ERIKA FÖLDESNÉ DUDÁS

NMR studies of folded and disordered proteins and bicelle systems

PhD Thesis

Supervisor:

Dr. Andrea Bodor, PhD habil.

Associate Professor, Eötvös Loránd University

Hevesy György Doctoral School

Head of Doctoral School: Prof. Dr. Attila G. Császár

Synthetic, Organic and Biomolecular Chemistry

Head of Doctoral Program: Prof. Dr. András Perczel

Eötvös Loránd University

Budapest, 2019

Table of Contents

Abbreviations

AFM	Atomic force microscopy
BPLED	Bipolar longitudinal eddy-current delay
CD	Circular dichroism
CHAPSO	3-([3-cholamidopropyl]dimethylammonio)-2-hydroxy-1-propanesulfonate
cmc	critical micelle concentration
COSY	Correlation spectroscopy
CSI	Chemical shift index
DBD	DNA-binding domain
DHPC	1,2-dihexanoyl-*sn*-glycero-3-phosphocholine
DLS	Dynamic light scattering
DMPC	1,2-dimyristoyl-*sn*-glycero-3-phosphocholine
DMPG	1,2-dimyristoyl-*sn*-glycero-3-phospho-(1'-rac-glycerol)
DMPS	1,2-dimyristoyl-*sn*-glycero-3-phospho-L-serine
DMSO	dimethyl sulfoxide
DPC	Dodecylphosphocholine
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
ERK	extracellular signal–regulated kinase
GST	glutathione S-transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High Performance Liquid Chromatography
HSQC	heteronuclear single-quantum coherense
IDP	intrinsically disordered protein
IDR	intrinsically disordered region

IUP	intrinsically unstructured protein
ITC	Isothermal titration calorimetry
LB	Luria-Bertani (broth)
LED	longitudinal eddy-current delay
MAPK	Mitogen-activated protein kinase
MAPKAPK	Mitogen-activated protein kinase activated protein kinase
MD	molecular dynamics
MES	2-(N-morpholino)ethanesulfonic acid
MG	molten globule
MK2	MAPK-activated protein kinase 2
MNK1	MAP kinase-interacting serine/threonine-protein kinase 1
Mts1	Metastasis-associated S100A4 protein
NFAT	Nuclear factor of activated T-cells
NMR	nuclear magnetic resonance (spectroscopy)
(het)NOE	(hetero)nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
OD	optical density
PBS	Phosphate-buffered saline
PFG-NMR	pulsed-field gradient NMR
PGSE	Pulsed Gradient Spin Echo
PMG	premolten globule
PRE NMR	paramagnetic relaxation enhancement NMR
RC	random coil
RDC	residual dipolar coupling
r_G	radius of gyration
r_H	hydrodynamic radius
RSK1	Ribosomal S6 kinase 1

SANS	small-angle neutron scattering
SASA	solvent accessible surface area
SAXS	small-angle x-ray scattering
SCS	secondary chemical shift
SDS	Sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	size-exclusion chromatography
smFRET	Single molecule fluorescence resonance energy transfer
SSP	secondary structure propensity
TAD	transactivation domain
TCEP	tris(2-carboxyethyl)phosphine
TOCSY	total correlation spectroscopy
wt	wild-type

Chapter 1

Introduction

During the course of my PhD work I got acquainted with various protein nuclear magnetic resonance (NMR) spectroscopy techniques. I applied them to characterize the structural propensities of intrinsically disordered proteins (IDPs) and to determine the molecular dimension, folding and shape of several biomolecules and the morphology of membrane mimetics (micelles and bicelles). Utilizing solution-state NMR spectroscopy methods I studied the conformational propensities of linear motifs from members of the Mitogen-Activated Protein Kinase Activated Protein Kinase (MAPKAPK) system and the Transactivation Domain (TAD) region of tumor suppressor protein p53 in complex with the metastasis-associated protein S100A4. Based on the assigment of homo- and heteronuclear two- and three-dimensional spectra I acquired several NMR parameters that enabled to probe the conformational preferences of these proteins.

During my PhD research I have developed a method to characterize the size and shape of biomolecules (folded, disordered and denatured peptides/proteins, micelles and bicelles) based on PFG-NMR measurements, later combined with Small-Angle X-Ray Scattering (SAXS) experiments. I proposed several primary and derived parameters and empirical relations that can be used to probe the structural characteristics of folded and disordered protein and peptide molecules in aqueous and denaturing media. The comprehensive study provided valuable information about the different behaviour of these protein families and also a bioanalytical tool for aggregation and molecular mass analysis.

Based on PFG-NMR and SAXS measurements, I also investigated the nature of peptide/protein-membrane interactions. Using simplified model peptide-bicelle systems the developed method is a simple and robust approach to screen the changes in global properties of the system – shape and size – upon protein-membrane interactions.

Chapter 2

Literature Review

2.1. IDPs

The native state of a protein is usually associated with a compact globular conformation and a rigid, highly ordered structure. However, at the turn of the century several studies [1-4] suggested that numerous proteins fulfill their functions without a well-defined, globular structure in an aqueous solution – i.e. they are native. Polypeptide chains may have segments that are not likely to form a stable three-dimensional structure; they are called intrinsically disordered regions (IDRs). Structured proteins lack IDRs while proteins with entirely disordered sequences without any tertiary structure are referred to as intrinsically disordered proteins (IDPs). IDRs were considered as passive linkers but it has been proved that they play important roles not only in posttranslational modifications but also in interactions with structured domains in other proteins as short linear peptide motifs (~3-10 residues). IDPs are widely spread in nature; by conservative estimates, about 10–35% of prokaryotic and about 15–45% of eukaryotic proteins contain significant disorder, that is, long disordered regions at least 30 residues in length [5]. Dunker and co-workers [6] distinguished 28 separate functions for disordered regions that can be summarized as molecular recognition, molecular assembly, protein modification, and entropic chains. A general correlation between intrinsic disorder and various diseases, such as cancer, diabetes, amyloidoses, and neurodegenerative and cardiovascular diseases, is supported by several bioinformatics analyses and by detailed studies on individual proteins [7-9]. IDP–partner interactions are attractive targets for drug development, as drugs can be designed to target either the IDP itself, or the binding site of the partner, through mimicking the disordered motif. The first approach is considered in the c-Myc-Max complex. C-Myc is a transcription factor involved in many types of cancer which has to bind to Max by mutual folding in order to bind to its targets. If inhibitors bind to c-Myc,

complex formation with Max is hindered [10]. As an example of the second approach, Nutlins have been discovered; they are small molecules that bind to the p53 binding site of MDM2, thus inhibiting the p53–MDM2 interaction and leading to re-activation of the p53 pathway [11] in cancer cells, promoting cell-cycle arrest, apoptosis, and growth inhibition. A Nutlin analogue is currently being evaluated in clinical trials [12]. These examples show that IDPs are involved in several key biological processes; therefore, it is of high importance to study them thoroughly with various biophysical methods to enable improved understanding of their functions.

Disordered proteins can be characterized by a high proportion of charged, polar and structure-breaking residues - also known as "disorder promoting amino acids" - (Ala, Arg, Gly, Gln, Ser, Pro, Glu, and Lys) and a low content of bulky hydrophobic and aromatic amino acids (Trp, Tyr, Phe, Ile, Leu, Val); they also possess fewer Cys and Asn residues, collectively termed "order promoting amino acids".

It is generally thought that the absence of structure is encoded by the amino acid sequence. The low mean hydrophobicity and high net charge promote disorder, because they provide less driving force for protein compaction while also contributing to charge–charge repulsion. This simple principle can be applied to predict intrinsically disordered or ordered nature of proteins by plotting the absolute net charge as a function of the mean normalized hydrophobicity, a plot denoted as a charge-hydropathy (CH) plot or Uversky plot (Figure 1) [13]. In this plot IDPs can be found in the high net charge-low hydrophobicity range.

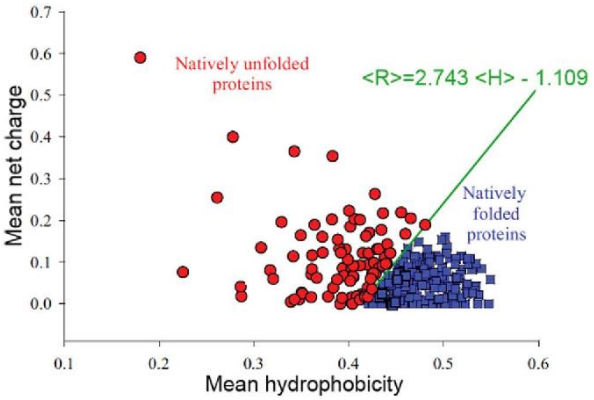


Figure 1. Charge-hydropathy plot of protein disorder. Absolute net charge vs mean hydrophobicity is plotted for disordered (red circles) and ordered (blue squares) proteins. The two sets are separated by a straight line ⟨R⟩ = 2.743⟨H⟩ − 1.109 shown as a green line. [14]

Protein disorder is also related to low sequence complexity. Based on these sequential characteristics sophisticated bioinformatic algorithms have been developed for predicting disorder from sequence: FoldIndex [15], PreLink [16], GlobPlot [17], DisProt [18], IDEAL [19], and MobiDB [20], PONDR [21], IUPred [22], DISOPRED [23] and RONN [24].

10 years ago, IDPs were considered as intrinsically unstructured proteins (IUPs) and based on the global structural level they were grouped random coil-like (RC), pre-molten globule type (PMG) and molten-globule type (MG) as shown in Figure 2. In the MG state native secondary structures exist although the protein molecule lacks a well-packed core. PMG represents a partially ordered version of the random coil with some residual secondary structure. The RC state has hydrodynamic dimensions typical of considerably unfolded polypeptide chains that are devoid of any ordered secondary structure. Far UV circular dichroism (CD) spectra provide reliable estimates of the secondary structure content, thus, ordered and molten-globule type proteins can be distinguished from random coils. Moreover, sharp peaks assigned to aromatic groups can be observed in near UV CD spectra when the protein is ordered, but these peaks are absent for molten globules and random coils due to motional averaging [25,26]. Thus, the combined use of near and far CD spectroscopy provides a method to distinguish between ordered, molten globule type and random coil proteins.

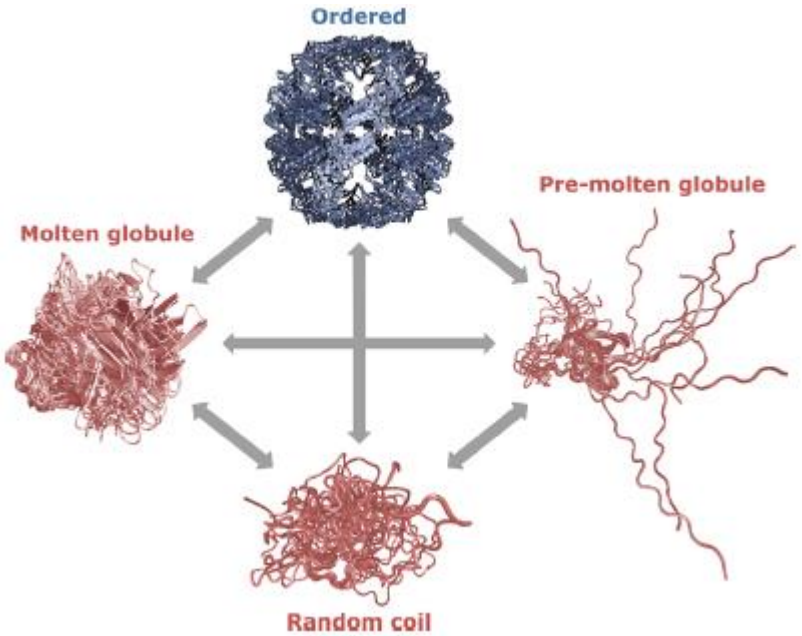


Figure 2. Protein quartet model of protein function. Function can arise from four different conformations of the polypeptide chain, or from transitions between any of the states [27].

The term “unstructured” in IUPs implied that they completely lack structure, but it was apparent that they possess short- and long-ranged organization which led to a change in terminology.

2.1.1. Characterization of protein disorder

Structural disorder can be detected and analyzed nowadays via many (bio)physical techniques, some are indirect, others are more quantitative in providing structural data. The most spectacular advance has been achieved through the application of NMR spectroscopy [28,29]. Many aspects of structural disorder can be detected directly by NMR, including local disorder, folding upon binding, and disorder in complex. The posttranslational modifications – that are important regulators of IDP function - and the interactions of IDPs can also be studied by NMR spectroscopy; in-cell methods enable the high-resolution NMR measurements of proteins in living cells. NMR is often combined with other techniques such as Small-Angle X-Ray Scattering (SAXS) [30] and Molecular Dynamics (MD) simulations [31]. SAXS assesses protein dimensions and shape by measuring the scattered X-ray intensity caused by the sample and the measured parameters indicate whether a protein is compact or unfolded [32,33]. Since there are several limits and pitfalls regarding SAXS data analysis and interpretation, much effort has been made to calculate theoretical SAXS curves from MD simulations [34] and coarse-grained modelling [35]. In the recent years IDPs were also studied at the single-molecule level. Single-molecule fluorescence resonance energy transfer (smFRET) [36] can measure dynamics and individual conformations of the unbound ensemble, intermediates during induced folding, and internal friction in the folding process. Atomic-force microscopy (AFM) [37,38] also enables either the visualization of conformational changes or the study of the energetics and dynamics of the structural ensemble of IDP molecules. From a large number of possible disordered conformers, a limited number of structural states are then selected. In contrast to the previous methods, detection of disorder using X-ray crystallography techniques is mainly indirect as it relies on missing electron density.

Certain approaches e.g. sodium dodecyl sulphate-polyacrylamide gel electrophoresis SDS-PAGE [39], size-exclusion chromatography SEC [40] may also provide important insight and/or fast information on the structural state of a protein or

its region. There is considerable interest in quantifying the thermodynamic forces that govern IDP interactions. Equilibrium dissociation constants (K_d) can be determined by fluorescence-based techniques when binding is extremely tight and the residue specific K_d values can be extracted by NMR titrations [41]. For many IDPs, binding is accompanied by a disorder-to order transition, leading to the hypothesis that the unfavorable entropy loss incurred by conformational restriction of the IDP in the bound state must be offset by a favorable decrease in enthalpy. It has been proposed that this entropy penalty may be mitigated in "fuzzy" IDP complexes [42], which retain some extent of disorder in the bound state, however, in this approach, the role of the solvent is not considered. The detailed thermodynamic information conveyed through temperature-dependent isothermal titration calorimetry ITC measurements provides exactly the experimental data needed to evaluate the energetics of coupled IDP folding and binding [43]. However, Poosapati *et al.* examined by ITC and CD measurements how folding and binding can be uncoupled for the transactivation domain of c-Myb and its ordered partner KIX [44]. Circular dichroism (CD) spectroscopy [45] is a simple and powerful technique, which allows for the assessment of the conformational properties of a protein or protein domain. IDPs have regions with secondary structure propensities and this secondary structure content and its changes can be estimated via CD spectroscopy, however, residue-specific information cannot be obtained.

2.1.2. Characterization by NMR

NMR spectroscopy enables atomic level characterization of an IDP, which usually starts by recording a one-dimensional proton spectrum where we can observe low dispersion of the proton resonance frequencies due to the similar chemical environment each residue has. Additionally, pulsed-field gradient NMR methods enable measurement of the translational diffusion coefficients, which can be starting values to derive hydrodynamic parameters (this will be discussed in detail in Chapter 3.2.1.) These parameters provide useful hints as IDPs possess peculiar hydrodynamic parameters as compared to folded proteins. Heteronuclear single quantum coherence (HSQC) is one of the most frequently utilized experiments in the IDP literature. It represents the starting point of resonance assignment from a variety of multidimensional experiments, out of which I focus on the 3D techniques. Resonance assignments can be made using the chemical shifts of ^{15}N and ^{13}C nuclei which reflect their local chemical environment. This

is the most common assignment approach, however besides this, methods exclusively based on ^{13}C detection have been reported [46-49] and recently HαCα–based measurements with band selective homonuclear decoupling are developed [50]. These approaches provide an independent strategy to simplify crowded spectra as well as to perform sequence-specific assignment. ^{15}N HSQC spectra are used to monitor interactions with unlabelled binding proteins and they can be acquired by using the natural abundance of the ^{15}N isotope, but normally isotopically labelled proteins are used. This chemical shift mapping may result in cross-peak shift, broadening or disappearance (broadening below detection limit) due to changes in chemical environment or exchange processes. Once a resonance assignment has been achieved, a variety of NMR parameters can be determined to characterize the structural and dynamic behavior at the residue level, thus obtaining sequence-specific structural information. NMR parameters that are sensitive to local structure, such as chemical shift, coupling constant, short-range nuclear Overhauser effect (NOE), heteronuclear Overhauser effect (hetNOE), relaxation, and residual dipolar coupling (RDC) values, can be determined; these can be complemented by long-range structural constraints obtained in paramagnetic relaxation enhancement (PRE) NMR measurements and SAXS experiments.

2.2. The MAPKAPK system

The mitogen-activated protein kinase (MAPK) cascades are evolutionary conserved, intracellular signal transduction pathways that control a large number of fundamental cellular processes including growth, proliferation, differentiation, motility, stress response, survival and apoptosis. [51] The mammalian MAPK family consists of extracellular signal-regulated kinase (ERK), p38, and c-Jun NH2-terminal kinase (JNK). Each MAPK signaling pathway consists of at least a MAPK, a MAPK kinase (MAP2K) and a MAPK kinase kinase (MAP3K). MAP3Ks phosphorylate and activate MAP2Ks, which phosphorylate and activate MAPKs.

Deviations observed in these pathways can be involved in many human diseases including Alzheimer's disease (AD), Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS) and various types of cancers [52-54].

Extracellular signals or mitogen stimulation activate the extracellular signal regulated kinase (ERK) pathway, which comprises a hierarchically organized kinase cascade [55]. ERK2 becomes phosphorylated by upstream MKK1/2 kinases, and the activated ERK1/2 activates ribosomal S6 kinase 1 (RSK1) by sequential phosphorylation events. Other MAPKAPKs such as MAPKAPK2 (MK2) or MAP kinase- interacting serine/threonine-protein kinase 1 (MNK1) directly phosphorylate downstream substrates. The three proteins are evolutionarily related with similar activation, but they play markedly different physiological roles [56].

In intracellular signaling networks the organization of physical protein-protein interactions is paramount for correct physiological function, and linear binding motifs play a crucial role in this process. They are less than 10-20 residues in length and they differ fundamentally from longer, globular protein domains in terms of their binding affinities and evolution. It is becoming accepted that their role in protein-protein interactions is comparable to the importance of classical interactions between globular domains.

Mitogen activated protein kinases (MAPKs) have a docking groove that interacts with linear motifs in binding partners e.g. MAPKAPKs. These upstream activators bind to the same MAPK surface as other substrates; therefore, linear motifs are needed to increase the specificity of the binding, since the docking grooves of different MAPKs show a high degree of similarity [57,58]. Only certain amino acid positions are required for MAPKAPK binding to open protein-protein interaction surfaces, these are the so-called stapling residues. (They form intra-peptide, side-chain mediated hydrogen(H)-bonds or stapling H-bonds.) Based on crystallography results MAPKAPK docking motifs bind in a similar fashion with similar H-bonds, but the intervening region conformations differ significantly. The linear motifs (RSK1, MNK1, MK2) clearly displayed alpha-helical conformation in their docking groove binding region or 3_{10} helical conformation in the intervening region between the stapling residues when bound to MAPKs, this region in MAPKAPK protein crystal structures is unstructured. As linear motifs are becoming promising pharmaceutical targets it was of great importance to study this discrepancy and their structural propensities in solution with NMR spectroscopy as it is a well-established technique to obtain atomic resolution structural information.

2.3. p53 and the p53-S100A4 complex

p53 is one of the most frequently mutated tumor suppressors in human cancers and as such has been intensively studied for a long time. It has been described as 'guardian of the genome' as it acts as a cell cycle regulator, involved in maintaining the genetic integrity. Other critical functions of p53 are control over the cell cycle machinery, apoptosis and DNA repair [59], and also a novel transcription-independent proapoptotic function has been revealed [60]. Stress signals, including heat shock, hypoxia and DNA damage trigger the activation of p53, and once activated, the protein conformation undergoes changes at both N- and C-terminal domains. The 393 residue long homo-tetramer p53 protein is encoded by the TP53 gene and each of its monomers is composed of an intrinsically disordered N-terminal transactivation domain (Met1-Asp42), a proline-rich domain with multiple copies of PXXP sequence (Asp61-Ser94), a central DNA binding core domain (Thr102-Lys292) and a C-terminal domain (Pro301-Asp393) containing a tetramerization domain (Asp324-Ala355), see Figure 3.

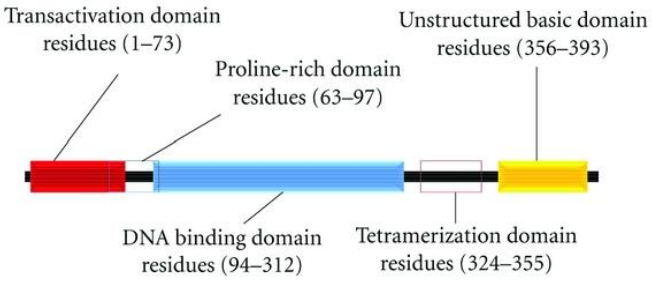


Figure 3. Human p53 can be divided into four domains, each with specific function.

The oligomerization domain and the DNA-binding domain of p53 have been extensively investigated and their structures are well determined [60-63]. The unbound full-length (73 residues) p53 TAD, largely unstructured in aqueous solution, contains a small fraction of secondary structure [64], as do many other TADs belonging to different proteins [65-67].

The N-terminal TAD region of p53 was studied in its full length [68] and in complex with several proteins (this will be discussed in detail in Chapter 4.2.) including Transcription factor II Human (TFIIH), PC4, 70 kDa subunit of human replication protein A (hRPA70), nuclear coactivator binding domain (NCBD) and Mouse double minute 2 homolog (Mdm2). Reactivation of p53 transcriptional activity has been sought as a novel

cancer therapeutic strategy, therefore, studying the structural propensities of the p53 TAD region in complexes is of great importance and the information obtained is also required to fully understand its functions. There is great interest in developing high-throughput assays to identify inhibitors of molecules that bind the transcription-activation domain of p53 [69].

The importance of the p53-S100A4 complex was demonstrated recently [70] since S100A4 has metastasis-inducing properties and its binding partners could be employed as optical biosensors. Their results also suggested that S100A4 might induce metastasis by influencing the function of p53 through binding it. S100A4 binds to the C-terminal tetramerization domain of p53, disrupting the tetramerization equilibrium of p53, thus, increasing the efficiency of its nuclear transportation. They speculated that high levels of S100A4 could promote p53 import to the nucleus and consequently increase the targeting of p53 for degradation and that this could be the basis of the causative effect of S100A4 in metastasis. In a paper from 2001, Grigorian *et al.* [71], for the first time claimed a physical and functional interaction between S100A4 and p53. In a study, Orre *et al.* showed that endogenous S100A4 and p53 interact in complex samples and showed that the interaction takes place in the cell nucleus [72]. In combination with the fact that reduced S100A4 levels results in increased p53 stability, their data suggested that S100A4 is involved in p53 ubiquitination and degradation. Van Dieck et al. investigated the ways various S100 proteins including S100A4 bind to different fragments of p53 in vitro and noted that beside the DNA binding domain there is an additional binding site common for all S100 proteins within p53, therefore, they analyzed the binding to the TAD region using fluorescence anisotropy experiments [73]. They determined the dissociation constants (K_d) and concluded that S100A4 showed medium binding affinity to the p53 TAD. Very few clinical studies have been performed where the expression level of p53 and S100A4 is determined in the same environment; however, a strong inverse correlation between S100A4 and p53 has been shown by immunohistochemistry in lung adenocarcinoma, suggesting that the level of S100A4 is higher in p53 wt tumours. A trend toward inverse correlation between S100A4 and p53 was also shown in a breast cancer cohort, where a higher level of S100A4 was found to be a negative prognostic factor. As there is great interest in identifying inhibitors of molecules that bind the TAD domain of p53 and binding partners of metastasis-inducing

S100A4 could be applied as biosensors, we aimed to study the – previously not investigated – structure of their complex.

2.4. Methods for molecular size and shape characterization

The molecular size is an important feature that provides information about the aggregation state, denaturation, conformational changes and folding. A key question in protein characterization is the determination of folding. Folded and disordered proteins bear different structures and shapes. Globular proteins are compact, and they have well-defined 3-D structure; they possess spherical, prolate or oblate shape. IDPs show resemblance to synthetic polymer chains and they can be described as 'wormlike' objects close to the random coil state.

Fully denatured folded proteins are expected to be similar to IDPs and the similarities or differences can be characterized by using hydrodynamic parameters. These parameters are valuable tools for quantitative description of folding by establishing empirical formulae that allow distinction between the different protein types and give estimation about the extent of denaturation which is important in understanding protein folding/misfolding. In the literature there are several commonly utilized techniques for evaluating the hydrodynamic parameters of a protein molecule: SAXS [30] is a well-established technique that provides low resolution structural characterization of biological macromolecules in solution. Moreover, it can probe the protein conformation and oligomeric state. However, the radius of gyration r_G which can be extracted directly from scattering curves (based on the so-called Guinier region) can inherently lead to high errors. The elution and hydrodynamic volumes from SEC [74] can also carry high errors. An unambiguous parameter of a protein conformational state is the hydrodynamic radius (r_H) derived from translational self-diffusion coefficients (D). Diffusion coefficients can be measured by Dynamic Light Scattering (DLS) [75], which is suitable for larger molecules or by diffusion NMR spectroscopy [76,77] which is advantageous for small and medium-sized proteins, on which we focused our studies.

A correlation between D and the molecular mass M is established via the Stokes-Einstein relation, assuming a hard sphere with radius r_H is moving through a continuum fluid and further, the occupied hydrodynamic volume is correlated with the mass according to Equation 1:

$$D = \frac{k_B \cdot T}{6 \cdot \pi \cdot \eta \cdot r_H \cdot F} = \frac{k_B \cdot T}{6 \cdot \pi \cdot \eta \cdot F} \cdot \left(\frac{4 \cdot \pi \cdot \rho \cdot N_A}{3}\right)^{\frac{1}{3}} \cdot M^{-\frac{1}{3}} \quad (1)$$

where k_B is the Boltzmann-constant, T the temperature, η the viscosity of the medium, F a form factor, ρ the effective density and N_A the Avogadro number. The F value incorporates several contributions: the different hydration properties based on the amino acid sequence and folding, the deviation from the ideal spherical shape and solvent effects. The applicability of the Stokes-Einstein equation for biomolecules will be discussed in Chapter 4.3.

Diffusion NMR spectroscopy is a versatile technique with a wide range of applications. There have been numerous publications related to the determination of the average molecular weight and the molecular weight distribution of polymers and dextranes from NMR diffusion experiments [78,79]. PFG-NMR is also routinely applied for the determination of D for small molecules and molecule mixtures. For biomolecules several empirical methods have been proposed for relating the molecular weight to the diffusion coefficient or the number of amino acids to the hydrodynamic radius of a certain protein using diverse methods (PFG-NMR, SEC, DLS, viscometry, gel filtration) and conditions. Danielsson proposed *D*-*M* correlation for sequentially similar disordered fragments (5-40 residues) of the Aβ peptide [80]. Groves *et al.* used five protein molecular weight standards in D_2O media [81]. *D*-*M* relations are based on primary parameters, however, equations based on the derived r_H-*N* enjoy wider application. Wilkins suggested two r_H-*N* relations, based on studies on various proteins from the range of 58-494 residues in their native and denatured states [82]. Marsh provided separate empirical relations for folded and denatured proteins, and IDPs based mostly on literature data from different experimental methods [83]. The logr_H-log*N* equations derived by Uversky *et al.* are separated for each protein types (native coils, pre-molten globules, molten globules, folded globular proteins and chemically denatured ones) and are based on gel-filtration chromatography, DLS and viscometry measurements under varied conditions [84]. Wang *et al.* observed a positive correlation between the number of residues of small cyclic peptides with therapeutic potential and their calculated hydrodynamic radii [85]. Evans proposed a simple method to estimate the propensity of aggregation and the molecular weight of small molecules based on DOSY spectra [86]. $D(M)$ type relations are rare and insufficient for biomolecules while $r_H(N)$ type formulae

carry several inconsistencies as they are based on literature data determined by different techniques or data was obtained under different experimental conditions. We therefore intended to perform a systematic study using a representative selection of folded, intrinsically disordered, and denatured proteins under highly controlled conditions for deriving reliable empirical formulae.

Not only the folding and the size but also the shape of a molecule can be estimated using diffusion NMR and the ratio r_G/r_H [87]. One way to calculate the radius of gyration (r_G) is from SAXS scattering curves [88], while the hydrodynamic radius (r_H) can be determined by PFG-NMR. Tentative numbers of the r_G/r_H ratio vary from 0.77 – being regarded as homogeneous hard spheres – to greater than 1.75 for linear polymer chains. Shape factor variations allude to molecular shape changes e.g. upon interaction.

2.5. Bicelles and bicelle-peptide systems

The cell membrane is a complex system that has multiple roles involving cellular transport, signaling processes, catalysis and interactions. Several proteins are involved in these essential biochemical processes. Therefore, it is crucial to investigate the physico-chemical ground (changes in structure and stability) of membrane-protein interactions to gain a better understanding of cell membrane functions.

The consequences of membrane-protein interactions are changes in biophysical properties, such as fluidity, curvature, variation of dimensions and shape. These variations become important factors during the absorption of active substances/drugs through the cell membrane. As several membrane proteins are considered as major drug targets and have bionanotechnological relevance, it is much interest in finding straightforward methods that reveal the nature of the overall interaction, and the change in morphology. Due to the complexity of the system it is a difficult task to study the interaction on the whole, thus, suitable mimetic media (Figure 4) and protein fragmentation has to be chosen that are applicable for the different biophysical investigations [89-91].

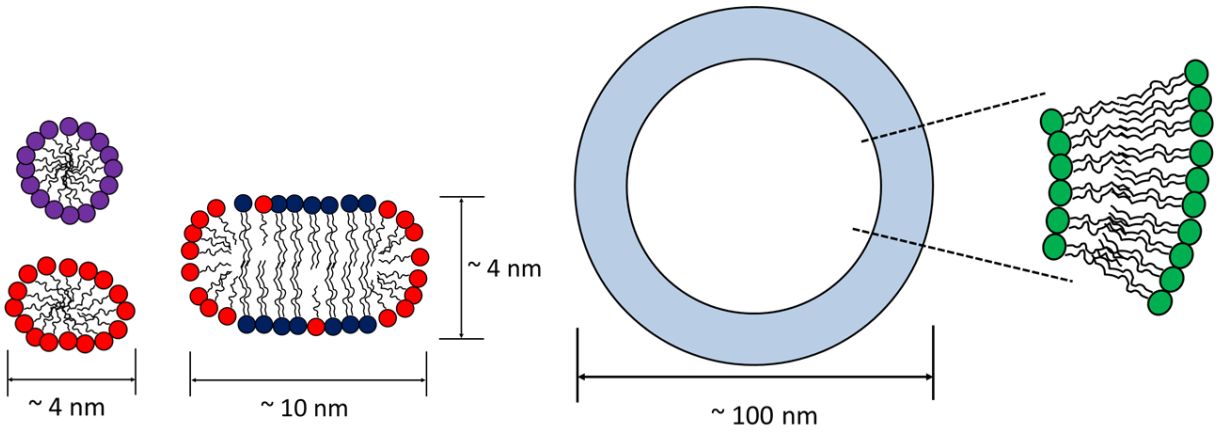


Figure 4. Membrane mimetics (SDS (top left) and DHPC (bottom left) micelle, DHPC/DMPC bicelle (middle), POPC vesicle (right)) with their typical sizes.

Membrane mimetics should meet several criteria: they should resemble the cell membrane bilayer as closely as possible and they should also have similar morphology to the biomembrane [90]. The most frequently used membrane mimetics are: micelles, bicelles and vesicles but reverse micelles, amphipols and recently, nanodiscs can be applied, as well. When the concentration of detergent molecules reaches a certain value, the critical micellar concentration (*cmc*), the excess of them will form micelles. In the energetically favored positioning of zwitterionic or amphiphilic molecules the hydrophilic head-groups point towards the aqueous solution while all the hydrophobic tails point towards each other. Detergent micelles like those composed of sodium dodecylsulphate (SDS) are spherical [92-95] and provide one of the simplest model membranes. Several other detergent molecules like DPC and DHPC also form micelles with slightly ellipsoidal or even worm-like [96,97] shapes. Relatively better NMR spectra can be obtained for micelles than larger membrane mimetics, however, they have a higher risk to destabilize a protein and their oversimplified structures might not mimick biological bilayers well enough (due to their curvature which distorts the peptide/protein binding to micelles and due to the lack of the bilayer).

Bicelles, which are bilayered mixed micelles, are more appropriate mimetics and they are composed of water-insoluble long-chained and short-chained lipid molecules. The long-chained lipids build up the bilayer and the short-chained lipids are located within the rim around the bilayer, resulting in a disk-like structure. Bicelles are characterized by the parameter q=[*short chain lipid*]/[*long chain lipid*]. If q<1, small disks are formed, that tumble isotropically within the solution, and they are used in

solution NMR. If q>2, large disk-like objects are formed, that are utilized in solid-state NMR studies. Another characteristic parameter for bicelles is the aggregation number, i.e. the numbers of long-chained and short-chained lipids in each bicelle. Bicelles represent an intermediate morphology between classical mixed micelles and lipid vesicles, combining some of the attractive properties of both model membrane systems. Like micelles, bicelles are noncompartmentalized, optically transparent, and effectively monodisperse. Consequently, it is much easier to achieve homogeneous mixing in bicelles than in lipid vesicles. On the other hand, bicelles have a much lower detergent content than classical mixed micelles and maintain some key bilayer properties that are absent in the latter systems. Biological membranes must provide a suitable environment in which to solubilize and maintain membrane protein function, while at the same time presenting an inert, nondenaturing surface to water soluble proteins. The most commonly used bicelles nowadays are made of dimyristoyl phosphatidylcholine (DMPC) and dihexanoyl phosphatidylcholine (DHPC) as under physiologyical conditions its size resembles the thickness of the membrane [90,98]. The original choices of CHAPSO or DHPC for use in generating bicelles was based, in part, upon the fact that their head groups have zwitterionic nature, thus, these molecules are mild detergents. A fraction of DMPC (typically 10-30%) can be replaced with negatively charged lipids (DMPG or DMPS) to mimic prokaryotic cell membranes.

Besides mimicking the cell membrane with bicelles we intended to simplify the membrane-active proteins as well; in this respect we chose two already characterized protein fragments (short peptides) as model peptides instead of the whole biomolecule. This synthetic transmembrane KALP23 peptide (GKKLALALALALALALALALKKA) contains hydrophilic residues at the N- and C-terminus, and alternating hydrophobic alanine and leucine residues forming helical structure in-between, providing a transmembrane segment. The 21 residue long KALP21 was shown to insert into bilayers, and KALP23 has the same behavior. The transmembrane insertion was confirmed through the measurement of the tilt angle characteristic for WALP23 – same sequence as KALP23 with the exception of the first residue – using fluorescence spectroscopy [99], solid-state NMR [100] and MD simulations [101].

The other peptide we chose was melittin - a major protein component of the venom from the European honeybee *Apis mellifera* [102]. The amino acid sequence is:

GIGAVLKVLTTGLPALISWIKRKRQQ. The peptide is well known for its strong hemolytic activity, but it is also harmful to other eukaryotic and prokaryotic cells [103]. In numerous studies melittin was used as a model to investigate membrane interactions and membrane disruption [102]. Melittin might be termed as a surface-active protein as it tends to lie at the surface of polar and apolar phases. However, there are several assumptions in the literature [104-106] regarding melittin insertion depending on the lipid composition of the bilayer and the lipid phase and hydration level.

Upon the membrane-peptide interactions several biophysical properties can change, and these variations can be followed by using various well-established methods. For *global* (hydrodynamic radius, radius of gyration) characterization a variety of techniques can be applied: PFG NMR, SANS/SAXS [107,108], DLS [109,110], TEM [107], fluorescence [91,112,113]. To determine *local* parameters (given atomic environments) ^{1}H, ^{2}H, ^{13}C, ^{31}P liquid and solid state NMR techniques [109,112,114-123] can be used.

Up to this point some attempts have been made to give a more precise description utilizing the synergy of methods, however, an NMR-SAXS based approach has not been applied yet. Therefore, we intended to develop a combined method to investigate changes in size and morphology for bicelles and bicelle-peptide systems.

Chapter 3

Experimental

3.1. Protein expression, purification and sample preparation

This sub-chapter is intended to give an overview of the expression and purification of the isotope-labelled p53 TAD region, the general sample preparation prior to diffusion NMR experiments and the production of solutions containing fast-tumbling bicelles.

3.1.1. Expression of ^{15}N and ^{13}C, ^{15}N-labelled p53

Constructs were transformed in BL21-(DE3) Rosetta cells (Novagen). After induction, cells were grown at 37°C for 3 h in LB medium. GST-tagged p53 TAD region was purified on a HiTrap Blue-Sepharose column (GE Healthcare). To cleave the GST-tag TEV-protease was added to the eluent and incubated for 3 h at room temperature. After TEV cleavage GST and TEV was precipitated by boiling the sample. The protein precipitation was spun down, and the supernatant was further purified by reverse-phase HPLC on a Jupiter 300 C5 column (Phenomenex). The expression of ^{13}C- and/or ^{15}N-labeled proteins was modified as follows. E. coli BL21(DE3) cells carrying the plasmid encoding p53 TAD region fragment were cultured in Luria Broth (LB) medium (1 L) at 37°C until the OD 600 reached 0.8 (MF = 3.5). The cells were harvested by centrifugation, washed thoroughly with sterile phosphate buffered saline solution (PBS) and resuspended in PBS (1 L) complemented with $^{15}NH_4Cl$ (1 g) and/or ^{13}C-glucose (4 g), $CaCl_2$ (0.1 mM) and $MgSO_4$ (2 mM). The culture was shaken for 4 h at 28 °C then the protein expression was induced with IPTG (0.5 mM) for 4 h at 28 °C.

NMR samples (505 µL) consisted of 1mM protein, 20 mM MES, 2 mM TCEP, 20 mM NaCl, 3 mM NaN_3, 10% D_2O and 5uL DSS at pH = 5.7. The required amount of lyophilized protein was measured by weight.

3.1.2. Sample preparation – PFG-NMR measurements

Diffusion measurements were performed on disordered and ordered protein fragments, miniproteins, peptides and proteins exhibiting diverse lengths, amino acid sequences and compactness. The samples were dissolved in 500 µL H_2O, 50 µL D_2O, and 5 µL DSS was added for chemical shift reference. For the denaturation process, the samples were lyophilized, and 500 µL 8 M urea solution was added to them as a denaturing agent. The protein sequences and measurement conditions are listed in the Appendix Table A2.

3.1.3. Membrane mimetic preparation

Phospholipids, 1,2- dimyristoyl-sn-glycero-3-phosphocholine (DMPC), 1,2-dimyristoyl-snglycero- 3-phospho-(1′-rac-glycerol) (DMPG) and 1,2-dihexanoyl-snglycero-3-phosphocholine (DHPC) were purchased as powder from Avanti Polar Lipids (Alabaster, AL, USA). All the lipids were used without further purification.

HEPES buffer stock solution was prepared by dissolving 0.065 g HEPES and 0.231g NaCl in 25 mL MilliQ water. pH adjustment of the buffer was done by adding a total of 90 µL 1 M NaOH, resulting in a pH value of 7.28. Bicelle stock solutions were prepared as described earlier by several cycles of heating/cooling [108]. The appropriate amounts of the long chain lipid DMPC (PC), or DMPC:DMPG=4:1 (PG) were dissolved carefully in 50 mM HEPES buffer solution by repeated vortexing until a homogeneous white slurry was formed; then an aliquot of the 1M DHPC stock solution was added. The obtained lipid mixture was subject to several cycles of heating (up to 70°C), cooling (to room temperature) and gentle vortexing, until a clear non-viscous solution was formed. In all samples the molar ratio was set to [DMPC] / [DHPC] = 0.5 (for neutral bicelles) or [DMPC(0.4) + DMPG(0.1)] / [DHPC] = 0.5 (for negatively charged bicelles). The total lipid concentrations were [long chain lipid] = 50 mM and [short chain lipid] = 100 mM. The pH of the solution was in all cases adjusted to 7.3 with 0.1 M NaOH solution.

The investigated model peptides were obtained were obtained from PolyPeptide Group (Strasbourg, France) and used without further purification. For the samples containing the peptide, the necessary amount of lyophilized peptide (powder) was measured by weight and mixed with the bicelle stock solution. This mixture was

subjected to freezing and lyophilisation, after which it was dissolved in distilled H_2O. NMR samples had the typical volume of 550 µl and contained 10% D_2O. Following the described preparation route, the samples were reproducible, as tested by 1D ^{31}P and ^{1}H NMR measurements.

3.2. NMR measurements

In this sub-chapter I will discuss the NMR experiments I carried out throughout my PhD research, assignment strategies and the NMR parameters that I utilized for protein structure characterization.

3.2.1. PFG-NMR experiments

The most common approach to measuring diffusion is to use the pulsed gradient spin echo (PGSE) NMR technique developed by Stejskal and Tanner [123], which was depicted in Figure 5. It is a modification of the Hahn spin echo pulse sequence in combination with gradient pulses for spatial encoding and decoding.

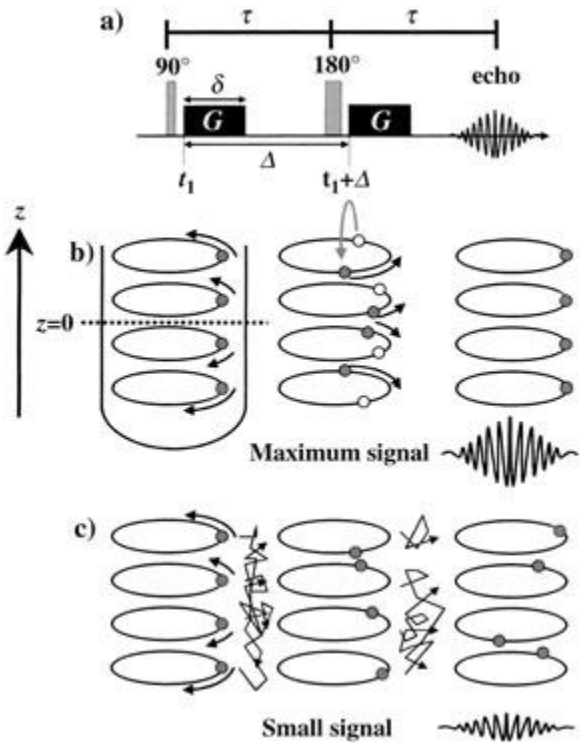


Figure 5. a.) Pulse sequence of the PGSE experiment. Phase shifts and signal intensities b.) in the absence of diffusion c.) in the presence of diffusion [77].

Initially, the net magnetization vector is oriented along the z axis. First, a 90° radiofrequency (RF) pulse is applied, which rotates the net magnetization to the x-y plane. At a given time point t_1 a pulsed field gradient of duration Δ and amplitude G is applied. Consequently, the spins experience a phase shift and their z positions are labelled. The application of the 180° pulse causes reversal of the sign of the phase angle and that of the precessing. A second pulsed gradient is applied at time $t_1+\Delta$, equal to the first in both duration and magnitude. At this point, two scenarios should be taken into consideration. If there is no diffusion, the spins do not undergo translation along the z axis and the phase shift of the spins after the first τ period is equal to that after the second τ period in magnitude. Therefore, the effects of the two pulsed field gradients cancel out, all the spins refocus, and maximum signal is obtained. If diffusion occurs, the phase shifts after the first period τ are not equal to those after the second period τ in magnitude. In the presence of diffusion, at times t_1 and $t_1+\Delta$, molecules are located at different positions along the z axis, hence they are situated in magnetic fields with different strengths. Therefore, their spins precess with angular frequencies specific to their positions. Phase angles increase, refocusing is less effective and the echo signal is smaller.

The signal intensity I can be described by the Stejskal-Tanner formula

$$I = I_0 \exp\left[-D\gamma^2\delta^2G^2\left(\Delta - \frac{\delta}{3}\right)\right] \qquad (2)$$

where I_0 is the initial signal intensity, D is the translational diffusion coefficient, γ is the gyromagnetic ratio, G is the pulsed gradient strength, δ is the duration of the pulse and Δ is the time separation between the pulsed field gradients.

If the gradient strength is incremented while δ, Δ and γ are kept constant, I can be measured, and D can be evaluated by fitting a Gaussian curve on the experimental data points Figure 6.

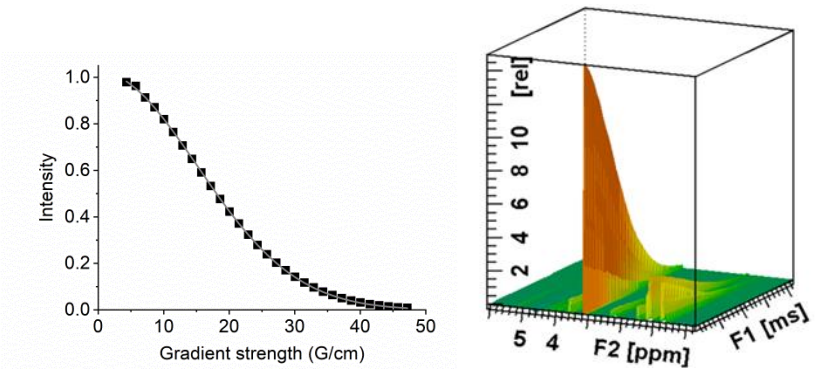


Figure 6. a.) Signal intensity as a function of the gradient strength, following a Gaussian curve. b.) Signal intensity as a function of the gradient strength in a 3D interpretation with the chemical shift as third axis.

In order to obtain reliable data, the signal has to be attenuated. The original measurement and data analysis developed by Stejskal-Tanner is not suitable for biomolecules with small diffusion coefficient values as they require long diffusion delays relative to the molecule's short transversal T_2 relaxation times. In PGSE experiments signal attenuation depends on T_2, therefore they are not applicable for determining diffusion coefficients of large molecules. To overcome this issue the PFG-stimulated echo (PGSTE) experiment was developed [124,125] where signal attenuation during the diffusion time depends on the molecular diffusion and longitudinal relaxation.

This PGSTE pulse sequence can be viewed as a modified version of the PGSE pulse sequence. The first r.f. pulse and the subsequent field gradient pulse appear to be identical in both sequences. To minimize signal losses due to transverse relaxation, another 90° pulse is applied, and spins diffuse during a diffusion delay, Δ. Subsequently, another 90° pulse is applied followed by a second z-gradient. In the absence of any diffusion, the signal attenuation is only due to gradient-independent processes, in particular longitudinal relaxation during Δ. Modifications of this sequence account for the effects of eddy currents following the gradient pulses: longitudinal eddy current delay (LED) in Figure 7, *middle* and bipolar pulse pair LED experiments in Figure 7, *bottom*. These types of experiments are most commonly applied to measurements of the diffusion of macromolecules or large aggregates.

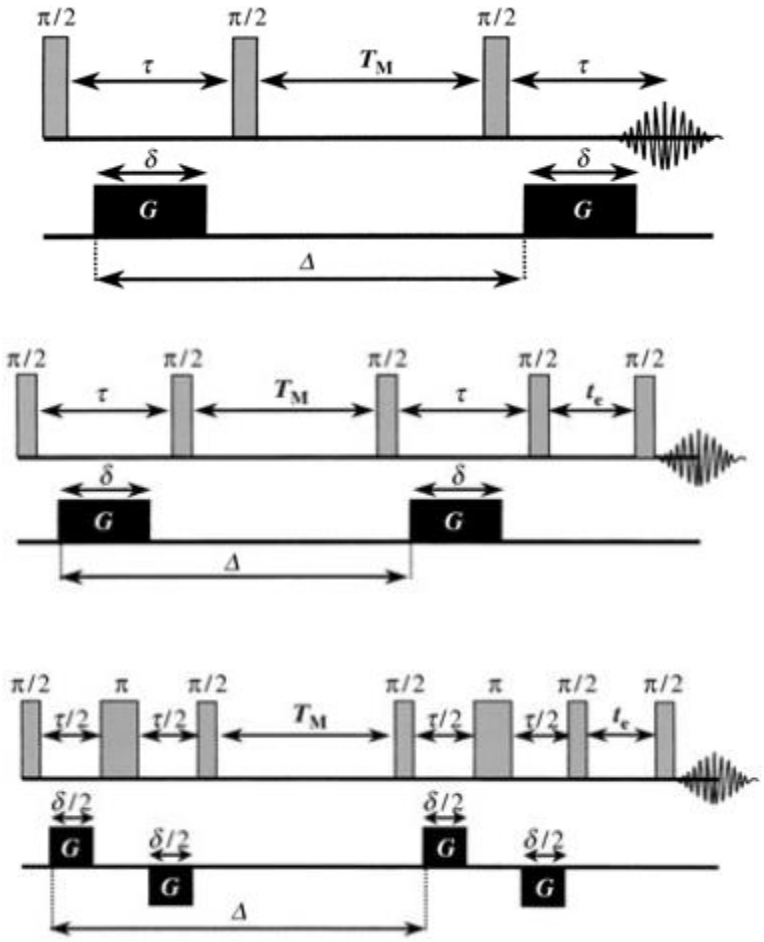


Figure 7. *Top*: The STE diffusion pulse sequence [124,125] *Middle*: the LED pulse sequence [126] *Bottom*: the bipolar LED (BPLED) pulse sequence [127]

Solvent suppression can readily be incorporated into such pulse sequences. As for our systems strong gradients and relatively long diffusion delays were required to obtain reliable data, we applied the PFG-STE stimulated echo approach with bipolar pulses with and without water suppression when necessary, see Figure 8.

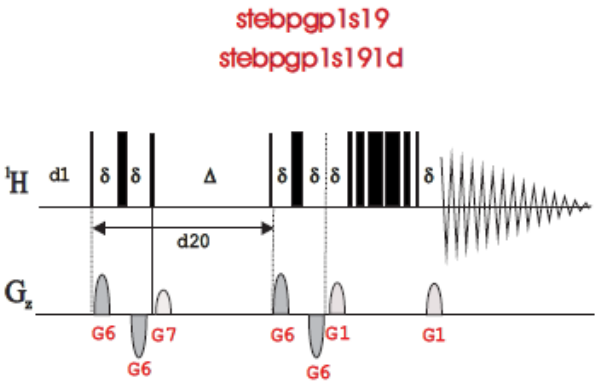


Figure 8. Pulse sequence of stebpgp1s19 (stimulated echo bipolar pulse-gradient pulses).

For each sample the lengths of diffusion delays and pulses were optimized and held constant during the experiment. The strength of the diffusion gradient was linearly incremented in 16 or 32 equal steps, varying between 5% and 95% of its maximum value. Each measurement was repeated at least two times to allow the uncertainty of the measurement to be estimated. The number of scans was adjusted for each sample to obtain reliable S/N ratios. A single Gaussian function was fitted to the decay of the aliphatic proton signals (integrals over two chosen regions in the 0.0-3.0 ppm range) based on the Stejskal-Tanner equation using the T1/T2 package of the TopSpin program. The *D* values were averaged over the integrated regions. Protein signals in denatured samples were perturbed by the extremely intensive urea signal; Gaussian fitting was done at higher *G* values after urea signal decay was complete. The resulting *D* values were averaged.

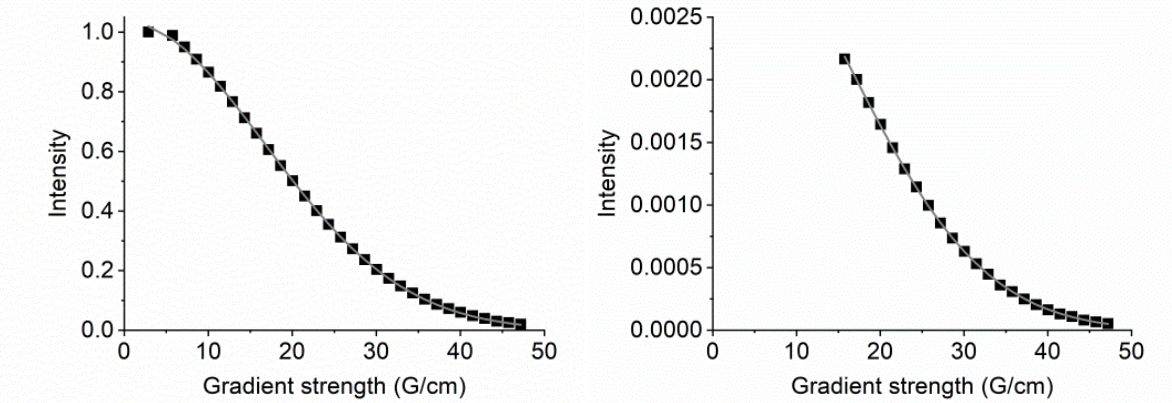


Figure 9. Signal intensity decays as a function of increasing gradient strengths for Tβ4 in aqueous solution (*left*) and Tβ4 in 8M urea solution (*right*).

Temperature calibration was done with ethylene glycol dissolved in 80% deuterated DMSO. The distance between the proton chemical shifts of the CH_3 and OH groups was $\Delta\delta$ = 1.51 ppm at 298 K. According to the following equation:

$$T_{real} = \frac{4.218 - \Delta}{0.009132} \qquad (3)$$

the real temperature was 287 K.

The gradient system must be calibrated in order to make any corrections due to possible mismatching of the components. The practical way to precisely calibrate gradient systems is to measure the diffusion coefficient of a well-known sample and compare the obtained results with the literature values. Gradient calibration was done

by using the Bruker standard 'doped water' sample, consisting of D_2O, 1% H_2O, 0.1% CH_3OH and 0.1 mg $GdCl_3$/ml D_2O (Table 1).

sample	D ($10^{-9}m^2/s$)
H_2O	2.299
D_2O	1.872
"Doped water"	1.910
DMSO	0.730

***Table* 1.** Bruker standards and their corresponding diffusion coefficients at 25°C.

The initial gradient calibration coefficient was 4257.64 Hz/G. With this initial setting the diffusion coefficient of doped water was $2.095 \cdot 10^{-9}$ m^2/s. We used the equation given by Bruker:

$$GCC_{new} = \sqrt{\frac{D}{D_{literature}}} GCC_{old} \qquad (3)$$

GCC_{new} was 4459 Hz/G.

As a result of careful optimization and calibration – including the gradient system, the temperature and the diffusion NMR measurement parameters – we were able to determine reliable, accurate diffusion coefficient values.

3.2.2. Assignment strategy

NMR spectroscopy is a powerful technique for studying protein structure and dynamics at the atomic level. The main goal in any NMR protein study is to obtain the chemical shift of each NMR-active spin, to perform chemical shift assignment [128,129]. A one-dimensional ^1H NMR spectrum of a protein is crowded with mutually overlapping lines due to the numerous protons in a protein. To overcome this problem higher-dimensional NMR experiments were introduced and isotope-labelling (^{15}N or $^{15}N/^{13}C$) techniques were developed. Two assignment methods were established: the sequential assignment procedure is based on 2D spectra and relies only on information about the amino acid sequence [130,131] of the protein; and the triple-resonance approach which is based on scalar couplings.

Linear motifs of MAPKAPKinases were unlabeled, therefore, I followed the sequential assignment method, which consists of two main steps. The first involves the identification of spin systems which belong to a particular type of amino acid. For this purpose, I used COSY and TOCSY experiments. COSY experiments are based on polarization transfer between J-coupled protons over two or three chemical bonds, while TOCSY enables transferring magnetization between all scalar coupled spins within a residue. To identify the systems of spin-spin coupled resonances of certain residues, random coil chemical shifts [132] and unique cross-peak patterns can be used (Figure 10).

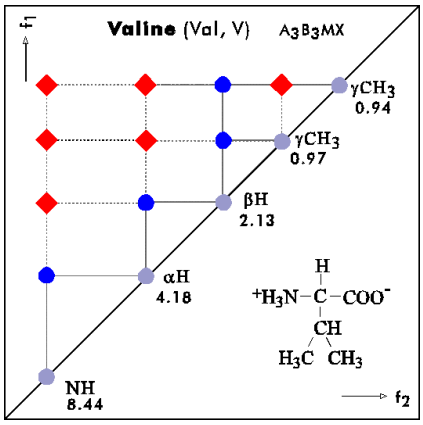


Figure 10. TOCSY-pattern of Valine. Blue dot: COSY cross-peak, red diamond: TOCSY cross-peak. [133]

Glycine has a unique cross-peak pattern with three coupled protons forming an AMX spin system which gives H(N)-Hα cross peaks with a characteristic 8 component shape. The H(N) chemical shift is correlated to two Hα shifts. The spin systems of alanine, isoleucine, leucine threonine and valine residues all contain an Hα, Hβ or Hγ atom coupled to one or two methyl groups. If there is a connection between the H(N) or Hα and the CH_3 group(s) then the H(N)-Hα cross-peaks can be assigned to one of these amino acid types. The remaining amino acids can be divided into groups with Hγ atom(s) and those without. The latter group can be denoted as type J and includes Asn, Asp, Cys, His, Phe, Ser, Tyr and Trp. These residues have Hβ resonances usually at ~2.5 ppm. Aromatic type-J amino acids, His, Phe, Tyr and Trp, have unique cross-peak patterns in the region of ~6-8 ppm. Lys, Arg, Met, Gln, Glu and Pro form the group of type U residues with two protons at the γ position coupled to Hβ. They also have Hβ resonances at ~2.2 ppm. The most important parameter of a TOCSY experiment is the isotropic mixing time required to transfer magnetization between scalar coupled spins

within a residue. With shorter TOCSY mixing times (20-25 ms) only H(N)-Hα peaks may be observed, however, as the mixing time is increased (up to ~100 ms) connectivities in the whole spin system will appear.

The second step of sequential assignment involves assignment of each identified spin system to a particular residue in the amino acid sequence. This is achieved by correlating an amino acid spin system with the neighboring residue spin systems, and relies on the spatial proximities observed in NOESY spectra. The nuclear Overhauser effect arises as a result of dipole-dipole cross-relaxation between two nuclei in spatial proximity (3-6 Å). The most beneficial NOE effects for sequential assignment involve the Hα and Hβ of residue i and the H(N) of residue (i+1) and the H(N)'s of residues i and (i+1). NOE cross-peak intensities depend on the torsion angles of the neighboring residues, this means that specific secondary structure elements are characterized by specific sequential NOE effects e.g. for an α-helix characteristic NOE cross-peaks include H(N)-H(N) of residues (i, i+1), (i, i+2) and (i, i+3), Hα-H(N) of (i i+3) and Hβ-H(N) of (i, i+1). Thus, longer range NOE effects are essential to identify regions of secondary and tertiary structure. The following stage of assignment includes the identification of stretches of sequential NOE effects as usually a full set of NOE connectivities from the N to C termini is not observed. Breaks in the sequential assignment occur for several reasons – similar sequential H(N) chemical shifts, proline residues in the sequence -, therefore, the sequential assignment is carried out in shorter segments within the protein sequence. However, as the size of the protein increases the occurrence of unique pairs of residues in the sequence decreases.

High protein fragment concentration in our MAPK linear motif samples enabled us to acquire not only 2D homonuclear COSY, TOCSY and NOESY spectra but natural abundance 2D ^{1}H-^{15}N HSQC, 2D ^{1}H-^{15}N SOFAST(band-selective optimized flip-angle short transient)-HMQC and ^{1}H-^{13}C HSQC spectra. These experiments allowed us to obtain a 2D heteronuclear chemical shift correlation map between directly-bonded ^{1}H and ^{13}C or ^{1}H and ^{15}N heteronuclei. Protein NMR samples bear typically lower concentrations; therefore, ^{15}N or $^{15}N/^{13}C$ labelling is needed to achieve appropriate signal intensities in 2D heteronuclear spectra. The ^{1}H-^{15}N HSQC spectrum should show one peak of each residue - with the exception of proline - correlating the nitrogen and the amide proton. In addition, pairs of cross-peaks with a common ^{15}N shift are observed

for each Asn and Gln side chain in the upfield region of both dimensions. A single peak appears for each Trp indole Hε1 in the downfield region of the spectrum. Arg, His and Lys side chain peaks may be present depending on the experimental conditions: pH and temperature. The expected number of peaks based on the amino acid sequence can be compared to the number of the observed peaks. 2D ^{1}H-^{15}N HSQC is the root experiment of most of the standard *triple-resonance* (^{1}H, ^{13}C, ^{15}N) NMR experiments. 2D ^{1}H-^{13}C HSQC spectra are mostly applied to verify assignments, however, I used them parallel to homonuclear spectra during the assignment process of MAPKAPK linear motifs. I observed several cross-peaks where Hα and H(N) chemical shifts of different residues were identical, but I could distinguish them based on their different Cα chemical shift values.

Limited chemical shift dispersion and highly overlapping proton signals in the NMR spectra of large proteins – such as the p53TAD-S100A4 complex - necessitate 3D experiments performed on ^{15}N or $^{15}N/^{13}C$ labelled samples [134-136]. The ^{15}N-edited 3D TOCSY-HSQC and NOESY-HSQC experiments can be utilized to assign side-chain protons. TOCSY-HSQC experiments provide through-bond spin system information while NOESY-HSQC spectra contain through-space NOE connectivities. These spectra include a subset of ^{1}H-^{1}H cross-peaks: only TOCSY or NOESY effects involving at least one H(N) are observed because the mechanism includes a heteronuclear magnetization transfer via $^{1}J(NH)$. The ^{1}H-^{1}H TOCSY or NOESY cross-peaks are separated into isolated planes on the basis of the ^{15}N chemical shift of the backbone amide group. Information in the 3D data set can be efficiently reduced to a 2D strip plot as the 3D matrix incorporates a high amount of noise. Spectra strips are extracted from ^{15}N planes. Each strip is centered around a $^{1}H(N)$ chemical shift in F3 dimension and the strip covers the full ^{1}H sweep width in the indirect ^{1}H dimension, F1. Extraction of 2D strip plots from a 3D matrix requires the identification of $^{1}H(N)(F3)$- $^{15}N(F2)$ positions at which the strips parallel to F1 should be extracted. Cross-peaks can be assigned in a 2D HSQC spectrum collected prior to the 3D experiment or in a 2D matrix created by projection of the 3D matrix along F1. The rules for assigning spin system type on the basis of the TOCSY-HSQC spectrum are identical to those used in two-dimensional counterpart.

The 3D HNHA spectrum contains H(N)-H(N) diagonal peaks and H(N)-Hα cross peaks. The ratio of these peak intensities provides quantitative information about the

$^{3}J_{HNH\alpha}$ coupling constant. In addition, it can be used to unambiguously identify Hα shifts for each residue.

The triple-resonance methods rely on coherence transfer via ^{1}J (and in part ^{2}J) couplings only to identify neighboring residues and they require uniformly $^{15}N/^{13}C$ double-labelled proteins. We performed Band-selective Excitation Short-Transient (BEST)-type triple-resonance experiments [137] to overcome the problem of long T1 spin-lattice relaxation times (usually of the order of one second for a backbone amide proton in a small- to medium-sized protein). Band-selective pulses can be used to excite only the backbone amide protons, while the majority of proton spins remain close to their thermodynamic equilibrium state and serve as an additional pathway for relaxation. In this case, the selective T1 for the protons of interest is shortened by a factor of three to five. The most simple experiment using this technique is the SOFAST HMQC [138].

Complete backbone assignment can be achieved on the basis of four triple-resonance experiments, HNCA, HNCO, HN(CO)CA and HN(CA)CO to exclusively correlate the resonances of the peptide backbone (HN(i), N(i), Ca (i), Ha (i), Ca (*i*-1) Ha (*i*-1), C0(*i*) and C0(*i*-1)) [131]. The nomenclature for these experiments reflects the magnetization transfer pathway of the experiments. Nuclei that are involved in magnetization transfers form the name of an experiment while spins, whose chemical shifts are not evolved are put in parentheses.

From the combination of HN(CO)CACB and HNCACB experiments the backbone resonance assignments and the sequential connectivities can be obtained. These experiments will provide the Ca and Cb chemical shifts to establish the sequential link between neighboring residues. Further, the Ca and Cb chemical shifts provide important information about the amino acid type - as several amino acids including Gly, Thr, Ser and Ala bear unique Cα and Cβ chemical shifts. In addition, there is a phase difference between the Cα and Cβ peaks which is particularly useful for discriminating the Cα and Cβ of Ser and Thr which often occur in the same spectral region. In addition to sequential assignments, these two experiments can, in principle, provide the secondary structure of the protein. However, for proteins larger than 15 kDa the HNCACB experiment becomes less sensitive, therefore, the more sensitive experiments HNCA and HN(CO)CA can be used in addition to establish the sequential connectivities. If the Ca

and Cb chemical shifts obtained from these four experiments still leave some ambiguities, the pair of HNCO and HN(CA)CO can be used to resolve the overlap. However, since the HN(CA)CO experiment is quite insensitive, this approach will be useful only in combination with a deuterated protein. Triple-resonance spectra include a relatively small number of peaks, generally one or two peaks per residue correlating 3 or 4 chemical shifts. The 3D spectra are reduced to strip plots with the help of different programs (CARA [139], CCPNMR [140]). The CCCONH experiment is specifically designed to correlate the ^{1}H and ^{15}N amide resonances of residue i with CA and all other carbon side-chain resonances of residue (i-1).

Recently very high dimensionality (6D and 7D) NMR measurements were performed in combination with automated analysis to produce backbone resonance assignments for full length tau protein and to overcome the resonance overlap in spectra of longer tau isoforms [141].

3.2.3. NMR parameters with structural information

Once a resonance assignment has been completed, a variety of NMR parameters can be determined to characterize the structural and dynamic behavior at the residue-level, thus obtaining sequence-specific structural information. In order to determine local secondary structural preferences, many methods have been described in the literature based on chemical shifts, coupling constants and short-range nuclear Overhauser effects (NOEs). I would like to focus on the parameters that I used for structural propensity characterization throughout my PhD work.

Chemical shifts represent one of the most direct NMR observables and they are valuable probes for local secondary structure. They have also emerged as one of the most effective structure constraints available for the refinement of disordered protein ensembles. Deviation of Hα, Cα, Cβ, and CO chemical shifts from random coil values (termed secondary chemical shift, SCS, or chemical shift index, CSI) are sensitive to local secondary structure and hence provide important insights into IDP structural propensities and the structures populated in the conformational ensemble in IDPs.

$$\mathrm{SCS} = \delta(\mathrm{measured}) - \delta(\mathrm{random\ coil}) \qquad (4)$$

Generally, residues located in β-sheets have negative Cα and positive Cβ secondary shifts, while amino acids in α-helices have positive Cα and negative Cβ secondary shifts. One potential problem associated with this kind of approach concerns incorrect frequency referencing, which can result in systematic errors in the secondary shifts. In order to address this problem, the Cα and Cβ chemical shifts (that shift in opposing directions for α-helical segments) can be used simultaneously to estimate the level of secondary structure in disordered proteins.

Secondary Structure Propensity (SSP) is a single score that combines chemical shifts from different nuclei and represents the expected fraction of structure. The contributions of different chemical shifts - Cα, Cβ, CO, Hα, H(N), and (H)N - are weighted by their sensitivity to the secondary structural elements mentioned previously. An SSP score of a given residue of 1 or -1 reflects fully formed α-helical or β-strand structure, respectively, while a score of 0.6 indicates that 60% of the conformers in the disordered state ensemble are helical at that position. SSP values are not completely quantitative, however, they can be expected to roughly correspond to the secondary structure propensity at a given position [142].

Chemical shifts and chemical shift perturbations are sensitive to the variations of chemical environments. Therefore, they can be utilized to identify binding site locations and ligand affinities via the chemical shift mapping method. In the standard experiment, an unlabeled ligand – a small molecule or another macromolecule - is added to the ^{15}N-labelled protein of interest, and each stage of the titration is monitored by acquiring a 2D ^{1}H-^{15}N HSQC spectrum. In the ^{1}H-^{15}N HSQC spectrum one cross-peak can be assigned to each (non-proline) residue in the protein, thus ^{1}H and ^{15}N chemical shifts can be determined at each titration point. Changes in the chemical shift values indicate the movement of peaks, and the extent of peak movement can be quantified by calculating the variation of individual ^{1}H and ^{15}N chemical shifts upon complex formation for each residue (Hódi, 2006)

$$\Delta\delta = \sqrt{\left[\delta(\mathrm{H})_{\mathrm{complex}} - \delta(\mathrm{H})_{\mathrm{free}}\right]^2 + \left[0.1\left\{\delta(\mathrm{N})_{\mathrm{complex}} - \delta(\mathrm{N})_{\mathrm{free}}\right\}\right]^2} \quad (5)$$

where $\delta(\mathrm{H})_{\mathrm{complex}}$ and $\delta(\mathrm{H})_{\mathrm{free}}$ are the amide proton chemical shifts and $\delta(\mathrm{N})_{\mathrm{complex}}$ and $\delta(\mathrm{N})_{\mathrm{free}}$ are the amide nitrogen chemical shifts of the free protein and the protein in

complex, respectively. Residues with the highest $\Delta\delta$ values are situated in the proximity of the binding site with cross-peaks shifted significantly in the HSQC spectrum.

For a protein(P) binding reversibly to a ligand(L) at a single site, given by P + L ↔ PL, forward and back reactions are characterized by rate constants k_{on} and k_{off} and the dissociation constant K_d is equal to [P][L]/[PL], where [P], [L] and [PL] are the concentrations of free protein, free ligand and complex, respectively. At equilibrium the forward and back rates are equal, implying that the dissociation constant K_d is also equal to k_{off}/k_{on}. This has an important effect on the NMR spectra of an exchanging system, thus, on the HSQC spectrum of a protein titrated with the ligand. When the exchange rate is slow on the chemical shift timescale, or in other words when k_{off} is significantly smaller than the chemical shift difference between free and bound protein, then the free protein signal gradually disappears and the bound protein signal appears as ligand is titrated in; the peak intensities reflect the concentrations of free and bound protein. On the other hand, when exchange is fast, i.e. when k_{off} is much greater than the chemical shift difference, then the signals will move smoothly from their position in the free spectrum to those in the bound spectrum, and the observed chemical shift is the population-weighted average of free and bound. When the exchange rate and the shift difference are similar, signals broaden and are shifted at the same time [143].

The chemical shifts of amide proton resonances also display a temperature-dependence [144]. They shift upfield as the temperature increases and this is conventionally described as a negative temperature coefficient. The generally accepted explanation of this effect is that in a hydrogen-bonded amide group, the carbonyl function causes the amide proton to be shifted downfield. If the temperature is increased, the hydrogen bond is weakened (lengthened on average) and the amide proton is shifted downfield to a lesser extent (i.e. a relative upfield shift). If an amide proton exchanges slowly and has a temperature coefficient more positive than 4.5 ppb/K, it is hydrogen bonded, while if it exchanges rapidly and has a temperature coefficient more negative than -4.5 ppb/K, it does not form a hydrogen bond. Usually amide protons have random-coil-type temperature coefficients in the range of -6 to -10 ppb/K [145]. Exceptionally great negative temperature coefficient values (<-10 bbp/K) imply that the exchange rate is fast, and the residue is not hydrogen bonded [145].

Structure determination of folded proteins by NMR spectroscopy is mainly based on the distance information obtained from homonuclear ^{1}H-^{1}H NOEs. NOE peaks can be converted to ^{1}H–^{1}H distance restraints used in structure calculations as they provide information on spatial relationships regardless of chemical connectivity. The NOE and the volume of the NOE peak depends on the internuclear distance ($\sim 1/r^6$), with an upper limit of about 5 Å. The peak volume can be converted to an internuclear distance based on a reference value which can be the fixed distance between geminal or aromatic ring protons. In practice, NOE-derived distances are often categorized into distinct ranges based on peak intensities such as 1.8–2.5, 1.8–3.5 and 1.8–5.0Å [131]. Sequential and medium-range NOEs can usually be observed in IDPs, but long-range NOEs are rarely present in NOESY spectra. The absence of long-range NOEs, however, does not exclude the existence of contacts too short-lived to allow sufficient buildup of NOE signals. NOEs can be obtained from 2D NOESY or 3D HSQC-NOESY type experiments.

3.3. Other techniques

Other experimental techniques that supplemented NMR spectroscopy are shortly discussed in the following sub-chapter.

3.3.1. SAXS

Small-angle X-ray Scattering is a powerful biophysical method to study the overall shape, conformation and assembly state of various biomolecular systems e.g. proteins and nucleic acids [33]. This technique generally provides low-resolution information. The basic scheme of a SAXS experiment is illustrated in Figure 11. A solution of macromolecules in a capillary or cuvette is illuminated by a collimated monochromatic X-ray beam from the radiation source and the intensity of the scattered beam is measured as a function of the scattering angle (2θ). The primary results of SAXS measurements are the scattering patterns that are generally recorded by two-dimensional detectors at high-quality instruments. These detectors allow for better statistical accuracy of the signal after radial averaging. An analogous measurement is performed on the same container filled by a reference sample – usually the solvent - as the 'background', and the latter scattering is then subtracted from that of the

macromolecular solution. The difference pattern arises from the dissolved particles and provides information about their structure.

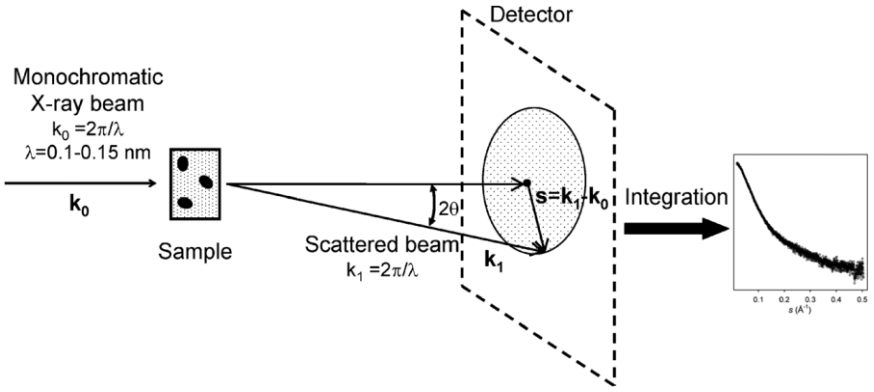


Figure 11. Basic scheme of a SAXS experiment [33].

The scattering of X-rays by a macromolecular solution is proportional to the number of biomolecules in the illuminated volume and to the 'contrast', which is the difference in the electron density of the solute and solvent. The useful signal of biomolecules in aqueous solution is rather small, as they are composed of light atoms. Therefore, the majority of biological SAXS experiments are conducted on synchrotron sources, which provide high brilliance X-rays.

The scattering patterns obtained by SAXS measurements are typically presented as radially averaged one-dimensional curves I(q). The low-q range – corresponding to small angles - of a scattering curve provides information about the large distances in the particle. Several relevant structural parameters (size, oligomeric state, overall molecular shape) can be obtained directly from these curves. Moreover, the radius of gyration, r_G can be directly obtained using the classical Guinier approximation, which is one of the most commonly used parameters to quantify the overall size of molecules in solution. In the Guinier approximation, r_G is obtained by a simple linear fit in logarithmic scale, assuming that at very small angles the intensity is represented as

$$\mathrm{I_{Guinier}(q) = I(0)e^{-\frac{q^2 r_G^2}{3}}} \qquad (6)$$

Shape information is provided by the middle-q range of the scattering curve. For membrane mimetic systems the whole curve can be fitted to obtain a more sophisticated r_G value, moreover, as SAXS intensities are well described by two-component ellipsoid models (Figure XY), with a dense outer shell corresponding to the detergent head

groups and a less electron dense hydrophobic core. The model (Figure 12) features an ellipsoidal core with semi-axes a and b, an outer shell of thickness t_a and t_b and electron densities ρ_{shell} and ρ_{core} [97].

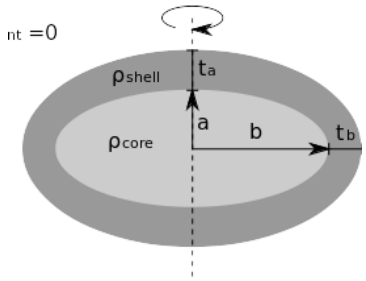


Figure 12. The ellipsoidal core-shell model.

3.3.2. Molecular Dynamics

Molecular dynamics (MD) has become a widespread tool for studying the structure-function relationship of proteins. Various systems can be simulated including entire proteins in solution with explicit solvent representations, large macromolecular complexes like ribosomes or nucleosomes or membrane embedded proteins [146]. An initial structure of the system is required for MD simulations, which can be obtained from either experimental structures or comparative modeling data. Simulations can be carried out at different levels of details of the system: atomistic representation can potentially lead to the best reproduction of the actual systems, however, coarse-grained representations are more beneficial when large systems are involved or long simulations are required. Solvent representation is of high importance in system definition and the explicit representation of solvent molecules is considered to be the most effective approach however, it increases the size of the simulated systems. Once the system is built, forces acting on every atom are obtained by deriving equations, the force-fields (complex equations that are easy to calculate), where potential energy is deduced from the molecular structure. Force-field representation simplifies various molecular features: springs ~ bond length and angles, periodic functions ~ bond rotations and Lennard–Jones potentials, and the Coulomb's law ~ van der Waals and electrostatic interactions, respectively. These simplifications assure that energy and force calculations are extremely fast even for large systems. Once the forces acting on individual atoms are obtained, accelerations and velocities are calculated based on classical Newton's law of motion and the atom positions are updated. As integration of

movement is done numerically, to avoid instability, a time step shorter than the fastest movements in the molecule should be used (1-2 fs for atomistic simulations). Microsecond-long simulations – close to the time scales of biological processes - require iterating over this calculation cycle more than 100 times. If coarse-grained strategies are applied, a more simplified representation of the system is used, much larger time steps are possible, and therefore the effective length of the simulations is dramatically extended. The disadvantage of this strategy is that the accuracy of the simulation ensemble is reduced.

Chapter 4

Results

4.1 MAPKAPK linear motifs

The essential role of mitogen-activated protein kinase (MAPK)→ MAPK-activated protein kinase (MAPKAPK) signaling events is well established, but the structural details of the complex formation were not known. RSK1 and MK2 MAPKAPKs are specifically activated by ERK2 and p38 MAPKs, respectively, whereas MNK1 is activated by both MAPKs in cells, the linear motif containing peptide from this protein bound both to ERK2 and p38α with equal (~0.5 μM) binding affinity (Table 2).

		MAPK		
		JNK	ERK2	p38
Substrate	NFAT	+	-	-
	MK2	-	-	+
	RSK1, RSK1_S/A	-	+	-
	MNK1	-	+	+

Table 2. MAPKs and their possible substrates. '+' denotes enzyme-substrate complex formation.

Linear motifs play a crucial role in protein-protein interactions and as MAPKinases have similar docking grooves, they were expected to contribute to binding specificity based on their structural differences. Crystal structures of several MAPK-MAPKAPK complexes were determined [147-149] and conformational discrepancies were observed for residues between stapling amino acids even if different linear motifs were bound to the same MAPKinase. These intervening regions bear α-helical or 3_{10} helical conformation in complex, but they were missing from the electron density map if only the linear motif was investigated, indicating unstructured regions. NMR spectroscopy is a suitable technique to gain atomic-level information about disordered protein regions, therefore, we intended to utilize it and examine structural propensities

of these peptide fragments in solution to see how their structure-function relationships can be explained (length and sequence given in Table 3). We also intended to investigate how the extent of disordered changes upon denaturation, thus we performed our measurements under denaturing conditions as well, using the commonly applied 8M urea solution as chaotropic agent.

Peptide fragment	Number of residues	Sequence
RSK1	24	PQLKPIESSILAQRRVRKLPSTTL
RSK1_S/A	24	PQLKPIEASILAQRRVRKLPSTTL
MNK1	19	MKLSPPSKSRLARRRRALA
MK2	20	IKIKKIEDASNPLLLKRRKK
NFAT	14	LERPSRDHLYLPLE

Table 3. Investigated peptide fragments, number of residues and sequence.

Variation of SCS values for Hα, Cα, and Cβ resonances along the peptide chain can reveal secondary structure propensities in unfolded and partly folded proteins. Here, we examined the behavior of several MAPKAPK peptides in solution using NMR-based secondary chemical shift (SCS) analysis using Equation 5 in Chapter 3.2.3. We determined the SCSHα, SCSCα and SCS(Cα-Cβ) values as they are the most unambiguous indicators of secondary structure propensities: for a region with α-helical tendency, at least four consecutive residues show characteristic values SCSHα<0, SCSCα >0 and SCS(Cα-Cβ)>0. In Figure 12 no tendency can be observed in the SCS values, they do not differ from the random coil values significantly. NFAT is a real IDP with small helical tendency in the D7-L9 region.

We can compare the secondary structure of isolated peptides (RSK1, MNK1, MK2) with their MAPK docking bound state (Figure 13). Notably, the RSK1 peptide displays some helical content in the intervening region between two anchor (or stapling) points (I6-R14, shown in yellow box; corresponding to roughly 20% of helix content compared to a fully helical peptide in this size). We cannot observe other regions showing tendency in their SCS values. Therefore, we can conclude that these regions do not have secondary structure propensity. MNK1 is disordered in its full length, while MK2 has a longer region with nascent helicity (D8-L15) which was supported with (i, i-2) and (i, i-3) cross-peaks in the NOESY spectrum.

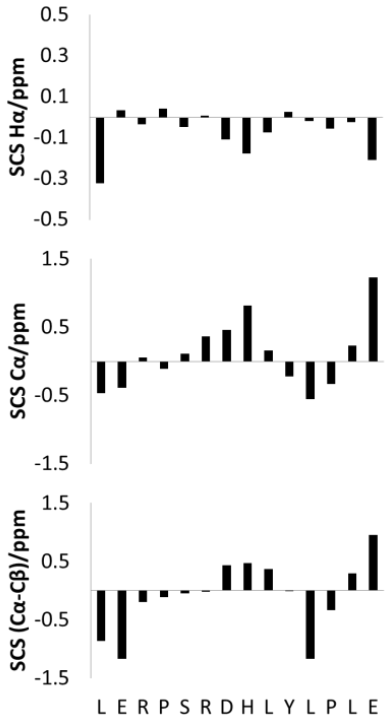


Figure 12. Sequence-corrected Hα (a), Cα (b), Cα-Cβ (c) secondary chemical shifts of NFAT in ppm. On graph (a) L1 and E14 SCS values are greater than the maximum value on the axis.

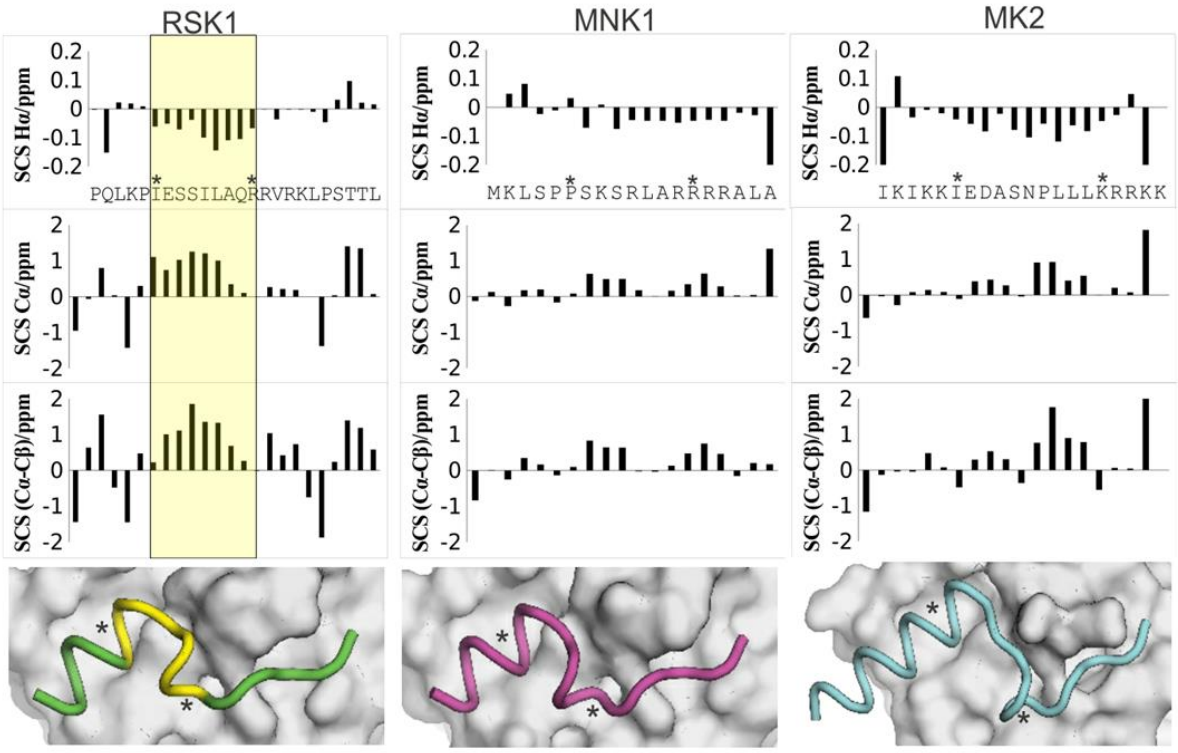


Figure 13. Structural characterization of MAPKAPK linear motifs in solution and in their MAPK bound state. The region with nascent helicity in the RSK1 peptide fragment is boxed in yellow. * indicates the positions of two anchor points. Bound-state figures were drawn by Gergő Gógl.

In order to examine the role of stapling amino acids, a mutant version of RSK1 was synthesized. The structural propensities of RSK1_S/A were also studied but no significant differences were observed compared to RSK1.

This analysis in combination with structure solution of MAPKs in complex with linear motif containing peptides and MAPK–MAPKAPK protein–protein complexes suggests that these MAPKAPK regions undergo disorder-to-order transition upon binding to the MAPK "docking" groove. Regions with nascent helicity were observed for linear motifs with specific binding, while the lack of these regions can be related to promiscuous binding. Thus, we can conclude that we established structure-specificity and specificity-function relationships in the investigated system.

We examined the effect of 8 M urea on the structure of three investigated linear motifs. SCS values of NFAT in aqueous solution and under denaturing conditions did not show significant discrepancies, they were closer to 0 indicating an increase in structural disorder. The same effect was observed for MK2, however, the most notable changes were detected in the region with helical propensity in RSK1_S/A: SCS values measured in aqueous solution were 2-3 times larger as in 8 M urea (Figure A1). We concluded that denaturation lead to a decrease in local residual structure elements; however, these minor changes do not influence the global hydrodynamic parameters (r_H(RSK1_S/A)=15.2 Å in an aqueous solution, while r_H(RSK1_S/A)=12.9 Å under denaturing conditions) .

Characteristic small signal dispersion (~0.6 ppm) on the assigned ^{1}H-^{15}N HSQC spectra of the five investigated peptide fragments (Figure 14) also supports our SCS analysis results of all MAPKAPK fragments being disordered in solution.

Thus, in conclusion we successfully explored the connection between the structural propensities and the function and binding of the studied linear motifs.

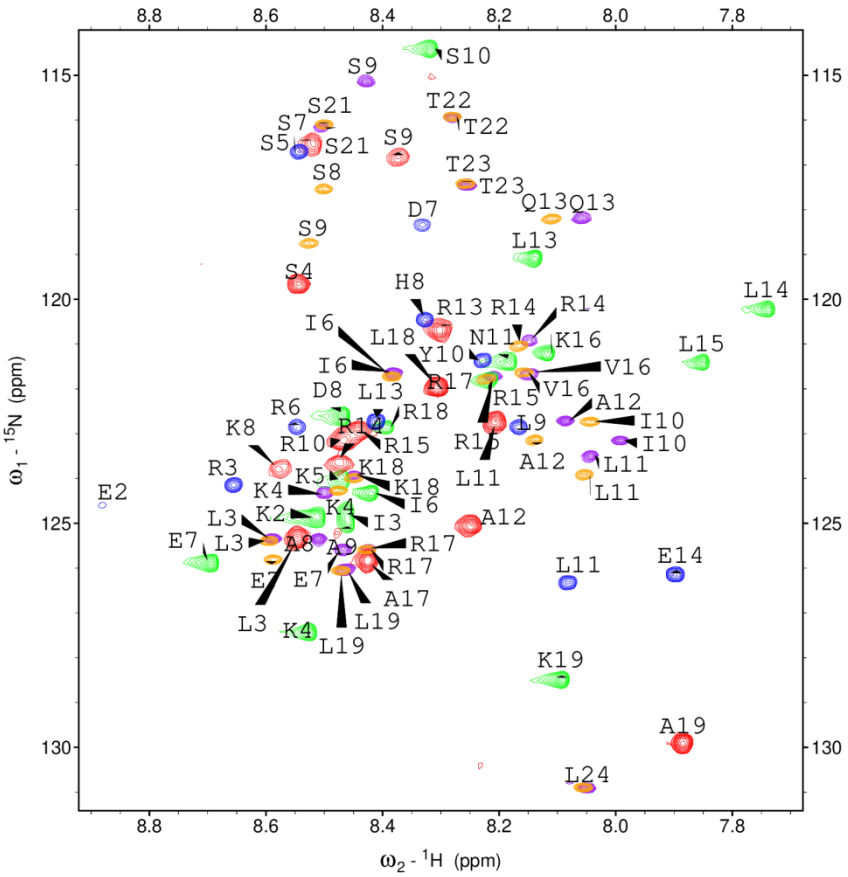


Figure 14. Assigned ^{15}N HSQC spectra of MK2, MNK1, RSK, RSK1_S/A és NFAT.

4.2. The p53-S100A4 complex

Various studies were conducted regarding the sctructure of the N-terminal TAD region of p53 in its full length and in complex with several proteins. One may speculate that being loosely folded is perhaps advantageous for TADs to carry out their promiscuous function, i.e. binding to diverse target proteins; yet, minimal structural elements found in different TAD members are not exactly the same, suggesting some kind of specificity. Lee *et al.* have carried out a detailed multidimensional NMR study on a uniformly ^{15}N-labeled full-length p53 TAD (gsMEEPQSDPSVEPPLSQETFSDLWKLLPEN NVLSPLPSQAMDDLMLSPDDIEQWFTEDPGP) and have found a preexisting amphipathic helix (18-26) and two nascent contiguous β-turns (40-44, 48-53) in its structure. They assumed that upon target binding the preexisting helix tightens into a stable helix and the two β-turns can be conversed into a helix. The full-length p53 TAD can be divided into two synergizing subdomains, an N-terminal subdomain consisting of residues 1–42 and a C-terminal subdomain containing residues 43–73 and both can bind to the same

target protein. However, the N-terminal domain binds more efficiently due to the amphipathic helix.

The most important binding partner of p53 N-terminal TAD domain is Mdm2, a ubiquitin ligase which interacts with p53TAD and ubiquitinates p53's C-terminal regulatory domain. In the absence of DNA damage, p53 cellular level is low as it activates the transcription of Mdm2 triggering its own degradation. DNA damage leads to the weakening of binding in the p53TAD-Mdm2 complex, resulting in increased p53 levels, activating DNA repair. If the binding is too tight – due to enhanced helicity of p53TAD – the timing and duration of these processes get altered, resulting in decreased tumor suppressor function of p53 [150,151]. Chen *et al.* used MD simulations and found that both AD1 and AD2 subdomains of p53TAD mediate its interaction with Mdm2, however, AD1 is the primary binding site, which folds into an amphipathic helix (residues 18-26) [152].

Lee et al. investigated the p53 TAD-NCBD complex utilizing NMR spectroscopy [68]. CBP inhibits the p53-MDM2 interaction and facilitates the p53-mediated stress response and it is also required for stabilization of the p53-DNA complex. They found that in full-length TAD, each subdomain folds upon binding to form a helical motif (P19-L25, P47-W53) connected by a loop (27-39) while the C-terminal region remains disordered. They assumed p53 TAD region as a clamp model with AD1 and AD2 as clamps connected by a disordered linker region. The p53-PC4 complex was also studied via ^{1}H-^{15}N HSQC chemical shift mapping [154]. The PC4-p53 interaction activates p53 for specific DNA binding, enhancing its transcriptional activity. Chemical shift changes were observed in the 35-57, 40-45 regions, while the cross-peaks in the 50-55 region disappeared, suggesting that residues in this region mediate interactions. L22 and W23 signals, and the 48-56 region showed the most significant chemical shift perturbations and Rajagopalan *et al.* concluded that residues in these parts of the protein are key determinants for binding. They also reported about an α-helix induced upon binding. ^{1}H-^{15}N HSQC chemical shift mapping was also utilized for the examination of the p53TAD-hRPA70 complex structure [154]. The hRPA–p53 complex disassociates in response to DNA damage and presumably provides a reservoir of p53 that is immediately available during the early stages of DNA repair. The largest signal intensity reductions were observed for residues 43-56, they became so weak they could no longer be detected.

Residues in the 39-59 range showed the largest resonance intensity changes suggesting that this region is essential for binding hRPA70. SCS analysis proved that residues 19-28 tend to form an amphipathic helix but contributed to binding to a much smaller extent. Cross-peaks assigned to I50, E51 and Q52 disappeared first during titration with hRPA70 suggesting the strongest interaction with it. The previously mention method was also performed on the p53TAD-TFIIH complex which is directly correlated with the ability of the p53 TAD to stimulate transcriptional elongation [155]. Chemical shift mapping resulted in the determination of an amphipathic alpha helical fold in the P47-T55 region, outside this region p53 was considered flexible. The p53 TAD domain they investigated encompassed both AD1 and AD2 subdomains and they found that AD1 is not involved in the binding while AD2 is necessary and sufficient.

S100A4 is a member of small, (10-20 kDa) Ca^{2+} binding-binding proteins which mostly exist as homodimers and it is overexpressed in a wide range of cancers [156-158]. In optical biosensor development it is of great interest to identify its binding partners and study their complexes [152]. In spite of the importance of the p53TAD-S100A4 interaction (see Chapter 2.3), there is only limited structural information available for their complex. To provide further insights into its structural basis we intended to study the atomic-level resolution solution structure of the complex via NMR spectroscopy. Based on previous literature p53 TAD structures in complex we aimed to examine if its structural variability is preserved in complex form, focusing on the presence of dynamic fuzzy regions in the complex.

Peak assignment and sequential connectivities were determined at 313 K from the analysis of three-dimensional standard HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, HN(CA)CO, and HNCO, and standard CCCONH, ^{1}H-^{15}N HSQC, ^{1}H-^{13}C HSQC measurements. NOE cross-peaks were assigned based on 3D HSQC-NOESY, HSQC-TOCSY spectra. ^{1}H chemical shifts were referenced to the internal 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) standard, whereas ^{15}N and ^{13}C chemical shifts were referenced indirectly via the gyromagnetic ratios.

To determine the binding sites on the p53 transactivation domain, we recorded ^{1}H-^{15}N HSQC of free p53TAD and p53TAD in complex. Both ^{1}H-^{15}N correlation spectra exhibit all the peculiar NMR features of an IDP, namely, a reduced chemical-shift

dispersion that is particularly pronounced in the ^{1}H dimension, and we can observe a high degree of signal crowding in the complex spectrum (Figure 15).

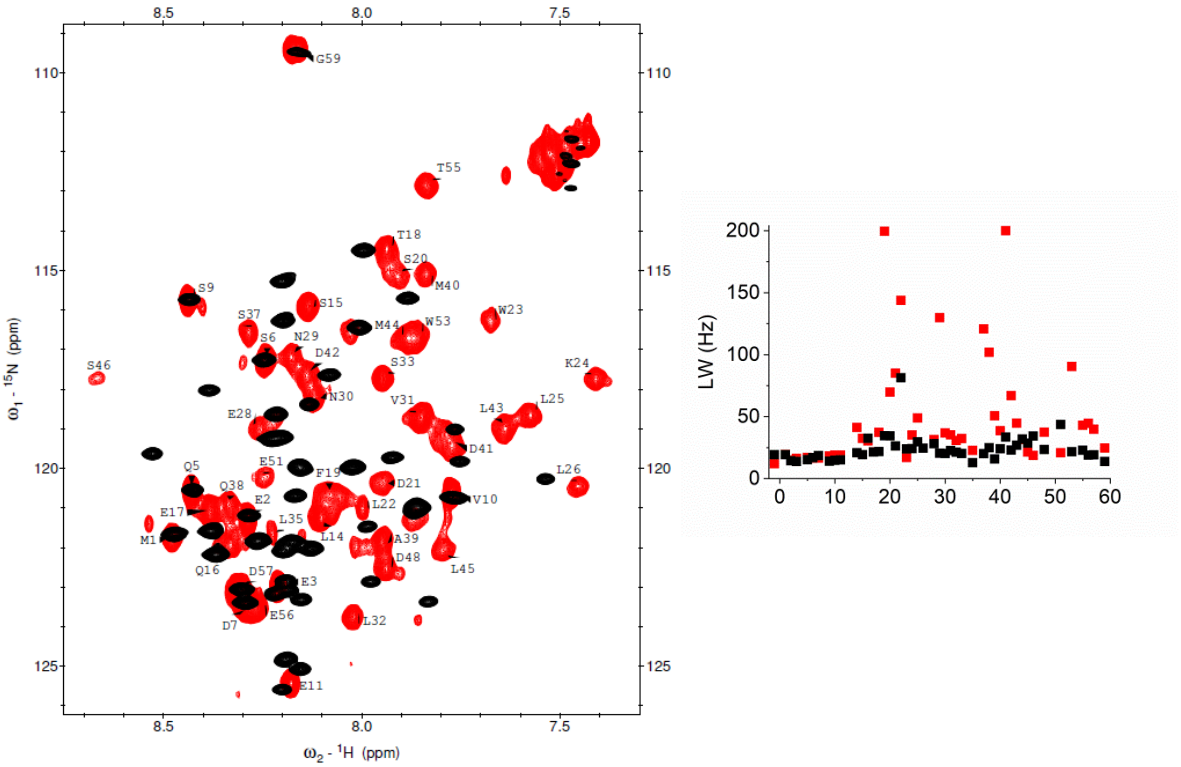


Figure 15. Left: Assigned ^{1}H-^{15}N HSQC spectra recorded at 313 K at 700MHz of ^{15}N-labeled p53 (black) and ^{15}N-labeled p53 complexed with S100A4 (red). Right: the signal width values as a function of the p53 residues.

Significant line broadening can be observed for cross-peaks in the 14-51 residue region upon complex formation due interaction and increase in size increase (from 7 kDa to 31 kDa). If the interacting ligand – in this case Ca^{2+}-bound S100A4 dimer - is of high molecular weight, then signals will broaden due to a slowing down of the tumbling rate. The dissociation constant (K_d) of p53TAD binding S100A4 is 0.76 μM [73] indicates fast exchange which leads to 'exchange broadening'. The linewidth of a peak is exchange broadened due to interference from P + L ↔ PL interconversion during the detection period of the NMR experiments. Other peaks undergo notable chemical shift changes. The signal width values - in the proton dimension - of free p53 are practically constant in the N- and C-terminal regions (-1-16 residues), and small deviations are characteristic in the intervening region, exceeding 40 Hz in only one case (Figure 15). A similar tendency is observable for the complex, however, two well-defined regions (17-31, 37-44) show significant deviations, indicating the presence of the TAD regions responsible

for binding. The signal width values in the proton dimension in these regions range from 40 to 200 Hz.

Binding regions in p53TAD can be detected by calculating the variation of individual chemical shifts upon complex formation. Those residues that possess the most accentuated shifts - shift change is larger than the standard deviation of the shift for all residues - are the ones involved in binding [143]. The cumulative $\Delta\delta$ values (combined from ^{1}H and ^{15}N chemical shifts) can be calculated using Equation 6.

Considerable chemical shift changes were revealed in the 20-31, 37-42 and 51-55 regions indicating the binding regions (Figure 16). Residues with the highest $\Delta\delta$ values are L26, M40 and W53.

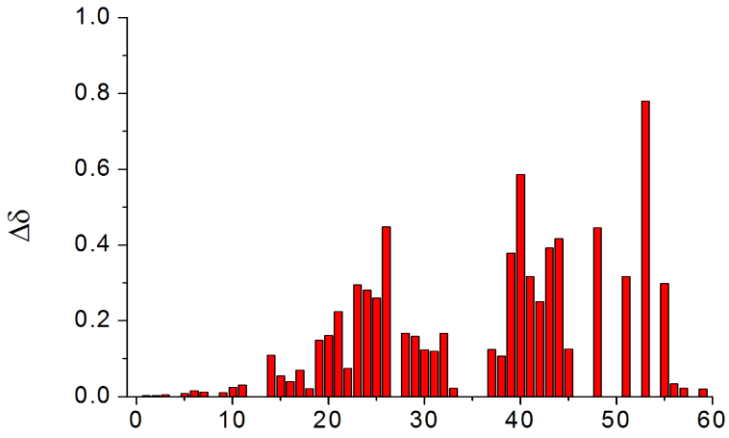


Figure 16. Cumulative Δα changes of p53TAD upon S100A4 binding.

The chemical shifts of $^{13}C\alpha$, $^{13}C\beta$, nuclei can be used to make a rapid, semi-quantitative identification of protein secondary structure by comparing them to random coil values and calculating the Secondary Chemical Shifts (SCSs). Resonance assignments were made for the $^{13}C\alpha$ and $^{13}C\beta$ nuclei, SCS values were determined and plotted in Figure 17.

It is well established that unbound IDPs, such as p53TAD, will often show some preference for the secondary structure that is observed when they form a protein complex: residues 18–24 form an amphipathic α-helix when bound to the MDM2 oncoprotein, NCBD and hRPA70, and region 47-55 also folds into an amphipathic α-helix when bound to NCBD and TFIIH. The SCS Cα and SCS(Cα-Cβ) values plotted in Figure 17 suggest the presence of transient α-helices spanning residues 14-30, 35-44 and 49-54. A transition from positive to negative SCS values between L45 and S46 indicate a β-turn.

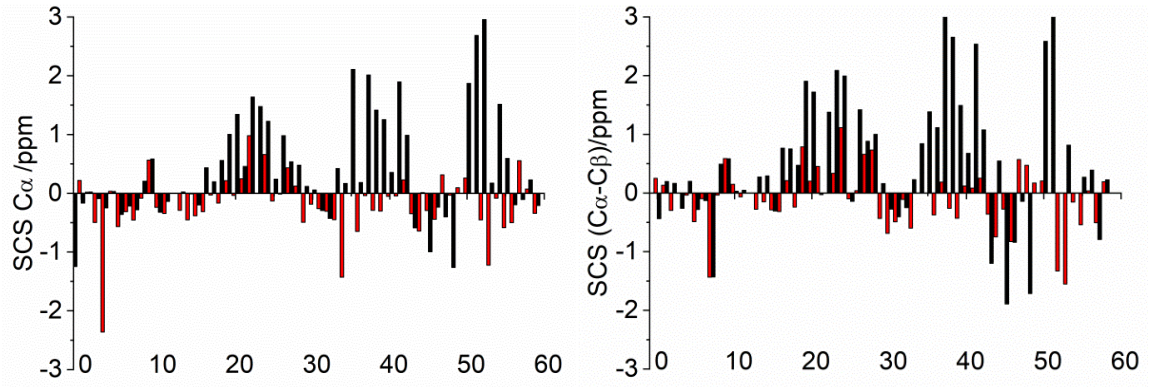


Figure 17. Plots showing secondary chemical shifts for p53TAD and p53TAD bound to S100A4 based on resonance assignments obtained at 313K. The random coil chemical shift standard used in the analysis was developed by Wishart et al. [132]. Left: Plot of SCS Cα values, right: plot of SCS(Cα-Cβ) values.

The presence of SSP at 313 K within p53-S100A4 was supported by comparing experimentally measured heteronuclear chemical shifts (Cα, and Cβ nuclei) with the corresponding random coil values utilizing the SSP approach.

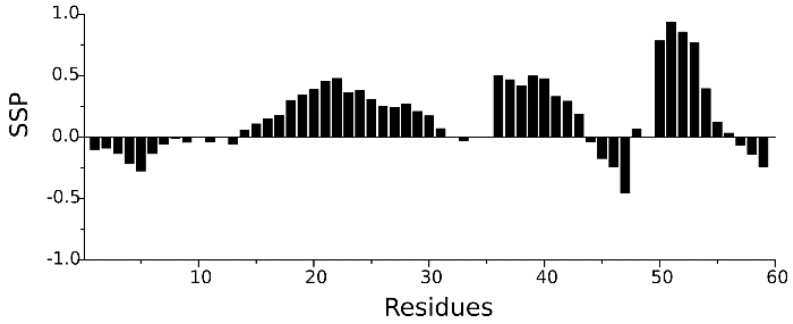


Figure 18. Prediction of structure and dynamics from chemical shifts. The secondary structure propensities (SSP) against residues sequence calculated from Hα, Cα and Cβ chemical shifts for p53 TAD in complex with S100A4.

Based on SSP results (Figure 18) three regions encompassing residues 14-31, 35-42 and 50-55 show preference for an α-helical structure, the strongest preference is observed for the region adjacent to the C-terminal part of p53TAD.

Correlations between amide proton temperature coefficients ($\Delta\delta$(N)H/ΔT) and hydrogen bonds were also investigated for free p53 and S100A4-bound p53 (Figure 19) by acquiring temperature-dependent ^{1}H-^{15}N HSQC spectra in the range of 278-328 K. With the temperature increase significant backbone amide ^{1}H chemical shift perturbations were observed. Less negative temperature coefficients are clustered in the two protein regions that exhibit a higher propensity to form α-helix, in agreement with the results of the other analyses.

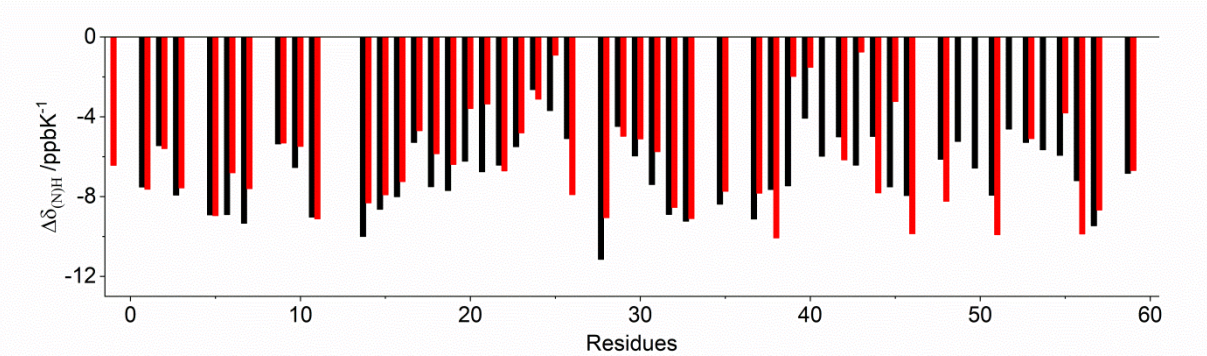


Figure 19. Temperature coefficient values in the 278-328K range against the residues of p53 (black) and p53-S100A4 (red).

The temperature dependence results suggest the presence of transient hydrogen bonds formed by residues S20, D21, K24, L25, A39, M40, L43, L45 and T55.

Binding regions with amphipathic α-helix forming tendencies were identified and 3D ^{1}H-^{15}N HSQC-NOESY spectra recorded to determine the spatial proximity of protons, and to collect distance constraints for structure calculation. Adequate peak intensities and resolution were observed in the disordered regions, however, due to signal broadening in the binding region limited characteristic helical (i; i+2) and (i; i+3) NOE data and structure restraints could be extracted. To overcome this problem 3D ^{1}H-^{15}N HSQC-NOESY spectra were collected at 303 K but no additional NOE data and structure constraints were obtained. Thus, NMR structure calculation was not possible and to address this issue MD simulations were performed. 15 proton-proton distances supported by strong NOE cross-peaks were sufficient as constraints in MD simulations and were used in setting up starting models: W23(N)H-D21Hβ, K24(N)H-D21Hβ, L25(N)H-D21Hβ, L25(N)H-E28Hβ, L26(N)H-E28Hβ, Q38(N)H-P36Hδ, A39(N)H-P36HD, A39(N)H-L43Hδ, M40(N)H-L43Hδ, D41(N)H-A39Hβ, D41(N)H-L43Hδ, D41(N)H-L43Hβ, D42(N)H-M40Hβ, M44(N)H-D42NH, L45(N)H-L43Hδ. The majority of the strong NOE signals originated from the second region with helical propensity. Volume integrals of intra-residual, non-sequential NOE cross-peaks in the helical regions were calculated to prove the spatial proximity (maximum 5 Å) of protons. Models from MD simulations were expected to fulfill as many NOE constraints as possible while maintaining the structural characteristics of the complex.

MD simulations were carried out by Dr. Dóra K. Menyhárd and they confirmed that several different conformationally stable models can describe the p53TAD-S100A4

complex. Two characteristically different arrangements were selected to represent the scope of variations (see Figure 20). The N-terminal segment is relatively similar in the two, with a stable helix running from Q16/E18-E25. In Model A none of the NOE constraints determined by the NMR experiments is violated, but the model is only occasionally helical in the L35-M44 region, while regions Q16-L34 and P49-Q54 are helical in the majority of the conformers. In Model B the C-terminal segment of p53 is shifted toward the solvent and the C-terminal tail of chain B of S100A4 is flipped to an open conformation, with a coupled increase of helicity in the middle of the p53 sequence (L35-M44). Of the 15 long distance NOE constraints only one is violated in this model.

Compliance with the NMR results is especially notable, since constraints were only applied in setting up starting models for the simulations, which were then carried out constraint-free. Model A and B together provide a plausible description for the S100A4 - p53 complex.

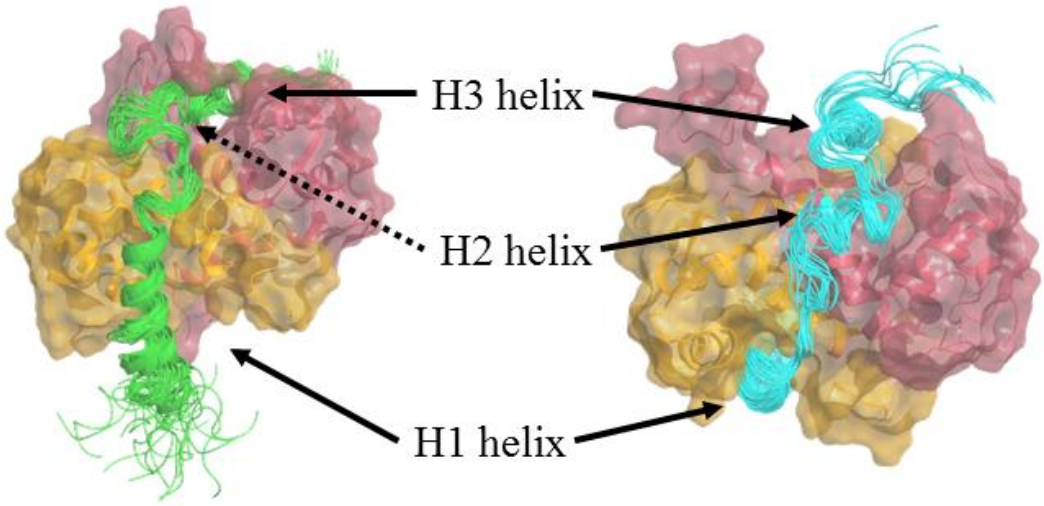


Figure 20. Separate structures of Model A and Model B showing the binding surface and the helical segments. (Chain A of S100A4 is orange, chain B red, Model A is colored green, while Model B is shown in cyan). Molecular Dynamics simulations were performed by Dr. Dóra K. Menyhárd.

The C-terminal loops of both chains of S100A4 participate in forming the complex R99, K100 and K101 of chain A form H-bond contacts with D21, E28, D30, and V31 of p53 (in the first helical region) in both models, while the C-terminal Q97, K100 and K101 of chain B interact with M40, D41 and D42 of p53 (in the second helical region) in Model A, and are immersed in the solvent in Model B. H-bonds are also observed between residues S20-G16, S20-K24, K24-D21, L25-D21, L25-L22, L25-E28, L25-N29, M40-S37, M40-E43 and T55-E51. There is also a very specific hydrophobic fit between

F45, L62, F78, M81, M85 of S100A4 chain A and F19, L22, L26 and P27 – residues in the first helical region - of p53 in both models. H-bond contacts and hydrophobic interactions are reflected in temperature coefficient values for several residues including S20, D21, K24, L25, A39, M40 and T55.

Backbone dynamics measurements and analysis were performed by Gyula Pálfy, therefore, I only intend to summarize these results shortly and compare them with structural characteristics: backbone relaxation analysi results showed that p53TAD fragment in complex with S100A4 has a highly mobile long N-terminal (M1-P13) and short C-terminal (E56-P60) region which do not participate in complex formation. It has three helices with rigid structure in regions S20-E28, P36-P47, E51-T55. Relatively dynamic regions can be found between the helices (N29-L35 and D48-I50) and in the region which connects the N-terminal flexible end with the first helix (L14-F19). These regions form loops between the helices. There are three residues showing conformational exchange located in the first helix (W23, K24) and in the second flexible loop (L32).

We can conclude that p53 TAD fragment in complex with S100A4 has a remarkably mobile long N-terminal (1-13) and short C-terminal (56-60) regions that are not involved in complex formation. Three α-helices with rigid structure were observed in regions S20-E28, P36-P47, E51-T55 with relatively dynamic regions between them – forming loops - and between the N-terminal end and the first helix. Characteristic temperature coefficient values determined via NMR measurements were in agreement with H-bonds and hydrophobic interactions or with conformational exchange (S20, D21, K24, L25, A39, M40 and T55). The results demonstrate the significance of combining various methods to successfully characterize protein structure in ‘challenging’ systems. Our results indicated that both TAD1 and TAD2 subdomains of p53 undergo disorder-to-order transition upon complex formation and mediate the binding to S100A4. TAD1 and TAD2 regions fold to form helical motifs and they function as clamps connected by a disordered linker. A clamp model of binding has been observed for other intrinsically disordered proteins, leading to an enhancement of binding affinity compared to that of the isolated motifs and providing flexibility and adaptability in molecular interactions.

The p53TAD interacts with several proteins, thus we can compare its structural propensities in the various complexes (See Appendix Figure A2): The TAD1 region is the

primary binding site for Mdm2 and Mdmx and folds into an amphipathic helix (18-26) upon complex formation. The p53TAD interacts with replication protein A (RPA) and the Tfb1/p62 subunit of TFIIH by way of the AD2 subdomain. Binding to RPA causes AD2 to fold to two short helices, located between residues 41 and 44 and between residues 47 and 55. A transient amphipathic helix (19-28) contributes to binding, as well, but to a much smaller extent. Residues 47-55 also fold into a helical structure when AD2 binds to the Tfb1 subunit of TFIIH. Binding to PC4 and CBP/p300 is also dominated by interactions with AD2. Lee et al. showed that both AD1 and AD2 interact directly with the protein NCBD, and each subdomain forms a helical motif (19-25, 47-53) upon binding [68]. Residues 40-45 of p53, which adopt helical structure in the RPA complex, also contribute to the binding to the protein NCBD without formation of regular secondary structure. The helical regions of the p53TAD are very similar in the complexes with the NCBD, MDM2, RPA, and Tfb1, suggesting that formation of a conserved local structure is a feature of p53 recognition. As the results showed atomic resolution characterization via NMR spectroscopy plays an important role in understanding protein function.

4.3. A diffusion NMR based bioanalytical method to distinguish folded, disordered and denatured proteins

The molecular size of a protein is a parameter of high importance describing the aggregation state, conformational changes and the degree of (un)folding. It provides a useful tool to differentiate folded, disordered and denatured proteins by determining their hydrodynamic parameters. These parameters can be quantified by establishing empirical formulae. Empirical relations from the literature show discrepancies, therefore, we intended to propose more sophisticated ones by performing a systematic study based on PFG-NMR measurements. We aimed to derive empirical variations of D as a function of molecular mass M, thus, we had to examine whether the Stokes-Einstein relation (Equation 1) is applicable for our systems. In practice, we had to investigate the contributions incorporated in form factor F. For biomolecules in aqueous solution the influence of the solvent on F is described by the Gierer and Wirtz formula.

$$F = \left(\frac{3\alpha}{2} + \frac{1}{1+\alpha}\right)^{-1} \text{ where } \alpha = \frac{r_{\mathrm{H}}(\text{solvent})}{r_{\mathrm{H}}(\text{solute})} \qquad (8)$$

Since solvent water molecules are much smaller than the solute molecules, α~0 and the value of *F*=1. To test the occurrence of molecular crowding in our samples, the *D* of water in the presence and absence of protein was also measured and in the case of 1mM lysozyme we obtained identical values suggesting the lack of this effect. In addition, we determined the diffusion activation energy by measuring the temperature dependence of the diffusion coefficient and using the corresponding form of the Arrhenius equation.

$$D(T) = D_0 \mathrm{e}^{-\frac{E_\mathrm{a}}{RT}} \tag{9}$$

which can be linearized to

$$\ln D(T/\mathrm{K}) = -\frac{E_\mathrm{a}}{R}\frac{1}{T} + \ln D_0(T/\mathrm{K}) \tag{10}$$

We fitted a linear on the data points and found that the slope gives a 20.6 kJ/mol value in perfect agreement with the value obtained by Roos *et al.* for the diffusion of water molecules (20 kJ/mol) [156]. Based on these results, we concluded that molecular crowdedness is negligible and also that the continuum model is viable for typical protein/peptide samples

It should be noted that several proteins have a tendency to form aggregates and it is necessary to test whether the investigated system is in a monomeric state or not. In our samples – in the μM-mM concentration range - the monomeric condition was fulfilled. With selected proteins we examined the concentration dependence of the diffusion coefficient and found that the variation was within the error limit.

We performed our systematic study at 287 K analyzing peptides and proteins with diverse length, charge-distributions and amino acid sequence (sequence, conditions, characterization is given in Appendix Table A2). Already the 1D ^{1}H spectra provide qualitative information about discrepancies between folded and disordered molecules. The dispersion of signals in the spectrum of the folded protein is far beyond the envelope of signals seen in the spectrum of the IDP. This clearly reflects that nuclei in proteins with well-defined structure are subject to numerous different types of microenvironments while nuclei in IDPs can be found in similar types of chemical environments corresponding to the random coil state (Figure 21).

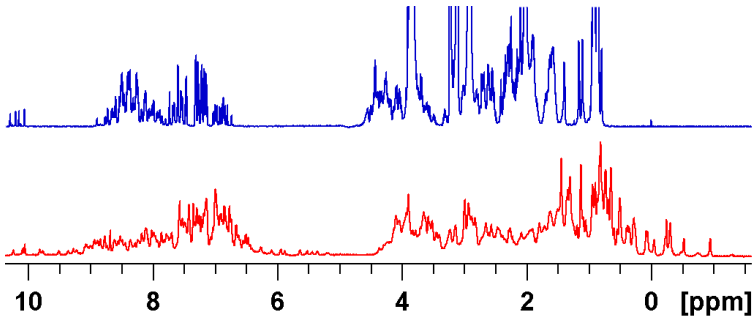


Figure 21. 1D ^{1}H NMR spectra of a folded protein (red) and an IDP (blue)

The translational diffusion coefficients were determined using PFG-NMR measurements, as described in detail in the Experimental section (Figure A3). The variation of the calculated D values as function of molecular mass for the studied 12 folded proteins and 10 disordered fragments is presented on Figure 22.

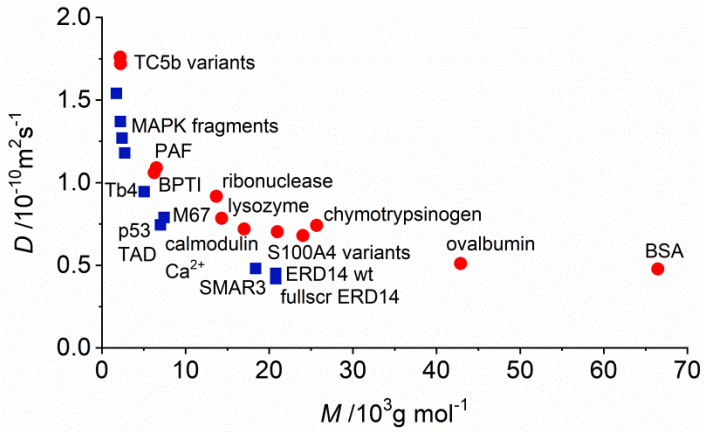


Figure 22. Variation of the translational diffusion coefficients with molecular weight for folded proteins (red circles) and IDPs (blue squares).

A rapid decay is observed for both protein families until cca 20000 g/mol; above this threshold there is no significant variation of D values with the molar mass. This phenomenon can be explained by the cooperative effect of weak molecular interactions, and is in agreement with the earlier observation that the maximum size of protein domains approaches 200 residues [157,158]. A clear discrepancy is also observable between the behavior of the two protein classes and the fitting of these decays resulted in the following empirical formulae:

$$D(\text{folded}) = 3.16 \cdot 10^{-9}(M[\text{g/mol}])^{-0.381}\ \text{m}^2\text{s}^{-1} \qquad (11)$$

$$D(\text{IDP}) = 6.78 \cdot 10^{-9}(M[\text{g/mol}])^{-0.507}\text{m}^2\text{s}^{-1} \qquad (12)$$

For easy and rapid bioanalytical application of molecular mass determination the linearized equations should be used:

$$\log D\ (\text{folded}) = -0.381 \log M - 8.499 \quad (13)$$

$$\log D\ (\text{IDP}) = -0.507 \log M - 8.169 \quad (14)$$

These formulae can be applied for the estimation of folding, thus, for distinguishing between different protein types. A 10% aggregation or change in molecular weight leads to a 'shift' in the diffusion coefficient value, which is beyond the error limit, therefore, this extent of aggregation can already be determined by using the proposed formulae. One must note that the two protein classes have a significant difference in shape. Folded systems resemble more a compact sphere with the exponent 0.381 similar to the theoretical value from Equation 1, while the exponent 0.507 determined for IDPs is indicative of more elongated, loose structures. Scaling parameters of 0.52-0.55 were established for poly(ethylene-oxide) (PEO) in D_2O and 0.55 in H_2O solutions, therefore, a similarity to polymer solutions can be assumed.

Diffusion coefficients are the starting point in the evaluation of the effective hydrodynamic radius r_H (Equation 1). In order to obtain the most reliable data we checked both methods enlisted in the literature:

i) the absolute method is based on direct calculation from the Stokes-Einstein relation and requires exact knowledge of solvent viscosity at the given temperature. Our measured viscosity values for this purpose are enlisted in Table 4 for water and for 8M urea solution.

T (K)	Water		8M urea	
	shear stress (Pa)	η (Pa·s)	shear stress (Pa)	η (Pa·s)
283.0	0.26	1.29E-03	0.43	2.17E-03
288.0	0.23	1.15E-03	0.38	1.88E-03
293.0	0.20	1.01E-03	0.34	1.69E-03
298.0	0.18	8.91E-04	0.30	1.52E-03
303.0	0.17	8.31E-04	0.28	1.39E-03
308.0	0.15	7.42E-04	0.25	1.25E-03
313.0	0.14	6.92E-04	0.23	1.14E-03

Table 4. Measured water and 8M urea viscosity values (shear rate was constant = 200 s^{-1})

ii) the relative method avoids the viscosity issue by including an internal reference molecule in the protein solution. Traditionally dioxane is used for this purpose as it is inert and does not interact with the protein, presenting one resonance peak at 3.54 ppm and constant r_H. However, the exact hydrodynamic radius value contradictory in literature: both 1.7 Å [159] and 2.12 Å [82] are used. Following this approach, the r_H of the macromolecule is calculated from the ratio of measured diffusion coefficients $r_H^{prot}=(D^{dioxane}/D^{prot})\cdot r_H^{dioxane}$. The disadvantage of this method is the peak overlap with protein resonances; causing higher uncertainties and larger errors in the evaluation.

On one hand, the r_H value from the relative method is in better accordance with the reference r_H value for lysozyme, but for ovalbumin it shows bigger discrepancy. On the other hand the relative method shows significantly larger error values for both proteins (±2.0 and ±1.4 Å) compared to the absolute method (±0.4 and ±0.5 Å). Peak overlap with protein resonances is another disadvantage of the relative approach. Our data suggest that in case of knowing the proper viscosity values, there is no need for a reference molecule, therefore, in our further analyses we employed the absolute method

It has been already discussed, that literature predominantly focuses on the relationship between r_H and residue number. This originates from polymer theory [160] where the polypeptide chain, similarly to a polymer, is treated as freely jointed chain, which consists of N statistical segments with given dimensions connected by virtual bonds. However, the hydrodynamic radius is already a derived number, which inherently carries the errors of the applied method of measurement and the spherical assumption, Moreover, the application of residue numbers ignores differences in molecular size and weight between peptides with equal chain lengths but different amino acid composition. Therefore we believe that it is more appropriate to directly compare the measured, primary parameters, in our case diffusion coefficients. Nonetheless, we also derived residue-based correlations using Equation 1 for the comparison with literature data:

$$\mathrm{r_H(folded)} = 3.405\mathrm{N}^{0.382}\ \text{Å} \quad (15)$$

$$\mathrm{r_H(IDP)} = 3.128\mathrm{N}^{0.492}\ \text{Å} \quad (16)$$

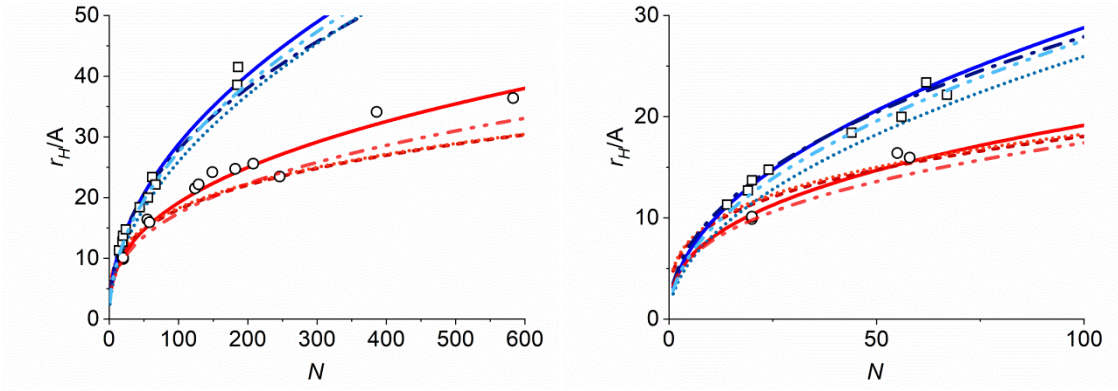


Figure 23. Left: Variation of r_H values with the residue number for folded (open black circles) and intrinsically disordered proteins (open squares). Continuous curves represent fittings of empirical relations for folded (red) and disordered proteins (blue) determined in the present study (solid line), by Wilkins (short dashed line), Uversky (dash-dot-dot line), Bernadó (dash-dot) and by Marsh (dotted line). Right: Zoomed for N=0-100 region.

The validity of our equations in comparison with literature was tested (Table A5). It is expected that more compact folded proteins can be described with smaller r_H values but literature data (Figure 23) show discrepancies, especially for moderate N values. Our r_H values for IDPs are higher than the ones predicted by formulae from literature, a difference of 4-6 Å is observed for proteins with 50-200 residues. For small (typically N<100) folded systems our predicted values fall between literature predictions, and we obtained systematically higher r_H values for N >100 residues (differences are 7-8 Å). These disparities could be explained by the fact that literature relations are based on data collected under different conditions and employing diverse methods. Discrepancies might originate from the diversity in molecular shape, therefore we tried to calculate shape factors using a set of folded proteins with already determined and deposited 3D-structures. Based on pdb coordinates the HYDROPRO [161] software calculates the corresponding D_{calc} value from the diffusion tensor, and gives the radius of gyration r_G as well. The r_G/r_H ratio reports on molecular shape, the characteristic value for a hard sphere is 0.77[87]. The higher the r_G/r_H ratio, the more pronounced is the distortion from the spherical shape (see Table 5). Molecules with elongated shape possess shape factor values close to 1.00, thus r_H will become higher than expected (ovalbumin and BSA). Consequently, we suggest the application of $D(M)$ relations for analytical purposes.

Protein/PDB code	D_{calc}(m^2s^{-1})	D_{exp}(m^2s^{-1})	r_G/r_H
TC5b/1l2y	1.41E-10	1.76E-10	0.94
dPAF/2mhv	1.21E-11	1.06E-11	0.79
ribonuclease/2e3w	8.08E-11	9.17E-11	0.83
lysozyme/1lys	9.00E-11	7.84E-11	0.65
S100A4wt/1m31	6.54E-11	6.80E-11	0.77
chymotrypsinogen/1ex3	7.61E-11	7.41E-11	0.70
ovalbumin/1ova	3.84E-11	5.10E-11	1.00
BSA/3v03	3.74E-11	4.80E-11	1.03

Table 5. Selected folded proteins from PDB; derived and measured diffusion coefficients; the calculated r_G/r_H ratio

Another approach to analyze diffusion data is by expanding r_H in Equation 1. This way shape information will be incorporated in the effective density. For X-ray structural analysis protein density is a key factor and several theoretical and experimental approaches were proposed. For small organic molecules a single value of 620kg/m^3 can be a good approximation [84], but in case of proteins the density will depend on the chain length, and will be influenced by shape, solvation and flexibility. The Voronoi method [162] distinguishes between buried atoms, that are inaccessible by solvent molecules (zero solvent accessible area), and exposed atoms with varying number of water neighbors. For a set of folded proteins the average density value was found 1.47g/cm^3 for buried atoms and 0.87 for exposed atoms [162].

As expected, the effective densities of our folded systems show higher values than IDPs and their denatured forms, and no specific variation with the molecular weight can be observed (Figure 24). Indeed, for native globules the density is more or less independent of the chain length [163].

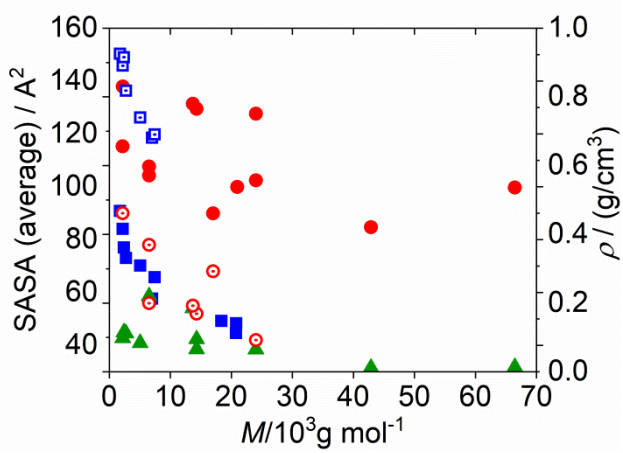


Figure 24. Average SASA values of folded (empty red circle with a dot) and disordered (empty blue square with a dot) proteins and effective/hydrodynamic density of folded (red circle), disordered (blue square) and denatured (green triangle) proteins plotted against the molecular weight

For the more solvated IDPs the $\rho(\mathrm{kg/m^3}) = 569 \cdot M(\mathrm{kg/mol})^{-0.52}$ decay was obtained by fitting our measured data. This expression is in agreement with the theoretical description for the density variation of macromolecules in an "ideal" solvent suggesting the -0.50 power of M [163]. Moreover, studies on proteins under highly denaturing conditions (6M GdnHCl) describe an apparent density variation of M having -0.64--0.66 exponents [163]. The correlations we observe indicate that IDPs behave similarly to denatured unfolded proteins.

Several experimental studies were dedicated to characterize the unfolded states under denaturing conditions [82,83,163]. Yet, it is still debated when is the denaturation end-point achieved and which chaotropic agent to use. We chose the most commonly applied 8M urea media and investigated its effect on the behavior of 7 IDPs and 6 folded proteins. We expected that regardless of the initial structure under high denaturant concentrations the protein will be present in a completely unfolded "random coil" conformation. Consequently, a single linear equation could be fitted on all denatured data points on the linearized representation. However, our data points are more spread (see Figure 25), and more likely to be fitted by two linear equations.

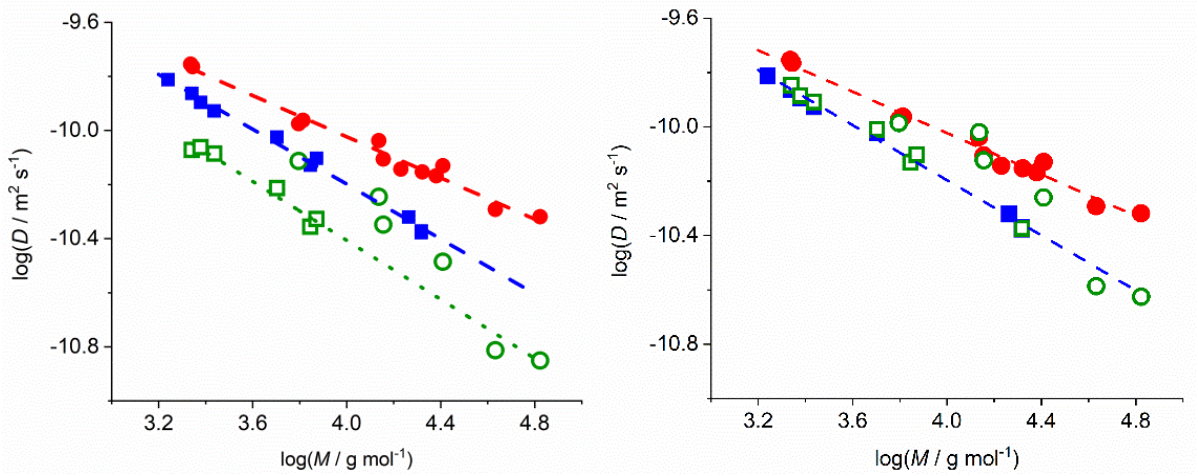


Figure 25. Logarithmic representation of diffusion coefficients as function of molecular weight for folded proteins (red circles), IDPs (blue squares), left: denatured folded proteins (open green circles) and IDPs (open green squares) without viscosity correction, right: viscosity corrected denatured proteins.

The log*D* variation of denatured IDPs as a function of log*M* is similar to the one obtained for aqueous solution, and the two fitted lines are almost parallel. If we want to compare the real extent of denaturation, then a viscosity correction is required. The ratio between the viscosities of aqueous and denaturing (8M urea) solutions is practically temperature independent (see Table 4) and can be applied to obtain a viscosity corrected value for *D*. This viscosity ratio can be determined also experimentally measuring the diffusion coefficient of dioxane in the given solution and in another sample under denaturing conditions. Our results show, that corrected values for denatured IDPs are very similar to the corresponding values measured in aqueous solution (Figure 25). For folded proteins almost no change can be observed if disulphide bridges are present in the molecule (they are not disrupted by urea). When a protein starts to denature, then its diffusion coefficient will fall between the folded and IDP lines, while corresponding data points of completely denatured systems will be situated along the disordered line. These experiments reveal two features:

1) IDPs cannot be further denatured in 8M urea, in accordance with the above depicted effective density variation. However, decrease in local structural tendencies can be detected – as seen in Chapter XY for MAPK linear motifs under denaturing conditions – but these do not affect the global hydrodynamic parameters which represent the highly mobile ensemble. Moreover, no significant differences were found between r_H values calculated for IDPs in the

present work and r_H values from the available literature variation for denatured proteins (Figure 26).

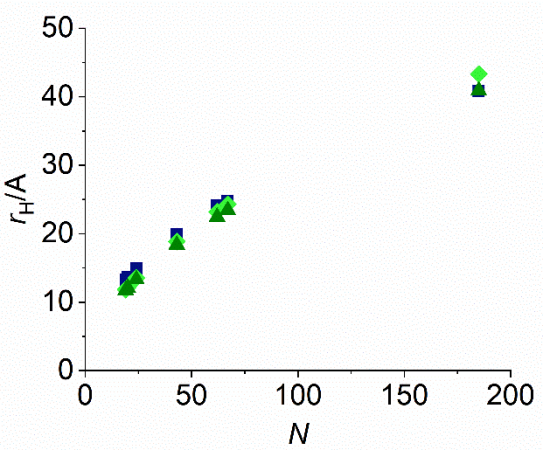


Figure 26. Calculated IDP r_H values (blue squares) using the $r_H(N)$ relation of the current study, denatured protein r_H values based on the relations determined by Wilkins (light green diamonds) and Marsh (green triangles).

2) The ^{1}H 1D spectrum provides qualitative information (Figure A4), whether the structure is collapsed or unaltered, while the diffusion measurement reveals whether the final unfolded state has been achieved under the applied denaturing conditions. If not, then the degree of denaturation can be estimated.

In conclusion we provided new logarithmic relations that can be utilized to gain reliable information about protein structure and compactness. With the obtained data we are able to characterize aggregation state and shape and their unfolding during denaturation. The measurements are easy to carry out; insensitive to isotopic labelling of the protein (keeping in mind the corresponding molecular mass correction). The method provides both bioanalytically and chemically useful results, and potentially enhances the use of diffusion NMR experiments in the structural characterization of proteins.

4.4. Combining NMR and SAXS – parameter optimization

The shape of a molecule can be assessed by determining the shape factor r_G/r_H, which is described as a useful, sensitive indicator of the molecular conformation and morphology. For reference, a solid sphere yields a typical shape factor value of $(3/5)^{0.5} \approx 0.77$ [167]. As the shape becomes more elongated r_G increases relative to r_H and r_G/r_H

increases; shape factors above 1.4 indicate a shape characteristic for denatured proteins while values above 1.7 refer to rod-like molecules. Various shape factor values were reported for polymers in solution, 1.5 [168], 1.26 [87(a)], 1.17 [87(a)]. The shape factor is widely applied for characterizing the shape of polymers [168-170] dendrimers [87(a),171], and proteins [172-177] in solution.

PFG-NMR measurements provide translational diffusion coefficients and r_H values – as presented in Chapter 3.2.1 - while r_G values can be obtained by performing SAXS experiments. No attempts have been made in the literature so far to carry out measurements utilizing these two techniques on the same sample under the same conditions; however, data collected this way are more reliable than using results from other studies, particularly for calculating the shape factors. For this reason, we intended to combine the methods. But first, we needed to optimize the measurement conditions (concentration, ionic strength) on lysozyme samples. SAXS experiments and data analysis was carried out by Dr. András Wacha and Dr. Attila Bóta. Dr. András Wacha determined the radius of gyration from the small-angle X-ray scattering curves, performed shape determination and he generated protein model structures.

In the concentration range of 0.25-2 mM a small increase is observable in the r_H values of lysozyme (Figure 27). The scattering curves indicate the presence of repulsive electrostatic interparticle interaction that causes reduction in the intensity at the small q-range. This is disadvantegous, because the Guinier range is distorted, precluding the accurate determination of the radius of gyration and the reconstruction of the shape.

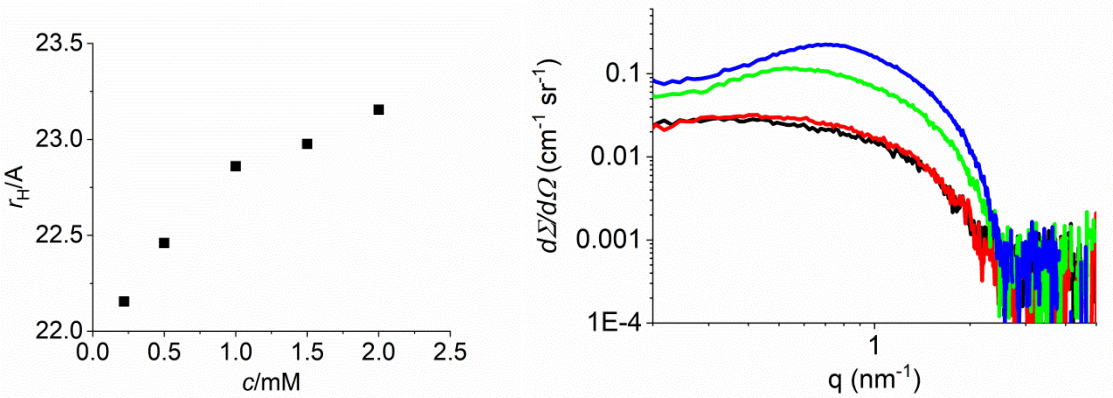


Figure 27. Left: Hydrodynamic radii as a function of concentration for lysozyme. Right: Concentration-dependent scattering of lysozyme solutions with concentration values: 2mM (blue), 1mM (green), 0.33mM (red) and 0.2mM (black). The scattering curves were measured by Dr. András Wacha.

To overcome this issue one can dilute the sample or increase the ionic strength of the solution. Reducing the protein concentration partially resolves this problem, although the effect is still visible at the lowest concentration and dilution comes with the cost of weaker scattering signal. Considering this and our main goal, we decided to follow the second approach and added NaCl to the sample to resolve the inter-particle interaction to shield the electrostatic repulsion between the protein molecules.

c_{NaCl} mM	D ($10^{-10}m^2s^{-1}$)	r_H (Å)
0	1.16±0.04	19.9±0.4
150	0.92±0.01	19.4±0.8

Table 5. Diffusion coefficient and hydrodynamic radius values of lysozyme in solutions with different ionic strength.

As it can be seen in Table 5 the apparent hydrodynamic radius is not affected by the ionic strength. As the slope of the SAXS curves at low angles at the two highest salt concentrations are approximately zero, we concluded that the issue of inter-particle interactions was resolved (Figure 28).

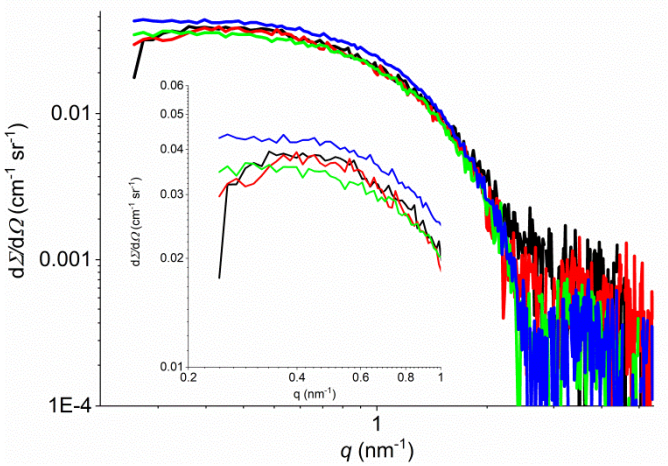


Figure 28. Full scattering curves of lysozyme in solutions with various NaCl concentration: 0mM (black), 25mM (red), 77mM (green), 150mM (blue). The scattering curves were measured by Dr. András Wacha.

After careful parameter optimization we determined the shape factor of the thoroughly studied protein lysozyme. The Guinier radius was found to be 14.2±0.1Å and the hydrodynamic radius was 19.4±0.8Å yielding a ratio of 0.73 which is very near to 0.77, the literature value for a solid sphere [87]. A specific region of the scattering curve

can be used for shape determination and the model resulting from it matches remarkably with the literature structure [178] (Figure 29).

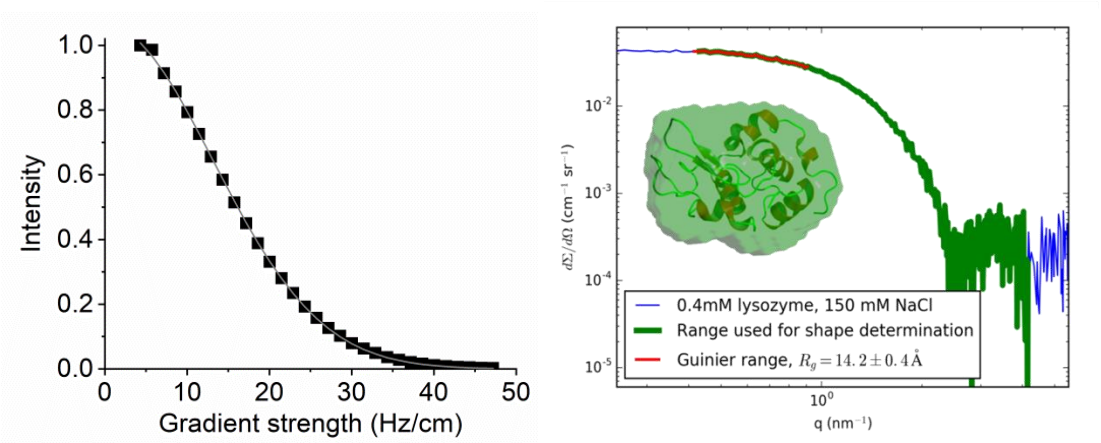


Figure 29. Left: PFG-NMR data evaluation for lysozyme: signal intensity decays as a function of increasing gradient strengths. Black squares represent experimental data points, continuous grey curve represent Gaussian fitting on the data points. Right: The result of shape determination performed by Dr. András Wacha and the SAXS scattering curve of lysozyme with specific regions for Guinier fitting (red) and shape determination (green). The scattering curve was measured by Dr. András Wacha.

To investigate the applicability and/or the limits of the approach, we decided to test our methodology on a protein which undergoes shape change induced by altered solution conditions. We chose the highly conserved protein calmodulin (CaM) which has various, highly important functions in cellular regulation. The name is an abbreviation of *cal*cium *modul*ated prote*in*, because its shape changes upon binding Ca^{2+} ions, making it a key participant in the calcium-induced signalling pathways.

The crystal structures of calcium-free ('apo') and calcium-bound CaM shown in Figure 30 (results of Kuboniwa *et al.* [179], PDB ID 1cfd and Babu, Bugg, and Cook [180], PDB ID 3cln) show the characteristic property of this molecule: a Ca^{2+}-induced shape change. In the apo conformation the two terminal domains of the dumbbell-shaped protein are linked by a flexible loop. Calmodulin can bind four Ca^{2+} ions (two at each end), thereby forming two so-called EF-hand motifs, essentially, making two hydrophobic pockets accessible for grabbing non-polar targets [181]. During this, the originally flexible linker adapts an α-helix conformation, increasing the distance between the two ends of the molecule.

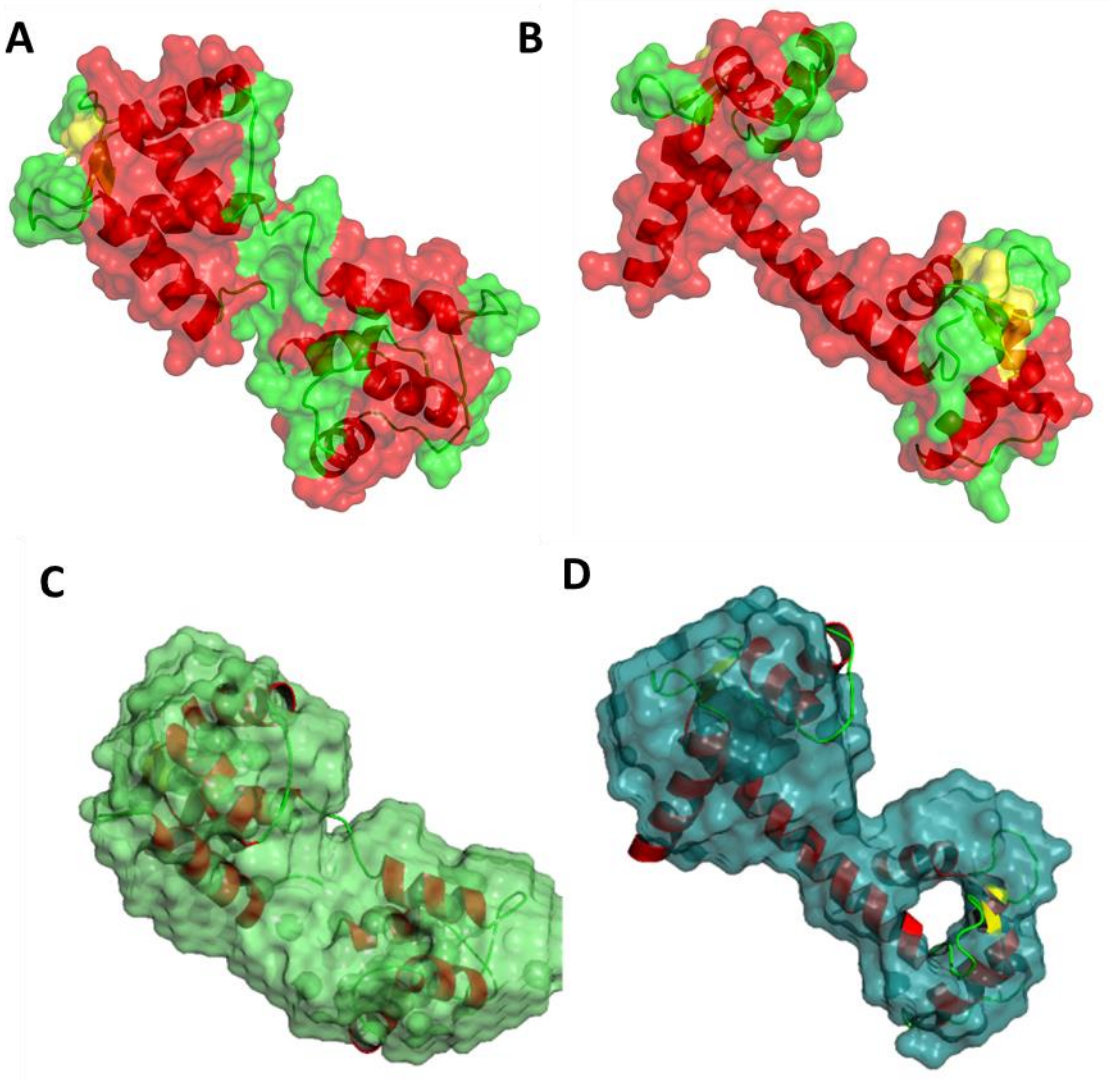


Figure 30. Structure of calmodulin. A: calmodulin in the apo state, B: Ca^{2+}-bound calmodulin. The coloured envelope is the Van derWaals surface, determined from atomic positions. Colours correspond to the secondary structure: α-helix (red), β-sheet (yellow), loop (green). Structures were generated by the PyMOL program. C, D: Calmodulin structures determined by Dr. András Wacha from SAXS measurements aligned with the corresponding crystal structures. Left: apo conformation. Right: Ca^{2+}-bound state.

Even though calmodulin undergoes considerable conformational changes upon calcium binding, it is not reflected in the r_H values (24.5±0.1 Å for the apo form and 24.4±0.1 Å for the Ca^{2+}-bound form). This observation is further supported by the two similar r_G values 19.4±0.6 Å and 20.7±0.6 Å. *Ab initio* shape determination yielded in two models that can be seen on Figure 30: the two end-domains in the Ca^{2+}-less/apo conformation are bent towards each other, while they are clearly separated in the Ca^{2+}-bound state, as if waiting for the target to grab.

In conclusion, the Ca^{2+}-induced changes in structure and size of human calmodulin are reflected in the models based on fitting the dedicated range of the scattering curve, however, they cannot be simply measured by using NMR or fitting the Guinier region of SAXS curves, as they do not lead to changes in the global parameters of the protein, in accord with the results of Panjkovich *et al* [182].

These studies demonstrated that our developed and validated NMR-SAXS approach can successfully provide new, scientifically relevant information. During our investigation, we ensured the reliability of our data by optimizing the reaction conditions and by carrying out the NMR and SAXS measurements on the same samples.

4.5. Bicelles and bicelle-peptide systems

During the course of my PhD work I intended to find a straightforward and powerful method to screen and characterize the variations in global properties – shape, size and curvature - of membrane mimetic systems occurring upon interactions with proteins. I aimed to utilize the previously described synergy of NMR and SAXS methods for further progress and give a more precise description. I explored the effect of two topologically different peptides, namely: the surface-active peptide melittin and the model transmembrane peptide KALP23 and tested the applicability of deriving shape information from the r_G/r_H ratio.

SAXS measurements and data analysis was carried out by Dr. András Wacha and Dr. Attila Bóta. They determined the radius of gyration from the small-angle X-ray scattering curves and he also performed ellipsoidal model fitting while I measured the translational diffusion coefficients and calculated the hydrodynamic radii.

PFG-NMR measurements were carried out as described in Section 3.2.1. Signal variations of all lipid environments were analyzed. As expected, they followed a Gaussian decay (Figure 31) and they were subject to a series of non-linear fitting based on the Stejskal-Tanner equation (Equation 2).

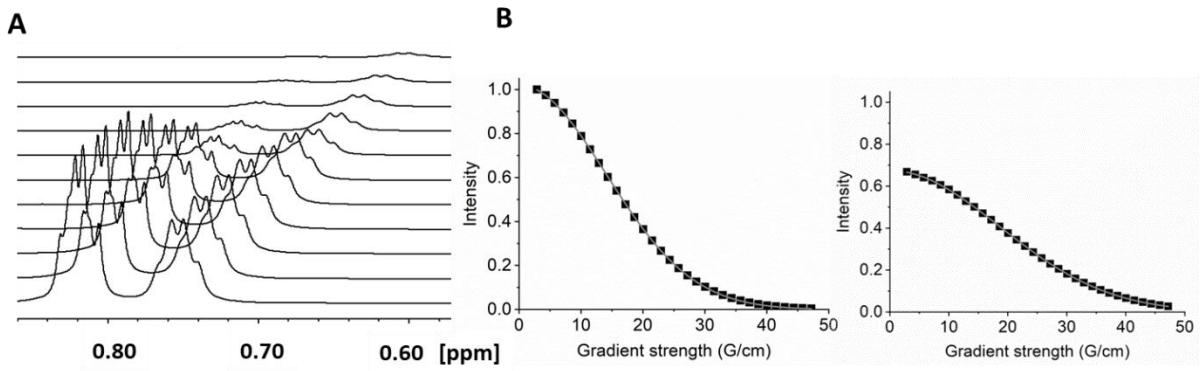


Figure 31. A) Signal decay in the methyl region of a neutral bicelle at 298K. B) Diffusion data evaluation: intensity decays as a function of increasing gradient strengths for regions *left:* alkyl-CH_3 (DHPC) and *right:* alkyl-CH_3 (DMPC).

The obtained translational diffusion coefficient was used to determine the apparent hydrodynamic radius r_H according to the modified Stokes-Einstein equation (Equation 17):

$$r_H = \frac{k_B \cdot T}{6 \cdot \pi \cdot \eta \cdot F \cdot D} \qquad (17)$$

where we augment Equation 2 by the form factor F. For spherical molecules F=1.00, and for bicelles possessing oblate shapes its value can be calculated from the Perrin formula [183]:

$$F_P = \frac{\sqrt{p^2-1}}{\sqrt[3]{p^2}+\tan^{-1}\sqrt{p^2-1}} \qquad (18)$$

where p is the ratio of the longer axis to the shorter axis. In the literature different values were reported for bicelles [115,122], but as a first approximation we used the F=1.00 assumption.

To validate the NMR-SAXS combined method on membrane mimetic systems and to obtain reproducible, biologically relevant data, several parameters were optimized. For this purpose, we first investigated the DHPC micelles that are known to have an elongated structure [96]. Temperature influences the formation of micelles and bicelles according to the corresponding phase diagram [184]. We tested the temperature stability for a q=0 150mM DHPC solution (only the short-chained lipid component was present) under physiological conditions in 10mM HEPES, 150mM NaCl media. In the

288-313K temperature range minor variations in the calculated r_H values were observed, meaning that the micelles retain their structure under these circumstances.

We found that the nature of the buffer altered the r_H values, in the presence of phosphate buffer (pH=5.5) additional small peaks appeared due to DHPC degradation. On the other hand, in the 10 mM HEPES, 150 mM NaCl environment DHPC micelles appeared to be stable.

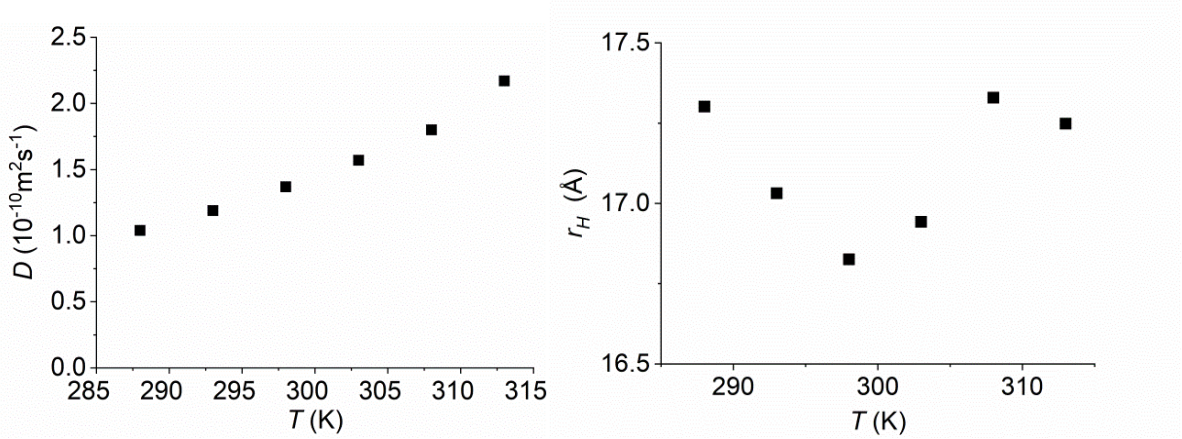


Figure 32. Left: Diffusion coefficients measured for 150mM DHPC solution, right: the calculated hydrodynamic radii as a function of temperature.

Under the optimal experimental conditions, micelles are present in the solution and at 298K the calculated hydrodynamic radius was found to be 16.8±0.1 Å. The SAXS measurement of the same sample provided an 18.1±0.5 Å radius of gyration. The resulting shape factor is 18.1/16.8=1.08, characteristic for objects with an elongated shape. The temperature dependence does not affect r_G as the same value was determined at 310K; while the hydrodynamic radius changed to 17.3±0.2 Å. The resulting shape factor is 1.09, identical to the value obtained at 298 K. This is in good agreement with previous findings [96] based on SANS measurements, where the authors suggested that the micellar structure of DHPC can be well represented by a prolate ellipsoid with two uniform regions. In conclusion, we successfully validated our combined NMR-SAXS method for membrane mimetics.

On the other hand, PFG-NMR measurements might be helpful in obtaining micellar aggregation numbers. Our previous work [185] conducted on folded (valid for approximately spherical molecules) and disordered (elongated, highly mobile molecules) proteins at 287 K resulted in establishment of empirical *D(M)* variations.

This means for an experimentally determined D=(1.04±0.01)·10^{-10} m^2/s for micelle M can be calculated. In this case M=8138 g/mol is obtained if a spherical and M=3886 g/mol if an unfolded molecule is assumed. As the formed DHPC micelles can present shapes that are in between these limits, aggregation numbers in the 9-18 region are expected. These values are below the previously determined data, but if one considers the 25,35,40 [97] values then the calculated molecular weights will no longer be in agreement with the hydration radius data, that were determined by different methods presenting the same outcome. Based on the correlation between the diffusion coefficients, viscosities and temperatures we calculated the M and N values for micelles with spherical and unfolded molecule assumptions at different temperatures. M and N values obtained with the spherical assumption showed small decrease in the 283-303 K temperature range (from 8864 to 7297 g/mol and from 20 to 16 molecules) above this temperature no change was observed in the aggregation numbers (17). Molecular weights with the unfolded molecule assumption showed a similar tendency with the increasing temperature, while there were no variations in the aggregation numbers (8) except at 283 K N=9.

The presence of micelles in bicelle solutions is an important issue that should be taken into consideration. Micelles are formed in the solution above their cmc value. For DHPC this value is 14.0mM in aqueous solution at 298 K [117], but it can be influenced by several factors, such as temperature, ionic strength, type of buffer. Van Dam *et al.* investigated the temperature dependence via fluorescent labelling and detected small variation of cmc between 14-15 mM [184]. We used ^{1}H NMR diffusion measurements and tested both the effect of temperature and ionic strength on *cmc* values. The determination of cmc values is based on evaluating the translational diffusion coefficients at increasing DHPC concentrations. For this purpose, integrated intensities of each region (multiplets) from the ^{1}H spectrum can be used. No shift in peak positions was observed with increasing concentration with the exception of the methyl group region (Figure 33).

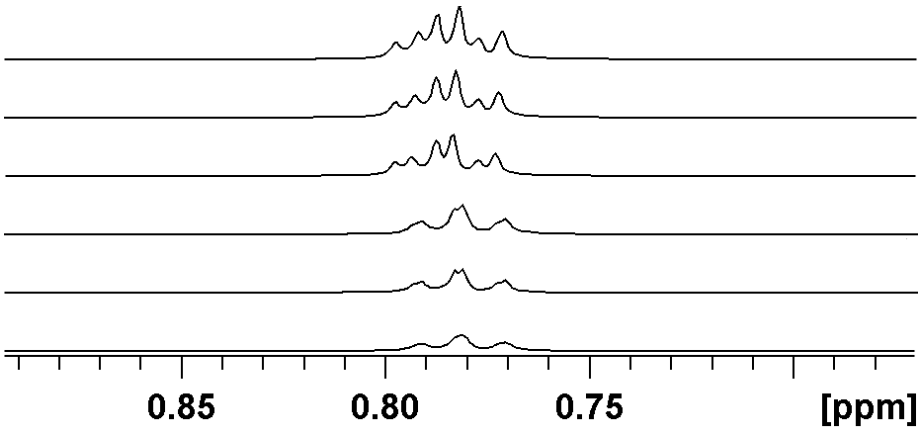


Figure 33. 1D ^{1}H spectra of DHPC molecules below and above the critical micelle concentration. (From bottom to top: 9mM, 12mM, 15mM, 21mM, 24mM, 27mM).

In monomeric form the terminal methyl groups of the two alkyl chains are in a similar chemical environment, therefore, a corresponding triplet (coupling to the neighboring protons at 3-bond distance) with $^{3}J_{HH}$= 6.6 Hz is observed. Once micelles are formed, the two chains are no longer in chemically equivalent environments, and two triplets with t1 $^{3}J_{HH}$= 7.2 Hz and t2 $^{3}J_{HH}$= 7.1 Hz coupling are obtained (Figure 34).

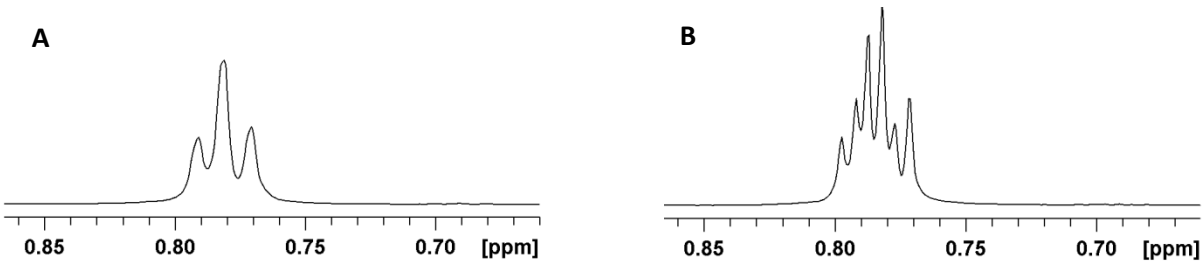


Figure 34. ^{1}H NMR spectra of the alkyl chain methyl groups for DHPC solutions at concentrations 9mM and 27mM

Moreover, while only DHPC monomers are present in the solution the diffusion coefficient has no concentration dependence; its value only starts to decrease once micelles are formed. Linear fitting of these two regimes on the *D* vs. concentration diagram provides two lines with intersection at cmc. We measured *cmc* values at two temperatures working in 10 mM HEPES media with altered NaCl amounts (Table 6). Results show that temperature has a more pronounced effect; cmc values decrease at higher temperatures regardless of the ionic strengths of the solutions. Changes in NaCl concentration show a smaller effect at 298 K, the observed decrease of *cmc* with increasing ionic strength is more pronounced at 310 K.

	cmc (mM)	
	T (K)	
c_{NaCl}(mM)	298	310
0	15.9	12.9
50	15.0	11.6
150	14.9	9.1

Table 6. Experimentally determined *cmc* values under various conditions.

For our investigation, we intended to prepare small isotropic bicelle solutions with a q=0.5 ratio. We applied ^{31}P NMR spectroscopy as a straightforward approach for assessing the obtained bicelle composition. Even though both lipids bear a phosphatidylcholine group, in the 1D ^{31}P spectrum, distinct signals can be detected, and each can be assigned to one of the two molecules. Their integrated intensities reflect the ratio of the DMPC:DHPC concentrations. In our case the ratio was found to be q=0.68 which is reasonably close to the desired value (Figure 35).

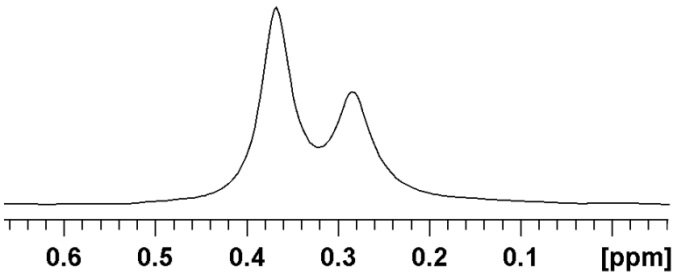


Figure 35. ^{31}P NMR spectra of neutral bicelles with relative integral values 1.00 (δ_{DHPC}=0.40ppm) and 0.68 (δ_{DMPC}=0.26ppm)

The bilayer-forming DMPC lipid is assumed to be exclusively incorporated in bicelles, while DHPC is in excess and can be dynamically distributed between bicelles and micelles, or exists as free monomer. Accordingly, we characterized sample composition, i.e. the non-bicelle bound DHPC concentration, via ^{1}H NMR measurements. A typical ^{1}H 1D spectrum of a DHPC/DMPC bicelle system is shown of Figure 36C.

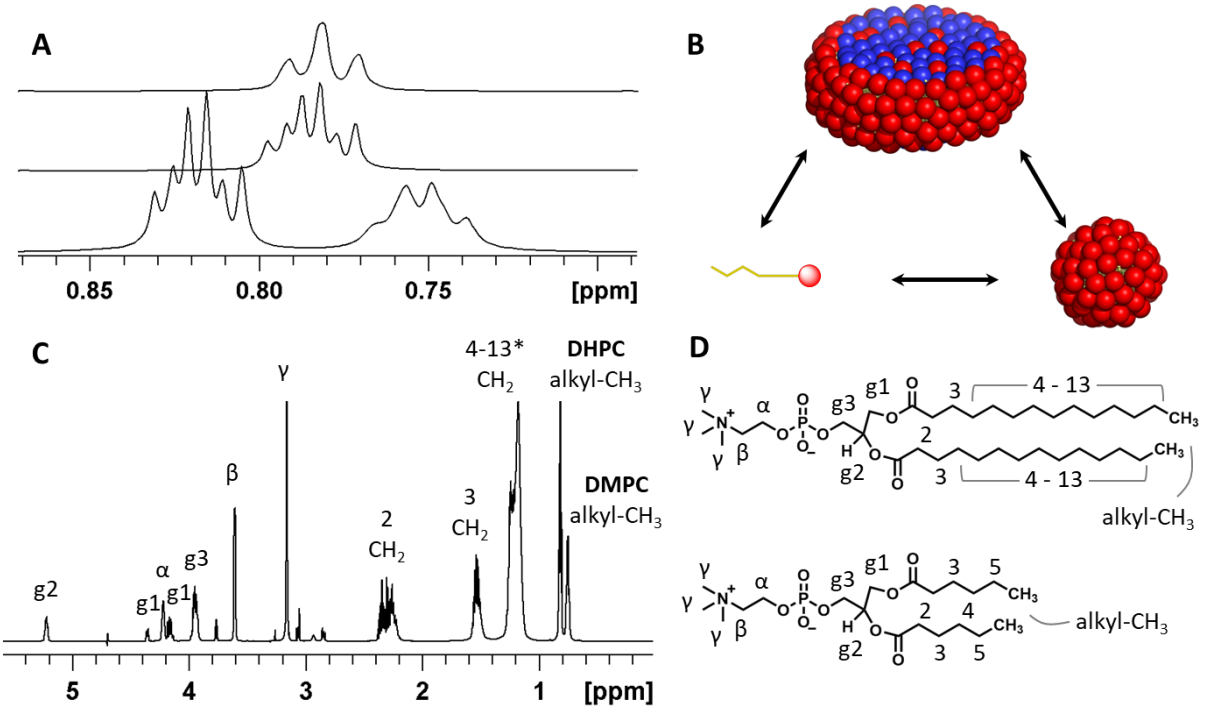


Figure 36. A) Methyl region of DHPC monomer (top, c_{DHPC}=9mM), DHPC micelle (middle, c_{DHPC} =27mM) and DHPC:DMPC bicelle (bottom, c_{DHPC}=100mM) ^{1}H 1D spectra. B) Possible distribution of the DHPC molecules. C) Assignment of the ^{1}H NMR spectrum for a PC solution with q = 0.5 and 150 mM total lipid concentration. The asterisk (*) indicates overlapping proton environments of CH_2 groups from both DMPC (4-13) and DHPC (4-5) molecules. D) Chemical structures of DMPC and DHPC.

As DHPC and DMPC molecules differ only in the length of the alkyl chain the ^{1}H spectra of the bicelle is comprised of highly overlapping peaks. Assignment of the different chemical environments was done utilizing homo- and heteronuclear 2D (COSY, TOCSY, HSQC, HMBC) measurements based on earlier findings [185]. The spectral resolution provided by the applied 700MHz field spectrometer allowed the separation of methyl protons (alkyl-CH_3) in DHPC and DMPC molecules, and a multiplet pattern around 0.8 ppm is detected (Figure 36 A and C). The intensities of the multiplets reflect a 1:2 ratio in accordance with DMPC:DHPC concentrations (q=0.5). A slight downfield shift is observed for the DHPC alkyl-CH_3 multiplet compared to the position of the corresponding multiplet in DHPC micelles, an indicator of decreased shielding. The coupling constant values are t1 $^3J_{HH}$= 7.1 Hz and t2 $^3J_{HH}$= 7.1 Hz similar to those observed for micelles. As no other methyl peaks are detected, all DHPC molecules in the solution (as monomer, in micelle, in bicelle) appear to be involved in a fast exchange on the NMR time-scale. These distinct methyl proton resonances are advantageous in the

interpretation of diffusion measurements, as for each multiplet a different *D* value can be obtained (see Table 7). DMPC is considered to exist exclusively in bicelles, thus DCH3-alkyl(DMPC) represents the diffusion coefficient of the bicelles. Diffusion coefficients calculated for other spectral regions ($2CH_2$ and $3CH_2$ region) differ as they are linear combinations of $D_{\text{alkyl-CH3}}$ (DHPC) and $D_{\text{alkyl-CH3}}$ (DMPC) weighted by the number of protons associated with the corresponding molecule.

^{1}H	2 CH_2	3 CH_2	alkyl-CH_3 (DHPC)	alkyl-CH_3 (DMPC)
region (ppm)	2.556-2.038	1.704-1.399	0.879-0.785	0.785-0.673
$D(10^{-11}m^2s^{-1})$	8.13	8.23	9.68	5.47

Table 7. Characteristic ^{1}H spectra regions with measured diffusion coefficient values.

The concentration of free DHPC molecules can be calculated by:

$$c_{free}^{DHPC} = \frac{D_{\text{alkyl-CH3}}\ (\text{DHPC}) - D_{\text{alkyl-CH3}}\ (\text{DMPC})}{D_{monomer}^{DHPC} - D_{\text{alkyl-CH3}}\ (\text{DMPC})} c_{total}^{DHPC}$$

where c_{free}^{DHPC} refers to concentration of DHPC molecules not bound to bicelle, $D_{\text{monomer}}^{\text{DHPC}}$ represents the diffusion coefficient of the monomeric DHPC determined by separate measurements under the same experimental conditions. c_{total}^{DHPC} represents the total amount of DHPC. Values for empty bicelles and for the different bicelle-peptide systems were summarized in Table 8. In all cases the concentration of free DHPC molecules is below the determined *cmc* values, thus no micelle formation is expected and our investigated solutions contained only bicelles and DHPC monomers. However, the presence of model peptides alters the concentration of free DHPC in the solution, the more peptide introduced the higher is the free DHPC concentration. We assume that incorporation or interaction with the peptide results in liberation of short chain lipid molecules. The highest influence was detected for 2 mM KALP23. Table 8 shows these values calculated for empty PC (DHPC/DMPC) and PC/PG (DHPC/DMPC/DMPG) bicelles and for the different bicelle-peptide systems at two temperatures. At 298 K in all cases the concentration of the free DHPC molecules is below the determined *cmc* values, and no micelles are formed. The presence of peptide has an increasing effect on the free DHPC concentration, the highest influence is observed for PC bicelles with 2 mM

KALP23. This phenomenon indicates that upon interaction with the peptide – regardless of its topology – some amount of short chain lipid is released. This observation is valid also for measurements conducted at 310 K, however for each bicelle type almost in all cases the free DHPC concentration exceeds the *cmc* value; meaning an amount of micelle is present in the solution. Still, this value is low (< 0.2 mM) taking an average N aggregation number for the calculation.

Bicelle concentration can be obtained by using the following equation:

$$c_{\text{bicelle}} = \frac{c_{\text{DMPC}}[\text{mol}/dm^3]}{N(\text{DMPC})}\ \text{mol/dm}^3$$

where c_{DMPC} is the DMPC concentration in the solution and N(DMPC) is the DMPC aggregation number (the number of DMPC molecules in a single bicelle). However, the aggregation number for small isotropic bicelles is debated, and several approaches were suggested in the literature for its evaluation:

Vold and Prosser [186] proposed the 'ideal bicelle model' where DHPC and DMPC are perfectly segregated even above T_{m} (main transition temperature of the long-chained lipid)and they obtained a formula by considering the ratio between the surfaces occupied by bilayers and edges. Based on this approach we calculated the surface areas and divided them by the headgroup areas of single lipid molecules. However, the surface occupied by DHPC molecules in edges is not well known, it has been estimated as 102 A^2 for DHPC micelles [96], 65 Å^2 for monolayers [187] and 60 Å^2 for DMPC molecules in multibilayers [188]. The corresponding aggregation numbers were N(DMPC)=47 and N(DHPC)=60 with a(DHPC)=100 Å^2 and N(DHPC)=94 if a(DHPC)=65 Å^2. In the latter case the q~0.5 condition is fulfilled.

Triba *et al.* [189] proposed the 'mixed bicelle model' assuming that the two lipids are mixed in each domain of the disk. They recommended calculating lipid aggregation numbers by dividing the bilayer and edge volumes with the volumes of DMPC and DHPC molecules considering ε and ε'parameters: the fraction of DHPC molecules inside the DMPC-rich section (bilayer) and the fraction of DMPC molecules inside the DHPC-rich section (rim). (The latter can be neglected while the former has a value of 0.02-0.03 at T=298 K.) This approach resulted in N(DMPC, bilayer)=68, N(DHPC, edge)=113, N(DHPC, bilayer)=2 and c(bicelle)=0.74 mM, q=0.6.

Situ et al. [190] calculated the lipid aggregation number as a function of the effective *q* factor based on the mixed bicelle model using literature parameters [189]. According to their calculations, if *q*~0.5 then *N*(DHPC)=75 and *N*(DMPC)=25 at *c*(bicelle)=2 mM. However, *N*(DMPC)/*N*(DHPC)=0.3, thus, the long-chained lipid and short-chained lipid ratio is not in accordance with the original *q* value.

Lee et al. [191] suggested the hemistroidal model of protein-loaded bicelles. Based on this model they initially obtained *N*(DHPC)=280 and *N*(DMPC)=140. For our systems this approach provided *N*(DHPC)=300 and *N*(DMPC)=164 and *c*(bicelle)=0.3 mM (with KALP23 peptide incorporated in the bicelle). It should be noted that due to inconsistencies between the measured overall rotational correlation time and the calculated molecular weight they modified their model leading to *N*(DHPC)=92 and *N*(DMPC)=46.

As literature data is contradictory, we tried to evaluate reliable aggregation numbers from molecular weight values derived from diffusion data. Translational diffusion coefficients measured for PC bicelles (Table 8) are in agreement with reported literature values, they can be considered as starting points for molecular weight estimation - based on our proposed empirical (log*D*-log*M*) formulae Equation 13, Equation 14 performing viscosity and temperature corrections. The calculated values are ~ 81 kDa for folded, spherical molecules (Eq. 13) and 22 kDa for disordered, elongated (Eq. 14) systems. The shape of the bicelles is expected to be in between these categories and the *q*=0.5 condition has to be fulfilled, therefore, *N*(DHPC)=92 and *N*(DMPC)=46 with *M*=73 kDa suggested by Lee *et al* [191], or the *N*(DHPC)=102 and *N*(DMPC)=47 with slightly decreased *q*=0.47 representing *M*=78 kDa given by Vold *et al* [186]can be considered as realistic aggregation numbers. Based on these values bicelle concentration in our solutions was cca 1mM. In several cases there are discrepancies between the calculated and the experimental values, thus the authors suggest modifications in their models and it is still debated which bicelle model is generally applicable.

Changes in shape and morphology upon addition of model peptides to PC and PC/PG bicelles was tested by PFG-NMR and SAXS measurements at two temperatures (298 K, 310 K). The results are summarized in Table 8. The measured bicelle diffusion coefficients agree well with literature values ($5.3\pm0.1\cdot10^{-11}$ by Björneras *et al*. [185] with

150 mM PC concentration at 298 K). Bicelles have larger hydrodynamic radius at 310 K, the difference is ~4 Å in accordance with previous findings by Lind *et al*. [121]. Their explanation was that there might be considerable mixing of DMPC and DHPC in small isotropic bicelles, which potentially leads to a larger amount of molecules constructing the bilayered part of the aggregate, thereby making them larger. Introduction of peptides – especially the transmembrane KALP23 – results in the increase of r_H [108]. r_G values were determined from the Guinier region of the SAXS scattering curves. Even though these values carry high uncertainties, their trend is similar to that of r_H.

Bicelle type	Peptide	T (K)	c^{DHPC}_{free} (mM)	D (10^{-11} m^2 s^{-1})	r_H (Å)	r_G (Å) Guinier
PC	-	298	11.1	5.46 ± 0.02	42.2 ± 0.2	34.3 ± 4.1
	KALP23 (1 mM)		11.3	5.03 ± 0.03	45.9 ± 0.2	42.7 ± 4.6
	KALP23 (2 mM)		14.8	4.44 ± 0.03	51.9 ± 0.2	-
	melittin (1 mM)		11.8	5.52 ± 0.01	41.8 ± 0.1	33.3 ± 3.3
	melittin (2 mM)		14.3	5.61 ±0.02	41.1 ±0.2	-
	-	310	11.3*	7.26 ± 0.02	46.1 ± 0.1	33.2 ± 2.7
	KALP23 (1 mM)		11.8*	6.08 ± 0.07	55.1 ± 0.6	42.3 ± 10.1
	KALP23 (2 mM)		14.6*	5.26 ± 0.02	63.7 ± 0.3	-
	melittin (1 mM)		12.0*	6.88 ± 0.01	48.7 ± 0.1	37.5 ± 6.2
	melittin (2 mM)		14.3*	6.74 ± 0.01	49.7 ± 0.1	-
PC/PG	-	298	10.0	5.35 ± 0.03	43.1 ± 0.3	33.5 ± 3.8
	KALP23 (1 mM)		10.7	4.74 ± 0.02	48.6 ± 0.1	37.9 ± 6.4
	melittin (1 mM)		11.1	4.95 ± 0.01	46.6 ± 0.1	35.8 ± 3.7
	-	310	8.9	8.16 ± 0.40	41.1 ± 2.7	36.3 ± 6.9
	KALP23 (1 mM)		10.9*	6.13 ± 0.08	54.6 ± 0.1	40.7 ± 5.5
	melittin (1 mM)		11.4*	5.91 ± 0.05	56.7 ± 0.2	36.7 ± 3.6

Table 8. Bicelle type, interacting model peptides, temperature, the free DHPC concentration (* values are > *cmc*) measured diffusion coefficients, hydrodynamic radii, radii of gyration calculated from the Guinier approximation. Missing values indicate the failure of the Guinier fit due to bicelle-bicelle interactions caused by high sample concentration.

For this reason, we attempted to determine r_G from the entire scattering curve using the two-shell ellipsoid model (Figure 12). This model features an ellipsoidal core with two semi-axes a,b that represent the hydrophobic interior of the bicelle, while the outer shell - with thickness t_a and t_b - mainly consists of the lipid headgroups. The parameters a,b,t_a,t_b are obtained by fitting the scattering intensities. Addition of the

model peptides influence the morphology of the bicelle which is reflected in these parameters (Table 9).

bicelle type	peptide	T (K)	a (Å)	b (Å)	t_a (Å)	t_b (Å)	r_G (Å)fitted
PC	-	298	12.0 ± 0.1	36.7 ± 0.3	14.3 ± 0.2	3.0 ± 0.1	40.0 ± 0.1
	KALP23 (1 mM)		11.9 ± 0.1	43.7 ± 0.7	13.4 ± 0.2	1.6 ± 0.2	40.0 ± 0.1
	KALP23 (2 mM)		11.9 ± 0.2	48.4 ± 2.0	12.5 ± 0.4	1.0 ± 0.5	40.0 ± 0.2
	melittin (1 mM)		11.9 ± 0.1	38.4 ± 0.4	14.4 ± 0.2	2.2 ± 0.1	38.8 ± 0.1
	melittin (2 mM)		11.6 ± 0.2	39.2 ± 0.7	14.7 ± 0.4	1.7 ± 0.2	37.3 ± 0.1
	-	310	12.3 ± 0.2	35.6 ± 0.6	13.1 ± 0.3	2.8 ± 0.2	40.1 ± 0.2
	KALP23 (1 mM)		-	-	-	-	-
	KALP23 (2 mM)		11.9 ± 0.2	43.4 ± 1.5	11.5 ± 0.4	2.1 ± 0.4	39.5 ± 0.3
	melittin (1 mM)		12.3 ± 0.2	37.7 ± 0.8	12.8 ± 0.4	2.1 ± 0.3	38.8 ± 0.2
	melittin (2 mM)		12.4 ± 0.2	39.9 ± 0.8	12.3 ± 0.3	1.3 ± 0.2	37.6 ± 0.2
PC/PG	-	298	11.9 ± 0.1	36.1 ± 0.4	14.6 ± 0.3	2.9 ± 0.2	39.2 ± 0.1
	KALP23 (1 mM)		12.2 ± 0.2	45.7 ± 1.0	12.5 ± 0.3	1.0 ± 0.2	39.7 ± 0.2
	melittin (1 mM)		12.5 ± 0.1	39.0 ± 0.6	12.8 ± 0.3	1.8 ± 0.2	38.0 ± 0.1
	-	310	12.5 ± 0.1	35.9 ± 0.5	12.7 ± 0.3	2.2 ± 0.2	39.2 ± 0.1
	KALP23 (1 mM)		12.5 ± 0.2	44.1 ± 1.3	11.4 ± 0.3	1.3 ± 0.3	40.3 ± 0.2
	melittin (1 mM)		12.8 ± 0.1	36.2 ± 0.3	11.5 ± 0.1	1.6 ± 0.1	37.6 ± 0.1

Table 9. Bicelle composition, interacting peptides, temperature, semi-axes of the fitted ellipsoid (a, b), thickness of the lipid head groups (t_a, t_b), r_G obtained from fitting the lenticular core-shell model

Uncertanties of the fitted ellipsoid parameters are below 4% for a,b,t_a and ~10% for t_b. As for all studied systems (a+t_a) is smaller than (b+t_b) we can conclude that bicelles possess oblate shapes. The maximum length of the fatty acid chain of DMPC is 17.9 Å and the length of the headgroup is 9 Å, therefore, the thickness of the bilayer is expected to be 2(17.9+9)=53.8 Å. In our model 2(a+t_a), which corresponds to the bilayer thickness, falls between 50-52 Å. This is not only in agreement with the values given by both Glover [111], but also supports the validity of the lentil model.

Further morphological changes can be analyzed based on Table 8 and 9. Upon the addition of model peptides (regardless whether it is transmembrane, or surface active) b increases and t_b decreases, indicating that the rim becomes narrower while the shape becomes more elongated. This alludes to the migration of DHPC molecules from the rim

to the bilayer region. This hypothesis is also supported by a slight increase in free DHPC concentration values. The increase in *b* is reflected in the increasing hydrodynamic radii.

Stepwise addition of the transmembrane peptide KALP23 resulted in increase of r_H for both PC and PC/PG bicelles, at 298 K with 10 Å (PC), and 5 Å (PC/PG), and with 18 Å (PC) and 13 Å (PC/PG) at 310 K, respectively. Addition of 1 mM and 2 mM KALP23 resulted in bicelle size increase by 3.7 Å and 6 Å at 298 K, and 9 Å and 8.6 Å at 310 K, while the *b* axis is changed by 14 and 9 Å at 298 K and by 16 Å at 310 K. These changes are accompanied by a slight decrease in t_b, i.e the thickness of the outer shell. r_G values obtained from the Guinier fitting follow the behavior of r_H, while fitted r_G values stay practically unaltered within error margins.

The size of empty, zwitterionic bicelles – regarding their determined r_G values - shows no change at the two studied temperatures. Introduction of 1 mM KALP peptide had no effect on bicelle size - according to the Guinier region fitting - while increase in the amount of melittin caused a decrease of 1 Å at 1 mM and 3 Å at 2 mM peptide concentration. For charged bicelles the addition of KALP resulted in a small increase in r_G (~1 Å) at both temperatures, while melittin caused a 1-1.5 Å decrease. Attention has to be paid to the PC/PG bicelle data interpretation as the net charge of the peptide under the experimental conditions (pH) can cause major variations due to electrostatic interactions. Therefore, directly comparing the results originating from different peptides is not recommended.

The parameters gained from SAXS data evaluation – *a*, *b*, t_a, t_b - allowed us to calculate Perrin shape factors for each investigated system based on Equation 18. According to Liebau *et al.* [115], Perrin shape factors between 1.04 and 1.22 are typical values for bicelles. Chou *et al.* [117] argued that, for q = 0.5 bicelles, aspect ratios $((b+t_b)/(a+t_a))$ are below 2, lead to Perrin shape factors close to 1. Biverstahl *et al.* [122] used a substantially higher shape factor of 1.22. For our investigations all calculated values are in the range of 1.01-1.04, in accordance with the results of Liebau and Chou.

The ratio of the global parameters is widely applied for estimating molecular shape. In case of hard spheres the characteristic r_G/r_H value is 0.77 (a complete derivation of this value is given in the Appendix, Figure A5). However, while the r_G parameter is obtained from measurements that do not detect the hydration layer, PFG

NMR measurements determine the effective r_H, that also incorporates the hydration layer. Independent measurements report on 10 Å water layer moving with the bicelle [193,194] Accordingly, when we calculated the r_G/r_H ratios this water layer thickness was subtracted from the experimentally determined r_H values.

bicelle composition	peptide	*T* (K)	r_G(Guinier)/r_H	r_G(fitted)/r_H
DHPC DMPC	-	298	1.07±0.13	1.24±0.01
	KALP23 (1 mM)		1.19±0.13	1.11±0.01
	KALP23 (2 mM)		-	0.95±0.01
	melittin (1 mM)		1.05±0.10	1.22±0.12
	melittin (2 mM)		-	1.20±0.01
	-	310	0.92±0.07	1.11±0.01
	KALP23 (1 mM)		0.94±0.22	-
	KALP23 (2 mM)		-	0.74±0.01
	melittin (1 mM)		0.97±0.16	1.00±0.01
	melittin (2 mM)		-	0.95±0.01
DHPC DMPC DMPG	-	298	1.01±0.11	1.18±0.01
	KALP23 (1 mM)		0.98±0.17	1.03±0.01
	melittin (1 mM)		0.98±0.10	1.04±0.00
	-	310	1.17±0.24	1.26±0.08
	KALP23 (1 mM)		0.91±0.12	0.90±0.00
	melittin (1 mM)		0.87±0.09	0.81±0.00

Table 10. Calculated shape factors.

r_G values were determined with significant uncertainties from the Guinier region of the SAXS curves and this error propagates in the shape factors(Table10). Regardless, all values suggest elongated shapes for bicelles. Addition of transmembrane peptide to PC or PC/PG bicelle solutions results in small decrease of the shape factor, while for the surface active peptide similar tendency is observable but the effect is negligible. This conclusion might be misleading, as r_G values from the fitted model reflect no variation upon peptide addition. However, this observation is contradicted both by increase in r_H from independent NMR measurements and also by a significant increase in parameter *b*. Based on these findings, even though it would be rather elegant to describe shape variation by a single variable – the shape factor -, this approach is unsuccessful. Therefore, we support conclusions drawn from the r_H and (b+t_b) parameters.

In conclusion, we successfully combined two techniques – NMR spectroscopy and SAXS – for studying the size and shape of bicelle systems. NMR measurements enabled a throughout characterization of DHPC micelles complemented by the determination of corresponding *cmc* values under various experimental conditions, DHPC distribution in bicelle solutions and diffusion coefficients and hydrodynamic radii of micelles and bicelles. Fitting of the core-shell lentil model to the complete scattering curve provided r_G values and valuable parameters reflecting bicelle shape. We showed that the shape factor - the r_G/r_H ratio – might lead to misleading conclusions, thus, we suggested the appliaction of r_H and (b+t_b) value for global characterization. The presence of the model peptides affected several global parameters of the bicelle. The transmembrane KALP23 induced size increase and elongation, while the surface-active melittin did not perturb neutral bicelles. We showed that careful data interpretation is needed for the negatively charged bicelles. It is now advisable to directly compare the results originating from the different peptides; comparisons should be done with respect to the 'empty' bicelle system. Our study demonstrated the usefulness of the synergy of NMR and SAXS for the characterization of bicelle size and morphology.

Summary and Conclusions

Protein NMR spectroscopy is a valuable tool not only in studying structure, structural propensities and local environments in proteins but also in the determination of global parameters describing molecular dimensions, folding, shape and morphology. During my PhD years I characterized the structural tendencies of several IDPs utilizing solution-state NMR spectroscopy and I developed a method based on PFG-NMR experiments to study the size and shape of various biomolecules.

I examined the structure of several linear motifs through which MAPKAPKs form specific complexes with their activating MAPKs. Secondary Chemical Shift (SCS) data of MK2, RSK1 and RSK1_S/A were consistent of preference for an α-helical propensity in short regions while NFAT and MNK1 were found to be unstructured. Under denaturing conditions SCS values decreased for each peptide indicating reduced residual structural tendencies which do not influence the highly mobile ensemble. The structural propensities of the examined linear motifs contribute to understand their binding specificity (nascent helicity - specific binding, structural disorder - promiscuity) and thus, the build-up of routes for signal flow in the MAPK system

Based on my experiences on the MAPKAPK peptide fragments I continued my PhD research with a more complex system: Upon p53 TAD interaction with the metastasis associated Ca^{2+}-loaded wt S100A4 I detected disorder-to-order transition – similar to other p53TAD complexes. Due to the high flexibility of p53TAD I found that structural characterization was possible only via an NMR-MD approach supported by dynamics analysis. Results obtained by the synergy of these methods revealed that disorder-to-order transition affects the TAD1 and TAD2 domains and three α-helices are formed in these regions (S20-E28, P36-P47, E51-T55) connected by relatively dynamic loop regions (L14-F19, N29-L35, D48-I50) but most of the p53TAD segments remain unstructured. This so-called clamp model provides flexibility and adaptability in molecular interactions and structural similarities of p53TAD in different complexes suggest the importance of conserved local structure in its recognition.

In the characterization of proteins, the degree of folding and the molecular dimensions are key questions. I intended to quantify these parameters by measuring the self-diffusion coefficients (*D)* via diffusion NMR experiments. I have carried out the measurements under the same experimental conditions with a set of relevant proteins and I was able to establish empirical log*D*-log*M* relations which are useful and reliable bioanalytical tools for aggregation and molecular mass analysis. I performed experiments under denaturing conditions – in 8M urea solution – which revealed that IDPs cannot be further denatured (decrease in local structural tendencies is not reflected in their global hydrodynamic parameters) while for folded proteins they indicated whether the final unfolded state had been achieved, or if not then what was the extent of denaturation.

In a research collaboration we combined the PFG-NMR technique with Small-Angle X-Ray Scattering and determined both r_H and r_G values on the same sample under the same conditions. We utilized these values to follow changes in global parameters of bicelle-peptide model systems: changes in size were reflected in values r_H, while the morphology of the bicelles was characterized by shape factors r_G/r_H and parameters a, b, t_a and t_b from fitting a lenticular core-shell model on the entire SAXS scattering curve. We showed that the shape factor approach failed for bicelles and bicelle-peptide systems, however, the r_H and (b+t_b) parameters were applicable for global characterization. Our results indicated that the transmembrane model peptide induced elongation and size increase; the surface-active peptide did not perturb neutral bicelles. Thus, our study demonstrated the versatility of the NMR-SAXS approach for investigating bicelle size and morphology.

In summary, I proved that solution-state NMR methods are versatile and can be utilized for both local and global characterization of proteins and protein-bicelle systems.

Összefoglalás

A fehérje NMR értékes módszer nemcsak fehérjék szerkezetének, szerkezeti hajlamainak és lokális szerkezeti jellemzőinek tanulmányozására, hanem a globális paramétereik meghatározására is, melyek összefüggésben vannak a méretükkel, feltekeredettségükkel, alakjukkal és morfológiájukkal. PhD munkám során több rendezetlen fehérje (IDP) szerkezeti hajlamát jellemeztem, illetve kifejlesztettem egy PFG-NMR alapú módszert biomolekulák méretének és alakjának vizsgálatára.

Vizsgáltam olyan lineáris motívumok szerkezeti tendenciáit, melyeken keresztül a MAPKAPKinázok specifikus komplexet képeznek aktiváló MAPKinázaikkal. Az MK2, RSK1 és RSK1_S/A esetében kapott SCS (Másodlagos Kémiai Eltolódás) adatok α-helikális preferenciára utaltak rövid régiókban, míg az NFAT és az MNK1 teljes hosszában rendezetlennek tűnt. Denaturáló közegben az SCS értékek minden peptidfragmens esetében csökkentek, ami a másodlagos szerkezeti hajlam csökkenését jelzi, de ez nem befolyásolja a mobilis szerkezetsokaságot és annak globális hidrodinamikai paramétereit. A vizsgált lineáris motívumok szerkezeti sajátságai hozzájárulnak specifikus kötődési tulajdonságaik megértéséhez (naszcens helicitás – specifikus kötődés, teljes rendezetlenség – promiszkuus kötődés) és ezáltal a MAPK rendszer jelátadási útvonalainak jobb megismeréséhez.

A MAPKAPK peptidfragmensek tanulmányozása során szerzett ismereteimet egy bonyolultabb rendszer vizsgálata során alkalmaztam: A p53TAD régió és a metasztázissal összefüggésben lévő Ca^{2+}-kötött S100A4 kölcsönhatása során rendezetlen-rendezett átmenet lejátszódását tapasztaltam, hasonlóan más p53TAD komplexekhez. A p53TAD régió nagymértékű flexibilitása miatt a szerkezet jellemzéséhez az NMR spektroszkópia és az MD szimulációk ötvözésére volt szükség, dinamikai analízisből származó eredményekkel alátámasztva. A módszerek szinergiájából származó eredmények azt mutatták, hogy a rendezetlen-rendezett átmenet a TAD1 és TAD2 doméneket is érinti, és három α-hélix kialakulásával jár e régiókban (S20-E28, P36-P47, E51-T55). A hélixeket dinamikus hurokrégiók kötik össze (L14-F19, N29-L35, D48-I50) de a p53TAD nagy része rendezetlen marad. Ez az ún. csipesz(clamp)-modell flexibilitást és alkalmazkodóképességet biztosít a kölcsönhatások során és a p53TAD szerkezetének hasonlósága különböző fehérjékkel alkotott

komplexekben arra enged következtetni, hogy a konzervált lokális szerkezet a p53 felismerésének egy jellegzetessége.

A fehérjék jellemzésénél a feltekeredettség mértéke és a molekulaméret kulcsfontosságú paraméterek, amiket diffúziós együtthatók (D) diffúziós NMR mérések segítségével történő meghatározásával szerettem volna számszerűsíteni. Ugyanolyan kísérleti körülmények között végeztem a méréseket egy reprezentatív fehérjekészleten, és empirikus log$D$-log$M$ összefüggéseket kaptam, amik hasznos és megbízható bioanalitikai eszközök aggregáció és molekulatömeg tanulmányozására. Méréseket végeztem denaturáló körülmények között is – 8 M karbamidoldatban – amik azt mutatták, hogy az IDP-ket nem lehet denaturálni (lokális szerkezeti hajlamaik csökkenhetnek, de ez nem befolyásolja a globális hidrodinamikai paramétereiket), míg rendezett fehérjéknél ezek a mérések jelzik, hogy elértük-e a végső rendezetlen állapotot, vagy ha nem, milyen mértékű a denaturáció.

Egy kutatási együttműködés keretében ötvöztük a PFG-NMR módszert a Kisszögű röntgenszórással (SAXS), és ugyanazon a mintán, ugyanolyan körülmények között meghatároztuk az r_H (hidrodinamikai sugár) és r_G (girációs sugár) értékeket. Ezek alapján szerettük volna követni a globális paraméterek változását bicella-peptid modellrendszerek esetében: a méretbeli változásokat az r_H jelezte, míg a morfológia megváltozására az alakfaktorokból r_G/r_H és a, b, t_a, t_b paraméterekből következtettünk. Utóbbi paramétereket úgy kaptuk, hogy egy lencse-modellt illesztettünk a teljes szórási görbére. Bemutattuk, hogy az alakfaktoron alapuló megközelítés nem alkalmazható bicellák és bicella-peptid rendszerek esetében, de az r_H és (b+t_b) paraméterek alkalmazhatóak a globális jellemzésre. Eredményeink alapján a transzmembrán peptid esetében méretnövekedésre és a bicella alakjának megnyúlására következtettünk, míg a felületaktív peptid nem okozott számottevő változást a semleges bicellák esetében. Mindezek alapján tehát azt láthatjuk, hogy az NMR-SAXS megközelítés sikeresnek bizonyult bicellák méretének és morfológiájának vizsgálatára.

Összefoglalásként tehát elmondható, hogy eredményeimmel bizonyítottam, mennyire sokoldalú módszer az oldatfázisú NMR spektroszkópia, hiszen nemcsak lokális, hanem globális jellemzésre is használható fehérjék és fehérje-bicella rendszerek esetében.

Acknowledgements

First of all, I would like to thank my supervisor, Andrea Bodor for her guidance, help and advices throughout my PhD work. I am grateful for the time she invested in showing me the practices of NMR spectroscopy and sample preparation.

I would like to express my thanks to all the members of the Laboratory of Structural Chemistry and Biology and the MTA-ELTE Protein Modelling Research Group, I have learned a lot while working with them.

I am grateful to our research collaborators: Dóra K. Menyhárd for carrying out the MD simulations in the p53-S100A4 project, Gyula Pálfy for the dynamics analysis, András Wacha and Attila Bóta for the SAXS measurements, data analysis and our fruitful discussions.

I am to thank A. Perczel, Gy. Batta, K. Liliom, L. Nyitray, A. Reményi, Á. Tantos, P. Tompa and L. Mäler for the protein samples and I. Bányai and L. Novák for the viscosity measurements in our diffusion studies.

I would like to acknowledge my closest co-worker during this PhD project, Andris, my office-mate for 7 long years. I learned a lot from him about NMR spectroscopy, spectrum processing and assignment. We had the 'most musical' office with the utmost bad jokes in the research group.

I would like to thank my parents for their love and for supporting me and my decisions in every way they could. I am grateful to my sister, Eszti for providing me 'the background' to make it through my PhD years.

I am to thank my friends for always being there for me.

And last, but most importantly, I am grateful to my husband, Tamás for his support, encouragement, patience and love – and I thank him for being him.

References

[1] Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; J. Ausio, Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; E. C. Garner, E. C.; Obradovic, Z. *J. Mol. Graph. Model.* **2001**, 19, 26.

[2] Tompa, P. *Trends Biochem Sci.* **2002**, 10, 527-33.

[3] Uversky, V. N.; Gillespie, J. R.; Fink, A. L. *Proteins* **2000**, 41, 415-427.

[4] Wright, P. E.; Dyson, H. J. *J. Mol Biol.* **1999**, 293, 321-31.

[5] Tompa, P. *Structure and Function of Intrinscially Disordered Proteins*; Taylor & Francis Group, 2010.

[6] Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradovic, Z. *Biochemistry* **2002**, 41, 6573.

[7] Mishra, P. M.; Uversky, V. N.; Giri, R. *J. Mol. Biol.* **2018**, 430, 2372–2388.

[8] Bhasne, K.; Sebastian, S.; Jain, N.; Mukhopadhay, S. *J. Mol. Biol.* **2018**, 430, 2508–2520.

[9] Bignon, C.; Troilo, F.; Gianni, S.; Longhi, S. *J. Mol. Biol.* 430, **2018** 2493–2507.

[10] Metallo, S. *Curr. Opin. Chem. Biol.* **2010**, 14, 481–488.

[11] Vassilev, L. T.; Vu, B.T.; Craves, B.; Carvajal, D.; Podlaski, F.; Filipovic, Z.; Kong, N.; Kammlott, U.; Lukacs, C.; Klein, C.; Fotouhi, N.; Liu, E.A. *Science* **2004**, 303, 844–848.

[12] Vu, B.; Wovkulich, P.; Pizzolato, G.; Lovey, A.; Ding, Q.; Jiang, N. *ACS Med. Chem. Lett.* **2013**, 4, 466–469.

[13] Habchi, J.; Tompa, P.; Longhi, S.; Uversky, V. N. *Chem. Rev.* **2014**, 114, 6561−6588.

[14] Dunker, A. K.; Oldfield, C. J.; Meng, J.; Romero, P.; Yang, J. Y.; Chen, J. W.; Vacic, V.; Obradovic, Z.; Uversky, V. N. *BMC Genomics* **2008**, 9 (Suppl.2), S1.

[15] Prilusky, J.; Felder, C. E.; Zeev-Ben-Mordehai, T.; Rydberg, E. H.; Man, O.; Beckmann, J. S.; Silman, I.; Sussman, J. L. *Bioinformatics* **2005**, 21, 3435.

[16] Coeytaux, K.; Poupon, A. *Bioinformatics* **2005**, 21, 1891.

[17] Linding, R.; Russell, R. B.; Neduva, V.; Gibson., T. J. *J. Nucleic Acids Res.* **2003**, 31, 3701.

[18] Sickmeier, M.; Hamilton, J. A.; LeGall, T.; Vacic, V.; Cortese, M. S.; Tantos, A.; Szabo, B.; Tompa, P.; Chen, J.; Uversky, V. N.; Obradovic, Z.; Dunker, A. K. *Nucleic Acids Res.* **2007**, 35, D786.

[19] Fukuchi, S.; Sakamoto, S.; Nobe, Y.; Murakami, S. D.; Amemiya, T.; Hosoda, K.; Koike, R.; Hiroaki, H.; Ota, M. *Nucleic Acids Res.* **2012**, 40, D507.

[20] Di Domenico, T.; Walsh, I.; Martin, A. J.; Tosatto, S. C. *Bioinformatics* **2012**, 28, 2080.

[21] Ferron, F.; Longhi, S.; Canard, B.; Karlin, D. *Proteins* **2006**, 65, 1.

[22] Dosztanyi, Zs.; Csizmok, V.; Tompa, P.; Simon, I. *Bioinformatics* **2005** 21, 3433-3434.

[23] Ward, J. J.; McGuffin, L. J.; Bryson, K.; Buxton, B. F.; Jones, D. T. *Bioinformatics* **2004**, 20, 2138.

[24] Yang, Z. R.; Thomson, R.; McNeil, P.; Esnouf, R. M. *Bioinformatics* **2005**, 21, 3369.

[25] Ohgushi, M.; Wada, A. *FEBS* **1983**, 164, 21-24.

[26] Dolgikh, D. A.; Gilmanshin, R. I.; Brazhnikov, E. V.; Bychkova, V. E.; Semisotnov, G. V.; Venyaminov, S. Y.; Ptitsyn, O. B. *FEBS Letters* **1981**, 136, 311-315.

[27] van der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R. J.; Daughdrill, G. W.; Dunker, A. K.; Monika Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D. T.; Kim, P. M.; Kriwacki, R. W.; Oldfield C. J.;, Pappu, R. V.; Tompa, P.; N. Uversky, V. N.; Wright, P. E.; Babu, M. M. *Chemical Reviews* **2014**, 114, 6589-6631

[28] Novacek, J.; Zidek, L.; Sklenár, V. *Journal of Magnetic Resonance* **2014**, 241, 41–52.

[29] Jensen, M. R.; Ruigrok, R. W. H.; Blackledge, M. *Current Opinion in Structural Biology* **2013**, 23, 426–435.

[30] Graewert, M. A.; Svergun, D. I. *Current Opinion in Structural Biology* **2013**, 23, 748–754.

[31] Karplus, M.; McCammon J. A. *Nature Structural Biology* **2002**, 9, 9.

[32] Zheng, W.; Best, R. B. *J Mol Biol* **2018**, 430, 2540–2553.

[33] Bernado, P.; Svergun, D. I. *Mol. BioSyst.* **2012**, 8, 151–167.

[34] Henriques, J.; Arleth, L.; Lindorff-Larsen, K.; Skepö, M. *J Mol Biol* **2018**, 430, 2521–2539.

[35] Cragnell, C.; Rieloff, E.; Skepö, M. *J Mol Biol* **2018**, 430, 2478–2492.

[36] Chen, H.; Rhoades, E. *Current Opinion in Structural Biology* **2008**, 18, 516–524.

[37] Puchner, E. M.; Gaub, H. E. *Current Opinion in Structural Biology* **2009**, 19, 605–614.

[38] Deniz, A. A.; Mukhopadhyay, S.; Lemke, E. A. *J. R. Soc. Interface* **2008** 5, 15–45.

[39] Walker, J. M. *The Protein Protocols Handbook;* Humana Press, 2009.

[40] Brusotti, G.; Calleri, E.; Colombo, R.; Massolini, G.; Rinaldi, F.; Temporini, C. *Chromatographia* **2018**, 81, 3–23.

[41] Gibbs, E. B.; Showalter, S. A. *Biochemistry* **2015**, 54, 1314−1326.

[42] Fuxreiter, M. *J Mol Biol* **2018**, 430, 2278-2287.

[43] Berlow, R. B.; Dyson H. J.; Wright, P. E. *J Mol Biol* **2018**, 430, 2309–2320.

[44] Poosapati, A.; Gregory, E.; Borcherds, W. M.; Chemes, L. B.; Daughdrill, G. W. *J Mol Biol* **2018**, 430, 2389–2402.

[45] Micsonai, A.; Wien, F.; Kernya, L.; Lee, Y.; Goto, Y.; Réfrégiers, M.; Kardos, J. *PNAS* **2015**, 24, 3095–3103.

[46] Bertini, I.; Duma, L.; Felli, I. C.; Fey, M.; Luchinat, C.; Pierattelli, R.; Vasos, P. R. *Angew. Chem. Int. Ed.* **2004**, 43, 2257 –2259.

[47] Bermel, W.; Bertini, I.; C. Felli, I. C.; Kummerle, R.; Pierattelli, R. *Journal of Magnetic Resonance* **2006**, 178, 56–64.

[48] Bermel, W.; Felli, I. C.; Gonnelli, L.; Kozminski, W.; Piai, A.; Pierattelli, R.; Zawadzka-Kazimierczuk, A. *J Biomol NMR* **2013**, 57, 353-61.

[49] Bertini, I.; Isabella C. Felli, I. C.; Gonnelli, L.; Kumar M. V. V.; Roberta Pierattelli, R. *ChemBioChem* **2011**, 12, 2347 – 2352.

[50] Haller, J. D.; Bodor, A.; Luy, B. *J Magn Reson.* **2019**, 302, 64-71.

[51] Plotnikov, A.; Zehorai E.; Procaccia S.; Seger, R. *Biochim Biophys Acta.* **2011**, 1813, 1619-33.

[52] Gehart, H.; Kumpf, S.; Ittner, A.; Ricci, R. *EMBO reports* **2010**, 11, 834-840.

[53] Huang, P.; Han, J.; Hui, L. *Protein Cell* **2010**, 1, 218–226.

[54] Kim, E. K.; Choi, E. *Biochimica et Biophysica Acta* **2010**, 1802, 396–405.

[55] Johnson, G. L.; Lapadat, R. *Science* **2002**, 298, 1911–1912.

[56] Cargnello, M.; Roux, P.P. *Microbiol Mol Biol Rev* **2011**, 75, 50–83.

[57] Garai, A.; Zeke, A.; Gógl, G.; Törö, I.; Fördös, F.; Blankenburg, H.; Bárkai, T.; Varga, J.; Alexa, A.; Emig, D.; Albrecht, M.; Reményi, A. *Sci Signal.* **2013**, 5, 245.

[58] Neduva, V.; Russell, R. B. *FEBS Letters* **2005**, 579, 3342-3345.

[59] Levine, A. J. *Cell.* **1997**, 88, 23-31.

[60] Vaseva, A. V.; Moll, U. M. *Biochimica et Biophysica Acta* **2009**, 1787, 414–420.

[61] Cho, Y.; Gorina, S.; Jeffrey, P. D.; Pavletich, N. P. *Science* **1994**, 5170, 346-355.

[62] Clore, G. M.; Ernst, J.; Clubb, R.; Omichinski, J. G.; Kennedy, W. M. P.; Sakaguchi, K.; Appella, E.; Gronenborn, A. M. *Nature Structural Biology* **1995**, 2, 321–333.

[63] Lee, W.; Harvey, T. S.; Yin, Y.; Yau, P.; Litchfield, D.; Arrowsmith, C. H. *Nature Structural Biology* **1994**, 1, 877–890.

[64] Chang, J.; Kim, D.-H.; Lee, S. W.; Choi, K. Y.; Sung, Y. C. *J. Biol. Chem.* **1995**, 270, 25014–25019.

[65] Hua, Q.; Jia, W.; Bullock, B. P.; Habener, J. F.; Weiss, M. A. *Biochemistry* **1998**, 37, 5858–5866.

[66] Hi, R.; Osada, S.; Yumoti, N.; Osumi, T. *J. Biol. Chem.* **1999**, 274, 35152–35158.

[67] Tell, G.; Perrone, L.; Fabbro, D.; Pellizzari, L.; Pucillo, C.; De Felince, M.; Acquaviva, R.; Formisano, S.; Damante, G. *Biochem. J.* **1999**, 329, 395–403.

[68] Lee, H.; Mok, K. H.; Muhandiram, R.; Park, K-H.; Suk, J-E.; Kim, D-H.; Chang, J.; Sung, Y. C.; Choi, K. Y.; Han, K-H. *The Journal of Biological Chemistry* **2000**, 275, 29426-29432.

[69] Xiong, Y.; Wu, Y.; Luo, S.; Gao, Y.; Xiong, Y.; Chen, D.; Deng, H.; Hao, W.; Liu, T.; Li, M. *Scientific Reports* **2017**, 7, Article number: 9185.

[70] Chen, H.; Fernig, D. G.; Rudland, P. S.; Sparks, A.; Wilkinson, M. C.; Barraclough, R. *Biochemical and Biophysical Research Communications* **2001**, 286, 1212–1217.

[71] Grigorian, M.; Andresen, S.; Tulchinsky, E.; Kriajevska, M.; Carlberg, C.; Kruse, C.; Cohn, M.; Ambartsumian, N.; Christensen, A.; Selivanovai, G.; Lukanidin, E. *The Journ of Biological Chemistry* **2001**, 276, 22699–22708.

[72] Orre, L. M.; Panizza, E.; Kaminskyy, V. O.; Vernet, E.; Graslund, T.; Zhivotovsky, B.; Lehtio, J.Oncogene **2013**, 32, 5531–5540.

[73] van Dieck, J.; Fernandez-Fernandez, M. R.; Veprintsev, D. B.; Fersht, A. R. *The Journal of Biological Chemistry* **2009**, 284, 13804–13811.

[74] Hong, P.; Koza, S.; Bouvier, E. S. P. *Journal of Liquid Chromatography & Related Technologies* **2012**, 35, 2923-2950.

[75] Murphy, R. M. *Current Opinion in Biotechnology* **1997**, 8, 25-30.

[76] Morris G. *Encyclopedia of Nuclear Magnetic Resonance*, John Wiley & Sons, Ltd, 2002.

[77] Cohen, Y.; Avram, L.; Frish, L. *Angew. Chem.* **2004**, 44, 520-554.

[78] Guo, X.; Laryea, E.; Wilhelm, M.; Luy, B.; Nirschl, H.; Guthausen, G. *Macromol. Chem. Phys.* **2017**, 218, 1600440.

[79] Kuzmina, N. E.; Moiseev, S. V.; Krylov, V. I.; Yashkir, V. A.; Merkulov, V. A. *Journal of Analytical Chemistry* **2014**, 69, pp. 953–959.

[80] Danielsson, J.; Jarvet, J.; Damberg, P.; Gräslund, A. *Magn. Reson. Chem.* **2002**, 40, S89-S97.

[81] Groves, P.; Palczewska, M.; Molero, M. D.; Batta, G.; Canada, F. J.; Jimenez-Barbero, J. *Analytical Biochemistry* **2004**, 331, 395–397.

[82] Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J. *Biochemistry* **1999**, 38, 16424-16431.

[83] Marsh, J. A.; Forman-Kay, J. D. *Biophys. Journ.* **2010**, 98, 2383-2390.

[84] Uversky, V. N.; Santambrogio, C.; Brocca, S.; Grandori, R. *FEBS Letters* **2012**, 586, 70-73.

[85] Wang, C. K.; Northfield, S. E.; Swedberg, J. E.; Harvey, P. J.; Mathiowetz, A. M.; Price, D. A.; Liras, S.; Craik, D. J. *J. Phys. Chem. B* **2014**, 118, 11129−11136.

[86] Evans, R.; Deng, Z.; Rogerson, A. K.; MacLachlan, A. S.; Richards, J. J.; Nilsson, M.; Morris, G. A. *Angew. Chem. Int. Ed.* **2013**, 52, 3199 –3202.

[87] For characteristic r_G/r_H values see: (a) Tande, B. M.; Wagner, N. J.; Mackay, M. E.; Hawker, C. J.; Jeong, M. *Macromolecules* **2001**, 34, 8580-8585. (b) Hofmann, H.; Andrea Soranno, A.; Borgia, A.; Gast, K.; Nettels, D.; Schuler, B. *PNAS* **2012**, 109, 16155-16160. (c) Kesimer, M.; Gupta, R., *Methods* **2015**, 87, 59–63. (d) Dias, J.; Renault, L., Perez, J.; Mirande, M. *The Journal of Biological Chemistry* **2013**, 288, 23979 –23989. (e) Aivaliotis, M.; Samolis, P.; Neofotistou, E.; Remigy, H.; Rizos, A. K.; Tsiotis, G. *Biochimica et Biophysica Acta* **2003**, 1615, 69–76.

[88] Wacha, A.; Varga, Z.; Bóta, A. *J. Appl. Cryst.* **2014**, 47, 1749-1754.

[89] Sanders, C.R.; Prosser, *Structure* **1998**, 6, 1227–1234.

[90] Warschawski, D. E.; Arnold, A. A.; Beaugrand, M.; Chartrand, A. G. E.; Marcotte, I. *Biochimica et Biophysica Acta* **2011**, 1808, 1957–1974.

[91] Killian, J. A. *FEBS Lett.* **2003**, 555, 134–138.

[92] Cabane, B.; Duplessix, R.; Zemb, T. *J. Phys. France* **1985,** 46, 2161-2178.

[93] Eriksson, J. C.; Ljunggren, S.; Henriksson, U. *J. Chem. SOC.,* **1985**, 81, 833-868.

[94] Zhao, J.; Fung, B. M. *Langmuir* **1993**, 9, 1228-1231.

[95] Sanders, C. R.; Oxenoid, K. *Biochimica et Biophysica Acta* **2000**, 1508, 129–145.

[96] Lin, T-L.; Chen, S-H.; Gabriel, N. E.; Roberts, M. F. *J. Am. Chem. SOC.* **1986**, 108, 3499-3501.

[97] Lipfert, J.; Columbus, L.; Chu, V. B.; Lesley, S. A.; Doniach, S. *J. Phys. Chem. B* **2007**, 111, 12427–12438.

[98] Mitra, K.; Ubarretxena-Belandia, I.; Taguchi, T.; Warren, G.; Engelman, D. M. *Proc. Natl. Acad. Sci. USA* **2004**, 101, 4083–4088.

[99] Holt, A.; Koehorst, R. B. M.; Rutters-Meijneke, T.; Gelb, M. H.; Rijkers, D. T. S.; Hemminga, M. A.; Killian, J. A. *Biophysical Journal* **2009**, 97, 2258–2266.

[100] Vostrikov, V. V.; Grant, C. V.; Daily, A. E.; Opella, S. J.; Koeppe, R. E.; *J. AM. CHEM. SOC.* **2008**, 130, 12584–12585.

[101] Ozdirekcan, S.; Etchebest, C.; Killian, J. A.; Fuchs, P. F. J. *J. AM. CHEM. SOC.* **2007**, 129, 15174-15181.

[102] Dempsey, C. E. *Biochimica et Biophysica Acta*, **1990,** 1031, 143-161.

[103] Steiner, H.; Hultmark, D.; Engström, A.; Bennich, H.; Boman, H. G. *Nature* **1981**, 292, 246-248.

[104] Terwilliger, T. C.; Weissman, L.; Eisenberg, D. *Biophysical Journal* **1982**, 37, 353-361.

[105] Yang, L.; Harroun, T. A.; Weiss, T. M.; Ding, L.; Huang, H. W. *Biophysical Journal* **2001**,81, 1475-1485.

[106] Luchette, P. A.; Vetman, T. N.; Prosser, R. S.; Hancock, R. E. W.; Nieh, M.-P.; Glinka, C. J.; Krueger, S.; Katsaras, J. *Biochimica et Biophysica Acta* **2001**, 1513, 83–94.

[107] Midtgaard, S. R.; Pedersen, M. C.; Kirkensgaard, J. J. K.; Sørensen, K. K.; Mortensen, K.; Jensen, K. J.; Morrison, L. A. *Soft Matter* **2014**, 10, 738–752.

[108] Bodor, A.; E. Kover, K.; Mäler, L. *Biochimica et Biophysica Acta* **2015**, 1848, 760–766.

[109] Morrison, E. A.; Henzler-Wildman, K. A. *Biochimica et Biophysica Acta* **2012**, 1818, 814–820.

[110] Caillon, L.; Lequin, O.; Khemtémourian, L. *Biochimica et Biophysica Acta* **2013**, 1828, 2091–2098.

[111] Glover, K. J.; Whiles, J. A.; Wu, G.; Yu, N.; Deems, R.; Struppe, J. O.; Stark, R. E.; Komives, E. A.; Vold, R. R. *Biophysical Journal* **2001**, 81, 2163-2171.

[112] Bocchinfuso, G.; Palleschi, A.; Orioni, B.; Grande, G.; Formaggio, F.; Toniolo, C.; Park, Y.; Hahm, K-S.; Stella, L. *J. Pept. Sci.* **2009**, 15, 550–558.

[113] Unnerstale, S.; Lind, J.; Papadopoulos, E.; Maler, L. *Biochemistry* **2009**, 48, 5813–5821.

[114] Unnerstale, S.; Maler, L.; Draheim, R. R. *Biochimica et Biophysica Acta* **2011**, 1808, 2403–2412.

[115] Liebau, J.; Ye, W.; Maler, L. *Magn. Reson. Chem.* **2017,** 55, 395–404.

[116] Wu, H.; Su, K.; Guan, X.; Sublette, M. E.; Stark, R. E. *Biochimica et Biophysica Acta* **2010**, 1798, 482–488.

[117] Chou, J. J.; Baber, J. L.; Bax, A. *Journal of Biomolecular NMR* **2004**, 29, 299–308.

[118] Diller, A.; Loudet, C.; Aussenac, F.; Raffard, G.; Fournier, S.; Laguerre, M.; Grelard, A.; Opella, S. J.; Marassi, F. M.; Dufourc, E. J. *Biochimie* **2009**, 91, 744–751.

[119] Carlotti, C.; Aussenac, F.; Dufourc, E. J. *Biochimica et Biophysica Acta* **2002,** 1564, 156– 164.

[120] Mineev, K. S.; Nadezhdin, K. D.; Goncharuk, S. A.; Arseniev, A. S. *Langmuir* **2016,** 32, 6624−6637.

[121] Lind, J.; Nordin, J.; Maler, L. *Biochimica et Biophysica Acta* **2008**, 1778, 2526–2534.

[122] Biverstahl, H.; Andersson, A.; Graslund, A.; Maler L. *Biochemistry* **2004,** 43, 14940-14947.

[123] Stejskal, E. O.; Tanner, J. E. *J. Chem. Phys.* **1965**, 42, 288–292.

[124] Tanner, J. E. *J. Chem. Phys.* **1970**, 52, 2523–2526.

[125] Dixon, A. M.; Larive, C. K. *Appl. Spectrosc.* **1999**, 53, 426A–440A.

[126] Gibbs, S. J.; Johnson, C. S. Jr., *J. Magn. Reson.* **1991**, 93, 395–402.

[127] Wu, D.; Chen, A.; Johnson, C. S. Jr., *J. Am. Chem. Soc.* **1993**, 115, 4291 – 4299.

[128] Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley & Sons Inc, 1991.

[129] Redfield C. *Protein NMR*; Springer, 2015.

[130] Wüthrich, K. *The Journal of Biological Chemistry* **1990**, 36, 22059-22062.

[131] Bax, A. *Annu. Rev. Biochem.* **1989**, 58, 223-56.

[132] Wishart, D. S.; Bigam, C. G.; Holm, A.; Hodges, R. S.; Sykes, B.D. *J. Biomol. NMR* **1995**, 5, 67-81.

[133] http://www.bp.uni-bayreuth.de/NMR/nmr_aminotocsy.html (visited on May 12, 2019)

[134] Oschkinat, H.; Muller, T.; Dieckmann, T. *Angew. Chem. Int. Ed.* **1994**, 33, 277-293.

[135] Sattler, M.; Schleucher, J.; Griesinger, C. *Progress in Nuclear Magnetic Resonance Spectroscopy* **1999**, 34, 93–158.

[136] Kay, L. E.; Marion, D.; Bax, A. *Journal of Magnetic Resonance* **1989**, 84, 72-84.

[137] Lescop, E.; Kern, T.; Brutscher, B. *Journal of Magnetic Resonance* **2010**, 203, 190-198.

[138] Schanda, P.; Forge, V.; Brutscher, B. *PNAS* **2007**, 104, 11257–11262.

[139] Keller, R. L. J. *The Computer Aided Resonance Assignment*; CANTINA Verlag, 2004.

[140] Vranken, W. F.; Boucher, W.; Stevens T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E.D. *Proteins* **2005**, 4, 687-96.

[141] Narayanan, R.L.; Dürr, U.H.; Bibow, S.; Biernat, J.; Mandelkow, E.; Zweckstetter, M.J. *Am Chem Soc.* **2010**, 132, 1906-7.

[142] Marsh, J. A.; Singh, V. K.; Jia, Z.; Forman-Kay, J. *Protein Science* **2006**, 15, 2795–2804.

[143] Williamson, M. P. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2013**, 73, 1-16.

[144] Andersen, N. H.; Neidigh, J. W.; Harris, S. M.; Lee, G. M.; Liu, Z.; Tong, H. *J. Am. Chem. Soc.* **1997**, 36, 8547-8561.

[145] Baxter, N. J.; Williamson, M. P. *J Biomol NMR* **1997**, 9, 359.

[146] Hospital, A.; Goni, J. R.; Orozco, M.; Gelpi, J. L. *Adv Appl Bioinform Chem.* **2015**, 8, 37–47.

[147] ter Haar, E., P.; Prabhakar, X. L.; Lepre, C*. J Biol Chem* **2007**, 282, 9733-9.

[148] Garai, A.; Zeke, A.; Gógl, G.; Törő, I.; Fördős, F.; Blankenburg, H.; Bárkai, T.; Varga, J., Alexa, A.; Emig, D.; Albrecht, M.; Reményi, A. *Sci. Signal*. **2012**, 245, 74.

[149] Alexa, A.; Gógl, G.; Glatz, G.; Garai, A.; Zeke, A.; Varga, J.; Dudás, E.; Jeszenői, N.; Bodor, A.; Hetényi, Cs.; Reményi, A. *Proceedings of the National Academy of Sciences* **2015**, 9, 2711-2716.

[150] Follis, A. V.; Llambi, F.; Ou, L.; Baran, K.; Green, D. R.; Kriwacki, R. W. *Nature Structural & Molecular Biology* **2014,** 21, 535–543.

[151] Borcherds, W.; Theillet, F-X.; Katzer, A.; Finzel, A.; Mishall, K. M.; Powell, A. T.; Wu, H.; Manieri, W.; Dieterich, C.; Selenko, P.; Loewer, A.; Daughdrill, G. W. *Nature Chemical Biology* **2014**, 10, 1000–1002.

[152] Chen, H. F.; Luo, R*. J. Am. Chem. Soc.* **2007**, 129, 2930-2937.

[153] Rajagopalan, S.; Andreeva, A.; Teufel, D. P.; Freund, S. M.; Fersht, A. R*. J Biol Chem.* **2009**, 284, 21728-37.

[154] Vise, P. D.; Baral, B.; Latos, A. J.; Daughdrill, G. W. *Nucleic Acids Research*, **2005**, 33, 2061–2077.

[155] Di Lello, P.; Miller Jenkins, L. M.; Jones, T. N.; Nguyen, B. D.; Hara, T.; Yamaguchi, H.; Dikeakos, J. D.; Appella, E.; Legault, P.; Omichinski, J. G. *Molecular Cell* **2006**, 22, 731–740.

[156] Boye, K.; Mælandsmo, G. M. *The American Journal of Pathology* **2010**, 176, 528-535.

[157] Helfman, D. M.; Kim, E. J.; Lukanidin, E.; Grigorian, M. *British Journal of Cancer* **2005**, 92, 1955–1958.

[158] Mishra, S. K.; Rahman, H.; Saleem, S. M. *Cancer Metastasis Rev* **2012**, 31, 163.

[159] Roos, M.; Ott, M.; Hofmann, M.; Link, S.; Rössler, E.; Balbach, J.; Krushelnitsky, A.; Saalwächter, K*. J. Am. Chem. Soc.* **2016**, 138, 10365−10372.

[160] Makhatadze, G. I.; Privalov, P. L. *Advances in Protein Chemistry* **1995**, 47, 307-425.

[161] Eilers, M.; Shekar, S. C.; Shieh, T.; Smith, S. O.; Fleming, P. J. *PNAS* **2000**, 97, 5796-580.

[162] Jones, J. A.; Wilkins, D. K.; Smith, L. J.; Dobson, C. M. *Journal of Biomolecular NMR* **1997**, 10, 199–203.

[163] Flory P. *Statistical Mechanics of Chain Molecules*, Interscience, 1969.

[164] Ortega, A.; Amorós, D.; Garcia de la Torre, J. *Biophys. J.* **2011**, 101, 892-898.

[165] Voronoi, G. *Journal für die Reine & Angewandte Mathematik* **1908**,134, 198–287.

[166] Tcherkasskaya O.; Uversky, V. *PROTEINS: Structure, Function, and Genetics* **2001**, 44, 244–254.

[167] Burchard, W.; Schmidt, M.; Stockmayer, W. H. *Macromolecules* **1980**, 13, 1265-1272.

[168] Wu, C.; Zhou, S. *Macromolecules* **1995**, 28, 5388—5390.

[169] Wang, X.; Goh, S. H.; Lu, Z. H.; Lee, S. Y.; Wu, C. *Macromolecules* **1999,** *32,* 2786-2788.

[170] Wan, W-M.; Sun, X-L.; Pan, C-Y. *Macromolecules* **2009**, 42, 4950–4952.

[171] You, Y-Z.; Hong, C-Y.; Pan, C-Y.; Wang, P-H. *Adv. Mater.* **2004**, 16, 1953-1957.

[172] Blackburn, E. A.; Fuad, F. A. A.;.Morgan, H. P.; Nowicki, M. W.; Wear, M. A.; Michels, P. A. M.; Fothergill-Gilmore, L. A.; Walkinshaw, M. D. *Biochemical and Biophysical Research Communications* **2014**, 450, 936–941.

[173] Receveur-Bréchot, V.; Durand, D. *Current Protein and Peptide Science*, **2012**, 13, 55-75.

[174] Uversky, V. N.; Li, J.; Souillac, P.; Millett, I. S.; Doniach, S.; Jakes, R.; Goedert, M.; Fink, A. L. *The Journal of Biological Chemistry* **2002**, 277, 11970–11978.

[175] Damaschun, G.; Damaschun, H.; Gast, K.; Gernat, C.; Zirwer, D. *Biochim. et Biopshy. Acta* **1991**, 1078, 289-295.

[176] Permyakov, S. E.; Millett, I. S.; Doniach, S.; Permyakov, E. A.; Uversky V. N. *PROTEINS: Structure, Function, and Genetics* **2003**, 53, 855–862.

[177] Borgia, A.; Kemplen, K. R.; Borgia, M. B.; Soranno, A.; Shammas, S.; Wunderlich, B.; Nettels, D.; Best, R. B.; Clarke, J.; Schuler, B. *Nature Comm* **2015**, 6:8861.

[178] Wang, J.; Dauter, M.; Alkire, R.; Joachimiak, A.; Dauter, Z. *Acta Cryst.* **2007**, 63, 1254-1268.

[179] Kuboniwa, H.; Tjandra, N.; Grzesiek, S.; Ren, H.; Klee, C. B.; Bax, A. *Nat. Struct. Mol. Biol.* **1995**, 2, 768–776.

[180] Babu, Y. S.; Bugg, C. E.; Cook, W. J. J. Mol. Biol. **1988**, 294, 191-204.

[181] Gopalakrishna, R.; Anderson, W. B. *Biochemical and Biophysical Research Communications* **1982**, 2, 830-836.

[182] Panjkovich, A.; Svergun, D.I. *Phys. Chem. Chem. Phys.* **2016**, 18, 5707-5719.

[183] Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry, Part II: Techniques for the Study of Biological Structure and Function*; W. H. Freeman, **1980**.

[184] van Dam, L.; Karlsson, G.; Edwards, K. *Biochimica et Biophysica Acta* **2004**, 1664, 241– 256.

[185] F. Dudas, E.; Bodor, A. *Anal. Chem.* **2019**, 91, 4929–4933.

[186] Björneras, J.; Nilsson, M.; Mäler, L. *Biochimica et Biophysica Acta* **2015**, 1848, 2910–2917.

[187] Vold, R. R.; Prosser, R. S.; Deese, A. J. *J. Biomol. NMR* **1997**, 9, 329–335.

[188] Tausk, R. J. M.; Esch, J. V.; Karmiggelt, J.; Voordouw, G.; Overbeek. J. T. G. *Biophys. Chem*. **1974**, 1, 184–203.

[189] Koenig, B. W.; Strey, H. H.; Gawrisch, K. *Biophys. J.* **1997**, 73, 1954 –1966.

[190] Triba, M. N.; Warschawski, D. E.; Devaux, P. F. *Biophysical Journal* **2005**, 88, 1887-1901.

[191] Situ, A. J.; Schmidt, T.; Mazumder, P.; Ulmer, T. S. *Journal of Molecular Biology* **2014**, 21, 3670-3680.

[192] Lee, D.; Walter, K. F. A.; Bruckner, A.; Hilty, C.; Becker, S.; Griesinger, C. *J. AM. CHEM. SOC.* **2008**, 130, 13822–13823.

[193] Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry,* W. H. Freeman and Co., 1980.

[194] Losonczi, J. A.; Prestegard, J. H. *J. Biomol. NMR* **1998**, 12, 447– 451.

Appendix

Figure A1. Sequence-corrected Hα (a), Cα (b), Cα-Cβ (c) secondary chemical shifts of RSK1_S/A in aqueous solution (blue) and under denaturing conditions (red) in ppm.

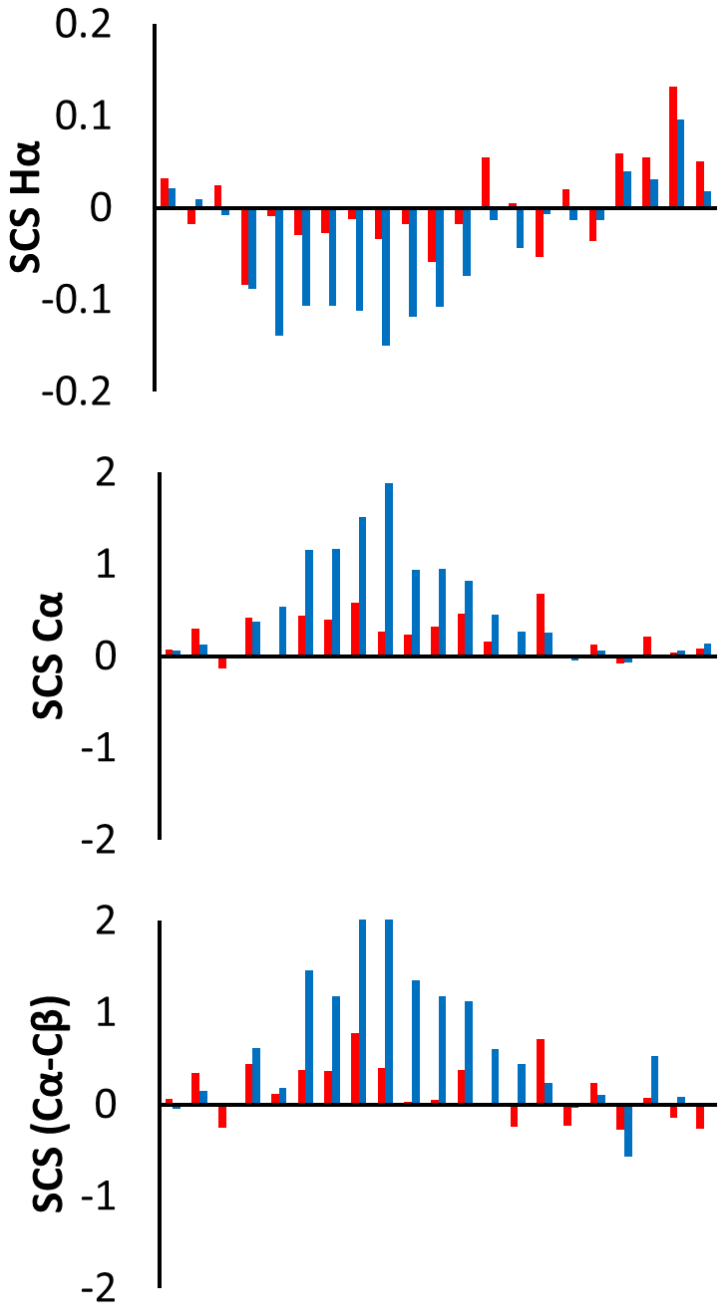


Figure A2. p53 TAD regions 1-29 (TAD1), 30-47 (linker) and 48-56 (TAD2) cut from PDB structures aligned with corresponding regions of the simulated p53TAD-S100A4 models (Model A and Model B). Model A/Model B, 1ycq, 1ycr, 2b3g, 2gs0, 2k8f, 2l14, 2ly4, 2z5t, 2mwy, 2mzd, 2ruk, 2z5s, 3dab, 3dac, 4hfz, 5hp0, 5hou, 5hpd

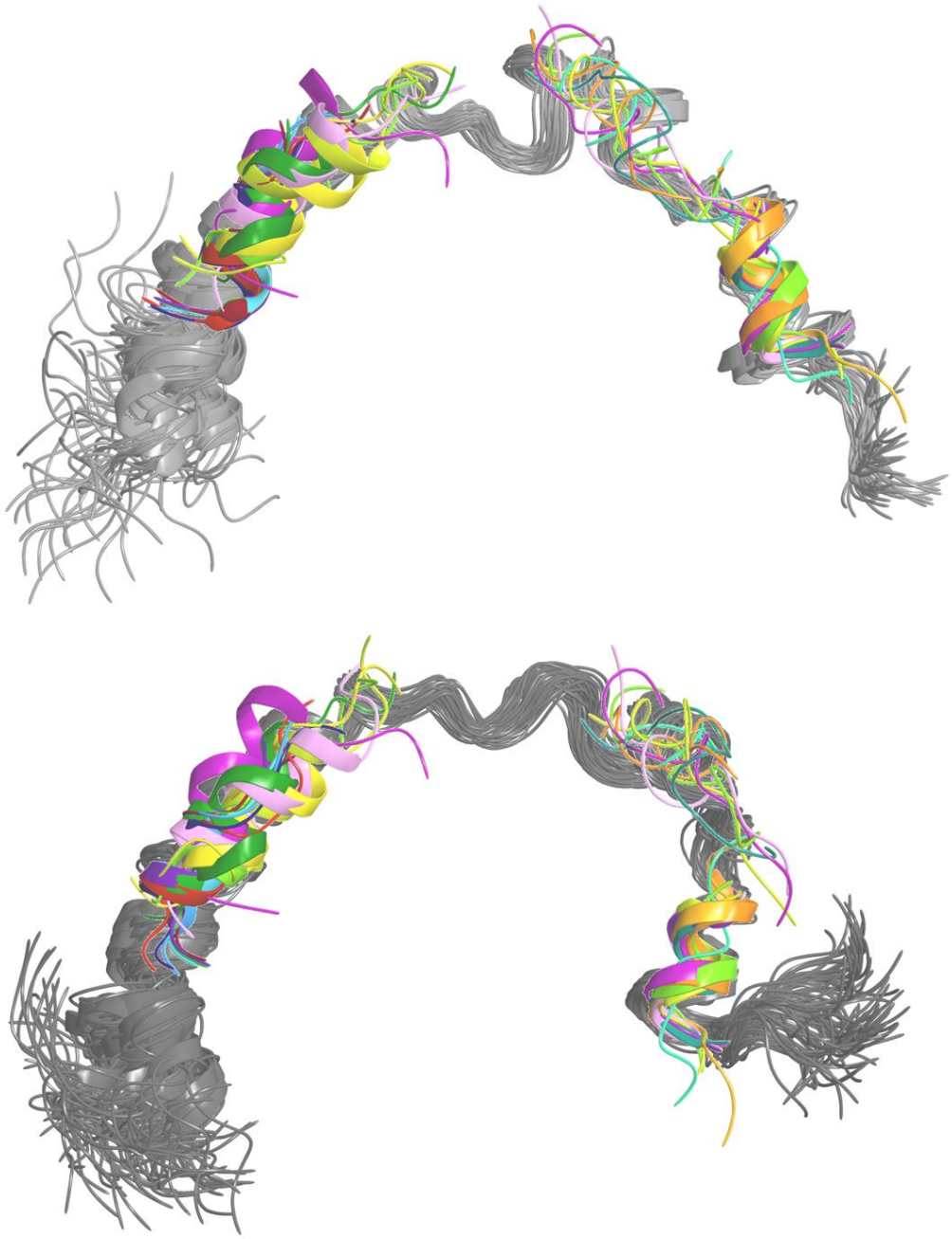
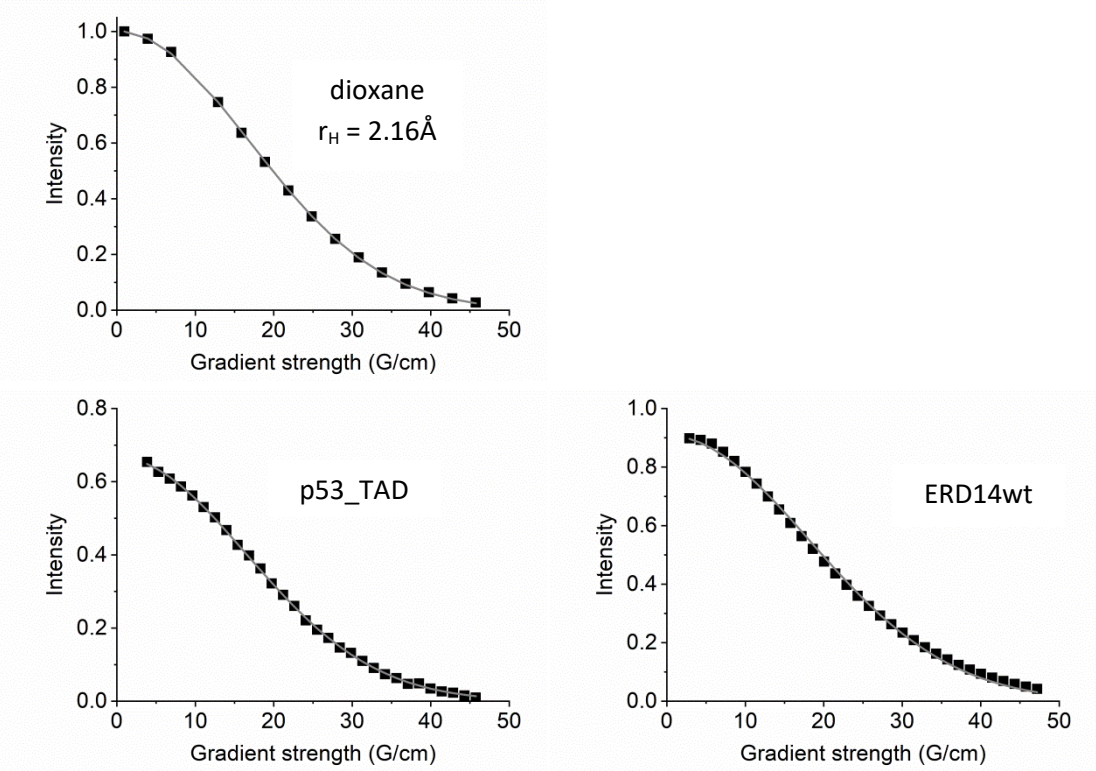
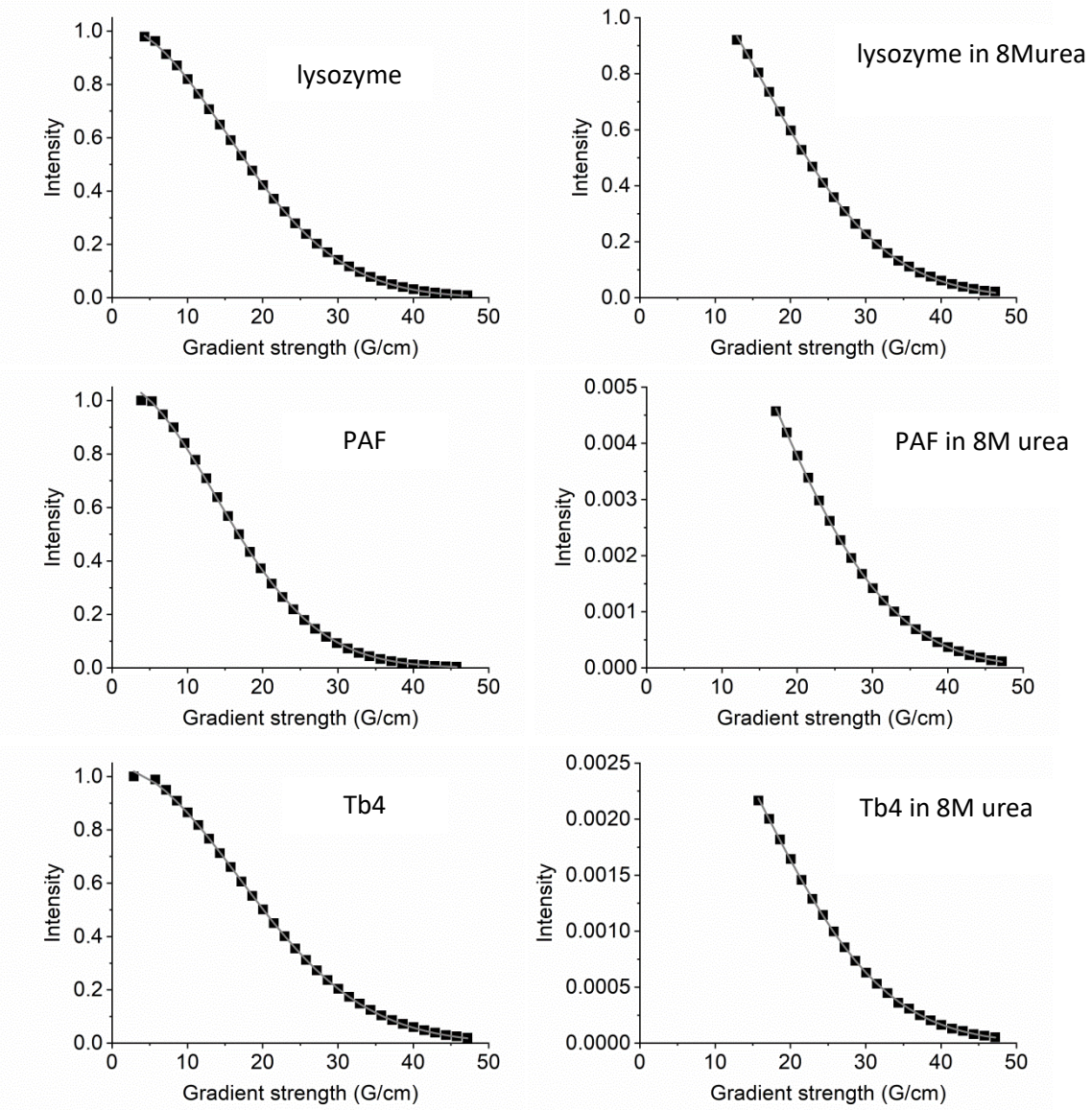


Figure A3. Examples of data evaluation for several selected proteins: intensity decays as a function of increasing gradient strengths.

Peak intensities were obtained from the aliphatic proton region (< 3ppm chemical shift range), as the amide proton exchange can influence the intensities in the –NH region, and the solvent water peak might affect the neighboring environments. Black squares represent the experimental data points and the grey solid line is the fitted curve based on the equation:

$$I = I_0 \exp[-D\gamma^2\delta^2G^2\left(\Delta - \frac{\delta}{3}\right)]$$





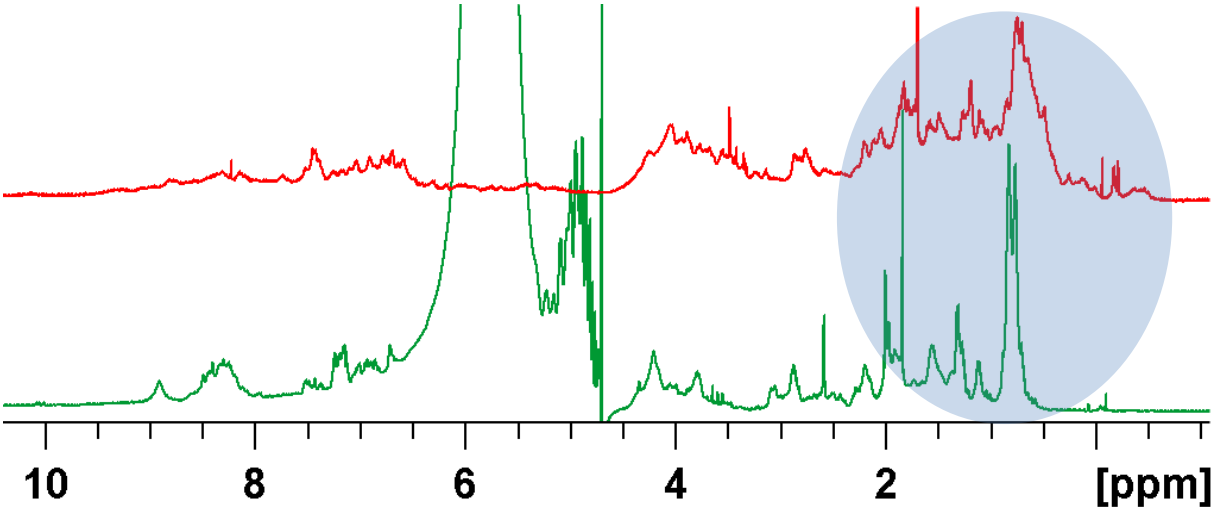
Protein signals in the denatured samples were perturbed by the extremely intensive urea signal; Gaussian fitting was done starting from higher (15-18G/cm) G values after urea signal decay was complete.

Validation of the PFG-NMR measurements and the absolute method for r_H calculation:

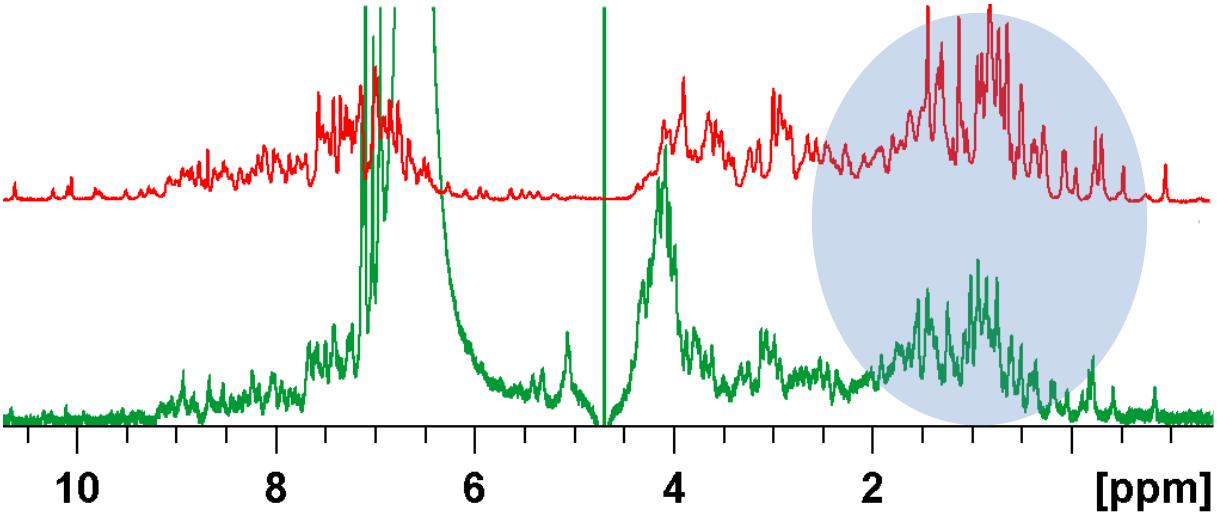
	T (K)	r_H literature (Å) /method/	r_H measured (Å)
lysozyme	293.0	19.4 /SAXS/	20.4±0.4
glucose	298.0	2.8 /optical viscometry/	2.9±0.1
lactose	298.0	3.9/optical viscometry/	3.8±0.1

Figure A4.

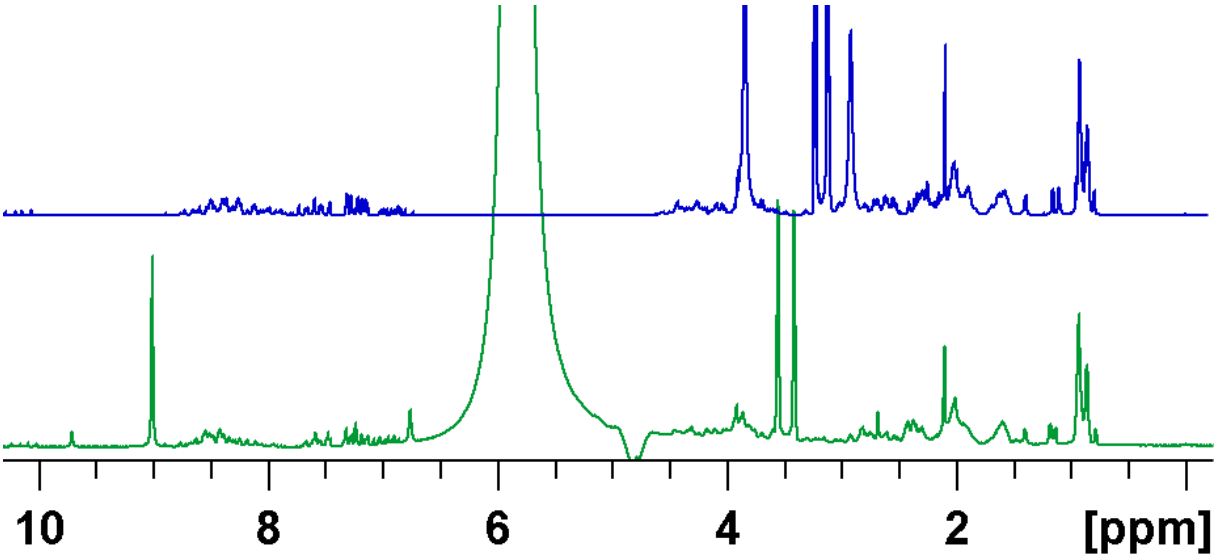
a) 1D ^{1}H spectra of ovalbumin in aqueous solution (red) and in 8M urea (green). The collapse of the structure as a consequence of denaturation is highlighted in the aliphatic region.



b) 1D ^{1}H spectra of lysozyme in aqueous solution (red) and in 8M urea (green). No collapse of the resonances is observed, as highlighted for the aliphatic region.



c) 1D ^{1}H spectra of p53TAD in aqueous solution (blue) and in 8M urea (green). No real variation of the spectra is observed.



d) 1D ^{1}H spectra of Tb4 in aqueous solution (blue) and in 8M urea (green).

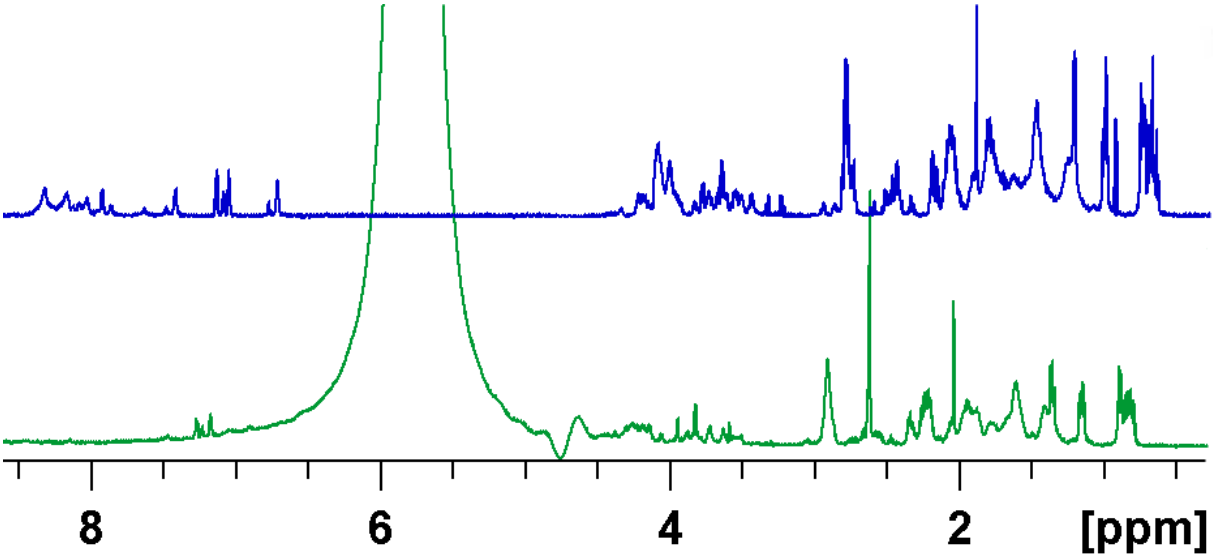


Figure A5. Complete derivation of the characteristic r_G/r_H=0.77 value for a spherical object (r_G is denoted as R_G while r_H is denoted as r)

I. The radius of gyration in case of a solid sphere rotating around *a single axis*:

$$R_G^2 = \frac{2}{5}R^2$$

II. If *rotation is about random axes*, then for an inhomogeneous body:

$R_G^2 = \frac{\int r^2 \rho(r) dr^3}{\int \rho(r) dr^3}$ (the generalized I=mR² description of the moment of inertia)

and for a homogeneous body: $R_G^2 = \frac{\int r^2 dr^3}{\int dr^3}$ where $\int dr^3 = V = \frac{4}{3}\pi r^3$ **(A)**

Moving to polar coordinates, with φϵ[0,π] and ϕϵ[0,2π]

$$r^2 \cdot sin\theta \cdot dr \cdot d\theta \cdot d\varphi$$

we have: $\int r^2 dr^3 = \int r^4 \cdot sin\theta \cdot dr \cdot d\theta \cdot d\rho = -2 \int r^4 \cdot cos\theta \,{}^{\pi/2}/_0 \cdot dr \cdot d\rho =$

$$2 \int r^4 \cdot 2\pi \cdot dr = \frac{4}{5}\pi r^5 \qquad \textbf{(B)}$$

On the basis of (A) and (B):

$R_G^2 = \frac{\frac{4}{5}\pi r^5}{\frac{4}{3}\pi r^3} = \frac{3}{5}r^2$ $\qquad$ $\frac{R_G}{r} = \sqrt{\frac{3}{5}} = 0.77$

Table A1. Empirical D(M) and r_H(N) relations from literature (f-folded, IDP, den-denatured)

1. Danielsson (2002)

$$D = (6.06 \pm 0.01) \times 10^{-9} M^{-(0.44 \pm 0.02)} m^2 s^{-1}$$

Sequentially similar peptides, 298K, pH = 6.33-7.47, buffer 10 mM Na-phosphate

2. Groves (2004)

$$\log D = -0.366 \log M - 8.484$$

5 proteins, measured in D_2O at 298K.

M (EMW): effective molecular weight

3. Bernado (2009)

$$r_H{}^{IDP} = (3.53 \pm 0.01) N^{0.449 \pm 0.01}$$

Simulation, coil database, SAXS

4. Wilkins (1999)

native folded:

$$r_H{}^{f} = (4.75 \pm 1.11) N^{0.29 \pm 0.02}$$

denatured:

$$r_H{}^{den} = (2.21 \pm 1.07) N^{0.57 \pm 0.02}$$

Different buffers and different denaturing agents, 293K, pH = 2.0-7.2

5. Marsh (2010)

$$r_H^{f} = 4.92 N^{0.285}$$
$$r_H^{den} = 2.33 N^{0.549}$$
$$r_H^{IDP} = 2.49 N^{0.509}$$

Based on literature data, PFG-NMR and SEC, different conditions

6. Uversky (2012)

$$\log r_H^{NC} = (0.454 \pm 0.017) + (0.493 \pm 0.008) \times \log N$$
$$\log r_H^{PMG} = (0.587 \pm 0.029) + (0.402 \pm 0.012) \times \log N$$
$$\log r_H^{MG} = (0.629 \pm 0.051) + (0.334 \pm 0.021) \times \log N$$
$$\log r_H^{f} = (0.525 \pm 0.012) + (0.358 \pm 0.005) \times \log N$$
$$\log r_H^{c} = (0.385 \pm 0.017) + (0.543 \pm 0.007) \times \log N$$

NC: native coil, PMG: pre-molten globule, MG: molten globule, f: folded, c: chemically unfolded

Based on viscometry, gel filtration and DLS methods

7. Evans (2013)

$$D = \frac{k_B T\left(\frac{3\alpha}{2} + \frac{1}{1+\alpha}\right)}{6\pi\eta\sqrt[3]{\frac{3M}{4\pi\rho_{eff}N_A}}} \quad \text{where} \quad \alpha = \sqrt[3]{\frac{M_S}{M}}$$

Small molecules, in D_2O, $CDCl_3$, d_6-DMSO, d_8-toluene, d_4-methanol

Where M stands for the molecular weight of the small molecule investigated

8. Nygaard (2017)

$$\frac{r_G}{r_H} = \frac{0.216(r_G - 4.06N^{0.33})}{N^{0.60} - N^{0.33}} + 0.821$$

Based on Flexible-Meccano, MD simulations and HYDROPRO calculations

Table A2. Overview of the investigated proteins: residue number, molecular weight, amino acid sequence, net charge, number of negatively and positively charged residues, theoretical pI and the pH value of the present study.

	Name	N	M (g/mol)	Sequence	Net charge	# of neg. charged res.	# of pos. charged res.	pI	pH
	IDP								
1	NFAT	14	1738	LERPSRDHLYLPLE	-1	-3	2	5.45	6.20
2	MNK1	19	2195	MKLSPPSKSRLARRRRALA	7	0	7	12.6	6.20
3	MK2	20	2392	IKIKKIEDASNPLLLKRRKK	6	-2	8	10.66	6.30
4	RSK	24	2732	PQLKPIESSILAQRRVRKLPSTTL	4	-1	5	11.72	6.00
5	Tb4	43	5053	MSDKPDMAEIEKFDKSKLKKTETQEKNPLPSKETIEQEKQAGES	-2	-11	9	5.02	7.05
6	p53 TAD	62	7000*	GSMEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPNNIEQWFTEDPGP	-11	-12	1	3.43	5.00
7	M67	67	7440*	RKLQRELEDATETADAMNREVSSLKNKLRRGDLPFVVPRRMARKGAGDGSDEEVDGKADGAEAKPAE	-1	-15	14	5.45	6.00
8	SMAR3	168	18370	MGTSSPALKPASESFGEKKEEGKKRMYHVSEEPELNVSPLDAVEKVHETIPKEDPHSIPENTNFPAGIEPQSNIPETPPPGYLSEDGETSDHQMNHSMDAGSPNLSPAAAEKGITGEPKHFPEQVFSSDVAVVVDKESILEPPKKTERLEHHHHHHEPEADTAGHELS	-19	-33	14	4.94	6.38
9	ERD14 wt	185	20786	MAEEIKNVPEQEVPKVATEESSAEVTDRGLFDFLGKKKDETKPEETPIASEFEQKVHISEPEPEVKHESLLEKLHRSDSSSSSSSEEEGSDGEKRKKKKEKKKPTTEVEVKEEEKKGFMEKLKEKLPGHKKPEDGSAVAAAPVVVPPPVEEAHPVEKKGILEKIKEKLPGYHPKTTVEEEKKDKE	-9	-46	37	5.4	7.70
10	fullscrERD14	186	20773	MKTGKLPEETSAFGKSNESIVGETKEDKMEESPEPIVSKSKVEEQEEKKEKSTEALGVKPLPEEYASRVEPGPHAEVFDGEKVRHGGEEPGFLPAHGKSSVKTEKESAIEVRHETTEPELVEQKKLKPPIHKPKEDLEDVEHKEGVVDKKTLPVKDLSKIKSEKMKDSEAVPKAKSKEKEPKKFEA	-7	-44	37	5.58	6.50

* : ^{15}N-labelled proteins

NFAT, MNK1, MK2 and RSK are denoted as 'MAPK fragments' on Figure 22

	Name	N	M (g/mol)	Sequence	Net charge	Nr of neg. charged res.	Nr of pos. charged res.	pI	pH
	Folded proteins								
1	TC5b	20	2169	NLYIQWLKDGGPSSGRPPPS	1	-1	2	3.8	7.00
2	TC5bS13E	20	2211	NLYIQWLKDGGPESGRPPPS	0	-2	2	6.8	7.00
3	PAF	55	6250	AKYTGKCTKSKNECKYKNDAGKDTFIKCPKFDNKKCTKDNNKCTVDTYNNAVDCD	5	-8	13	8.93	6.00
4	BPTI	58	6517	RPDFCLEPPYTGPCKARIIRYFYNAKAGLCQTFVYGGCRAKRNNFKSAEDCMRTCGGA	6	-4	10	9.24	4.80
5	ribonuclease	124	13690	KETAAAKFERQHMDSSTSAASSSNYCNQMMKSRNLTKDRCKPVNTFVHESLADVQAV CSQKNVACKNGQTNCYQSYSTMSITDCRETGSSKYPNCAYKTTQANKHIIVACEGNPY VPVHFDASV	6	-10	16	8.93	4.30
6	lysozyme	129	14313	KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDGSTDYGILQ INSRWWCNDGRTPGSRNLCNIPCSALLSSDITASVNCAKKIVSDGNGMNAWVAWRNR CKGTDVQAWIRGCRL	5	-8	13	8.96	3.20
7	calmodulinCa2+	149	16997	MADQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQDMINEVDA DGNGTIDFPEFLTMMARKMKDTDSEEEIREAFRVFDKDGNGYISAAELRHVMTNLGEK LTDEEVDEMIREADIDGDGQVNYEEFVQMMTAK	-24	-38	14	4.09	7.30
8	S100A4d13Ca2+	182	20966	GSHMACPLEKALDVMVSTFHKYSGKEGDKFKLNKSELKELLTRELPSFLGKRTDEAAFQK LMSNLDSNRDNEVDFQEYCVFLSCIAMMCNE	-6	-30	24	5.33	6.20
9	S100A4wt	208	24018	GSHMACPLEKALDVMVSTFHKYSGKEGDKFKLNKSELKELLTRELPSFLGKRTDEAAFQK LMSNLDSNRDNEVDFQEYCVFLSCIAMMCNEFFEGFPDKQPRKK	-2	-34	32	5.85	6.00
10	chymotrypsinogen	246	25678	CGVPAIQPVLSGLIVNGEEAVPGSWPWQVSLQDKTGFHFCGGSLINENWVVTAAHCG VTTSDVVVAGEFDQGSSSEKIQKLKIAKVFKNSKYNSLTINNDITLLKLSTAASFSQTVSAV CLPSASDDFAAGTTCVTTGWGLTRYANTPDRLQQASLPLLSNTNCKKYWGTKIKDAMI CAGASGVSSCMGDSGGPLVCKKNGAWTLVGIVSWGSSTCSTSTPGVYARVTALVNWV QQTLAAN	4	-14	18	8.52	3.50
11	ovalbumin	386	42881	MGSIGAASMEFCFDVFKELKVHHANENIFYCPIAIMSALAMVYLGAKDSTRTQINKVVR FDKLPGFGDSIEAQCGTSVNVHSSLRDILNQITKPNDVYSFSLASRLYAEERYPILPEYLQC VKELYRGGLEPINFQTAADQARELINSWVESQTNGIIRNVLQPSSVDSQTAMVLVNAIV FKGLWEKAFKDEDTQAMPFRVTEQESKPVQMMYQIGLFRVASMASEKMKILELPFASG TMSMLVLLPDEVSGLEQLESIINFEKLTEWTSSNVMEERKIKVYLPRMKMEEKYNLTSVL MAMGITDVFSSSANLSGISSAESLKISQAVHAAHAEINEAGREVVGSAEAGVDAASVSEE FRADHPFLFCIKHIATNAVLFFGRCVSP	-12	-47	35	5.19	6.10
12	BSA	583	66463	DTHKSEIAHRFKDLGEEHFKGLVLIAFSQYLQQCPFDEHVKLVNELTEFAKTCVADESHA GCEKSLHTLFGDELCKVASLRETYGDMADCCEKQEPERNECFLSHKDDSPDLPKLKPDP NTLCDEFKADEKKFWGKYLYEIARRHPYFYAPELLYYANKYNGVFQECCQAEDKGACLL PKIETMREKVLTSSARQRLRCASIQKFGERALKAWSVARLSQKFPKAEFVEVTKLVTDLT KVHKECCHGDLLECADDRADLAKYICDNQDTISSKLKECCDKPLLEKSHCIAEVEKDAIP ENLPPLTADFAEDKDVCKNYQEAKDAFLGSFLYEYSRRHPEYAVSVLLRLAKEYEATLEE CCAKDDPHACYSTVFDKLKHLVDEPQNLIKQNCDQFEKLGEYGFQNALIVRYTRKVPQ VSTPTLVEVSRSLGKVGTRCCTKPESERMPCTEDYLSLILNRLCVLHEKTPVSEKVTKCCT ESLVNRRPCFSALTPDETYVPKAFDEKLFTFHADICTLPDTEKQIKKQTALVELLKHKPKA TEEQLKTVMENFVAFVDKCCAADDKEACFAVEGPKLVVSTQTALA	-13	-99	86	5.82	6.60

Table A3. The investigated proteins, the measured translational diffusion coefficients and the calculated effective hydrodynamic radii

		D (m^2s^{-1})	r_H (Å)
IDP			
1	NFAT	(1.54±0.02)E-10	11.7±0.1
2	MNK1	(1.37±0.02)E-10	13.1±0.2
3	MK2	(1.27±0.03)E-10	14.1±0.3
4	RSK	(1.18±0.01)E-10	15.2±0.1
5	Tb4	(9.45±0.10)E-11	19.0±0.2
6	p53 TAD	(7.44±0.01)E-11	24.2±0.0
7	M67	(7.89±0.04)E-11	22.8±0.0
8	SMAR3	(4.78±0.04)E-11	37.6±0.3
9	ERD14 wt	(4.25±0.07)E-11	42,3±0.6
10	fullscrERD14	(4.19±0.10)E-11	42.9±0.9
folded			
1	TC5b	(1.76±0.04)E-10	10.2±0.3
2	TC5bS13E	(1.72±0.03)E-10	10.4±0.2
3	PAF	(1.06±0.03)E-10	17.0±0.5
4	BPTI	(1.09±0.03)E-10	16.5±0.4
5	ribonuclease	(9.17±0.04)E-11	19.6±0.1
6	lysozyme	(7.84±0.04)E-11	22.9±0.4
7	calmodulinCa2+	(7.19±0.09)E-11	25.0±0.3
8	S100A4d13Ca2+	(7.03±0.16)E-11	25.6±0.6
9	S100A4wt	(6.80±0.11)E-11	26.4±0.4
10	chymotrypsinogen	(7.41±0.04)E-11	24.2±0.5
11	ovalbumin	(5.10±0.03)E-11	35.2±1.0
12	BSA	(4.80±0.03)E-11	37.4±0.2

Table A4. The investigated proteins, their corresponding residue numbers (N) and calculated effective hydrodynamic radii based on the present study and previous empirical relations.

		N	measured r_H (Å)	calculated r_H (Å)	r_H Marsh (Å)	r_H Uversky (Å)	r_H Wilkins (Å)
IDP							
1	NFAT	14	11.7	11.5	9.5	10.4	
2	MNK1	19	13.1	13.3	11.1	12.1	
3	MK2	20	14.1	13.7	11.4	12.5	
4	RSK	24	15.2	14.9	12.6	13.6	
5	Tb4	43	19.0	19.9	16.9	18.2	
6	p53 TAD	62	24.2	23.8	20.3	21.8	
7	M67	67	22.8	24.8	21.2	22.6	
8	SMAR3	168	37.6	38.9	33.8	35.6	
9	ERD14wt	185	42.3	40.8	35.5	37.3	
10	fullscrERD14	186	42.9	40.9	35.6	37.4	
folded							
1	TC5b	20	10.2	10.7	11.6	9.8	11.3
2	TC5bS13E	20	10.4	10.7	11.6	9.8	11.3
3	PAF	55	17.0	15.7	15.4	14.1	15.2
4	BPTI	58	16.5	16.1	15.7	14.3	15.4
5	ribonuclease	124	19.6	21.5	19.4	18.8	19.2
6	lysozyme	129	22.9	21.8	19.7	19.1	19.4
7	calmodulinCa2+	149	25.0	23.0	20.5	20.1	20.3
8	S100A4D13Ca2+	182	25.6	24.9	21.7	21.6	21.5
9	S100A4wt	208	26.4	26.2	22.5	22.6	22.3
10	chymotrypsinogen	246	24.2	27.9	23.6	24.0	23.4
11	ovalbumin	386	35.2	33.1	26.9	28.2	26.7
12	BSA	583	37.4	38.8	30.2	32.7	30.1

r_H Wilkins[82], r_H Marsh[83], r_H Uversky[84]

ADATLAP

a doktori értekezés nyilvánosságra hozatalához*

I. A doktori értekezés adatai

A szerző neve: **Földesné Dudás Erika**

MTMT-azonosító: **10055041**

A doktori értekezés címe és alcíme: **NMR studies of folded and disordered proteins and bicelle systems**

DOI-azonosító[46]: **10.15476/ELTE.2019.091**

A doktori iskola neve: **Hevesy György Kémia Doktori Iskola**

A doktori iskolán belüli doktori program neve: **Szintetikus kémia, szerves és biomolekuláris kémia**

A témavezető neve és tudományos fokozata: **Dr. Bodor Andrea, PhD Habil.**

A témavezető munkahelye: **ELTE-TTK, Analitikai Kémiai Tanszék**

II. Nyilatkozatok

1. A doktori értekezés szerzőjeként
a) hozzájárulok, hogy a doktori fokozat megszerzését követően a doktori értekezésem és a tézisek nyilvánosságra kerüljenek az ELTE Digitális Intézményi Tudástárban. Felhatalmazom a Természettudományi kar Dékáni Hivatal Doktori, Habilitációs és Nemzetközi Ügyek Csoportjának ügyintézőjét, hogy az értekezést és a téziseket feltöltse az ELTE Digitális Intézményi Tudástárba, és ennek során kitöltse a feltöltéshez szükséges nyilatkozatokat.
b) kérem, hogy a mellékelt kérelemben részletezett szabadalmi, illetőleg oltalmi bejelentés közzétételéig a doktori értekezést ne bocsássák nyilvánosságra az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban;
c) kérem, hogy a nemzetbiztonsági okból minősített adatot tartalmazó doktori értekezést a minősítés (*dátum*)-ig tartó időtartama alatt ne bocsássák nyilvánosságra az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban;
d) kérem, hogy a mű kiadására vonatkozó mellékelt kiadó szerződésre tekintettel a doktori értekezést a könyv megjelenéséig ne bocsássák nyilvánosságra az Egyetemi Könyvtárban, és az ELTE Digitális Intézményi Tudástárban csak a könyv bibliográfiai adatait tegyék közzé. Ha a könyv a fokozatszerzést követőn egy évig nem jelenik meg, hozzájárulok, hogy a doktori értekezésem és a tézisek nyilvánosságra kerüljenek az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban.
2. A doktori értekezés szerzőjeként kijelentem, hogy
a) az ELTE Digitális Intézményi Tudástárba feltöltendő doktori értekezés és a tézisek saját eredeti, önálló szellemi munkám és legjobb tudomásom szerint nem sértem vele senki szerzői jogait;
b) a doktori értekezés és a tézisek nyomtatott változatai és az elektronikus adathordozón benyújtott tartalmak (szöveg és ábrák) mindenben megegyeznek.
3. A doktori értekezés szerzőjeként hozzájárulok a doktori értekezés és a tézisek szövegének plágiumkereső adatbázisba helyezéséhez és plágiumellenőrző vizsgálatok lefuttatásához.

Kelt: London, 2019.05.09.

Földesné Dudás Erika
.....
a doktori értekezés szerzőjének aláírása

**ELTE SZMSZ SZMR 12. sz. melléklet*