Lehrstuhl für Organische Chemie II der
Technischen Universität München

Biochemical Methods for NMR Investigations of Large Proteins

Sylvain Tourel

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:		Univ.-Prof. Dr. Christian F. W. Becker
Prüfer der Dissertation:	1.	Priv.- Doz. Dr. Gerd Gemmecker
	2.	Univ.-Prof. Dr. Sevil Weinkauf

Die Dissertation wurde am 19.06.2008 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 22.09.2008 angenommen.

To my parents
and Matthias

« Ce n'est pas parce que les choses sont difficiles que nous n'osons pas, c'est parce que nous n'osons pas qu'elles sont difficiles. »

« Nicht weil es schwer ist, wagen wir es nicht, sondern weil wir es nicht wagen, ist es schwer. »

« It is not because things are difficult that we do not dare, it is because we do not dare that they are difficult. »

(Lucius Annaeus Seneca) Sénèque

Zusammenfassung

Die NMR-Spektroskopie stellt neben der Strukturbestimmung eine der wichtigsten Methoden zur Untersuchung der Dynamik von Proteinen dar. In dieser Arbeit werden verschiedene Methoden beschrieben, mit deren Hilfe auch Proteine oberhalb der momentan limitierenden Größe von ungefähr 100kDa untersucht werden können. Neben der selektiven Markierung von Proteinen, welche unter anderem zu einer geringeren Überlappung der Signale im Spektrum führt, wird ein neuer Ansatz für die Präparation von reversen Mizellen vorgestellt. Durch den Transfer von Proteinen in diese Art von Mizellen werden deren Relaxationseigenschaften drastisch verbessert. Eine Kombination dieser Techniken stellt eine vielversprechende Methode zur Untersuchung von großen Proteinen und Protein-Komplexen dar.

Abstract

The basic motivation of this work is to make measurements of large proteins accessible to nuclear magnetic resonance, allowing studies of large protein complexes at atomic resolution. Because nuclear magnetic resonance also allows measurements of protein dynamics, it is a powerful biophysical method and complementary to X-ray crystallography.

Our particular approach aims to extend the size limit of proteins suitable for NMR studies, which is currently estimated at 100kDa.

The results of this work describe an improved strategy for isotopic segmental labeling of proteins as signal filter for reducing signal overlap. It introduces a new approach of reverse micelles preparation via phase transfer and thus provides a powerful and widely applicable method for successful encapsulation of proteins in reverse micelles.

Practical examples of this approach demonstrate the feasibility of combining several biophysical and biochemical methods for successfully studying large proteins.

Acknowledgements

The work presented in this thesis was prepared from March 2005 until June 2008 under the supervision of PD Dr. Gerd Gemmecker and Prof. Dr. Horst Kessler at the Chemistry department of the Technical University of Munich.

I would like to thank Prof. Dr. Horst Kessler personally for, his trust, his interest and his financial support.

In addition I sincerely thank Dr. Gerd Gemmecker for his technical support and his motivation for the challenging projects.

I would like to thank Prof. Dr. Gerhard Wagner and Dr. Philipp Selenko from the Harvard Medical School for their full support and for the great working atmosphere.

I thank very particularly my colleague and friend Sandra Groscurth for the great technical discussions, team work, and for her full support during the good as well as the bad periods. This work would have not been possible without our close collaboration.

I am indebted to numerous people who have influenced this work in one way or the other (in alphabetical order): Dr. Hari Arthanari, Johannes Beck, Dr. Murray Coles, Lucas Doedens, Janine Eckardt, Katie Edmonds, Andreas Enthart, Alexander Frenzel, Dr. Dominique Frueh, Dr. Rainer Haessner, Timo Huber, Dr. Dmitri Ivanov, Peter Kaden, Maura Kilcommons, Jochen Klages, Dr. Tom Malia, Dr. Assen Marintchev, Dr. Monica Lopez, Dr. Monika Oberer, Florian Opperer, Dr. Ricard Rodriguez, Dr. Chikako Suzuki, Dr. Koh Takeuchi, Dr. Vincent Truffault and Mona Wolff.

Contents

Contents 1
Abbreviations 1
General Introduction 1
Chapter 1: *In-vivo* intein segmental labeling 1
1. Summary 1
2. Introduction 2
3. Materials and methods 4
3.1. Plasmid construction 4
3.2. Expression and purification of segmentally labeled HSP90 6
3.3. Expression and purification of segmentally labeled eIF4A 8
4. Results and discussions 9
5. Conclusions 30
6. Litterature 31
Chapter 2: Thermal shift assay 1
7. Summary 1
8. Introduction 2
9. Theory 3
10. Material and methods 4
10.1. Method implementation 7
11. Results: unfolding/refolding study of GB1 9
12. Application to natively unfolded protein A-synuclein 18
13. Conclusion and future developments 23
14. References 24
Chapter 3: Reverses micelles 27
15. Summary 27
16. Introduction 27
17. Material and methods 33

17.1. Proteins 33
17.2. Surfactant 34
17.3. Solvent 35
17.4. CD spectroscopy 36
17.5. Recovery of partially unfolded transferred cCytochrome C. 42
17.6. Efficient extraction of hemoglobin 45
18. Conclusion and future developments 50
19. Literature 51
Final Conclusion 56
Appendix 58

Abbreviations

ADP	Adenosine diphosphate
Amp	Ampicillin
AOT	Aerosol OT, sodium bis(2-ethylhexyl) sulfosuccinate
Ara	Arabinose
ATP	Adenosine triphosphate
BL21 (DE3)	*E. coli* protein expression strain
$C_{12}E_4$	Dodecyl tetraethylene glycol ether
Carb	Carbenicillin
CD	Circular dichroism
CTAB	Cetyltrimethylammonium bromide
CTD	C-terminal domain
dd	Double distillated
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylene diamine tetraacetic acid
eIF4A	Eukaryotic initiation factor 4A
eIF5E	Eukaryotic initiation factor 5E
GB1	Protein G, binding domain 1
GdmCl	Guanidinine chloride
HSP90	Heat shock protein 90
HSQC	Heteronuclear single quantum coherence
IPTG	Isopropyl • -D-thiogalactoside
Kan	Kanamycin
LB	Lucia – Bertani broth
M9	Minimal medium for protein labeling
MD	Middle domain
M_r	Molecular weight
MWCO	Molecular Weight Cut Off
NAC	Non-Abeta component
Ni-NTA	Immobilized nickel affinity chromatography material
NMR	Nuclear magnetic resonance
Npu DnaE	Intein fragments from *Nostoc punctiforme*
NTD	N-terminal domain

OD_{600}	optical density of liquid medium at 600 nm
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PD	Parkinson Disease
PDB	Protein Data Bank
pI	isolelectric point
pSKBAD2	Plasmid containing GB1 fragment for ligation with Npu controlled via AraBAD promoter and ampecillin resistance gene
pSKDuet1	Plasmid containing GB1 fragment for ligation with Npu controlled via T7 promoter and Kanamycin ressitance gene
pSTDuet51	Plasmid containing HSP 90 NTD fragment for ligation with Npu controlled via T7 promoter and Kanamycin resitance gene
Q-PCR	Quantitative Polymerase Chain Reaction
RF	restriction free
RM	Reverse Micelles
RNA	Ribonucleic acid
SDS	Sodium Dodecyl Sulfate
Ssp DnaE	Intein fragments from *Synechocystis sp.*
TEV	Tobacco Etch Virus
T_m	Melting temperature
TRIS	tris(hydroxymethyl)aminomethane
TSA	Thermal shift assay
UV	Ultra Violet
wt/vol	weight per volume

General Introduction

Modern methods in nuclear magnetic resonance (NMR) spectroscopy provide a wealth of information on the structural and kinetic properties of different biological molecules. Unfortunately, multidomain proteins have long been excluded from modern NMR studies due to their large sizes and because of the resulting spectral complexity with high degrees of resonance overlap. For these reasons NMR spectroscopy has largely focused on isolated protein domains in the past. While those analyses have provided strong evidence that many protein domains exhibit functional and structural characteristics that are indiscernible from their full-length protein propensities, information about inter-domain orientations or about the interactions between individual domains could not be obtained experimentally. Here we present in the first chapter a protocol that allows to selective isotope-labeling of individual protein domains in multi-domain proteins, for the purpose of analyzing them with modern NMR methods. The described approach enables single protein domains to be studied in full-length protein contexts. While advancements in NMR methodology, most notably the introduction of transverse relaxation-optimized spectroscopy (TROSY) methods[1], have greatly extended the size limit of proteins suitable for NMR analyses [2,3], the presented method of domain-selective isotope labeling reduces the spectral complexity of large multi-domain proteins to single domain polypeptides.

Nevertheless high resolution NMR of large proteins suffers from adverse relaxation properties due to their slow tumbling in solution.

Reverse micelles (RMs) are aggregates in which nanoscale droplets of a polar liquid, usually water, are surrounded by a surfactant layer in a nonpolar continuous phase. Nonpolar continuous phases such as alkanes provide low viscosity to

reverse micelles, directly influencing their overall tumbling and therefore the protein ‘s NMR properties. They are widely used as media for reactions or protein encapsulation, in which the extent of confinement or the presence of a surfactant interface play a central role to the protein structure. Our studies on reverse micelle preparation have focused on the effects of protein extraction via phase transfer, which are determined by charged interactions. Specifically, we have examined the effects of water on the secondary structure and a strategy to influence the water amount within the reverse micelles.

Biophysical characterization of proteins is a key factor to understanding biology at a chemical level. Understanding protein folding/unfolding improves our knowledge of protein stability, protein aggregation or fibril formation. Therefore factors influencing protein folding are very important, and high throughput screening assays can facilitate corresponding studies. The thermal shift assay takes advantage of an environmentally sensitive fluorescence dye, such as Sypro Orange®, and follows its signal changes while the protein undergoes thermal unfolding. The assay is widely used for screening compound libraries for ligands of targets proteins. The ligand-binding affinity of any potential inhibitor can be assessed from the shift of the unfolding temperature (T_m) obtained in the presence vs. absence of the potential inhibitor.
The assay can also be used to detect und study essential conditions for protein folding unstability. We have applied this strategy to determine buffer conditions which destabilize GB1 allowing unfolding/refolding NMR experiments via thermal denaturation in a temperature range compatible with NMR experiments.

In brief, this work provides improvements and development of biophysical and biochemical protocols, facilitating NMR experiments for large proteins as well as protein folding.

[1]. Wider, G. & Wüthrich, K. NMR spectroscopy of large molecules and multimolecular assemblies in solution. Curr. Opin. Struct. Biol., 9, 594-601 (1999)

[2]. Yu, H. Extending the size limit of protein nuclear magnetic resonance. Proc. Natl. Acad. Sci. USA 96, 332-334.

[3]. Iwai H, Zuger S, Jin J, Tam PH. Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiforme. FEBS Lett. 2006 Mar 20;580(7),1853-8.

Chapter 1:

In-vivo intein segmental labeling: development and application to heat shock protein 90 and eukaryotic initiation factor

1. Summary

Nuclear magnetic resonance (NMR) provides, contrary to X-ray crystallography, more information on the kinetics, interactions and conformational state of polypeptide chains. The interdomain structural reorganisation caused by protein-protein or protein-ligand binding events of large, multidomain proteins is essential for the understanding of biological signal transduction pathways, as well as for drug design strategies intervening with those.

Since isolated domains of proteins tend to be structurally identical to the original multidomain protein, NMR structural studies are usually focused on single domains with a reasonable size for NMR. Nevertheless studies of full length proteins would have the benefit of giving insight into the structural interactions between domains, which are often the molecular basis for regulation and signal transduction.

In order to make such systems accessible for NMR, we have adapted and optimized an intein ligation method based on the recent *in vivo* ligation protocols developed by Hideo Iwai[1][2].

2. Introduction

Molecular biology has focused on understanding protein functions and interactions on a structural level. For this purpose scientists have developed and improved biophysical methods, allowing the observation of events and rearrangements. Such techniques as NMR, infrared spectroscopy, fluorescence spectroscopy, electrophysiology or crystallography have pushed molecular biology to a field of excellence.

NMR is strongly established in the field of life science. This success is certainly the result of major and constant developments in sampling strategies, acquisition and processing methods in the last 20 years [1] [2]. Still NMR has not reached yet its full capacity and method developments are a key of the NMR success.

With the introduction of isotopic labeling in the mid 90s, the sensitivity of NMR has increased drastically, making nitrogen and carbon spins available for protein NMR studies. But NMR has maintained one of its main difficulties: the protein size. Large proteins have disadvantageous physical properties, which have a strong impact on the measurement results. Therefore NMR research groups have mainly focused on single domain proteins and structures published and stored in the PDB database rarely excess the 30 kDa (or even 20 kDa) limit. Characterizations of biological pathways were rarely studied using NMR, except for reasonably small full length proteins.

The recent development of TROSY (transverse relaxation-optimized spectroscopy) [1], RDC (residual dipolar couplings), and new labeling strategies have presumably shifted this limit of protein structure determination up to 100 kDa. This is naturally a theoretical value, while there are practically no examples in PDB.

Polypeptide ligation is one of this recent labeling strategies with very interesting potential for the understanding of protein-protein interactions on a structural level in a full length protein. Such methods are based on reducing the number of signals,

and hence signal overlap in NMR spectra, which greatly simplifies the structural analysis of large proteins.

In vitro ligation for segmental isotopic labeling has been tested and established, but it has limited application since it requires laborious preparation and optimization of the ligation conditions.

In contrast, the *in vivo* approach has shown promising results and has been used successfully for different applications. The research group of Hideo Iwai has recently presented a novel *in vivo* ligation protocol with very promising results [3]. This method is based on sequential expression of the two precursor fragments, coded in two different expression systems. The ligation relies on the trans-protein splicing activity of split inteins. The Iwai group was able to apply this technique using the split intein DnaE from *Synechocystis sp.* strain PCC6803 (Ssp Dna E) and from *Nostoc punctiforme* (Npu DnaE) [3].

The work presented here has focused on application and optimization of this promising approach to achieve protein overexpression yields suitable for NMR. We have adapted and combined several methods from cloning to expression protocols to develop a fast and simple strategy accessible to a large number of biochemistry groups.

3. Materials and methods

3.1. Plasmid construction.

The plasmids pSKDuet1 and pSKBAD2, containing the gene of the split DnaE intein from *Nostoc punctiforme* (Npu intein), were provided from Hideo Iwai [3].
The DNA construct for the 220 first amino acids of the N-terminal portion of HSP90 (a gift from Prof. Buchner, Technical University of Munich) was amplified using the oligonucleotides:
Nco_his_TEV_Hsp90_NTD_1
5'CGCCTCGCCATGGGGCATCACCATCACCATCACGAAAACCTGTATTTTCAG
GGCATGGCTAGTGAAACTTTTGAATTTC-3'
and 220_Hsp90_NTD_BamH1
5'-GGTACCTGGATCCGGAATTGGAACTTCCTTTTCAA-3'.

The gene was ligated using the Nco and BamH1 sites into pSKDuet1, which allows the expression of the resulting HSP90 NTD-InteinN fusion protein with an N-terminal hexahistidine tag, cleavable with a TEV site. The resulting plasmid pSTDuet51 contains a T7/lac promoter and the fusion gene, of which expression can be induced by isopropyl •-D-thiogalactoside (IPTG). It also contains the kanamycin resistance gene (Kan^R), the RSK origin of replication (RSF ori) and the lacI gene (Figure 1).

The residues 229 to 709 of HSP90 (MD-CTD) were amplified using the oligonucleotides:
Kpn1_NN_Hsp90_MTD_229
5'-CCGCCTCGGGTACCACCACCAAGAAGGATGAGGAAAAGAAGGA-3'
And 709_HSP90_CTD_st_Hind
5'-GGTACCTAAGCTTCTAATCTACCTCTTCCATTTCGGTG-3'.

The plasmid pSTBad3 was constructed by introducing the gene of HSP90 MD-CTD with the Kpn1 and HindIII sites into the expression vector pSKBAD2, containing the ColE1 origin of replication (ColE1 ori), the phage M13 origin of replication, the gene araC (araC) and the ampicillin resistance gene (Amp^R) (Figure 1).

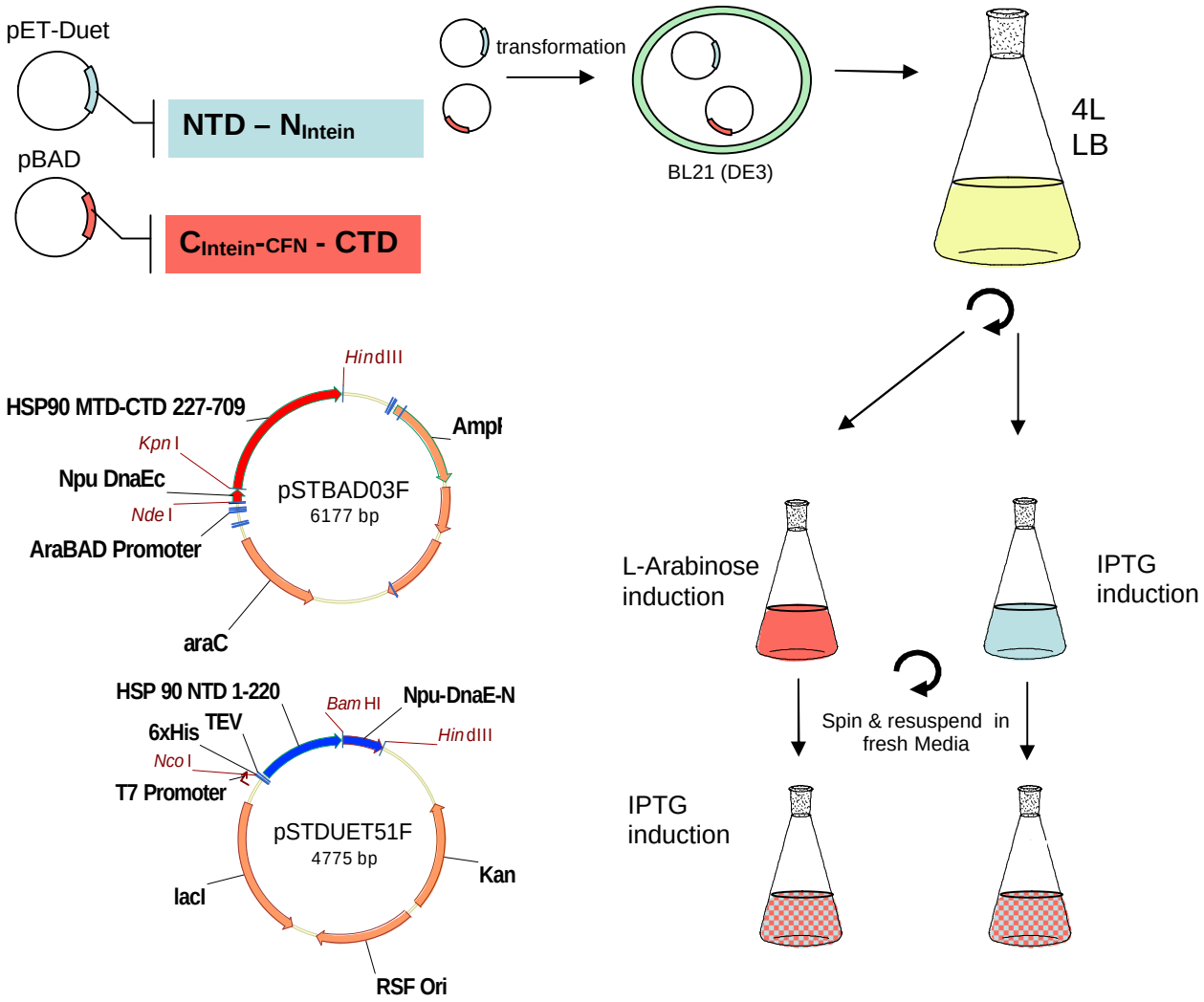


Figure 1: *In-vivo* segmental labeling protocol. Plasmid map of HSP90 Fragments

For the cloning of the genes coding for the N-terminal domain of the elongation initiation factor 4A (eIF4A-NTD) and the C-terminal domain (eIF4A-CTD), resp., a novel cloning method known under “restriction free cloning technique” (RF cloning) [5] was used . This technique allows the cloning of any PCR amplified gene into a target gene without the help of restriction enzymes based cloning. This technique relies principally on the QuickChange™ PCR (Stratagene) of a previously amplified gene instead of oligos.

Using this method the eIF4A-NTD (residue 1 to 237) was amplified from the gene coding for the full length eIF4A (gift from Gerhard Wagner) using the primers #ST007 (5’-CCTGTATTTTCAGGGCATGTCTGCGAGCCAGG-3’) and #ST008 (5’- CCCCATTCGGATTCTTGTCAAGTGTTTAAGC-3’) and cloned into pSTDuet1 between the TEV cleavable His-tag and the Npu inteinN without any additional nucleotides. The resulting plasmid pSTDuet50 allows the expression of eIF4A NTD fused with Npu inteinN at the C-terminus.

The C-terminal domain (eIF4A-CTD) was amplified using the oligonucleotides #ST006 (5’-GCTATCTGGTACCCTGGAGGGTATCCGCCAG-3’) and 4ac-st_hind (5’-GGTACCTAAGCTTAGATGAGGTCAGCAACATTGAG-3’). The PCR product was finally introduced into the pSKBAD2 using Kpn1 and HindIII sites to form pSTBAD44, which allows the expression of the eIF4a-CDT fused with the split Npu inteinC at its N-terminus.

3.2. Expression and purification of segmentally labeled HSP90

E. coli BL21 (DE3) (Invitrogen) was transformed by the plasmid pSTDuet51 and pSTBad03 for the protein expression. The cells were inoculated with 2% overnight culture at 37 ºC in 4L Lucia – Bertani broth (LB) supplemented with 50µg/ml carbenicillin and 50µg/ml kanamycin to an OD_{600} of 0.6-0.8. The cells were spun down for 20 min 900g and resuspended in 1L 40% deuterated M9 medium with

$^{14}NH_4Cl$ (2g/l) and D-Glucose (2g/l) as the sole nitrogen and carbon sources, supplemented with 50µg/ml carbenicillin and 50µg/ml kanamycin. The culture was shaken for 1 hour to adjust the cells to the new environment, then the previous step was repeated 2 times with M9 containing 70% D_2O (^{15}NH4Cl (2g/l), D-Glucose (2g/l)) and 99% D_2O ($^{15}NH_4Cl$ (2g/l), [^{2}D]D-Glucose (2g/l)), resp.. The cell density was checked to make sure that the cells were still growing and the N-terminal fragment was induced with a final concentration of 1mM IPTG for 4 hours.

Subsequently, the cells were spun down for 20 min at 900g and resuspended in 2L LB medium. The culture was shaken for 1 hour at 30 ºC before the second induction to give time for the cells to recover the precedent step. The expression of the C-terminal fragment was induced with a 0.2% (wt/vol) L-arabinose for 15h. The cells were harvested and frozen at -80 ºC.

The cells were resuspended in a lysis buffer (50mM TRIS, 300mM NaCl, 10mM imidazole, pH 8.0) and lysed using sonication on ice. The lysate was loaded on a Ni-NTA column (Qiagen) and washed with a washing buffer (50mM TRIS, 300mM NaCl, 30mM imidazole pH 8.0). The hexahistidine-tagged proteins were eluted with 250mM imidazole and the eluted fraction was then loaded on a column loaded with Sephadex G-75 (Amersham Biosciences), previously equilibrated with 40mM Hepes, 300mM KCl, pH=7.5. The ligated product was loaded again to a Ni-NTA column to get off residual unwished proteins and eluted with 250mM imidazol. The eluate was desalted using an IgG Sepharose 6 Fast Flow (Amersham Biosciences) again 40mM Hepes, 150mM NaCl, pH 7,5 and concentrated for NMR measurements.

3.3.Expression and purification of segmentally labeled eIF4A

E. coli BL21 STAR (DE3) (Invitrogen) was transformed by the plasmids pSTDuet50 and pSTBad44 for the protein expression. The cells were inoculated at 37 ºC with 2% overnight culture in 4L Lucia – Bertani broth (LB) supplemented with 50µg/ml carbenicillin and 50µg/ml kanamycin to an OD_{600} of 0.6-0.8. The cells were spun down for 20 min 900g and resuspended in 1L 95% deuterated M9 medium with 99% D_2O, $^{15}NH_4Cl$ (2g/l), D-glucose (2g/l) and ^{15}N Celton-N (2g/l) (Spectra stables isotopes) as the sole nitrogen and carbon sources, supplemented with 50µg/ml carbenicillin and 50µg/ml kanamycin. The culture was shaken for 1 hour to adjust the cells to the new environment. The cell density was checked to make sure that the cells were still growing and the N-terminal fragment was induced with a final concentration of 1mM IPTG for 4 hours.

Subsequently, the cells were spun down for 20 min at 900g and resuspended in 2L LB medium. The culture was shaken for 1 hour at 37 ºC before the second induction to give time for the cells to recover from the precedent step. The expression of the C-terminal fragment was induced with a 0.2% (wt/vol) L-arabinose for 15h. The cells were harvested and frozen at -80 ºC.

The cells were resuspended in a lysis buffer (50mM TRIS, 300mM NaCl, 10mM imidazole, pH 8.0) and lysed using sonication on ice. The lysate was loaded on a Ni-NTA column (Qiagen) and washed with a washing buffer (50mM TRIS, 300mM NaCl, 30mM imidazole pH 8.0). The hexahistidine- tagged proteins were eluted with 250mM imidaizole and were desalted using an IgG Sepharose 6 Fast Flow (Amersham Biosciences) again 75mM NaCl, 2mM DTT, 1mM EDTA, pH=7.2. The collected fraction was loaded on an Ressource Q ion exchange column (Pharmacia Biotech), previously equilibrates with the previous buffer. The protein was eluted by increasing NaCl concentration stepwise. The eluate fraction was desalted using an IgG Sepharose 6 Fast Flow (Amersham Biosciences)

against 20mM TRIS, 150mM NaCl, 2mM DTT, 4mM $MgCl_2$ pH 7,5 and concentrated for NMR measurements.

4. Results and discussions

Intein mediated polypeptide ligation is commonly presented as a attractive tool in biochemistry, especially for NMR, also if intein is usually associated as an ineffective method, difficult to set up and time consuming. The complicated approach of the *in vitro* method is certainly responsible for this, but native split inteins engineered *in vivo* behave much more efficiently than *in vitro* and react quite independently from their environment. Thus polypeptide ligation becomes accessible to a large number of biochemistry laboratories. Nevertheless, this tactic is often criticized, since native split inteins introduce or mutate residues of the target protein sequence, because specific sequences are required for the ligation.

For example, the native split intein from DnaE from Synechocystis species strain PCC6803 (Ssp DnaE) needs the sequence CFNK in the C-extein and interleaves this sequence after trans-protein splicing reaction (only CFN in case of DnaE from Nostoc punctiforme PCC 73102 [3]). It has been shown that some single mutations of these sequences can be accomodated without complete loss of trans-splicing, but sometimes with a decrease of efficiency [3][4]. In regular cloning strategies, as employed previously in the plasmids designed by Iwai et al. [3][4], the gene of the target fragment and the split intein are cloned using 3 restriction enzyme sites. The central cloning site inserts a mutation of two amino acids in case of a 6 bp recognition site, like the BamH1 site (ggatcc) which codes after transcription for the amino acids GS. Therefore the finale ligated protein contains 7 residues that differ from the original sequence: 4 amino acids for the restrictions sites and 3 for the intein interleaved residues. These additional mutations are often undesirable and this problem is solved by additional cloning steps based on QuickChange™ (Stratagene) to delete or mutate the residues. Every additional

step makes cloning (for intein purposes) more time and cost consuming. Therefore we were looking for simple cloning strategies which prevent additional mutations due to the restriction site used for cloning. Restriction-free (RF) cloning is described by the authors as ”a simple, universal method to precisely insert a DNA fragment into any desired location within a circular plasmid, independent of restriction sites, ligation, or alterations in either the vector or the gene of interest” [5]. The technique is very similar to QuickChange™ (Stratagene) but instead of introducing small insertions/deletions or mutations of few bases, it inserts a complete gene without additional unwanted extra residues. We have applied RF cloning successfully to several samples and it has strongly simplified our cloning strategy and shown to be a very powerful general way of cloning, especially for intein purposes.

In our hands, the *in-vivo* native intein ligation method has shown simple and powerful capacities and we have employed this technique on several protein fragments with success.

Escherichia coli expression stems were double transformed (Figure 1), with two plasmids containing two different promoters, antibiotic resistances, and origins of replication. This configuration allows induction control of one of the target genes in an independent matter using the T7 and AraBAD promoters.

The expression procedure is described in the material and methods section and follows a typical protocol (Figure 1). The independent promoters allow to use two pathways where each promoter can be induced first. Individual fragments were expressed first to control the proper expression and as references. Then each promoter was induced sequentially (Figure 2 A and B) and in parallel (Figure 2 C) to figure out if the induction order during the expression affects the ligation. The first induction using L-arabinose of the C-terminal domains of the heat shock protein 90 (MTD-CTD HSP 90) fused with Npu DnaE seems to repress the expression of the N-terminal of HSP90 (NTD HSP90) induced with IPTG (Fig. 2A).

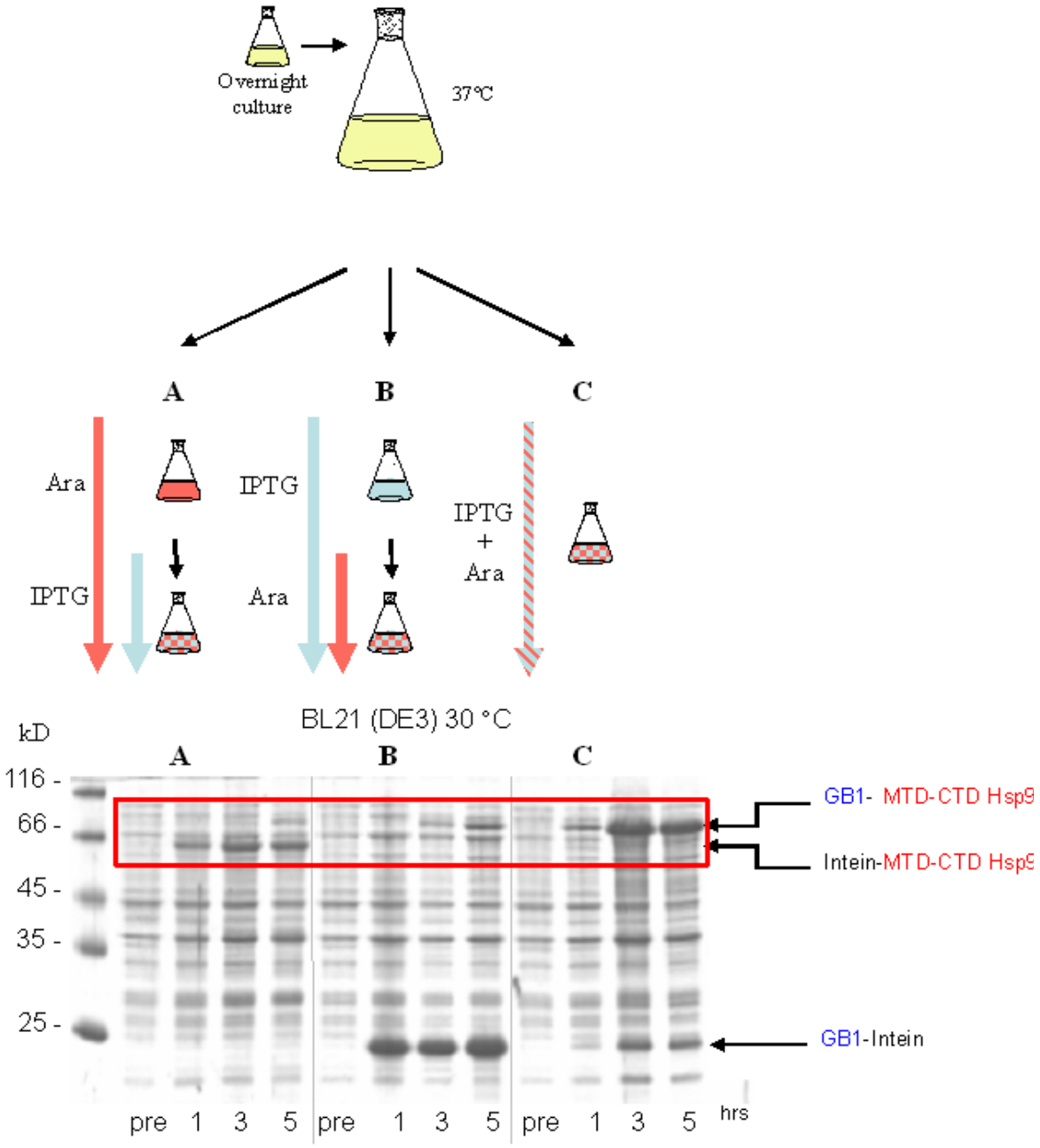


Figure 2: Studies on fragment stability expression time and temperature. The SDS loaded samples correspond to the soluble fractions. Light blue corresponds to the N-terminal fragment to be ligated. Red corresponds to the C-terminal fragment to be ligated.

The NTD HSP90 is very poorly expressed (not detectable using Coomassie blue) and not consumed for ligation since no full-length protein could be observed at the expected size (Fig. 2A). To verify this result, we have proceeded to a Ni-NTA purication, followed by a Western plot (data not shown), which has confirmed that no ligated product was formed and only very little amount of NTD HSP90 was expressed. In contrast, when the NTD HSP90 was induced first ((Fig. 2B), the induction of the AraBad promoter didn't affect the overexpression of the MTD-CTD HSP90, and the two protein fragments could ligate.

In the case of the elongation initiation factor 4A (eIF4a) we were not able to detect ligated products even when both plasmids were induced simultaneously (Figure 3) in the regular BL21 (DE3) strain. After scrambling expression conditions without success, we tried alternative expression stems available on the market. The BL21 (DE3) STAR seems to solve this problem and both domains could be expressed and ligated properly and effectively. This strain carries a mutated *rne* gene (*rne*131) which encodes a truncated RNase E enzyme that lacks the ability to degrade mRNA, resulting in an increase in mRNA stability.

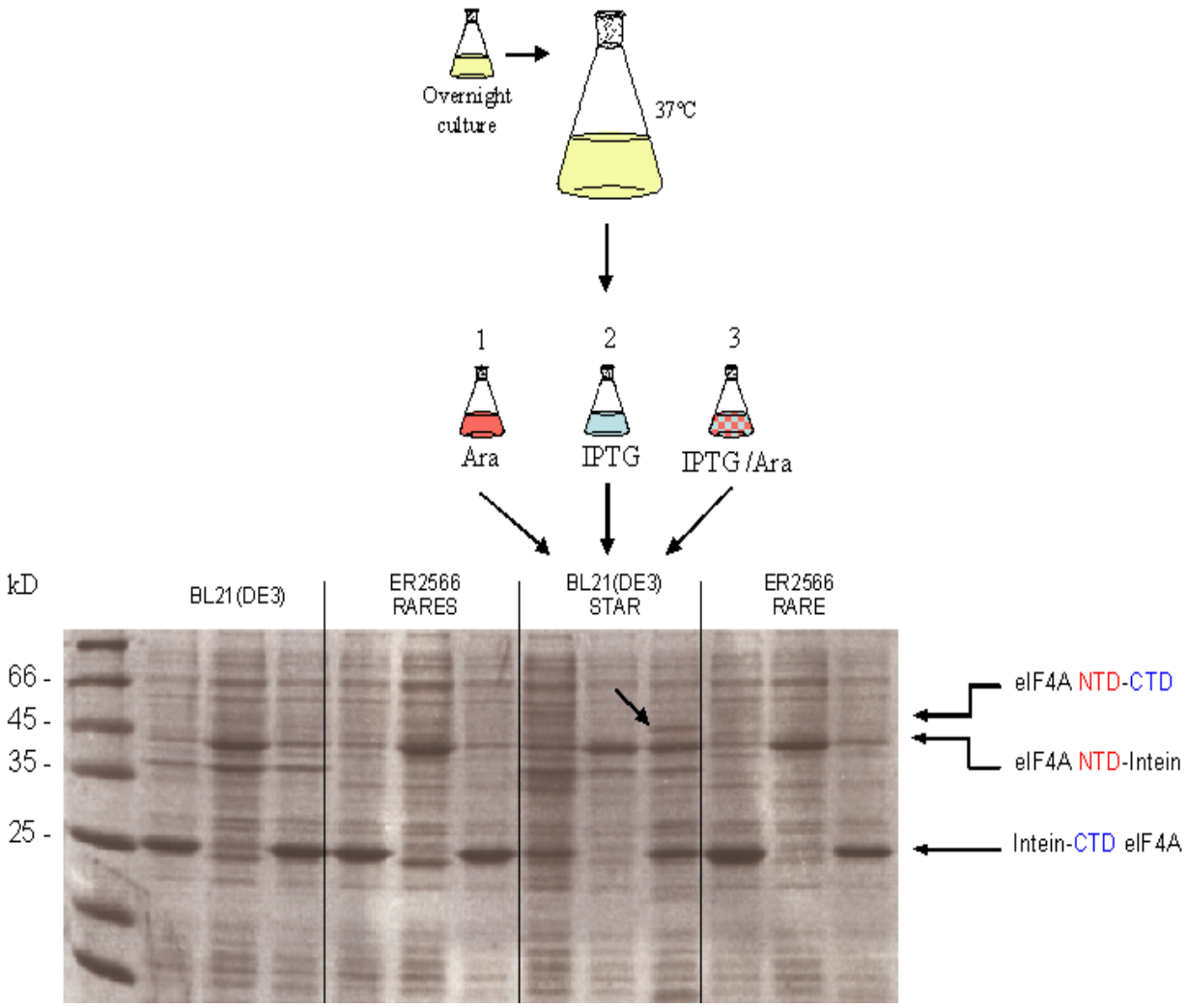


Figure 3: Strain ligation dependence experiments. The SDS loaded samples correspond to the soluble fractions. Light blue corresponds to the N-terminal fragment to be ligated. Red corresponds to the C-terminal fragment to be ligated.

Theoretically, the induction pathway should not be relevant for the ligation efficiency. Nevertheless, in our hands we have observed that induction of the T7 promoter first improves the reaction yield. The vector controlled via the T7 promoter contains the RSF origin of replication, in a low replication number. In contrast the vector under the araBad promoter, which uses the colE1 origin of replication, has a high replication number. This difference affects the amount of each plasmid in the cell and it might explain why the fragment under the T7 promoter doesn't express well after a previous induction of the AraBad promoter. After induction of the AraBad promoter, the amount of mRNA after the DNA transcription is high due to the high copy number of the plasmid. When the amount of transcribed RNA reaches a critical concentration, the cell's RNA degradation pathway is activated to protect the cell by expressing RNAse. The expression is not necessarily affected, since the number of DNA templates available is high. Nevertheless, when the second plasmid (under T7 promoter) is induced the amount of transcribed mRNA is lower compared to the previous plasmid, and the RNAse, previously expressed, degrades rapidly the new mRNA, which reduces the overall expression of the second fragment, especially if the mRNA of the target gene is not highly stable. As mentioned before the BL21 (DE3) STAR strain lacks the ability to degrade mRNA by RNAse E, this might explain the expression and ligation improvements observed in BL21 (DE3) Star.

Since each fragment is expressed sequentially, the solubility of the first expressed fragment and its stability over the whole expression process is crucial. Figure 4 demonstrates the difference of protein stability over time at different temperatures. This aspect is a key factor for the success of this method. Therefore we recommend analyzing protein solubility and stability over time and temperature for each fragment and for the full length protein, to determine optimal expression conditions.

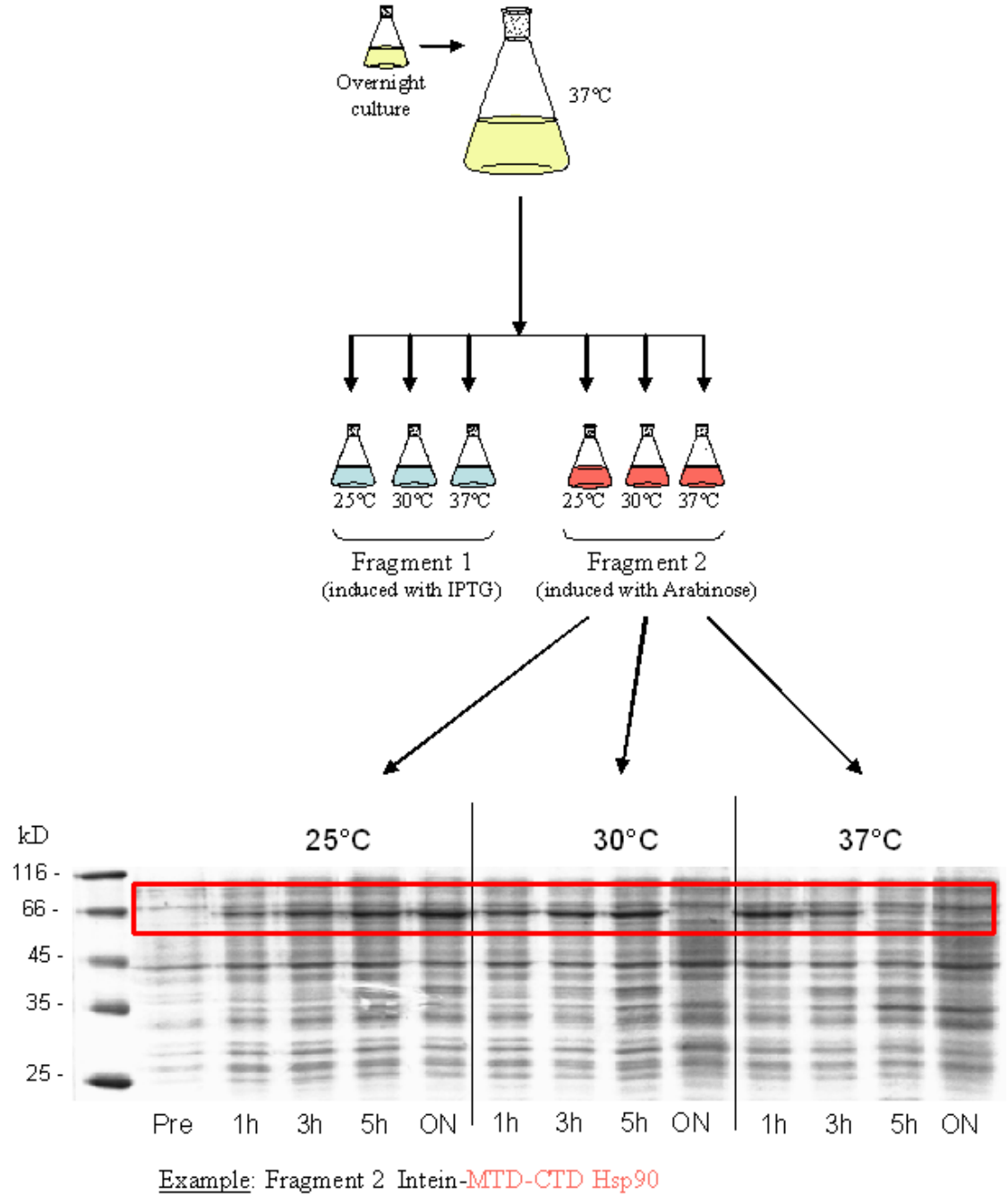


Figure 4: Studies on fragment stability expression time and temperature. The SDS loaded samples correspond to the soluble fractions. Light blue corresponds to the N-terminal fragment to be ligated. Red corresponds to the C-terminal fragment to be ligated.

In general, with the standard protocol for protein expression the amount of protein expressed in minimal medium is only about 30% of the amount expressed in full medium. To achieve high yields of protein expression, the conditions and especially the composition of full as well as of minimal medium have to be optimized before isotopic labeling. To get nearly the same yield of protein in minimal medium as in full medium, Marley et al. have introduced an extremely useful protocol [7].

A large amount of cell mass is generated in full medium and then used to inoculate minimal medium at high cell densities. Basically the cells grow rapidly in full medium until they have reached a sufficient OD and the cell mass is resuspended in minimal medium, allowing high cell densities with similar log phase behaviour. The labeling cost is significantly reduced and the isotopic incorporation is >90% for ^{15}N and >95% for ^{13}C [7]. To minimize the incorporation of unlabeled material available, the cells were incubated 1 hour in minimal medium before induction with L-arabinose as well as with IPTG. This additional step also allows the cells to recover and to adapt their metabolism to minimal medium.

This technique has been applied for *in-vivo* selective labeling of proteins and the expression yield of pure ligated protein increased about 3-fold for eIF 4A (up to 35mg/L) and 2-fold for HSP 90 (up to 55 mg/L) in full medium. The cell mass was generated in 4 L of full medium and resuspended in 1 L of minimal medium supplemented with 1g Celtone algae extract (Spectra Isotopes) for adaptation and higher yield. In perdeuterated minimal medium, expression yields of 35mg/L and 18 mg/L of HSP 90 and eIF4A, resp., were achieved. This amount is suitable for NMR purposes with good cost efficiency for perdeuterated samples. The final cell density after complete expression reaches an OD_{600} of 7 to 10.

Using heteronuclear single-quantum coherence (HSQC) spectra we have investigated HSP90 samples of which only the N-terminal fragments were

expressed in minimal medium and ligated at the C-terminus with its middle and C-terminal domain (Fig. 5), reconstituting the full length protein (84kDa). Since the C-terminal domain of HSP90 is responsible for the homodimerization of HSP90, the final size of the protein is 170kDa. The size of the protein affects strongly the overall tumbling of the protein, influencing the relaxation properties, making it very unfavorable for NMR experiments. By increasing the temperature to 37°C, we were able to improve significantly the relaxation properties and therefore the HSQC quality (Figure 5), demonstrating the feasibility of segmental labeling for large proteins. Unfortunately the quality obtained is not yet sufficient for structural biology experiments.

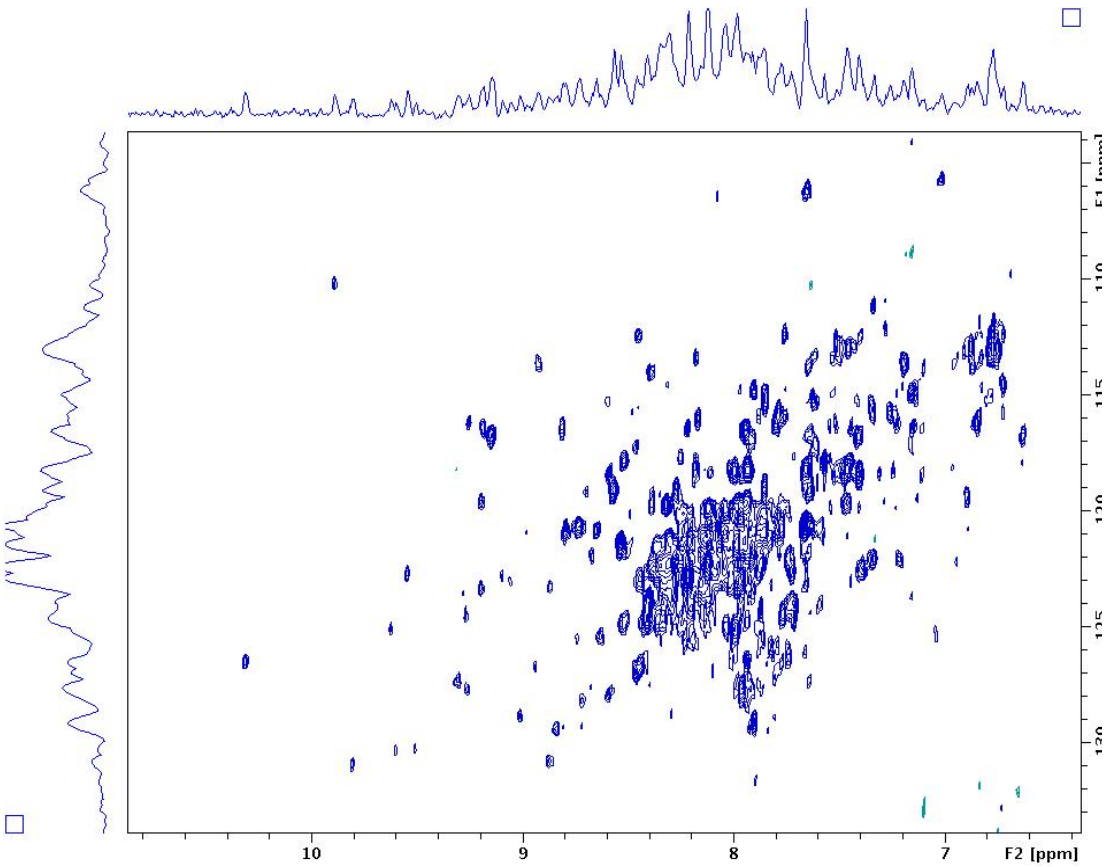


Figure 5: Characterization of segmental labeling by NMR spectroscopy. HSQC spectrum of HSP90 170kDa segment of (^{15}N,^{2}D) labeled N-terminal domain, while the rest of the protein was unlabeled (^{14}N,^{1}H). The experiment was carried out in PBS at 37°C at a protein concentration of 0.2mM using a BRUKER 900MHz spectrometer equipped with a cryoprobe.

The proteineIF4A (46 kDa) is a helicase involved in unwinding RNA secondary structure during translation initiation. It is a two-domain protein which requires both domains to maintain complete helicase activity [7]. Each domain contains several conserved motifs, involved in ATP-binding/hydrolysis as well RNA binding (Figure 6). The domains are connected together via a 4 to 6 residues linker and this linker is critical for the helicase activity. EIF4A is a potential target for elongation initiation inhibition in cancer therapies.

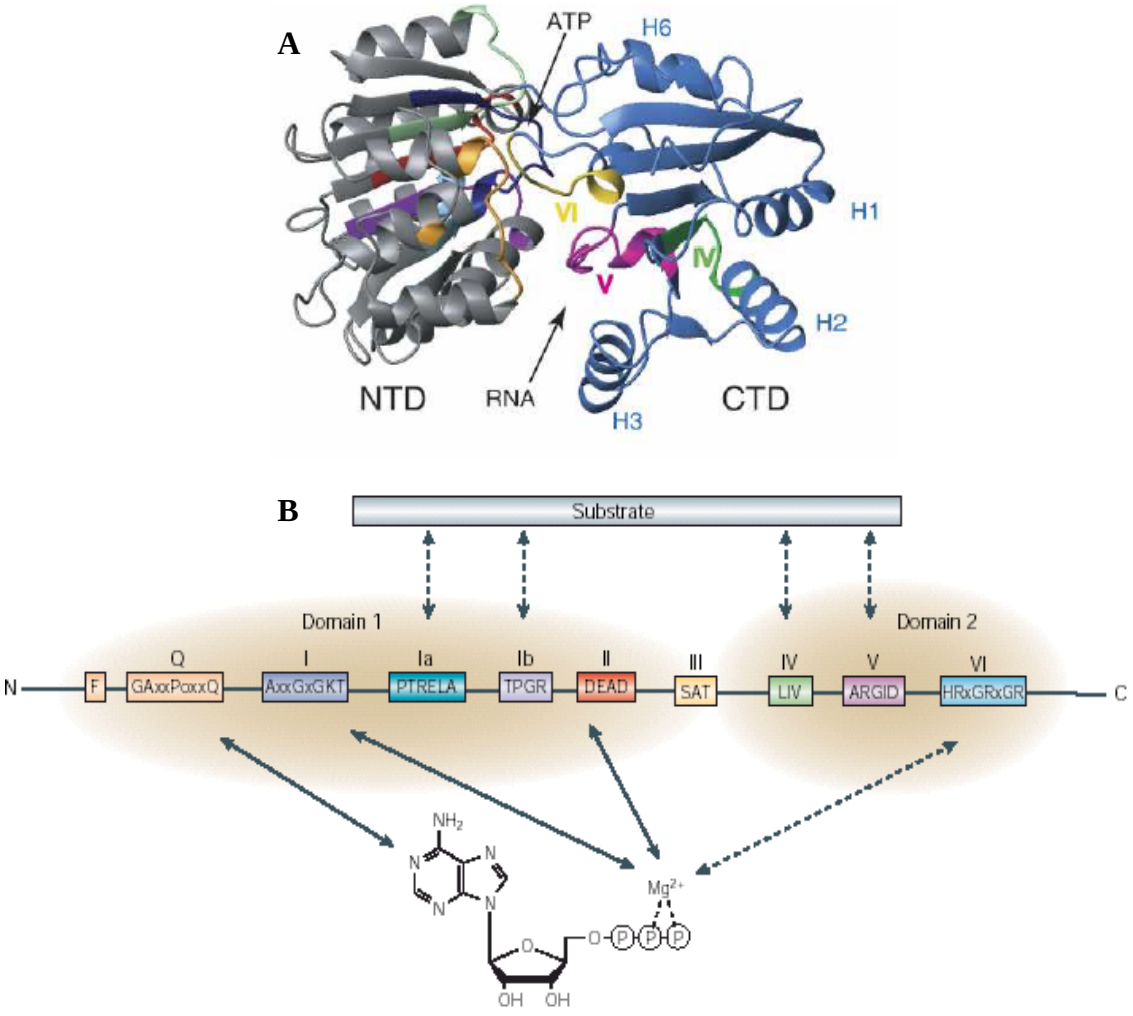


Figure 6: A. 3D structure of eIF4A illustrated the ATP and RNA binding regions. B. Correlations between the binding regions and ATP/RNA

The protein eIF4A has been extensively studied with NMR [9] and chemical shift assignments of the separated domains are available in the collaboration group of Gerhard Wagner at the Harvard Medical School in Boston.
The ATP and RNA binding titration of the N-terminal domain and C-terminal domain have shown chemical shift perturbation indicating binding and structure rearrangement at the binding site. Nevertheless, RNA unwinding can be achieved only via interaction with both domains, limiting the importance of those binding studies.
Measurements of the fully deuterated full length protein were unfortunately unsuccessful due to strong signal overlapping (data not shown). Intein segmental labeling is particularly interesting for such cases, by reducing signal overlap, focusing at a single domain in full length conformation.

The HSQC of the segmentally labeled N-terminal domain is shown in Figure 7. For this measurement, we tried to implement a novel pulse sequence [10] based on isotopically discriminated NMR spectroscopy. Both domains are ^{15}N labeled, but only one domain with ^{13}C. ^{15}N HSQC spectra for both domains can be recorded and separated via a ^{13}C filter. Unfortunately, due to some labeling “cross-telk” of the expressed fragments, we were not able to separate the HSQCs completely, so sparious signals and signal from the C-terminal domain remain in the recorded HSQC (marked with a grey cross in the following figures).

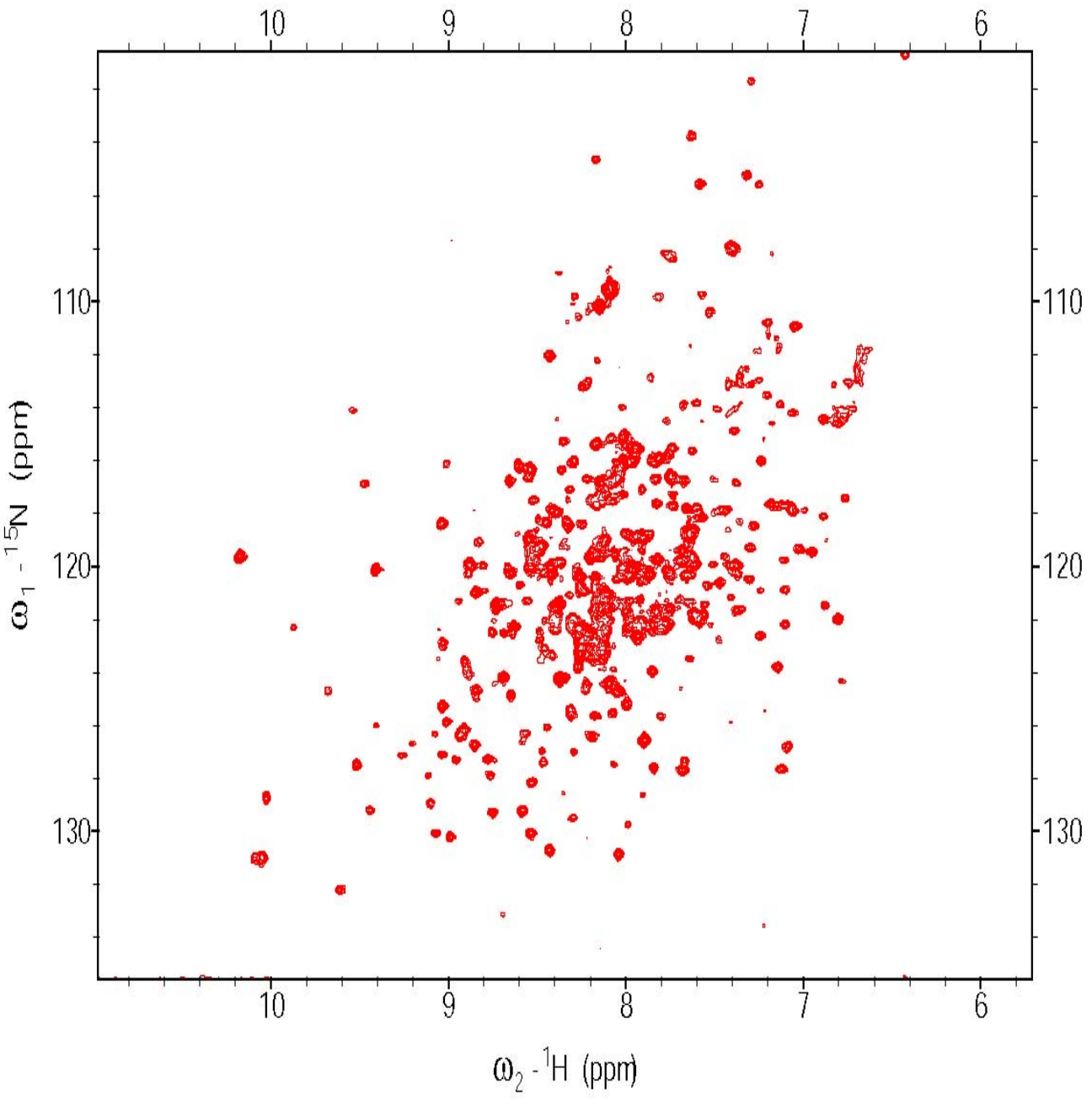


Figure 7: HSQC spectrum of the segmentally labeled N-terminal domain of eIF4A (^{15}N,^{13}C,^{1}D), which was ligated to C-terminal domain of eIF4A (^{15}N,^{12}C^{1}D) at its C-terminus, total MW 47kDa. The experiment was carried out in PBS at 23°C at a protein concentration of 200µM using BRUKER 600MHz spectrometer equipped with a cryoprobe.

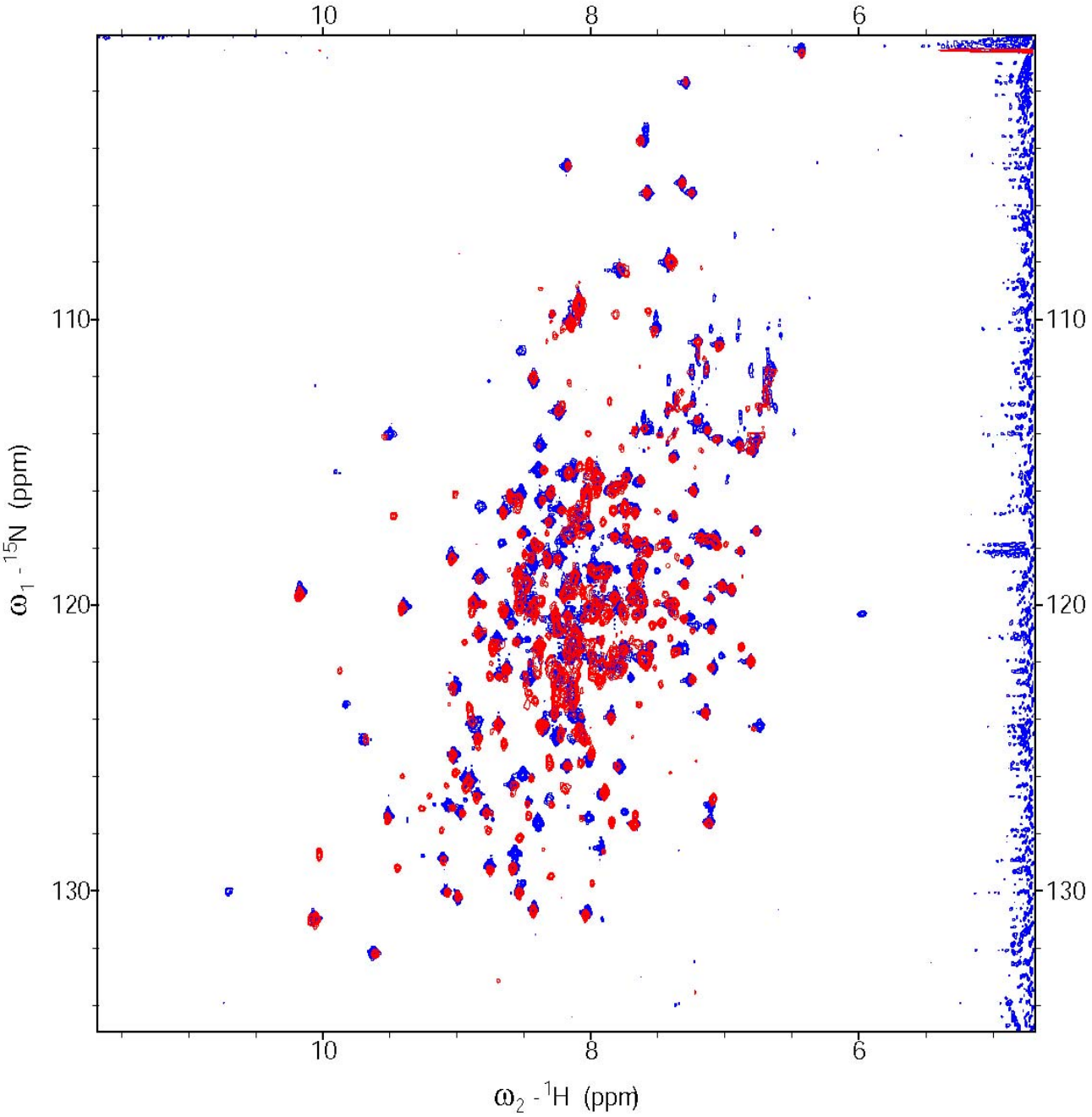


Figure 8: Overlayed NMR spectra of segmentally labeled eIF4A 47kDa. Red: HSQC spectrum of the segmentally labeled N-terminal domain of eIF4A (^{15}N,^{13}C,^{1}D) , which was ligated to the C-terminal domain of eIF4A (^{15}N,^{12}C^{1}D) at its C-terminus. Blue: reference HSQC spectrum of the N-terminal domain of eIF4A (^{15}N,^{1}D). The experiment was carried out in PBS at 23°C at a protein concentration of 200µM using a BRUKER 600MHz spectrometer equipped with a cryoprobe.

By overlay comparison (Figure 8), several chemical shift perturbations can be noticed, indicating that the structural environment of the N-terminal domain of eIF4A is different when this domain is fused to the C-terminal domain. This suggests the existence of interactions between the two domains even in the absence of further ligands as ATP or RNA.

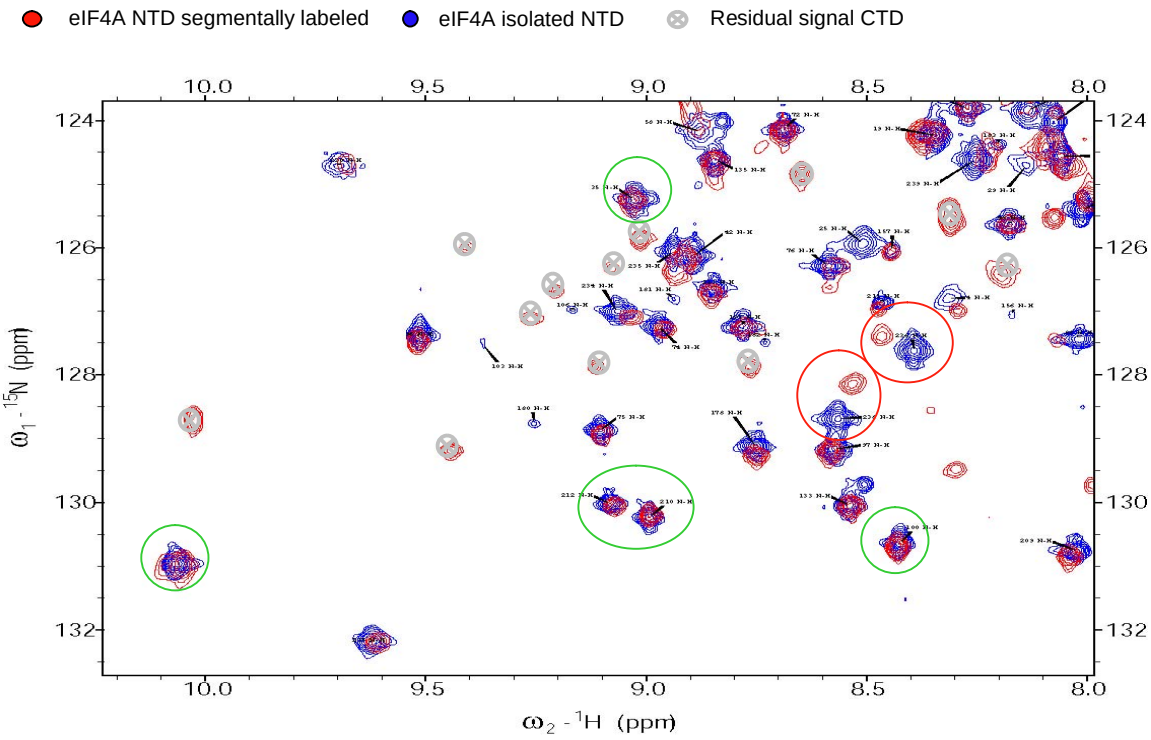


Figure 9: Overlayed NMR spectra of segmentally labeled eIF4A 47kDa. Red: HSQC spectrum of the segmentally labeled N-terminal domain of eIF4A ($^{15}N,^{13}C,^{1}D$) , which was ligated to the C-terminal domain of eIF4A ($^{15}N,^{12}C^{1}D$) at its C-terminus. Blue: reference HSQC spectrum of the N-terminal domain of eIF4A ($^{15}N,^{1}D$). The experiment was carried out in PBS at 23°C at a protein concentration of 200µM using BRUKER 600MHz spectrometer equipped with a cryoprobe. Grey crosses correspond to residual signals of the C-terminal domain.

Figure 9 illustrates in a narrower window that the change of the chemical environment does not affect all the residues (red circles), as demonstrated by unaffected signals (green circle) of the isolated N-terminal domain.
Some additional assignment experiments will have to be run to clarify the chemical shift assignment in the full length sample. Therefore, the following results only correspond to a qualitative analysis of chemical shift perturbations. The assignment used for this analysis corresponds to chemical shifts similarities based on the chemical shift assignment of the isolated N-terminal domain.

By plotting (Figure 10) the signal perturbation on the N-terminal domain structure based on no, low, middle and strong shift when the C-terminal domain is fused and can interact with the N-terminal domain, it is possible to determine which part of the structure is affected by this interaction.
First of all it is of importance to notice that most of the residues are not or little affected by the C-terminal interaction.
The residues strongly affected by this interaction (red and black) are mainly involved in the ATP binding regions like the Q, I and III box. This suggests structural rearrangements of the ATP binding pockets via C-terminal interaction facilitating ATP binding.

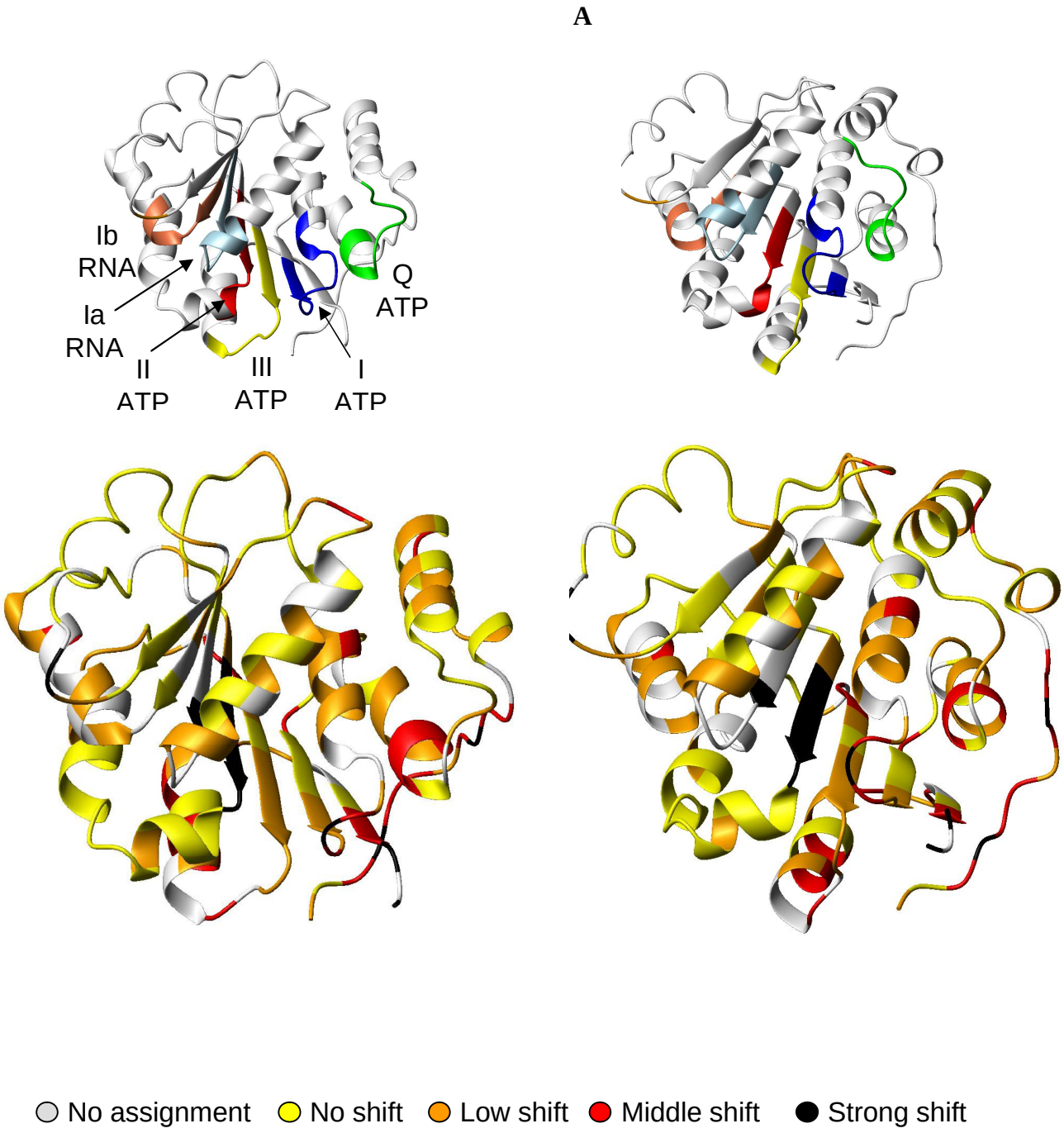


Figure 10: Qualitative analysis of chemical shift differences between the isolated form of the N-terminal domain (reference value) and the segmentally labeled N-terminal domain. Low shift corresponds to half line broad perturbation. Middle shift corresponds to full line broad perturbation. Strong shift corresponds to two line broad perturbation or more. **A:** structure indicates the ATP/RNA binding regions under two different view angles. B: same structures plotted with binding signal perturbations.

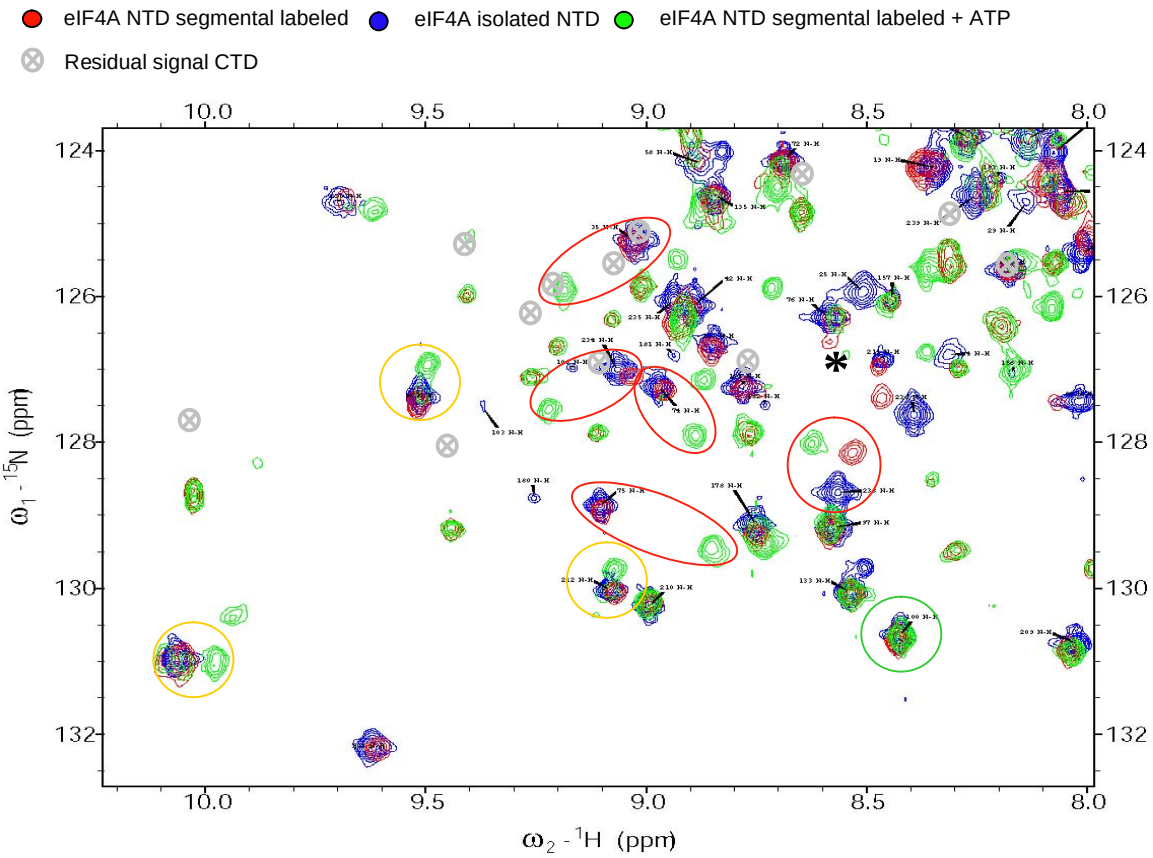


Figure 11: Overlayed NMR spectra of segmentally labeled eIF4A 47kDa. Red: HSQC spectrum of the segmentally labeled N-terminal domain of eIF4A ($^{15}N,^{13}C,^{1}D$) , which was ligated to the C-terminal domain of eIF4A ($^{15}N,^{12}C^{1}D$) at its C-terminus. Blue: reference HSQC spectrum of the N-terminal domain of eIF4A ($^{15}N,^{1}D$). Green: HSQC spectrum of the segmentally labeled N-terminal domain of eIF4A ($^{15}N,^{13}C,^{1}D$) with 800µM ATP The experiment was carried out in PBS at 23°C at a protein concentration of 200µM using BRUKER 600MHz spectrometer equipped with a cryoprobe. Grey crosses correspond to residual signals of the C-terminal domain.

Additional signal perturbations are observable when the segmentally labeled sample is titrated with ATP (Figure 11). Several signals retain their chemical shift, but most of the signals are affected, suggesting strong structural rearrangements by ATP binding.

It is also interesting to notice that signals which were already affected by the C-terminal domain interaction are again affected by the ATP binding. This is the case for the residue labeled with * in the figure 11. This observation confirms that the C-terminal domain interacts with the N-terminal domain influencing the structure to facilitate ATP binding.

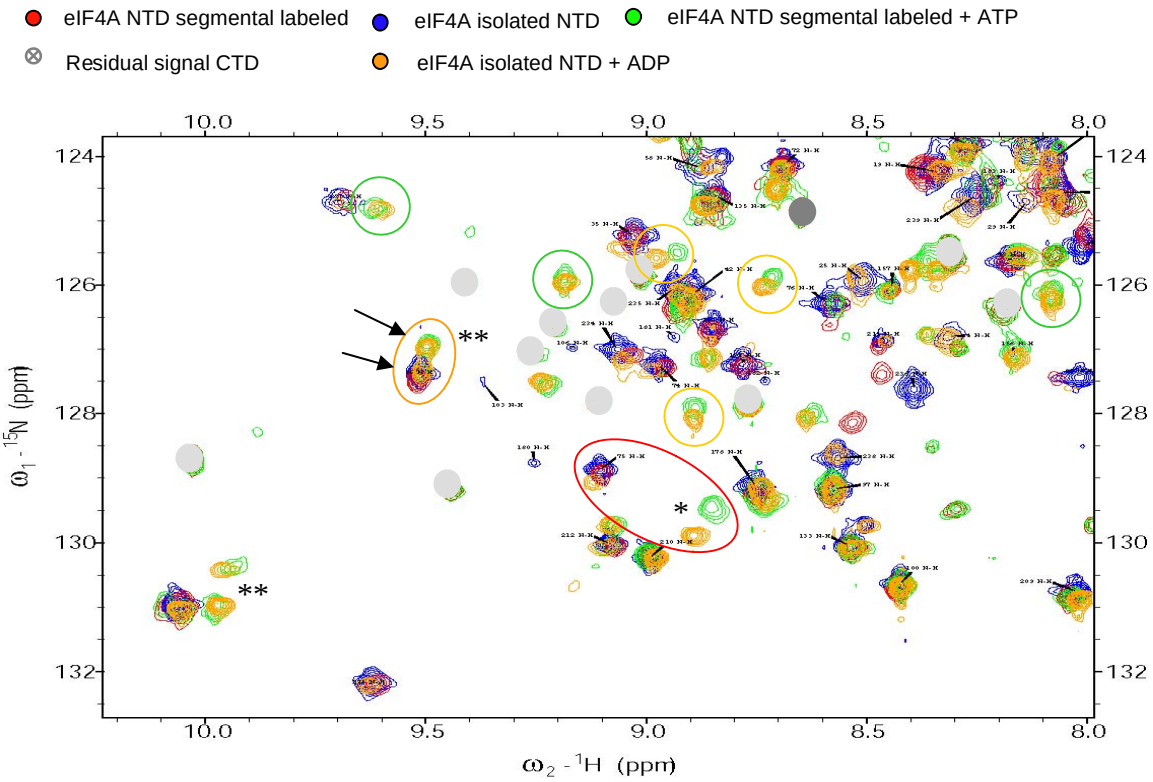


Figure 12: Overlayed NMR spectra of segmentally labeled eIF4A 47kDa. Red: HSQC spectrum of the segmentally labeled N-terminal domain of eIF4A ($^{15}N,^{13}C,^{1}D$) , which was ligated to C-terminal domain of eIF4A ($^{15}N,^{12}C^{1}D$) at its C-terminus. Blue: reference HSQC spectrum of the N-terminal domain of eIF4A ($^{15}N,^{1}D$). Green: HSQC spectrum of the segmentally labeled N-terminal domain of eIF4A ($^{15}N,^{13}C,^{1}D$) with 800µM ATP. Orange: HSQC spectrum of the N-terminal domain of eIF4A ($^{15}N,^{1}D$) + 800µM ADP. The experiment was carried out in PBS at 23°C at a protein concentration of 200µM using BRUKER 600MHz spectrometerequipped with a cryoprobe. Grey crosses correspond to residual signals of the C-terminal domain.

By comparing the ATP bound form of eIF4A labeled segmentally and the isolated N-terminal domain bound to ADP (Figure 12), it is interesting to notice that both protein mostly assume a very similar conformation (red and orange circles), bur also show some differences. This is the case, e.g. for the signal labeled with an asterisk. Also where the (ATP-)free forms of the two proteins have similar chemical shifts, when ATP is bound, the chemical environment is affected differently inducing chemical shift perturbations and therefore probably a different conformation.

It is also interesting to note that in the ADP bound conformation of the isolated N-terminal domain, two conformations (bounded and unbound) are observable at 800μM ADP (**). This might be due to the lower binding affinity of ADP to eIF4A and/or the lower affinity to the isolated N-terminal domain.

An important aspect for the eIF4A structure analysis is its inhibition mechanism via hippuristanol [11] (Figure 13).

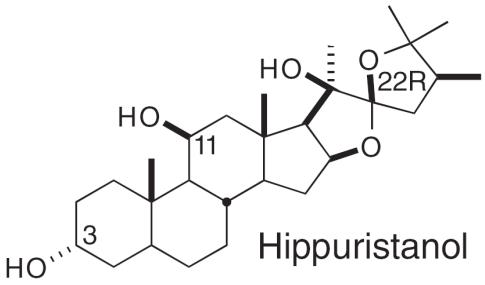


Figure 13: Chemical structure of hippuristanol

Hippuristanol is a cytotoxic, polyoxygenated steroid and is a eukaryotic translation initiation inhibitor prepared from an extract from the coral *Isis hippuris* [11]. Hippuristanol inhibits eIF4A ATPase activity in a dose-dependent manner [11] but is not a competitive ATP inhibitor.

NMR chemical-shift perturbation studies on uniformly labeled N-terminal and C-terminal domains of eIF4A showed substantial chemical shift changes for resonance signals of the C-terminal domain of eIF4A upon addition of hippuristanol, indicative of tight binding. Almost no changes were observed when hippuristanol was added to the N-terminal domain.

Figure 14 illustrates the influence of hippuristanol on the eIF4A N-terminal domain.

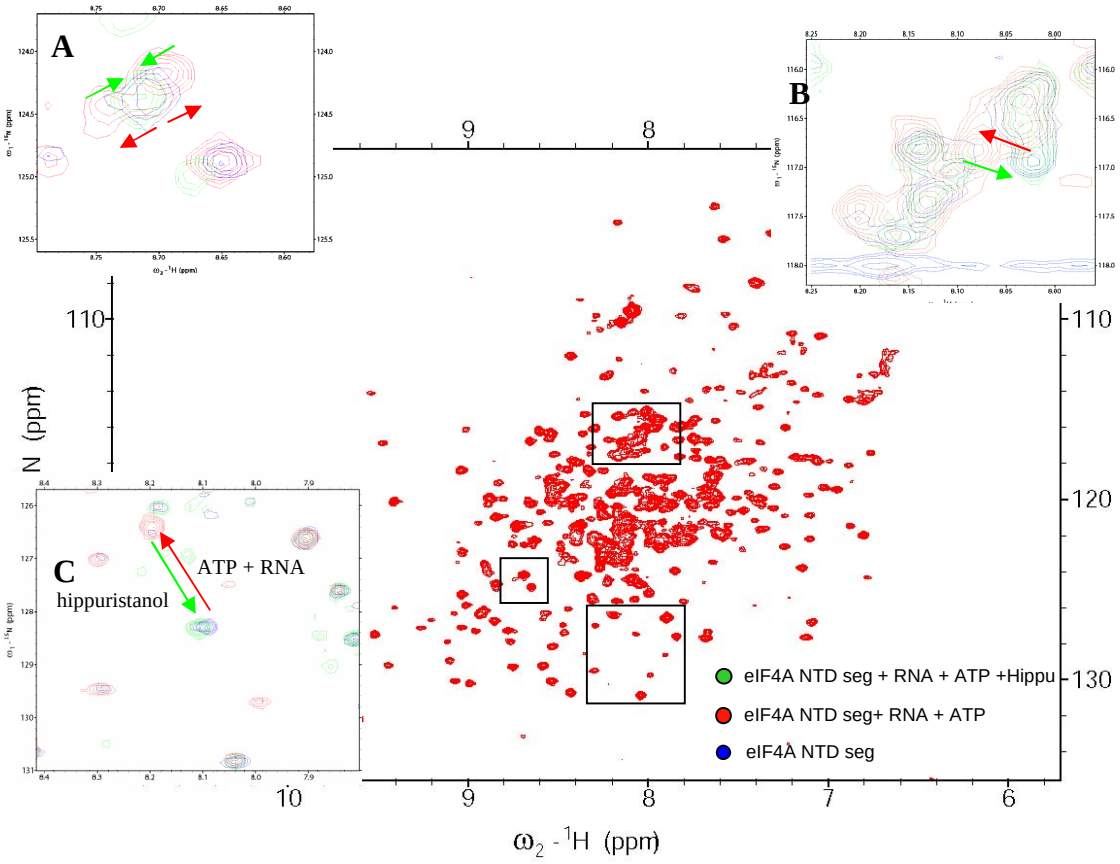


Figure 13: Overlayed NMR spectra of segmentally labeled eIF4A 47kDa. Blue: reference HSQC spectrum of the N-terminal domain of eIF4A (^{15}N,^{1}D) (eIF4A NTD seg). Red: HSQC spectrum of the segmentally labeled N-terminal domain of eIF4A (^{15}N,^{13}C,^{1}D) , titrated with 800µM ATP and 800µM RNA. (eIF4A NTD seg + RNA + ATP) Green: HSQC spectrum of the segmentally labeled N-terminal domain of eIF4A (^{15}N,^{13}C,^{1}D) with 800µM ATP, 800µM RNA and 400µM hippuristanol (eIF4A NTD seg + RNA + ATP +Hippu) . The experiment was carried out in PBS at 23°C at a protein concentration of 200µM using BRUKER 600MHz spectrometer equipped with a cryoprobe.

Hippuristanol seems to be capable to partially reverse the structural rearrangement of the N-terminal domain bound to ATP and RNA, since some the chemical shift perturbations return to their original position after inhibition (Figure 13- B and C). This can also be observed for chemical shifts which split after ATP and RNA binding (Figure 13-A).

Since the binding of the inhibitor hippuristanol occurs at the C-terminal domain and this inhibition is not competitive to ATP/RNA binding, hippuristanol must promote structural rearrangements that come closer to the free form of eIF4a. It is interesting to notice that the structural rearrangement of the C-terminal domain is strong enough to induce also structural rearrangements in the N-terminal domain.

Without the intein segmental labeling approach, these phenomena would not have been so easily observable. A thorough analysis will require additional measurements as well as additional assignments to clarify which residues are really involved in these structural rearrangements, and to finally arrive at a deeper understanding of the RNA helicase mechanism of eIF4A.

5. Conclusions

Intein mediated segmental isotopic labeling has reached maturity and its application provides challenging experiments. The *in vivo* ligation approach presented in this report is particularly useful for several applications, especially due to its simple laboratory implementation.

Naturally the final sequence of the ligated protein is mutated at the ligated region, which should preferably be situated in a linker region. But this might influence the protein structure, stability, and activity. Therefore we recommend achieving biological functional assays to determine the influence of this mutation to the protein activity.

It is interesting to notice that intein segmental labeling open new insights to multi domain proteins, allowing atomic resolution studies of internal domain interactions, which might not been detectable with isolated domains. eIF4A is a particularly interesting example for inter-domain interactions and its influences to the ATP/RNA binding mechanisms and kinetics. Obviously, additional experiments on eIF4A are necessary to fully understand the unwinding mechanism of RNA, but our results has demonstrated the importance and necessity of NMR experiments of biological active multi domain protein.

The *in vivo* ligation method has shown high efficiency, especially in term of ligation yield. Nevertheless some aspects e. g. labeling scrambling or fragments degradation limit sometime its application. Therefore we are currently working on a in-vitro protocol comparable to the *in-vivo* approach extended the range of intein segmental labeling application.

.

6. Literature

[1]. Wider, G. & Wüthrich, K. NMR spectroscopy of large molecules and multimolecular assemblies in solution. Curr. Opin. Struct. Biol., 9, 594-601 (1999)

[2]. Yu, H. Extending the size limit of protein nuclear magnetic resonance. Proc. Natl. Acad. Sci. USA 96, 332-334.

[3]. Iwai H, Zuger S, Jin J, Tam PH., Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiforme. FEBS Lett. 2006 Mar 20;580(7):1853-8..

[4]. Zuger S, Iwai H., Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies. Nat Biotechnol. 2005 Jun;23(6):736-40.

[5]. van den Ent F, Lowe J., RF cloning: a restriction-free method for inserting target genes into plasmids. J Biochem Biophys Methods. 2006 Apr 30;67(1):67-74.

[6]. David Kennell, Processing Endoribonucleases and mRNA Degradation in Bacteria J Bacteriol. 2002 September; 184(17): 4645–4657.

[7]. J. Marley, M. Lu, C. Bracken, A method for efficient isotopic labeling of recombinant proteins. J Biomol NMR. 2001 May; 20(1):71-75.

[8]. Pause, A. and Sonenberg, N. Mutational analysis of DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. *EMBO J.* **11**: 2643-2654 (1992)

[9]. M. Oberer, A. Marintchev, and G. Wagner Structural basis for the enhancement of eIF4A helicase activity by eIF4G. Genes Dev. 19(18):2212-23 (2005).

[10]. Golovanov AP, Blankley RT, Avis JM, Bermel W. Isotopically discriminated NMR spectroscopy: a tool for investigating complex protein interactions in vitro. J Am Chem Soc. 2007 May 23;129(20):6528-35.

[11]. M.-E. Bordeleau, A. Mori, M. Oberer, L. Lindqvist, L. S Chard, T. Higa, G. J Belsham, G. Wagner, J. Tanaka & J. Pelletier: Functional characterization of IRESes by an inhibitor of the RNA helicase eIF4A. Nat Chem Biol. 2006 Apr;2(4):213-20.

Chapter 2:

Thermal shift assay and application to biochemistry for NMR purpose

7. Summary

Biophysical characterization of proteins is a key factor to understand biology at chemical level. Understanding protein folding/unfolding improves our knowledge of protein stability, protein aggregation or fibril formation. Nevertheless factors influencing protein folding are important and high throughput screening assays facilitate such studies. Thermal shift assay takes advantage of an environmentally sensitive fluorescence dye, such as Sypro Orange, and follows its signal changes while the protein undergoes thermal unfolding. The assay is widely used for screening compound libraries for ligands of targets proteins. The ligand-binding affinity of any potential inhibitor can be assessed from the shift of the unfolding temperature (Tm) obtained in the presence vs absence of the potential inhibitor.

The assay can also be used to detect und study essential conditions for protein folding unstability. We applied this strategy to determine buffer conditions, which destabilize GB1; allowing unfolding/refolding NMR experiments via thermal denaturation in a temperature range compatible with NMR spectrometers.

8. Introduction

Indeed protein homogeneity, stability, and solubility are critical factors affecting aggregation, folding, or probability of protein crystal formation. Structural genomics projects have increased considerably the number of proteins, available in sufficient amount for structural and functional studies. However, protein sampling as shown to be a critical and limiting factor for suitable measurement. Especially the treatment of sample at high protein concentration is important, since the buffer provides environment which maintains the protein stable and minimizes aggregation. Several factors such pH, ionic strength, additives precipitant and protein concentration, lead to different protein stability behavior. Unfortunately many proteins are found to be unstable under a typical buffer condition and therefore large number of buffer conditions should be tested in order to maintain the protein in favorable conditions.

Although functional assays can be used to optimize buffers, assays using biophysical properties, such as homogeneity, solubility, and stability, are often preferred since the results are more accurate and reproducible, especially in case of structural genomics , where the biological function of the target protein is often not know before the structural analysis. The thermal shift assay described in this chapter is one of the biophysical assays allowing to determine the stabilizing proper of buffers for a target protein.

9. Theory

Thermal shift assay is a thermofluor-based high-throughput measurement originally developed by Pantoliano and coworkers (US patent 6,020,141) as development tool for drug discovery to allow rapid binding identification of ligand libraries to target proteins [1]. The structural genomic community involved in structural studies employed this method to optimized stability of proteins [2].

In principal folded and unfolded protein can be distinguished through exposure to a hydrophobic fluoroprobe,: e.g, SYPRO® Orange (Sigma) (see Figure 1). The probe is quenched in aqueous solution but will preferentially bind to the exposed hydrophobic interior of an unfolding protein leading to a sharp decrease in quenching and therefore an increase in fluorescence emission, which can be studied as a function of temperature. The typical result of thermally induced unfolding is characterized by the typical two-state model with a sharp transition between the folded and the unfolded state. The resulting transition defines the melting temperature T_m, which represents the temperature at which the protein unfolds. The melting temperature can be as well determined using circular dichroism (CD), by observing the specific wavelength characteristic for secondary structure as a function of temperature. Melting temperatures measured with both techniques correlate well for several proteins and therefore the thermal shift assay can proposed use as an alternative method to CD.

This method is currently used for screening purposes like identification of ligands [1] or optimal buffer screening for NMR purposes [2].

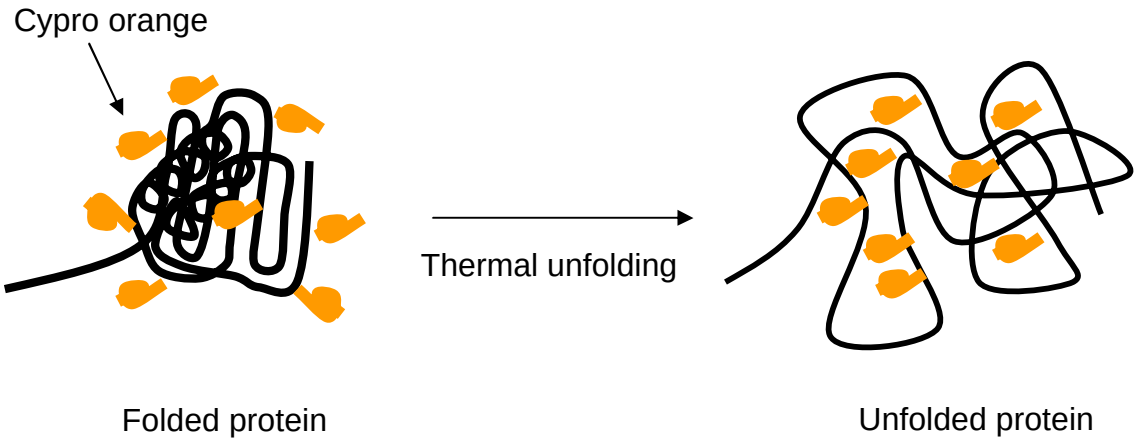


Figure 1: Thermally induced unfolding representation and environment changes of the fluoroprobe Sypro Orange.

10. Material and methods

The thermal shift assay (TSA) can be performed conveniently in a commercially available real-time PCR instrument [3], which allows screening 96 samples in 1½ hour.

The real-time PCR instrument used for the current study is a Q-PCR ABI 7900HT machine from Applied Biosystems. This machine contains continuous wavelength detection from 500-660 nm, which allows the use of multiple fluorophores in a single reaction. It was gratefully provided from the Institute of Chemistry and Cell Biology (ICCB) at the Harvard Medical School in Boston.

The PCR cycle template used for the analysis is shown in Figure 2. For the reversibility experiments of GB1, the PCR cycle shown in Figure 3 has been set up.

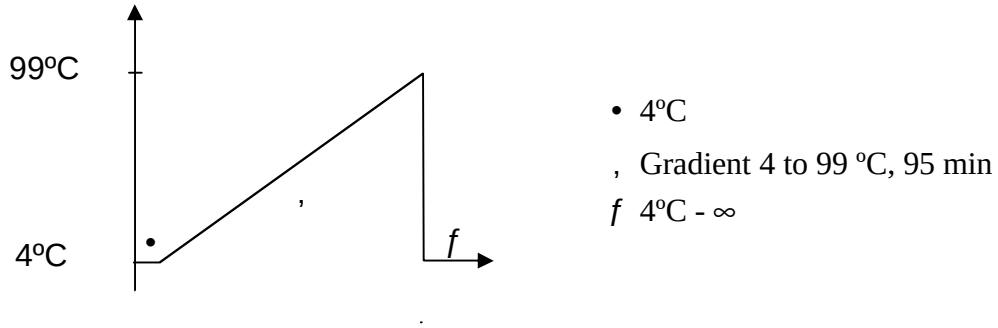


Figure 2: Thermal cycle used for the thermal shift assay.

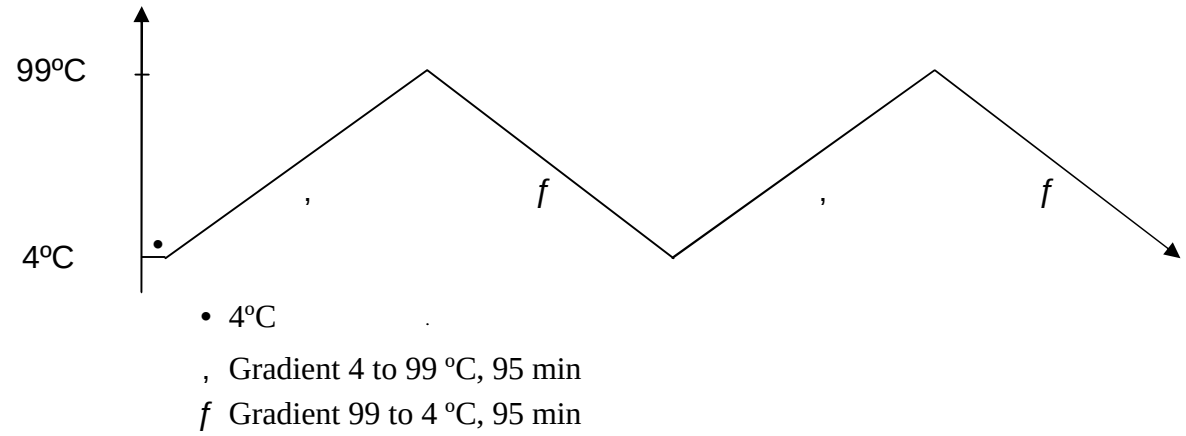


Figure 3: Thermal cycle used for the thermal shift assay during reversibility

For the assay only a relatively small amount of protein (5-25 µg per well) is needed, which is very attractive for high-throughput biophysical characterizations.

SYPRO® Orange protein gel stain is commercially available from Sigma Aldrich (S5692) as 5000x concentrate and is a protein staining agent for gel electrophoresis which is more sensitive (1-2 ng) than generally used coomassie blue. The commercial stock was diluted 50 times and aliquoted at -20ºC.

Each well contains:

- 4 to 25µg of protein,
- 1µL 50x concentrate Sypro Orange stock solution,
- 25µL 2x concentrated buffer
- X µL ddH_2O up to 50µL

100 µL Micro Amp Fast Optical 96-well reaction plates with barcode from Applied Biosystems have been used for the measurements and were sealed with a supplied plastic foil to prevent sample evaporation.

Data were imported using a self programmed Microsoft Excel macro and the fluorescence intensity was plotted as a function of temperature. The macro allows defining the wavelength to be plotted, since the maximal emission wavelength can be shifted due to the buffer conditions. Depending on the protein hydrophobicity, the fluorescence emission can reach a level that is sufficient to saturate the detector. Since the fluorescence emission has a Gaussian shape, the plotted intensity can then be reduced by shifting the wavelength until the signal doesn't reach saturation any more.

NMR measurements have been realized at a BRUKER 500 MHz spectrometer from BRUKER equipped with a TXI probe; the temperature has been calibrated using pure methanol. Before any titration series, the ^{1}H and ^{15}N 90° pulses have been measured and implemented in a standard HSQC pulse sequence.

Specific measurement parameters will be given together with the HSQC spectra presented in the Results section.

10.1. Method implementation

In order to implement and test the technique, 96-well sample plates were prepared with the protein Eukaryotic Initiation Factor 5E (eIF5E) (full length and truncated version) at various pH values (4.5 to pH=9). The sample mixture was set up as described previously in Materials and Methods, and the assay was run with the cycle shown in Figure 2.

TSA results of the full and truncated construct are shown in Figure 1-A and 1-B, resp. The melting curve clearly demonstrates that the pH range 6 to 8.5 does not strongly influence the protein stability. The highest melting temperature is found for pH=7, which agrees well with previous experiments (data not shown).

At pH=5 the protein loses its secondary structure and no thermal unfolding can be observed. In contrary, pH 5.5 and 9 are not sufficient to unfold the protein but the protein is destabilized, since the T_m at this pH is lower.

In case of eIF5E, the optimal pH for NMR purposes would be pH=7. Additional experiments have also shown that a specific eIF5E binding peptide stabilizes the protein structure (Figure 4). This result was already suggested from NMR data, but TSA clearly demonstrates the positive influence of the peptide to the protein stability.

In conclusion the protocol implemented here is suitable to study buffer stability influences to proteins in a fast and simple manner. (An Excel macro has been also developed to import and analyze the resulted data more conveniently).

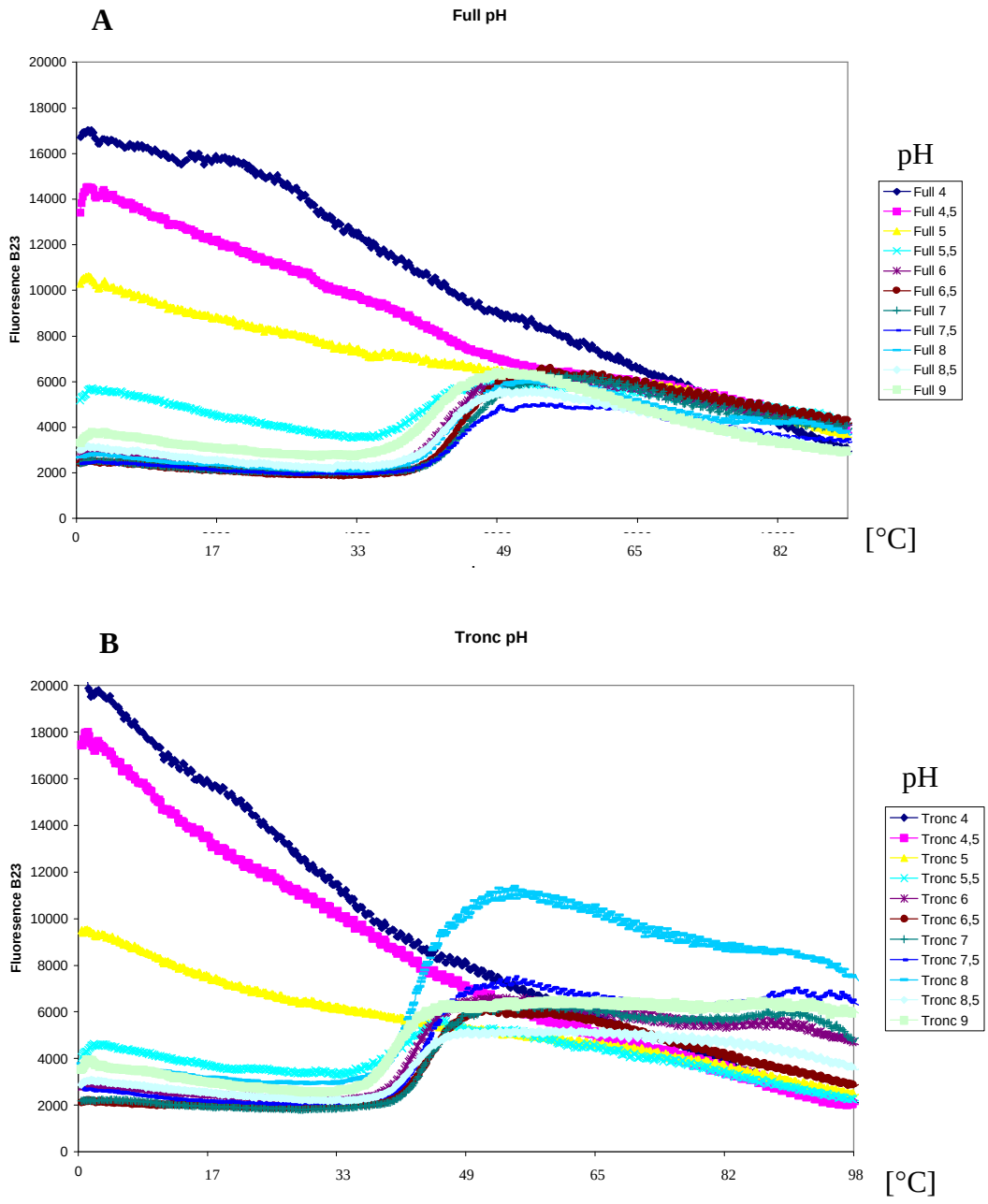


Figure 3: pH TSA matrix (5 to 9) of eIF5E (A) and its truncated form (B).

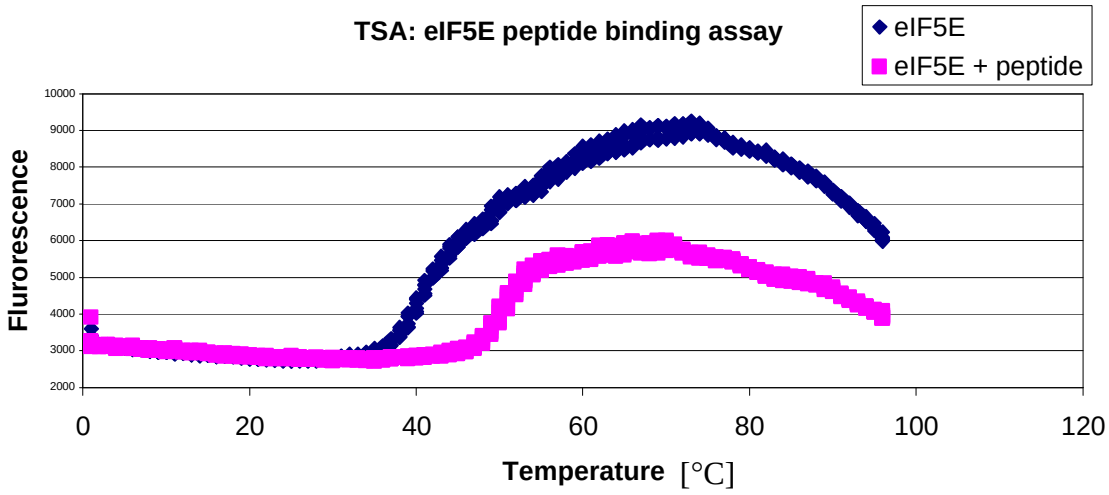


Figure 4: TSA binding assay of eIF5E and its target peptide

11. Results: Unfolding/refolding study of GB1

In order to test the limits of this method, we have continued to test different types of proteins and their folding as influenced by denaturizing agents. In case of complete protein unfolding, no T_m is expected. To verify this hypothesis, TSA was run with the protein Eukaryotic Initiation Factor 4G (eIF4G) fused with GB1 at the N-terminus. Buffers containing guanidinium chloride at 0, 1, 2, 3, 4, 5, and 6M resp. respectively were used to denaturate the protein. Figure 5 show the resulting annealing curves at corresponding concentrations of GdmCl.

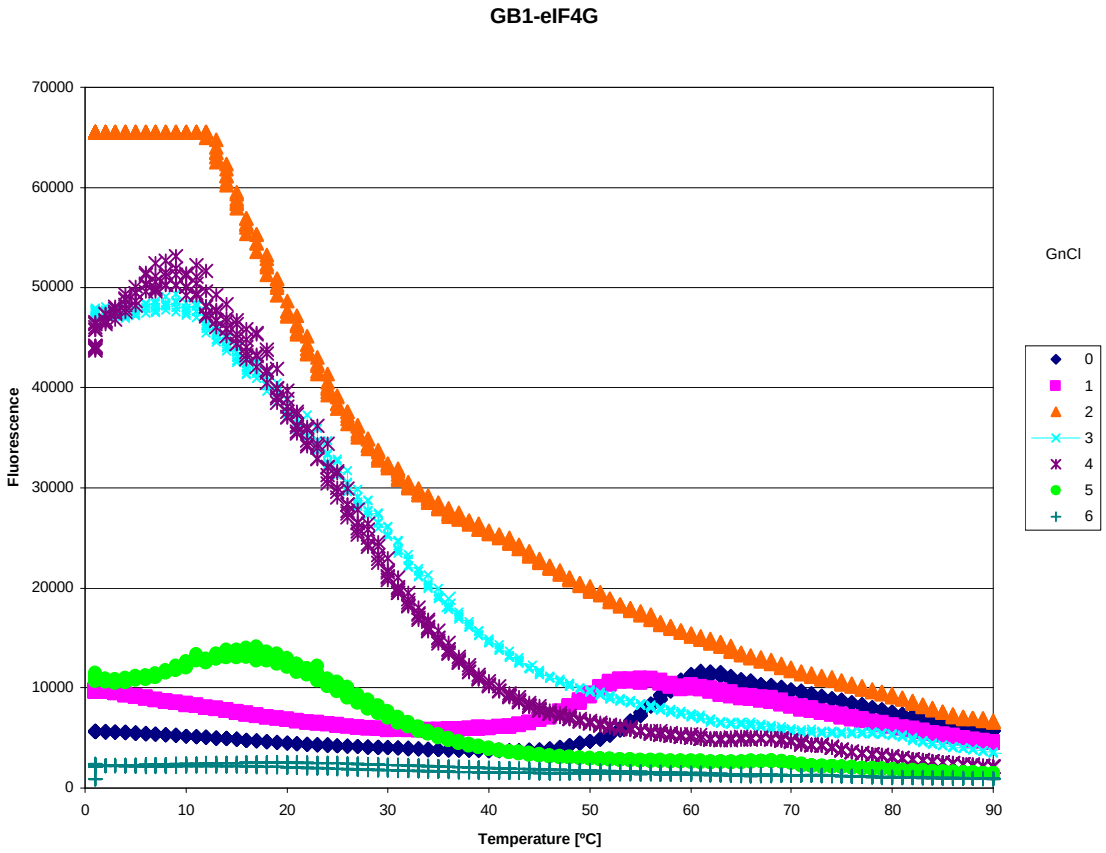


Figure 5: TSA of eIF4G fused with GB1 at the C-terminal domain at various GdmCl concentrations

By analyzing the results of the TSA, we observe a decrease of the T_m at 1M GdmCl and a complete loss of T_m at 2M GdmCl. Curiously the melting curve reappeares at GdmCl concentrations of 3 to 5M and disappears again at 6M GdmCl, which corresponds to complete denaturation. The melting curve observed at 3 to 5M GdmCl was suspected to be the possible denaturation of the GB1 domain of the fusion protein.

TSA of GB1 was run using the same denaturing buffers and the thermo cycles described previously in figure 3.

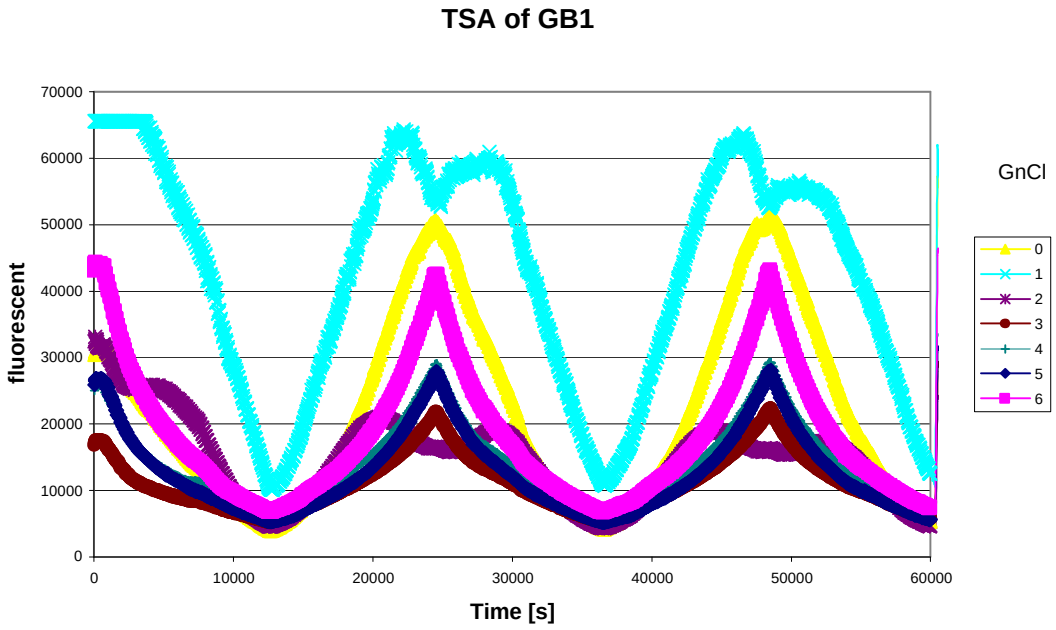


Figure 5: TSA of GB1. Folding reversibility experiment

Figure 5 is the resulting analysis of the TSA. GB1 seems to be so stable that at native conditions the protein doesn't show heat denaturation. First at 1M to 2M GdmCl an annealing temperature is observed (T_m= 28ºC - figure 6) and at higher concentrations the protein seems to be completely denaturated since the curves don't show heat denaturation.

An interesting property of GB1 is that the protein seems to be able to almost completely refold after the heat denaturation in the cooling step. The TSA of GB1 at 2M GdmCl is shown in figure 7. While the fluorescence intensities after each heating/cooling step decrease, due to possible aggregation, this decrease is small and a large majority of GB1 refolds. Based on the TSA results we can conclude that GB1 has the property of spontaneous refolding itself after heat denaturation at 2M GdmCl.

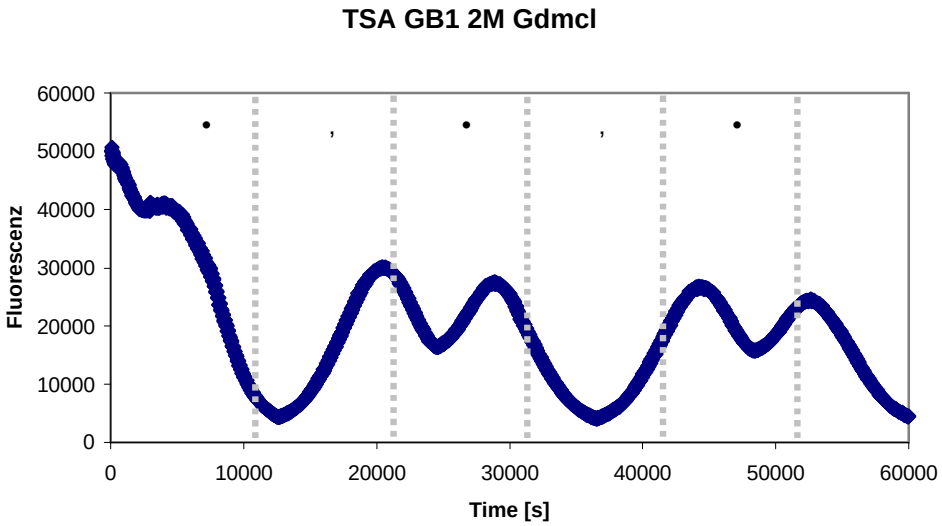


Figure 6: TSA of GB1 at 2M GdmCl, folding reversibility experiment.

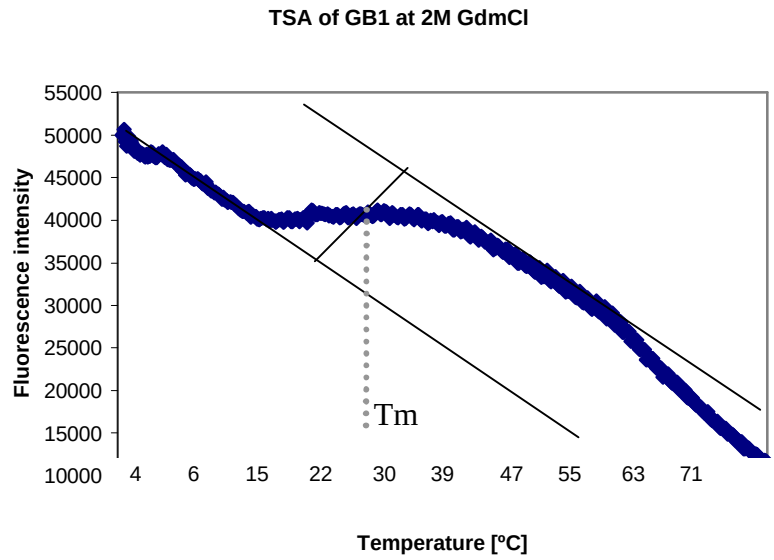


Figure 7: TSA of GB1 at 2M GdmCl and determination of T_m

To verify the particular folding properties of GB1 in 2M GdmCl, we have studied the defolding/refolding process of GB1 using NMR at 200µM protein. A series of HSQC was recorded at different temperatures to simulate the thermal denaturation. In order to assign resonance peaks and therefore follow the defolding/refolding process, the first HSQC was recorded at 4°C and reassigned

using the assignment published in Selenko et al. [4]. Minimal chemical shifts were observed and 84% of GB1 signals were reassigned (Spectra 1).

The temperature was stepwise increased 4°C until the resulted resonance peaks in HSQC looked completely unfolded, which appeared at 48°C. Some of the recorded spectra are shown in the Figure 8.

The reversibility of the defolding/refolding process has been verified by measuring several times at 36°C and 48°C, and the HSQC remains same after several cycles, (data not shown).

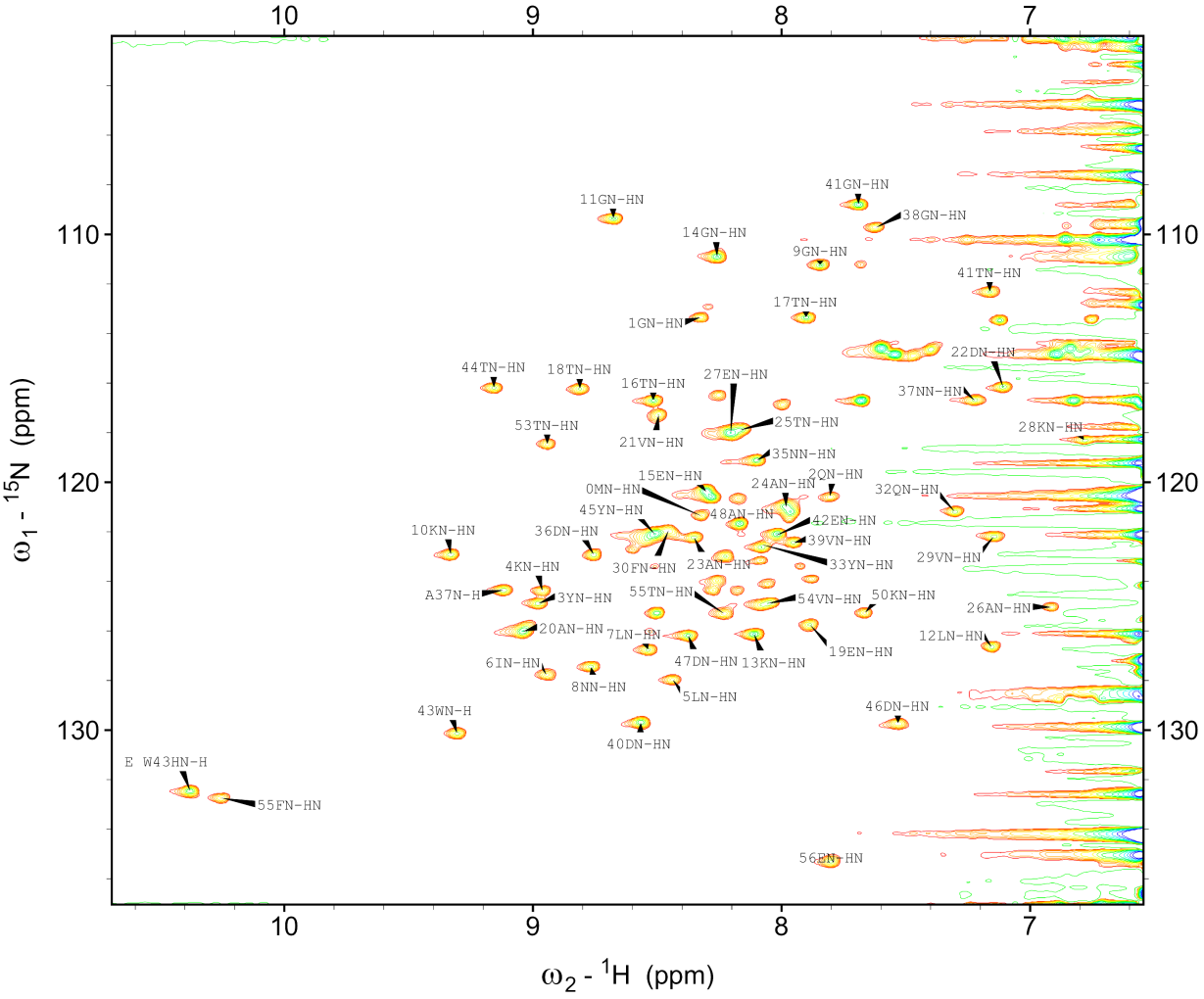


Figure 8: HSQC of GB1, 500MHz, 8 Scans, 4 °C, 200µM, 2M GdmCl, pH=6.4.

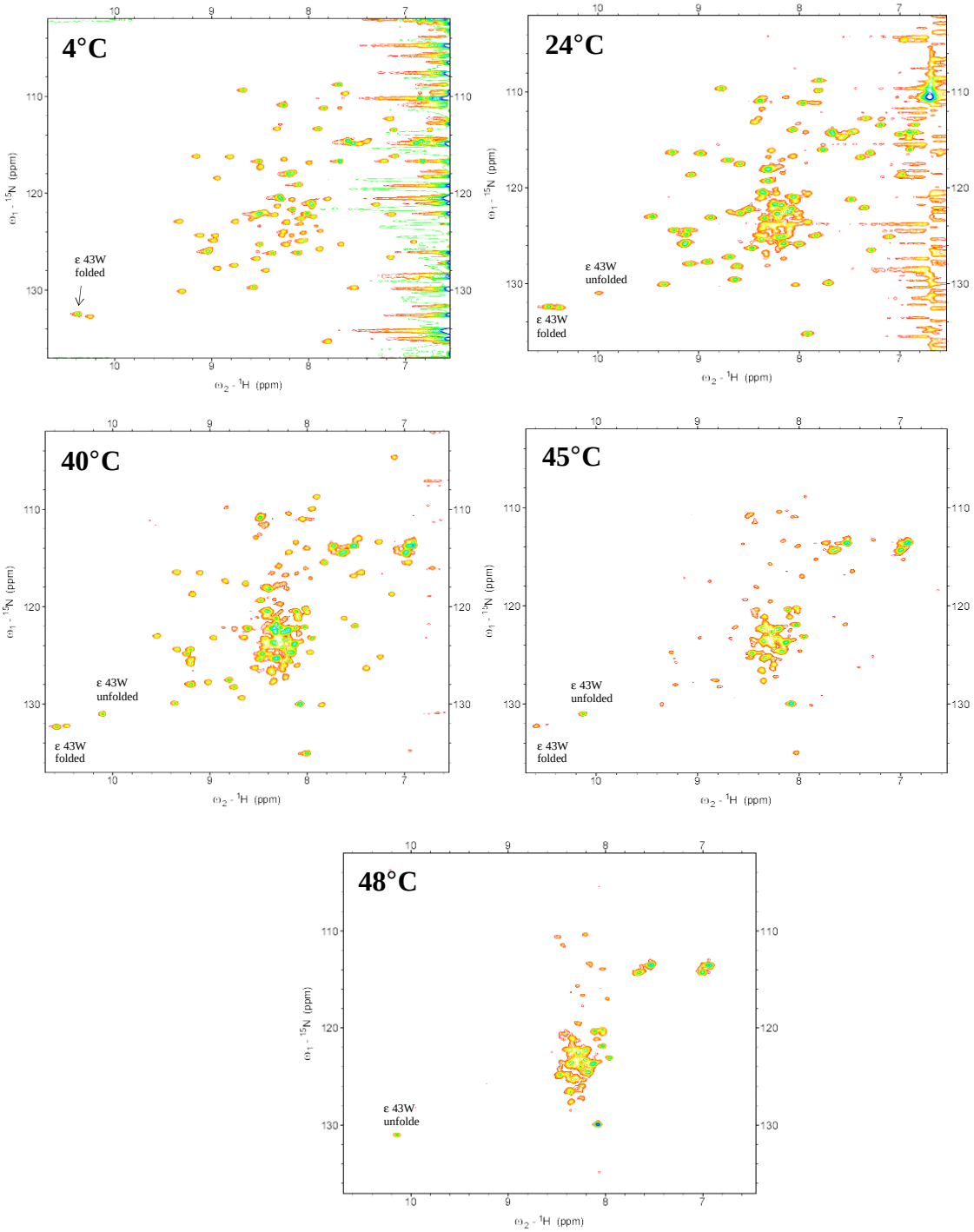


Figure 9: HSQC spectra of GB1, 500MHz, 8 scans at various temperatures, 200µM, 2M GdmCl, pH=6.4.

The unfolding process is starting already at low temperature and covers a large temperature range, but is significantly happening between 40°C and 48°C.

The unfolding process seems to be happening with slow exchange, since the folded (10.55; 132.3 at 36°C) and unfolded W43signals (10.07; 131 at 36°C) retain their chemical shifts and only the relative intensity is affected by the temperature.

The volume ratio of both peaks has been plotted in Graph 8, and characterizes the folded/unfolded population of W43. It is not necessarily completely representative of the full protein folding but it gives an idea about the refolding process.

The defolding/refolding process was characterized by following the peak shifts and peaks were considered completely unfolded when their signals were not detectable anymore.

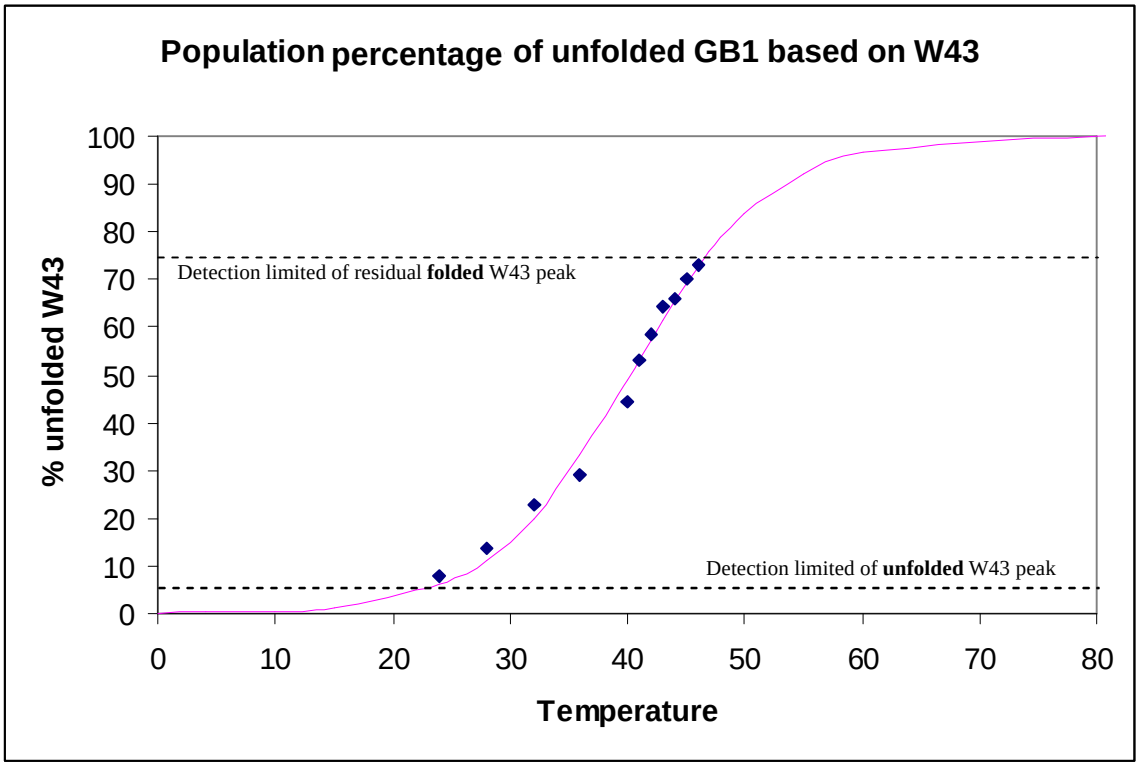


Figure 7: Population repartition of folded/unfolded GB1 population depending on the temperature.

Using this strategy we were able to follow at which temperature residues were completely unfolded or which residues refold first (Figures 8 and 9).

We were not able to observe any unfolding pattern during thermal unfolding in contrast to the refolding experiment, in which the helix clearly refolds first. 60% (6 of 10) of the assigned helix signals are already showing signals in the folded state at 47°C (Figure 5). Almost no signals of β-sheet residues are present at this temperature, suggesting that only the helix is partially formed.

At 46°C in figure 5, several signals assigned in the β-sheet region appear back to their folded shifts.

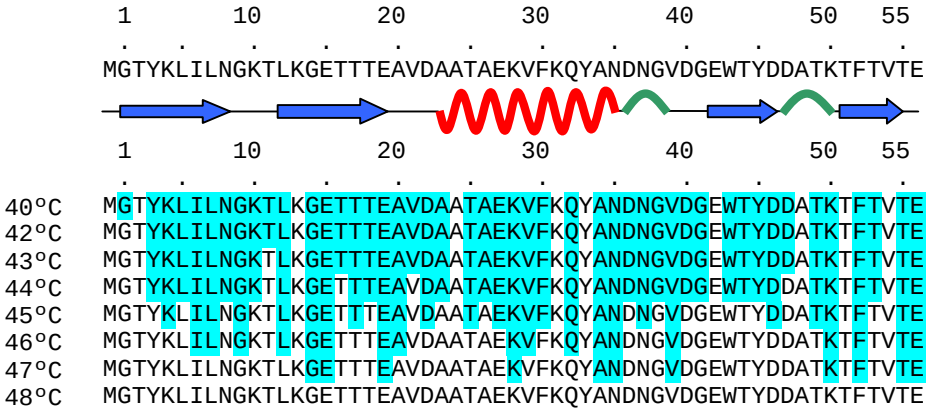


Figure 8: Residue unfolding analysis of GB1 via thermal unfolding

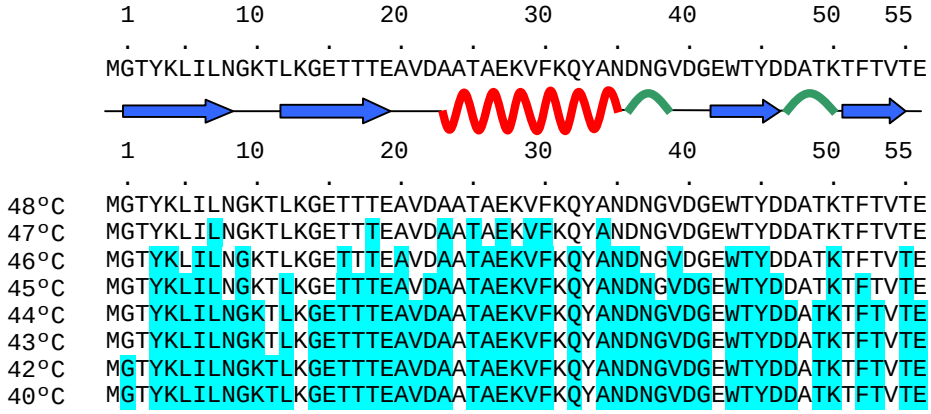


Figure 9: Residue refolding analysis of GB1 via thermal refolding.

By comparing Figures 8 and 9, we can observe that the unfolding and refolding processes are different, since corresponding HSQCs at same temperature are different.

Conformation and dynamics of GB1 during folding have been previously studied [5][6] using similar techniques to destabilize the protein. In one study GdmCl was also used to induce unfolding, which was followed with CD and fluorescence of the tryptophan 43 as intrinsic reporter. Several folding states of GB1 have been characterized and the minimal concentration of GdmCl to unfolde GB1 was 4M; however, at 3M GB1 remains 50% folded [6]. It has also been shown that 2M GdmCl over 95% of folded molecules remains. Those results merge perfectly with the results observed with the TSA. 2M GdmCl destabilizes the protein to the limit where GB1 still remains almost completely folded. The temperature increase, during the thermal unfolding titration, destabilizes further the protein and simulates a higher amount of GdmCl.

Temperature titrations are very convenient in NMR spectroscopy compared to titrations with denaturing agents, which are problematic in terms of sample preparation and recovery.

In summary, TSA provides a convenient and high throughput way to find buffer conditions which easily allow studies of protein unfolding/refolding processes. Even the reversibility of the protein folding can be easily checked.

12. Application to natively unfolded • -synuclein protein

• -synuclein is a synuclein protein of unknown function primarily found in neural tissue, where it is seen mainly in presynaptic terminals. It is predominantly expressed in the neocortex, hippocampus, substantia nigra, thalamus, and cerebellum. It is predominantly a neuronal protein, but can also be found in glial cells.

• -synuclein is specifically upregulated in a discrete population of presynaptic terminals of the songbird brain during a period of song-acquisition-related synaptic rearrangement.[7]

Normally an unstructured soluble protein, • -synuclein can aggregate to form insoluble fibrils in pathological conditions characterized by Lewy bodies, such as Parkinson's disease, Lewy bodies, and multiple system atrophy (Figure 10). • -synuclein is the primary structural component of Lewy body fibrils. In addition, an • -synuclein fragment, known as the non-Abeta component (NAC), is found in amyloid plaques in Alzheimer's disease.

The aggregation mechanism of • -synuclein is quite uncertain. There is some evidence of a structured intermediate rich in beta structure that can be the precursor of aggregation and, ultimately, Lewy bodies.[8] [9] A single molecule study in 2008 suggests • -synuclein exists as a mix of unstructured, • -helix and beta-sheet rich conformers in equilibrium. Mutations or buffer conditions known to improve aggregation strongly increase the population of the beta conformer, thus suggesting this could be a conformation related to pathogenetic aggregation [10].

In rare cases of familial forms of Parkinson's disease there is a mutation in the gene coding for •-synuclein. Three point mutations have been identified thus far: A53T, A30P and E46K. In addition, duplication and triplication of the gene appear to be the cause of Parkinson's disease in other lineages.

Antibodies against •-synuclein have replaced antibodies against ubiquitin as the gold standard for immunostaining of Lewy bodies.

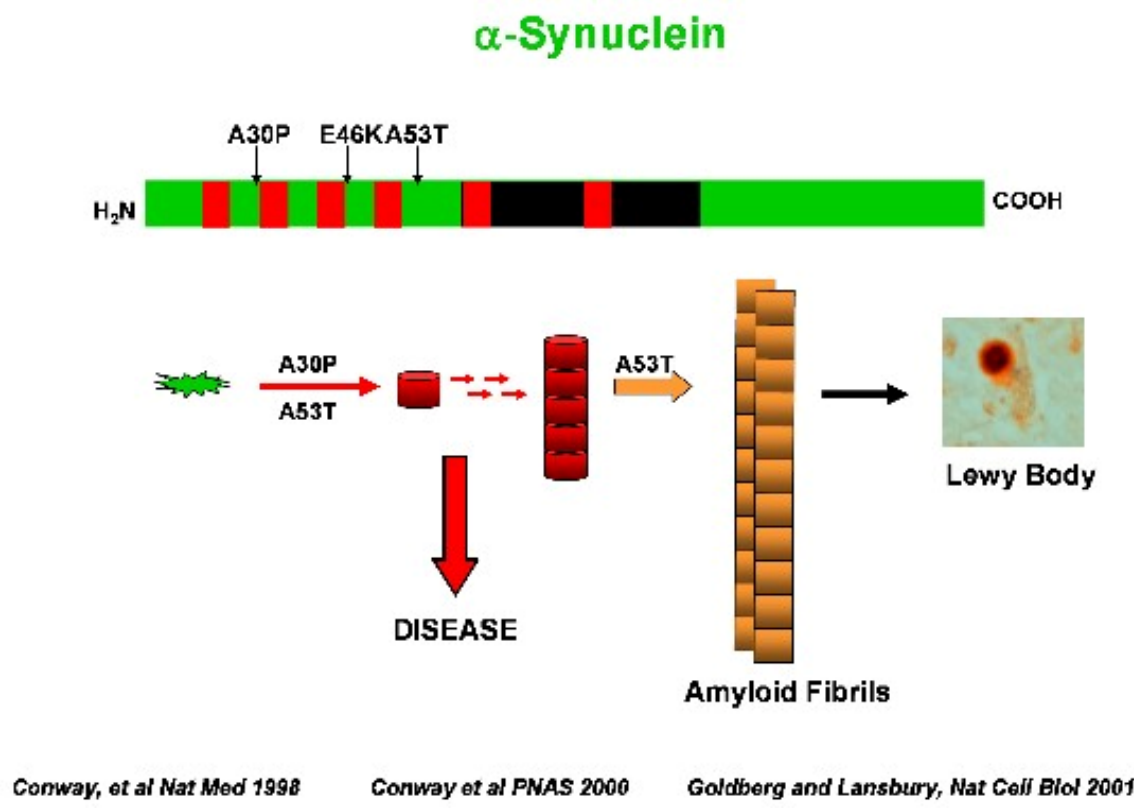


Figure 10: Amyloid fibrils and Lewis body formation of α Synuclein

The rate of • synuclein fibrillation is significantly accelerated by numerous environmental conditions and factors, including low pH and high temperature [11] [12], organic solvents [13], heparin and other glycosaminoglycans [14], and metal cations [11][12].

••-synuclein forms complexes with unstructured polycations [16] and the highly basic histone H1 [15] to form oligomers. Antony et al. [17] have recently reported that the cellular polyamines putrescine, spermidine, and spermine accelerate the aggregation and fibrillation of ••-synuclein to a degree that increases with the total charge, length, and concentration of the polyamine.

Polyamines are cellular stabilizers of nucleic acids and membranes and are essential for growth and differentiation. They mediate local and systemic functions in the central nervous system and are involved in neurodegenerative processes [18]. At high intracellular levels, spermidine and spermine are toxic [19] and can produce oxidative intermediates during polyamine retroconversion. For these reasons, it has been suggested [17] that at physiological concentrations and in a cellular context, these natural compounds may modulate the propensity of ••-synuclein to form fibrils.

The oligomerization process of ••-synuclein at structural levels might be the key study to understand the protein complexity and the key steps of Parkinson Diseasis. NMR is therefore an appropriate tool, since it provides information about conformational rearrangements and dynamics information at atomic resolution.

The observation of early step protein conformation changes due to environmental influences is essential to prevent oligomerization and fibrillation. Nevertheless ••-synuclein has a very large numbers of partners and therefore assays based on fibrillation limit are usually inappropriate for screening purpose.

FTIR spectroscopy, NMR spectroscopy, and circular dichroism (CD) are standard techniques for secondary structure characterization. However those techniques detect secondary structures based on • -sheet and/or • -helix as shown in Figure 11 for CD spectroscopy and therefore are not suitable for native unfolded protein like • -synuclein.

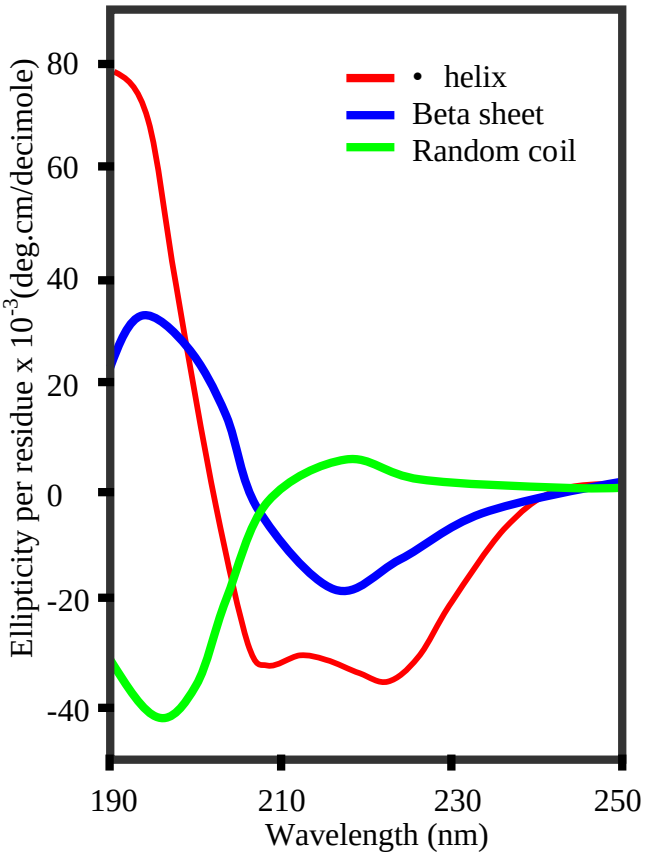
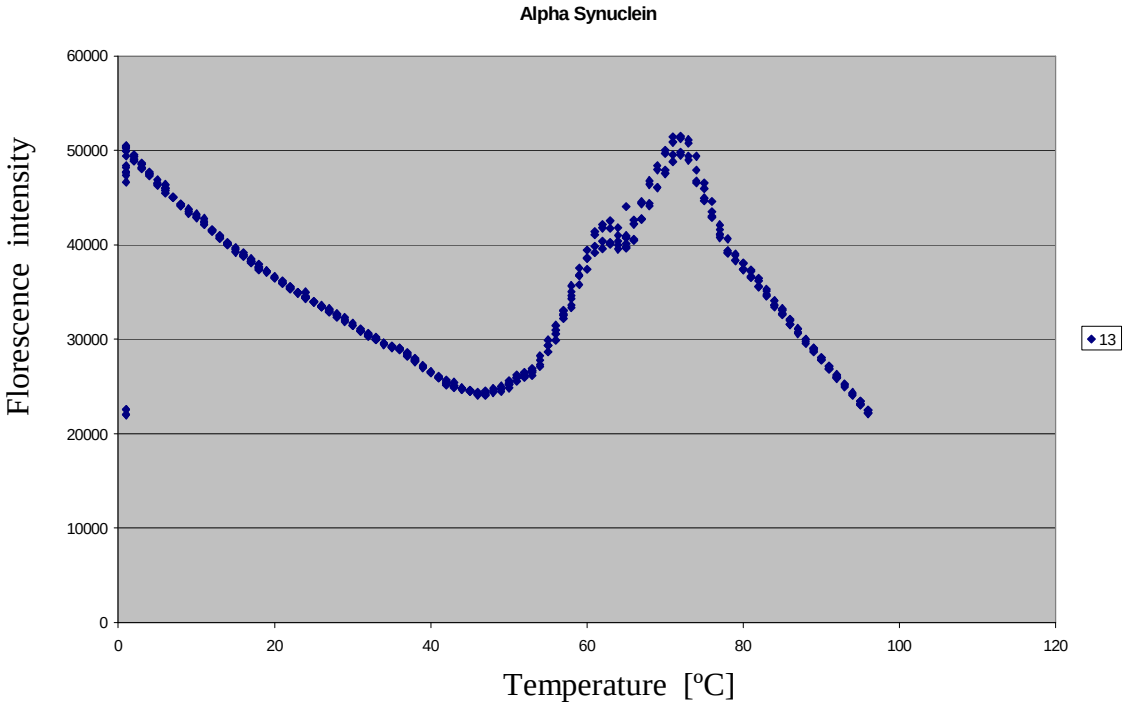


Figure 11: Reference CD spectra for red: • helix, blue: beta sheet, and green: random coil secondary structure.

Since TSA is based on structure rearrangements independent from secondary structure, it can readily detect any remaining fold in natively unfolded proteins. TSA is for that reason a unique tool for fast and convenient screening of natively unfolded protein.

As published recently [10], it seems that •-synuclein co-exists in equilibrium with partially folded conformers. In order to verify these results, TSA were performed on •-synuclein at several GdmCl concentrations. At 0M GdmCl, a melting temperature is observable (Figure 12) and confirms the existence of native fold in •-synuclein. At already 1M GdmCl, the protein seems to lose this fold since no Tm could be detected (data not shown).



Graph 12: TSA of native unfolded protein •-Synuclein at pH=6.4

Some screening experiments with polyamines with different incubation times have been measured and they show indeed some influence on T_m. Those experiments are still in a very early step and are therefore not presented in this report.

Nevertheless, this measurement shows the potential of TSA to investigate high throughput *in-vitro* conditions for •-Synuclein fibril formation and, in general, for characterizations of natively unfolded proteins.

13. Conclusion and future developments

In this chapter, we have demonstrated that biophysical measurements methods can be employ for different purpose, extending the range of potential applications.
Thermal shift assays is use extensively for protein ligand interaction in the pharmaceutical industry, illustrating the efficiency and reproducibility of this method.
The GB1 thermal unfolding/refolding experiment suggests interesting application in the protein folding field. NMR has the advantage of providing structural information at atomic resolution, but the thermal range for NMR measurement is unfortunately limited to the heat sensitivity of the electronic devices.

TSA permits extensive screening possibilities of protein denaturating/renaturating agents, providing condition at which the protein folding is still maintained but enough affected to able further denaturation via temperature increase.
Naturally GB1 is a very stable protein, with very convenient properties and similar studies on complete domain is probably not often applicable. But this approach is more relevant for unfolding/refolding experiments on protein motifs, which are much smaller.
This approach could also provide folding advancement and parameters influencing prtein refolding from inclusion bodies.

We would also point the completely new possibility of native unfolded protein characterization, not possible with other biophysical techniques at high throughput. This is also interesting to notice that native unfolded proteins are not necessarily a random coil peptide chain freely tumbling in solution, but they can adopt structural shape via amino acid interactions, like in case of • Synuclein and it is detectable with TSA. This open great opportunity for high throughput screening for fibril formation, reducing the ligand/agent candidates for NMR experiments

14. References

[1]. M. W. Pantoliano, E. C. Petrella, J.D. Kwasnoski, V. S. Lobanov, J. Myslik, E. Graf, T. Carver, E. Asel, B.A. Springer, P. Lane, F.R. Salemme, High-density miniaturized thermal shift assay as a general strategy for drug discovery, J. Biomol. Screen 6 (2001) 429-440.

[2]. U. B. Ericsson, B. M. Hallberg, G. T. DeTitta, N. Dekker, P. Nordlund, Thermofluor-based high-troughput stability optimization of proteins for structural studies, J. Anal. Biochem. 357 (2006) 289-298.

[3]. M. C. Lo, A. Aulabaugh, G.X. Jin, R. Cowling, J. Bard, M. Malamas, G. Ellesrad, Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery, Anal. Biochem. 332 (2004) 153-159.

[4]. Selenko P, Serber Z, Gadea B, Ruderman J, Wagner G. Quantitative NMR analysis of the protein G B1 domain in Xenopus laevis egg extracts and intact oocytes. Proc Natl Acad Sci U S A. 2006 Aug 8;103(32):11904-9.

[5]. Ding K, Louis JM, Gronenborn AM. Insights into conformation and dynamics of protein GB1 during folding and unfolding by NMR. J Mol Biol. 2004 Jan 30;335(5):1299-307.

[6]. Tcherkasskaya O, Knutson JR, Bowley SA, Frank MK, Gronenborn AM. Nanosecond dynamics of the single tryptophan reveals multi-state equilibrium unfolding of protein GB1. Biochemistry. 2000 Sep 19;39(37):11216-26.

[7]. George JM, Jin H, Woods WS, Clayton DF. Characterization of a novel protein regulated during the critical period for song learning in the zebra finch. Neuron (1995) 15, 361-372.

[8]. Vladimir N.Uversky, Jie Li, Anthony L. Fink Evidence for a partially folded intermediate in •-synuclein Fibril Formation; J. Biol. Chem. 2001, 276(14), 10737-10744.

[9]. Kim HY, Heise H, Fernandez CO, Baldus M, Zweckstetter M Correlation of Amyloid fibril beta-structure with the unfolded state of •-synuclein. Chem Bio Chem 2007; 8: 1671–1674.

[10]. Massimo Sandal, Francesco Valle, Isabella Tessari, Stefano Mammi, Elisabetta Bergantino, Francesco Musiani, Marco Brucale, Luigi Bubacco, Bruno Samorì. Conformational Equilibria in Monomeric •-Synuclein at the Single-Molecule Level; PLoS Biology 2008;6(1).

[11]. Uversky VN, Lee HJ, Li J, Fink AL, Lee SJ Stabilization of partially folded conformation during •-synuclein oligomerization in both purified and cytosolic preparations. J Biol Chem (2001a) 276: 43495–43498.

[12]. Hoyer W, Antony T, Cherny D, Heim G, Jovin TM, Subramaniam V. Dependence of •-synuclein aggregate morphology on solution conditions. J Mol Biol (2002) 322: 383–393.

[13]. Munishkina LA, Phelan C, Uversky VN, Fink AL. Conformational behavior and aggregation of •-synuclein in organic solvents: modeling the effects of membranes. Biochemistry (2003) 42: 2720–2730.

[14]. Cohlberg JA, Li J, Uversky VN, Fink AL. Heparin and other glycosaminoglycans stimulate the formation of amyloid fibrils from •-synuclein in vitro. Biochemistry (2002) 41:1502–1511.

[15]. Goers J, Manning-Bog AB, McCormack AL, Millett IS, Doniach S, Di Monte DA, Uversky VN, Fink AL. Nuclear localization of •-synuclein and its interaction with histones. Biochemistry (2003) 42: 8465–8471.

[16]. Goers J, Uversky VN, Fink AL. Polycation-induced oligomerization and accelerated fibrillation of human •-synuclein in vitro. Protein Sci. (2003) 12: 702–707.

[17]. Antony T, Hoyer W, Cherny D, Heim G, Jovin TM, Subramaniam V. Cellular polyamines promote the aggregation of •-synuclein. J Biol Chem (2003) 278: 3235–3240.

[18]. Morrison LD, Kish SJ. Ornithine decarboxylase in human brain: influence of aging, regional distribution, and Alzheimer's disease. J Neurochem (1998) 71: 288–294.

[19]. Auvinen M, Passinen A, Andersson LC, Holtta E. Ornithine decarboxylase activity is critical for cell transformation. Nature (1992) 360: 355–358.

Chapter 3

Reverses micelles preparation via phase transfer

15. Summary

The NMR community works actively on methods extending range of protein accessible for NMR studies. Factors influencing the physical properties have extensively studied and are nowadays perfectly understand. Viscosity influences the general tumbling of proteins, affecting the relaxation properties, which provoke strong signal broadening. Extensive protein deuteration is well established and improves the relaxation properties for large protein.

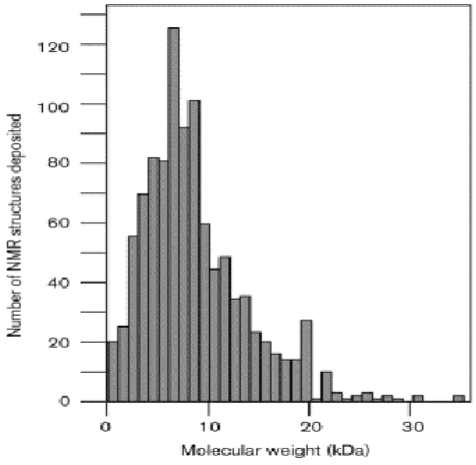
Low viscosity solvents have been suggested to ameliorate the tumbling of large properties. The protein is encapsulated in reverse micelles, protected from the hydrophobic solvent. Several successful strategies have been used to achieve encapsulation, but we will focus on phase transfer, rarely used for this purpose. The aim of this work was to establish a simple protocol widely useable by NMR community. We present in this chapter the preliminary results, suggesting future successful applications for NMR.

16. Introduction

Nulear magnetic resonance of proteins is an efficient method for studying protein dynamics, but it is suffers of low sensitivity in case of large proteins. Therefore NMR spectroscopy has been limited to relatively small proteins or protein domains. Large proteins are subject to severe line broadening due to adverse

relaxation properties. This is the main reason why NMR structures deposited in the PDB database generally do not exceeded a molecular weight of 35kDa, and even very rarely 20 kDa (Figure 1).

The magnetization of large proteins relaxes faster, resulting in less time for signal detection. Extended deuteration [2], optimized pulse sequences [1], and/or isotopic labeling strategies [2] have improved and extended the size limit of proteins suitable for NMR. By using these techniques it has been possible to study proteins in complex with the 900 kDa chaperone GroES-GroEL [3].



Molecular weight distribution of the NMR structures deposited in the PDB as 12/1997

P. Gunter, *Q. Rev. Biophys*, 1998, v31, p145

Relaxation times provide information on the atomic level about motions within a molecule. In case of protein dynamics studies, the ^{15}N isotope is the preferred nucleus, because its relaxation time is simple to correlate to molecular motions. The T1 and T2 relaxation times are measured using HSQC based experiments. Motions in the 10 ps to 10 ns range are the most convenient to detect, but slower motion rates (10 μs to 100 ms) can also be studied.

Since nitrogen is mainly found in the backbone chain of the protein, the relaxation time of ^{15}N reflects the motion of the backbone chain, which is unfortunately the less flexible part of the protein [22]. Relaxation measurements on ^{13}C improve the studies of dynamics due to the higher flexibility of the side chains.

Relaxation rates are dependent on the correlation time, which is directly influenced by solvent viscosity •, temperature T and volume/shape V of the protein, as shown in equation 1.

$$(1)\ T_c = \frac{\eta V}{kT} \quad \text{, (for spherical protein).}$$

Since the correlation time depends on the temperature, NMR experiments are often run at higher temperature. Since viscosity itself also decreases with an increase of temperature, the temperature increase affects equation 1 twice, resulting in a strong decrease of the correlation time.

Unfortunately protein folding is also temperature dependent, resulting from a decrease of stability of secondary structures and therefore promoting partial unfolding and aggregation. The range of proteins tolerating a significant temperature increase is unfortunately quite small, which strongly limits this strategy for NMR measurements.

Independently from temperature, viscosity decrease can be a strategy to influence the correlation time. The solubilization of proteins in organic solvents using ionic and non-ionic surfactants has been well documented [15], [16], [17], [18]. A common technique for this purpose employs reverse micelles, whereby the protein is sequestered within the aqueous core [19], [20] or is localized at the surfactant interface [21], and is therefore in an environment that is distinct from the continuous organic phase. (Figure 2)

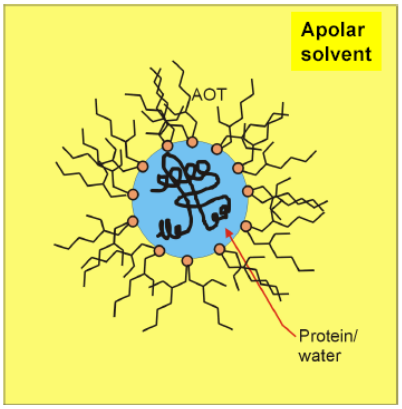


Figure from W. Westler at NMRFAM

Figure 2: Representation of protein encapsulate in reverse micelles solubilized in low viscosity solvent.

This aqueous core contains the protein buffer and therefore salts, buffers, and/or reducing agents. It provides a protected and suitable aqueous environment required for the native protein fold.

The extremely low viscosity of alkanes favors fast overall tumbling of the reverse micelles, resulting in strong improvement of the protein's relaxation properties and therefore signal-to-noise. The protein behaves like a much smaller protein (• 10-15 kDa for original size of 100 kDa), even allowing measurements without perdeuteration.

Much larger species, such as bacterial cells [16] and soy bean mitochondria [15], have been solubilized in reverse micelles as well, but required the addition of much more surfactant.

Organic solubilization of proteins has been achieved using the liquid-liquid (phase-transfer), solid-liquid and injection-extraction methods, in which protein from an aqueous phase is extracted into an organic phase containing very low concentrations of a surfactant [4] [17] (Figure 3).

The most commonly used surfactant for organic solubilization is Aerosol OT (AOT) in combination with isooctane. AOT is an anionic amphiphilic surfactant and is particularly well suited for organic solubilization of proteins in reverse micelles due to its high solubility in hydrophobic solvents. AOT also has a high

propensity to incorporate large amounts of water into the organic phase during phase extraction.

The feasibility of using reverse micelles in biological NMR studies has first been demonstrated and validated in 1998 [5] using Ubiquitin, prepared in a variety of alkane solvents. Since then several experimental protocols with anionic as well as cationic surfactants for reverse micelles preparation for NMR purposes have been published [6-14].

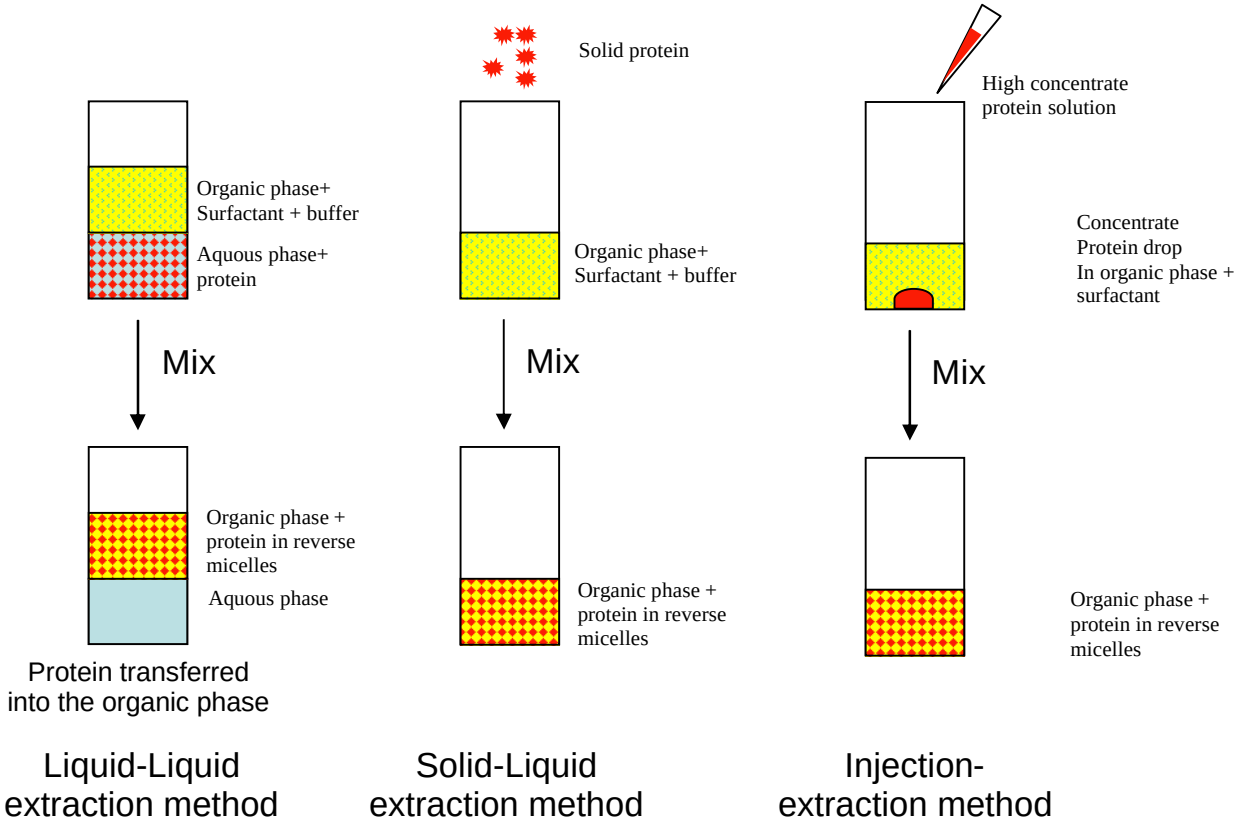


Figure 3: Protein extraction methods for reverse micelles preparation.

Nevertheless most of the protocols developed in the last decade have focused on solid-liquid extraction or the injection-extraction method. These strategies have mostly the advantage that the quantities of buffer and protein transferred into the reverse micelles can be easily controlled and adjusted, which greatly simplifies the overall procedure.

The buffer conditions as well as the buffer volume solubilized in the reverse micelle are influencing the local secondary structure of the encapsulated protein. By using the phase transfer method, the amount of transferred water/buffer cannot be controlled, usually increasing the total water content; therefore the protein secondary structure is strongly affected in the reverse micelle.

While this procedure is more favorable for most protein samples, this method is rarely used and mostly only for purification via 2 extraction steps: extraction/re-extraction in continuous process [23][24]. This approach is incompatible with sample preparation for NMR purpose since most of the protocols extract a very low protein concentration via a continuous process. The final concentration required for NMR experiments is in the 150 to 250 μM range, so therefore several mg of protein would have to be encapsulated.

Although several reverse micelle sample preparations for NMR purposes have been successfully established and published [10-14], the size of the encapsulated proteins rarely exceeded 60-70 kDa, since NMR signal overlap becomes more severe with increasing protein size. This problem can be overcome by using the in-vivo segmental labeling method presented previously in this thesis. Combining the two strategies will allow NMR measurements of protein sizes over 100 kDa.
This chapter describes an improved version of the phase transfer method for NMR sample preparation. We tried to focus on a protocol as simple as possible, without special apparatus and suitable for a large number of proteins. We will describe the

extraction process, the local protein secondary structure recovery strategy, as well as reverse micelle purification. We will also suggest possible approaches for finalizing the sample preparation for NMR purposes.

17. Material and methods

17.1. Proteins

In the last decade, cytochrome C has turned out to be a reference for reverse micelle preparation, which is certainly due to its physical properties. Cytochrome C has a size of 12 kDa, is highly soluble, can be lyophilized and is readily commercially available. But probably the most interesting characteristic of this protein is its ability to bind heme, which confers to the protein a red color. In case of protein denaturation due to direct contact with the hydrophobic solvent, the iron ion is released and the protein looses its red color. The efficiency of protein transfer can be easily evaluated without extensive physical measurements.

Regarding the purpose of this protocol to encapsulate large proteins, our interest in cytochrome C was limited. A large number of proteins with various sizes have been studied extensively and fully reported [25]. Here, hemoglobin has been chosen because of its physical properties similar to cytochrome C. Hemoglobin is a homo-tetramer containing an iron binding pocket in each monomer unit. The overall size of the complex is 64kDa. The protein is commercially available in a lyophilized form and it is extracted from bovine blood. In case of unsuitable reverse micelle extraction conditions, the complex is destabilized and each unit releases the iron ion, inducing a color shift from red to colourless.

17.2. Surfactant

A large range of surfactants are commonly used for protein encapsulation. As mentioned before, AOT is the most widely used surfactant for this purpose and therefore was mainly used in our assay. Nevertheless Peterson et al. [9] have demonstrated the advantage of using a mixture of anionic, cationic and neutrally charged surfactants (Figure 4) for protein micelle preparation. They claim that the local charge of the protein chain is influenced by the charge of the surfactant, especially if both the amino acid and surfactant have the same charge. This may explain the influence of water content on local secondary structure perturbation. By using such mixtures the protein surface charges are equilibrate with respectively charged surfactant. This mixture system was successfully applied to encapsulate cytochrome C in its folded form [9], which is not possible by using only AOT. (Figure 4)

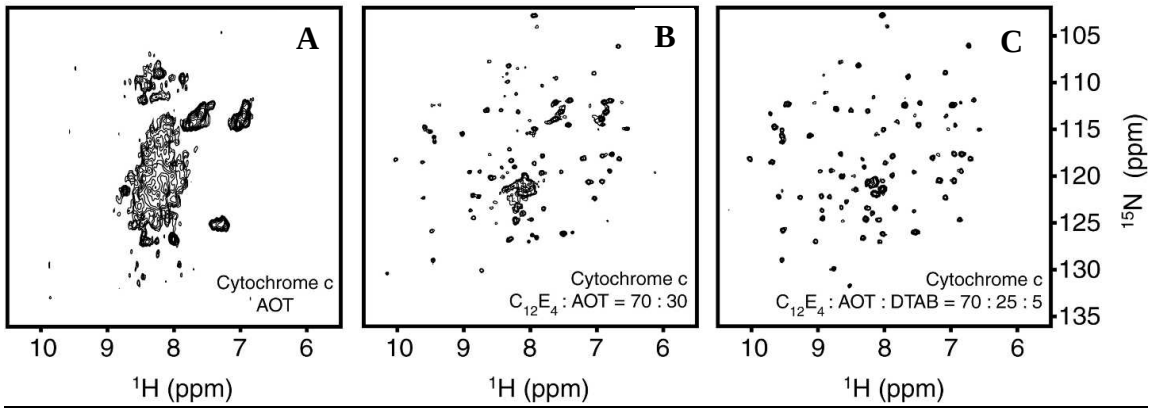
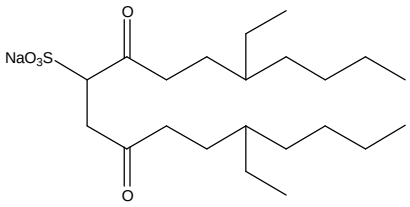


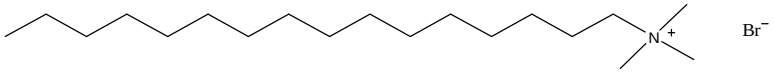
Figure 4: [9] Performance of novel surfactant mixtures for NMR spectroscopy of encapsulated proteins dissolved in low-viscosity fluids. (A) ^{15}N-HSQC spectrum of cytochrome c solubilized in 50 mM sodium acetate (pH 5.0), and encapsulated with a water loading of 10 in 100% AOT reverse micelles in pentane. (B) ^{15}N-HSQC spectrum of cytochrome c solubilized in 50 mM sodium acetate (pH 5.0), and encapsulated with a water loading of 10 in 70% $C_{12}E_4$/30% AOT reverse micelles in pentane. (C) ^{15}N-HSQC spectrum of cytochrome c solubilized in 50 mM sodium acetate (pH 5.0), and encapsulated with a water loading of 10 in 70% $C_{12}E_4$/25% AOT/5% DTAB in pentane. Spectra were collected on a 600-MHz Varian Inova-class spectrometer equipped with a cold probe.

Therefore we used the same surfactant mixture system:

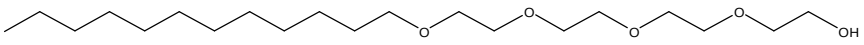
Anionic: AOT (sodium bis(2-ethylhexyl)sulfosuccinate)



Cationic: CTAB (cetyltrimethylammonium bromide)



Neutral : $C_{12}E_4$ (dodecyl tetraethylene glycol ether)



17.3. Solvent

The length of the alkane chain does not strongly influence the reverse micelle formation [15]. Hexane is liquid and has a relatively slow evaporation rate compared to pentane, which makes it easier to handle. For that reason we used hexane for testing our reverse micelle preparations.

Reverse micelle preparation was done at room temperature and any temperature changes were avoided. The surfactant mixture was always dissolved in the organic phase. The protein phase was prepared by resolubilization of the lyophilized protein into the buffer of interest.

The volume ratio between the organic phase and the water phase was 1:1 in case of the cytochrome C experiments.

17.4. CD Spectroscopy

CD spectroscopy in the "far-UV" spectral region (190-250 nm) can be used to determine the secondary structure content of proteins. At these wavelengths the chromophore is mainly the peptide bond, and the CD signal arises when it is located in a regular, folded environment.

• -helix, beta-sheet, and random coil structures each give rise to a characteristic shape and magnitude of CD spectrum. This is illustrated in Figure 5, which shows spectra for poly-lysine in these three different conformations. The approximate fraction of each secondary structure type that is present in any protein can thus be determined by analyzing its far-UV CD spectrum as a sum of fractional multiples of such reference spectra for each secondary structure type.

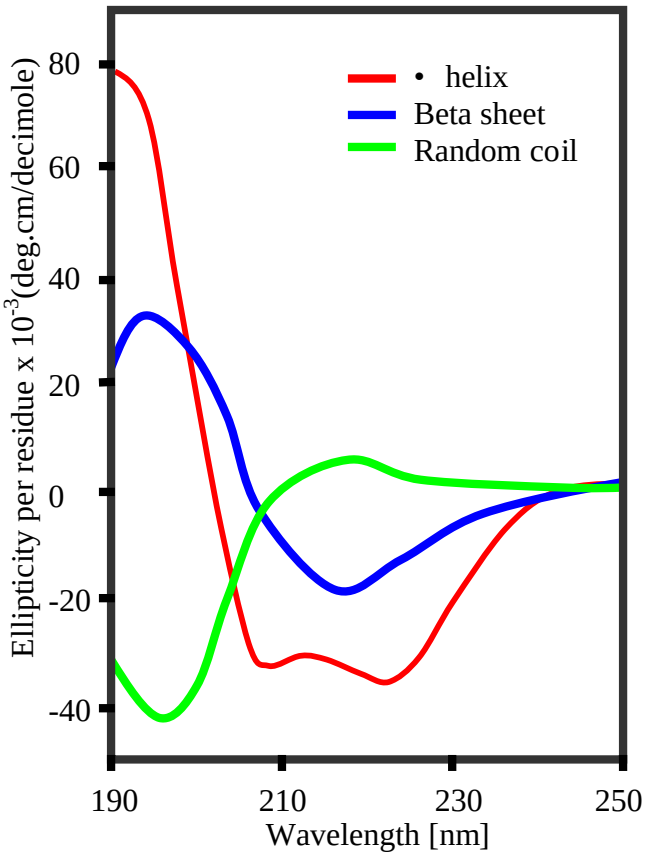


Figure 5: Reference CD spectra for red: • helix, blue: beta sheet, and green: random coil secondary structure.

Like all spectroscopic techniques, the CD signal reflects an average over the entire molecular population. Thus, while CD can determine that a protein contains about 50% •-helix, it cannot determine which specific residues are involved in the •-helical portion.

Far-UV CD spectra require 20 to 200 µl of solution containing 1 mg/ml to 50 µg/ml ofprotein, in any buffer which does not have a high absorbance in this region of the spectrum (high concentrations of DTT, histidine, or imidazole, for example, cannot be used in the far-UV region). The secondary structure of cytochrome C was followed at RT using CD spectroscopy with a JASCO J-715 spectropolarimeter (wavelength: 195-260 nm).

18. Results and discussions

First of all, since the local secondary structure of the protein can be affected by surfactant charge interactions, the CD spectra of fully folded cytochrome C in solution (Figure 6) is required as a reference for further experiments.

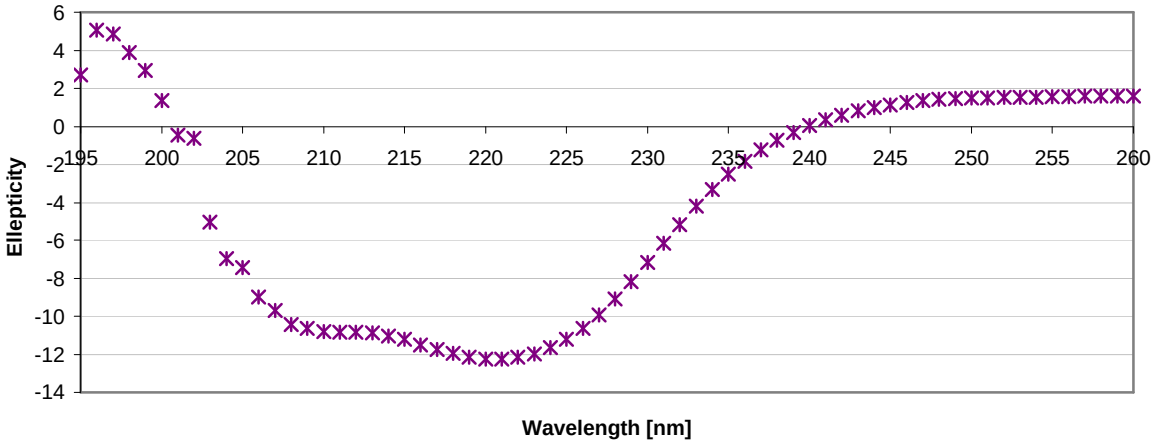


Figure 6: Circular dichroism UV spectrum of cytochrome c in 50mM TRIS, 150 mM NaCl, pH=9,6.

This spectrum shows two characteristic minima at 209 and 222 nm, which is typical for •-helical secondary structure (Figure 6). The NMR structure of cytochrome C is available at the Protein Data Bank (PDB) under the ident code *1j3s* (Figure 7). This structure clearly shows only •-helices. The DSSP secondary structure analysis [26] gives 42 % helical content (5 helices; 44 residues) and 1 % beta-sheet (2 strands; 2 residues).

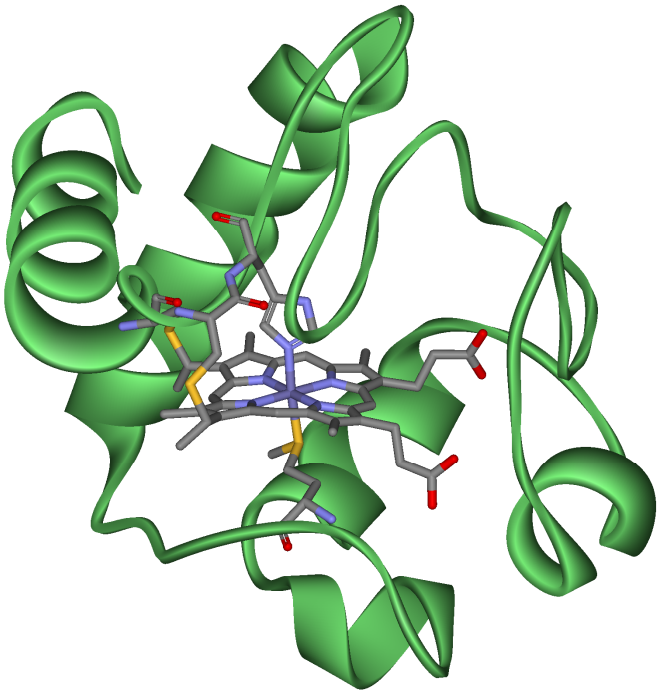
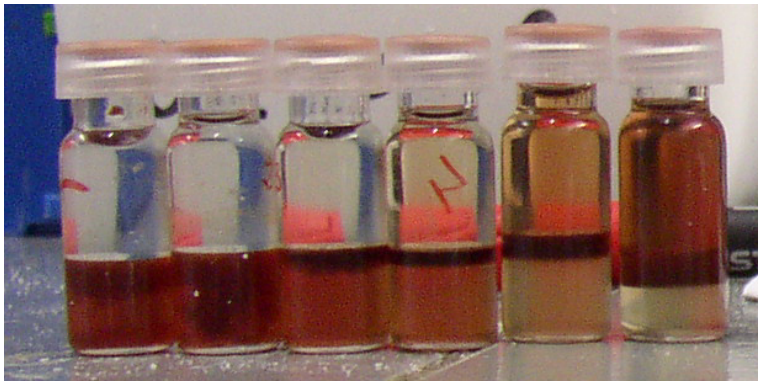


Figure 7: Cytochrome c with heme. Structure available at PDB under ident code 1j3s.

In order to characterize the influence of surfactant concentration on protein transfer, we have set up the following experiment (Table 1).

Table 1: Influence of AOT concentration

Sample	1	2	3	4	5	6
AOT [ ] in mM	0	25	50	75	100	150
% of Cyt C transferred	1	2	15	35	60	90



1 2 3 4 5 6

Figure 8: Visual results of cytrochrome C extraction at different AOT concentrations. Room temperature, 400 rpm / 30 s homogenization. Upper phase: organic phase (hexane), lower phase: 200 µM protein in buffer (50 mM TRIS, 50 mM NaCl, pH 9,6).

Visual results of the protein transfer are shown in Figure 8. The dark red region situated between the organic and water phase consists of aggregates of cytochrome C which has retained its heme molecule. We can also observe that some residual cytochrome C remains in the water phase. Figure 9 illustrates the influence of AOT concentration on the extraction yield and demonstrates that surfactant concentration is critical to transfer the protein with high efficiency.

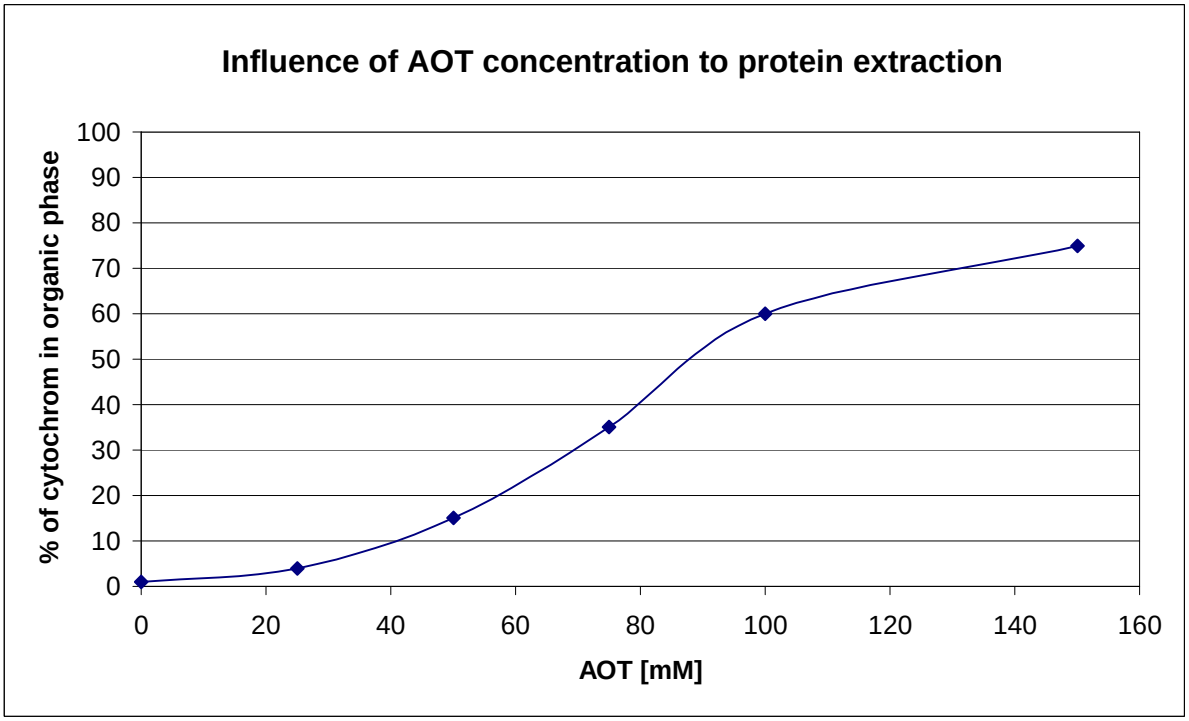


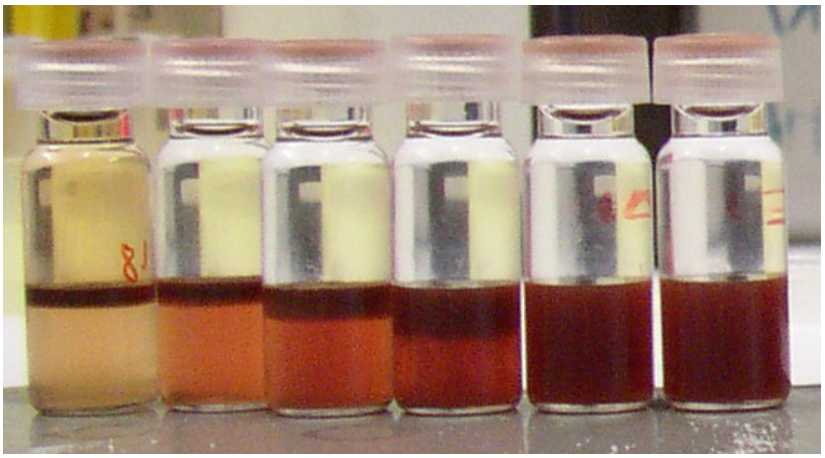
Figure 9: Influence of AOT concentration on percentage of cytochrome extracted.

Earlier studies have shown the influence of salts to extraction efficiency [24]. Those studies claim that an excess of salts will interact directly with the surfactant and decrease the extraction efficiency. Nevertheless salts are often required to maintain proteins soluble, prevent aggregation and promote formation of a two-phase system [9]. Therefore it is necessary to find a good compromise in salt

concentration. The next experiment illustrates this phenomenon (Table 2 and Figure 10).

Table 2: Influence of NaCl concentration

Sample	1	2	3	4	5	6
NaCl in mM	0	100	200	400	600	800
% of Cyt C transferred	50	15	7	1	0	0



1 2 3 4 5 6

Figure 10: Visual results of cytrochrome C extraction at different NaCl concentrations. Room temperature, 400 rpm / 30 s homogenization. Upper phase: organic phase (hexane + 50 mM AOT), lower phase: 200 µM Protein in buffer (50 mM TRIS, pH 9,6)

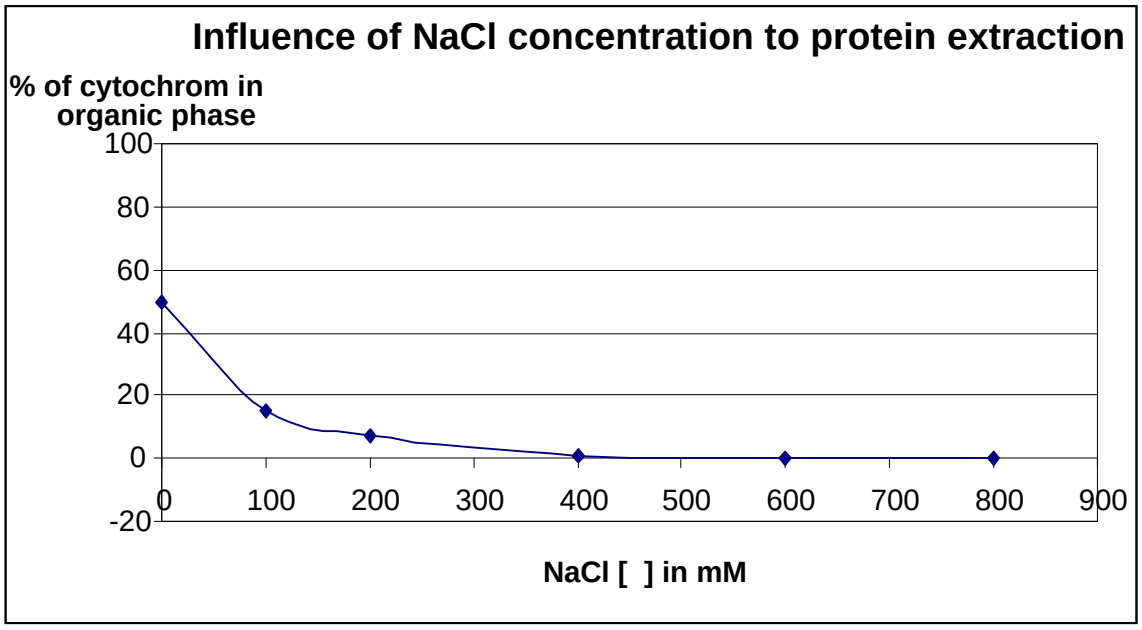


Figure 11: Influence of NaCl concentration on percentage of cytochrome extracted**.**

As a result, high salt concentration is indeed clearly inhibiting the extraction process (Figure 11). Nevertheless, we were able to transfer over 95% of cytochrome C with 150 mM AOT in the presence of 100 mM NaCl.

The aggregation observed during the transfer phase has been strongly reduced after we reduced the vortex speed to 200 rpm. Those aggregates were presumably formed by contact with hexane in case where the protein was not protected with surfactant.

18.1. Recovery of partially unfolded transferred Cytochrome C.

Unfortunately, as has beem demonstrated before [9], cytochrome C is unable to maintain its secondary structure in AOT reverse micelles. Nevertheless it seems to be possible to encapsulate cytochrome if the surfactant charges are properly equilibrated [9].

In the following experiment, we used a surfactant mixture of 60% AOT + 40% CTAB to extract cytochrome C. The lyophilized protein was dissolved in 50 mM TRIS, 100 mM NaCl, pH=9,6. The total amount of surfactant was 150 mM dissolved in hexane. The samples were prepared by adding an equal volume of surfactant-containing organic phase to a protein-containing aqueous phase.

One of the main problems in the phase transfer method is the control of the water amount in the reverse micelles. Since the water concentration directly affects the secondary structure, it is therefore difficult to extract the protein in a fully folded form. This phenomenon is illustrated in Graph 4 with the dark blue curve. The CD spectrum was recorded right after the extraction process. This curve has a minimum at 206 nm. This curve shape can be associated to a partially unfolded protein, since the shape doesn't fit a random coil profile. 1D NMR measurements have demonstrated (data not shown) that the reverse micelle water content was unpredictable and high.

This problem can be overcome by absorbing a large amount of water via dehydratation reagents like calcium chloride or sodium sulfate from the organic

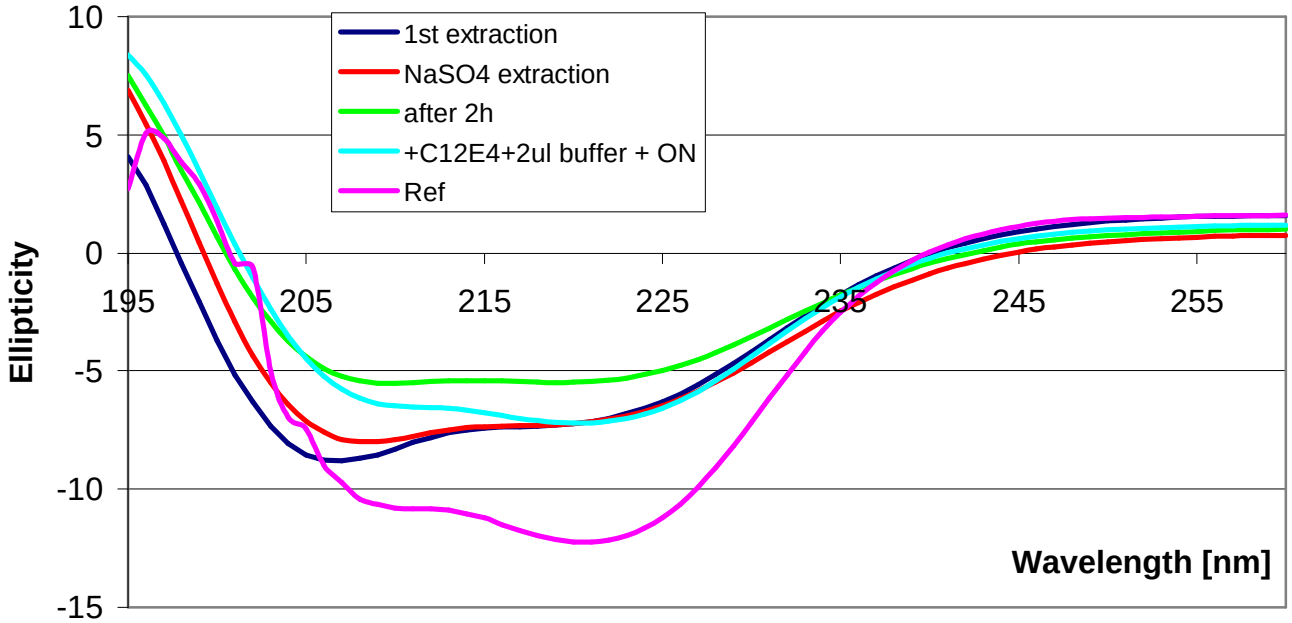


Figure 12: Circular dichroism UV spectra of cytochrome c at different extraction step at RT. Dark blue: direct after extraction; Red: after sodium sulfate extraction; Green: after 2 hour equilibrate; blue: after adding 124ul $C_{12}E_4$ + 2µL TRIS buffer and equilibrate overnight; Purple: reference CD of Cytochrome C in buffer.

phase after separation of the water phase.

Recent studies have demonstrated [27] that small molecules like denaturant, redox reagents, water etc can exchange from the reverse micelle to the hexane phase and vice versa. This transfer was occurring through a dialysis membrane with 3.5kDa cutoff, so that reverse micelles containing protein and reverses micelles containing reagents could not interact directly to each other.

This strategy was used successfully for refolding chemically denaturized proteins.. The denaturizing reagent concentration can be decrease via dialysis allowing the protein to refold [12]. They have also observed that water was transferred from the inner chamber to the outer chamber, reducing the water amount seven fold in 4 hours.

Excess of sodium sulfate was sufficient to extract the excess of water contained in the reverse micelles. The sodium sulfate was added slowly and stepwise until it does not form aggregates anymore. Since the encapsulated water amount seems to be unpredictable in the phase transfer method, we have not recorded the amount of sodium sulfate necessary to absorb the water excess. In Graph 4 the red curve was recorded right after this absorption step. The protein is still mostly unfolded but the shape of the curve has changed, suggesting that the reverse micelles have been influenced by this extra step. It is important to remember that deficiency of water promotes direct interaction of the hydrophilic charged head of the surfactant with the protein. This interaction can directly affect the secondary structure. Therefore it is difficult to determine if the water content is below or above the optimal concentration by measuring the CD spectrum. Nevertheless 1D NMR spectra (not shown) have confirmed the strong decrease of the total water amount.

The equilibrium of the reverse micelles is also strongly affected by this extraction step since the average size of the micelles decreases strongly with water deficiency. For that reason an equilibration pause of at least two hours is required to stabilize the mixture (Figure 12- light green curve). This curve shows two minima at the same characteristic wave lengths that the reference spectra of

cytochrome C. However the ellipticity ratio of the two wavelength is still not identified to the • helix shape.

In the next step, we have added 124µL $C_{12}E_4$ as suggested by Peterson et al. [9]. This amount corresponds to a final concentration of 5%. This new mixture is supposed to facilitate and adjust the charge interaction of the surfactant with the protein. Then protein buffer was added 1µL stepwise to the reverse micelle mixture until the shape of the CD spectrum merges with the reference spectrum. It is also important to remember that the system requires time to reach equilibrium and also if the measurements suggest proper folding after adding the buffer, the shape of the curve can change after several hours. Therefore we recommend carefully adding buffers and always waiting for equilibrium. In our hand if the CD spectrum remains unchanged overnight, we considered the reverse micelles in equilibrium (Figure 12 – light blue curve). It is interesting to notice that the final curve very high similarity to the reference curve.

In our knowledge this is the first time that cytochrome C was successfully encapsulated and folded into reverse micelles using the phase transfer method. This strategy should be applicable to any protein, and to demonstrate the applicability of this protocol we have continued working with hemoglobin (64kDa).

18.2. Efficient extraction of Hemoglobin

Hemoglobin is a good candidate for reverse micelle preparation given its high molecular weigh, classifying the protein as large and therefore difficult examine with NMR without extensive deuteration.

By switching from 12kDa to 64kDa we observed drastic changes in term of extraction efficiency.

Several studies in reverse micelles extraction [28][29] have focused on the specific interaction between the surfactant and the protein. The isoelectric point seems to drive the nature of this interaction. Below this point, the protein is positively charged and the protein will be attracted by negatively charged surfactant. In contrast above the isoelectric point the protein is negatively charged and the protein is attracted to positively charged surfactant. So the transfer of a protein solubilized at a pH below its isoelectric point will not occur with a positively charged surfactant.

Since the protein-surfactant electrostatic interaction is the driving force for transfer [30], the pH in the protein buffer will directly affect the transfer efficiency. The global charge of the protein increases when the gap between the pH and the pI increases [31].

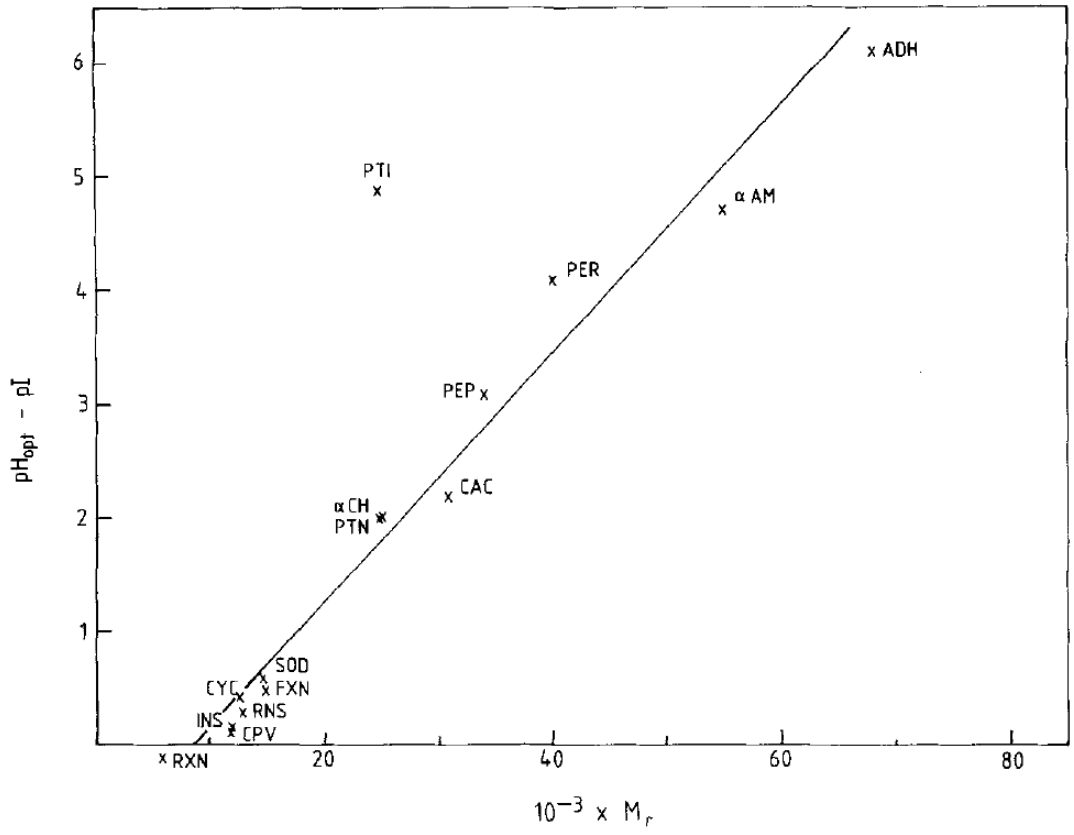


Figure 13: The difference between the pH where maximal solubilization occurs (pH_{opt}) and the isoelectric point (pI) as a function of Molecular weight M, ,for *trioctylmethylammonium chloride reversed micelles.* For abbreviations and data, see Table 3 [31]

Table 3: Compilation of data on protein solubilization in reversed micelles.PI values were measured at 4°C [31]

Protein	Notation	$10^{-3} \times M_r$	% S	pI	pH_{opt}	Transfer	pH_{opt}-pI
						%	
Alcohol dehydrogenase	ADH	68	58	6.8	12.9	40	6.1
α-Amylase	α-AM	55		5.3	10.0	95	4.7
BSA	BSA	68		4.9		0	
Carbonic anhydrase	CAC	31	71	5.8	8.0	32	2.2
α-Chymotrypsin	α-CH	25		8.5	10.5	6	2.0
Cytochrome *c*	CYC	12.5	54	9.9	10.3	55	0.4
Flavodoxin	FXN	15	35	3.8	4.3	70	0.5
Hemoglobin	MHB	60–64	80	5.4–6.5	10.5	15	5.1–4.5
Hexokinase	HEX	100		4.7–5.0		0	
Insulin	INS	12		5.9	6.0	5	0.1
Lysozyme	LYZ	12	97	10.9		0	
Parvalbumin	CPV	12	49	5.3	5.4	58	0.1
Pepsin	PEP	34		<1.0	4.1	8	3.1
Peroxidase	PER	40		6.9	11.0	35	4.1
Ribonuclease A	RNS	13.5	24	8.0	8.3	65	0.3
Rubredoxin	RXN	6	74	3.2	3.1	56	−0.1
Superoxide dismutase	SOD	14.5	87	5.3	5.9	4	0.6
Trypsin	PTN	25	61	10.0	12.0	25	2.0
Trypsin inhibitor	PTI	24.5	29	4.6	9.5	70	4.9

This phenomenon has been extensively study by Wolbert et al. [31]. They have studied uptake of 19 proteins as a function of pH of the aqueous solution. They found that the pH where the maximal solubilization occurs can be described by the relationship:

$$\mathbf{pH_{optimum} = pI + 0{,}11 \times 10^{-3}\, M_r - 0{,}97} \quad (1)$$

where M_r corresponds to the molecular weight (Figure 13).

So the larger the protein, the more charge is required to provide the energy required for the adaptation of the micelle size to the protein size. This relationship is applicable for CTAB as well as for AOT (Equation 1 applicable only for CTAB[31]).

So for each increase of 10kDa molecular weight, optimum uptake takes place about 1 pH unit further away from the isolelectric point. A protein with an M_r of 120 kDa and an isoelectric point of 1, will be transferred a pH 13 when CTAB is used as a surfactant. Based on these results, this is the limit of the size of an extractable protein.

This study suggests an optimal pH for extraction at 10.5 for hemoglobin with an extraction efficiency of 15%. So for our future experiments we could have followed these conditions. But the objective of this research is to achieve extraction of very large proteins (over 100kDa) in order to get the benefit of low viscosity for NMR studies.

The protein surface area exposed to the buffer increases (depending on the protein shape) with the protein size. Therefore the number of required charged residues increases with the protein size. By using anionic and cationic surfactants, negatively as well positively charged residues can interact with the surfactant mixture improving the uptake of the protein. Unfortunately this effect was limited in case of hemoglobin, also at the pH recommended for extraction. Over 75% of the protein could not get extracted and almost 50% of the total concentration aggregated.

The concentration range of surfactant used for reverse micelle preparation is 100-200mM, which corresponds to the upper concentration limit where the reverse micelles are not too close to each other. At higher surfactant concentration the viscosity will increase and affect the tumbling of the micelles.

Nevertheless in order to encasulate hemoglobin with high yield, we tried to increase the surfactant concentration.

At 500mM AOT/CTAB, 70% of hemoglobin was transferred into the organic phase. The aggregation level was still up to 20%. The volume ration between the water phase and the organic phase was 1:1. Transfer experiments were run at different pH and it seems that the pH is not influencing the uptake of the protein. This might be explained by the large excess of surfactant which can protect the protein from the hexane and transfer more efficiently the protein more efficiently. In order to demonstrate this theory, we reduced the volume ratio to 1:5 (1 volume protein solution: 5 volumes Hexane-AOT/CTAB) and were able to transfer 95% of hemoglobin with less than 2% aggregation. The total amount of hemoglobin extracted was 10mg starting with a protein solution of 1mg/mL. 200µL of this

protein solution was added to 1mL of Hexane-AOT/CTAB (500mM) in the bottom of an eppendorf tube. The two phases were first gently mixed by rotating 5 to 10 times and then vortex for 20s at 200rpm. The emulsion was spinned at 2000rpm for 2 min in a small desktop centrifuge at room temperature. This step separates the large excess of water which can not be taken up in the reverse micelles or the water amount instable in the micelles. The two phases were gently separated using micropipettes, and 200µL of protein buffer was added for the next extraction. The separated water phase is at this point colorless and the remaining organic phase turns red, indicating that the hemoglobin has been extracted.

These results demonstrate that this protocol seems to be suitable for the uptake of large protein into reverse micelle.

19. Conclusions

Reverse micelles preparations have high potential for NMR investigations, especially for studies of large proteins. This strategy has been extensively demonstrated and the proof of principle established more than a decade ago.

Nevertheless reverse micelles preparation requires a lot of expertise and the extraction methods successfully applied are unfortunately limited to a small range of proteins.

The protocol developed in this study is the first steppingstone for the larger application of reverse micelles preparation since it establishes a method suitable for a large number of proteins, depends on as few factors as possible and has a simple protocol.

Naturally, reverse micelles are mostly interesting for large multimeric proteins or segmentally labeled large proteins, so that measurements with little signal overlap can be achieved in spite of large proteins and complexes.

There are several problems which will have to be overcome to enable NMR measurements like reducing the large excess of surfactant and exchanging the hexane to ethane or propane.

Several strategies are now in progress like dialysis or lipophilic sephadex purification to separate empty micelles from protein filled reverse micelles.

20. Literature

[1]. Pervushin K, Riek R, Wider G, Wüthrich K, Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc Natl Acad Sci U S A. 1997 November 11;94(23):12366-71.

[2]. Markus MA, Dayie KT, Matsudaira P, Wagner G, Effect of deuteration on the amide proton relaxation rates in proteins. Heteronuclear NMR experiments on villin at 14T. J Magn Reson B. 1994 Oct;105(2):192-5.

[3]. Fiaux J, Bertelsen EB, Horwich AL, Wuthrich K. NMR analysis of a 900K GroEL-GroES complex. Nature. 2002 July 11;418(6894):207-211.

[4]. Melo EP, Aires-Barros MR, Cabral JM. Reverse micelles and protein biotechnology. Biotechnol Annu Rev. 2001; 7:87-129.

[5]. Wand AJ, Ehrhardt MR, Flynn PF., High-resolution NMR of encapsulated proteins dissolved in low-viscosity fluids. Proc Natl Acad Sci U S A. 1998 Dec 22;95(26):15299-302.

[6]. Valentine KG, Pometun MS, Kielec JM, Baigelman RE, Staub JK, Owens KL, Wand AJ. Magnetic susceptibility-induced alignment of proteins in reverse micelles. J Am Chem Soc. 2006 Dec 20;128(50):15930-1.

[7]. Pometun MS, Peterson RW, Babu CR, Wand AJ. Cold denaturation of encapsulated ubiquitin. J Am Chem Soc. 2006 Aug 23;128(33):10652-3.

[8]. Shi Z, Peterson RW, Wand AJ. New reverse micelle surfactant systems optimized for high-resolution NMR spectroscopy of encapsulated proteins. Langmuir. 2005 Nov 8;21(23):10632-7.

[9]. Peterson RW, Pometun MS, Shi Z, Wand AJ. Novel surfactant mixtures for NMR spectroscopy of encapsulated proteins dissolved in low-viscosity fluids. Protein Sci. 2005 Nov;14(11):2919-21. Epub 2005 Sep 30. Erratum in: Protein Sci. 2006 Apr;15(4):941.

[10]. Peterson RW, Lefebvre BG, Wand AJ. High-resolution NMR studies of encapsulated proteins in liquid ethane. J Am Chem Soc. 2005 Jul 27;127(29):10176-7.

[11]. Lefebvre BG, Liu W, Peterson RW, Valentine KG, Wand AJ. NMR spectroscopy of proteins encapsulated in a positively charged surfactant. J Magn Reson. 2005 Jul;175(1):158-62. Epub 2005 Apr 7.

[12]. Peterson RW, Anbalagan K, Tommos C, Wand AJ. Forced folding and structural analysis of metastable proteins. J Am Chem Soc. 2004 Aug 11;126(31):9498-9.

[13]. Babu CR, Flynn PF, Wand AJ. Preparation, characterization, and NMR spectroscopy of encapsulated proteins dissolved in low viscosity fluids. J Biomol NMR. 2003 Apr;25(4):313-23.

[14]. Ehrhardt MR, Flynn PF, Wand AJ. Preparation of encapsulated proteins dissolved in low viscosity fluids. J Biomol NMR. 1999 May;14(1):75-8.

[15]. Pfammatter N, Guadalupe AA, Luisi PL. Solubilization and activity of yeast cells in water-in-oil microemulsion. Biochem Biophys Res Commun. 1989 Jun 30; 161(3):1244-51.

[16]. Häring G, Luisi PL, Meussdoerffer F. Solubilization of bacterial cells in organic solvents via reverse micelles. Biochem Biophys Res Commun. 1985 Mar 29; 127(3):911-5.

[17]. Paradkar VM, Dordick JS. Purification of glycoproteins by selective transport using concanavalin-mediated reverse micellar extraction. Biotechnol Prog. 1991 Jul-Aug;7(4):330-4.

[18]. Bru R, Sánchez-Ferrer A, Garcia-Carmona F.A theoretical study on the expression of enzymic activity in reverse micelles. Biochem J. 1989 Apr 15;259(2):355-61.

[19]. Bratko D, Luzar A, Chen SH. (1988) Electrostatic model for protein / reverse micelle complexation. J. Chem. Phys., 89: 545-550.

[20]. Woll JM, Hatton TA. (1989) A simple phenomenological thermodynamic model for protein partitioning in reverse micellar systems. Bioprocess Eng., 4: 193-199.

[21]. Brandani S, Brandani V, Giacomo GD. (1994) A thermodynamic model for protein partitioning in reverse micellar systems. Chem. Eng. Sci., 49: 3681-3686.

[22]. Dayie KT, Wagner G, Lefèvre JF. Theory and practice of nuclear spin relaxation in proteins. Annu Rev Phys Chem. 1996; 47:243-82.

[23]. Cunha MT, Costa MJ, Calado CR, Fonseca LP, Aires-Barros MR, Cabral JM.Integration of production and aqueous two-phase systems extraction of extracellular Fusarium solani pisi cutinase fusion proteins. J Biotechnol. 2003 Jan 9;100(1):55-64.

[24]. Pires MJ, Cabral JM. Liquid-liquid extraction of a recombinant protein with reverse micelles. J Chem Technol Biotechnol. 1994 Nov;61(3):219-24.

[25]. Wolbert RB, Hilhorst R, Voskuilen G, Nachtegaal H, Dekker M, van't Riet K, Bijsterbosch BH. Protein transfer from an aqueous phase into reversed micelles. The effect of protein size and charge distribution. Eur J Biochem. 1989 Oct 1;184(3):627-33.

[26]. Kabsch W., Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983 Dec.

[27]. Ono T, Nagatomo M, Nagao T, Ijima H, Kawakami K. Nonaggregating refolding of ribonuclease A using reverse micellar dialysis. Biotechnol Bioeng. 2005 Feb 5;89(3):290-5.

[28]. Wu JZ, Bratko D, Blanch HW, Prausnitz JM. Interaction between oppositely charged micelles or globular proteins. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics. 2000 Oct;62(4 Pt B):5273-80.

[29]. Creagh AL, Prausnitz JM, Blanch HW. Structural and catalytic properties of enzymes in reverse micelles. Enzyme Microb Technol. 1993 May;15(5):383-92.

[30]. S. F. Matzke, A. L. Creagh, C. A. Haynes, J. M. Prausnitz, and H. W. Blanch. Mechanism of protein Solubilization in Reverse Micelles. Biotech. and Bioengin. 1992; 40:91-102.

[31]. Wolbert RB, Hilhorst R, Voskuilen G, Nachtegaal H, Dekker M, Van't Riet K, Bijsterbosch BH. Protein transfer from an aqueous phase into reversed micelles. The effect of protein size and charge distribution. Eur J Biochem. 1989 Oct 1;184(3):627-33.

Summary

Nuclear magnetic resonance (NMR) provides, contrary to X-ray crystallography, more information of the kinetic, interaction and conformation state of polypeptide chains. The interdomain structural reorganisation caused by protein-protein or protein-ligand binding events of large, multidomain proteins is essential for the understanding of biological signal transduction pathways, as well as for drug design strategies intervening with those.

Since isolated domains of proteins tend to be structurally identical to the original multidomain protein, NMR structural studies are usually focused on single domains with a reasonable size for NMR. Nevertheless studies of full length proteins would have the benefit of giving insight into the structural interactions between domains, which are often the molecular basis for regulation and signal transduction.

In **Chapter 1** we present an adapted and optimized intein ligation method based on the recent *in vivo* ligation protocols developed by Hideo Iwai[1][2] in order to make such systems accessible for NMR.

This protocol was applied to two large proteins, HSP90 (170kDa) and eIF4A successfully. Additional structure information extracted from the HSQC are discussed.

Biophysical characterization of proteins is a key factor to understand biology at chemical level. Understanding protein folding/unfolding improves our knowledge of protein stability, protein aggregation or fibril formation. Nevertheless factors influencing protein folding are important and high throughput screening assays facilitate such studies. Thermal shift assay takes advantage of an environmentally sensitive fluorescence dye, such as Sypro Orange, and follows its signal changes while the protein undergoes thermal unfolding. The assay is widely used for screening compound libraries for ligands of targets proteins. The ligand-binding

affinity of any potential inhibitor can be assessed from the shift of the unfolding temperature (Tm) obtained in the presence vs absence of the potential inhibitor.
The assay can also be used to detect und study essential conditions for protein folding unstability. In **Chapter 2** we applied this strategy to determine buffer conditions, which destabilize GB1; allowing unfolding/refolding NMR experiments via thermal denaturation in a temperature range compatible with NMR spectrometers. We have also suggested the possibility of studying native unfolded protein and fibril formation via TSA.

Factors influencing the physical properties have extensively studied and are nowadays perfectly understand. Viscosity influences the general tumbling of proteins, affecting the relaxation properties, which provoke strong signal broadening. Extensive protein deuteration is well established and improves the relaxation properties for large protein.
Low viscosity solvents have been suggested to ameliorate the tumbling of large properties. The protein is encapsulated in reverse micelles, protected from the hydrophobic solvent. Several successful strategies have been used to achieve encapsulation, but we will focus on phase transfer, rarely used for this purpose. The aim of this work was to establish a simple protocol widely useable by NMR community. We present in the **Chapter** 3 the preliminary results, suggesting future successful applications for NMR.

Appendix

Expression protocol for mediated Intein segmental isotopic labeling

MATERIALS

Pfu DNA polymerase

Bacto-Tryptone

Bacto-yeast extract

Sodium chloride

Imidazol

2,3-Dithiothreitol

L-Arabinose

Isopropyl-• -D-thiogalactopyranoside (IPTG) solution 1 M, sterile-filtered; stored frozen at -20°C

EQUIPMENT

PCR machine

5 L culture flasks

Low speed centrifuge with angle rotor or swingout-bucket for 500 or 1000 mL flask

Shaking incubators for e. coli

FPLC purification system with UV detector operating at the wavelength λ=280 nm.

UV/Vis spectrophotometer

Standard cell culture facilities

PROCEDURE

Cloning procedure. • **TIMING** 2 days

1 | Design primers for restriction free PCR cloning.

2 | Set up 50 µL regular PCR reaction in parallel for the 2 fragment genes

3 | Load reaction mixture on 1% agarose gel

4 | Control the size of your amplified PCR product and purify it using PCR purification kits.

• **CRITICAL STEP** Regular PCR purification kits use ethanol to wash the bound DNA. It is recommended to let the purification column equilibrate several minutes before adding the elution buffer. Alternatively, the purified PCR product can also be dried *in vacuo* and resuspended with water. Residual ethanol may inhibit the polymerase enzyme and reduce the efficiency of the second cloning step.

5 | Determine the DNA concentration of each purified PCR products and of the acceptor plasmids.

6 | Set up the RF-Cloning mixture as follows:

5–10 µL template (5 ng/ µL)

2–4 µL PCR (100 ng/ µL)

1 µL dNTP (10 mM)

1 µL PfuTurbo (2.5 U/ µL)

5 µL 10x reaction buffer

Water up to 50 µL

1. 95 °C - 30 s
2. 95 °C - 30 s
3. 55 °C - 1 min
4. 68 °C - 2 min/kb
5. GOTO 2 repeat 35 cycles
6. 68 °C - 10 min
7. Hold 4°C

• **CRITICAL STEP** The concentration ratio between the acceptor plasmid and the PCR insert is critical, since it can strongly affect the efficiency of the annealing step.

7 | Control the amplification by loading 5µL on 0.8% agarose DNA gel.

8 | Digest residual acceptor plasmid with 1uL (2.5 U) Dpn 1 for at least 2 h at 37 °C.

9 | Transform 5-10 µL in XL1-Blue or Top10 competent cells and plate on appropriate agarose plate. Incubate at 37 °C.

? TROUBLESHOOTING

Optimization of expression conditions • **TIMING** 2 days

10 | Co-transform the 2 plasmids containing the protein fragments in a regular expression stem like BL21 (DE3) and plate on dual antibiotic plate.

? TROUBLESHOOTING

11 | Inoculate 10 mL LB (supply with Kan/Carb) with 1 to 5 colonies and incubate at 37 °C overnight.

12 | Inoculate 600 mL LB (supply with Kan/Carb) with 1 mL overnight culture and incubate at 37 °C to OD_{600} = 0.6. Collect 500 µL for SDS gel (t = 0)

13 | Split culture in 6x 100 mL.

14 | Incubate each fragment at 25 °C, 30 °C and 37 °C for 30 min.

15 | Induce each fragment with 2 % (w/v) arabinose and 1 mM IPTG, resp.

16 | Collect 10 mL of each culture after 1, 3, 5 h and overnight expression for solubility tests. Measure the OD_{600} and spin at 5000 rpm for 5 min and store the pellet at -20 °C.

17 | Resuspend with 1 mL lysis buffer for each OD_{600} unit (i.e., for OD_{600} = 1.2, pipette 1.2 mL) on ice and sonicate 3x15 s with 10 s pause in between.

18 | Collect 100 µL and resuspend with same volume of 2x SDS sample buffer (total lysate fraction).

19 | Spin for 15 min at 14000 rpm at 4 °C. Collect 100 µL supernatant and resuspend with same volume of 2x SDS sample buffer (soluble fraction).

20 | Load 5 µL on SDS-page protein gel.

21 | Determine optimal expression conditions of single expression fragment.

Small scale consecutive induction test • **TIMING** 1 day

22 | Start overnight culture of co-transformed cells.

23 | Inoculate 30 mL LB with 1 mL overnight culture.

24 | Incubate at 37 °C until OD_{600} has reached 0.6 to 0.8.

25 | Split culture in 3x 10 mL into 50 mL sterile tubes.

26 | Incubate at the previously determined optimal temperature.

27 | Induce:

Tube A: 2 % w/v arabinose for 2 h, then 1 mM IPTG for another 2h.

Tube B: 1 mM IPTG for 2 h, then 2 % w/v arabinose for another 2h

Tube C: 1 mM IPTG, 2 % w/v arabinose for 4h.

28 | Spin cultures and run on SDS page gel.

29 | Decide on optimal induction order.

? TROUBLESHOOTING

Preparation of large scale expression • **TIMING** 1 day

30 | Inoculate 4 L LB with 80 mL overnight culture.

31 | Incubate at 37 °C until OD_{600} has reached 0.6 to 0.8.

32 | Spin at 4000 rpm for 15 min at room temperature.

33 | Resuspend in fresh minimal medium under specific isotopic conditions and incubate at the temperature determined in step 21 for fragment 1. We usually supply our minimal medium with 1 g L^{-1} ISOGRO™ or Celtone® Base Powder.

34 | Measure OD_{600} after 1 h to verify cell growth.

35 | Induce the first promoter to express the first fragment (determined at step 29).

36 | Choose expression time depending on conditions determined in step 21.

37 | Spin at 4000 rpm for 15 min at room temperature.

38 | Resuspend in fresh minimal medium under specific isotopic conditions and incubate at the optimal temperature for the ligated product if known, otherwise follows the optimal temperature for the second fragment (Step 21).

39 | Measure OD_{600} after 1 h to verify cell growth.

40 | Induce the second promoter to express the second fragment (determined in step 29)

41 | Choose incubation time as determined in step 33.

42 | Spin at 4000 rpm for 15 min at 4 °C.

43 | Confirm expression and ligation result on SDS page gel.

• **CRITICAL STEP** Make sure your cells are still growing after transferring from full to minimal medium.

Avoid any thermal shock during spinning and use a regular centrifuge (no vacuum).

? TROUBLESHOOTING

? TROUBLESHOOTING

For troubleshooting advice see **Table 1.**

TABLE 1 | Troubleshooting table.

PROBLEM	POSSIBLE REASONS	SOLUTION
No PCR product during the second amplification.	Purified PCR product contaminated with ethanol .	Dry the PCR product to eliminate residual ethanol traces.
Low co-transformation yield.	Low plasmid concentration or low cell competency.	Transformation using electroporation.
No or low ligated product.	Low expression or insoluble fragment. mRNA stability.	Check fragment solubility and adjust expression conditions. Or follow **extra procedure.** Expression in BL21 (DE3) STAR (RNAse E deficient).
Cross labeling contamination.	Residual material available at the induction time point.	Increase incubation time after resuspending the cell mass.

EXTRA Procedure: Small scale ligation test in different expression stem

TIMING 1 day

44 | Start overnight culture of co-transformed cells (for each stem).

45 | Inoculate 30 mL LB with 1 mL overnight culture.

46 | Incubate at 37 °C until OD_{600} has reached 0.6 to 0.8.

47 | Split culture in 3x 10 mL into 50 mL sterile tubes.

48 | Incubate at the previously determined optimal temperature.

49 | Induce:

Tube 1: 2 % w/v Arabinose = Reference fragment 2

Tube 2: 1 mM IPTG = Reference fragment 1

Tube 3: 2 % w/v arabinose + 1 mM IPTG

50 | Spin cultures and run on SDS-page gel.

51 | Check for ligation efficiency.

TIMING

Cloning required 3 days. Optimization of expression conditions (figure 2) takes 2 days. Small scale consecutive induction test (figure 3) takes 1 days. Preparation of large scale expression (figure 1) requires 1 day. Small scale ligation test in different expression stem (figure 4) takes 1 day. Each expression experiment requires overnight culture.