis for HX-MS experiments. *J. Am. Soc. Mass. Spectrom.* **2011,** *22*, 1906-1915.
87. Lou, X.; Kirchner, M.; Renard, B. Y.; Köthe, U.; Boppel, S.; Graf, C.; Lee, C.-T.; Steen, J. A.; Steen, H.; Mayer, M. P., Deuteration distribution estimation with improved sequence coverage for HX/MS experiments. *Bioinformatics* **2010,** *26*, 1535-1541.
88. Wei, H.; Ahn, J.; Yu, Y. Q.; Tymiak, A.; Engen, J. R.; Chen, G., Using hydrogen/deuterium exchange mass spectrometry to study conformational changes in granulocyte colony stimulating factor upon PEGylation. *J. Am. Soc. Mass. Spectrom.* **2012,** *23*, 498-504.
89. Rand, K. D.; Zehl, M.; Jensen, O. N.; Jørgensen, T. J., Protein hydrogen exchange measured at single-residue resolution by electron transfer dissociation mass spectrometry. *Anal. Chem.* **2009,** *81*, 5577-5584.
90. Burns, K. M.; Rey, M.; Baker, C. A.; Schriemer, D. C., Platform Dependencies in Bottom-up Hydrogen/Deuterium Exchange Mass Spectrometry. *Mol. Cell Proteomics* **2013,** *12*, 539-548.
91. Zhang, Q.; Chen, J.; Kuwajima, K.; Zhang, H.-M.; Xian, F.; Young, N. L.; Marshall, A. G., Nucleotide-induced conformational changes of tetradecameric GroEL mapped by H/D exchange monitored by FT-ICR mass spectrometry. *Sci. Rep.* **2013,** *3*.
92. Ranson, N. A.; Farr, G. W.; Roseman, A. M.; Gowen, B.; Fenton, W. A.; Horwich, A. L.; Saibil, H. R., ATP-bound states of GroEL captured by cryo-electron microscopy. *Cell* **2001,** *107*, 869-879.
93. Marcsisin, S. R.; Engen, J. R., Hydrogen exchange mass spectrometry: what is it and what can it tell us? *Anal. Bioanal. Chem.* **2010,** *397*, 967-972.

Chapter 2

The eIF3 complexes: functions, structure, production and purification

1 Pathway of eukaryotic translation initiation 51
2 The role of eIF3 in translation initiation 52
3 The structure of eIF3 53
3.1 MS revealed the overall interaction network of human eIF3 54
3.2 The atomic structure of yeast eIF3i:b(654-700) 56
4 Synthesis and purification of proteins of interest 59
References 62

1 Pathway of eukaryotic translation initiation

In all kingdom of life, the translation is the process by which mRNA is translated into proteins. Here a brief summary of this process is presented according to a review by Hinnebusch in 2006. [1] In eukaryotes, the initiation of this process is understood to undergo the following pathway: it begins with the production of free 40S ribosomal subunits (Figure 1). A ternary complex (TC) of eIF2, GTP and Met-tRNA$_i^{Met}$ binds to the 40S subunit, forming the 43S pre-initiation complex (PIC) through stimulation of eIF1, eIF1A, eIF3 and eIF5. The 43S PIC reacts with the m^{7}G-capped 5'-end of the mRNA, stimulated by the cap-binding protein eIF4E and its partners in the eIF4F complex, as well as the poly(A)-binding protein (PABP). eIF4G is the scaffold subunit of eIF4F and has additional binding sites for eIF4A (ATP-dependent RNA helicase) and (in mammals at least) for eIF3. Then comes the scanning of 43S PIC for an AUG start codon, a process stimulated by eIF1, eIF1a and eIF4G. When secondary structure is contained in the mRNA leader, this process also needs eIF4A, eIF4B and ATP hydrolysis. The GTP bound to eIF2 in the scanning complex is hydrolyzed to GDP in an eIF5-dependent manner; however, release of Pi from eIF2–GDP–Pi is prevented by eIF1 until an AUG start codon enters the ribosomal P-site and base-pairs with the anticodon of Met-tRNA$_i^{Met}$. eIF1 is released from the P-site upon r ecognition of the start codon, enabling the 48S PIC to assume a closed conformation and thereby it no longer scans the mRNA and it triggers the Pi release. The rest of the eIFs are released when the 60S subunit joins to form the 80S complex stimulated by eIF5B•GTP. Catalyzed by eIF2B, the eIF2•GDP turns back to eIF2•GTP for a new initiation cycle.

Recycling of eIF2•GDP by eIF2B is prevented under stress or starvation conditions by phosphorylation of a subdomain of eIF2, eIF2α. The reduction in TC assembly decreases the rate of general translation initiation, but accelerates the translation of specialized mRNAs, e.g. *GCN4* or *ATF4*, which contain upstream open reading frames (uORFs) in their leaders and encode stress-induced transcriptional activators. Assembly of eIF4F, and thus its binding to the mRNA cap, is also downregulated upon starvation or lack of insulin in mammalian cells by a mechanism involving the eIF4e binding protein (4E-BP), which competes with eIF4G for binding to eIF4E. The inhibitory function of 4E-BP is controlled by the mammalian kinase mTOR and is overcome in nutrient-repleted, insulin-treated cells as a mean of stimulating protein synthesis rates under these conditions.

Some mRNAs are translated by an alternative mechanism, i.e. internal initiation, in which the 40S ribosome binds directly to an internal ribosome entry site (IRES) located close to the start codon, circumventing the requirement for the m7G cap and eIF4E and, depending on the IRES, some or all of the other eIFs.

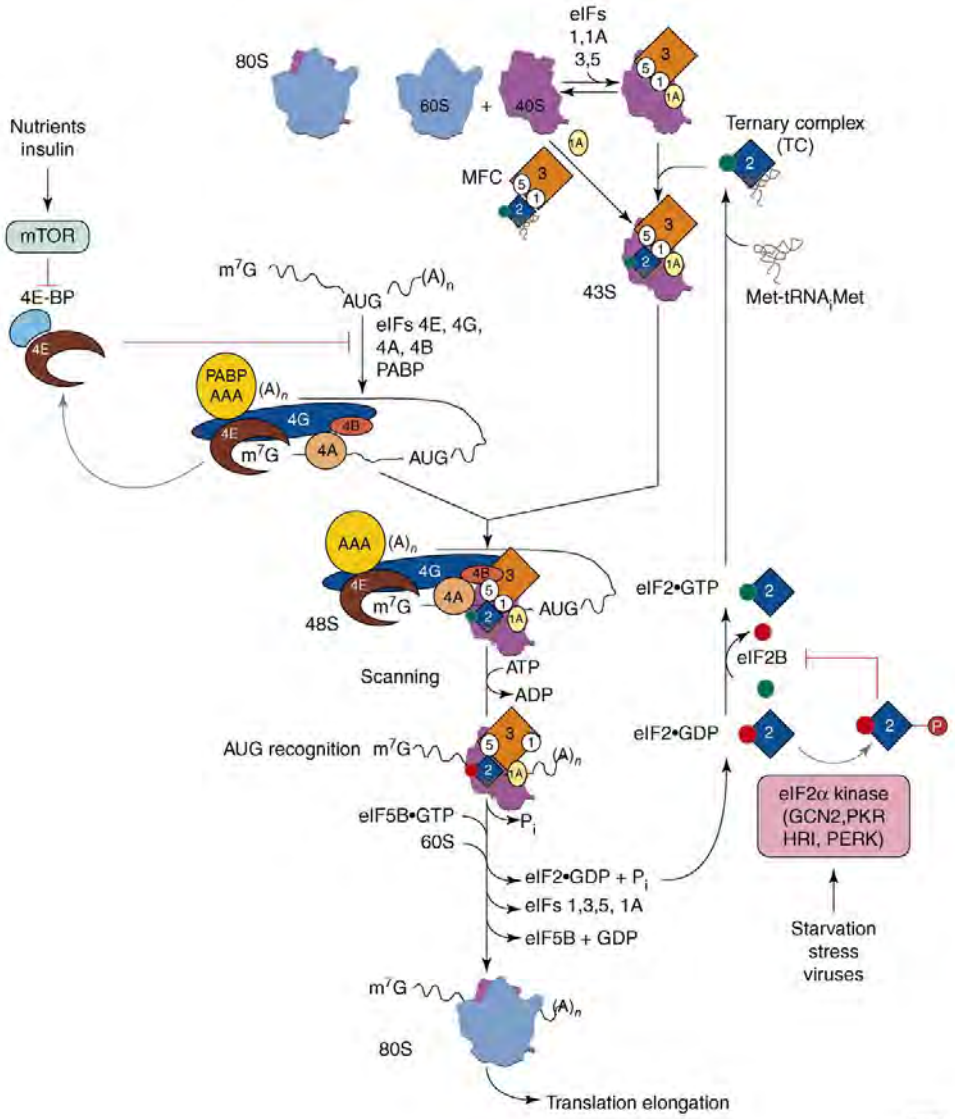


Figure 1 The eukaryotic translation initiation pathway. [1]

2 The role of eIF3 in translation initiation

In the above model, at least 12 different eukaryotic initiation factors (eIFs) are involved. eIF3 was one of the first identified initiation factors in the 1970s [2-3], and it is now proven that most of the reactions in the initiation pathway are stimulated by eIF3.

The best-studied function of mammalian eIF3 is that it c an bind to the 40S subunit and inhibits its interaction with the 60S subunit, which is crucial for the formation of the 43S PIC.

Recent studies showed that this function of eIF3 is enhanced by other factors including TC [4-5], a combination of eIF1 and eIF1A [6], and RNA oligonucleotides. Regarding the binding region, several cryo-EM studies revealed that in the interaction of mammalian eIF3 with the 40S subunit, one 'leg' of eIF3 extends close enough to the interface surface to disrupt the intersubunit contact activity with 60S. [7-8]

In the subsequent step, mammalian eIF3 is shown to promote TC assembly under conditions of limiting Met-$tRNA_i^{Met}$. In reconstituted mammalian systems, eIF3 stimulates binding of TC to the 40S subunit. [4-5] In a reconstituted yeast system, eIF3 and eIF5 both enhance TC binding to the 43S PIC, which has a lower affinity for TC than has the 48S PIC. [9] When it comes to the scanning of mRNA by 43S PIC, mammalian eIF3 stimulates mRNA binding to the 40S subunit in vitro [10], and this is implicated by yeast eIF3 studies showing that a mutation in eIF3b impairs mRNA binding by the 40S subunit in yeast extracts, which can be rescued by eIF3 or the eIF3a–eIF3b–eIF3c subcomplex. [11] Evidences also show that eIF3 functions as a specialized reinitiation factor in gene-specific translational control, and as a platform for signal transduction. [1]

Therefore, although eIF3 has no known enzymatic activities and might not directly interact with $tRNA_i^{Met}$ or the decoding sites on the ribosome, it is possibly acting as a large scaffold anchored to the solvent side of the 40S subunit with appendages reaching round to the interface surface. [1]

3 The structure of eIF3

Consistent with its diverse functions, eIF3 has a complex structure: the mammalian eIF3 is made up of 13 non-identical subunits as summarized in Table 1. [1] eIF3 from budding yeast contains orthologs of five mammalian eIF3 subunits: eIF3a, eIF3b, eIF3c, eIF3g and eIF3i, all of which are essential for translation in vivo. These five essential yeast subunits constitute a conserved 'core' complex that can execute the crucial functions of eIF3, stimulated by eIF3j, a sixth, nonessential, substoichiometric subunit of yeast eIF3.

Table 1 Summary of eIF3 subunits in selected eukaryotes. [1]

Unified nomenclature	S. cerevisiae	Functions
eIF3a[a]	TIF32	40S binding; eIF4B binding; MFC assembly; TC and mRNA recruitment
eIF3b[a]	PRT1	40S binding; MFC assembly; TC and mRNA recruitment; scanning
eIF3c[a]	NIP1	40S binding; MFC assembly; TC and mRNA recruitment; AUG recognition
eIF3d	–	
eIF3e	–	
eIF3f	–	Possible binding site for mTOR and S6K1
eIF3g[a]	TIF35	Binding eIF4B and CaMV TAV
eIF3h	–	
eIF3i[a]	TIF34	
eIF3j	HCR1	40S binding; MFC assembly
eIF3k	–	
eIF3l	–	
eIF3m	–	

[a]: Core subunits; S. cerevisiae: *Saccharomyces cerevisiae*

3.1 MS revealed the overall interaction network of human eIF3

Because of eIF3's overall structural complexity and the lack of facile genetic approaches, knowledge of many aspects of its structure and function remain elusive. Mass spectrometry on the intact eIF3 complex successfully generated a subunit interaction map. [12] Intact eIF3 was subjected directly to mass analysis in solutions with various ionic strengths. Under lower salt concentration (100 mM ammonium acetate), the principle series of charge state peaks illustrated that the complex exists as a homogeneous population, with all 13 subunits present at unit stoichiometry. The dissociation of eIF3i was already observed at this stage, indicating a relatively labile association within the intact complex. In addition, eIF3j, k and m dissociated from the complex when the ionic strength or the acceleration in the mass spectrometer was slightly increased, and these results suggested that these subunits locate on the periphery of the complex.

Polar and ionic interactions of subunit interfaces can be perturbed when the ionic strength is increased. Upon changing the ionic strength, a series of subcomplexes each comprising a small number of subunits ($\leq$5) were observed and assigned. Precise subunit –subunit contacts were derived by elucidating these results. For example, three subcomplexes: two dimers b:g and i:g together with the trimer (b:g:i) were detected in human eIF3, indicating the

arrangement of the trimer b:g:i with g in contact both with b and i. Upon using isolated eIF3 from yeast, pairwise dimers g:i and a:b and the trimer b:g:i were observed when the affinity tag is placed on subunit b (Figure 2). By tagging subunit c, additional subcomplexes were assigned as (c:b:i), b:i, b:g, and (c:b:g). These data from yeast can be derived into an interaction network for eIF3(a:b:c:i:g) with eIF3b serving as the scaffolding protein connecting the remaining subunits a,c,i and g (Figure 2).

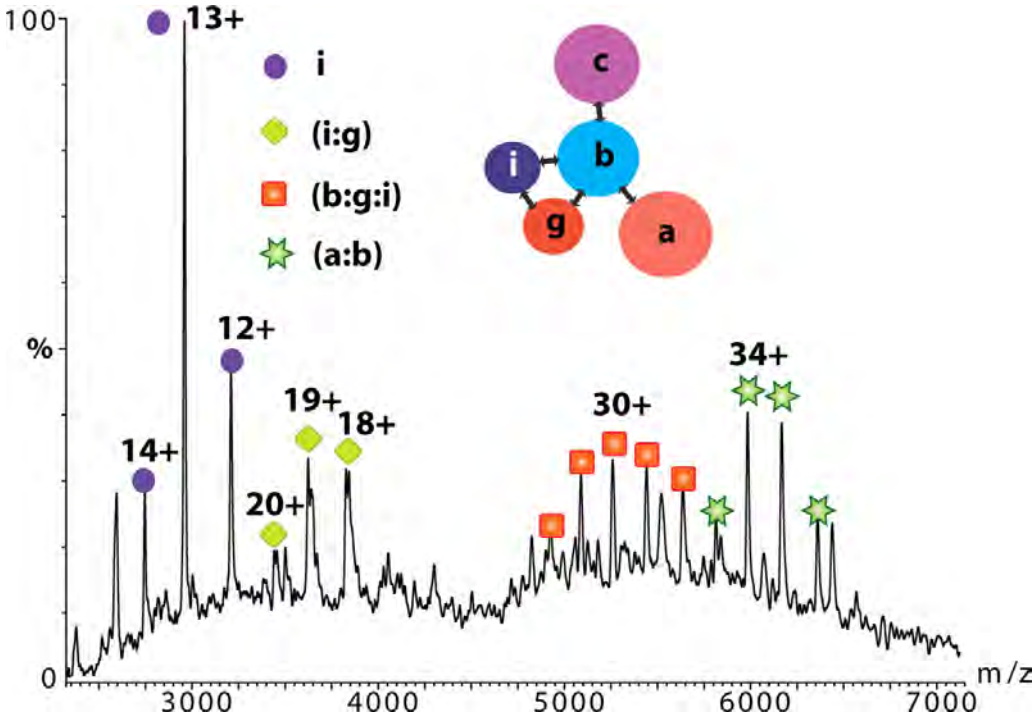


Figure 2 MS spectrum of the yeast eIF3 isolated by tagging subunit eIF3b. [12]

Upon similar derivation, three tightly associated modules were generated by manipulating the ionic strength, denoted as A(eIF3a:b:i:g), B(eIF3f:h:m), and C(eIF3k:l:e:d:c). eIF3c and h provide one of the electrostatic interactions between modules B and C, while the role of eIF3c is crucial in recruiting module C to A. Based on MS as well as immuno-precipitation (IP) result, interactions between eIF3b:f and b:h are important for the contact of the modules A and B. Taking this data together, the authors were able to propose a detailed interaction network for the stable interactions within the entire human eIF3 which is shown in Figure 3.

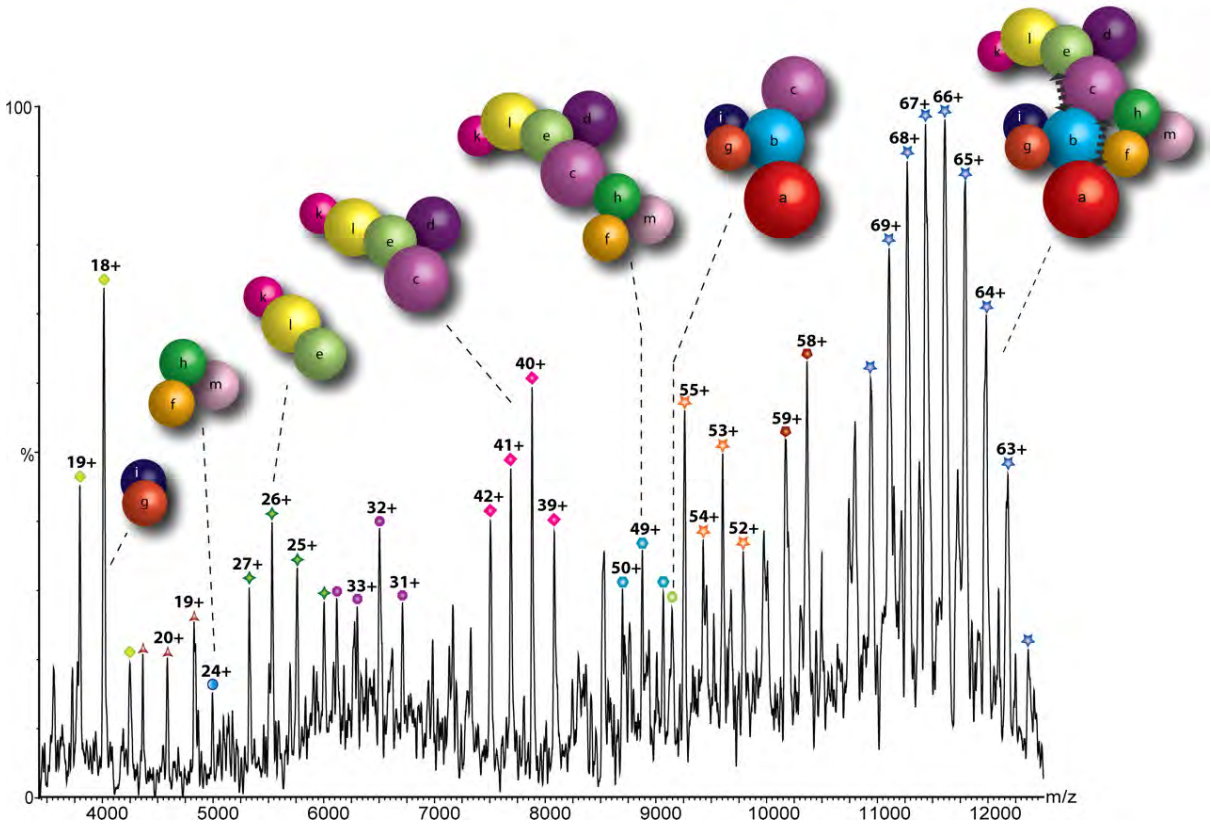


Figure 3 Model of the human eIF3 derived from 27 subcomplexes, IP data, and interactions identified in the yeast complex. The complex dissociates into distinct modules in response to changes in ionic strength; in this case, the spectrum shown was recorded at intermediate ionic strength (350 mM ammonium acetate). Arrows denote additional interactions not readily represented in this model.[12]

3.2 The atomic structure of yeast eIF3i:b(654-700)

As described above, the subcomplex of eIF3b:g:i participates in many steps during the initiation process. Despite its physiological significance, the structural information on this complex remains incomplete. The eIF3b is one of the major scaffolding subunits of eIF3 and associates with the other core subunits in both yeast and mammals. [12-14] Its N-terminal domain (NTD) contains a conserved RNA recognition motif (RRM) [15-16], which interacts with the C-terminal half of eIF3a and the NTD of eIF3j. Its middle domain is predicted to fold into two β-propeller structures [17], the second of which binds to eIF3c. Finally, the extreme C-terminal domain (CTD) of eIF3b is required for the binding of eIF3g and i. [14] While eIF3i is predicted to adopt a seven-bladed β-propeller structure made up of seven WD-40 repeats, eIF3g interacts with eIF3i and eIF3b through its NTD containing a predicted Zn-

finger domain [14], and its CTD contains another RRM not involved in any subunit-subunit interactions.

In 2012, during the course of this thesis, the first crystal structure of one subunit was reported among the core subunits in yeast eIF3: the eIF3i subunit in complex with the minimal CTD of eIF3b(654-700), and the binding region was resolved by solution X-ray crystallography. [18] The minimal C-terminal eIF3i-binding site of eIF3b(654–700) could be determined by X-ray crystallography, which does not contain any random coil segments that hinder crystallization. The crystallographic structure of this optimized yeast eIF3i:b(654–700) complex was determined at 2.2 Å.

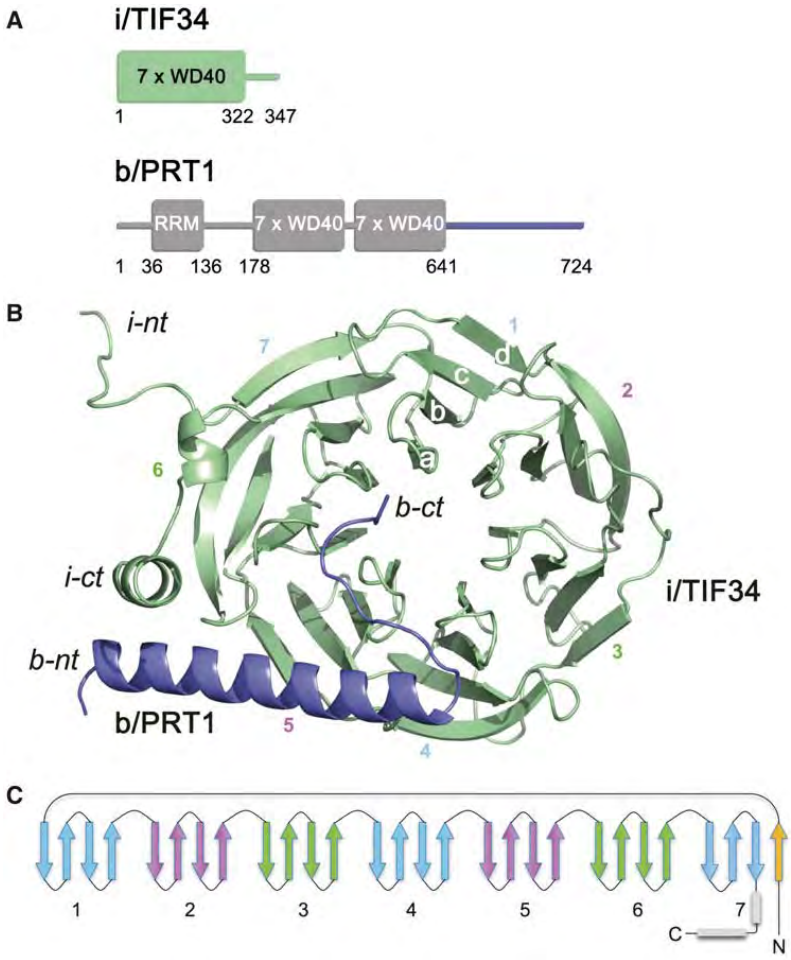


Figure 4 Structure of eIF3i:b(654–700). A: Schematic drawing of the predicted protein domains of yeast eIF3i (top) and eIF3b (bottom). The folded domains and their boundaries are indicated. Predicted unstructured regions are shown as lines. B: Overview of the structure: i (green) and b(654–700) (blue). C: Topology diagram of i fold. [18]

The eIF3i shows a seven-bladed β-propeller structure with a short and bent C-terminal α-helix (Figure 4 A, B). One strand of blade 7 is formed by N-terminal residues 1-7, while the

subsequent amino acid residues consecutively form blades 1–7 (Figure 4 C). The eIF3b(654–700) fragment binds to the bottom side of the β-propeller along the loops of blades 5 and 6 of eIF3i (Figure 4 B). Residues 663-689 in eIF3b form a long α-helix while the extended C-terminus (residues 690–699) advances towards the central cavity of the β-propeller in eIF3i.

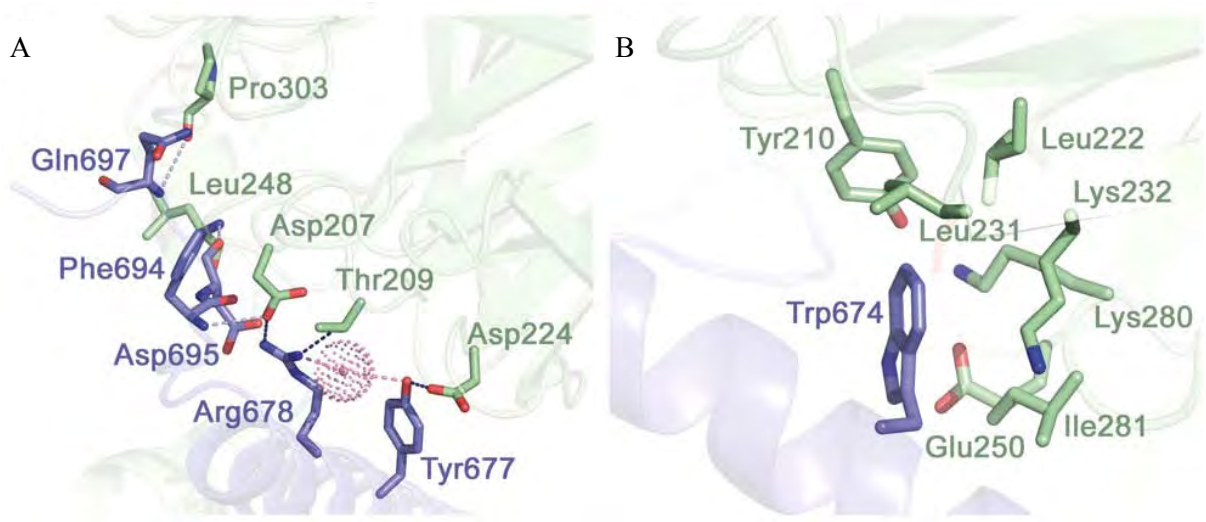


Figure 5 Molecular details of eIF3i:b interactions (eIF3i in green and eIF3b in blue). A: Y677 and R678 in eIF3b interact with D207, T209 and D224 in eIF3i via hydrogen bonds (shown in dark blue). Other residues making contacts (light blue) via main chain atoms are also shown in sticks. One water molecule (shown in dots) is in close proximity (light pink) to R678 (2.6Å) and Y677 (3.3 Å). B: eIF3b W674 is surrounded by hydrophobic and charged amino acid residues in a shallow pocket. [18]

This work also revealed the detailed binding surfaces according to the crystallographic results. Seven residues of the b(654–700) fragment are in close contact with eight eIF3i residues (<3.5 Å apart). Hydrogen bonds are formed between the most conserved residues Y677-D224 and R678-D207/T209 (Figure 5 A), and therefore these residues are crucial for eIF3i:b complex formation. The second important feature is the insertion of the aromatic ring of b W674 into a pocket of eIF3i formed by hydrophobic amino acids (L231 and I281 as well as the conserved L222) and the side chains of polar amino acids (Y210 and K280 as well as the conserved K232 and E250) (Figure 5 B). In addition, there is a series of interactions between main chain atoms of the extended C-terminal part of eIF3b from N690 and eIF3i residues in blades 5 and 6.

Despite this information, little is known about the structure and interactions of the subcomplex of eIF3b:g:i. In this work, we tried to unveil the binding of eIF3i:b and eIF3i:g using the method of HDX-MS and the partially resolved atomic structure of 3i:b.

4 Synthesis and purification of proteins of interest

In this work, three proteins or protein segments are independently studied: full-length subunit eIF3i with a His-tag at the N-terminus, C-terminal segment of eIF3b(641-724) with a His-tag t the C-terminus and N-terminal segment of eIF3g(42-170) with a His-tag at the N-terminus, the sequences being shown in Table 2. They will be referred to as eIF3i, eIF3bC3 and eIF3gC1ΔC through the work, respectively.

Table 2 Sequences of proteins synthesized in this work.

Protein / Mw	**Sequence**
eIF3i / 40918.7 Da	MGSSHHHHHHSSGLVPRGSHMKAIKLTGHERPLTQVKYNKEGDLLFSCSK DSSASVWYSLNGERLGTLDGHTGTIWSIDVDCFTKYCVTGSADYSIKLWD VSNGQCVATWKSPVPVKRVEFSPCGNYFLAILDNVMKNPGSINIYEIERDS ATHELTKVSEEPIHKIITHEGLDAATVAGWSTKGKYIIAGHKDGKISKYDVS NNYEYVDSIDLHEKSISDMQFSPDLTYFITSSRDTNSFLVDVSTLQVLKKYE TDCPLNTAVITPLKEFIILGGGQEAKDVTTTSANEGKFEARFYHKIFEEEIGR VQGHFGPLNTVAISPQGTSYASGGEDGFIRLHHFEKSYFDFKYDVEKAAEA KEHMQEAN
eIF3bC3 / 11324.5 Da	MLREWSAQFEEQDAMEADTAMRDLILHQRELLKQWTEYREKIGQEMEKS MNFKIFDVQPEDASDDFTTIEEIVEEVLEETKEKVELEHHHHHH
eIF3gC1ΔC / 16199 Da	MGSSHHHHHHSSGLVPRGSHMEKVHKSVAERKNWHKYGSEKGSPAGPSA VTARLGEEVELRLSRNWKQAEEERIQKEKASLTKTGLQCRLCGNDHMTM NCPFKTILSELSALEDPATNEGGVEAASEEKAGQVGGAGSIPGQYVPPSRR

If one compares the eIF3b CTD used in this work to the fragment for which a crystallographic structure was obtained, the eIF3b CTD encompasses residues 641-724 which is slightly extended both to the C and N –terminal sides compared to 654-700 fragment used in the work by Herrmannová et al. [18]

A detailed procedure of protein synthesis and purification is described in Part 5: Materials and methods. Briefly, for each of the above three proteins, the plasmid was transformed into cells and inoculated in E.coli. The harvested cells were disrupted by sonication, and the cell extract was centrifuged to get rid of cellular and membrane debris.

The supernatant was first subjected to fast protein liquid chromatography (FPLC) separation using a TALON affinity column. The resin binds specifically with the His-tag proteins, whereas the other proteins in the cell-extract went directly in the flowthrough. An example of

SDS-PAGE gel image of eIF3i after the TALON separation is shown in Figure 6 (left). eIF3i eluted in fractions 43-49 as marked in the black box, whereas the other proteins elutes in early fractions such as in fraction 2.

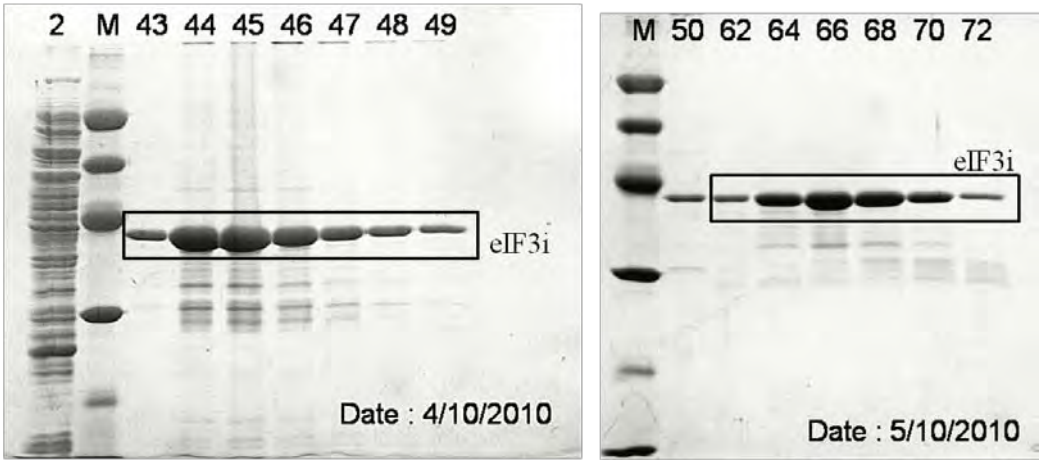


Figure 6 SDS-PAGE gel image of eIF3i after FPLC with TALON (left) and Superdex75 (right), marked with collected fraction numbers. Left: 2: proteins flowed through the colume. M: standard protein markers. 43-49: peak fractions containing eIF3i and de gradations. Right: M: standard protein markers. 50: peak fraction of aggregations. 62-72: peak fractions containing pure eIF3i.

The eluted peak fractions were then applied on a 20 mL Q-Hiload column, to which the nucleic acids bound and the protein directly flowed through. The collected fractions were concentrated up to 5 mL and applied onto a gel-filtration Superdex75 column. This column allowed the separation of larger size protein aggregates and the pure protein of interest was collected (example shown in Figure 6 (right)). The pure protein fractions were concentrated and the concentration was estimated by measuring absorbance at 280 nm.

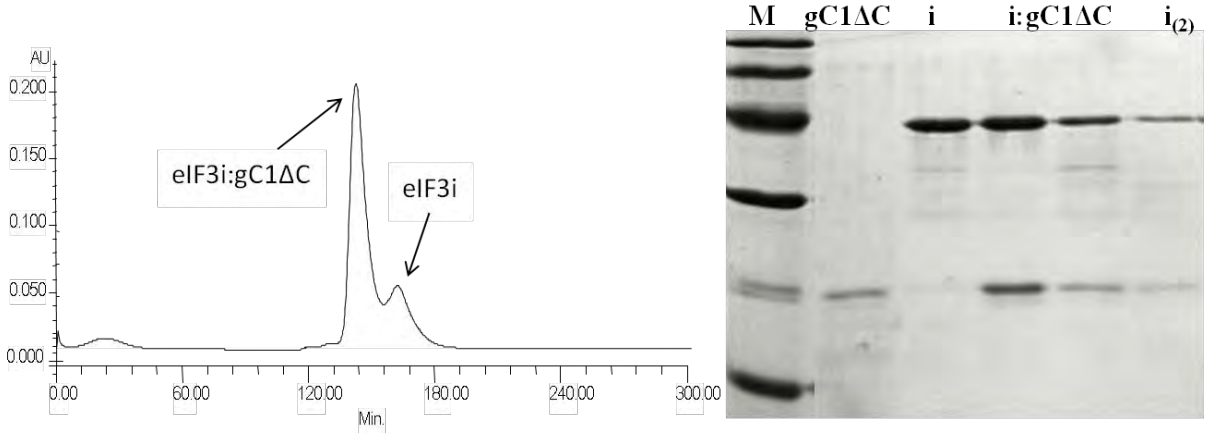


Figure 7 Separation of dimeric compex eIF3i:gC1ΔC. Left: FPLC spectrum with peaks of the eIF3i:gC1ΔC dimer (first) and excess of eIF3i monomer (second). Right: SDS-PAGE gel image. M: standard protein standard. gC1ΔC: reference band of eIF3gC1ΔC. i: reference band of eIF3i. i:gC1ΔC: two peak fractions from the first peak in the left spectrum. $I_{(2)}$: peak fraction from the second peak in the left spectrum.

Once the eIF3i, eIF3bC3 and eIF3gC1ΔC were independently purified, two dimeric complexes, eIF3i:bC3, eIF3i:gC1ΔC and a trimeric complex eIF3i:bC3:gC1ΔC were generated by mixing the subunits with a 20% excess of one subunit. The mixture was subjected to another FPLC separation with the Superdex75 column, and the complex eluted in earlier fractions. Figure 7 shows an example of eIF3i:gC1ΔC in the Superdex75 separation. As isolated on the gel, the fractions corresponding to the first peak contain a stoichiometric complex of eIF3i and eIF3gC1ΔC, indicating the presence of a stable dimer in these fractions.

References

1. Hinnebusch, A. G., eIF3: a versatile scaffold for translation initiation complexes. *Trends Biochem. Sci.* **2006,** *31*, 553-562.
2. Prichard, P.; Gilbert, J.; Shafritz, D.; Anderson, W., Factors for the initiation of haemoglobin synthesis by rabbit reticulocyte ribosomes. *Nature* **1970,** *226*, 511-514.
3. Benne, R.; Hershey, J., Purification and characterization of initiation factor IF-E3 from rabbit reticulocytes. *PNAS* **1976,** *73*, 3005-3009.
4. Chaudhuri, J.; Chowdhury, D.; Maitra, U., Distinct functions of eukaryotic translation initiation factors eIF1A and eIF3 in the formation of the 40 S ribosomal preinitiation complex. *J. Biol. Chem.* **1999,** *274*, 17975-17980.
5. Kolupaeva, V. G.; Unbehaun, A.; Lomakin, I. B.; Hellen, C. U.; Pestova, T. V., Binding of eukaryotic initiation factor 3 to ribosomal 40S subunits and its role in ribosomal dissociation and anti-association. *RNA* **2005,** *11*, 470-486.
6. Majumdar, R.; Bandyopadhyay, A.; Maitra, U., Mammalian translation initiation factor eIF1 functions with eIF1A and eIF3 in the formation of a stable 40 S preinitiation complex. *J. Biol. Chem.* **2003,** *278*, 6580-6587.
7. Siridechadilok, B.; Fraser, C. S.; Hall, R. J.; Doudna, J. A.; Nogales, E., Structural roles for human translation factor eIF3 in initiation of protein synthesis. *Science* **2005,** *310*, 1513-1515.
8. Hashem, Y.; des Georges, A.; Dhote, V.; Langlois, R.; Liao, H. Y.; Grassucci, R. A.; Hellen, C. U.; Pestova, T. V.; Frank, J., Structure of the Mammalian Ribosomal 43S Preinitiation Complex Bound to the Scanning Factor DHX29. *Cell* **2013,** *153*, 1108-1119.
9. Maag, D.; Fekete, C. A.; Gryczynski, Z.; Lorsch, J. R., A conformational change in the eukaryotic translation preinitiation complex and release of eIF1 signal recognition of the start codon. *Mol. Cell* **2005,** *17*, 265-275.
10. Hershey, J. W.; Merrick, W. C., 2 The Pathway and Mechanism of Initiation of Protein Synthesis. *Cold Spring Harbor Monograph Archive* **2000,** *39*, 33-88.
11. Phan, L.; Schoenfeld, L. W.; Valášek, L.; Nielsen, K. H.; Hinnebusch, A. G., A subcomplex of three eIF3 subunits binds eIF1 and eIF5 and stimulates ribosome binding of mRNA and tRNAiMet. *The EMBO journal* **2001,** *20*, 2954-2965.
12. Zhou, M.; Sandercock, A. M.; Fraser, C. S.; Ridlova, G.; Stephens, E.; Schenauer, M. R.; Yokoi-Fong, T.; Barsky, D.; Leary, J. A.; Hershey, J. W., Mass spectrometry reveals modularity and a complete subunit interaction map of the eukaryotic translation factor eIF3. *PNAS* **2008,** *105*, 18139-18144.
13. Fraser, C. S.; Lee, J. Y.; Mayeur, G. L.; Bushell, M.; Doudna, J. A.; Hershey, J. W., The j-subunit of human translation initiation factor eIF3 is required for the stable binding of eIF3 and its subcomplexes to 40 S ribosomal subunits in vitro. *J. Biol. Chem.* **2004,** *279*, 8946-8956.
14. Asano, K.; Phan, L.; Anderson, J.; Hinnebusch, A. G., Complex formation by all five homologues of mammalian translation initiation factor 3 subunits from yeast Saccharomyces cerevisiae. *J. Biol. Chem.* **1998,** *273*, 18573-18585.
15. ElAntak, L.; Tzakos, A. G.; Locker, N.; Lukavsky, P. J., Structure of eIF3b RNA recognition motif and its interaction with eIF3j: structural insights into the recruitment of eIF3b to the 40 S ribosomal subunit. *J. Biol. Chem.* **2007,** *282*, 8165-8174.
16. ElAntak, L.; Wagner, S.; Herrmannová, A.; Karásková, M.; Rutkai, E.; Lukavsky, P. J.; Valášek, L., The indispensable N-terminal half of eIF3j/HCR1 cooperates with its structurally conserved binding partner eIF3b/PRT1-RRM and with eIF1A in stringent AUG selection. *J. Mol. Biol.* **2010,** *396*, 1097-1116.

17. Marintchev, A.; Wagner, G., Translation initiation: structures, mechanisms and evolution. *Q. Rev. Biophys.* **2004,** *37*, 197-284.
18. Herrmannová, A.; Daujotytė, D.; Yang, J.-C.; Cuchalová, L.; Gorrec, F.; Wagner, S.; Dányi, I.; Lukavsky, P. J.; Valášek, L. S., Structural analysis of an eIF3 subcomplex reveals conserved interactions required for a stable and proper translation pre-initiation complex assembly. *Nucleic Acids Res.* **2012,** *40*, 2294-2311.

Part II

Results

Chapter 3

Method development in the bottom-up HDX approach

1 Initial workflow of HDX experiments and data processing steps 69
1.1 The HDX-MS experiments 69
1.2 Initial data processing workflow 70
1.2.1 Retention time correction 71
1.2.2 Establishment of the peptide reference list 72
1.2.3 Peak assignment for deuterated peptides 74
1.2.4 Calculation of deuteration level and kinetics plotting 77
2 Method improvements in the establishment of peptide reference list 77
2.1 Insufficiency of mass accuracy for peptic peptide identification 78
2.2 Pepsin specificity as another approach for peptic peptide identification 80
2.2.1 Probability of occurrence of a peptide 80
2.2.2 POc filter and mass error filter 83
2.2.3 Resolving ambiguous assignments 85
2.2.4 Reproducibility control 87
2.2.5 Method validation by MS/MS and evaluation 87
3 Method improvements in the processing of deuterated sample data 92
3.1 Mass recalibration 92
3.2 Program upgrades for peak assignments of deuterated samples 92
References 94

1 Initial workflow of HDX experiments and data processing steps

The basic methodology followed in this thesis derives from a previous work by Magalie Duchateau [1], including the exchange step and digestion strategy, the chromatography gradient, the FT-ICR mass measurement and the data processing. In this section, the methods followed for this work will be detailed, with an emphasis on the improvements added to the original methodology.

1.1 The HDX-MS experiments

A complete discussion of the experimental condition optimization has been performed in a previous PhD thesis in our group [1] and the adapted protocol in this work is described in details in Chapter 5 “Materials and methods”. Briefly, the experimental session starts with the incubation of a purified protein sample in D_2O on ice, where the deuterium exchange takes place. After various durations, the deuterated protein sample is mixed with an identical volume of 6% AcOH in H_2O on ice, which quenches the exchange reaction.

In the initial protocol, the deuterated protein samples were rapidly frozen by liquid nitrogen and stored at -80 °C, and are thawed shortly before the analysis. This was convenient for experimental handling and for saving protein samples, as all deuterated samples could be prepared with enough replicates within two days and a smaller amount of protein samples were consumed. However, back-exchange may be induced during the freezing and thawing steps. Thus in this work, after having encountered some major issues with the freeze / thaw protocol, instead of performing the deuteration of protein with the amount required for all the kinetic points and replicates and taking aliquots of proteins out of the deuteration solution, we decided to perform each deuteration experiment separately: deuteration of protein with the amount required for one LC-MS acquisition is performed, and the deuterated samples are digested by pepsin immediately after quenching. The peptic digests are then immediately injected into the nanoLC system and analyzed by FT-ICR MS. In a separate experiment, unlabelled proteins are digested by pepsin and measured similarly. These data are used as the reference data to allow the identification of the peptides within the deuterated samples.

1.2 Initial data processing workflow

Figure 1 illustrates the data processing workflow that was used in the previous work and adapted at the beginning of this work.

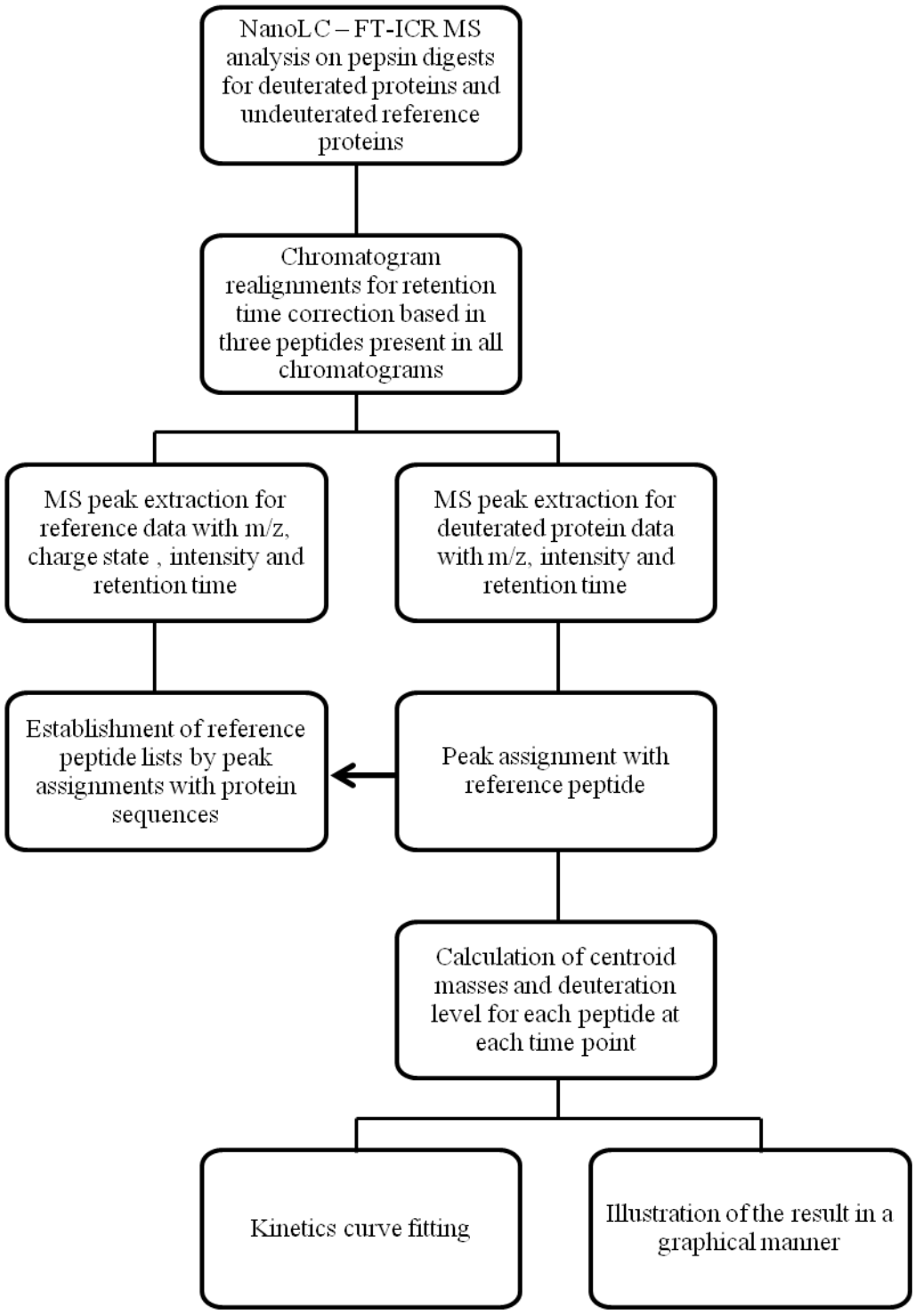


Figure 1 Initial data processing workflow.

1.2.1 Retention time correction

The data processing starts with the acquisition of the LC-MS spectra of unlabelled protein digests (referred to as "reference peptides" from here on) and exchanged protein digests. As the gradient of the chromatography was designed to be accomplished as quickly as possible, a retention time correction was necessary to account for some time shifts between chromatograms and allow the use of the retention time for peptide identification. For each protein or protein complex, three peptides were selected based on the following criterion: first, they should be present in all chromatograms; second, they should be peptic auto-digested peptides which present a low deuterium incorporation level. Their experimental retention times were extracted from the extracted ion chromatograms (EICs) as t_1, t_2 and t_3, respectively. Once all the t_1, t_2 and t_3 from the n sets of experimental data were listed, the average retention time of each peptide was calculated as $\bar{t_1}$, $\bar{t_2}$ and $\bar{t_3}$:

$$\bar{t} = \frac{\sum_1^n t}{n} \qquad (1)$$

The difference for each t and $\bar{t}$ was calculated as

$$\Delta t = t - \bar{t} \qquad (2)$$

and a least square approximation was carried out based on Δt and $\bar{t}$.

$$\Delta t = a \bullet \bar{t} + b \qquad (3)$$

where

$$a = \frac{\sum_1^3 (\bar{t_i} - \bar{\bar{t}})(\Delta t_i - \overline{\Delta t})}{\sum_1^3 (\bar{t_i} - \bar{\bar{t}})^2} \qquad (4)$$

$$b = \overline{\Delta t} - a \bullet \bar{\bar{t}} \qquad (5)$$

with $\bar{\bar{t}}$ representing the average value of $\bar{t}$. Thereby the chromatogram realignment parameters α and β for each chromatogram were calculated as

$$\alpha = - \frac{b}{a+1} \quad (6)$$

$$\beta = \frac{1}{a+1} \quad (7)$$

All experimental retention times were then corrected as

$$t_{corrected} = \beta \bullet t_{experimental} + \alpha \quad (8)$$

1.2.2 Establishment of the peptide reference list

The bottom-up HDX-MS is based on the measurement of deuterium exchange of each peptide in the enzyme (pepsin in most cases) digest. In order to correctly assign an exchanged peptide isotope envelope to a given peptide, a pre-established peptide reference list is required. This list will allow peptide identification based on its accurate mass and its retention time.

In the initial workflow, the peptide reference list was established as follows:

Experimental peaks were extracted and time-grouped from the raw LC-MS spectra of reference proteins with information on m/z, charge state, retention time and intensity for each isotopic distribution identified in the initial data. As LC/MS made it difficult to introduce internal calibration during the whole LC run, an internal recalibration was performed based on the theoretical peaks from an unspecific digest of the protein of interest. For this procedure, a 20 ppm mass accuracy was used for inclusion of peaks in the calibration list. After recalibration, assignment of experimental peaks to putative peptides (derived from the protein sequence) was also performed with a 20 ppm mass error tolerance.

Due to the extreme complexity of pepsin digest caused by its lack of strict specificity, this procedure was not accurate, and a high number of ambiguous peak identification always occurred. An ambiguous peptide, as used further in the manuscript, is designated as an experimental mass peak that can be assigned to at least two possible peptides from the protein (or proteins) of interest. The initial strategy to resolve the ambiguities was based on the literature preferential specificity of pepsin. According to

the nomenclature proposed by Schechter and Berger [2-3] that is used throughout this work, the amino acid residues of a protein digested by an enzyme are designated as P1, P2, P3 and P4 on the N-terminal of the cleaved peptide bond and P1', P2', P3', P4' on the C-terminal. The corresponding residues as the reacting sites on the enzyme of pepsin are designated as S1, S2, S3 and S4 on the N-terminal and S1', S2', S3', S4' on the C-terminal. Pepsin cuts preferably with Phe, Tyr, Trp, and Leu in the P1 or P1' sites (Figure 2, top). [4] Inhibition of cleavage is observed in the presence of Arg, Lys or His at the P3 site. This inhibition is also observed when Arg is located at position P1. In addition, the presence of Pro promotes cleavage of the bond at positions P3 and P4, but inhibits the cleavage when the latter is in position between P2 and P3' (Figure 2 bottom).

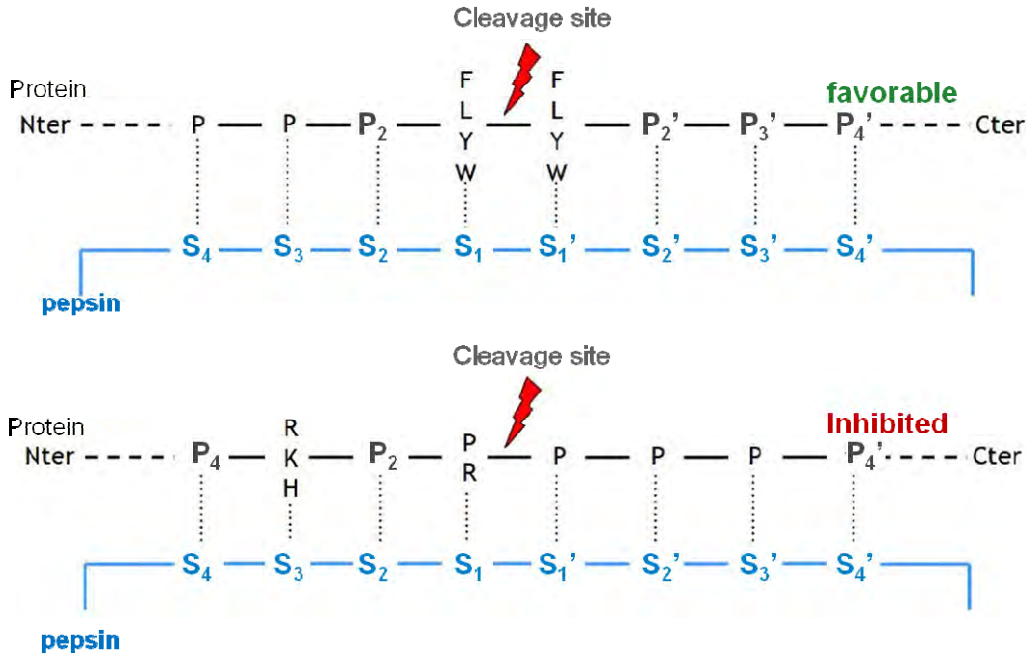


Figure 2 Scheme of experiential pepsin cleavage preference. Top: favorable cleavage sites. Bottom: inhibited cleavage sites.

Based on these empirical results, a score was assigned to each potential peptide. For each peptide, the cleavage probability was considered at both ends. For each end, for each residue located in a favorable cleavage position, a positive point was given; and for each residue located in an inhibited cleavage position, a negative point was given. A peptide was considered valid if both termini were rated greater than or equal to 1. Also a peptide is considered *potentially valid* if one terminus has a rating greater or equal to 1 and the other a score of zero. A less specific cleavage at either terminus

was therefore tolerated, and it was considered that the presence of two hydrophobic residues at the cleavage site compensated the presence of a Pro or a basic residue in inhibiting positions. The scoring process facilitated the selection of potential peptides in the cases of ambiguities. This method has been further improved in this work, which will be discussed in details in Section 2.

1.2.3 Peak assignment for deuterated peptides

Once the reference peptide list was established, it was used to assign the experimental data of deuterated samples. After peak extraction, an experimental peak list was generated containing m/z, realigned retention time (as described in 1.1) and intensity.

Such a peak list was difficult to process, as the mass shifted peak patterns by deuteration could not be readily assigned using the usual algorithms for high-resolution data, which validate a measured mass by comparison of the isotopic pattern with the reference peptides. Specific software was designed with the home-built *ProtDBSearch* program modules, in order to process the HDX-MS data obtained with ppm-order mass accuracy. Figure 3 shows the entry window of the software with all parameters that could be customized.

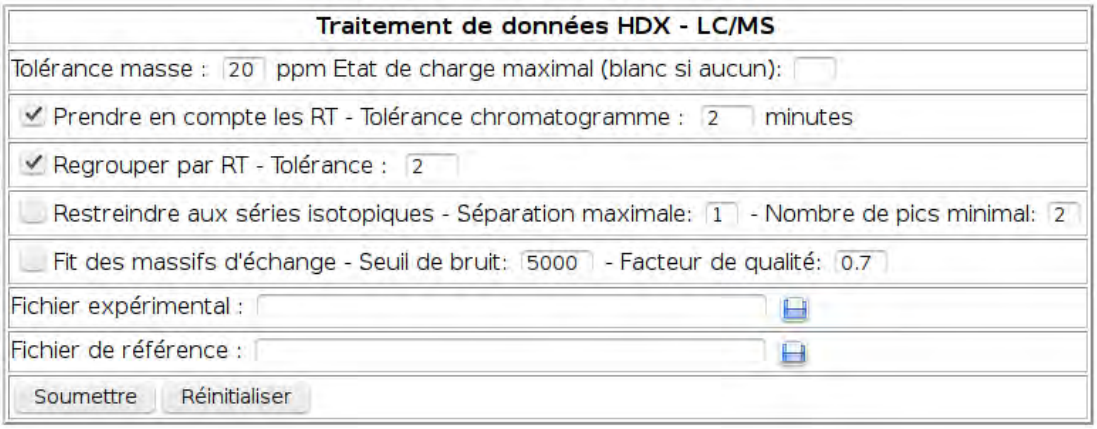


Figure 3 Parameter entry window of the home-built HDX data assignment program.

The first filter was to find peaks that were likely to belong to the deuterated state of a given peptide. The principle (Figure 4) was that for each reference peptide, a filter was constructed based on the theoretical masses of MH_i^{i+}, $MH_i^{i+}d1$, $MH_i^{i+}d2$, ... $MH_i^{i+}dn$ with $MH_i^{i+}dn$ representing the i times charged m/z value with n amide hydrogen exchanged by deuterium. The value n was determined as the sum of the maximum number of possible deuterium, and the number of isotopic peaks whose monoisotopic peak intensity was greater than 5% (calculated by a averagine peptide

of similar size [1][5]), with an additional 20% margin to allow for partial labeling of side chains.

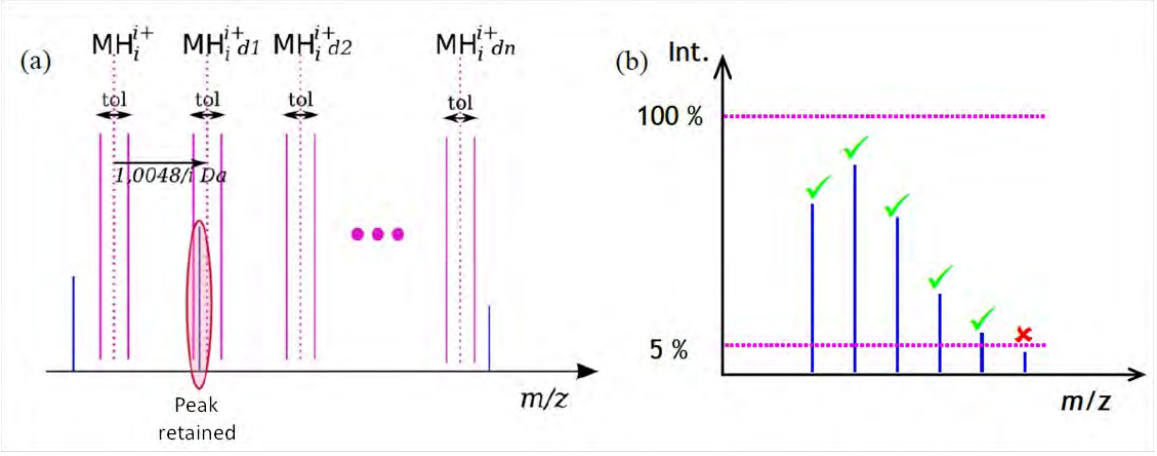


Figure 4 Schematic description of filters (pink) applied on experimental peaks (blue). (a) Filter to include peaks that may belong to a deuterated peptide. (b) Determination of the value n.

After the first filter, a series of potential experimental peaks is associated with each reference peptide. The following filters could be then selected to refine the assignment:

- Retention time

This filter requires that the experimental peaks lie within a certain window of retention time relative to the retention time measured for the reference peptide. A regrouping step was also possible at this stage. If a peak appears in several spectra within a given retention time range, the m/z values and intensities are replaced by the average value (weighted by the intensity) of the m/z and the sum of the intensities.

- Consistence of isotopic peak series

[1] Rigorously, the mass shift between peaks should be different between $^{1}H/^{2}H$, $^{12}C/^{13}C$ or $^{14}N/^{15}N$ as 1.006277, 1.003355 and 0.997035 Da, respectively. In the averagine simulation, an average shift of 1.0031 between the isotopic peaks for peptides without H/D exchange is applied. This raises a problem for the filter, because for low levels of deuteration, an average spacing is close to 1.0031, yet for high levels of deuteration, this shift moves gradually to 1.0063. A shift was chosen in our approach as the average value of 1.0048 Da. The mass error is thus 1.7 mDa for peaks without H/D exchange and -1.7 mDa for peaks with H/D exchange, so that for a peptide of 1000 Da, the mass error is within ±1.7 ppm. This error is not negligible; therefore in the current state of the program, a mass error tolerance of at least 5 ppm should be applied.

In an H/D exchange experiment, it was expected that the isotopic series appeared as a series of peaks separated by close to 1 Da. This was used as another filter by limiting a minimum number of peaks in a series and tolerance of missing peaks within the series. Figure 5 illustrated how a given series of peaks were selected based on two different peak series filters.

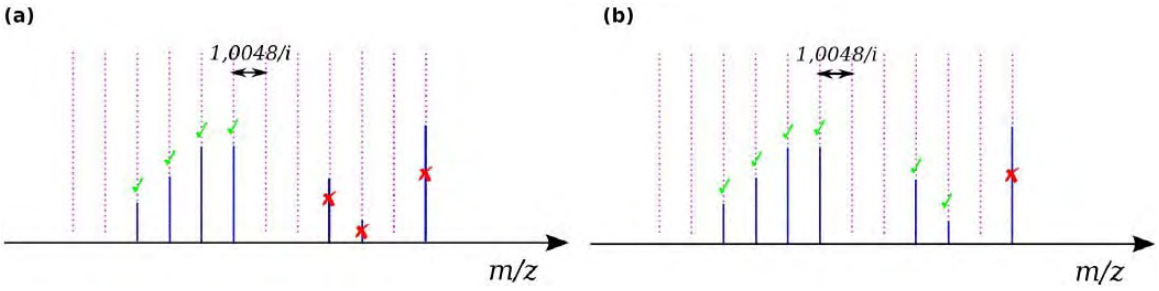


Figure 5 Illustration of the isotopic peak series filter. (a) selection upon criterion of at least three peaks and maximum one missing peak in the series. (b) selection upon criterion of at least two peaks and no missing peak in the series.

- Pseudo-Gaussian isotopic distribution

The last filter implemented in the program selects peak series with intensities close to a Gaussian distribution. For HDX that follow a EX2 kinetics, a quasi-Gaussian distribution can be expected, which does not apply to EX1 kinetics. In addition, extreme cases (without exchange or full exchange) are not Gaussian distributed.

However, with a sufficiently low tolerance ($\chi^2 > 0.7$), this filter can discriminate noises from true signals. It particularly compensates the weaknesses of the previous filters, which were not based on intensities, thus a "hole" in a series (e.g. a background noise with low intensity in the position of an intermediate mass) may not be readily identified, while the series was obviously incomplete or false. The Gaussian filter is sensitive to sudden changes in intensity (downward or upward) in a peak series and can therefore retain only series with a profile of reasonable intensities.

The program outputs a list of peptides present in the LC/MS spectra with the deuterated mass (intensity-weighted average mass of the isotopic peak series) and a complete list of triplets (RT, m/z and intensity) used to calculate the deuterated mass. In the situation where the triplets could be attributed to more than one peptide, a label is placed and a manual check by the user should be performed to resolve such ambiguities.

1.2.4 Calculation of deuteration level and kinetics plotting

Once validated, the deuteration level (%D) of each peptide was calculated as:

$$\%D = \frac{deuterated\ mass - MH^{+}_{reference}}{\#D_{max}} \quad (9)$$

where the maximum number of exchanged deuterium ($\#D_{max}$) for a given peptide was calculated as:

$$\#D_{max} = \#\text{amino acid residues} - 1 - \#\text{Pro} \quad (10)$$

with #amino acid residues being the number of total amino acid residues in the peptide and #Pro being the proline residues in this sequence.

When the deuteration levels for a given peptide at different time points were all acquired, they could be plotted against the time, giving a series of points with a positive reaction profile of single or double exponential. A curve fitting could be carried out according to Eq. 11 in order to determine the rate constants, which has been described in 3.4.3 in Chapter 1.

$$\%D_t = A_0 + A_1(1 - \exp[-k_1 t]) + A_2(1 - \exp[-k_2 t]) \quad (11)$$

2 Method improvements in the establishment of peptide reference list

At the beginning of this work in October 2010, an FT-ICR mass spectrometer of the type Apex III (Bruker, Bremen, Germany) coupled with a 7 Tesla magnet was used for MS data acquisition, and the data processing adapted the workflow as described above. This instrument was upgraded to an Apex Q platform (Bruker, Bremen, Germany) in January 2011, and again to a SolariX platform (Bruker, Bremen, Germany) in March 2012. A lthough the superconducting magnet and the ICR cell remained the same, these changes in the ion source indeed increased the sensitivity of the mass spectrometer, which in turns required changes and improvements to the data processing method.

2.1 Insufficiency of mass accuracy for peptic peptide identification

The number of experimental peaks massively increased when the instrument was updated from Apex III to Apex Q. A typical LC/MS on a reference protein gave 100-200 peaks when it was acquired on the Apex III, and the sample of the same pretreatment resulted in around 1600 peaks on the Apex Q.

A large number of peak assignments are thereby generated. An example of the reference sample of eIF3i:bC3 is taken to demonstrate the method. Acquired on the APEX Q instrument, this data set provides 1596 peaks. It brings around 1600 matches after peak assignment to the protein sequence as described in 1.2. The number of assignments within each 1 ppm gap is plotted against the mass error as shown in Figure 6 in red. A clear clustering towards the small error region can be observed, but a large number of assignments remain in the large error regions, indicating a high "baseline" of false identifications.

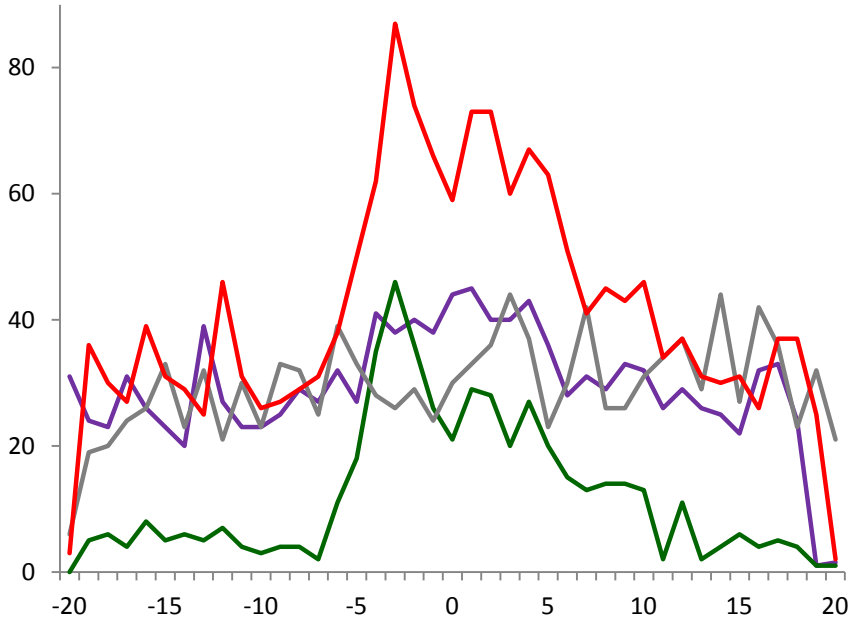


Figure 6 Error distributions of an example data set of eIF3i:bC3. The number of the assigned peaks within certain error range is plotted against mass error. Red: all assignments based on the real protein sequence. Grey: all assignments based on the randomized sequence. Purple: assignments based on the real protein sequence with POc values lower than the POc limit. Green: assignments based on the real protein sequence with POc values higher than the POc limit.

To better access the level of false identification, a randomized sequence was generated. This sequence is composed with the same amino acid residues as

eIF3i:bC3 but in a totally random order. The example data set is assigned with this "fake" sequence, giving around 1200 matches. These matches distributed along the mass error as the grey curve in Figure 6. No obvious clustering is observed for these assignments and an average of 30 assignments using the randomized sequence within each 1 ppm region indicates that within the assignments using the real protein sequence, similar number of false assignments is likely to be embedded.

The larger the protein is, the more complex the digest becomes, and therefore the more assignments can be matched in the LC/MS data. Such increases in the number of assignments are shown in Table 1, together with the increases of false assignments analogized by the number of assignments using the randomized sequences.

Table 1 Examples of peak assignments for proteins with different sizes.

	eIF3bC3	eIF3i:bC3	aIF2Mj
No. of AAs	93	460	817
Mw (kDa)	11.3	52	91.4
No. of assigned peaks with real sequence (±5 ppm)	224	664	778
No. of assigned peaks with randomized sequence (±5 ppm)	21	429	455

The standard strategy to reduce these false assignments is tandem MS. In order to retain the RT information for the peak assignments of the HDX data, LC-MS/MS is usually carried out in the same LC condition as for the HDX sample. Such a fast gradient brings difficulty to the MS/MS experiment as the number of peptides coeluting within a short time range can be much more than those with a long gradient, which is the common case for standard proteomics MS. The harsh LC condition, as well as the fact that peptic peptides are much more difficult to fragment than tryptic peptides, make it difficult to generate a p eptide reference list with high sequence coverage and rich sequence overlaps.

2.2 Pepsin specificity as another approach for peptic peptide identification

2.2.1 Probability of occurrence of a peptide

We have described the method in the initial data processing workflow using pepsin specificity to validate the identified peptides in 1.2. This prototype of specificity filter is further developed is this work. In 2008, a study was reported based on a large number of MS/MS results of porcine pepsin digests under conditions compatible with hydrogen/deuterium (H/D) exchange (<1°C, <30 s on immobilized pepsin column) [6]. The observation of favorable and inhibited cleavage sites is coherent with previous observations, and importantly, this statistical analysis provided a table of probability of cleavage at different sites (Table 2). The residue on P1 is read from the column and the residue on P1' is read from the line. For example, the cleavage between F-Y is observed in 65% of the cases, while between Y-F it is observed in 33% of the cases.

Table 2 Pepsin specificity by P1-P1' positions [6].

	F	L	M	C	E	W	Y	D	A	Q	N	T	S	G	V	I	K	H	R	P	Ave.
Y	65%	68%	30%	25%	70%	71%	36%	48%	48%	47%	39%	23%	31%	17%	11%	7%	0%	0%	0%	0%	33%
F	85%	84%	64%	75%	53%	40%	33%	37%	38%	24%	28%	21%	5%	9%	8%	8%	0%	0%	0%	0%	28%
W	60%	60%	50%	57%	63%	-	50%	17%	17%	33%	38%	17%	23%	17%	24%	30%	0%	0%	0%	0%	28%
I	65%	63%	62%	40%	36%	15%	20%	42%	30%	40%	25%	18%	13%	9%	9%	0%	2%	0%	0%	0%	24%
M	83%	58%	42%	0%	29%	33%	25%	20%	30%	20%	0%	11%	0%	11%	6%	0%	0%	0%	0%	0%	21%
V	50%	53%	61%	21%	31%	10%	16%	25%	28%	23%	16%	11%	12%	0%	4%	2%	0%	0%	0%	0%	19%
L	64%	56%	66%	36%	21%	29%	25%	12%	16%	7%	4%	8%	4%	13%	7%	1%	0%	0%	0%	0%	18%
C	20%	54%	0%	50%	36%	22%	0%	33%	6%	33%	25%	0%	20%	12%	0%	18%	0%	0%	0%	0%	17%
A	55%	54%	38%	18%	35%	8%	9%	13%	14%	7%	9%	7%	6%	4%	3%	2%	0%	0%	0%	0%	16%
E	42%	45%	29%	20%	9%	24%	19%	6%	6%	2%	0%	0%	4%	0%	4%	6%	0%	0%	0%	0%	10%
D	44%	46%	38%	0%	11%	21%	17%	5%	5%	0%	5%	2%	2%	4%	0%	0%	0%	0%	0%	0%	10%
R	42%	34%	26%	29%	9%	13%	16%	9%	8%	6%	4%	0%	5%	3%	0%	0%	0%	0%	0%	0%	10%
N	42%	45%	7%	0%	13%	0%	11%	0%	4%	5%	0%	0%	4%	0%	0%	0%	0%	0%	0%	0%	9%
S	52%	42%	22%	0%	4%	6%	14%	2%	5%	0%	0%	5%	2%	0%	3%	0%	0%	0%	0%	0%	9%
T	31%	27%	25%	35%	3%	29%	4%	12%	5%	3%	6%	9%	3%	2%	0%	0%	0%	0%	0%	0%	9%
H	43%	33%	29%	17%	6%	0%	10%	15%	22%	15%	0%	11%	3%	0%	0%	0%	0%	0%	0%	0%	9%
K	47%	33%	32%	0%	12%	13%	0%	2%	3%	3%	0%	4%	0%	2%	0%	2%	0%	0%	0%	0%	8%
Q	33%	26%	17%	0%	11%	0%	20%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	6%
P	17%	24%	18%	20%	8%	0%	6%	0%	5%	0%	0%	0%	2%	0%	11%	2%	0%	5%	0%	0%	6%
G	28%	7%	8%	6%	1%	0%	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	0%	0%	0%	3%
Ave.	46%	44%	35%	23%	20%	17%	16%	13%	13%	10%	9%	7%	5%	4%	4%	2%	0%	0%	0%	0%	14%

By itself, these data do not allow readily an inversion. The percentages given above represent only the fraction of sites for which *at least* one cleavage was observed in the experimental data set relative to the total number of these types of sites within the observed data. This is not directly extendable to the proportion of cleavage at a given position since this value does not include peptides that have been observed to contain the intact cleavage site. Our method is based on a postulate: the data given above, although not strictly measuring the probability that pepsin cleaves at a given site, can

be used as such. It is quite obvious that for the lowest values of 0% is the same. But on the other hand, the highest values are likely to be overestimated, as in the conditions used for pepsin digestion in H/D exchange, only partial digestion is generally performed, and rates of cleavage on the order of 85% at a given site are quite rare.

Nevertheless, the choice was done to perform a statistical prediction of peptides based on the values given in the above table, to check if such an approach was valuable. In the case of good results, a further work could encompass including a true "probability of cleavage" at a specific site. This work would not be straightforward at all: the intensity of the peptides should be taken into account, which would be problematic considering varying ionization and measurement efficiencies between peptides, inclusion of both very small and very large peptides in the data set, …

Assuming that we can use the probabilities given in Table 2 as the ratio of cleavage versus non-cleavage at a given site, we therefore define a term, probability of occurrence (POc), representing the probability for a given peptide from a known protein sequence to be present in the peptic digest. The algorithm to generate this POc value is as follows:

A table containing all the sequences that have to be processed is kept and associated with its probability $P_{b\text{-}e}$ (b is the index of the N-terminal residue and e that of the C-terminal residue). This table is initialized with the complete protein and a probability of 1.

The possible cleavage sites are processed from the N-terminal to the C-terminal, although the results are the same if the processing is done the other way around. The peptides are denoted by their starting and ending position relative to the numbering of residues in the sequence.

1°) The probability of cleavage at this position ($P_{current}$) is looked up in Table 2, based on the residues present in position P1 and P1' relative to the current site.

2°) For each peptide in the sequences table that have the current cleavage position (all peptides except those that have their C-terminus before the current positions), which has a starting (N-terminal) position b and an ending position e (C-terminal) three peptides are generated:

- A N-terminal peptide ranging from position b to the current position, the probability of which is calculated as:

$$P_{b-current} = P_{b-e} \times P_{current} \quad (12)$$

- A C-terminal peptides ranging from the current position the position e for which the probability is calculated as follows:

$$P_{(current+1)-e} = P_{b-e} \times P_{current} \quad (13)$$

- The probability for the current peptide to remain non-cleaved is also updated to remove the probability of cleavage described previously, thus the new $P_{b\text{-}e}$ is updated as follows:

$$P_{b-e} = (1 - P_{current}) \times P_{b-e} \quad (14)$$

3°) As several peptides in the table can lead to the same peptide after cleavage, the table is merged by adding together the probabilities for all peptides having the same N-terminal and C-terminal position.

4°) In order to improve algorithm speed, peptides with a score below a given threshold are removed from the peptide sequence list. The user can enter the lowest POc value that should be considered as relevant (10^{-12} is used in this work). The threshold is set based on this threshold and on the length of the protein at:

$$Threshhold = \frac{1}{\#_{residues\ in\ protein}} (User\ defined\ minimum)^2 \quad (15)$$

5°) Peptides that have their C-terminal end before the current position will not be altered further and are placed in a separate list to speed up further processing.

6°) Iteration goes back to step 1 increasing current position by one unit, until the C-terminus is reached.

A schematic of the algorithm for a 4 residue peptide is provided in Figure 7 to help understanding the process.

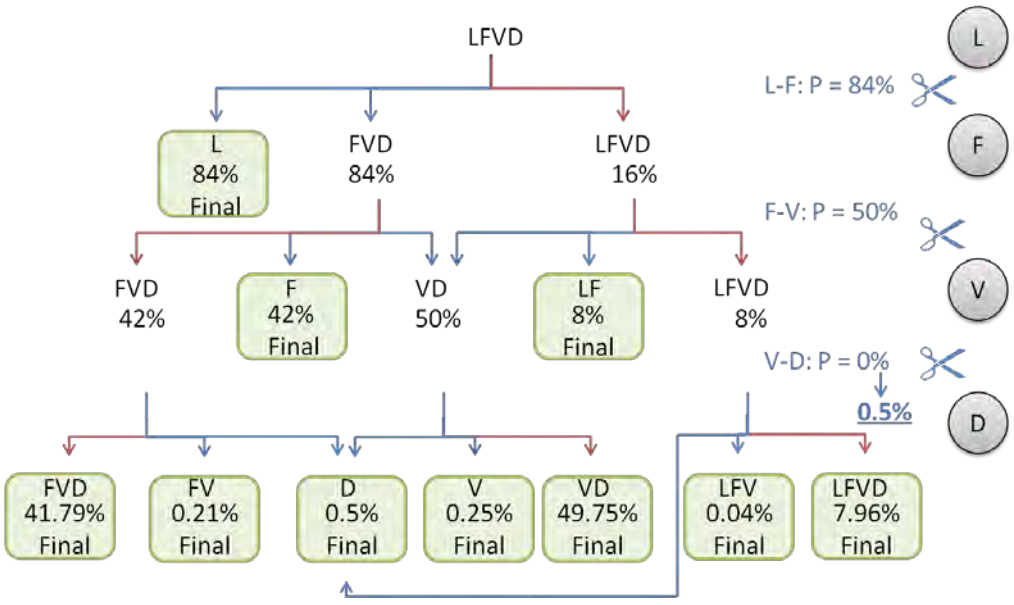


Figure 7 Example scheme of the algorithm for POc generation.

As can be seen in Figure 7, the two step process allows mimicking a processing by enzyme. The resulting values lead to having the probability of cleavage for each site equal (up to the approximation level) to the sum of POc values for peptides C- (or N-) terminated at this site.

As will be described further, the initial values were exactly those found in the Table 2. After some work and validation, some MS/MS validated peptides were found that did contain a cleavage position for which a value of 0% was present in the table. This was to be expected, as quite obviously the quality of Table 2 is dependent on the statistics for each cleavage position. The choice was then done to increase **all** values at 0% in Table 2 to 0.5%. This did not strongly alter the results for most peptides, but led to consider some peptides that had a high probability for cleavage at one end and a 0% probability at the other end, which were otherwise attributed a POc of 0.

2.2.2 POc filter and mass error filter

Once an LC/MS data set is assigned with the protein sequence to a list of peptides all labeled with the POc value, a POc filter is used as a first selection. As shown in 2.1, five randomized sequences are generated and used to assign the experimental data, and the average number of resulting matches, $\#peptide_{rand}$, is taken as the analog of the false assignment rate. The POc threshold is then set as that the number of

assignments with POc values below this limit is equal to the number of false assignment:

$$\#peptide_{rand} = \#peptide_{low\ POc} \quad (16)$$

and these low-POc assignments are excluded from the list. The green and purple curves in Figure 6 show the population distribution of the assignments with POc values above and below the threshold, respectively. The excluded population distributes randomly along the mass error value, while the POc-survived population clusters clearly within the small mass error range with a much lowered baseline, i.e. a significantly reduced false identification level.

Within the high-POc assignments, a narrower mass error tolerance is then applied, decided upon the performance of the instrument. Throughout this work, a ±7 ppm mass filter is used on data acquired in the APEX Q F T-ICR MS, although this is higher than the commonly achievable mass resolution for FT-ICR mass spectrometer. It is fixed based on the trend in the green curve in Figure 6, as the assignments cluster within the ±7 region.

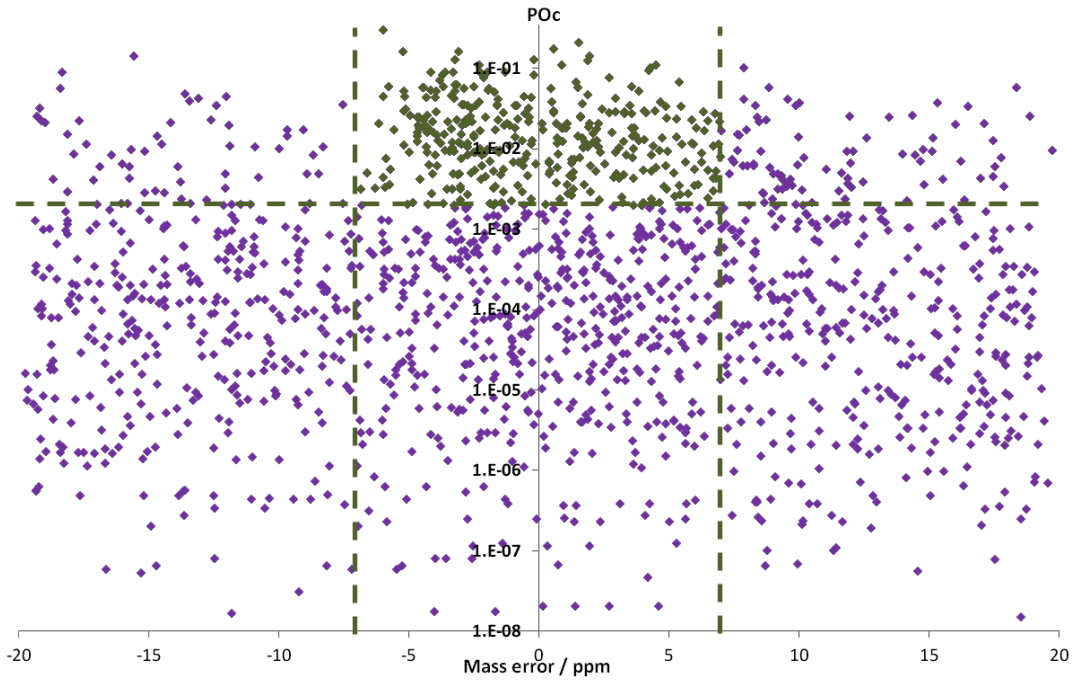


Figure 8 Plotting of peak assignments with POc values against mass errors. Green: assignments survived in both filters. Purple: assignments excluded by either of the two filters. The green dotted line shows the POc threshold for this data set.

Figure 8 is a detailed plotting of the POc value against the mass error for each assignment in an example data set. The points in purple are excluded due to low POc and/or high mass error, whereas the green points are kept at this stage. The random distribution of the purple data points is a proof of the efficiency of these filters.

2.2.3 Resolving ambiguous assignments

Due to the complexity of the pepsin digest, it often occurs that one experimental peak can be assigned to more than one peptide, the mass differences of which are within the mass error tolerance and therefore not easy to distinguish. The above filters significantly reduced the number of ambiguities, however, not completely. Table 3 shows several examples of ambiguities present in the list after the POc and mass error selections. All these experimental peaks are present in both the spectra acquired in the HDX-MS standard condition as well as a validation LC-MS/MS run and detected by MS/MS (see experimental details in Chapter 5).

Table 3 Example of ambiguous peak assignments.

Experimental m/z	MH^+ of possible peptide	Mass error / ppm	POc	Peptide position	Sequence
590.2582 (eIF3i)	1179.5099	-0.54	0.027	99-109	WDVSNGQCVAT
	1179.5099	-0.54	0.0076	100-110	DVSNGQCVATW
813.3744 (eIF3bC3)	1625.7249	5.28	0.023	53-66	KIFDVQPEDASDDF
	1625.7249	5.28	0.014	52-65	FKIFDVQPEDASDD
1094.5107 (eIF3i)	1094.5153	-4.16	0.091	298-305	YHKIFEEE
	1094.5054	4.88	0.034	340-347	HHFEKSYF

On the other hand, for the examples in Table 3, each peak relates to two possible peptides that both survive in the previous two filtering steps. The peak at 590.26 can be assigned to two peptides in neighboring positions (eIF3i(99-109) or eIF3i(100-110)) with exactly the same mass. MS/MS detected the presence of eIF3i(99-109), which does have a higher POc value than the other. However, for the peak at 813.37, initially three peptides can be assigned: eIF3bC3(52-65), eIF3bC3(53-66) or eIF3bC3(63-76). The last peptide has a mass error of 3.87 ppm, but its POc value, 0.0003, is lower than the POc threshold, and is therefore excluded, which is indeed **not** detected by MS/MS, providing an example of the efficiency of the POc selection.

Figure 9 shows the MS/MS spectra of 813.37 assigned to the rest two neighboring sequences. The spectrum on top contains a rich series of b-ions (red) and some y-ions (blue) that firmly validates the presence of the peptide eIF3bC3(52-65), yet the peaks marked in green belong to the fragments of the other peptide eIF3bC3(53-66), indicating that these two peptides are both present and coelute in the long gradient conditions used for the MS/MS validation work.

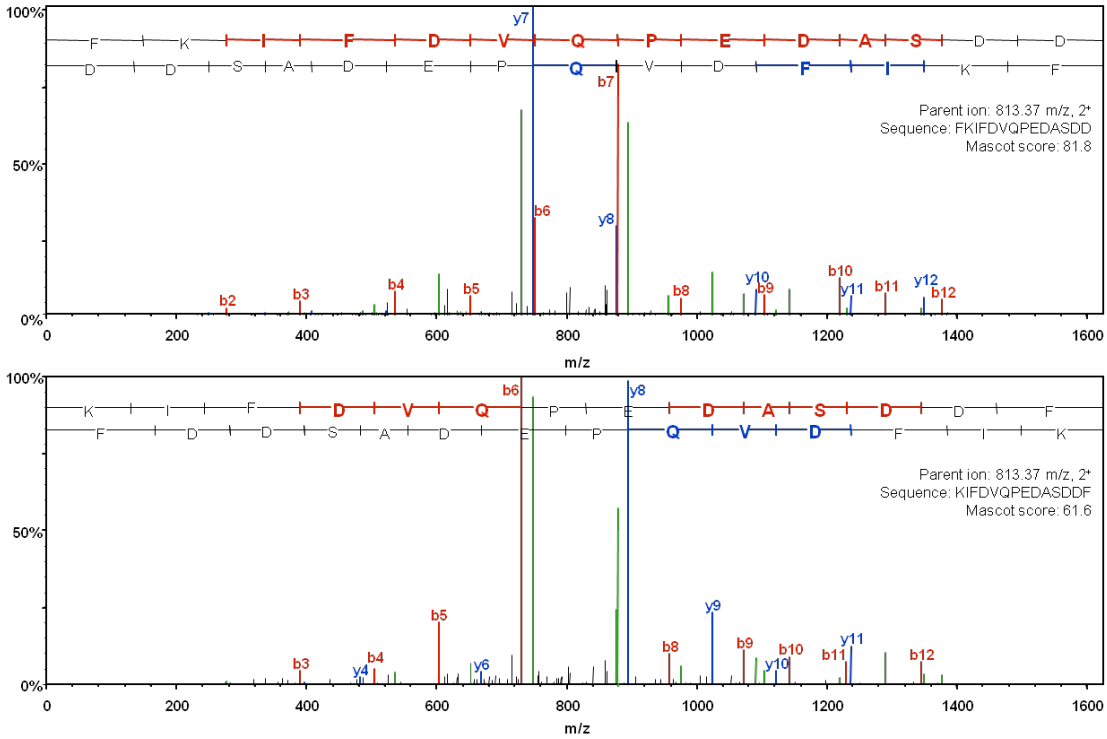


Figure 9 MS/MS spectra of parent ion 813.37 assigned to two neighboring peptides. Peaks labeled in red, blue and green correspond to b-ions, y-ions and ions from the other sequence, respectively.

Therefore, when it comes to the cases where one peak can be assigned to two (or more, although this was never observed) peptides with neighboring sequences of POc values both above the POc limit and of exactly the same mass, it can be true that both peptides are present in the digest, or at least one of them is present. In the latter situation, the peptide not present in the digest is indeed a false identification. Nevertheless, including such peptides in the reference list does not introduce significant error to the structure elucidation step; because the correct sequences are always in the reference list and the false sequences induce only a shift of one amino acid residues. On the contrary, if all such ambiguities are excluded, a decrease of sequence coverage or sequence overlaps cannot be avoided, and a loss of spatial

resolution becomes consequential. We thereby decided that in such cases, both sequences are retained in the reference list yet in a marked group, so that in later analysis appropriate caution will be taken in the data interpretation for these peptides and regions.

Table 3 also shows an example of another type of ambiguity, where one peak is assigned to two completely different sequences which both meet the criteria for the two filtering steps, for example the peak at 1094.51. MS/MS detected only the peptide of eIF3i(298-305) which has a POc value higher than the other. However in such cases it is harmful to hazard that the sequence with higher POc value should be the correct assignment, because the difference between the POc values can be rather insignificant. We decided to exclude the peaks with such ambiguities, in order not to introduce spatial errors for later analysis.

2.2.4 Reproducibility control

Due to the lack of definite specificity of pepsin, the peptic digestion is not strictly reproducible between different experimental replicates. However, given that the digestion condition remains the same throughout this work, a high reproducibility is still achieved. Therefore a reproducibility control is introduced as the last step to establish the reference list: the experiments for each reference protein are repeated five times, and a reference peptide is included in the list only when it is present in at least two replicates. The chemical noises which randomly occur can be further excluded by this filter; yet it does not discriminate against less probable peptides that may not be present in all replicates.

2.2.5 Method validation by MS/MS and evaluation

In order to validate the method described above, LC-MS/MS experiments were performed. It should be emphasized that the LC condition for MS/MS acquisition in this work is modified: to achieve maximum peptide detection a 2-hour gradient is applied instead of the fast (20 minutes) gradient in standard HDX conditions. After each MS scan, the top 20 peaks are selected for fragmentation. The results are then subjected to a Mascot database search with the following criterion: no specific enzyme, oxidation as variable modification, 10 ppm peptide tolerance, 0.5 D a fragment tolerance. The database search gives a large number of identified peptides,

many of which are repeated as the parent ions are fragmented more than once. After refining these replicates and applying a Mascot score limit of 20, a list of 385 MS/MS validated peptides is generated with an average Mascot ion score of 49 and standard deviation of 21. Compared to the average Mascot ion score of 26 reported in a recent work with MS/MS data of a similar-size protein acquired under HDX condition using the same mass spectrometer [7] as in this work, we obtained the MS/MS list with a higher score. This is most probably due to the extended LC gradient, which reduced the effect of critical data acquisition parameters, e.g. longer spectra acquisition time are possible and a larger number of peaks are fragmented.

Figure 10 shows the number of MS/MS validated peptides with different POc values. Most of the peptides identified in the MS/MS run are with a POc value above 0.001, which is coherent with the POc threshold derived from the analysis of the data shown in 2.2.2. This similar distribution of POc values comparing to Figure 7 is a primary proof of the validation of the POc value method as well as for the rationality of the POc filter. In fact, since the POc values are derived from a statistical analysis of HDX MS/MS results, it can be considered as a "theoretical" prediction of MS/MS, within the initial approximation assigning a similar value to probability of cleavage at a given position and probability of at least one cleavage at a given position. The distribution in Figure 9 therefore again confirms the accuracy of both Table 2 and the POc algorithm.

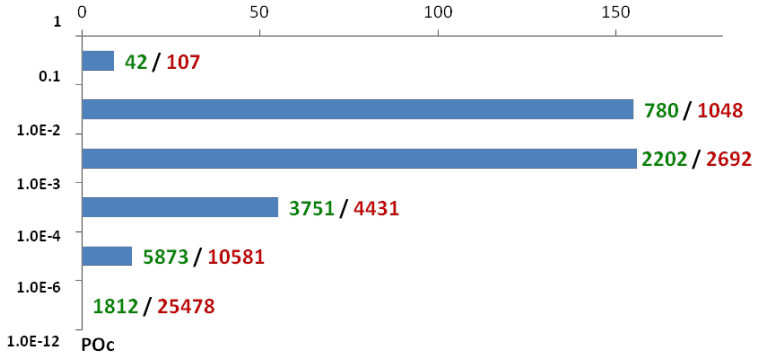


Figure 10 Population distribution of MS/MS validated peptides in eIF3i:bC3 in a given POc range. The blue bar represents the number of peptides detected within the POc range; the green number is the number of the theoretical peptides with 4-40 AAs in the POc range; the red number is the number of the theoretical peptides of any length in the POc range.

The peptides in this example retained after the filters are compared with the peptides validated by MS/MS. Among the 216 peptides from the example data set, 160 are

detected in the LC-MS/MS run. Figure 11 is the plot of the retained assignments from the example data set in comparison with the MS/MS identified peptides. It should be noted that one peptide sequence may corresponds to more than one assignment due to different charge state, and therefore there are 328 data points plotted in Figure 10.

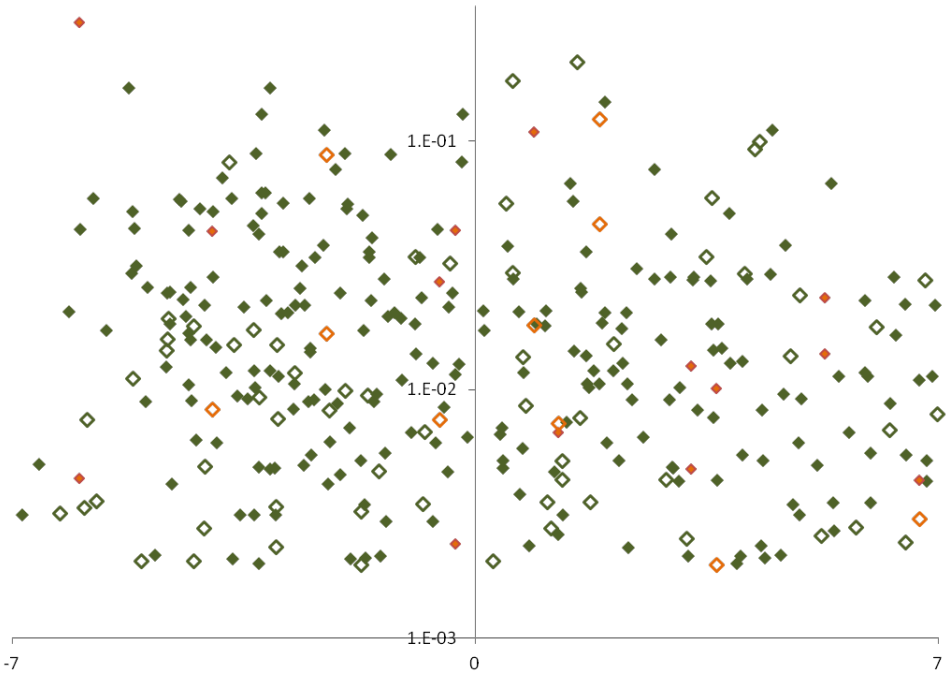


Figure 11 Plotting of peak assignments kept in the reference list with POc values against mass errors. Green square: unambiguous assignments validated by MS/MS. Empty green square: unambiguous assignment not detected by MS/MS. Orange square: ambiguous assignment with neighboring sequences validated by MS/MS. Empty orange square: ambiguous assignment with neighboring sequences not detected by MS/MS.

There are 12 peptides excluded either by the POc filter or the mass error filter which are detected by MS/MS, indicating that by excluding the false identifications with the filters, it is unavoidable that some real peptides are also excluded. These peptides mostly fall in the regions that are close to the POc or mass error thresholds. There are peptides detected by MS/MS, the peaks of which are not included in the example peak list; on the contrary, some assignments, as marked in the empty squares in Figure 10, are not detected by MS/MS. Both cases can be attributed to the experimental difference, e.g. digestion, fragmentation, detection. More interestingly, throughout this comparison, no f alse identification has been observed: when a MS peak is assigned to a peptide, it is either validated by MS/MS or not detected, but never proved by MS/MS to belong to another sequence excluded by the filters. This solidly

validates the method we developed, as it efficiently reduced the false identifications and resolved the ambiguous identifications.

The improved method to establish the reference peptide list is in principle based on two criteria: the mass accuracy and the calculated POc value. While the POc value is only dependent in the sequence of the peptide, the mass accuracy varies according to the protein size, the LC condition, the sample concentration, the instrumental and experimental parameters of the mass spectrometer, etc. In this work, the LC condition and the sample concentration remained constant; however, the upgrades of the MS instrument significantly increased its performance, as demonstrated before the upgrade from APEX III to APEX Q brought marginally higher detection sensitivity. The protein size also affect the mass error range, as illustrated in Figure 12.

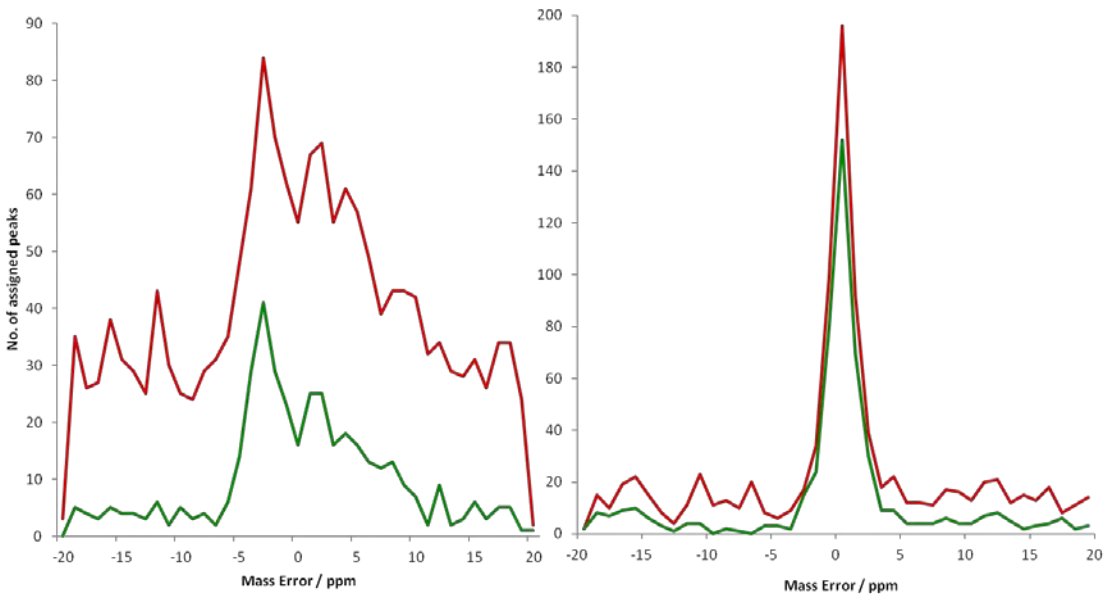


Figure 12 Error distributions of example data sets of eIF3i:bC3 (left) acquired with APEX Q and eIF3gC1ΔC (right) acquired with SolariX. The number of the assigned peaks within certain error range is plotted against mass error. Red: all assignments according to the real protein sequence. Blue: assignments that survived the POc selections.

The data set in Figure 12 (left) is the same example of eIF3i:bC3 (460 AAs) used to demonstrate the method, while the data set on the right is of eIF3gC1ΔC (149 AAs). Obviously, for eIF3gC1ΔC a narrow error distribution is present even in the initially assigned peaks without any selections (Figure 12, right, red curve). The baseline is efficiently lowered after the POc selection (Figure 12, right, green curve), and the mass error tolerance can be readily decided to be ±4 ppm. Whereas in the other example, a ±7 ppm mass error selection has to be applied.

Nevertheless, the data acquired with APEX Q and SolariX both meet the quality needed for the processing of the deuterated samples. Therefore in this work results acquired with both APEX Q and SolariX are equally used. Figure 13 shows the sequence coverage of the reference peptide lists for eIF3bC3, eIF3gC1ΔC and eIF3i. All lists have high sequence coverage and high spatial resolution due to rich sequence overlaps.



Figure 13 Sequence coverage of reference peptide lists of eIF3bC3 (top), eIF3gC1ΔC (middle) and eIF3i (bottom).

3 Method improvements in the processing of deuterated sample data

Generally, in this work the data processing other than reference peptide list establishment follows the previous workflow. A few modifications have been made on the HDX peak assignment and data representation to cope with the instrumental upgrade.

3.1 Mass recalibration

The FT-ICR mass spectrometer provides data with high mass accuracy and resolution. However, under the HDX condition, the mass accuracy is reduced due to the following reasons: the mass measurement with fast nanoLC gradient produces large number of coeluted peaks with various masses and charge states, and this leads to the space-charge effect often observed in FT-ICR MS that induces significant mass error [8-11], which is significant in measurements of high mass species with the highly sensible instrument. On the other hand, the harsh LC condition leads to a relatively poor chromatographic reproducibility, thus the retention time control is weak. Since the use of LC prohibits the internal calibration throughout the whole chromatograph, a recalibration step of the deuterated sample data is introduced in this work prior to the data assignment.

For each data set, four to six peaks with high intensity are selected. These peaks should be with a low deuteration level so that monoisotopic peak is present in all the samples at varying deuterium incubation times. They should also cover the entire mass range. The reference masses and the experimental masses are entered in a home-built program with the two calibration parameters (called ML1 and ML2 in the Bruker interface) used for the data acquisition. The recalibration parameters are generated by linear least squares fitting, and the whole peak list is then recalibrated accordingly.

3.2 Program upgrades for peak assignments of deuterated samples

Similar to the situation described in 2.1, the number of peaks in the deuterated samples also increased after the APEX Q replaced the APEX III. With samples of similar concentrations, typically 3000 – 5000 peaks were extracted from the APEX III LC-MS spectra, while from the APEX Q spectra, 30000 – 200000 peaks can be

extracted. The complexity of the spectra made it necessary to refine the criteria of peak assignment. Figure 14 shows the interface of the updated version of the HDX data assignment program used in this work.

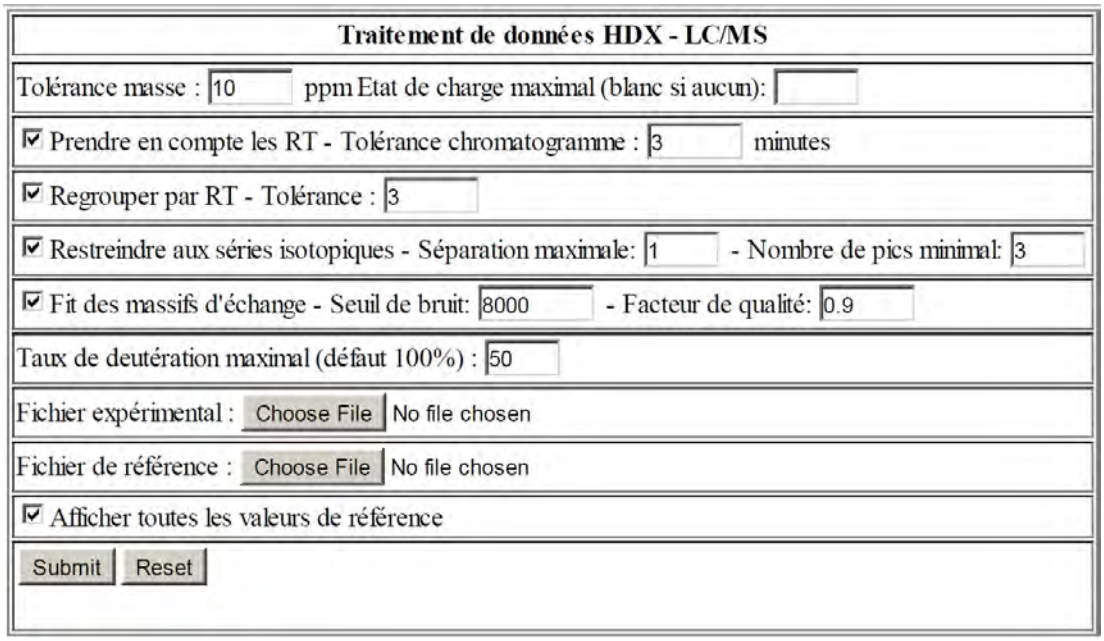


Figure 14 Updated parameter entry window of the home-built HDX data assignment program.

Instead of a high mass tolerance of 20 ppm which was applied previously, a 10 ppm tolerance is used as the peaks are recalibrated, yet this cannot be further reduced due to the variation of mass error caused by space-charge effect. A 3-minute window of retention time is used. Since the detection sensitivity has been enhanced, isotopic patterns of less intensive peaks are also better shaped. Therefore the restrictions are narrowed: missing peaks are no longer tolerated, a peak series is considered when at least three peaks are present, and the minimum Gaussian quality factor is fixed at 0.9 with an intensity threshold of 8000. In addition, a maximum deuteration level is set according to deuteration time, and this helps resolve the overlapping peaks which may belong to two different peptides.

Successfully assigned peaks leads to the deuteration levels of the peptides, which are then used to unveil the structural information of the proteins, and this will be discussed in the next chapter.

References

1. Duchateau, M. Détermination de surface d'interaction d'assemblages biologiques macromoléculaires par marquage au deutérium et analyse par spectrométrie de masse FT-ICR. Ecole Polytechnique, 2010.
2. Schechter, I.; Berger, A., On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967,** *27*, 157.
3. Schechter, I.; Berger, A., On the active site of proteases. III. Mapping the active site of papain; specific peptide inhibitors of papain. *Biochem. Biophys. Res. Commun.* **1968,** *32*, 898-902.
4. Keil, B., *Specificity of proteolysis*. Springer Verlag: 1992.
5. Senko, M. W.; Beu, S. C.; McLafferty, F. W., Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *J. Am. Soc. Mass. Spectrom.* **1995,** *6*, 229-233.
6. Hamuro, Y.; Coales, S. J.; Molnar, K. S.; Tuske, S. J.; Morrow, J. A., Specificity of immobilized porcine pepsin in H/D exchange compatible conditions. *Rapid Commun. Mass Spectrom.* **2008,** *22*, 1041-1046.
7. Burns, K. M.; Rey, M.; Baker, C. A.; Schriemer, D. C., Platform Dependencies in Bottom-up Hydrogen/Deuterium Exchange Mass Spectrometry. *Mol. Cell Proteomics* **2013,** *12*, 539-548.
8. BenDaniel, D.; Duke, C., Space-charge effects on electron tunneling. *Physical Review* **1966,** *152*, 683.
9. Easterling, M. L.; Mize, T. H.; Amster, I. J., Routine part-per-million mass accuracy for high-mass ions: space-charge effects in MALDI FT-ICR. *Anal. Chem.* **1999,** *71*, 624-632.
10. Jeffries, J.; Barlow, S.; Dunn, G., Theory of space-charge shift of ion cyclotron resonance frequencies. *Int. J. Mass Spectrom. Ion Processes* **1983,** *54*, 169-187.
11. Ledford Jr, E. B.; Rempel, D. L.; Gross, M., Space charge effects in Fourier transform mass spectrometry. II. Mass calibration. *Anal. Chem.* **1984,** *56*, 2744-2748.

Chapter 4

HDX analysis of the eIF3 complexes

1 Peptide selection, centroid mass verification and deuteration level calculation for deuterated samples 97
2 The eIF3i monomer and complexes of eIF3i:bC3 and eIF3i:gC1ΔC 102
2.1 HDX of eIF3i monomer 102
2.2 HDX of eIF3i in complex with eIF3bC3 106
2.3 HDX of eIF3i in complex with eIF3gC1ΔC 118
2.4 Discussion 127
References 129

As introduced in Chapter 2, the crystal structure of eIF3i in complex with the minimal CTD of eIF3b(654-700) was reported in 2012 [1], the second year of this PhD work. This significantly facilitates the structural study as deuteration information can be analyzed with the help of the atomic structure. In the complex, eIF3i shows a seven-bladed β-propeller structure with a C-terminal short, bent α-helix. One strand of the last blade 7 is formed by N-terminal residues 1-7, while the subsequent amino acid residues consecutively form blades 1–7.

The atomic structure of eIF3i alone is not reported, nor is the structure of the eIF3i:g binding surface. Yet there should not be huge structural difference between the free form of eIF3i and eIF3i in complex with eIF3b. In this part, the structure information by HDX-MS on the eIF3i subunit will be discussed in details upon binding with eIF3bC3 and eIF3gC1ΔC.

1 Peptide selection, centroid mass verification and deuteration level calculation for deuterated samples

The reference peptide list is established using the method described in Part 3. This list covers the whole eIF3i sequence except the first residue in the NTD and the five residues in the CTD. As shown in Figure 1 a large number of sequence overlaps are included in the list, ensuring a sub-peptide level of spatial resolution.

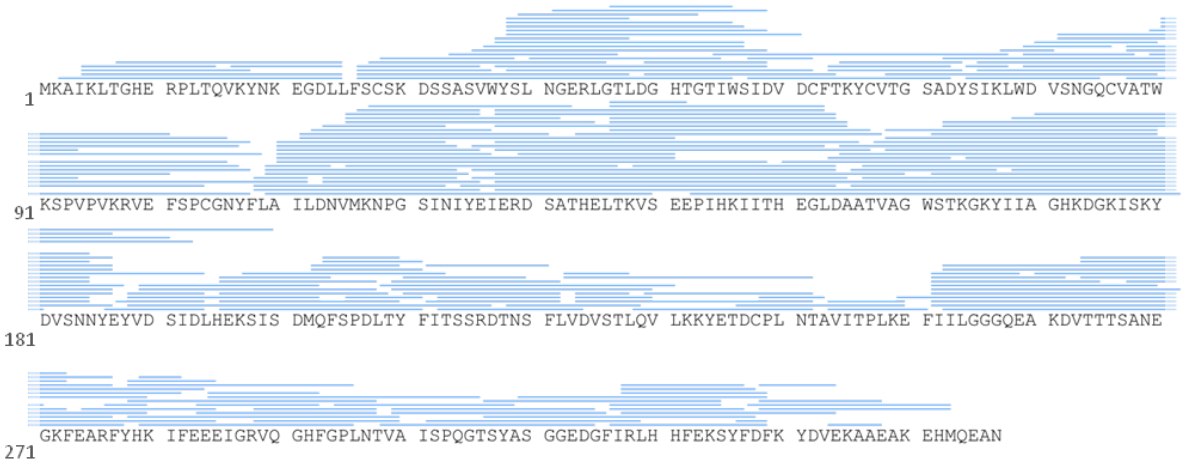


Figure 1 Sequence of eIF3i and sequence coverage of the reference peptide list.

After automated peak assignments and centroid mass calculation of the deuterated peptides according the reference peptide list, a limited number of representative

peptides are selected. These peptides are present in all time points and with relatively high signal intensity. In addition, these peptides are not coeluting with other peptides with similar masses which may interfere with the isotopic peak series. Generally, efficient peak assignment and accurate mass calculation are readily achieved for these peptides using the automated software. However, a manual check is necessary in the "abnormal" cases, i.e. when a significant difference in deuteration level occurs between peptides at similar regions or between replicates of same time points.

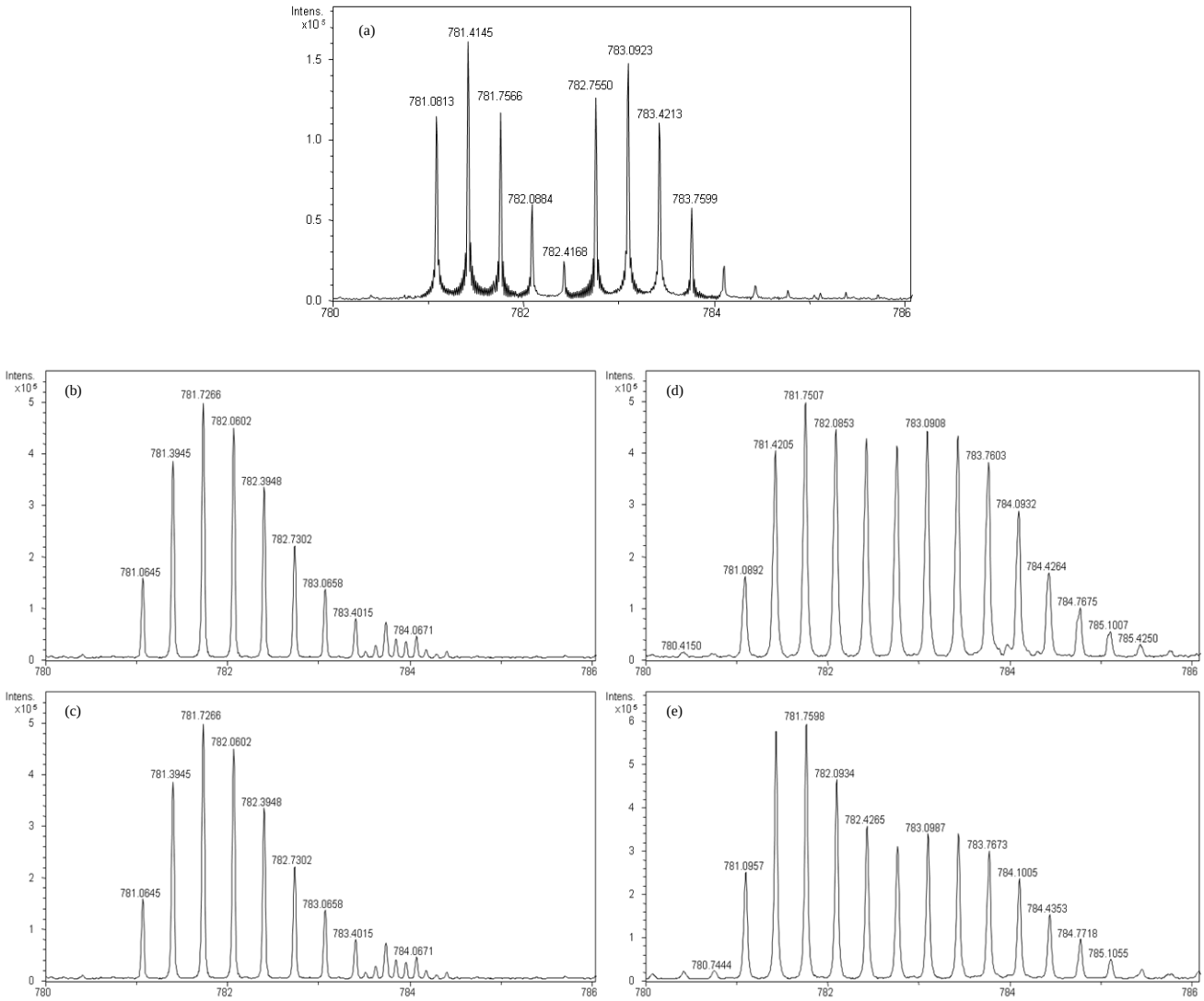


Figure 2 Mass spectra of peptide 89-108 in eIF3i (a) without deuteration and after (b) 5 s, (c) 1 min, (d) 15 min, (e) 30 min of deuteration.

Figure 2 illustrates a case for which automated analysis leads to wrong values. It is an example of spectra of a triply charged ion with 5 s (b), 1 min (c), 15 min (d) and 30 min (e) incubation in D_2O, as well as the reference spectrum of sample without deuteration (a). The original ion with monoisotopic peak at 781.1 (Figure 2 (a), first

peak series) is assigned to the peptide 89-108 of eIF3i, with MH^+ = 2341.1785. However, there is another adjacent triply charged peak series at 782.8 (Figure 2 (a), second peak series), corresponding to a coeluting peptide 263-282 of eiF3i, with MH^+ = 2346.1865. The two series are well separated in the spectrum without deuteration; yet when it comes to the spectra of deuterated samples, the peak patterns become quite complex. The overall mass shifts due to deuterium uptake are clearly observed. However, the isotopic peak series of the two peptides merge together very early in the kinetics. Initially (Figure 2 (b, c)), one could even assume that only the first peptide (monoisotopic peak at 781.1) is present. For later time points, at 15 min and 30 min of deuteration (Figure 2 (d, e)), two distributions seem to emerge and at first glance one might even consider it as a typical EX1 phenomenon. In the automated process, all the peaks are taken into the calculation of the centroid mass of peptide 89-108 as all of these peaks survive the selection process of the software, which leads to a completely invalid result. These two peaks could be separated by two Gaussian fittings. However, since there are other neighboring peptides present in the results, such ambiguous peak patterns are simply excluded from the data processing in order to accelerate the whole data analysis.

The use of FT-ICR MS and the reference list establishment method ensure peptides with a wide range in size to be monitored. Under common conditions, MS/MS is readily able to detect doubly and triply charged ions but vulnerable to detect peptide ions with higher charge states. Therefore, peptides with masses above 3000 Da are rarely present in the reference list established with MS/MS. In this work, MS/MS is not used to establish the peptide reference list, so that the size limitation increases up to 6000 – 7000 Da, which is no longer restricted by the instrument but by the nature of pepsin digest. The high resolution of FT-ICR MS enables it to precisely analyze ions with high charge state. Figure 3 shows the mass spectrum of a large peptide eIF3i(127-187) with MH^+ = 6756.9042 after 30 min deuteration (top) and a zoomed view of its 12^+ ion peak (bottom). The isotopic series are well separated even for the 12^+ ion, clearly demonstrating the high performance of the mass spectrometer.

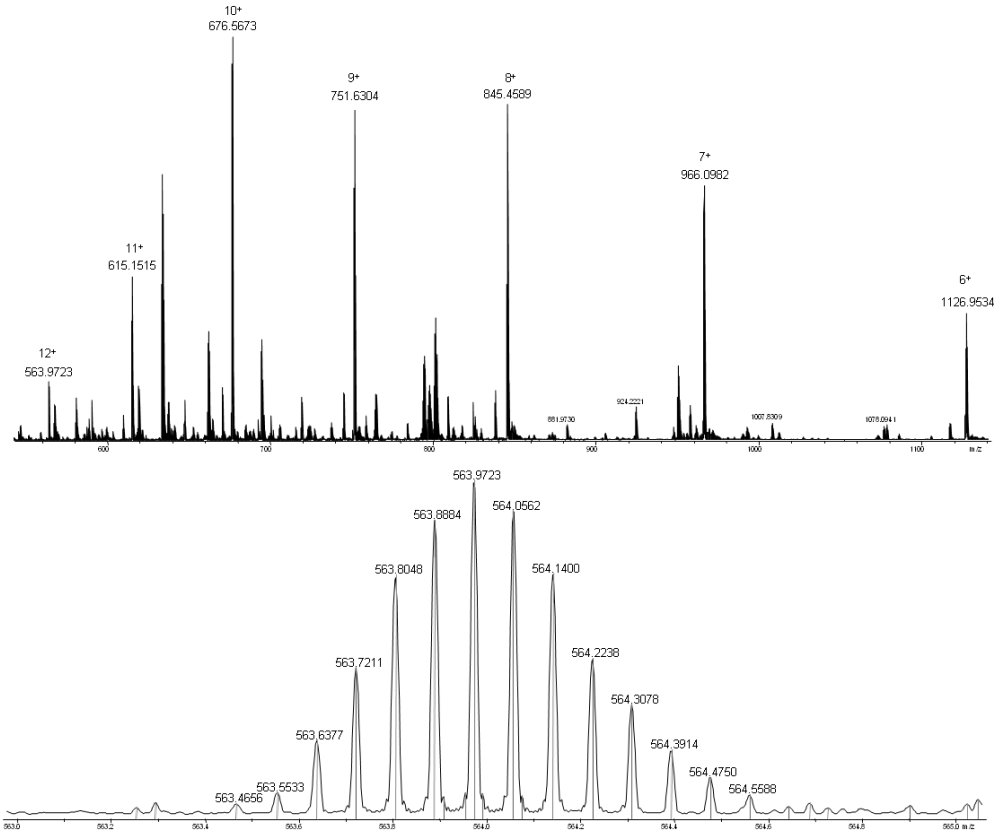


Figure 3 Mass spectrum of eIF3i(127-187) after 30 min deuteration with ions detected from 6^+ to 12^+(top), and zoomed view of its 12^+ ion peak series (bottom).

The deuteration level achieved in this work is, however, limited by the experimental conditions. As demonstrated in Part 1, back exchange of the deuterium to hydrogen can occur during the whole process after quench of the exchange reaction, and the extent of back exchange is affected by temperature and analysis time, as long as pH is kept stable. Due to the limitation of temperature-control devices and lack of automation, in this work the low temperature is achieved by on-ice handling of the pepsin digestion and ice-water bath of the nanoLC system. In addition, according to the gradient optimization result, the latest peaks are acquired 35 min after quenching, whereas in other works reported with advanced LC system the data acquisition can be finished within 10-15 min after the quench. [2-4] A relatively high level of back exchange is therefore unavoidable in this work.

Examples of high and low levels of deuteration are shown as mass spectra in Figure 4. Spectra (a) and (c) are acquired from reference samples without deuteration, and (b) and (d) are taken from samples after 30 min of deuteration. The deuteration level of eIF3i(243-261) (Figure 4 (a, c)) is calculated to be 45.9%, and that of eIF3i(87-108)

(Figure 4 (b, d)) is calculated to be 11.8%. Although various approaches have been introduced to correct the back exchange or normalize the deuteration level during data processing, here we use the deuteration level in the form of the absolute value in order to retain the comparability of data between replicates, different deuteration times and different protein complexes.

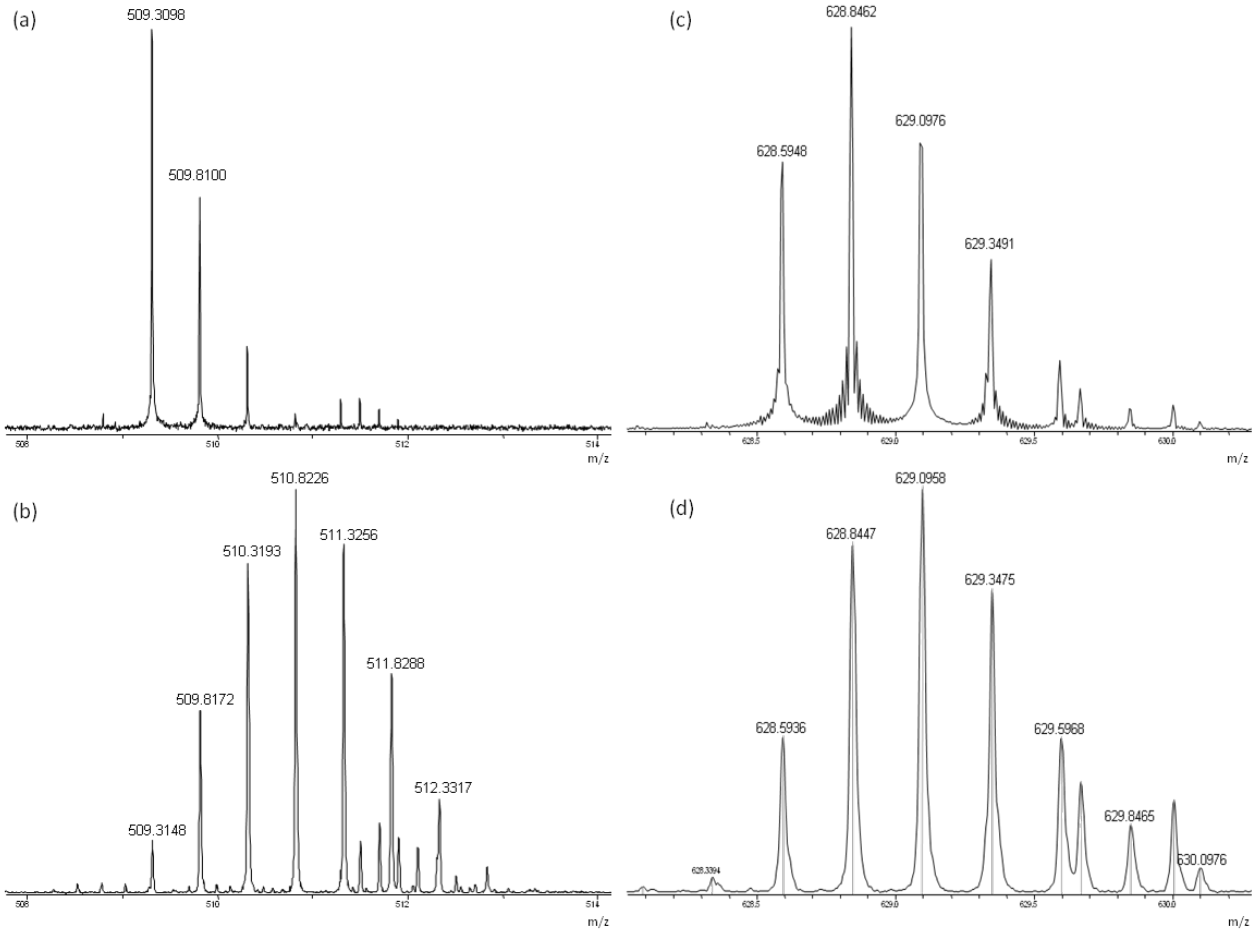


Figure 4 Mass spectra of peptides with high (left, eIF3i(243-261), MH^+ = 1017.5979, doubly charged) and low (right, eIF3i(87-108), MH^+ = 2511.2841, triply charged) levels of deuteration. The top spectra are from samples without deuteration and bottom spectra are from samples with 30 min of deuteration.

Due to the limit of time within this work, the selected peptides in data for 5 s, 1 min, 15 min and 30 min of deuteration were thoroughly verified as described above. These time points are significant as they provide information for short and long deuteration time. After the verification of the centroid masses of the selected peptides, the deuteration levels are calculated and analyzed together with the atomic structure.

2 The eIF3i monomer and complexes of eIF3i:bC3 and eIF3i:gC1ΔC

2.1 HDX of eIF3i monomer

The crystal structure of eIF3i monomer is still not reported, and the difficulty in crystallization indicates a possibly higher level of flexibility in the structure of this subunit. The deuteration level of eIF3i reflects this hypothesis. After calculation of the deuteration levels at various time points for the peptides in the digest of eIF3i, the HDX profiles are plotted on the protein sequence (Figure 5).

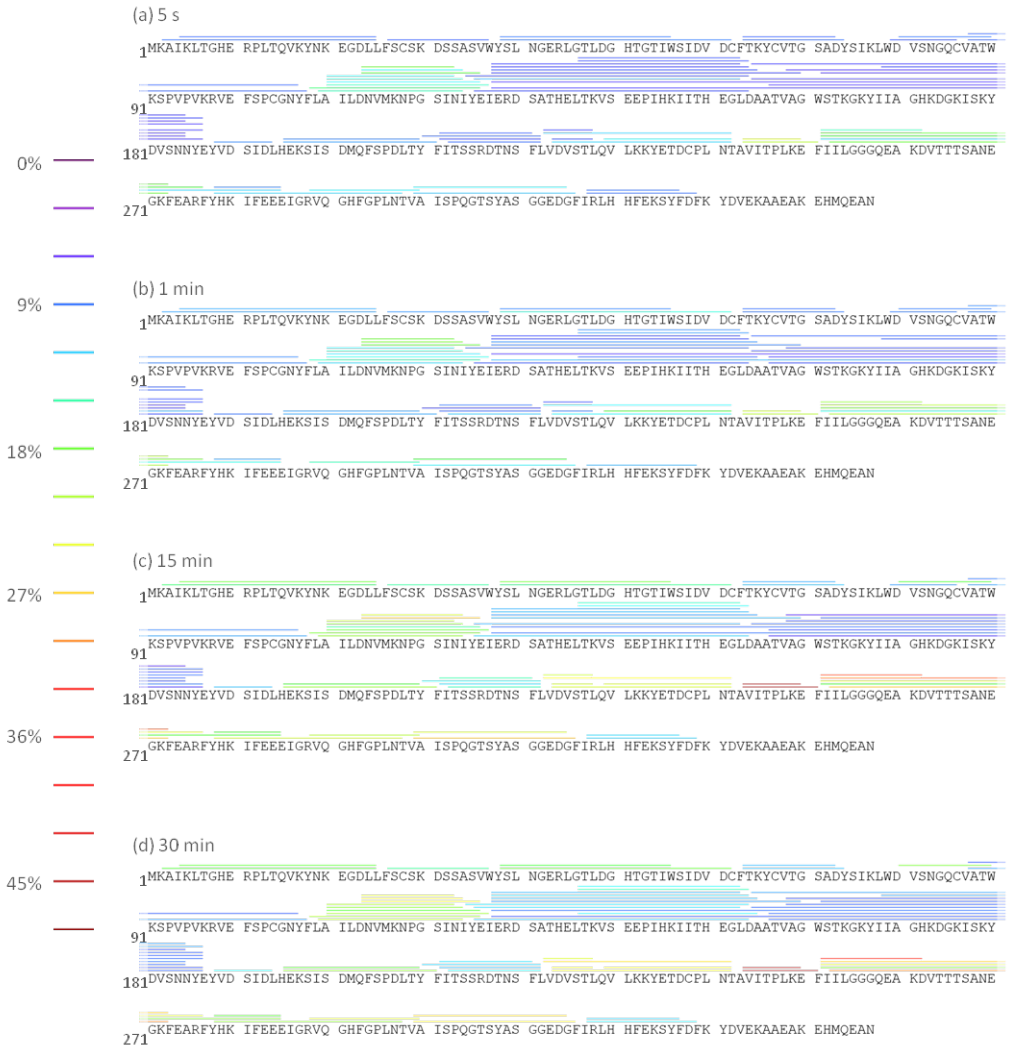


Figure 5 HDX profile of eIF3i monomer after (a) 5 s, (b) 1 min, (c) 15 min, (d) 30 min of deuteration. Peptides are color-coded according to the deuteration levels.

The HDX profile shows the deuteration level of each peptide monitored in this work after different time of deuteration. The population of these peptides is significantly

smaller to that of the initial reference list as shown in Figure 1, due to the fact that only peptides which are present in the majority of the time points are included. The spatial resolution is therefore compensated to some extent by higher accuracy and reproducibility. This profile provides a direct overview of the deuteration corresponding to the protein sequence, from which the first-hand information is generated.

After very short deuteration time (5 s as shown in Figure 5 (a)), the overall deuteration level remains below 20%. Higher deuteration levels are found in the regions of 108-125, 243-251 and 251-273, indicating higher dynamics and easier accessibility to the solvent. As the deuteration time increases, the peptides in the regions of 2-62, 108-125, 185-211, 222-242 and 262-298 gradually reach medium level of deuteration, i.e. after 30 min of deuteration these regions have 20-25% of deuteration. Regions of 63-74, 87-10, 126-188, 211-222 and 317-329 remain with low deuteration level below 15%, indicating well-protected and rigid structures which are not easily accessible by the solvent. The region of 243-273 consists of several peptides all showing high levels of deuteration over 40%, typical for loop regions and flexible dynamics and high solvent accessibility.

In order to correlate the deuteration information with the reported crystal structure, two approaches are carried out. The first attempt is to generate an average deuteration level for each amino acid residues. As such, if several peptides overlap in a certain region, the deuteration information carried by these peptides is integrated and a higher spatial resolution may be achieved. However, a loss of precise deuteration information may be induced at the averaging step. For example, after 30 min reaction, peptide 251-262 shows 32.8% of deuteration, whereas peptide 251-273 shows 27.0% of deuteration. If one chooses to average, the averaging of the two peptides gives a value of 29.9% for residues 251-262 and 27.0% for residues 263-273. This is indeed a problem since if we consider the shortest peptide, an accurate value of 32.8% should be considered for 251-262, and one could expand the process by requiring that for peptide 251-273, the total 27.0% should be an weighted average between 11 amide bonds at 32.8% (peptide 251-262) and 11 amide bonds at x percent (peptide 262-273). Application of this calculation leads to a value of 21.2% for the 262-273 region instead of 27%. This differential approach is technically correct but hard to completely automate: peptide overlaps are not always as simple as in the example

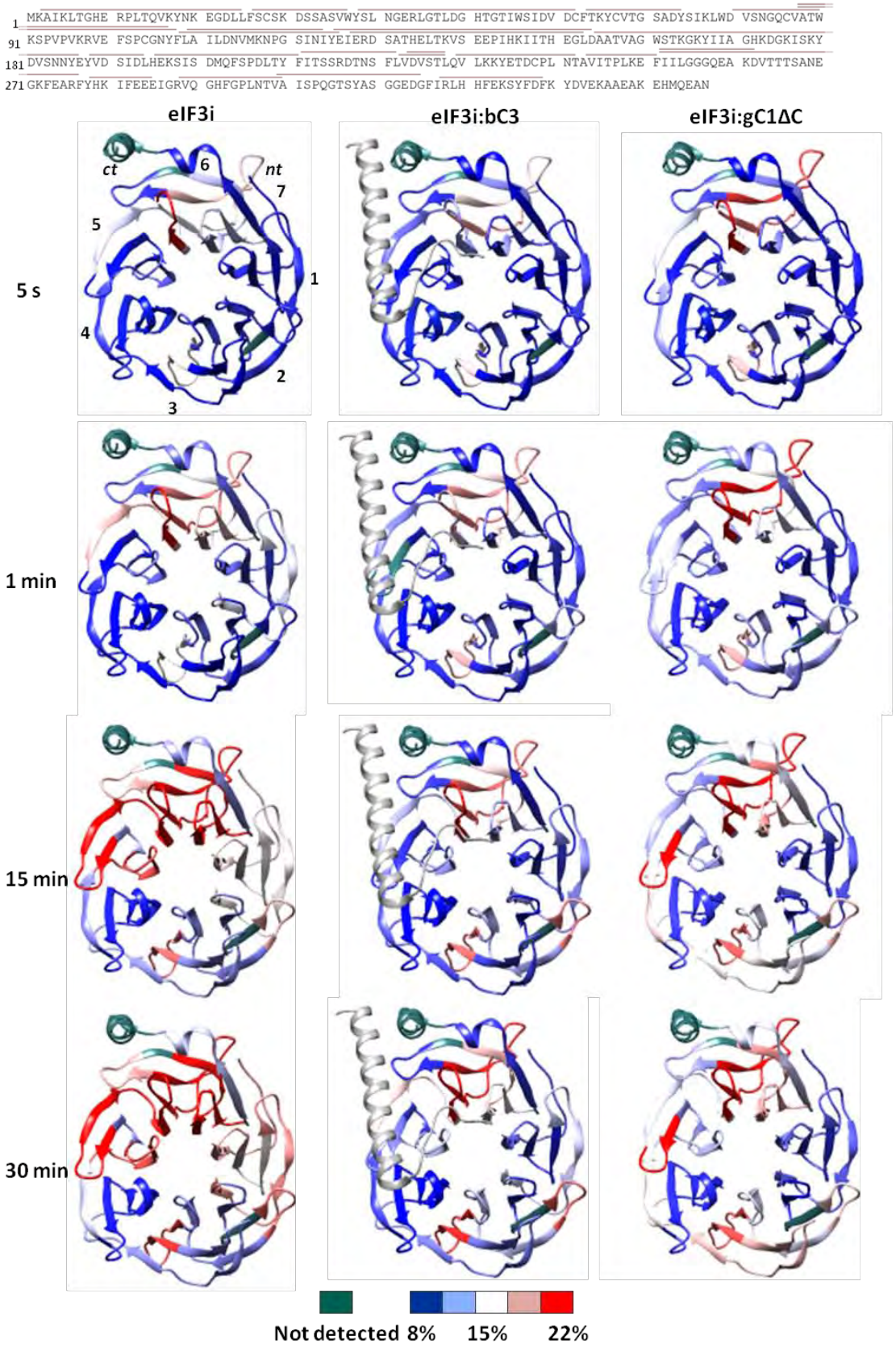


Figure 6 The HDX eIF3i in three forms after different times of deuteration color-coded on the crystal structure of eIF3i:b(654-700). The 7 blades of the β-propeller are marked and blue-white-red colors represent low-medium-high levels of deuteration, respectively. The selected peptide list is presented on the top.

taken here, and sometimes one would have to include 6-7 overlapping peptides, and the correct method to extract the final result with the inclusion of experimental errors is a difficult problem. [5-6]

Since the error induced by averaging the deuteration levels can be harmful, and extraction of overlapping values is difficult, this approach is not applied in this work. Instead, shortest peptides that cover each sequence region and not overlapping with each other are first used to plot on the structure model. This selected list of peptides are shown in Figure 6 (top). For the sequence gaps between these peptides, deuteration levels are taken from shortest peptides that cover these residues.

Figure 6 (left column) shows the time-dependent evolution of deuteration levels of the free from of eIF3i color-coded on the crystal structure model. It has to be noted that the reported structure is in complex with eIF3b(654-700), the stability of certain regions maybe slightly different from the structure of the eIF3i monomer. Indeed, the highly structured β-propeller is in agreement with the initial low deuteration levels in Figure 6 (a) and (b). Blade 6 already shows a higher deuteration level due to its rich content of loop structure. However, with the evolution of time, the 7 blades show different level of deuteration. Blades 1, 5, 6 and 7 are highly deuterated while blades 2, 3 and 4 remain in a low deuteration level. The highly deuterated residues 108-126 at blade 3 may result from the lack of spatial resolution as this value is derived from a single long peptide covering the loop region in addition to the residues in blade 3, which falsely raises the deuteration of the neighboring β-sheets. Despite this region, the HDX of eIF3i monomer leads to the hypothesis that the N- and C-terminal blades may have a higher level of flexibility and undergo rapid open-close motions, while the inner blades are more rigid and well protected from the solvent access.

2.2 HDX of eIF3i in complex with eIF3bC3

When eIF3i interacts with eIF3bC3 in the complex, the HDX in the eIF3i domain changes significantly. An overall decrease of deuterium incorporation is observed in each time point, as the binding of eiF3bC3 inhibits the interaction of eIF3i with the solvent water.

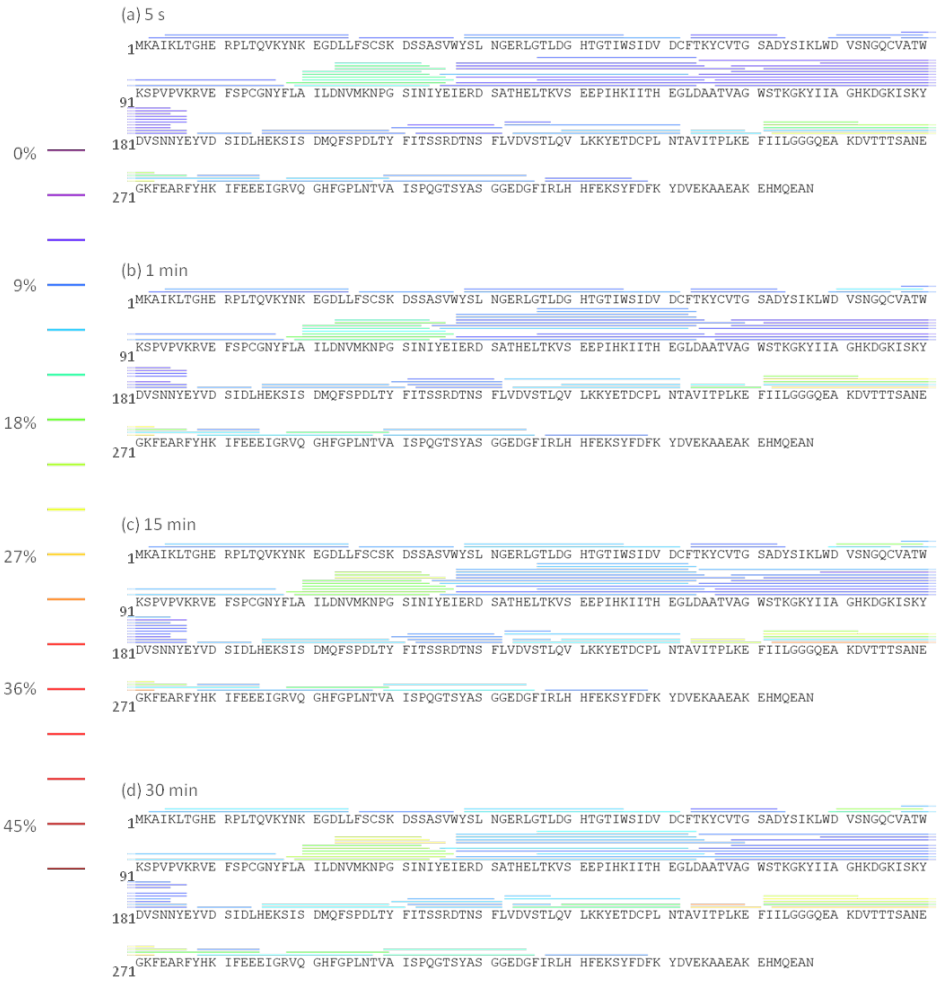


Figure 7 HDX profile of eIF3i in the complex of eIF3i:bC3 after (a) 5 s, (b) 1 min, (c) 15 min, (d) 30 min of deuteration. Peptides are color-coded according to the deuteration levels.

The HDX profile of eIF3i in the complex of eIF3i:bC3 (Figure 7) shows decreases in deuteration levels at various regions. The regions of 63-74, 87-107, 126-188, 211-222 and 317-329 remain low deuteration levels throughout the whole time course, which are similar to the behaviors of these regions in eIF3 monomer. This similarity indicates that these regions are not on the binding surface of eIF3i and eIF3bC3.

Figure 6 (middle column) illustrates the time-dependent evolution of deuteration level for eIF3i in complex with eIF3bC3 color-coded on the crystal structure model of eIF3i:b(654-700). Blades 1, 5, 6 and 7 show significantly lower deuteration level in the complex indicating that they interact intensively with eIF3b. In addition, the binding of eIF3bC3 may stabilize the whole β-propeller structure of eIF3i, as the overall deuteration level decreases in comparison to the eFI3i monomer.

In order to more precisely analyze the binding regions, the difference of the deuteration levels for each peptide at the same time point are plotted in Figure 8 for data after 5 s and 30 min deuteration. Such plotting provides a clear overview of the HDX behavior upon the complex formation. Notably, the data points reflect the change in the peptide level, rather than the single residue level. In order to avoid misleading errors introduced in the experimental and data processing steps, a threshold of ±3% is applied so that only difference larger than this is considered structurally significant. The higher values in the plot indicate more significant protection in the structure upon the binding with eIF3bC3. According to the nature of deuterium exchange on amide hydrogens, the protection can be attributed to the direct formation of hydrogen bonds between eIF3i and eIF3bC3, to the inhibition restraining access of water molecules towards the amide hydrogen on eIF3i by nearby residues of eIF3bC3, or to a higher ordered and less flexible conformation, induced by the complex formation. Overlapping peptides with slight residue shifts show similar changes and can cross-validate each other, e.g. peptides 222-228 and 223-228 decrease by 12.1% and 13.2% in deuteration levels after 30 min upon eIF3bC3 binding, respectively. On the other hand, significant difference between peptides which share part of the sequence can be informative: for example, peptide 251-262 shows a decrease of 10.5% in deuteration level after 30 min, whereas peptide 251-273 only shows a decrease of 2.5%, and this indicates that the regions 262-273 is not directly protected by the binding, although this region is not separately observed as a single peptide.

Representative peptides are selected and analyzed with more precise time points and two replicates at each time point: deuteration levels after 5 s, 30s, 1 min, 2 min, 4 min, 8 min, 15 min and 30 min reaction are studied. Kinetic curves are thereby generated and compared with the structure model reported as eIF3i:b(654–700) [1]. It has to be stressed that the eIF3bC3 fragment used in our work contains residues 640-726, a

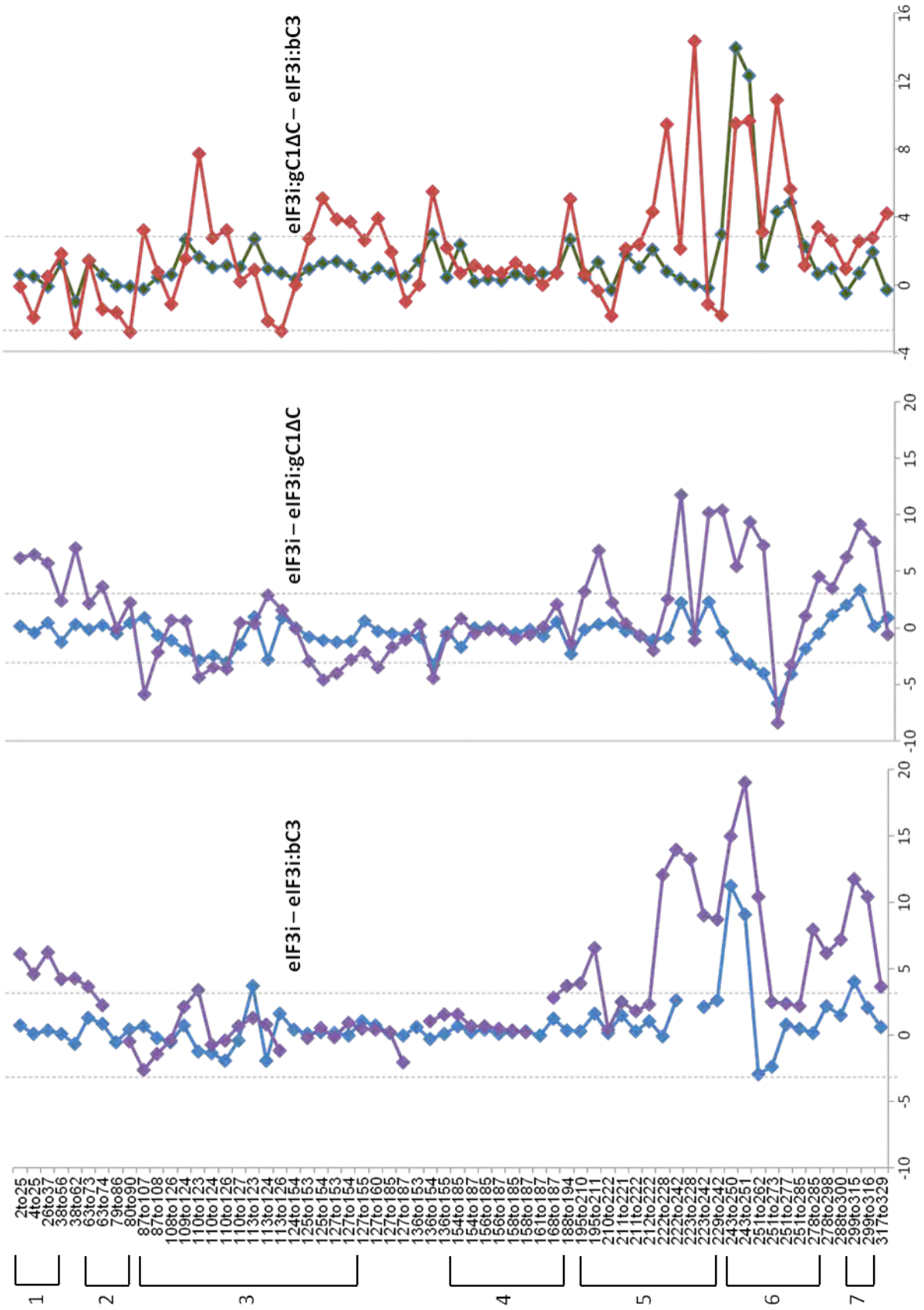


Figure 8 Differences of the deuteration level of peptides in eIF3i of different forms after 5 s (blue or green) and 30 min (purple or red) of deuteration.

more extended fragment than in the crystallized structure. It consists of the long α-helix of residues 663-689 which is shown to bind to the bottom side of the b-propeller along the loops of blades 5 and 6 of eIF3i, the unstructured chain of 690-699 shown in the crystal structure that advances towards the central cavity of the eIF3i b-propeller, as well as the N-terminal residues 640-653 and the C-terminal residues of 700-726 which have not been crystallized yet.

The peptide eIF3i(243-251) (Figure 9) displays significant change in deuteration levels upon the formation of the complex. In the eIF3i monomer, this peptide rapidly reaches high levels of deuteration within the first few minutes. A relatively large deviation between replicates indicates a highly flexible structure in this region easily accessible by the solvent (Figure 9, bottom-right, blue and green). In the eIF3i:bC3 complex, much lower deuteration levels with less deviation are observed (Figure 9, bottom-right, red and purple). These behaviours agree with the structure shown in Figure 9, bottom-left. The peptide 243-251 consists of several residues on a β-sheet and a loop in blade 6. This loop closely interacts with the C-terminal loop structure of eIF3b, via main chain atoms contact between W248 from eIF3i and F694 and D695 from eIF3b (Figure 9, bottom-middle). This interaction clearly stabilizes the flexible loop in eIF3i, and limits the contact of its amide hydrogens with the solvent water molecules.

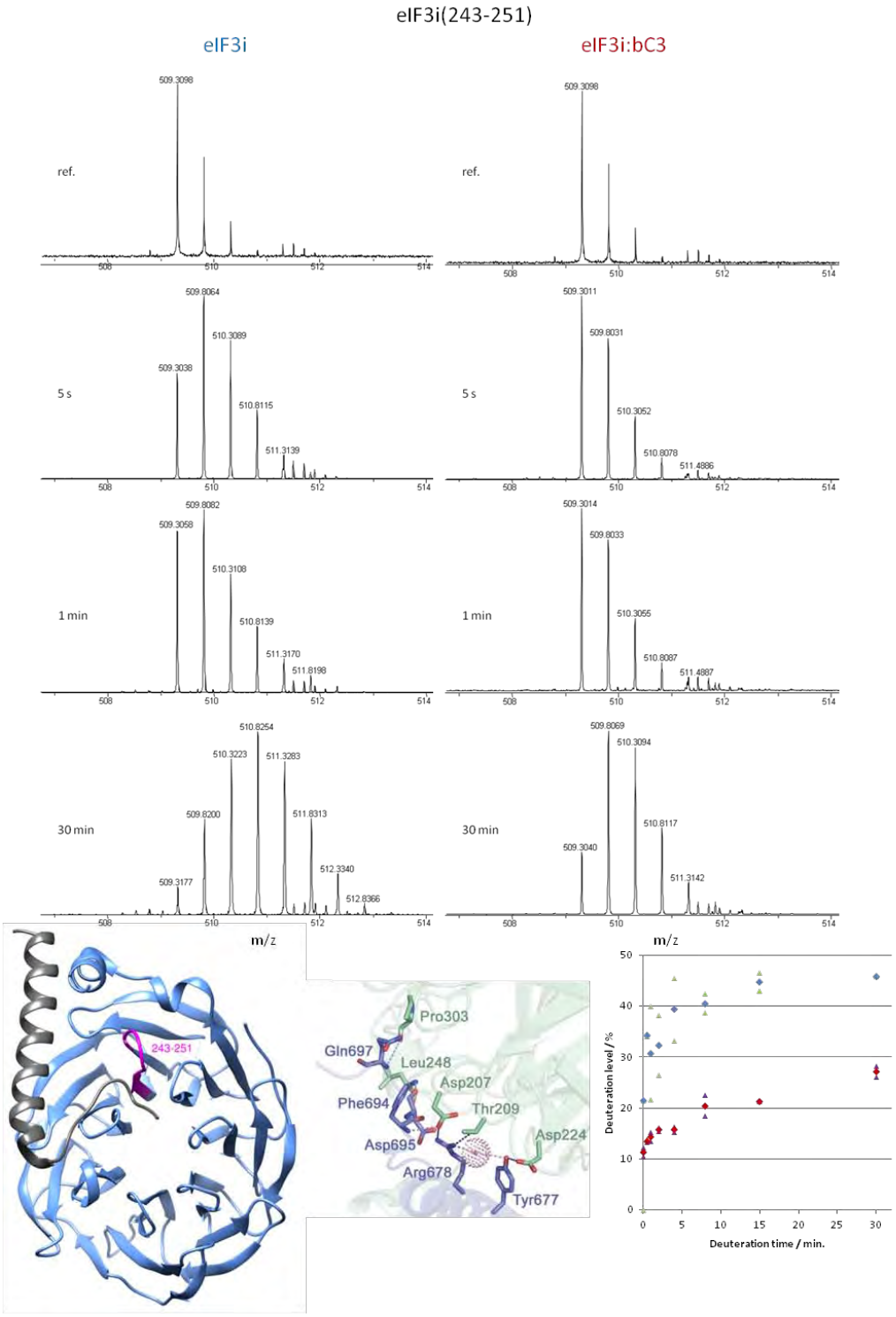


Figure 9 HDX of eIF3i(243-251). Top: mass spectra of the isotopic envelop of doubly charged ion from eIF3i monomer (left) and eIF3i in the complex (right) before deuteration (ref.) and after different times of deuteration. Bottom-left: highlight of the position of this fragment on the crystal structure. Bottom-middle: interacting regions with molecular details(eIF3i: green; eIF3b: blue) [1]. Bottom-right: Plot of deuteration levels as the average (square) of two replicates (triangle) against the deuteration time. Blue and green marks represent eIF3i monomer while red and purple marks represent eIF3i in the complex.

A similar agreement between HDX data and the reported structure is also found for peptides eIF3i(195-210) and eIF3i(222-228), which cover regions that directly interact with eIF3b. As shown in Figure 10, both peptides show medium levels of deuteration around 20% after 30 min reaction in the eIF3i monomer (top- and middle-left, blue and green). This may be due to the fact that peptide 195-210 locates near the central cavity of the β-propeller and 222-228 consists of rich β-sheet structure, and therefore they are both under protection from direct solvent access. Nevertheless, significant decreases of deuteration levels occur upon the binding of eIF3bC3 (top- and middle-left, red and purple). D207, T209 and D224 from eIF3i form hydrogen bonds with Y677 and R678 from eIB3b via side chains (Figure 10, bottom-left) [1]. In addition, the aromatic ring of eIF3b W674 inserts into a pocket of eIF3i formed by hydrophobic amino acids (L231, I281 and the conserved L222) and the side chains of polar amino acids (Y210 and K280 as well as the conserved K232 and E250) (Figure 10, bottom-right) [1].

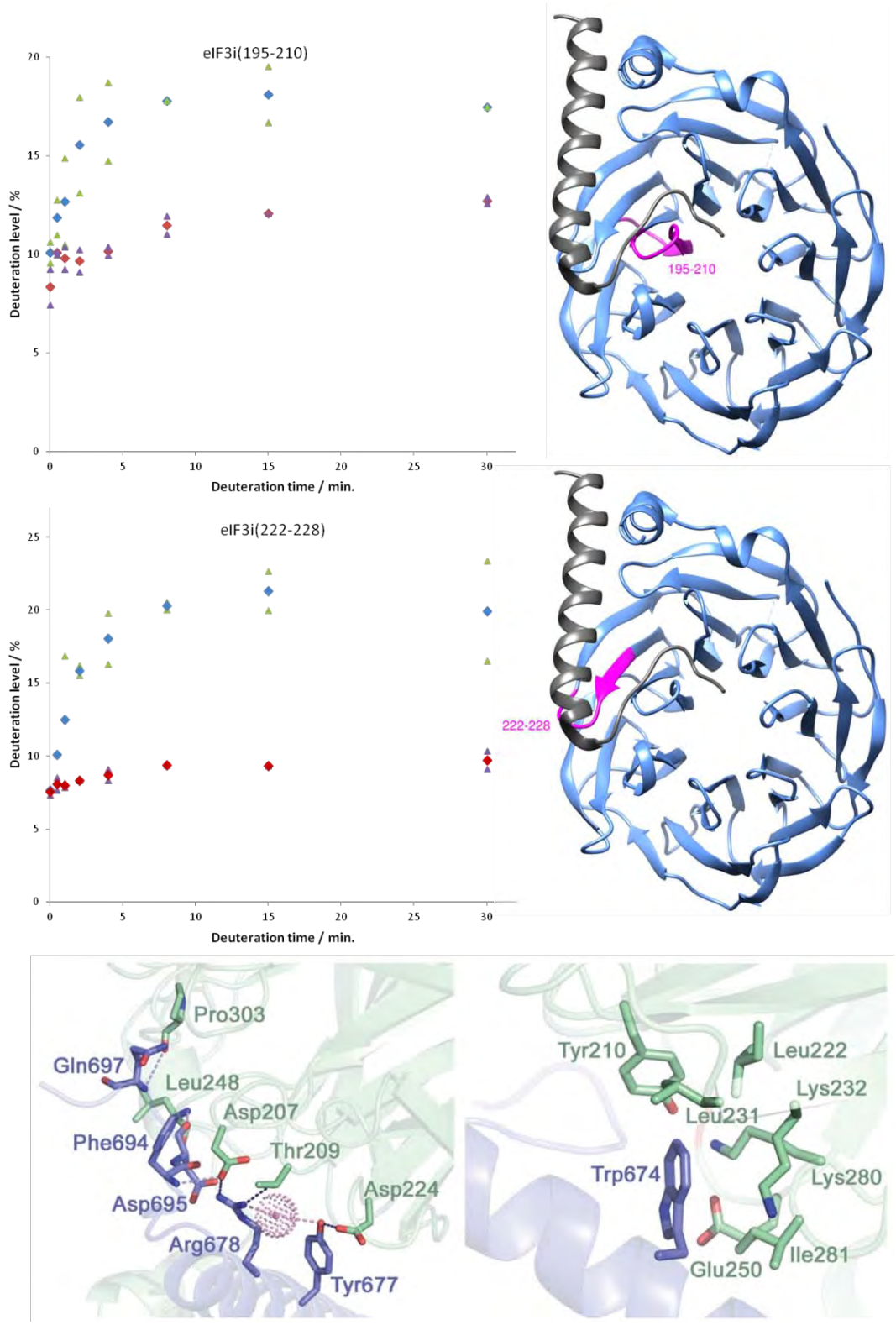


Figure 10 HDX of eIF3i(195-210) and eIF3i(222-228) in comparison with structure model. Top- and middle-left: Plots of deuteration levels with the average (square) of two replicates (triangle) against the deuteration time. Blue and green marks represent eIF3i monomer while red and purple marks represent eIF3i in the complex. Top- and middle-right: highlight of the position of the fragments on the crystal structure. Bottom: interacting regions with molecular details(eIF3i: green; eIF3b: blue) [1].

The HDX of another peptide, the N-terminal 4-25 in eIF3i, is shown in Figure 11. It is reported that one β-sheet of the last blade 7 is formed by N-terminal residues 1-7, which contains the N-terminus of peptide 4-25, while the subsequent residues consecutively form blades 1-7 [1]. The rich content of β-sheet structure explains the medium levels of deuteration of this peptide in the eIF3i monomer. The C-terminus of this peptide located in the central cavity of the β-propeller is likely to interact with the eIF3b C-terminus, which extends further towards the loop region in eIF3i (4-25). Since this region of eIF3b(701-724) is not included in the reported crystal structure of eIF3i:b(654-700), in the structure model (Figure 11 bottom-left) no direct interaction can be observed. This hypothesis might explain the significant decrease of deuteration levels upon the complex formation, and it complements the interaction information that was not achieved by previous reports. Peptide eIF3i(26-37) and eIF3i(38-56) also exhibit such significant decreases of deuteration (Figure 8), indicating that the long C-terminus may extend further towards blade 1 in eIF3i.

Similar HDX behaviour is also observed in the peptide eIF3i(288-300) (Figure 12). The relatively high levels of deuteration and deviations between replicates agree with a long, flexible loop structure in the middle of the peptide. When the complex forms, the deuteration levels decrease. The C-terminal β-sheet structure locates close to the central cavity of eIF3i, and is therefore likely to interact with the C-terminus of eIF3b. Another hypothesis is that the N-terminus of this peptides interacts with the unreported eIF3b(640-653) region, indicating the possibility that the N-terminus of eIF3bC3 bends towards the outer surface of blade 6 and 7 from eIF3i.

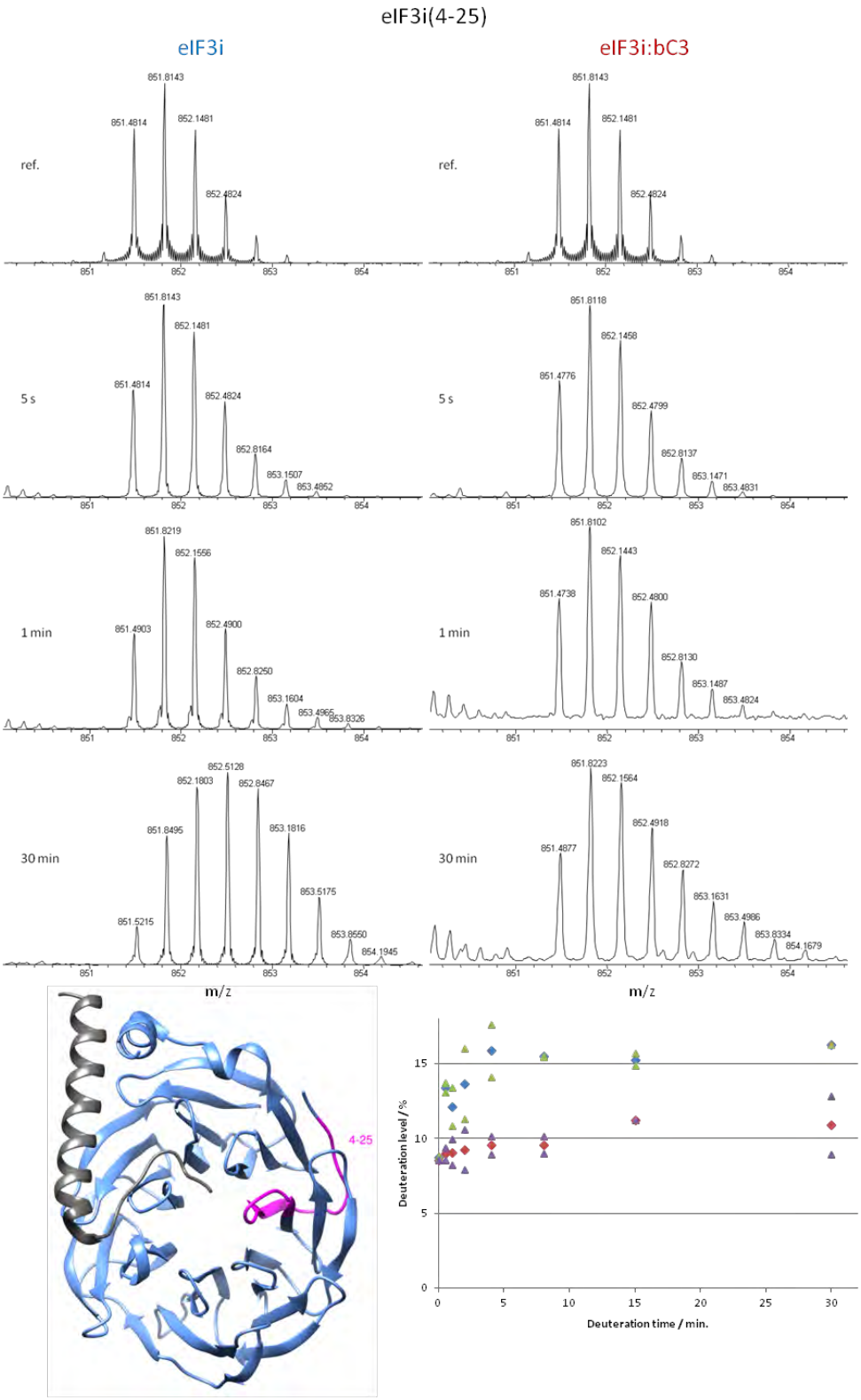


Figure 11 HDX of eIF3i(4-25). Top: mass spectra of the isotopic envelop of doubly charged ion from eIF3i monomer (left) and eIF3i in the complex (right) before deuteration (ref.) and after different times of deuteration. Bottom-left: highlight of the position of this fragment on the crystal structure. Bottom-middle: interacting regions with molecular details(eIF3i: green; eIF3b: blue) [1]. Bottom-right: Plot of deuteration levels as the average (square) of two replicates (triangle) against the deuteration time. Blue and green marks represent eIF3i monomer while red and purple marks represent eIF3i in the complex.

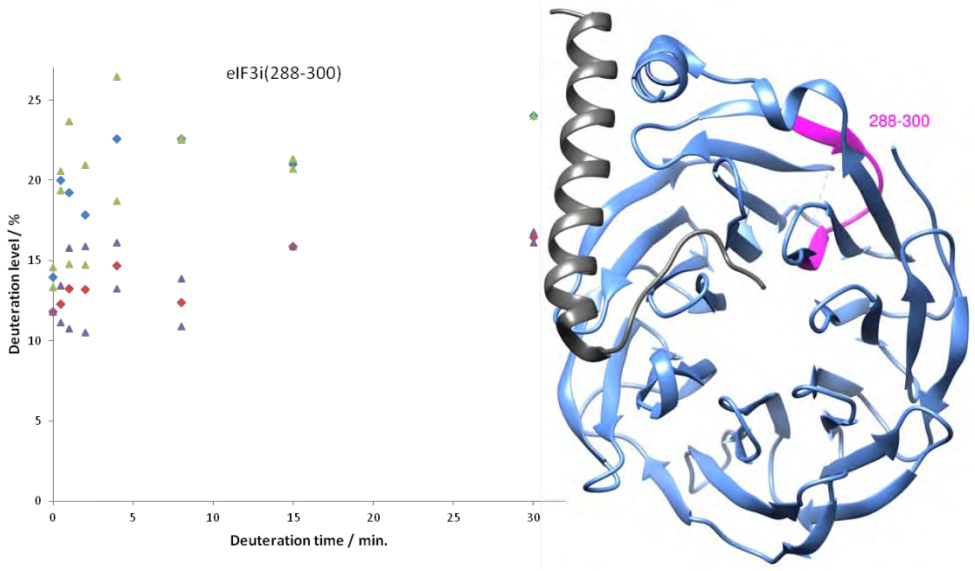


Figure 12 HDX of eIF3i(288-300). Left: Plot of deuteration levels as the average (square) of two replicates (triangle) against the deuteration time. Blue and green marks represent eIF3i monomer while red and purple marks represent eIF3i in the complex. Right: highlight of the position of this fragment on the crystal structure.

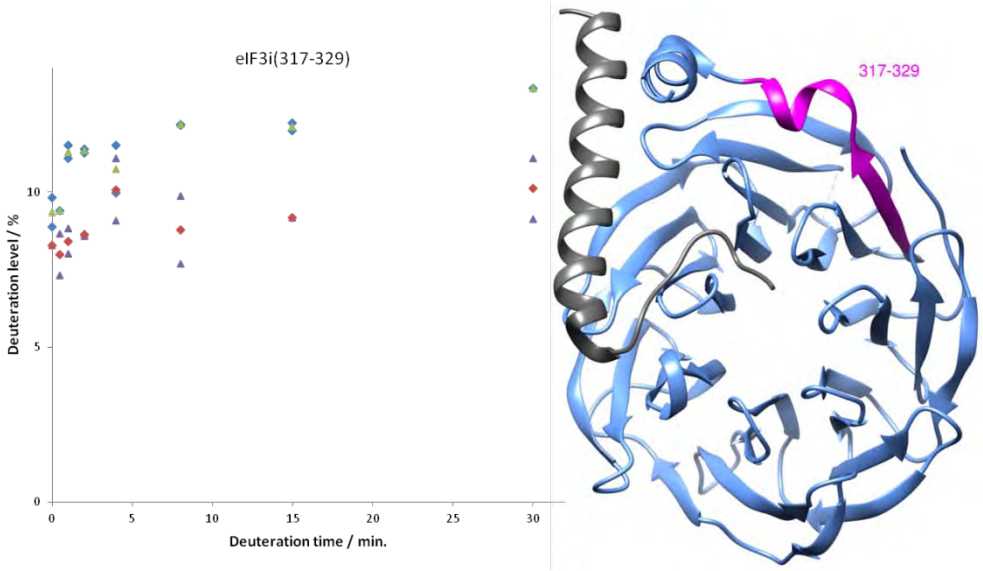


Figure 13 HDX of eIF3i(317-329). Left: Plot of deuteration levels as the average (square) of two replicates (triangle) against the deuteration time. Blue and green marks represent eIF3i monomer while red and purple marks represent eIF3i in the complex. Right: highlight of the position of the peptide on the crystal structure.

The second hypothesis finds its co-evidence from eIF3i(317-329) (Figure 13). In eIF3 monomer, eIF3i(317-329) shows low levels of deuteration as it consists of a β-sheet and a short α-helix. Decrease in deuteration levels is observed after complex formation. This leads to the similar hypothesis that it is protected via interaction with

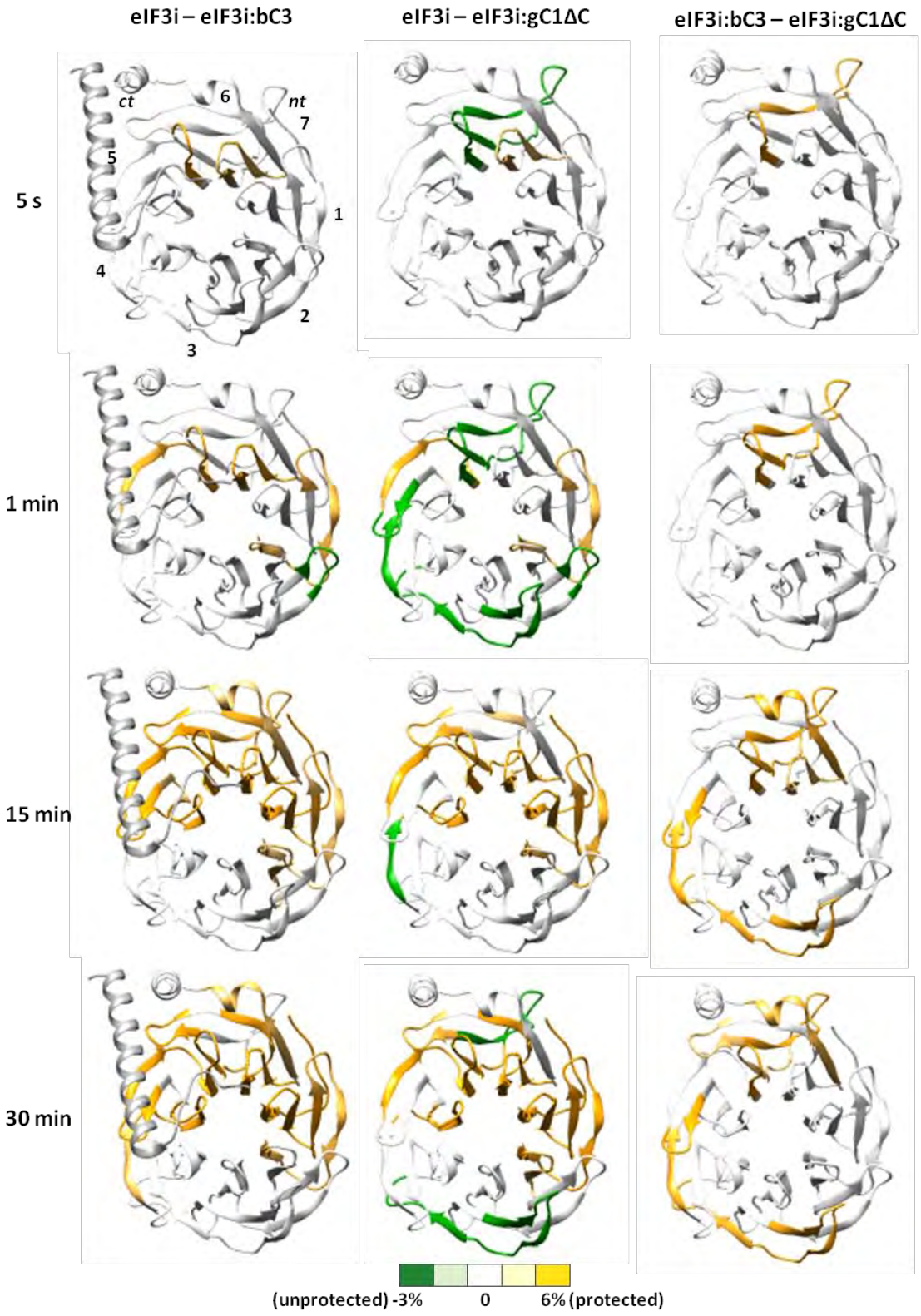


Figure 14 The change of deuteration level after different times of deuteration for eIF3i in three forms, color-coded on the crystal structure of eIF3i:b(654-700). Regions in yellow show increase in deuteration level, in white show no c hanges and green show decrease in deuteration level.

the N-terminus of eIF3bC3, as it is distant from the central cavity of eIF3i which may interact with the C-terminus of eIF3b. However, it is also possible that the α-helix in the C-terminus of eIF3i is less structured in the monomer, and the complex formation stabilizes this structure, protecting the distant residues in eIF3i(317-329) so that the deuteration level decreases.

A summarized overview of the interacting regions is illustrated in Figure 14, where the change of deuteration level for each peptide before and after the complex formation is color-coded on the structure. Detailed interaction information that have been discussed above can be reviewed from this figure: regions near the central cavity of eIF3i interact with the C-terminal of eIF3bC3, the loops on blades 5 and 6 interact with the long α-helix in the middle of eIF3bC3, and the outer β-sheets on blades 6 and 7 may involve in the interaction with the eIF3bC3 N-terminus.

2.3 HDX of eIF3i in complex with eIF3gC1ΔC

Due to the limits of time, in this work only one data point of eIF3i:gC1ΔC(42-170) is analyzed at each time point. The lack of replicates leads to the loss of deviation estimation for the deuteration level, and therefore a high threshold needs to be applied when considering the significance of the deuteration change upon complex formation. Nevertheless, since the experiments on eIF3i:gC1ΔC were taken continuously within one day, the instrumental condition remained highly stable, lowering the errors that may be induced during the experiments.

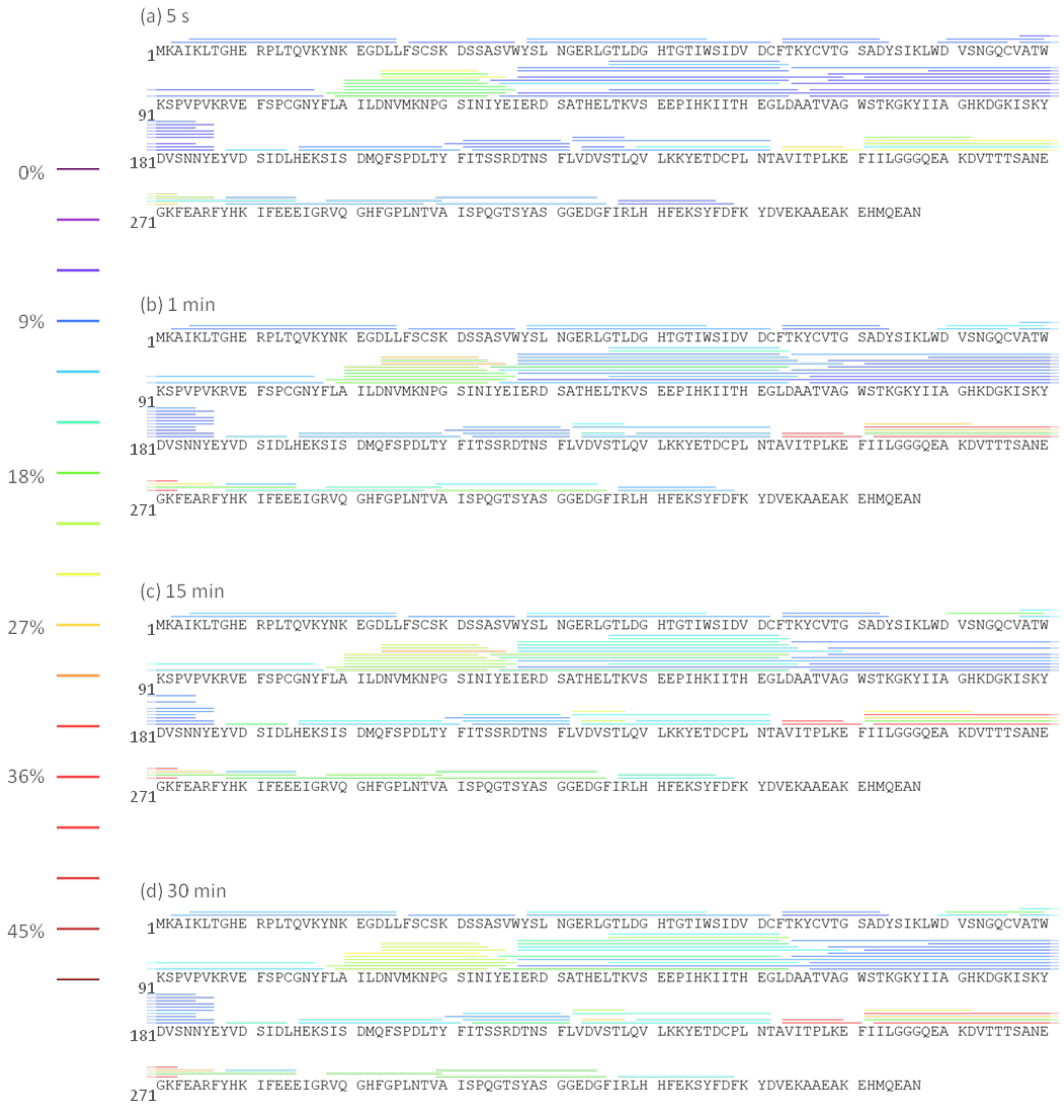


Figure 15 HDX profile of eIF3i in the complex of eIF3i:gC1ΔC after (a) 5 s, (b) 1 min, (c) 15 min, (d) 30 min of deuteration. Peptides are color-coded according to the deuteration levels.

The same list of selected peptides as for eIF3i monomer and the eIF3i:bC3 complex was thoroughly verified. The deuteration level for each peptide is color-coded on the

protein sequence (Figure 15). Along the time course, the N-terminus of eIF3i, eIF3i(2-74) remains less deuterated, as well as the regions of 154-187 and 210-222. Regions of 80-107, 125-153, 188-209, 230-242 and 317-329 show medium-low deuteration levels, while regions of 108-124, 222-229 and 243-276 show high deuteration levels. Parts of these regions are similar to the changes observed on eIF3i in complex with eIF3bC3. This might indicate that the binding surface of eIF3b and eIF3g on eIF3i are close in space to each other, or that binding leads to similar changes in eIF3i dynamics.

Similarly to the data processing of eIF3i:bC3, the deuteration levels of selected peptides in eIF3i from the complex of eIF3i:gC1ΔC are plotted on the structure model of eIF3i (Figure 6, right column). The inner parts of blades 2, 4 and 5 remain less deuterated even after long deuteration times, and this is coherent with the behaviors in eIF3i monomer and eIF3i:bC3. The rest of the structure shows different HDX in comparison with eIF3i monomer (Figure 6, left column), and these changes of solvent protection or destabilization reveal some clues on the binding of eIF3gC1ΔC and are discussed in details.

In order to focus on the significant binding regions, the deuteration change of each peptide before and after complex formation is plotted in Figure 8 from data of 5 s and 30 min reaction. Similar decreases in deuteration levels are shown in N-terminal residues 4-37 in eIF3i:gC1ΔC as in eIF3i:bC3, which may be caused by conformation stabilization upon the bindings. Such coincident decreases in both complexes are also observed in the regions of 195-211, 229-242, 271-282, 278-285 and 288-316. The region of 263-271 in eIF3i:gC1ΔC also exhibits a decrease of deuteration, however less significant than in eIF3i:bC3. These similarities in deuteration behaviour strongly support the primary hypothesis from Figure 15 and Figure 6 (right column) that the two complexes share close binding regions. On the other hand, several other peptides in the regions of 222-228, 251-277 and 317-329 exhibit no decreases in deuteration levels although their neighbouring peptides are less deuterated. These regions are informative to raise the spatial resolution of the binding region elucidation.

Increases in deuteration levels are observed in various peptides in the regions of 87-107 and 125-155 in eIF3i:gC1ΔC. Such significant increases occur exclusively in the presence of eIF3gC1ΔC. Figure 16 shows the locations of these regions on the crystal

structure as well as an unchanged region of 154-187, and the mass spectra of peptide eIF3i(125-154) (left) and eIF3i(154-187) (right) are shown in Figure 17 in different conformations after 30 min deuteration, as labelled on each spectrum. The spectra of these two peptides from the trimer of eIF3i:bC3:gC1ΔC are also extracted and shown here, although the complex has not been analyzed thoroughly yet. Clearly, in eIF3i:gC1ΔC (16.6%) as well as in eIF3i:bC3:gC1ΔC (17.0%), eIF3i(125-154) shows higher deuteration level than in eIF3i monomer (12.0%) and eIF3i:bC3 (11.5%). Notably, the experiments of eIF3i:gC1ΔC and eIF3i:bC3:gC1ΔC were taken in May 2013, while those of eIF3i monomer and eIF3i:bC3 were taken in December 2012 and January 2013. In order to distinguish the changes from instrumental errors, a neighbouring peptide, eIF3i(154-187), is similarly analyzed as a control example. The deuteration levels of eIF3i(154-187) in the monomer, the complex with eIF3bC3, the complex with eIF3gC1ΔC and the trimeric complex are 8.7%, 8.0%, 9.6% and 9.0%, illustrating that despite the experimental deviation, the deuteration change in the peptide eIF3i(125-154) is structurally significant. The change occurs in both complexes containing eIF3gC1ΔC, and this indicates that the binding of eIF3gC1ΔC leads to a destabilization of the region of eIF3i(125-155) located on the outer part of blade 3. Obviously, this region should not be at the direct binding surface, otherwise the deuteration should be prevented by the binding of eIF3gC1ΔC. However, it may be linked to the binding regions so that its conformation is influenced by the complex formation.

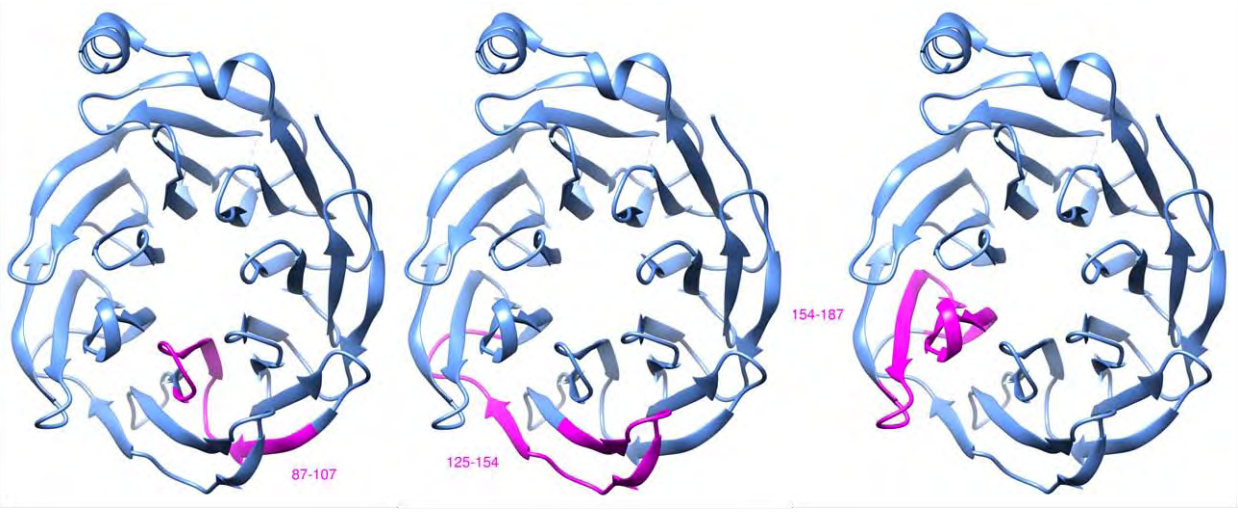


Figure 16 Highlights of the positions of the peptides eIF3i(87-107), eIF3i(125-154) and eIF3i(154-187) on the crystal structure.

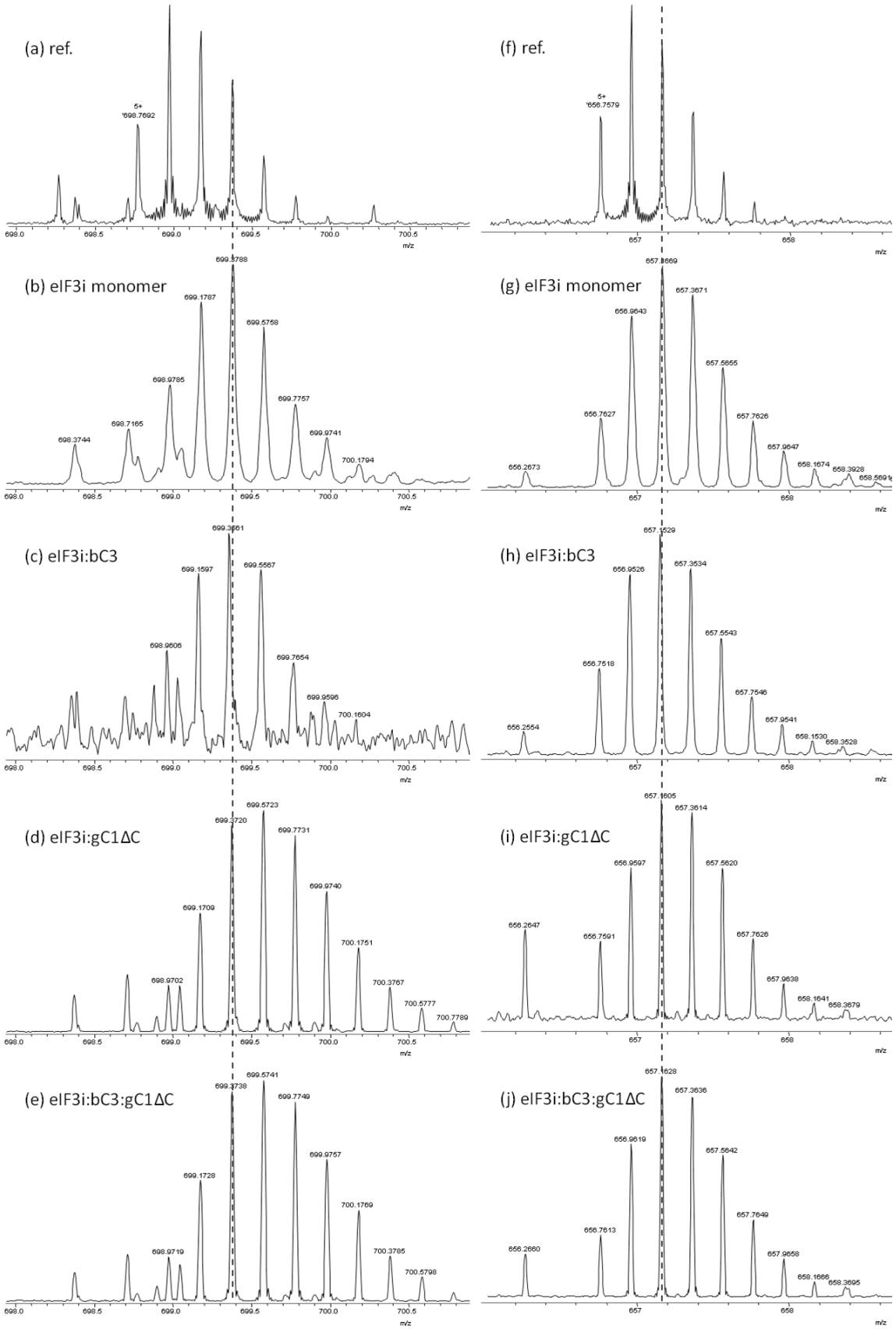


Figure 17 Mass spectra of eIF3i(125-154) (left) and eIF3i(154-187) from different conformations (shown as labelled) after 30 min deuteration.

Besides this region, several other peptides which have significance in the structural elucidation of eIF3i:3g are analyzed in further details and the kinetic plots are generated. These peptides lead to more precise insights of the binding regions.

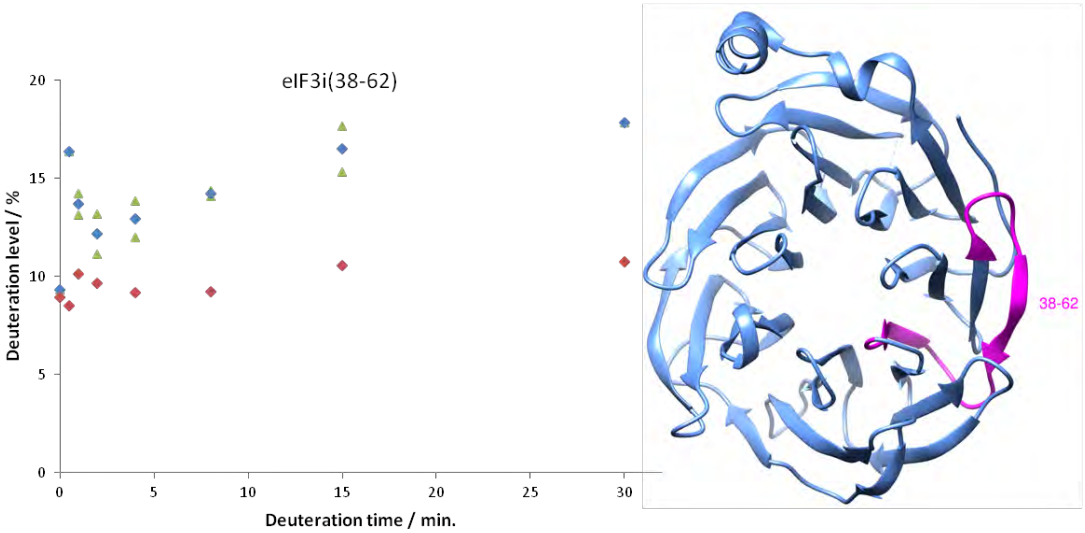


Figure 18 HDX of eIF3i(38-62). Left: Plot of deuteration levels as the average (blue) of two replicates (green) for eIF3i monomer and one set of data for eIF3i:gC1ΔC (red) against the deuteration time. Right: highlight of the position of this peptide on the crystal structure.

The HDX of eIF3i(38-62) is shown in Figure 18, where the deuteration levels decrease upon the binding of eIF3gC1ΔC. Such decreases of deuteration levels are also observed in eIF3i:bC3, which may be attributed to the interaction of the eIF3bC3 C-terminus with the C-terminus of eIF3i(38-62) near the central cavity of eIF3i, or to the stabilization in the N-terminus of eIF3i in blade 1 upon binding with eIF3bC3. As the binding location of eIF3bC3 and eIF3gC1ΔC should not completely overlap, the interaction region of eIF3gC1ΔC is likely to be at the outer surface of blade 1 and the adjacent loops. On the other hand, it is also possible that the decreases are caused by structural stabilization, rather than by the direct binding of the two subunits.

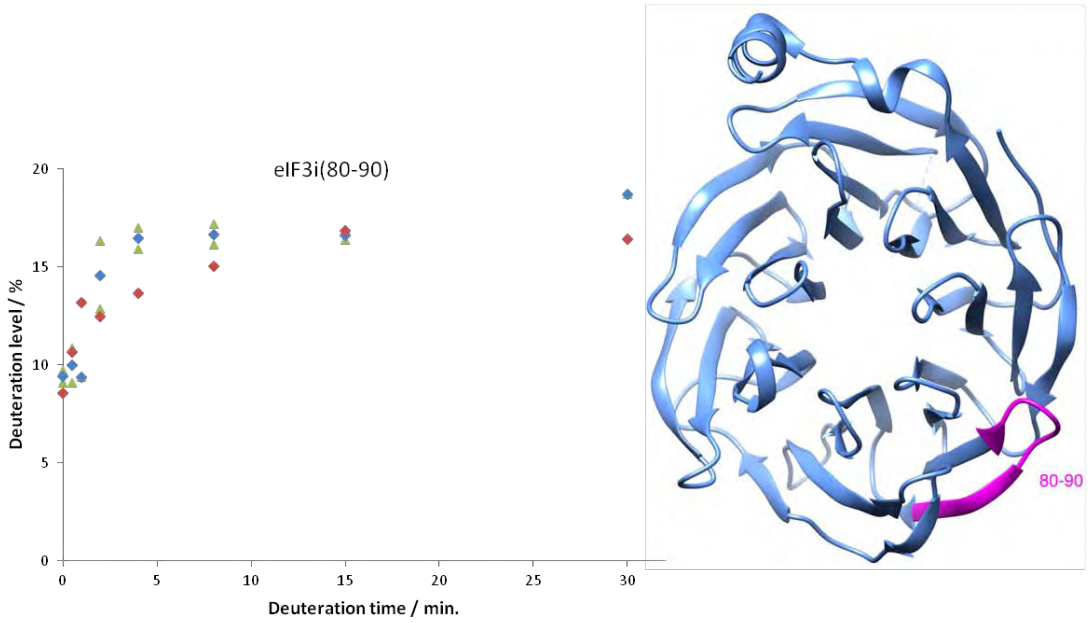


Figure 19 HDX of eIF(80-90). Left: Plot of deuteration levels as the average (blue) of two replicates (green) for eIF3i monomer and one set of data for eIF3i:gC1ΔC (red) against the deuteration time. Right: highlight of the position of this peptide on the crystal structure.

Spatially close to the regions of eIF3i(38-62) is the peptide eIF3i(80-90) as shown in Figure 19, locating in the loop and outer β-sheet of blade 2 in eIF3i. In the kinetic plots in Figure 20 (left), no significant difference is observed after 15 min deuteration, but the exchange in the complex is slower than in the eIF3i monomer as shown by the data points at the short deuteration times. This region is, therefore, not likely to be the direct binding region.

The peptides shown in Figure 20, 21, 22 and Table 1 together provide binding information on blades 5, 6 and 7 of eIF3i. As shown on the top of Figure 20, peptide eIF3i(223-228) exhibits no change in deuteration upon the binding of eIF3gC1ΔC, which can be clearly explained by the top-right structure model as the highlighted eIF3i(222-228) tightly binds to eIF3b(654-700), inhibiting the interaction of this region with eIF3g. The peptide of eIF3i(223-242), however, exhibits significant decreases in deuteration level in the complex. The kinetic plot for peptide eIF3i(229-242) is not shown here as it is not present in all time points; yet it also shows less deuterated in some data points: 23.3% and 24.7% of deuteration are observed in the eIF3i monomer after 15 min and 30 min reaction, respectively, corresponding to 12.7% and 14.3% in eIF3i:gC1ΔC. The kinetic plots in Figure 20 and the partially observed

deuteration change in eIF3i(229-242) indicates the direct interaction in the outer β-sheet of blade 5 and the long loop region between blades 5 and 6.

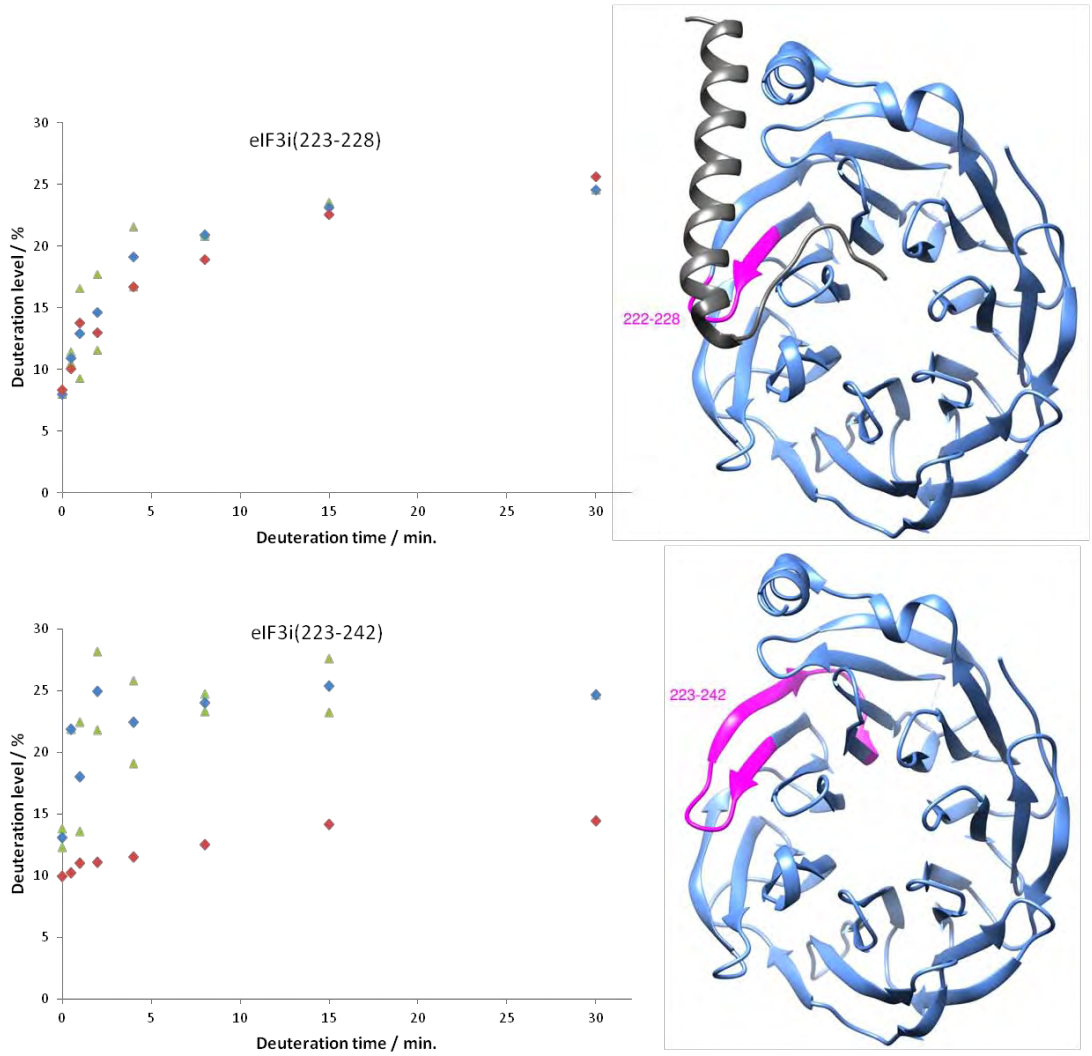


Figure 20 HDX of eIF3i(223-228) (top) and eIF3i(223-242) (bottom). Left: Plots of deuteration levels as the average (blue) of two replicates (green) for eIF3i monomer and one set of data for eIF3i:gC1ΔC (red) against the deuteration time. Right: highlights of the position of the peptides on the crystal structure.

The regions of eIF3i(251-285) is located on blade 6. Peptide eIF3i(251-262) is less deuterated in the complex (Table 1), whereas an increase in deuteration level is observed in the peptide eIF3i(251-273) (Figure 21, top and Table 1) as well as in the peptide eIF3i(251-277) (Table 1). There is no separate information for any peptides in 263-277 (Figure 22, left), but by comparison of the information in Table 1, it is clear that peptide 251-262 is involved directly in the binding region, while 263-277, the unstructured loop and part of the β-sheet is destabilized by the complex formation. The peptide eIF3i(278-285) again exhibits lower deuteration levels in the complex

(Figure 21, bottom). As this peptide is located on the same side of eIF3i(251-262) in blade 6, this co-supports the hypothesis that the binding of eIF3gC1ΔC occurs exclusively with the “left” part of blade 6.

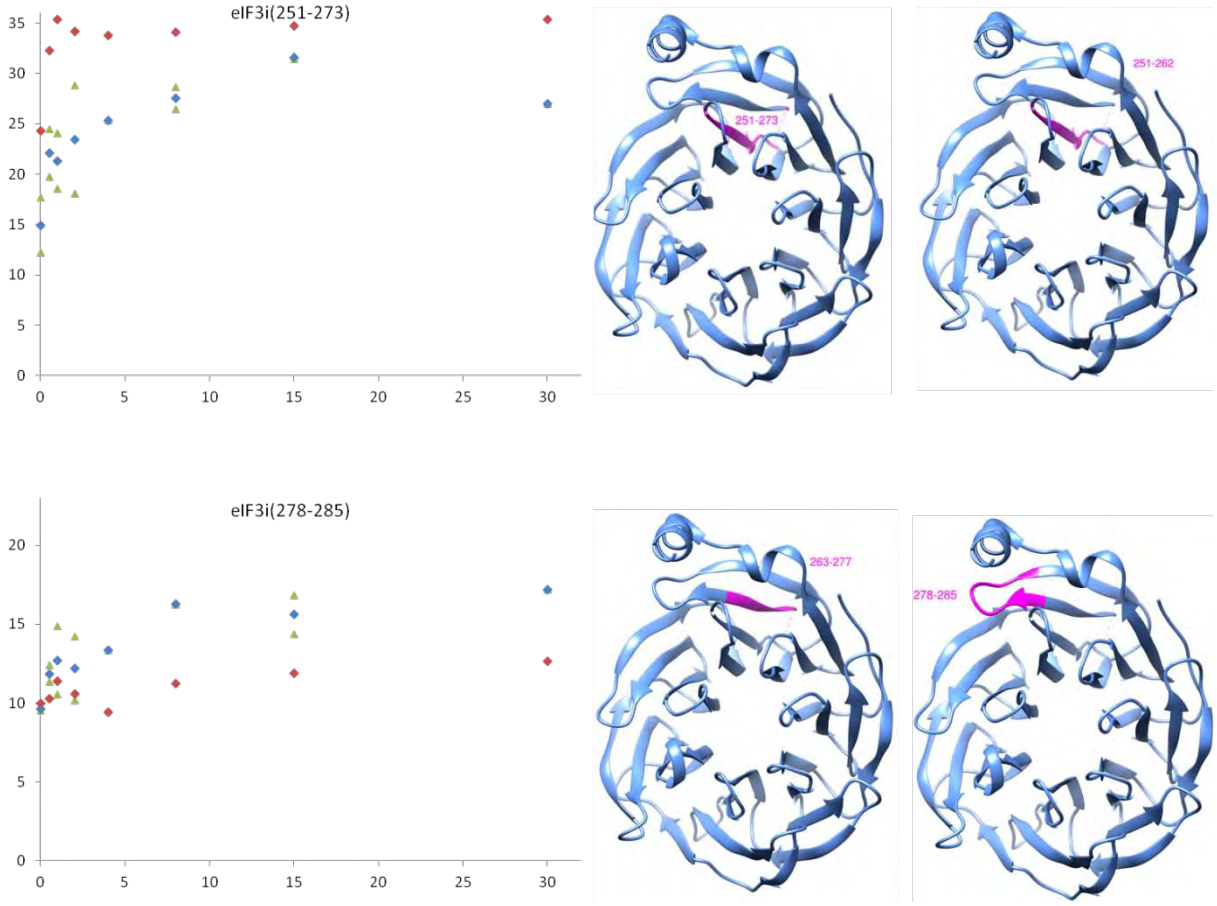


Figure 21 HDX of eIF3i(251-273) and eIF3i(278-285). Left: Plots of deuteration levels as the average (blue) of two replicates (green) for eIF3i monomer and one set of data for eIF3i:gC1ΔC (red) against the deuteration time. Right: highlights of the positions of the peptides on the crystal structure.

Table 1 Deuteration levels of selected peptides.

	eIF3i monomer			eIF3i:gC1ΔC		
time	5 s	15 min	30 min	5 s	15 min	30 min
251-262	14.7%	32.1%	32.8%	18.7%	25.2%	25.5%
251-273	17.7%	31.5%	27.0%	24.3%	34.7%	35.4%
251-277	16.5%	23.9%	23.7%	20.5%	26.3%	27.0%
288-300	13.4%	21.3%	24.0%	11.3%	18.3%	17.7%
299-315	13.5%	23.3%	26.4%	10.1%	16.1%	17.2%
317-329	8.9%	12.2%	13.3%	8.0%	14.2%	13.9%

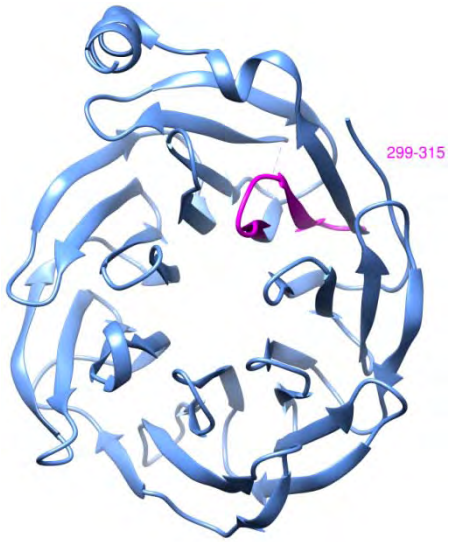


Figure 22 Highlight of eIF3i(299-315) on the crystal structure.

Peptide eIF3i(288-300) consists of part of the outer β-sheet of blade 6, the loop between blades 6 and 7 and the inner β-sheet of blade 7（Figure 12, right), and it also exhibits significantly lower deuteration levels in the complex (Table 1). As demonstrated above with the evidence of eIF3i(263-277), it is not likely that the N-terminus of 288-300 is in the binding region. Rather, as the C-terminal β-sheet structure locates close to eIF3i(251-262), it is more probable that this region interacts with eIF3gC1ΔC and causes the decrease of the deuteration level. The decrease in deuteration level in eIF3i(299-315) (Table 1) supports that the binding occurs in the inner side of blade 7 (Figure 22), while the similar deuteration levels before and after complex formation for peptide eIF3i(317-329) (Figure 13, right) indicates that the outer β-sheet and the short α-helix are not in the binding region.

A summarized view of the HDX change upon the formation of eIF3i:gC1ΔC is shown in Figure 14 (middle column). It has to be noted that some details discussed above cannot be revealed directly due to the peptide-level resolution. Nevertheless, this figure provides a direct overview of the HDX change in the two forms of eIF3i. Intensive red regions are destabilized by the binding of eIF3gC1ΔC, e.g in blades 2 and 3. White and pink regions are not involved in the interaction and do not change in the HDX behavior, whereas the blue regions show significant decreases in deuteration levels and should be in the binding surface.

2.4 Discussion

The partial interaction map of eIF3i:b has been unveiled by X-ray crystallography [1], however the monomeric eIF3i has not been crystallized, and the binding information was only reported on a restricted fragment of eIF3b. In this work, HDX-MS first shows a possible conformation change of eIF3i upon complex formation with eIF3bC3. The “inner” blades 3 and 4 exhibit low deuteration levels along the time course, while the N- and C-termini show higher deuteration levels. This may be attributed to a higher flexibility at both termini, and a frequent open-close motion of the β-propeller structure of eIF3i monomer.

The HDX behavior upon the complex formation of eIF3i:bC3 is coherent with the reported structure of eIF3i:b(654-700). The reported binding regions in eIF3i:b(654-700) are significantly less deuterated, either due to the hydrogen bond formation or spatial shielding of eIF3bC3. In addition, there are several regions showing significant HDX changes which are not in the binding surface of eIF3b(654-700), and further interaction information of the eIF3b(640-653) and eIF3b(701-735) with eIF3i can be elucidated. The validated correlations between the structure model of eIF3i:b(654-700) and the HDX of peptides in known binding regions serve at this stage as the inner controls. The extended N-terminal fragments in eIF3bC3 may bend towards the outer surface of blades 6 and 7 in eIF3i, while the C-terminus of eIF3bC3 may interact with the central cavity of the eIF3i β-propeller and extend further towards blade 1.

The complex of eIF3i:g still lacks atomic structure model from X-ray crystallography, nor is there any high-resolution EM information. However, it has been reported that the subunits of eIF3i, eIF3b and eIF3g each interacts with the others [7]. Therefore the binding surface of eIF3i:b and eIF3i:g may be close to each other. The HDX of eIF3i:gC1ΔC is coherent with this hypothesis: significant deuteration changes are observed in both complexes in blades 1, 5, 6 and 7. The known interaction sites of eIF3i with eIF3b are informative to unveil the interaction of eIF3g, as although the binding surfaces may be close to each other, the binding sites should be exclusive: one residue in eIF3i is not likely to bind with both eIF3b and eIF3g. Thorough analysis of the information from the above various methods provides some clues of the structure of eIF3i:gC1ΔC: eIF3gC1ΔC interacts with blades 1, the outer loop of blade 2, inner β-sheets of blade 7 and the “right” half of blade 6. It causes a

significant destabilization of blade 2 and 3, and it does not interact with blades 4 or 5, the “left” half of blade 6 and the outer β-sheet of blade 7.

A comparison of deuteration upon the formation of eIF3i:bC3 and eIF3i:gC1ΔC is shown in Figure 14 (right column), with the previous comparisons between the free form of eIF3i and the complexes, as labeled on top of each structure. The interaction surfaces of the two complexes can be reviewed: regions in white are either not in the binding surfaces as in blades 2 and 4, or strongly interact with both eIF3b and eIF3g as in blade 1 and 7. Regions in blue shows that the deuteration levels decrease more significantly in eIF3i:bC3 and therefore is on its binding surface as in blades 4, 5 and 6, or that there is an increase of deuteration level in eIF3i:gC1ΔC as in blade 3. Regions in red show stronger decrease in deuteration level in eIF3i:gC1ΔC as in blades 1 and 6, indicating that these are the exclusive binding regions of eIF3g.

In summary, eIF3bC3 and eIF3gC1ΔC share close binding surface with eIF3i. They form a shell-like protection on both sides of blades 1, 5, 6 and 7, from the central cavity to the outer surface. Blades 2, 3 and 4 in eIF3i are not in the binding surface of either complex. This side of eIF3i is more likely to be the functional region as it remains accessible to other proteins.

HDX-MS probes the structural information based on the different behavior of deuterium exchange for amide hydrogens at different locations. It correlates the protein structure with the HDX of given peptides, but is not possible to generate complete structure model or interaction map using only this method. However, with the help of even incomplete structural information provided by other methods such as X-ray crystallography, EM and native MS, HDX-MS can be powerful and sensitive to probe the interactions of protein complexes.

References

1. Herrmannová, A.; Daujotytė, D.; Yang, J.-C.; Cuchalová, L.; Gorrec, F.; Wagner, S.; Dányi, I.; Lukavsky, P. J.; Valášek, L. S., Structural analysis of an eIF3 subcomplex reveals conserved interactions required for a stable and proper translation pre-initiation complex assembly. *Nucleic Acids Res* **2012,** *40*, 2294-2311.
2. Zhang, Q.; Chen, J.; Kuwajima, K.; Zhang, H.-M.; Xian, F.; Young, N. L.; Marshall, A. G., Nucleotide-induced conformational changes of tetradecameric GroEL mapped by H/D exchange monitored by FT-ICR mass spectrometry. *Sci. Rep.* **2013,** *3*.
3. Wei, H.; Ahn, J.; Yu, Y. Q.; Tymiak, A.; Engen, J. R.; Chen, G., Using hydrogen/deuterium exchange mass spectrometry to study conformational changes in granulocyte colony stimulating factor upon PEGylation. *J. Am. Soc. Mass. Spectrom.* **2012,** *23*, 498-504.
4. Trelle, M. B.; Hirschberg, D.; Jansson, A.; Ploug, M.; Roepstorff, P.; Andreasen, P. A.; Jørgensen, T. J., Hydrogen/deuterium exchange mass spectrometry reveals specific changes in the local flexibility of plasminogen activator inhibitor 1 upon binding to the somatomedin B domain of vitronectin. *Biochemistry* **2012,** *51*, 8256-8266.
5. Cravello, L.; Lascoux, D.; Forest, E., Use of different proteases working in acidic conditions to improve sequence coverage and resolution in hydrogen/deuterium exchange of large proteins. *Rapid Commun. Mass Spectrom.* **2003,** *17*, 2387-2393.
6. Mazon, H.; Marcillat, O.; Forest, E.; Vial, C., Local dynamics measured by hydrogen/deuterium exchange and mass spectrometry of creatine kinase digested by two proteases. *Biochimie* **2005,** *87*, 1101-1110.
7. Zhou, M.; Sandercock, A. M.; Fraser, C. S.; Ridlova, G.; Stephens, E.; Schenauer, M. R.; Yokoi-Fong, T.; Barsky, D.; Leary, J. A.; Hershey, J. W., Mass spectrometry reveals modularity and a complete subunit interaction map of the eukaryotic translation factor eIF3. *PNAS* **2008,** *105*, 18139-18144.

Part III

Methods

Chapter 5

Materials and methods

1 NanoLC system in HDX-MS 135
2 FT-ICR MS 137
2.1 The principles of FT-ICR MS 138
2.2 Factors ensuring the high performance of FT-ICR MS 140
2.3 The evolution of instruments during this work 141
2.3.1 APEX III 141
2.3.2 APEX Q 142
2.3.3 SolariX 142
3 Experimental protocols 144
3.1 Protein production and purification 144
3.1.1 Production of eIF3i, eIF3bC3 and eIF3gC1ΔC 144
3.1.2 Protein purification 145
3.1.3 Complex purification 146
3.1.4 Protein storage 146
3.2 Hydrogen/deuterium exchange reaction 146
3.3 NanoLC-MS 146
3.4 LC-MS/MS 148
References 150

In this work, proteins were produced and purified by the author in our collaborating laboratory, the Biochemistry Laboratory in Ecole Polytechnique. The HDX-MS experiments were carried out in the host laboratory of DCMR. After deuteration and pepsin digestion, the peptide mixtures were desalted and separated using a nanoLC system, ionized by a nanoESI ion source and analyzed by a 7 T FT-ICR MS.

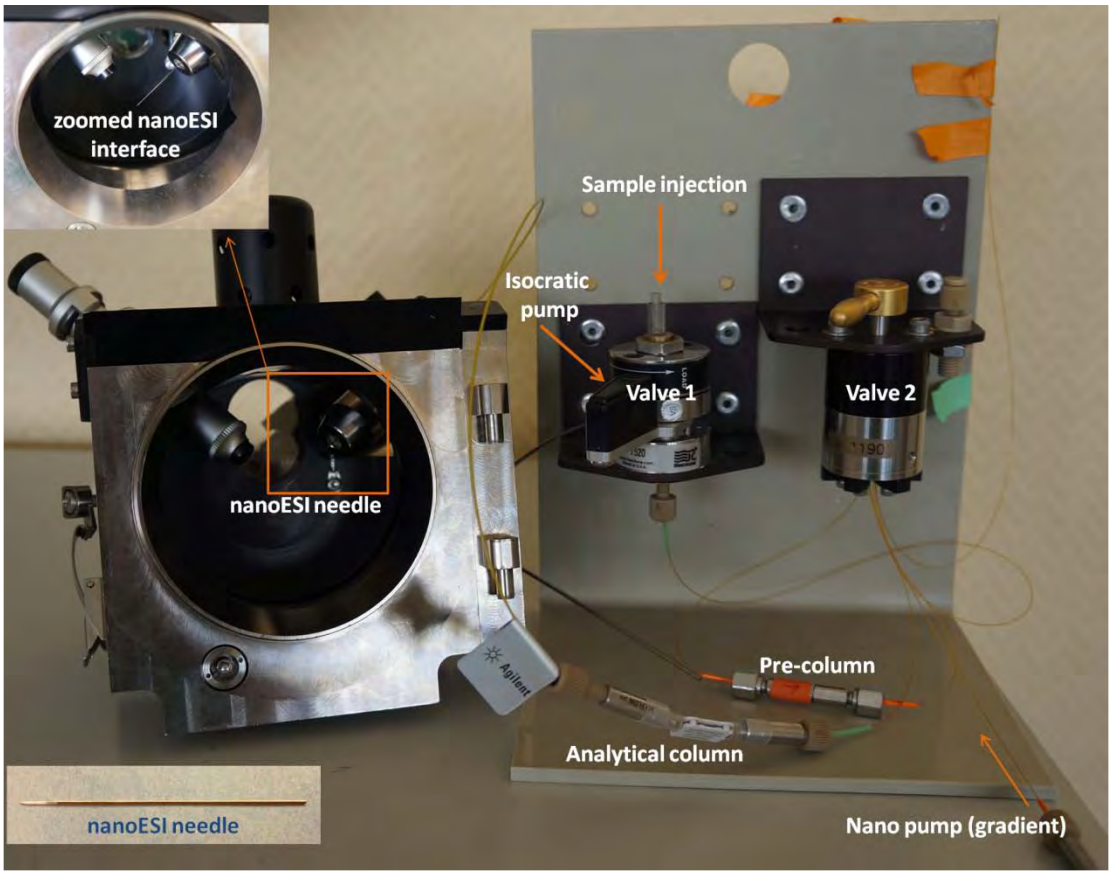
The LC-MS/MS experiments were carried out in the Proteomics Platform in Pasteur Institute, Paris, using a LTQ Orbitrap Velos (Thermo Fisher) MS connected to a nano-HPLC Ultimate 3000 (Dionex).

The instrumentation of the FT-ICR MS played an important role throughout the method development and data processing of this work, and is therefore discussed in details in this part. The other part of the instrumentation will also be described briefly.

1 NanoLC system in HDX-MS

The nanoLC instrument used in the HDX-MS part of this work consists of a home-built double-valve column system connected to an Agilent Technologies 1100 Series nanoLC system. In the double valve system, the injection valve is a microsample injector (Rheodyne 7520) with a syringe port and an internal injection loop volume of 0.5 µl. It is connected to a 6 channel valve (Micro Valve Upchurch Scientific M-485) as shown in Figure 1. This second valve has a dead volumn of 35 nL. Silica capillaries (Polymicro Technologies) of two different inner diameters are used: 20 µm for the parts involved in the gradient (Figure 1, red) and 75µm for the remaining connections (Figure 1, blue). The valves are manually switched in order to inject the sample into the sample loop or load the sample from the loop onto the precolumn for valve 1, and change between washing the precolumn and elute the sample through the analytical column for valve 2.

A Zorbax 300SB-C_{18} column, which is 5 mm long and 0.3 mm in internal diameter, is installed as the precolumn to desalt and concentrate the sample. The stationary phase of this column comprises octadecyl carbon chain (C_{18})-bonded silica particles with an average diameter of 5 µm. The analytical column is a Zorbax 300SB-C_{18} column of 5 cm length and 0.075 mm inner diameter and the diameter of the silica particles in the stationary phase is 3.5µm.



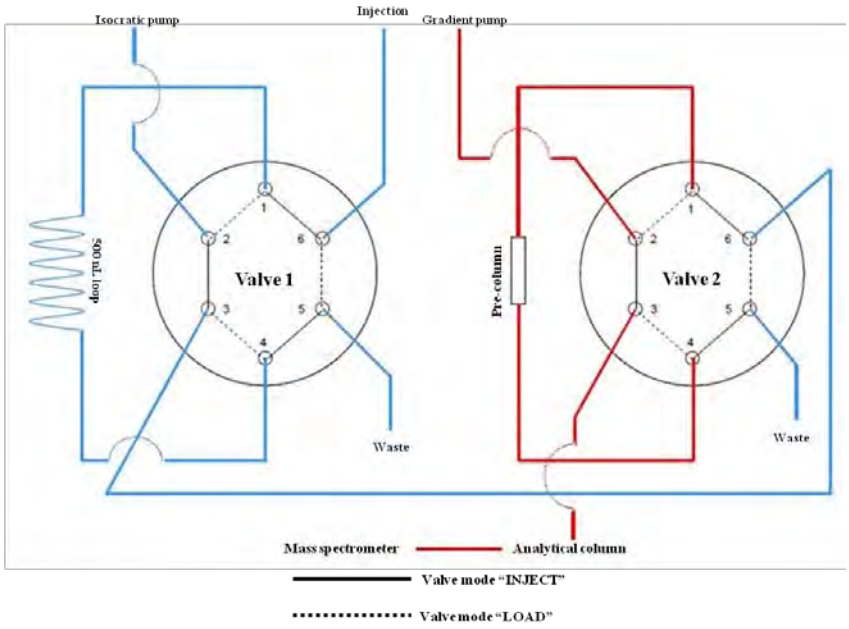


Figure 1 Scheme of the double-valve system.

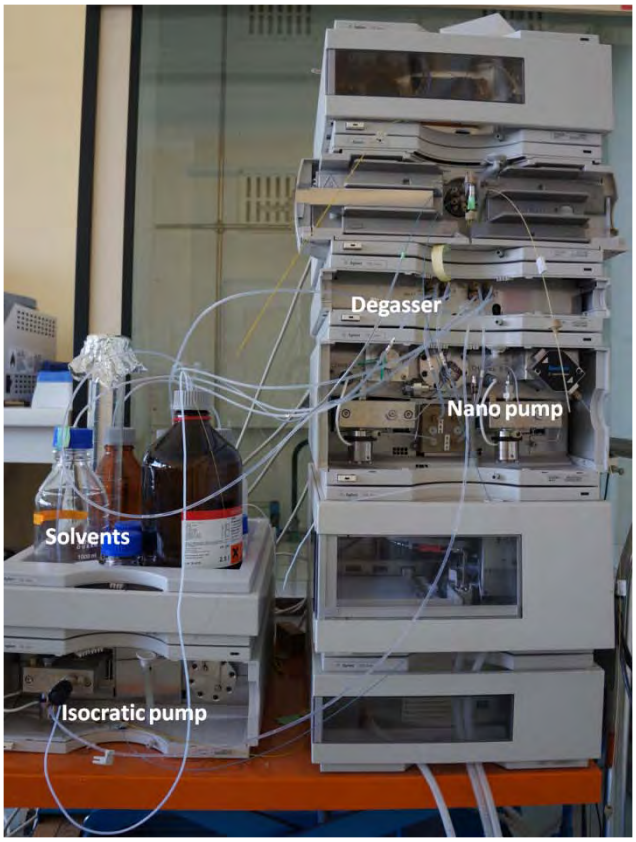


Figure 2 Photo of the Agilent Technologies 1100 Series nanoLC system.

The gradient is controlled by an Agilent Technologies 1100 Series nanoLC system (Figure 2). Inorganic solvent and organic solvent are separately degassed, and transferred to the nano pump where they are mixed according to the gradient. The Agilent nano-LC system is based on an active split of the primary flow generating the secondary nano-flow after mixing.

The interface between the nanoLC and the FT-ICR MS, i.e. the nano-electrospary (nanoESI) is generated in a nanoESI source using PicoTip Emitter needles purchased from New Objective (Woburn, MA, USA), which has an outer diameter of 360μm, an internal diameter of 75μm and an outlet of 10μm for the spray formation. The sprayed ions are transferred into the FT-ICR MS and analyzed.

2 FT-ICR MS

FTMS, including FT-ICR MS and Orbitrap MS, has the highest resolution among all the modern MS techniques, which is crucial to this work. The FT-ICR MS was first introduced in 1974 by Melvin Comisarow and Alan Marshall [1], and nowadays with

the development of the instrumentation as well as the electromagnet it is playing an powerful role in various applications. Numerous reviews and textbooks have been published to introduce its principle, application and perspectives. [2-4] Here only a brief illustration of the instrumental principle and relevant evolution during this work will be presented in this part.

2.1 The principles of FT-ICR MS

FTMS is based on the technique of ion cyclotron resonance (ICR) mass spectrometry: a charged particle will precess in a magnetic field at a frequency depending on its m/z value [1, 3-4].

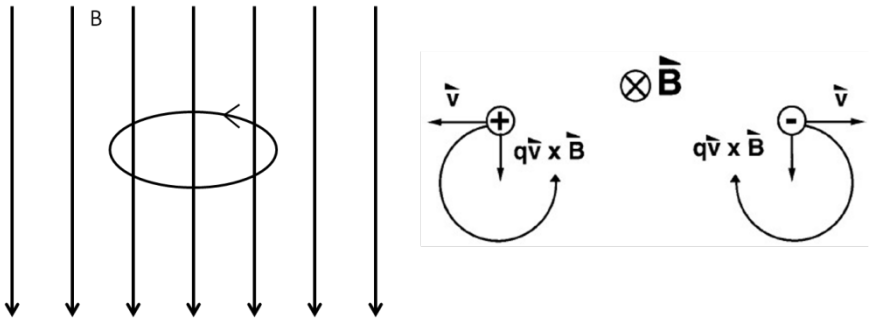


Figure 3 Precessing ion in a magnetic field. [3]

The moving ion is subject to a force $\vec{F}$ given by:

$$\vec{F} = q\vec{v} \times \vec{B} \qquad (1)$$

where q, $\vec{v}$ and $\vec{B}$ represent the charge, velocity and uniform magnetic field, respectively. The ion moves in a circle with a radius (r) which is perpendicular to the magnetic field as shown in Figure 2, and the positive and negative ions orbit in opposite trajectories. If the opposite direction of the magnetic field is denoted as the z-axis, i.e. $\vec{B} = -B_0\vec{k}$, the ion velocity in the xy plane, v_{xy} can be determined according to Equation 1 as

$$\frac{mv_{xy}^2}{r} = qv_{xy}B_0 \qquad (2)$$

with m as the mass of the particle. The angular velocity, ω (in rad/s), about the z-axis is defined as

$$\omega = \frac{v_{xy}}{r} \quad (3)$$

so that Equation 2 can be deduced as

$$\omega = \frac{qB_0}{m} \quad (4)$$

and further as

$$v_c = \frac{\omega_c}{2\pi} = \frac{1.535611 \times 10^7 B_0}{m/z} \quad (5)$$

Equation 5 is the so-called cyclotron equation, in which the frequency v_c is in Hz, B_0 is in Tesla, m is in u and z is the charge state of the particle. Thus ions with the same m/z value have the same frequency v_c given that the magnetic field remains constant, independent of their velocity. This is the essential feature that allows the ICR frequency to be used for m/z determination.

According to Equation 2, the orbital radius is dependent on the velocity and the m/z. Ions of thermal energy have thus a radius:

$$r = \frac{1}{qB_0}\sqrt{2mkT} \quad (6)$$

where k is the Boltzmann constant and T is the temperature in K. This indicates that the thermal radius is proportional to the square root of m and reciprocal to the magnetic field.

However, the frequency is not readily detected if the ions oscillate at the thermal radius. A spatially uniform electric field oscillating at or near the ICR frequency has to be applied, in order to accelerate the ions to a larger orbital radius detectable by the detector (Figure 4, left). The excitation pulse makes the ions reach the same radius close to the detection plates, ensuring all ions to have coherent movement, which generate a current that is detected by the detection plates (Figure 4, right). The time-domain ICR signals are Fourier transformed into frequency domain magnitude spectra that are, as demonstrated in Equation 5, reciprocal to m/z.

The cubic cell was initially developed and widely used as the ICR cell, which consists of six plates arranged in a cubic shape, in three opposite pairs as the trapping, excitation and detection plates. In our instrument, a cylindrical cell is installed, in

which the trapping plates remain planar while the excitation and detection plates are cylindrical.

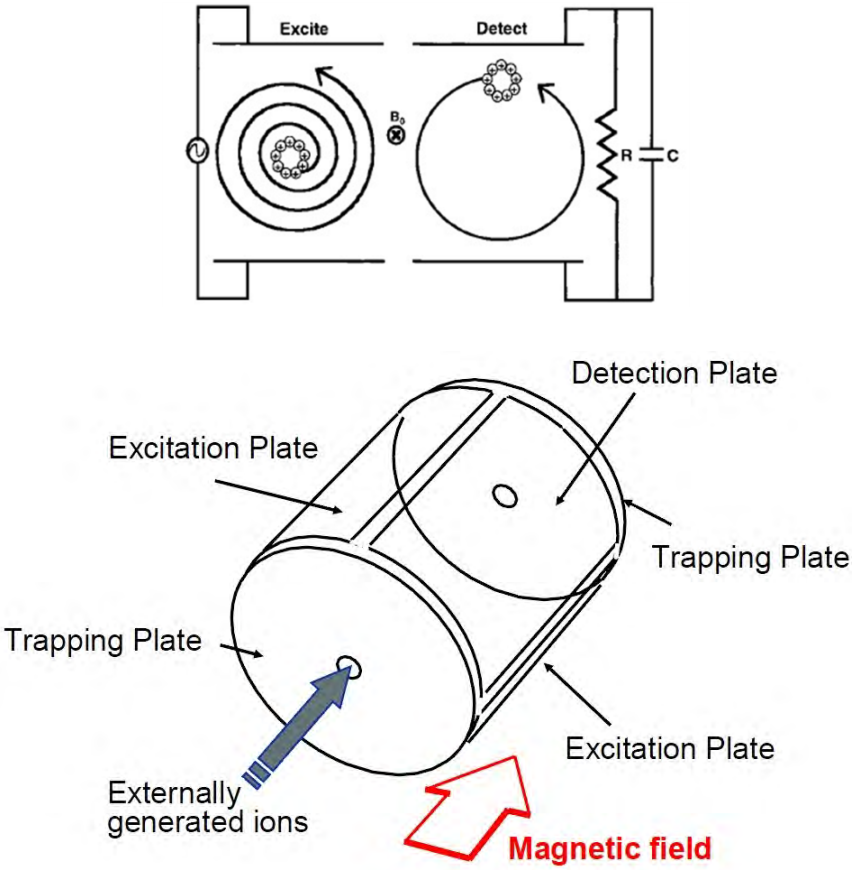


Figure 4 Scheme of the excitation and detection in the ICR cell (top) [3] and illustration of a cylindrical ICR cell (from Bruker document).

2.2 Factors ensuring the high performance of FT-ICR MS

The resolution in FT instruments is usually defined as the full width of a spectral peak at half-maximum peak height, termed as $\Delta m_{50\%}$ in mass-domain FT-ICR spectra. The resolving power is thus defined as $m/\Delta m_{50\%}$. Deduction of the above equations gives the expression of the resolving power:

$$\frac{m}{\Delta m_{50\%}} = -\frac{qB_0}{m\Delta\omega_{50\%}} \qquad (7)$$

where $\Delta\omega_{50\%}$ is frequency domain resolution. The resolving power differs thus with the mass as well as the frequency resolution, the latter depending on the acquisition time. Thus the correlation of these factors can be illustrated as in Figure 5. Generally,

higher magnetic field and longer acquisition time leads to higher resolution for the FT-ICR MS, while the resolving power decreases with the increase of the mass.

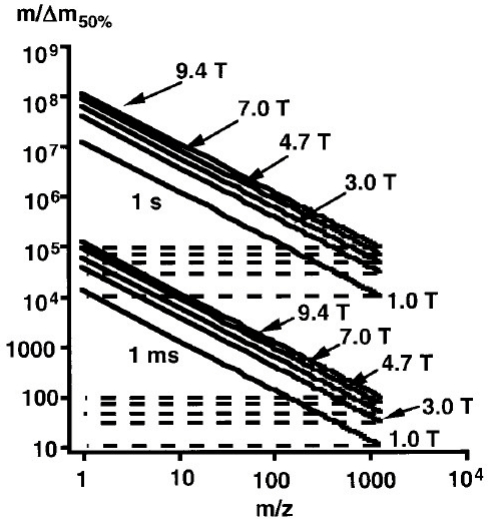


Figure 5 Correlation of the mass resolving power of FT-ICR MS with the magnetic field, m/z and acquisition time [3].

2.3 The evolution of instruments during this work

This work started in October 2010, with a 7 T FT-ICR APEX III mass spectrometer (Bruker, Germany). The instrument was upgraded to APEX Q (Bruker, Germany) in January 2011, and again upgraded to SolariX (Bruker, Germany) in April 2012, whereas the ICR cell and the magnet remained the same. Although as demonstrated above, it is the magnetic field of the ICR cell that primarily determines the resolution of the FT-ICR MS, the earlier stages of the instrument, including the ion source, the ion storage and transfer elements, are also crucial to achieve the high performance.

2.3.1 APEX III

In the initial APEX III instrumentation, ions formed by the nano-ESI source as described above are transferred, through a set of electrostatic lenses, to the center of the cell locating in the center of the superconductive electromagnet. A hexapole storage is inserted after the capillary in the ESI source, and before electrostatic transfer system, in order to accumulate ions before sending them into the ICR cell, where the ions are analyzed and detected by measuring the cyclotron frequency.

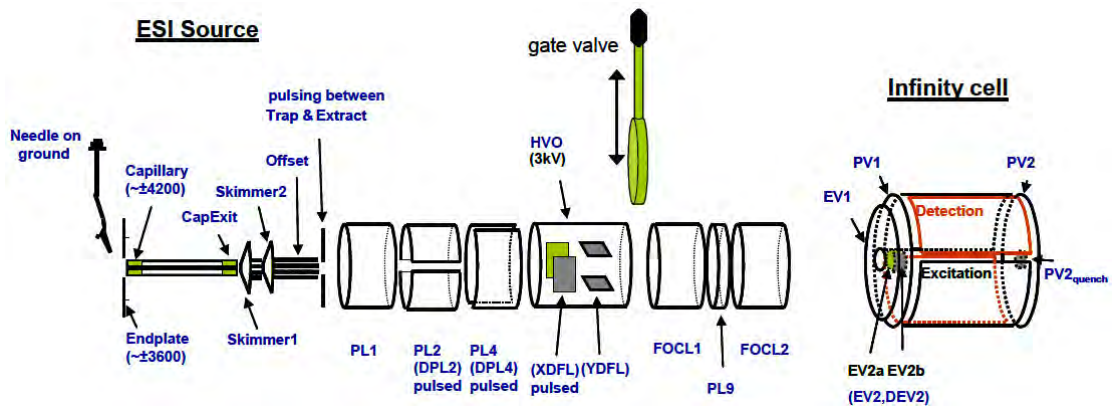


Figure 6 Scheme of the FT-ICR APEX III MS. (From Bruker document)

2.3.2 APEX Q

The first instrumental upgrade was the replacement of APEX III by APEX Q hybrid q-FT-ICR instrument (Figure 6). Several features essentially increased the performance of the FT-ICR MS. An ion funnel system, located after the ESI source capillary allowed an improved transfer of the ions through the pressure restriction systems. A quadrupole was added in order to allow possible selection of parent ions. A collision hexapole was also added to allow CID fragmentation of the quadrupole selected parent ions. Significantly higher sensitivity was achieved upon this upgrade as demonstrated in Chapter 3: marginally larger numbers of ions were detected which led to an increase in the sequence coverage of the HDX analysis.

2.3.3 SolariX

In April 2012, a new generation Bruker SolariX package was installed to replace the APEX Q system (Figure 7). An orthogonal change of the ion transfer direction was introduced after the ion source, which reduces the population of neutral particles entering the ion funnel system. Also important for this work is the addition of a new hexapole ion guide as well as a second ion funnel which again increased the efficiency of the ion transfer.

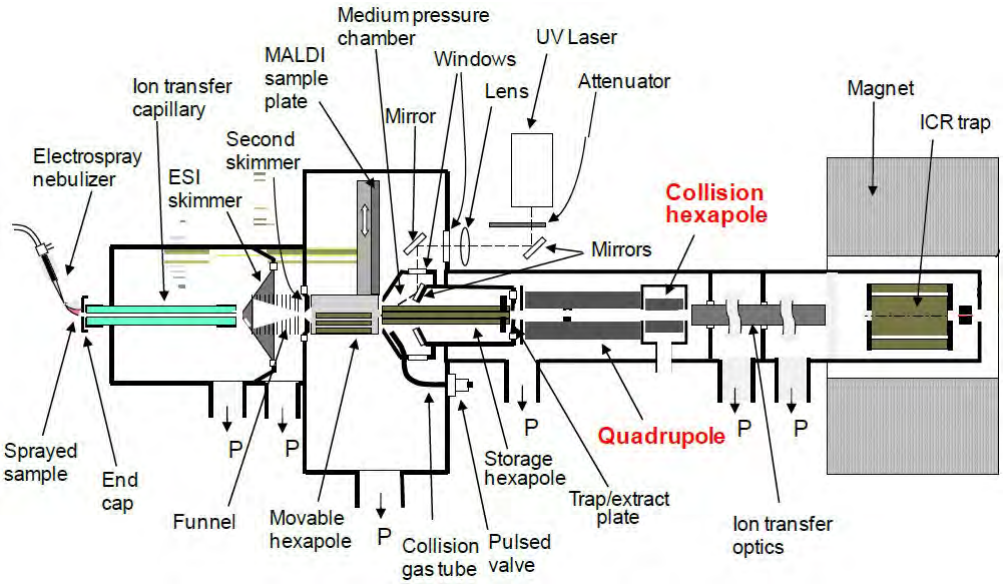


Figure 7 Scheme of the FT-ICR APEX Q MS. (From Bruker document)



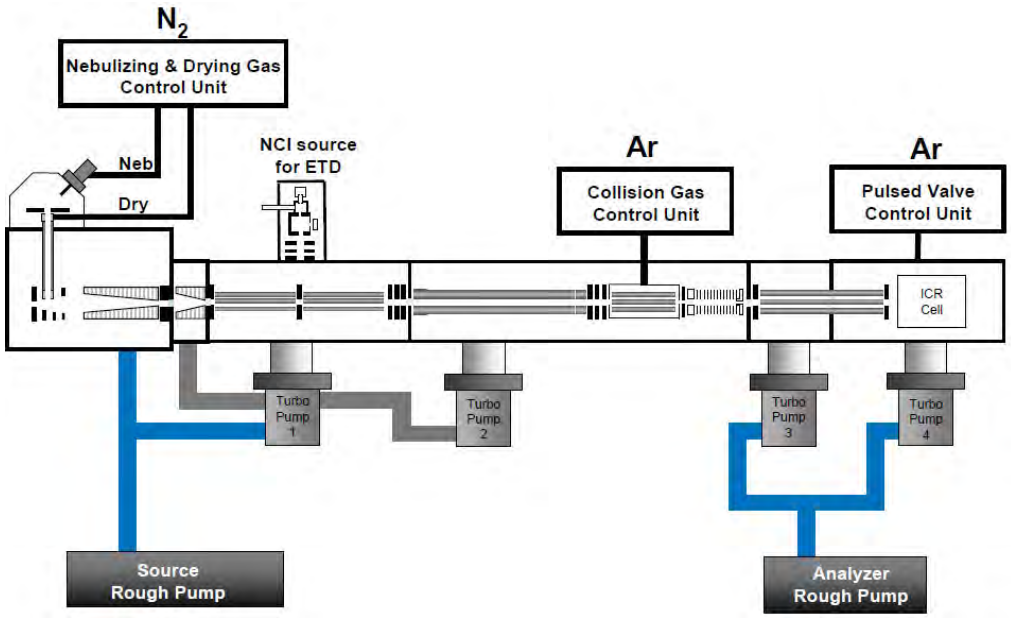


Figure 8 Photo (top) and scheme (bottom) of the FT-ICR SolariX MS (from Bruker document).

3 Experimental protocols

Water used in this work was purified using a Millipore Direct-Q 3. D_2O, acetic acid, formic acid, acetonitrile, pepsin and other chemicals were purchased from commercial suppliers of Sigma, Aldrich, Merck or Roth.

3.1 Protein production and purification

3.1.1 Production of eIF3i, eIF3bC3 and eIF3gC1ΔC

The plasmid pET3a-lpa-eIF3i was transformed by electroporation into BL21(DE3)-Rosetta competent cells (40 µL) and inoculated overnight in 1L 2xTY media at 18 °C. The next morning, 1 mM IPTG was injected to induce the expression of the eIF3i gene at 18 °C for one day.

The plasmids pET32b-eIF3bC3 and pET15b-lpa-eIF3gC1ΔC were independently used to transform by electroporation BL21(DE3)-Rosetta competent cells. The transformed cells were used to inoculate autoinductible ZY media at 18 °C overnight. The cells were let grown 3 days until the turbidity of the media was high enough (Abs <= 6).

The proteins of eIF3i and eIF3gC1ΔC have tags on the N-termini as shown in Table 1 for the initial 20 residues, i.e. the wild-type protein sequences starts from the 21st residue. eIF3bC3 has a his-tag on the C-terminus, i.e. the last six residues shown in Table 1.

Table 1 Sequences of the proteins.

Protein / Mw	**Sequence**
eIF3i / 40918.7 Da	MGSSHHHHHHSSGLVPRGSHMKAIKLTGHERPLTQVKYNKEGDLLFSCSKDSSASV WYSLNGERLGTLDGHTGTIWSIDVDCFTKYCVTGSADYSIKLWDVSNGQCVATWK SPVPVKRVEFSPCGNYFLAILDNVMKNPGSINIYEIERDSATHELTKVSEEPIHKIITHE GLDAATVAGWSTKGKYIIAGHKDGKISKYDVSNNYEYVDSIDLHEKSISDMQFSPDL TYFITSSRDTNSFLVDVSTLQVLKKYETDCPLNTAVITPLKEFIILGGGQEAKDVTTTS ANEGKFEARFYHKIFEEEIGRVQGHFGPLNTVAISPQGTSYASGGEDGFIRLHHFEKS YFDFKYDVEKAAEAKEHMQEAN
eIF3bC3 /	MLREWSAQFEEQDAMEADTAMRDLILHQRELLKQWTEYREKIGQEMEKSMNFKIF

11324.5 Da	DVQPEDASDDFTTIEEIVEEVLEETKEKVELEHHHHHH
eIF3gC1ΔC / 16199 Da	MGSSHHHHHHSSGLVPRGSHMEKVHKSVAERKNWHKYGSEKGSPAGPSAVTARLG EEVELRLSRNWKQAEEERIQKEKASLTKTGLQCRLCGNDHMTMNCPFKTILSELSAL EDPATNEGGVEAASEEKAGQVGGAGSIPGQYVPPSRR

3.1.2 Protein purification

- Cell extraction

The harvested cells were centrifuged to separate from the supernatant, resuspended into the breaking buffer (10 mM MOPS at pH 6.7, 500 mM NaCl, 2 mM β-mercaptoethanol, 0.1 mM PMSF and antiprotease cocktail, 30 mL for 1 L culture) and sonicated for 600s (1s pulse ON, 1s pulse OFF) to break the cell membrane. The cell extract was then centrifuged in a JA-20 rotor at 20 000 rpm at 4°C for 30 minutes.

- FPLC purification and quality control

An FPLC system from Biorad controlled by the BioLogic DuoFlow software was used to purify the proteins.

A TALON (2 mL, Clonetech) column was equilibrated with buffer A (10 mM MOPS at pH 6.7, 500 mM NaCl, 2 mM β-mercaptoethanol, 25 mM imidazole). The supernatant was loaded on the TALON with 60 mL buffer A at a flow rate of 0.5 mL/min, washed with 30 mL buffer A (15 column volumes) at 1 mL/min and the protein was then eluted with 20 mL (10 column volumes) of buffer B (10 mM MOPS at pH 6.7, 500 mM NaCl, 2 mM β-mercaptoethanol, 250 mM imidazole) at 1 mL/min.

The eluted peak fractions were collected, and applied on a Q-Hiload equilibrated with the buffer (10 mM MOPS at pH 6.7, 500 mM NaCl, 2 mM β-mercaptoethanol) to get rid of nucleic acids (flow of 2.5 mL/min).

The flow-through peak fractions of Q-Hiload was then concentrated by centrifugation at 4 °C to less than 5 mL using 20 mL centrifugation tubes of 10000MWCO or 30000MWCO from Vivaspin. The sample was then applied on a Superdex75 column equilibrated with the buffer (10 mM MOPS pH 6.7, 200 mM NaCl, 2 mM β-mercaptoethanol) and washed with 120 mL of the buffer at a flow rate of 0.4 mL/min.

The purities of the peak fractions corresponding to the protein of interest were controlled by SDS-PAGE gel after each TALON and Superdex75 column.

The samples were then concentrated to 50-150 mM by centrifugation at 4 °C using 5 mL centrifugation tubes of 10000MWCO or 30000MWCO for the HDX-MS experiments, and the concentrations were estimated by the UV-absorbance at 280 nm using Nano Drop 1000 UV-Vis spectrometer.

3.1.3 Complex purification

eIF3i:bC3, eIF3i:gC1ΔC and eIF3i:bC3:gC1ΔC were produced by mixing the independently purified proteins. The same Superdex75 column was then used to purify the complex mixture containing the protein complex and the excess of one of the proteins (usually the one with smaller size to ensure larger difference in retention times). The complex peak fractions were collected, controlled by SDS-PAGE, and concentrated to 50-150 mM for the HDX-MS experiments.

3.1.4 Protein storage

Purified proteins were stored in buffer solution (200 mM NaCl) at -20 °C. Thaw-freeze cycles were avoided, i.e. once the proteins were thawed for HDX-MS experiments they were consumed freshly and not frozen again.

3.2 Hydrogen/deuterium exchange reaction

Protein samples were diluted 1:20 with 200 mM NaCl D_2O solution, and incubated on ice for different times: 5 s, 30 s, 1 min, 2 min, 4 min, 8 min, 15 min, 30 min, 3 h, 5 h, 7 h, 24 h. The deuteration reaction was quenched by mixing 22.5 µL of the deuterated proteins solution with 22.5 µL of 6% acetic acid H_2O solution on ice. 5 µL of 40 µM pepsin in H_2O solution was added to digest the protein and the peptide mixture was injected into the nanoLC system for separation.

3.3 NanoLC-MS

The nanoLC workflow was optimized in the previous PhD work [5], as shown in Figure 8. Solvent A (0.5% formic acid H_2O solution) and solvent B (0.5% formic acid

in a mixed solvent of H_2O and acetonitrile (10 : 90)) were used in the LC gradient (Table 1). In order to minimize delays due to dead volumes in the pumps / split / nanoflow sensor and capillaries circuit, the LC gradient was started before the sample was prepared: thus two minutes before the pepsin digestion, the nanoLC run was started. The protein digest was injected at 6.5 min and started to be loaded onto the precolumn at 7 min, and after the 2 min wash it started to be eluted through the analytical column at 9 min.

The MS acquisition triggered by the LC system was started at 10 min, with detailed parameters shown in Table 2. Peptide signals began at around 15 min in the LC time scale (which means 13 minutes after digestion started) and ended at around 40 min. A 25 min wash with Solvent A ensured the elution of possible carry-overs and prepared the columns for the next run.

Data of both reference samples and deuterated samples were acquired in the same condition and processed as described in Part II.

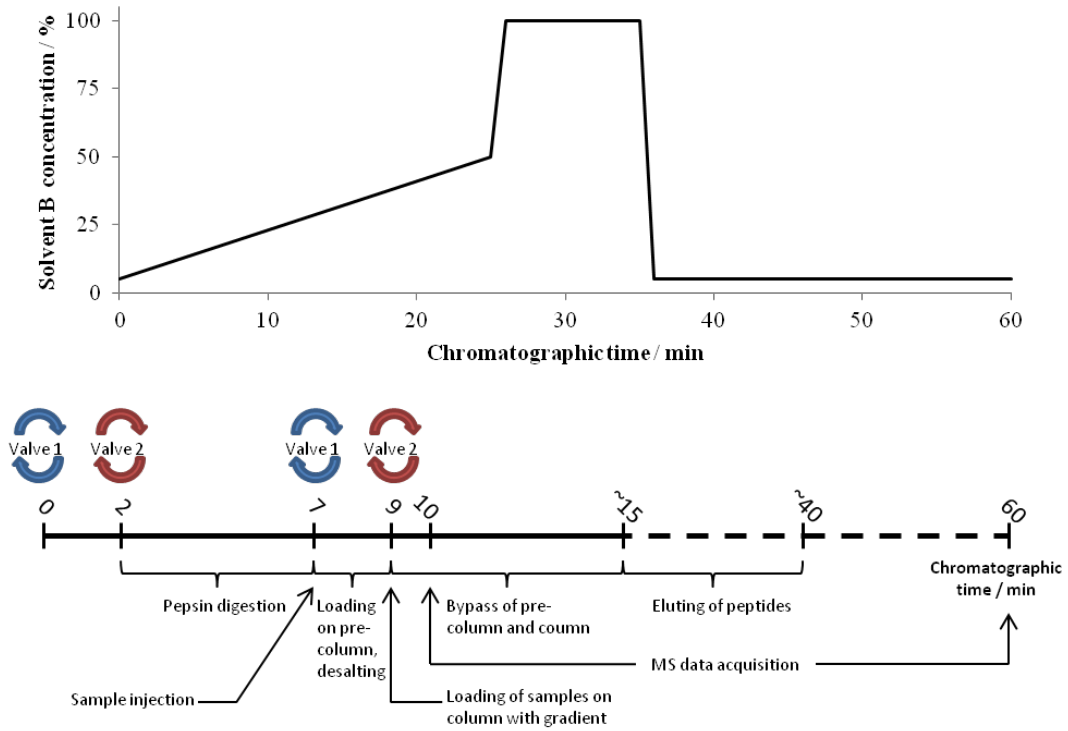


Figure 9 Time course of a whole nanoLC run, including the various stages of the experiment.

Table 2 NanoLC gradient in HDX-MS experiments.

Time / min	Solvent B / %	Time / min	Solvent B / %
0	5	35	100
25	50	36	5
26	100	60	5

Table 3 Parameters used in the acquisition by FT-ICR MS.

Ion source		Acquisition and accumulation	
Capillary	-850 V	Detection mode	Broadband
End plate offset	0 V	TD (Acquisition)	128 k
Dry gas	4.0 L/min	Mass range	498 – 2000
Drying gas temperature	120 °C	Estimated resolving power	41000 (at m/z 400)
		Average spectra	1
		Source accumulation	0.01 s
		Ion accumulation time	0.1 s
		Ion cooling time	0.01 s
		Time of flight	0.9 ms

3.4 LC-MS/MS

The LC-MS/MS experiments on the reference protein digests were carried out in a nano-HPLC-Orbitrap system. Proteins were diluted to 5 µM in 3% acetic acid, and 40 µL of the diluted protein samples were digested by 5 µL pepsin solution of 40 µM. 5 µL of the protein digests were injected to the HPLC system, separated and analyzed in the following condition:

The 2-hour LC gradient is shown in Table 3, with the solvent A composed of water: acetonitrile: formic acid (98: 2: 0.1%) and solvent B of water: acetonitrile: formic acid (20: 80: 0.08%) at a flow rate of 0.3 µL/min. Such a long gradient was required to achieve maximal coverage in terms of number of peptides, but of course it is not a gradient accessible for an HDX experiment.

The MS spectra were acquired with a m/z range of 300 – 2000. The top 20 peaks with signal intensity above 5000 in each MS spectrum were selected for MS/MS

fragmentation, using CID with isolation width of 2 Da and a normalized collision energy of 35 V.

Table 4 HPLC gradient in LC-MS/MS experiments.

Time / min	Solvent B / %	Time / min	Solvent B / %
0	2	109	99
3	2	110	2
98	55	120	2
99	99		

LC-MS/MS peak lists were extracted and submitted to a Mascot search with the following criterion: no specific enzyme, oxidation as variable modification, 10 ppm peptide tolerance, 0.5 Da fragment tolerance. The Mascot results were viewed using Scaffold Viewer, and the peptides with MS/MS of Mascot ion scores higher than 20 were considered valid.

References

1. Comisarow, M. B.; Marshall, A. G., Fourier transform ion cyclotron resonance spectroscopy. *Chem. Phys. Lett.* **1974,** *25*, 282-283.
2. Marshall, A. G., Milestones in Fourier transform ion cyclotron resonance mass spectrometry technique development. *Int. J. Mass spectrom.* **2000,** *200*, 331-356.
3. Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S., Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom. Rev.* **1998,** *17*, 1-35.
4. Marshall, A. G.; Hendrickson, C. L., Fourier transform ion cyclotron resonance detection: principles and experimental configurations. *Int. J. Mass spectrom.* **2002,** *215*, 59-75.
5. Duchateau, M. Détermination de surface d'interaction d'assemblages biologiques macromoléculaires par marquage au deutérium et analyse par spectrométrie de masse FT-ICR. Ecole Polytechnique, 2010.

Conclusion and perspectives

The initial objective of the thesis was to obtain HDX data on the eIF3 translation initiation complex. The long term objective was to reach a stage at which analysis of the entire eIF3 complex composed of 5 sub-units would be feasible. Of course this was an ambitious goal and the methodology was not yet available in the laboratory to reach such a goal. In the current work, progress was done on the binding regions between only three subunits eIF3b, eIF3g and eIF3i.

One of the challenges met in the course of this work is related to changes in instrumentation: not less than three FT-ICR hardwares were used in the course of this work. These changes in hardware have impeded data acquisition: it proved necessary to complete a whole series of measurements on the same version of mass spectrometer, since the changes in sensitivity and mass accuracy meant that the reference list and data analysis had to be adapted for each change in instrumentation. On the other hand, these hardware changes also helped in improving the data analysis steps: improved sensitivity meant a large increase in the number of peptides that could be analyzed, thus leading to improved sequence coverage with more overlapping peptides. This led to the proposal of a new peptide assignment scheme, based on mass accuracy and a probability of occurrence (POc) scoring system. These steps were absolutely necessary: as shown in Chapter 3, the increase in system size would have necessarily led to an increase in the total number of peptides and therefore an increase number of instances for which mass accuracy alone would not be sufficient for peptide assignment. This work has also shown that sensitivity leads to a huge number of monoisotopic peaks that are present in a single pepsin digest (over 1600 for the eIF3i:3bC3 complex) Of course, all these peaks do not belong to pepsin digest peptides, but still a fairly large number of peptides (300) can be identified within this digest.

As shown in Chapter 3, progress in instrumentation has also led to a large improvement in mass accuracy. Although FT-ICR mass spectrometry is regularly cited as a very accurate mass measuring equipment, there are some limitations, which are oftentimes overlooked. One such limitation is the space-charge effect, which occurs when the number of ions present in the cell reaches an upper limit. In the pepsin digests, a very large distribution of peak intensities was observed. Thus a compromise had to be done between keeping the ion accumulation time at a low level, and thus missing low intensity peptides but keeping high mass accuracy, or using a long accumulation time with better peptide coverage but lower mass accuracies. The internal calibration procedure used in this work could not cope with this problem, as the calibration ions (pepsin self-digest peptides) are not present in every

spectrum of the whole chromatogram. The most recent instrumentation (SolariX) has a new electronics system, which improved the management of space charge effects by increasing cell sensitivity. This led to a large improvement in mass accuracy over the whole chromatogram, and a narrow mass distribution is now achieved (see Chapter 3). Further improvements now require that the analysis software are able to make efficient use of this sub-ppm mass accuracy in processing HDX data: as stated in Chapter 3, the current data processing scheme is limited by the ($^{13}C - {}^{12}C$) vs (^{2}H-^{1}H) mass difference. The Marshall group have proposed a clever solution to this issue [1], but it will require adapting our current data processing software.

From a biological point of view, this thesis has helped progress in the deciphering of the interaction regions between eIF3i, eIF3g and eIF3b. As detailed above, parallel work [2] has achieved a crystal structure for eIF3i with a truncated form of eIF3b. Our work is well correlated with these results and can provide the basis for understanding the binding between eIF3i and eIF3g, for which very few is known. At the point of writing this manuscript, the data for the eIF3b:eIF3g:eIF3i complex are still under acquisition. Based on a few points, it is possible to extract the fact that most of the total structure is not strongly perturbed in the ternary complex, but hopefully we could reach is a complete picture within a few months. Based on our results on both the potential binding sites but also on the exchange kinetics which reflect the dynamics and structure of proteins in solution, our biologist collaborators are preparing new protein constructions, hoping to be able to obtain crystals that could lead to a crystal structure of this binding region.

References

1. Kazazic, S.; Zhang, H. M.; Schaub, T. M.; Emmett, M. R.; Hendrickson, C. L.; Blakney, G. T.; Marshall, A. G., Automated Data Reduction for Hydrogen/Deuterium Exchange Experiments, Enabled by High-Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *J Am Soc Mass Spectr* **2010,** *21*, 550-558.
2. Herrmannová, A.; Daujotytė, D.; Yang, J.-C.; Cuchalová, L.; Gorrec, F.; Wagner, S.; Dányi, I.; Lukavsky, P. J.; Valášek, L. S., Structural analysis of an eIF3 subcomplex reveals conserved interactions required for a stable and proper translation pre-initiation complex assembly. *Nucleic Acids Res* **2012,** *40*, 2294-2311.