1949

Understanding structure and dynamics with advanced NMR and *in silico* methods

Thesis for the Degree of Doctor of Philosophy (PhD)

by **Tamás Milán Nagy**

Advisor: Prof. Katalin E. Kövér

UNIVERSITY OF DEBRECEN
Doctoral Council of Natural Sciences and Information Technology
Doctoral School of Chemistry
Debrecen, 2022

Hereby I declare that I prepared this thesis within the Doctoral Council of Natural Sciences and Information Technology, Doctoral School of Chemistry, University of Debrecen in order to obtain a PhD Degree in Natural Sciences at Debrecen University. The results published in the thesis are not reported in any other PhD theses.

Debrecen, 08.08.2022 ……

signature of the candidate

Hereby I confirm that Tamás Milán Nagy candidate conducted his studies with my supervision within the K/5 Doctoral Program of the Doctoral School of Chemistry between 2017 and 2021. The independent studies and research work of the candidate significantly contributed to the results published in the thesis. I also declare that the results published in the thesis are not reported in any other theses. I support the acceptance of the thesis.

Debrecen, 08.08.2022 ……

signature of the supervisor

Understanding structure and dynamics with advanced NMR and *in silico* methods

Dissertation submitted in partial fulfilment of the requirements
for the doctoral (PhD) degree in Chemistry

Written by *Tamás Milán Nagy* (certified chemist)

Prepared in the framework of the Doctoral School of Chemistry of the University of Debrecen
(Chemistry and chemical biology of carbohydrates and heterocycles, program K/5)

Dissertation advisor: *Prof. Katalin E. Kövér*

The official opponents of the dissertation:

Dr. .....
Dr. .....

The evaluation committee:

chairperson:	Dr. .....
members:	Dr. .....
	Dr. .....
	Dr. .....
	Dr. .....

The date of the dissertation defense: 2022.

”

TUDÓS

Nézd, nézd, hogyan forr, nézd, miként ragyog,
Itt-ott tünékeny alakok mozognak,
Ezen meleg, e jól elzárt üvegben,
A vegyrokonság és ellenhatás
Mind összevág, és kényszerülve lesz
Engedni az anyag kivánatomnak. [...]

Mit gúnyolódtok, nem látjátok-é,
Egy szikra kell csak, és életre jő?

ÁDÁM

De azt a szikrát, azt honnan veszed?

TUDÓS

Csak egy lépés az, ami hátra van.

”

Madách Imre: Az ember tragédiája / The tragedy of man

Acknowledgements

This work would not have been possible without the devoted guidance of my supervisor, *Prof. Katalin E. Kövér,* who has been supporting me for almost a decade now.

I will never forget the first steps toward academic research with *Dr. István Timári,* who encouraged me to work in the NMR lab.

I would like to thank *Prof. Mónika Fuxreiter*, *Prof. Ole W. Sørensen*, and *Prof. Ferenc Erdődi* for the fruitful collaboration.

Many thanks to all the members of the *NMR facility* for the supportive atmosphere. *Prof. Gyula Batta, Dr. Krisztina Fehér* and *Dr. Ádám Fizil* for the inspiring and motivating discussions. *Dr. Tamás Gyöngyösi* for the pleasant hours spent together with liquid nitrogen and for the unforgettable conference experiences.

I am grateful for the Ph.D. grant provided by the *Richter Gedeon Talentum Foundation* and the financial support from the *New National Excellence Program by the Hungarian Government* and *the European Regional Development Fund (projects GINOP-2.3.2-15-2016-00008 and GINOP-2.3.2-15-2016-00044).*

I am thankful to *my friends* who understood me and showed that there is life outside the lab.

My family deserves endless gratitude. I appreciate the patience of my parents for my early scientific explorations and the occasional chemical explosions in the backyard. I am grateful to my brother and sisters for accompanying me on the journey and finally to my grandmother who fuelled my chemical and scientific curiosity.

Contents

Abbreviations

BANGO	small angle or ***β***-T**ANGO**
BIRD	**BI**linear **R**otation **D**ecoupling
BIG-BIRD	**B**iselective **I**ndependent **G**yrations **BIRD**
DOSY	**D**iffusion-**O**rdered **S**pectroscopy
HMBC	**H**eteronuclear **M**ultiple **B**ond **C**orrelation
HSQC	**H**eteronuclear **S**ingle **Q**uantum **C**oherence
H2OBC	**H**eteronuclear **2**-bond and **O**ne-**B**ond **C**orrelation
IDP	**I**ntrinsically **D**isordered **P**rotein
IDR	**I**ntrinsically **D**isordered **R**egion
MD	**M**olecular **D**ynamics
MEF2	**M**yocyte **E**nhancer **F**actor-2
NanoDSF	Nanoscale **D**ifferential **S**canning **F**luorometry
NMR	**N**uclear **M**agnetic **R**esonance
NOAH	**N**MR by **O**rdered **A**cquisition using $^{1}\boldsymbol{H}$-detection
NOE	**N**uclear **O**verhauser **E**ffect
NORD	**NO R**elaxation **D**elay
RCI	**R**andom **C**oil **I**ndex
RMSD	**R**oot **M**ean **S**quare **D**eviations
RMSF	**R**oot **M**ean **S**quare **F**luctuations
SASA	**S**olvent **A**ccessible **S**urface **A**rea
SEA XLOC	**S**eperate **E**cho and **A**nti Echo **X** nucleus for **lo**ng-range **c**ouplings
WT	**W**ild-**T**ype
ZQ	**Z**ero **Q**uantum
2BOB	**2**-**B**ond and **O**ne-**B**ond correlations
2Q	**D**ouble **Q**uantum

1. Introduction and objectives

Part I: characterizing MEF2D dynamical mutants

The biological function and activity of proteins are strongly related to their three-dimensional structure. However, due to the conformationally diverse nature of protein chains, these native structures are not static, they exhibit fluctuations over time and might change upon interaction with other molecules. It is crucial to consider the contribution of protein dynamics when elucidating biological mechanisms, especially in the case of the unique family of intrinsically disordered proteins, drawing significant scientific attention in the past decades.

The amino acid sequence is changing over time through evolution often resulting in altered structure and dynamics. Gene technology allows us to „accelerate" this evolution via targeted mutations to synthesize proteins with better properties. During my doctoral years, we were studying the MEF2D transcription factor with mutations designed to modify its dynamical nature.

The MEF2 protein family is playing role in complex cellular processes via gene regulation. In muscular and nervous tissues, they participate in cell proliferation, embryonic development and apoptosis. The MEF2D isoform is crucial in muscle cell differentiation processes, and it was shown that MEF2D knock-out mouses suffer from cardiac dysfunction. A short motif (286-292, *TEDHLDL*) called the β-domain is located in the disordered C-terminal region, a site for binding partner proteins and transcriptional co-regulators. This conserved site is regulated by alternative splicing and proved to be crucial for the maximum transcriptional activity via its role in transactivation.

In collaboration with the Fuxreiter group, we hypothesized that without a well-defined structure, the biological function of the MEF2D β-domain is mainly determined by dynamics. We have introduced directed mutations to perturb the dynamical profile of the β-domain and the flanking residues and we followed the corresponding biological responses. During the doctoral work, we planned a detailed study of the structure and dynamics of β-domain mutant MEF2D mini-proteins with promising preliminary biological responses and bioinformatical scores.

Part II: developing combined NMR experiments

NMR spectroscopy is a widely used, interdisciplinary technique and an invaluable tool for investigating biologically relevant compounds as well. The measurement techniques are being constantly developed with advancements in hardware (*e.g.*, larger magnets, cryoprobes), pulse programme design and data collection methods (*e.g.*, non-uniform sampling, NUS).

Our group is working on the development of NOAH type (NMR by Ordered Acquisition using ^{1}H-detection) NMR super sequences allowing the concatenation of different experiments with reduced measurement time and rich spectral information. During my Ph.D., our objective was to combine NMR experiments useful for the complete resonance assignment of small- and medium-sized molecules. Another goal was to significantly reduce the corresponding measurement time with efficient pulse sequence design. The main starting point was to reduce the relaxation time, a long delay component crucial for the magnetization recovery between the repetition of experiments. Secondly, we also aimed to completely eliminate the recovery delay with utilizing the Ernst-angle concept, unique excitation techniques and pulse sequence design strategy.

2. Theoretical framework

2.1 Proteins and their function

Although the origin of life is still unclear, it is known for decades that the genome encodes almost every biological property of an organism. This means a massive sequence of nucleic acids organized in double-stranded Deoxyribonucleic Acid (DNA), storing the genetic information in unique sequences or units called the genes. In 1958, Francis Crick published the central dogma of molecular biology, describing the flow of information at the molecular level. In simple terms, throughout the transcriptional process, DNA (Deoxyribonucleic Acid) is converted into RNA (Ribonucleic acid), and then the translation in the ribosomes finishes the gene expression manifesting in the production of proteins.

Stabilized by a special covalent bond, the peptide bond, long chains of connected amino acids assemble proteins. Naturally, there are 20 different kinds of these building blocks available, resulting in an almost infinite number of possible permutations. *E.g.*, the first discovered peptide hormone insulin is considered a small protein containing 51 amino acids. This size allows 20^{51} possible combinations of the residues showing the undeniable diversity of proteins only in terms of combinatorics. Through evolution, life benefits from this opportunity and gives a large variety of biological roles to these macromolecules: structural functions (collagen), enzymes (trypsin), transporters (hemoglobin), members of signaling pathways, hormones (oxytocin), antibodies of the immune system (immunoglobulin) are just a few to mention. Even they serve as regulators of their production during gene expression (transcription factors).

2.2 Structural and dynamical views

The classical structure-function paradigm states that the molecular structure of a protein strongly determines its biological role and activity in an organism. The genes indirectly dictate folding information encoded in the amino acid sequence; this way, the primary structure determines the behavior

and the role of a protein we know from Anfinsen[1]. Controlled by thermodynamical driving forces or by specific post-translational modifications, proteins go through conformational changes adopting a well-defined three-dimensional structure. This involves the formation of local order by secondary structures via repeating elements (alpha-helices, β-sheets, turns, coils). The next level is the tertiary structure, the three-dimensional arrangement of the mentioned building blocks with exact coordinates. Sometimes, assembly of several units of polypeptide chains occurs, forming a quaternary structure also. At the end of the folding process, they reach their native state, the final biologically active form. This is often a relatively rigid, well-defined globular state that is usually stabilized by intramolecular interactions, such as disulfide bonds, salt bridges, and hydrogen-bond networks.

Back in 1894, Fischer introduced the well-known „lock and key" hypothesis[2], elucidating the mechanisms of substrate binding. The binding pockets are located in specific parts of the protein structures, being small regions, motifs of amino acid residues that are providing a unique milieu for binding ligands, catalyzing reactions, translocating ions, or forming surfaces for the assembly of complexes. As a key only fits one type of lock, *i.e.* a substrate can only interact with a specific active site of a protein that possesses a static structure. This idea was revised later since *e.g.*, enzymes are not necessarily ligand-specific; they can catalyze the reactions of several different substrates. The induced-fit model[3] from 1958 states that a shape transformation of the binding pocket precedes the formation of the enzyme-substrate complex, making progression in the understanding of biological function.

The energy landscape is a surface associating the total energy of every possible configuration of a system. In the case of molecules, through the conformational fluctuations, proteins are constantly moving towards states with the lowest possible energy. The folding process drives the structure to the global minimum that could be represented by a funnel on the surface.

There are several experimental techniques to investigate the three-dimensional structure of proteins with different precision and resolution. Most of them are spectroscopic techniques, gaining data by interacting the sample with specific electromagnetic radiation. UV-spectroscopy utilizes the excitation of electron-states in protein chromophores (amide bond, aromatic residues).

Some conformational changes can be determined from the resulting absorption spectra. More sensitive is the IR spectroscopy being able to probe the prominent secondary structural motifs via the excitation of vibrational states of the peptide bonds. Exploiting the optical activity, chiroptical techniques (ECD, VCD, ROA) provide more intricate knowledge about the secondary and tertiary structure also. Diffraction techniques were first used to determine the molecular structure of proteins, with the leading role of X-ray diffraction (XRD). Reaching the end of 2020, 88 % of Protein Data Bank structures were elucidated by this method. With lower resolution, imaging the molecular shape, Small Angle X-ray Scattering (SAXS) is also gaining more attention. A cutting-edge resolution can be also achieved by Nuclear Magnetic Resonance (NMR) spectroscopy, explained in detail in *chapter 2.5*. Finally, regarding Electron Microscopy (EM) techniques, mainly cryo-EM in protein science is blooming in recent years, resulting in a Nobel Prize in Chemistry (2017).

The structural conception of protein molecules and their function dominated science for several decades. However, the advances in biochemical and structural determination techniques forced the scientific community to rebuild the available models persistently. The emerging number of the isolated and/or expressed proteins and their investigation at atomic level with increasing depth described new phenomena and mechanisms. Although the structural conception is fundamental, this static view became only a simplified model underestimating the role of dynamics. Protein motions arise in various time scales: in atomic depth, fs-ps vibrations, and librations, then in the ns-ms regime, side-chain rotations, and aromatic ring flipping. Proteins switch their conformations in the ms-s scale, driving large-scale transformation by folding and unfolding processes. A single rigid conformation is often insufficient to understand the molecular mechanisms of enzyme catalysis, transcription, molecular recognition, and other biological roles. A statistical conformational ensemble is the most precise and necessary model, considering local fluctuations, large-scale structural motions (*e.g.*, domain motions), and interconversion of these states.

Moreover, several studies highlight the role of internal dynamics in enzyme catalysis. Correlations were found between the 'promoting vibrations', 'coupled promoting motions[4] or *e.g.*, the specific movements of loops around

the active site and between the catalytic turnover rate[5]. Mutation of the residues responsible for these specific motions could alter the biological activity of an enzyme via dynamics.

2.3 The era of intrinsically disordered proteins (IDP)

Emerging numbers of contradicting observations appeared in the last decades with the discoveries of unorthodox proteins[6]. They are often characterized by the resistance to denaturation by chemicals[7] or heat treatment[8], by crystallization issues for X-ray crystallography or missing segments in the diffraction pattern[9], by strange SDS-PAGE behavior with ambiguity in molecular weight determination[10]. Spectroscopic techniques show no signs of or low propensity of secondary structures (*e.g.*, CD), and poor NMR chemical shift dispersion. These phenomena were previously reported when globular proteins lost their structure by unfolding after the treatment with denaturing agents.

The members of this protein family are maintaining partial or complete structural disorder and a high level of dynamics under physiological conditions. Surprisingly, being in an unusual native state, they possess essential biological functions playing roles in several areas of cell life. The literature is referring them as natively unfolded, intrinsically disordered (naturally they are unfolded which is coded in their primary amino acid sequence), partially folded (referring to more structured members with disordered regions), cloud, mobile, dancing (flexible, dynamical nature), malleable, chameleon (many possible conformations, promiscuity for a high number of ligands) proteins. The diverse nomenclature clearly shows a large universe of not yet fully discovered, unique compounds with the well-known hardship and uncertainty in their characterization and distinction from globular, structured proteins.

The novel partially or entirely unfolded members of the protein family cannot be treated with rigid models and with single biological function; one must consider them as a more complex structural ensemble with enhanced biological promiscuity. The energy landscapes of IDPs are significantly different from globular, folded proteins; the surface is relatively flat, lacking funnels and large global minimums (*Figure 1*). Fluctuating rapidly between

several states, IDPs are continually changing their conformations on a large scale due to the low energy barrier of transformations. Thus, the displacement of atomic positions and the backbone and sidechain dihedrals changes rapidly, often without fluctuating around any characteristic equilibrium values.

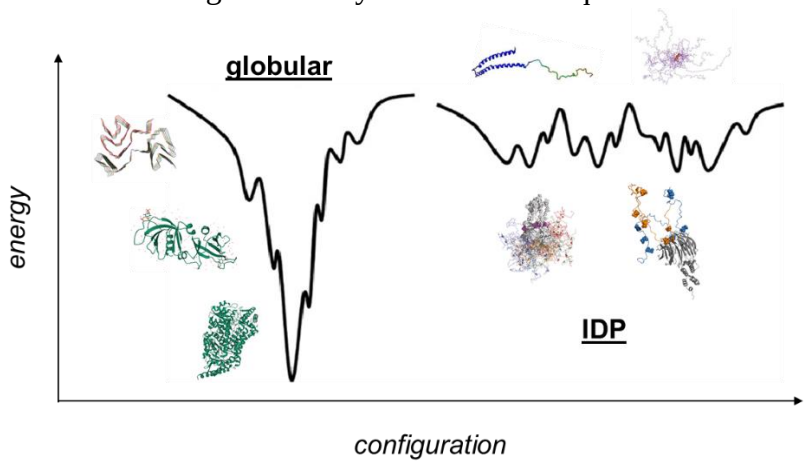
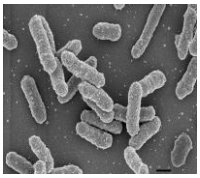
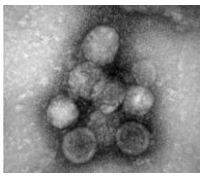


Figure 1. *Differences between the energy landscape of globular and disordered proteins, as shown in the simplified energy profile.*

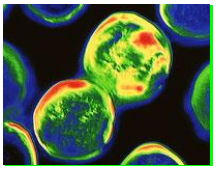
Structural properties are encoded in the amino acid sequence, thus the extent of fully or partially unfolded proteins can be estimated in the proteome. Bioinformatical sequence analysis methods are available that reliably predict disorder, based only on the amino acid sequence and composition. In 2005, in a ground-breaking study, Keith Dunker revealed a proteome analysis[11] showing that the occurrence of disorder is more common in eukaryotes. Comparing different species, the more advanced a life form is, the higher the occurrence of intrinsically disordered regions (defined as segments with more than 30 residues) (*Table 1*).



Yersinia pestis (5%)



HCoV-NL63 (7%)



Archaeoglobus fulgidus (6%)

Arabidopsis thaliana (23%)

Mus musculus (28 %)

Homo sapiens (32-51%)

Table 1. *Occurrence of disordered proteins in different species based on bioinformatic analysis*[11,12].

Due to the complexity of the disordered nature, the classification of IDPs is still having ambiguities. However, with strict definitions, it is accepted that IDPs (more than 30% of disordered residues) build up 32% of the human proteome, while the inclusion of structured proteins with intrinsically disordered regions (IDRs) adds up to 51%[12].

It is well-established that IDPs play diverse roles in the maintenance of cell life. Showing promiscuity for ligands, IDPs are suitable for fine-tuning systems in molecular interaction networks taking significant roles in signal transduction pathways and cell-cycle control[13]. Besides molecular recognition, they often participate in the formation of intermolecular self-assemblies (*e.g.* tau protein[14]). As the D^2 concept (disorder in disorders)[15] states that their malfunction could be the cause of several known pathological conditions. A more specific analysis from the Giansanti group reveals[12], that IDPs are mainly cancer-related (tumor-suppressors p53[16] and BRCA1[17] (breast-cancer) malfunction), on the other hand, IDPRs are rather involved in neurodegenerative (synucleinopathies like Alzheimer's, Parkinson's diseases) and cardiovascular illnesses (thrombin-hirudin complex[15]). Uyar and co-workers discovered that 22% of disease-related mutations are located[18] in the disordered regions.

The molecular explanation for the native disorder is hiding in the unique biophysical characteristics strongly related to unusual amino acid compositions.

Characteristic differences can be observed[19] when comparing the residual distribution between folded proteins and IDPs in the databanks. Specifically, in IDPs there is a high propensity for amino acids with polar or charged sidechains (Arg, Gln, Ser, Lys) and those with structure-breaking properties (Ala, Pro). A high net charge is also typical since the ionic sidechains often remain uncompensated. In contrast to the 'disorder promoting' amino acids, the bulky, hydrophobic (Ile, Val, Leu), especially aromatic residues (Trp, Phe, Tyr) are neglected, reducing the possibility for the formation of hydrophobic cores, and with that eliminating a significant driving force of folding. Since cysteines (Cys) are also rare in the sequence of IDPs, chains are unable to form structure-stabilizing disulfides bonds.

Although they naturally exist in disordered forms, the interaction with partners often induces the formation of a partially or fully folded state. Different binding modes are described by the Fuxreiter group; disorder-to-order (folding-upon-binding), disorder-to-disorder transition and 'fuzzy binding' (polymorphic, conditional folding, dynamic)[20]. The interactions with other molecules are often mediated by specific, non-complex sites, called short linear motifs (SliMs), consisting of 6-12 residues. The molecular environment, the context[21] has a significant contribution to the regulation and modulation of function, for instance, via the charge distribution, structural, conformational changes, chain dynamics varying over time or even allostery and phase-separation. The neighboring amino acids in the sequence (flanking region) with a critical role usually involve ± 20 residues around the SliMs.

2.4 The Myocyte Enhancer Factor-2 (MEF2) proteins and the isoform MEF2D

The MEF2 transcription factor family is playing various roles in cellular life. Abundant in many tissues like skeletal and cardiac muscles, the central nervous system, and bone cells, their functions involve differentiation, embryonic development, morphogenesis, cell death, proliferation and adaptive responses. Four different types have been found so far in vertebrates: MEF2A, MEF2B, MEF2C and MEF2D[22-25]. They interact with Myogenic Regulatory Factors (MRFs), responsible for regulating the early stages of myogenesis, thus

having a significant role in muscle cell differentiation. Loss-of-function studies revealed that due to cardiac dysfunctions, mice without MEF2A and MEF2C were not viable; on the contrary, MEF2D null mice lived, but with reported cardiac defects[26,27].

As members of the MADS-box transcriptional regulators, they share a highly conserved N-terminal region responsible for DNA binding, dimerization, and interaction with other factors. The protein core has a MADS-box (MCM1, agamous, deficiens, serum response factor) domain (residues 3-57) and an adjacent MEF2-specific site (residues 58-86) with well-known structural properties investigated with both X-ray and NMR techniques. While the different MEF2 members show high-level homology in the N-terminal, the MEF2 genes encode a large, carboxyl-terminal region (87-511) that hides domains crucial for transcriptional activation, as evidenced by the mutations of MEF2C[28]. This large C-terminal region shows a high-level of disorder based on IUPRED[29] sequence analysis (*Figure 2*) and NEXTPROT[30] data.

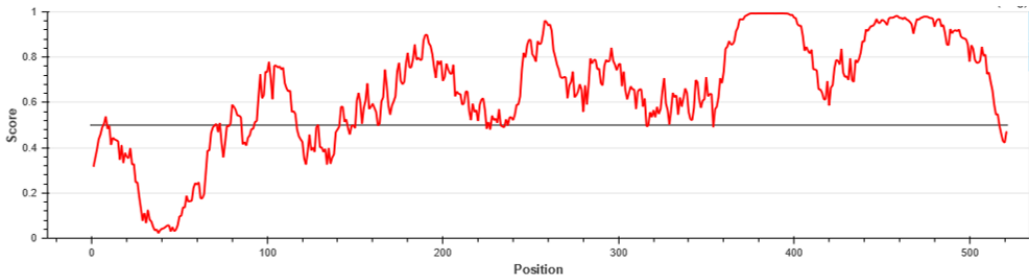


Figure 2. *The IUPRED analysis indicates that the MEF2D protein mainly possesses extended, disordered regions (scores above 0.5) aside from the DNA binding core and the MEF2-specific site (1-87).*

The first 3D structure of the MEF2D protein in interaction with an oligonucleotide (*Figure 3*) has been published recently[31]. Typical for IDPs and IDRs, the MEF2D C-terminal region is abundant in poly-proline and proline-glutamine sequential hotspots (regions 252-255, 451-456 and 365-404). A short, conserved β-domain was also identified in 2005 with an acidic character and a disordered nature (residues 286-292) and was described as a crucial site for transactivation[32]. The expression of this region is regulated by alternative

splicing with the activated exon inclusion during myocyte differentiation, resulting in more potent transactivator MEF2s[33].

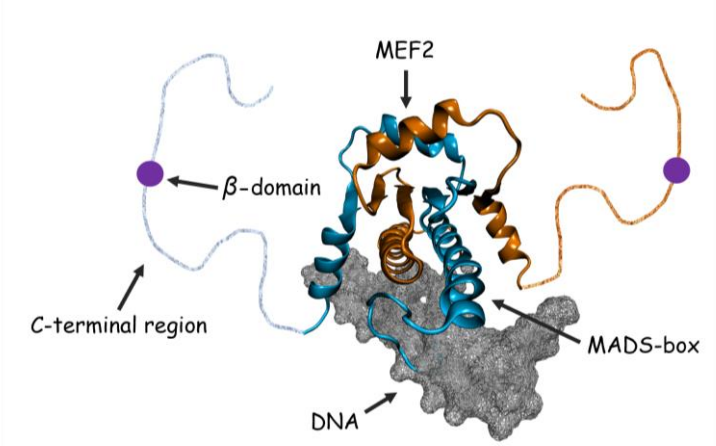


Figure 3. *The X-ray structure of the MEF2D-dimer (blue/orange) binding core domain with an oligonucleotide (grey)[34]. The MADS-box (3-57) and the MEF2-specific site (58-86) are responsible for DNA-binding and dimerization. The dotted lines represent the unknown, unstructured part of the protein (C-terminal region, 87-511), including the β-domain (286-292) with transactivation roles (purple). The amino-acid sequence of the β-domain and the flanking residues (265-301) can be found at the bottom.*

The available structural data and the preliminary sequential analysis indicate MEF2D as a promising model for studying the role of disordered, flexible sites in protein interactions and function. Prior to this work, in vitro, cell culture experiments from the Fuxreiter group confirmed the functional relevance of the *β*-domain. Lacking this short motif in the transfected MEF2D, the transcriptional activity was significantly reduced in C2C12 myoblasts compared to cells with the wild-type MEF2D (*Figure 4, blue vs. red bar).* Furthermore, biological responses could be triggered by point mutations targeting the *β*-domain with perturbing its dynamical behavior. Eight MEF2D variants (*v1-v8*) were transfected into HEK293 and C2C12 cells and as a result, modified transcriptional activities were observed again (*Figure 4*).

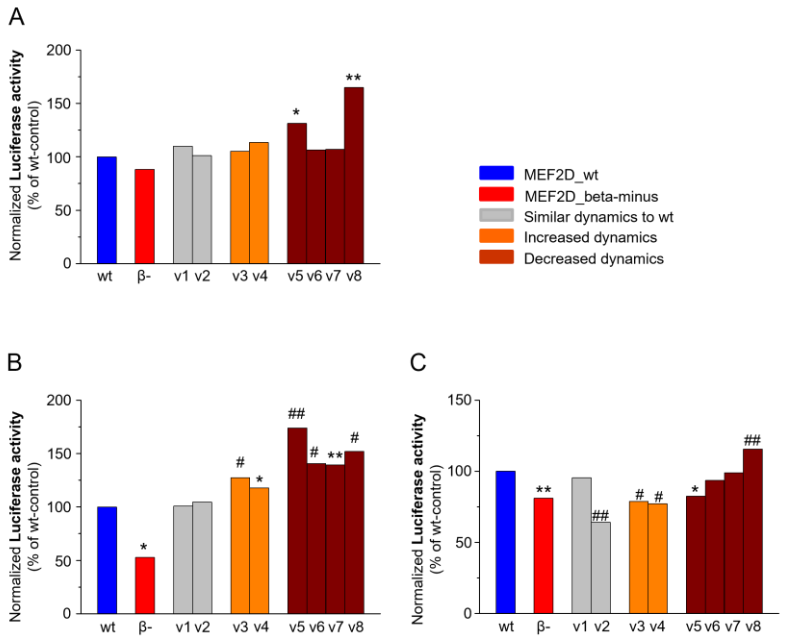


Figure 4. *Transcriptional activity of MEF2D variants with distinct dynamical properties in the β-domain. Normalized transcriptional activity parameter in HEK293 cells (A), non-differentiated C2C12 myoblasts (B), and differentiated C2C12 myotubes (C) expressing wild-type (blue) and different MEF2D variants with similar (grey), increased (orange), or decreased (brown) dynamical properties, and with the absence of β-domain (red). Experiments were repeated four times, and in each, Luciferase activity was measured in three sample replicates. (Significance *p<0.05; **p <0.01; #p<0.005; and ## p <0.001). Measurements were carried out by the Fuxreiter group.*

The myogenic differentiation program has also been altered as reported by monitoring the MEF2D target gene and regulatory factor expression levels. As an example, the expression of target gene APP (Amyloid Precursor Protein) and myogenic regulatory factor myogenin was enhanced at the early stage of differentiation in the case of mutants with decreased dynamics. Impact on the level of other MEF2D partner proteins was observed at a later phase of differentiation.

C2C12 cell differentiation was also followed by transmission light microscopy visualizing long, multinucleated myotubes (*Figure 5*). The fusion index (myocytes with two or more nuclei/all nuclei) showed the same trend in the transcriptional activity: no change was observed with mutants with identical dynamics to the *wt*, decreased activity with enhanced dynamics and increased differentiation with the decreased dynamics.

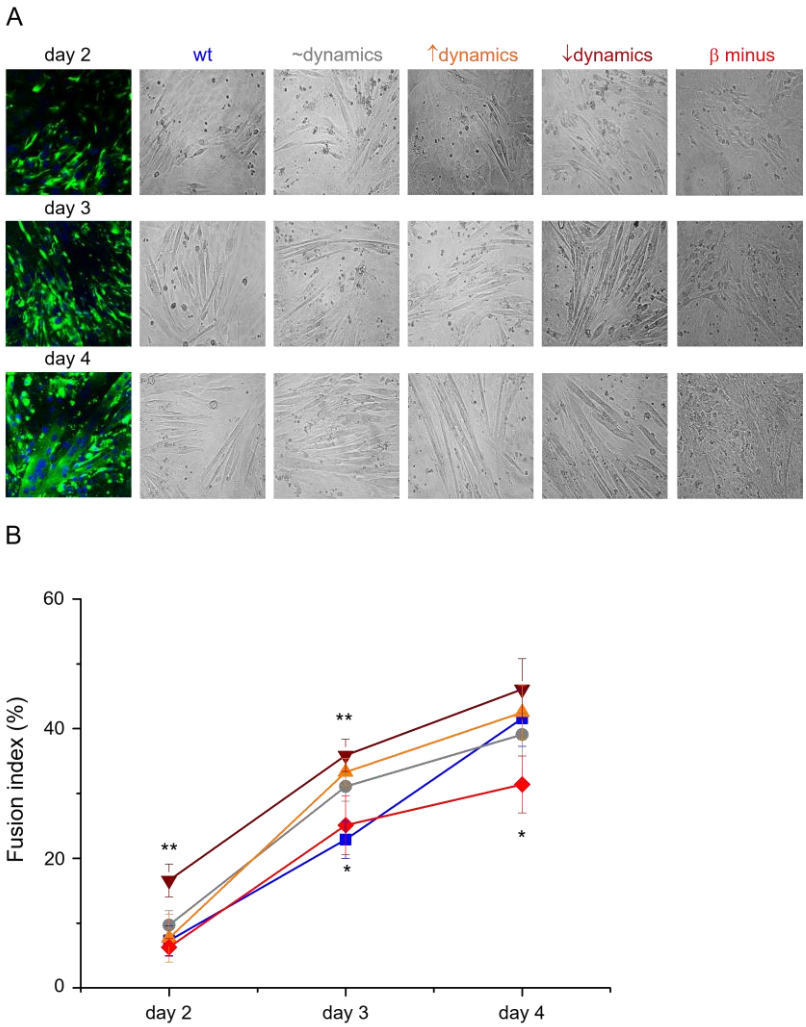


Figure 5. *Characterization of myotube differentiation in cell cultures overexpressing MEF2D variants with modified β-domain dynamics. Representative fluorescent and transmitted images from the C2C12 cultures during early differentiation (A). An increasing number of multinucleated, long myotubes correlates with more pronounced differentiation progress. The scale bar is 50 μm. (B) Averaged fusion index of C2C12 cells transfected with different MEF2D variants representing similar (grey), increased (orange) and decreased (brown) dynamical properties compared to the wt (blue). MEF2D without β-domain is also presented (red). Significance *p<0.05; **p <0.01. Measurements were carried out by the Fuxreiter group.*

2.5 Nuclear Magnetic Resonance (NMR) spectroscopy and its recent advancements

Nuclear Magnetic Resonance (NMR) Spectroscopy is a widespread technique available for chemistry, structural biology, material science, and other interdisciplinary fields for 75 years now. The capability to determine molecular properties at atomic resolution, including chemical constitution, stereochemistry, conformation, motions at different timescales, diffusion, molecular interactions, *etc.*, makes NMR an invaluable tool with never-ending efforts to constant development. However, molecular size will always be a limiting factor, but starting from the investigation of small organic molecules initially, small and compact proteins became observable with technology advancements after a few decades in the '80s. The emerging field of bioNMR can handle 10-40 kDa molecular weights, reliably elucidating the structure, dynamics, and interactions of proteins. 8% of the submitted protein structures in the PDB database were determined by NMR spectroscopy, being the second experimental method after XRD. Cutting-edge papers provide studies on large molecular assemblies like proteosomes[35], molecular engines[36] and measurements in living cells (in-vivo or in-cell NMR)[37]. To date, NMR spectroscopy is the leading tool for the high-resolution structural and dynamical investigation of intrinsically disordered proteins since X-ray diffraction suffers from missing electron densities of the dynamical regions.

Despite the progression, the NMR investigation of larger molecules (> 40 kDa) often ends up in signal overlapping and rapid relaxation resulting in signal broadening with missing peaks from the spectra. Moreover, due to the timescale of the investigated molecular phenomena, NMR parameters primarily represent only the most-populated states, leaving hidden conformers/forms. However, the missing information can be complemented by other experimental (MS, ESR, cryoEM, SAXS, SANS, ECD, VCD) or computational methods (MD, QM/MM, bioinformatics) to gain a complete understanding of protein structural landscape and the nature of internal motions and/or interactions.

NMR method development strategies are focused on boosting the quality of the acquired spectral data with increased sensitivity and resolution. Recent NMR experiments provide additional spectral information and

significant shortening of measurement times. Furthermore, with the simplified datasets, automated resonance assignment and analysis is an option saving invaluable time for research.

Advanced methods exploit the latest hardware capabilities (cryoprobe), the enhanced sampling protocol of the data, and optimized pulse sequence design. Utilizing multiple signal receivers, parallel acquisition NMR (PANSY[38]), the PANACEA[39] experiment and the interleaved acquisition techniques (UTOPIA NMR[40]) simultaneously detect multiple nuclei providing multiple correlation data necessary for complete resonance assignment and structure elucidation. Single receiver methods are also being developed in the past two decades, with the basic idea of magnetization sharing between experiments. The first attempts exploited different magnetization reservoirs through polarization transfer via the Nuclear Overhauser Effect (NOE[41]) or a short isotropic mixing scheme[42]. The rapid SOFAST (band-Selective Optimized Flip-Angle Short-Transient) and ASAP-HMQC (Acceleration by Sharing Adjacent Polarization) measurements utilize these strategies, respectively. The efficient magnetization share of these experiments allowed the reduction of the relaxation delay, the most time-consuming part of the pulse sequences.

Uniquely, the NOAH (NMR by Ordered Acquisition using ^{1}H-detection) concept[43] exploits different magnetization pools for each module. Using only one receiver, multiple experiments are nested into one NMR super sequence and detected sequentially with only one relaxation delay applied at the beginning of the sequences. The preceding experiment exhausts a magnetization pool and simultaneously prepares and preserves another one for the following module. As an example, the magnetization of protons attached to ^{13}C can be distinguished from the ones attached to ^{12}C or ^{15}N with appropriate isotope-selective spin manipulations and can be distributed between experiments. Combining the key 2D experiments in the appropriate order (HSQC, HMBC, COSY, NOESY *etc*.) up to 5, a complete set of data for structure determination of small- to medium-sized molecules can be recorded more rapidly than running the different experiments separately in the traditional way.

In general, NMR data recording (acquisition) involves the repeated detection of noise, creating a huge spectral dataset with minor valuable

information (resonance peaks). Further improvements can be achieved by reducing the number of the acquired data points by applying a random acquisition strategy. Although the obtained spectra are incomplete, there are various statistical methods to reconstruct them without losing the relevant resonance peaks of the studied molecule. Due to the high resolution that can be achieved in a short measurement time, this technique is getting more attention in the application and is termed the Non-Uniform Sampling (NUS) of the NMR data[44].

2.6 Details of the individual NMR experiments incorporated in the combined sequences

The classical HMBC experiment is used to extract correlations between protons and carbons (or other heteronuclei) separated by two, three or even four covalent bonds. Recent alternative methods are providing solutions for the inherent problems of HMBC experiment; SEA-XLOC (Separate Echo and Anti Echo X nucleus for long-range couplings), H2BC[45] (Heteronuclear 2 Bond Correlation) and its 2BOB and H2OBC variant.

H2BC is selective on the detection of two-bond heteronuclear correlations of protonated carbons, those that are often missing from HMBC spectra due to the frequently vanishingly small values of two-bond coupling constants. The correlations of bonds over two are eliminated, and many other features like pure absorption peak shapes, short pulse sequence length, and independence of long-range $^{1}H-^{13}C$ (or any other X nucleus) coupling constants are useful feature of the experiment.

SEA XLOC is a special HMBC-type experiment used to distinguish between heteronuclear two- and three-bond correlations for all carbon multiplicities. Comparing to other solutions (H2BC, HAT-HMBC[46], $^{2}J,^{3}J$-HMBC[47]), the SEA XLOC sequence is unique since it has no limitations for different spin systems and can be effectively utilized for quaternary carbons. The distinction between two- and three-bond correlations relies on the sign of the homonuclear and heteronuclear (*e.g.*, ^{1}H-^{13}C) coupling constants and the formation of multiplets due to the pertinent passive couplings of the different coherence states (ZQ, 2Q). In the case of three-bond correlations, the width of

ZQ multiplet is determined by the difference of the three-bond ($^3J_{HH}$) and the two-bond couplings ($^2J_{CH}$), while the width of 2Q multiplet is their sum. Since $^3J_{HH}$ couplings are always positive and $^2J_{CH}$ couplings are mostly negative, $^3J_{HH}$-$^2J_{CH}$ is larger than ^3J_H + $^2J_{CH}$ meaning that ZQ multiplets are wider than 2Q multiplets. Regarding two-bond correlations, the situation is the opposite, because ZQ multiplet width is determined by $^3J_{HH}$ - $^3J_{CH}$ and 2Q width is determined by $^3J_{HH}$ + $^3J_{CH}$, thus, ZQ peaks are narrower than 2Q. The 2BOB[48] (2-Bond and One-Bond correlations) sequence evolves the heteronuclear two-bond and one-bond correlations in one experiment. While HMBC lacks the information on the number of intervening bonds, 2BOB merges H2BC and HSQC type correlations in one spectrum or in two (or four) separate ones after proper editing and processing. To provide a direct connectivity map, H2OBC (Heteronuclear 2-bond and One-Bond Correlation) was developed as a simple modification of 2BOB. The distinction between one- and two-bond correlations is carried out with peak phase/magnitude editing, and so connectivities between protonated ^{13}C-s can be effortlessly identified. Consequently, if H2OBC is recorded for a compound, COSY experiments can be discarded since the H2OBC dataset contains heteronuclear one-bond correlation (HSQC) and, indirectly, ^1H connectivity (COSY) information as well.

The advanced experiments mentioned above provide additional spectral data compared to the classical and more conventional experiments for complete homo- and heteronuclear resonance assignment of molecules. Therefore, SEA XLOC and H2BC/H2OBC were chosen as constituting experimental modules of the developed combined pulse sequences presented in this work.

2.7 Molecular dynamics (MD) calculations of proteins

Besides different *in silico* conformational sampling techniques (*e.g.* grid searches, Monte Carlo simulation), molecular dynamics calculations (MD) are utilized to study chemical and biological systems at the atomic level predicting microscopical and macroscopical properties. Classical MD simulations are based on molecular mechanics using force fields to evaluate the energy of the system. Unlike the more accurate, but time and resource-consuming quantum chemical methods, certain approximations are applied in

force fields based calculations to efficiently handle more complex molecular systems like proteins. Firstly, the atoms are defined as spherical particles with nuclei merged with electrons (Born-Oppenheimer approximation) and their motions are calculated by solving the classical Newtonian equations. Secondly, molecular features, e.g, atom types, masses, charges, bond lengths as well as bond and dihedral angles are defined in parameter sets, called the force fields. During the simulation, the energy of the system is calculated using potential functions considering non-bonded (electrostatic and van der Waals) and bonded interactions between particles. In small, iterative steps, after the energies are obtained, the forces acting on atoms are determined, and knowing their velocity and acceleration, the positions are updated[49].

The large number of protein force fields (*e.g.*, *GROMOS*, *AMBER*[50], *OPLS3*, and *CHARMM*[51]*)* indicate that MD calculations are widely used to refine X-ray or NMR structures, to study conformational changes, dynamics, folding mechanisms or interactions with proteins or other molecules. Moreover, with the growing interest in intrinsically disordered proteins, the demand for atomistic simulations increased to supplement the limited information provided by the experimental methods. Although existing MD protocols for proteins can handle IDPs, they are often inaccurate in reproducing compactness or secondary structure content and experimental data (*e.g.*, NMR couplings). For more realistic computational outputs, fine-tuned force fields have been developed recently. For instance, in AMBER ff14SB[52] mainchain and sidechain potentials were modified, while CHARMM C36m[53] uses its TIP3P water model to produce more extended structures with a better fit to the available experimental and computational data on disordered proteins.

The MD simulation samples the potential energy surface by generating a conformational ensemble of the molecule with the initial structure usually obtained from experiments (X-ray, NMR). A well-established calculation protocol requires careful preparation to produce realistic results. First, a geometry optimization by energy minimization is required for the starting structure to find the closest local energy minimum. After the multistep process converges, short equilibration MD calculations are needed to set macroscopic thermodynamic properties. Once the system reaches stability in terms of *e.g.*, the total energy, density, temperature, structural fluctuations, long production

calculations can be started to generate a trajectory (molecular ensemble) that will be the source for analysis.

In certain cases, the conformational sampling represented by the canonical ensembles is not sufficient. There are several existing techniques for the enhanced sampling of the protein energy landscape: replica exchange, Gaussian accelerated and smoothed or scaled molecular dynamics simulations. The basic concepts behind these protocols are to run simulations with enhanced temperatures or to modify the energy potentials to flatten the conformational surface to overcome conformational energy barriers followed by the reconstruction of canonical ensembles at standard conditions. In scaled MD algorithms[54] an arbitrary weighing factor is applied, and after the production, the canonical ensemble is recovered with reweighing.

3. Methodology

3.1 NMR measurements

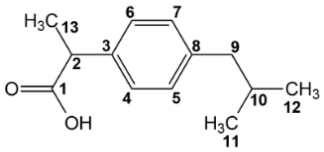
The NMR experiments were run on a Bruker Avance Neo 700 MHz spectrometer equipped with a Prodigy TCI cryoprobe. The spectra were processed and evaluated using the TopSpin 3.0 software. Details on the compounds, sample preparation (*Table 2*) and the experimental parameters (*Table 3*) are listed in the following tables.

Table 2. *Samples used in the NMR investigations.*

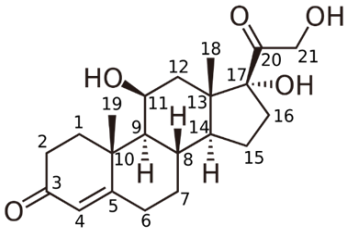
MEF2D NMR samples (37aa peptides)
wild-type *(wt):* SRKPDLRVITSQAGKGLMHHLTEDHLDLNNAQRLGVS, 4109 g/mol 2 mM in 520 µl H_2O + 30 µl D_2O PBS + DSS
variant no.3 *(var3):* SRKPDLRVITSQAGKGLMHHLTEDHLDKNNAQRAGVS, 4082 g/mol 2 mM in 520 µl H_2O + 30 µl D_2O PBS + DSS
variant no.4 *(var4):* SRKPDLRVITSQAGKGLMHHLTEDHLDNNNHQRLGTS, 4178 g/mol 2 mM in 520 µl H_2O + 30 µl D_2O PBS + DSS
variant no.8 *(var8*): SRKPDLRVITSQAGKGLMHHLTEDHLDLDNWQRLGVS, 4225 g/mol 2 mM in 520 µl H_2O + 30 µl D_2O PBS + DSS
PBS: Phosphate Buffer Solution, 20 mM Na_2HPO_4/NaH_2PO_4 100 mM NaCl, pH=6. DSS: Sodium 3-(trimethylsilyl)propane-1-sulfonate, 2 µl from 2mg/500µl stock solution.

NMR samples for BANGO, NORD experiments

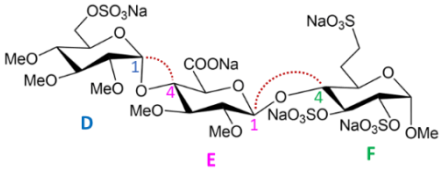
ibuprofen 206 g/mol



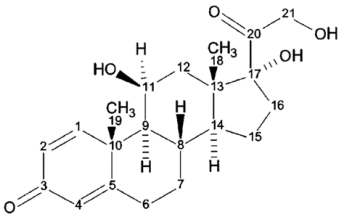
hydrocortisone 362 g/mol



trisaccharide: 1030 g/mol



prednisolone 360 g/mol



pentasaccharide: 1800 g/mol

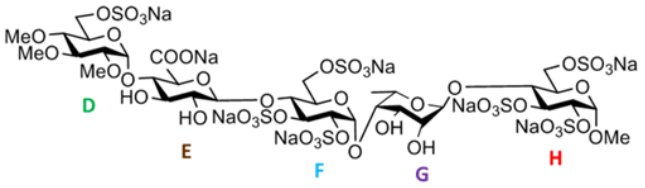


Table 3. *Summary of the experimental parameters typically used in NMR experiments.*

MEF2D		
resonance assignment		
^{1}H-^{1}H TOCSY (1)	**^{1}H-^{1}H ROESY (2)**	**^{1}H-^{13}C HSQC (3)**
ek_mlevgpph19_15N ns = 16 TD = 4096x800 d9 = 80 ms aq = 0.24 s ^{15}N-decoupling expt = 4 h	*ek_easyroesy_wg_15N* ns = 48 TD = 4096x800 p7 = 100 ms aq = 0.24 s (200 ms mixing) ^{15}N-decoupling expt = 21.5 h	*hsqcetgpsi2* ns = 16 TD = 2048x512 aq = 0.11 s expt = 4 h
^{1}H-^{13}C HSQC-TOCSY (4)	**^{1}H -^{15}N HSQC (5)**	**^{1}H -^{15}N HSQC-ROESY (6)**
hsqcetgpml ns = 120 TD = 2048x400 d9 = 80 ms aq = 0.12 s expt = 21.5 h	*fhsqcf3gpph* ns = 8 TD = 2048x64 aq = 0.12 s expt = 0.25 h	*ek_hsqcetf3gp_easyroesy* ns = 128 TD = 2048x64 p7 = 100 ms ROESY mixing = 200 ms aq = 0.12 s expt = 4 h
^{15}N-relaxation (T_1, T_2, hetNOE)		
^{1}H-^{15}N HSQC (T_1) (7)	**^{1}H-^{15}N HSQC (T_2) (8)**	**^{1}H-^{15}N HSQC (hetNOE) (9)**
hsqct1etf3gpsi3d.2.t2.be ns = 8 TD = 2048x64x12 d1(ms)= 10, 20, 40, 80, 120, 160, 200, 300, 400, 600, 1200, 2400 aq = 0.18 s expt = 8.5 h	*hsqct2etf3gpsi3d* ns = 8 TD = 2048x64x12 CPMG trains (ms): 17, 34, 51, 68, 85, 102, 119, 153, 204, 254 aq = 0.18 s expt = 8 h	*hsqcnoef3gpsi3d* ns = 80 TD = 2048x64x2 d1 = expt aq = 0.18 s expt = 23.5 h
Diffusion measurements, oligomerization		
^{1}H DOSY (10)	**^{1}H-^{13}C HSQC (11)**	
ledbpgppr2s ns = 128 TD = 16392x32 d1 = 3s d20 = 75 ms p30 = 2.1 ms gpz6 = 2-95% aq = 1 s expt = 4.5 h	*hsqcetgpsi2* ns = 32 TD = 2048x512 aq = 0.14 s expt = 7 h	

<table>
<tr><th colspan="3">Combined experiments (BANGO, NORD)</th></tr>
<tr><td>^{1}H-^{13}C BANGO {HMBC}-{H2OBC} (12)
(trisaccharide)</td><td>^{1}H-^{13}C BANGO {HMBC}-{H2OBC} (13)
(ibuprofen)</td><td>^{1}H-^{13}C BANGO {SEA XLOC}-{H2OBC} (14)
(ibuprofen)</td></tr>
<tr><td>noah2_hmbc_2BOB_ek.nus
BANGO (90°), ^{1}H 90°
Δ_{HMBC} = 83 ms,
T_{H2OBC} = 23 ms,
$TD_{HMBC/H2OBC}$ = 1024x256
ns = 1, NUS = 64 (25%)
d_1 = 1.7 s
aq = 0.14 s
expt = 5 min</td><td>ek_hmbc_bango_2BOB_in2
BANGO (90°), ^{1}H 90°
Δ = 83 ms,
T_{H2OBC} = 23 ms,
$TD_{HMBC/H2OBC}$ = 2048x512
ns = 1
d_1 = 1.7 s
aq = 0.17 s
expt = 11 min</td><td>ek_seaxloc_bango_2BOB_in2
BANGO (90°), ^{1}H 90°
$\Delta_{SEA\ XLOC}$ = 83 ms,
T_{H2OBC} = 23 ms
$TD_{SEA\ XLOC/H2OBC}$ = 2048x512
ns = 1
d_1 = 1.7 s
aq = 0.17 s
expt = 42 min</td></tr>
<tr><td colspan="2">^{1}H-^{13}C BANGO {SEA XLOC(ZQ)}-{SEA XLOC (2Q)}-{H2OBC}(15) (ibuprofen)</td><td>^{1}H-^{13}C BANGO {SEA XLOC}-{HMBC}-{H2OBC} (16)
(prednisolone)</td></tr>
<tr><td colspan="2">ek_seaxloc_2bango_2BOB_in
BANGO (β^{I}_{1}=150°, β^{I}_{2}= 90°)
$\Delta_{SEA\ XLOC}$ = 83 ms, T_{H2OBC} = 23 ms
ns = 16
d_1 = 1.6 s
TD = 2048x1536
aq = 0.17 s
expt = 6 h
adiabatic ^{13}C inversion pulse:
CAWURST-20</td><td><i>ek_seaxloc_hmbc_2bango_2BOB_in3</i>
BANGO (β^{I}_{1}=150°, β^{I}_{2}= 90°)
$\Delta_{SEA\ XLOC/HMBC}$ = 83 ms, T_{H2OBC} = 23 ms
ns = 8
d_1 = 1.6 s
TD = 2048x512
aq = 0.17 s
expt = 3 sh
adiabatic ^{13}C inversion pulse:
CAWURST-20</td></tr>
<tr><td>^{1}H-^{13}C NORD {HMBC}-{H2OBC} (17)
(trisaccharide)</td><td>^{1}H-^{13}C NORD {HMBC}-{H2OBC} (18)
(pentasaccharide)</td><td>^{1}H-^{13}C NORD {SEA XLOC}-{HMBC}-{2BOB} (19) (hydrocortisone)</td></tr>
<tr><td>ek_hmbc_2BOB_nord_adia2
BANGO (120°)
BIG-BIRD (ν=20°)
Δ_{HMBC} = 83 ms,
T_{H2OBC} = 23 ms
$TD_{HMBC/H2OBC}$= 1024x512
ns = 16
d_1 = 0 s
aq = 0.14 s
expt = 1.5 h
adiabatic ^{13}C inversion:
CAWURST-20</td><td><i>noah2_hmbc_2BOB_nord_ad2.nus</i>
BANGO (120°)
BIG-BIRD (ν=20°)
Δ_{HMBC} = 83 ms, T_{H2OBC} = 23ms
NUS = 64 (25%)
$TD_{HMBC/H2OBC}$ = 1024x256
$ns_{HMBC/H2OBC}$ = 4
d_1 = 0 s
aq = 0.14 s
expt = 6 min
adiabatic ^{13}C inversion pulse:
CAWURST-20</td><td><i>ek_seaxloc_hmbc_2BOB_nord_adia6</i>
BANGO (β^{I}_{1}=110°, β^{I}_{2} =120°)
BIG-BIRD (ν=20°)
$\Delta_{SEA\ X./HMBC}$ = 83 ms,
T_{2BOB} = 23 ms,
$TD_{SEA\ X./HMBC}$ = 2048x512
TD_{2BOB}= 2048x128
d_1 = 0 s
aq = 0.23 s
expt= 42 min
adiabatic ^{13}C inversion pulse:
CAWURST-20</td></tr>
<tr><td colspan="3">Parameters</td></tr>
<tr><td colspan="3">ns: number of scans, TD: time domain data points, resolution in F2xF1, d1: relaxation delay, d9: TOCSY or ROESY mixing time, d20: DOSY diffusion delay, p30: DOSY gradient pulse length, gpz6: DOSY gradient pulse strength range, expt: experimental time. The temperature was set to 298 K in all experiments.
NUS: number of NUS points and their percentage,
Δ, T: time of evolution under heteronuclear long-range and homonuclear proton-proton couplings
$\beta^{I}_{1,2}$: first and second BANGO angle
ν: BIG-BIRD angle.</td></tr>
</table>

3.2 The setup of MD calculations

The appropriate selection of the force field and solvent model is crucial in creating an MD simulation for a specific system. There are many options for proteins; however, there are just a few optimized for IDPs. The secondary structural elements are parametrized and involved in most classical biomolecular force fields, and with those naturally disordered chains tend to fold upon the simulation. To circumvent the formation of permanent, well-defined secondary structures and to avoid a funnel-like energy landscape, *CHARMM36M*[55] a modified *CHARMM36* force-field is one of the best available possibilities. The recommended water model is a modified three-point *TIP3P* parametrized for disordered systems. As reported recently[56], the small, disordered NPBP protein of the Ebola virus was studied successfully with this protocol.

The simulation systems for MEF2D were built with rigorous steps. Since the NMR-structural calculations resulted in ambiguous and unvalidated ensembles, the starting structures for MD were obtained by linear chains of the peptides created *in silico* using the LEaP module of AMBERTools 17. The linear chains were used as an input for the graphical user interface of the CHARMM project[57] to build the simulation box. The terminals were capped by non-charged, non-polarizable acetyl (ACE) and *N*-methyl (NME) groups. The protonation states of the residues were set according to the literature on MEF2D and to NMR spectroscopic data regarding His sidechains. Although the NMR samples were dissolved in PBS and 100 mM NaCl, to reduce the interaction with solvent ions, only the necessary number of neutralizing counterions (Na^+ and Cl^-) were added avoiding possible electrostatic artifacts. Ion placement was carried out randomly at 5 Å separation from each other and with Monte-Carlo method of the CHARMM GUI to find optimal positions. The solute was surrounded by a 12 Å water layer in a simulation box with a truncated octahedron geometry, saving invaluable computational time.

Before production runs, the linear geometries were carefully optimized and equilibrated. First, the minimization in 2000 steps was performed, switching from steepest descent to conjugate gradient algorithm after 1000 steps to remove atomic clashes. Then an equilibration was carried out using

constant total energy MD with an NVT ensemble for 50 ps while increasing temperature from 0 K to 50 K using Langevin dynamics as a thermostat and restraining backbone atoms by positional harmonic restraint with a 4 kcal/mol/Å force constant. The system was further relaxed in a 1 ns long MD simulation using an NPT ensemble with isotropic position coupling using a Berendsen barostat while increasing temperature from 50 K to 310 K during 500 ps and keeping it at this temperature for another 500 ps and thus allowing the density to stabilize at around 1 g/cm^3. During this second relaxation step, backbone atoms were restrained using a harmonic potential with 1 kcal/mol/Å force constant. Finally, by releasing the restraints, a third relaxation step was carried out for 1 ns with an NPT ensemble at 310 K and 1 bar. After the equilibration, production MD was carried out with the length of 100 ns individual runs using an NPT ensemble at 310 K and 1 bar. They were repeated three times from the same starting structure to reach maximal conformational sampling. Coordinate snapshots were saved for analysis every 10 ps. The translational center of mass motions was removed in a wrapping procedure every 1000 steps.

All production MD simulations were performed using the AMBER simulation engine version 16 implemented on GPUs[58]. The cut-off used for non-bonded interactions was 8 Å. The particle-mesh Ewald procedure[59] was used to describe long-range electrostatic interactions. The SHAKE algorithm[60] was applied to keep the bond lengths of hydrogen atoms rigid, allowing a time step of 2 fs to be used.

Studying the oligomerization of *var8*, replicas of classical MD calculations were repeated, but with larger simulation boxes containing 7 copies of the peptide. The same linear starting chains as in the monomer simulations were positioned equidistantly. Since oligomerization processes are much slower than intramolecular conformational rearrangements, there was a need for enhancing the conformational sampling of the simulation. With the last snapshot of the classical trajectories, replicas of scaled MD simulations were carried out scaling down the free energy surface with the arbitrary factor of 0.7.

After the collection of MD datasets, all trajectories were analyzed with the cpptraj[61] program package of AMBER and visualized in VMD[62]. Every

three replicas were checked for major errors, concatenated and further processed together.

Python software mdout_analyser.py was used to analyze the stability of the trajectories. The quality of the energy optimizations was assessed by validating the values of the final energy-gradients and checking the distance of the counterions from the protein chains before and after minimization. No unusual behavior of these parameters was observed. An equilibrium in the production simulations was reached as the total potential energy and temperature along the trajectory settled. Finally, the structural relaxation is shown by the root mean square deviation (RMSD) of atomic positions from the starting conformation over time (*Figure 6*).

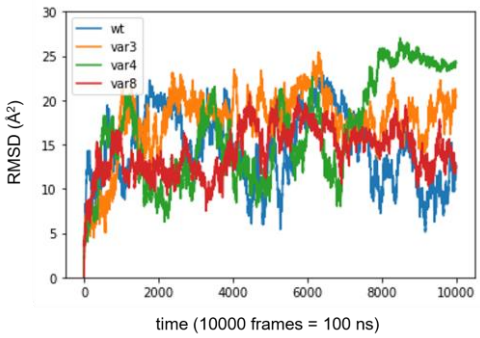


Figure 6. *Comparing the stability of MD simulations for each peptide by the RMSD of atomic positions with the reference of the starting structure over time in an example 100ns trajectory.*

Cluster analysis was carried out on the full trajectories using the DBSCAN method[63] implemented in cpptraj module of AMBER. The minimum number of conformations to form a cluster was 100 with a minimum distance of 4 Å between them. The *STRIDE*[64] algorithm was utilized to determine secondary structure elements in the ensembles and the representative conformers. Salt bridges were determined in VMD using the geometrical criterion of oxygen-nitrogen distance cut-off of 3.2 Å. For determining the Solvent Accessible Surface Area (SASA) a 1.4 Å radius extension was used. The oligomerization state was assessed by using distance-based criteria. Structures were considered oligomers when the distance between any of their atoms is less than a cutoff value, which was set to 4 Å. For the analysis python scripts described in a recent work of Samantray et. al [65] were utilized.

4. Results part I: characterizing MEF2D mutants

4.1 Designing MEF2D mutants and mini-proteins

From the set of previously investigated mutants, three were selected for further studies based on the most prominent biological responses. Transcription factors denoted as MEF2D *var3* and *var4* showed a decrease in biological activity, while *var8* significantly elevated the early-stage differentiation compared to the wild type. The mutations were deliberately designed to cause a divergence in the β-domain dynamical behavior and to disclose the molecular mechanisms behind the altered biological function.

The design strategy of MEF2D mutants was based on modifying the biophysical characteristics of the β-domain and its flanking region so that the transactivation is modulated through altered interaction patterns. Firstly, two variants were created with point mutations targeting the β-domain to modify dynamics via increased flexibility and enhanced propensity for the disorder.

Amino acids with charged or polar sidechains (*e.g., Lys, Arg, Glu, Asn*) are abundant in intrinsically disordered proteins; thus, introducing them will preferably induce the disorder[66]. The other way is to eliminate the hydrophobic residues (*e.g., Trp, Leu, Met*) since the formation of hydrophobic cores - buried in the inner part of globular proteins – is a main driving force for folding. With merging the two approaches, more mobility is expected from variants *var3* and *var4*. Mutation *Leu*292*Lys* (*var3*) switches a bulky, hydrophobic residue to a charged one. *Ala*295*Hys* and *Val*300*Thr* (*var4*) introduce new, polar amino acids in the sequence. *Leu*298*Ala* (*var3*) aims to reduce the hydrophobicity caused by the long, apolar side chain of *Leu* by truncating it to the short side chain of *Ala*. One variant *(var8)* was designed with opposite goals. In *var8,* the hydrophobic residues remained and *Trp*, an amino acid abundant in structured protein regions, was introduced to enhance order, compactness, and to decrease dynamics with its large, bulky aromatic sidechain.

The MEF2D transcription factor is a complex biomolecule consisting of 511 amino acid residues. Since molecular size and poor chemical shift dispersion of highly unstructured regions significantly limit the applicability of NMR techniques, thus a compromise has been made with the syntheses of

representative *β*-domain mini-proteins. Owing to the reduced size and complexity, both the experimental (NMR) and computational (MD) techniques benefit in terms of data production and analysis time.

Based on the biological studies, four mini-protein models with the most promising results were selected for detailed structural analysis. To preserve the native structural and dynamical characteristics of the conserved *TEDHLDL* *β*-domain, neighboring (flanking) residues were included in the synthetic peptides comprising 37 amino acids (*Figure 7*).

Figure 7. *The sequence of the wild type and mutant peptides with 37 amino acids. Residues (L) labeled with ^{15}N are in red, sites of mutations are shown in green. The position of the short β-domain is indicated with purple color in the corresponding region of the MEF2D transcription factor. wt: wild type, varx: variant x.*

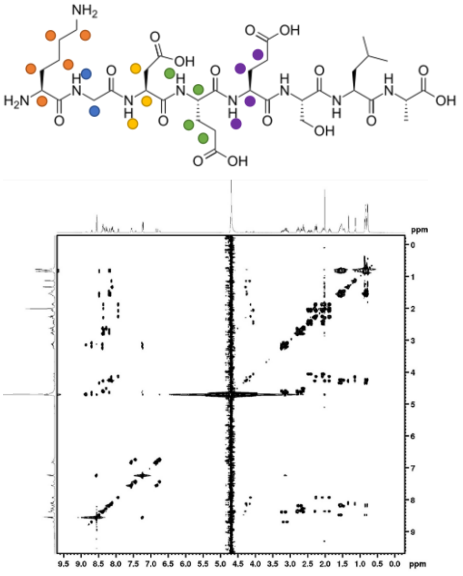
The structural effects of mutations were verified with a bioinformatical prediction method, in which the disorder score is calculated based on the protein sequence input. The IUPRED[29] software predicted the wild type and two variants (*var3*, *var4*) as highly disordered, while *var8* possessed scores under 0.5 indicating preferences for ordered structure in the C-terminal part (*Figure 8*). The global distinction between peptides is clear with the order of *var8* < *wt* < *var3* < *var4* regarding the level of disorder.

Figure 8. *IUPRED bioinformatical disorder prediction using the amino acid sequences of the MEF2D mini-proteins as an input.*

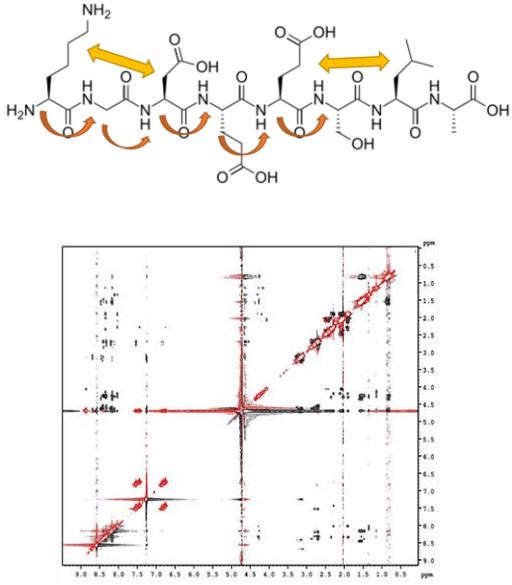
4.2 Complete NMR resonance assignment

The observed NMR data are originating from the individual magnetic resonances of the NMR active nuclei in the sample. The peaks of a complex spectrum are the 'footprints' of protein backbone and sidechain atoms, which must be identified prior to structural analysis. Since the studied MEF2D variants were ^{15}N isotopically labeled on the *Leu* amide nitrogens, the classical homonuclear NMR resonance assignment strategy was followed with a minor help of the ^{1}H-^{15}N correlations. To gather the required spectral data, several 2D homo and heteronuclear (^{13}C, ^{15}N) measurements were performed. All the TOCSY, ROESY and HSQC-type spectra used for the assignment with the corresponding structural information can be found in *Figure 9*.

^{1}H-^{1}H TOCSY



^{1}H-^{1}H ROESY



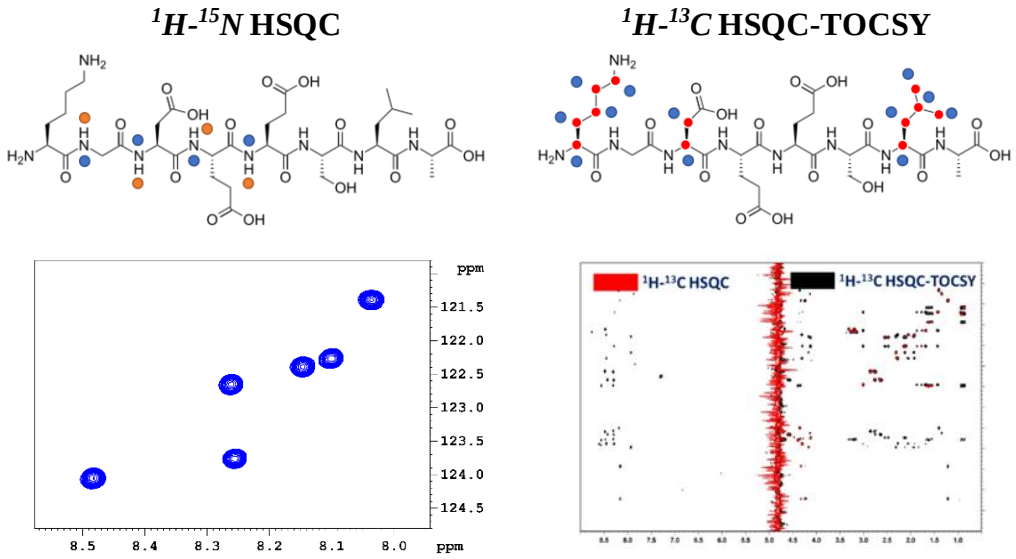


Figure 9. *Conventional 2D homo- and heteronuclear NMR measurements utilized in the resonance assignment. The example spectra were recorded on peptide var8 and the relevant connectivity information is provided and indicated by colored dots in the structures above. ^{1}H-^{1}H TOCSY: separation of ^{1}H spin systems labeled with different colors. ^{1}H-^{1}H ROESY: sequential information (orange arrows) and spatial proximities between residues (yellow arrows). ^{1}H-^{15}N HSQC: correlations between amide protons and ^{15}N-labeled nitrogens. ^{1}H-^{13}C HSQC-TOCSY: separation of ^{1}H (blue dots) spin systems and additional ^{13}C (red dots) chemical shift data. The overlayed red (HSQC) and black (HSQC-TOCSY) spectra demonstrate that the combined HSQC and TOCSY measurements provide more spectral information compared to the conventional HSQC experiment alone.*

Typical for proteins with extended disordered regions, low signal dispersion of ^{1}H and ^{13}C resonances was observed for all compounds (*Figure 10*)**.** *Var8* backbone ^{1}H resonances are slightly better resolved compared to the others, which can be an initial sign of more compactness in structure and relatively decreased dynamics.

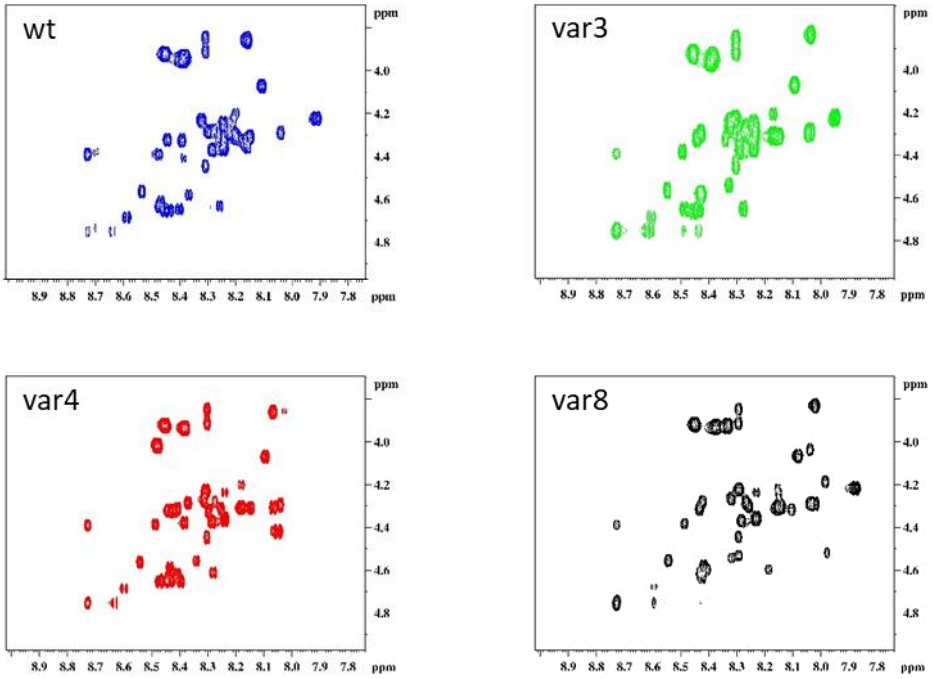


Figure 10. *Expanded NH-Hα backbone regions of the ^{1}H-^{1}H TOCSY spectra show the low signal dispersion characteristic for IDPs.*

In the TOCSY experiment, magnetization is transferred from the excited proton spins to their *J*-coupled partners, all along the proton-coupled spin system of an amino acid residue. The efficiency of the magnetization transfer depends on the pertinent proton-proton coupling constants and the duration of TOCSY mixing time. In protein/peptide NMR, proton resonances of the specific amino acid residues can be distinguished and separated based on the characteristic TOCSY peak patterns.

The ROESY experiment gives rise to cross-peaks in the corresponding 2D spectrum indicating spatial proximity (< 5 Å) between two spins. The amino acid sequence could be determined via the "walk-through" of $N_H(i)$-$H_\alpha(i-1)$ and/or $N_H(i)$-$N_H(i-1)$ ROE peaks between neighboring residues. In addition, distance constraints for structure calculation can be obtained from the strong, medium- and long-range ROE correlations.

If the proteins are labeled with ^{15}N, the ^{15}N-^{1}H HSQC correlation map could be a great help to identify the fingerprints of residues (N_H region) providing generally more well-resolved peaks than the ^{1}H-^{1}H correlation spectra. Additional information can be extracted from the combined experiments *e.g.*, ^{1}H-^{13}C HSQC-TOCSY and ^{1}H-^{15}N HSQC-TOCSY and -ROESY providing TOCSY or ROESY correlations at the resolution of heteronuclei, ^{13}C and ^{15}N, respectively. In overall, the NMR assignment of the studied peptide variants was carried out on both the backbone and sidechain resonances relevant for structural analysis, covering about 70-80% of the total number of proton (^{1}H) and 65-70% of the carbon (^{13}C) atoms.

4.3 Translational diffusion measurements (DOSY NMR)

The shape (compactness) and size (molecular weight) of molecules determine their motions in the media. If a protein is globular with a hydrodynamic radius at the minimum, its translational diffusion is faster than that of an unfolded, or partially folded protein of the same size possessing more extended structure and experiencing increased resistance with solvent molecules. The diffusion coefficients measured in DOSY NMR experiment mainly depend on the molecular mass, but in the case of proteins with the same weight, the differences in the molecular compactness can be assessed. In other words, if there is a difference between the diffusion coefficients of molecules of the same molecular mass, it could be ascribed to altered diffusion speed due to different molecular shape resulting in different hydrodynamic radius.

In our DOSY measurements (*Figure 11*) with peptides of 0.5 mM concentration, *var8* molecules diffused significantly faster ($2.95*10^{-10}$ m^2/s) compared to the others ($1.51\text{-}1.56*10^{-10}$ m^2/s). Since the studied peptides have almost identical molecular weights, the observed difference in the diffusion coefficients suggests a more compact structure of *var8* resulting in its faster movement in the media.

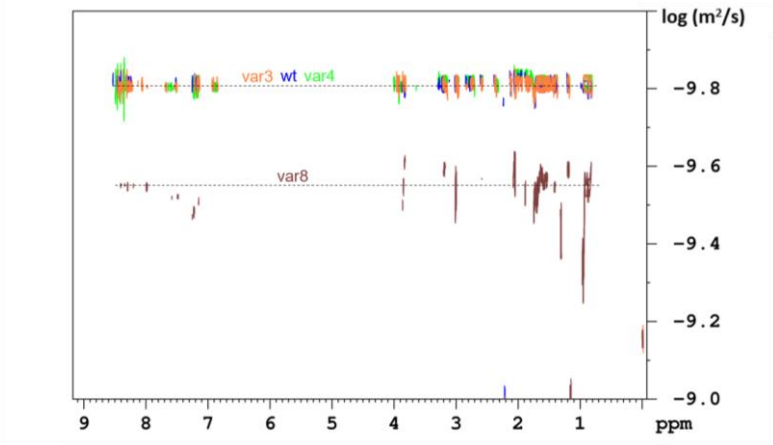


Figure 11. *Overlayed DOSY plots of 0.5 mM peptide solutions. Here, the 1D ^{1}H spectra (x axis, ppm) are plotted against the diffusion coefficients at a logarithmic scale (y axis, log(m^2/s) giving the diffusion fronts of the peptide variants. The values of the diffusion constants (D) were calculated by an exponential fit of the signal intensity decays obtained with increasing gradient strength: wt (blue, $D = 1.54*10^{-10} \pm 2.61*10^{-12}$ m^2/s), var3 (orange, $D = 1.56*10^{-10} \pm 3.47*10^{-13}$ m^2/s), var4 (green, $D = 1.51*10^{-10} \pm 1.75*10^{-12}$ m^2/s), and var8 (brown, $D = 2.95*10^{-10} \pm 1.15*10^{-10}$ m^2/s). The gradient constant for calculations was calibrated with lysozyme.*

4.4 Backbone dynamics from NMR chemical shifts

Chemical shifts are sensitive to the conformation of the protein chain. The C_α, C_β, H_α, H_β and amide H_N nuclei experience different magnetic shielding in alpha-helical, beta-sheet or in other folded structural environments resulting in characteristic resonance frequencies.

The chemical shifts corresponding to a completely unfolded chain are referred as the random-coil values. They can be calculated by several methods taking only the amino acid sequence as an input. The comparison with the experimentally measured shift values gives invaluable information about the structural and dynamical properties of the protein backbone. Regarding IDPs, the deviations from the random-coil values (secondary chemical shifts, SCS) provides a well-established way to estimate the level of disorder or to prove the

existence of transient secondary structures, local folds in certain regions of the protein chain.

The corresponding random-coil values were calculated with the Poulsen-method[67] for the C_α resonances of MEF2D peptides. *Figure 12* shows, that *var3* and *var4* shifts are closer to the random-coil values than those of *wt* and *var8*. The global level of disorder is *wt* < *var8* < *var3* < *var4* based on the cumulative average of SCS values.

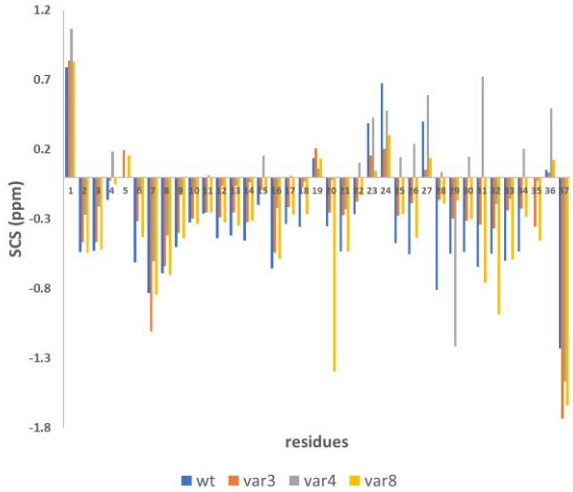


Figure 12. *Cαsecondary chemical shifts. The average deviations from the random coil values are: wt (0.47 ppm), var3 (0.34 ppm), var4 (0.29 ppm), var8 (0.43 ppm).*

The Random Coil Index (RCI) provides a more precise estimation for backbone dynamics[68]. The method uses only NMR chemical shifts with corrections (*e.g.*, smoothing the secondary shifts with neighboring residue corrections, scaling to reach numerical consistency). Once the residue indices are calculated, local dynamical preferences can be estimated. High mobility is assumed in the case of higher values, on the other side, low values suggest structural rigidity. The protocol takes the backbone chemical shifts of C_α, C_β, CO (carbonyl) carbon atoms, H_α protons, and amide nitrogens (*N*) as an input. Since the MEF2D peptides were ^{15}N-labeled only in the Leu residues and the CO shifts could be just partially assigned due to poor signal dispersion, only the C_α, C_β, and H_α resonances were used for calculations.

To compare the results with molecular dynamics (MD) simulations, the RCI values were converted to order parameters (S^2) describing the amide-bond vector motion in the molecular frame. The S^2 values are also in the range of 0-1, but as opposed to RCI, 1 means total rigidity.

The MD calculations of MEF2D β-domain proteins resulted in a broad range of order parameter values varying between 0.2-0.8 (*Figure 13*). Along the whole chain of the peptides, the disorder and flexibility are confirmed with S^2 values residing below 0.7. The global S^2 difference between *wt*, *var3* and *var4* based on the average values (0.43, 0.45, and 0.46, respectively) is within the error limit. A small enhancement of rigidity for *var8* (0.50) is introduced, compared to the *wt* (+ 0.07) on average and at the point of mutations. Namely, N29D and A31W show a significant increase in S^2 by + 0.15 and + 0.12, compared to the corresponding residues in the *wt*. L28K with + 0.08 enhancement in *var3* as well as L28N with + 0.11 in *var4* indicate a similar but slightly less pronounced increase of rigidity in the respective region. On the other hand, V36T in the C-tail of *var4* caused increased dynamics (-0.07).

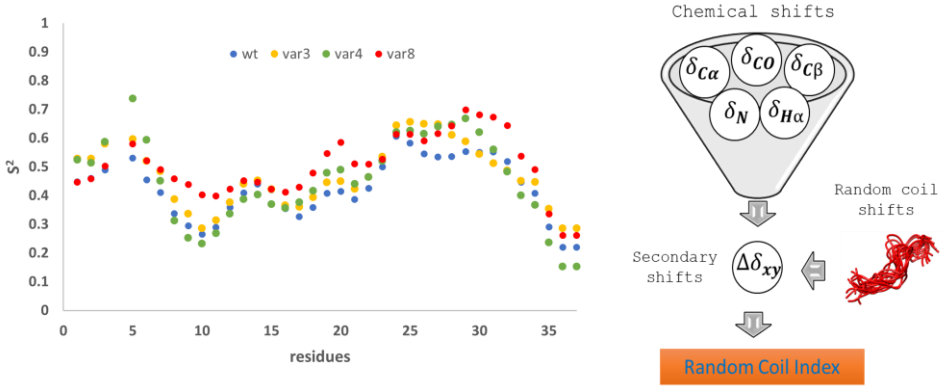


Figure 13. *Residue-wise S^2 order parameters (left) calculated from NMR chemical shift based RCI data (right). Averages calculated for all residues are 0.43, 0.45, 0.46, and 0.50 (wt, var3, var4, and var8), respectively.*

4.4 Probing dynamics with ^{15}N NMR-relaxation data

Usually, proteins are expressed in ^{13}C and ^{15}N containing media for NMR measurements. The double-labeled samples make the full and unambiguous NMR characterization feasible and at the same time provide extra dynamical information utilizing the relaxation properties of the heteronuclei. In the case of MEF2D models, only partial, selective ^{15}N-labeling of Leu residues was available, which allowed the characterization of molecular motions in the ps-ns regime only at the labeled sites using ^{15}N NMR relaxation measurements.

However, as the ^{1}H-^{15}N HSQC experiments reported, correlations corresponding to Leu284 of the β-domain were missing in the *wt* and *var4* spectra. Since the quality report and later the NMR structure elucidation confirmed the proper constitution of the synthesized peptides, peaks could be possibly vanished due to line broadening. This phenomenon is most likely caused by an intermediate chemical exchange between several conformational states existing in equilibrium, which is a common feature of disordered proteins. The exchange rate can be shifted with modifying the temperature or changing pH of the buffer solution, however, in the case of *wt* and *var4* models no new ^{1}H-^{15}N correlations could be observed after several attempts.

Firstly, the ^{15}N longitudinal (T_1) and transversal (T_2) relaxation times (or rates R_1, R_2, respectively) as well as heteronuclear NOEs (hetNOE) have been determined with well-established experiments[69]. Comparing the ratio of ^{15}N relaxation rates (R_2/R_1) significant differences are observed between *var8* and the other models (*Figure 14A*). While the latter three have almost identical R_2/R_1, *var8* shows significantly enhanced values. The faster R_2 relaxation might be explained by an increase in the global correlation time of the observed species due to their aggregation or oligomerization and also to the intermediate timescale of these equilibrium processes. This will be discussed in-depth in the next chapter.

The Lipari-Szabó interpretation of the relaxation data[70] normally gives further insights into the local protein backbone dynamics. However, it is not applicable in the case of IDPs, where the timescales of the internal motions and the global molecular rotations are in the same regime and so cannot be separated from each other. Instead, in this special case, the reduced spectral

density mapping[69] can be used to probe the high-frequency motions in the ps-ns timescale.

The dynamical properties of the *β*-domain and the adjacent flanking regions were characterized utilizing the *^{15}N*-labeling of *Leu* residues in the peptide models. The *^{15}N* T_1, T_2 and hetNOE data of MEF2D variants were processed to calculate the value of reduced spectral density function at proton (*^{1}H: $J(0.87\omega_H)$*), nitrogen (*^{15}N:* J(N)) and zero (J(0)) frequencies. All peptides show a high propensity for enhanced flexibility/dynamics with $J(0.87\omega_H)$ values between $2*10^{-11}$ and $4*10^{-11}$ ns/rad with only minor deviations at the different *Leu* positions (*Figure 14B*). Only the C-terminal region shows larger differences between *wt* and *var8*, which is in agreement with the expected reduced backbone dynamics of *var8*. Unfortunately, due to the missing (exchange broadened) peak of *Leu21* in *wt* and *var4* the complete analysis of data could not be achieved. The Lefèvre plot[69] (J(ωN) vs. J(0), *Figure 14C*) is a visual interpretation of relaxation data indicating dominant fast internal motions of the *wt* and the variants. Additionally, a significant exchange contribution is evidenced for *var8*.

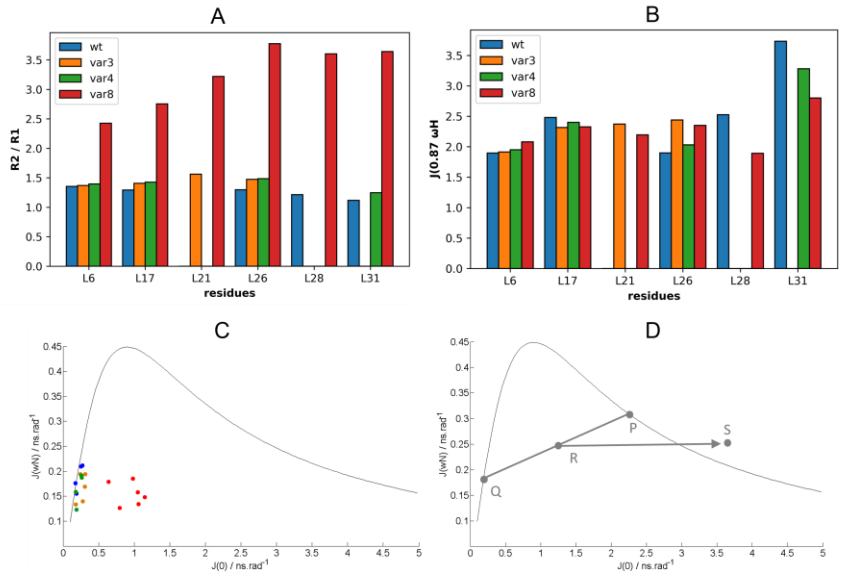


Figure 14. *Ratios of R_2/R_1 relaxation rates (A). Values of Reduced Spectral Density function at proton frequency $J(0.87\omega_H)$ at the positions of the ^{15}N-labeled Leu residues (B).) J(ωN) vs. J(0) (Lefèvre plot, C) and its interpretation (D). Where the curve is representing all possible cases of single-motion dynamics. Point Q: fast internal motions are dominant. Point P: only the isotropic tumbling determines the motions. Ideally, experimental values are expected between Q and P (e.g., R). If there is an exchange contribution to the relaxation, the measured values are shifted towards point S.*

4.5 Investigating aggregation/oligomerization properties

Particular attention was paid to the characterization of the oligomerization properties of the MEF2D peptides under the experimental conditions of previous NMR measurements. In the samples, neither visible physical changes such as precipitation nor other invisible sample changes that could cause the alteration of spectral data over time (chemical shifts, intensity) were observed. The protein solutions remained stable for a long time, usually for weeks and were often reused after lyophilization. After the suspicion of the presence of protein aggregates or oligomers in *var8* samples, examinations were performed by modifying conditions, like concentration and solvent types for all variants.

Firstly, DOSY measurements carried out in the range of 2.0 mM – 0.5 mM showed that *var8* diffusion constants had a concentration dependence with peptide species diffusing faster at lower concentrations (*Figure 15*). Unfortunately, only weak signals could be acquired for the 0.25 mM samples which were unsuitable for analysis, and no more solid peptides were available to test higher concentrations. Nevertheless, the concentration range between 0.5 – 2.0 mM was enough to shift the equilibrium between different forms of *var8*. The *wt* and *var3*, *var4* variants, on the other hand, did not show changes in diffusional behavior in that concentration range. Thus, it is very likely, that in contrast to *var8*, they did not form oligomers under the studied conditions.

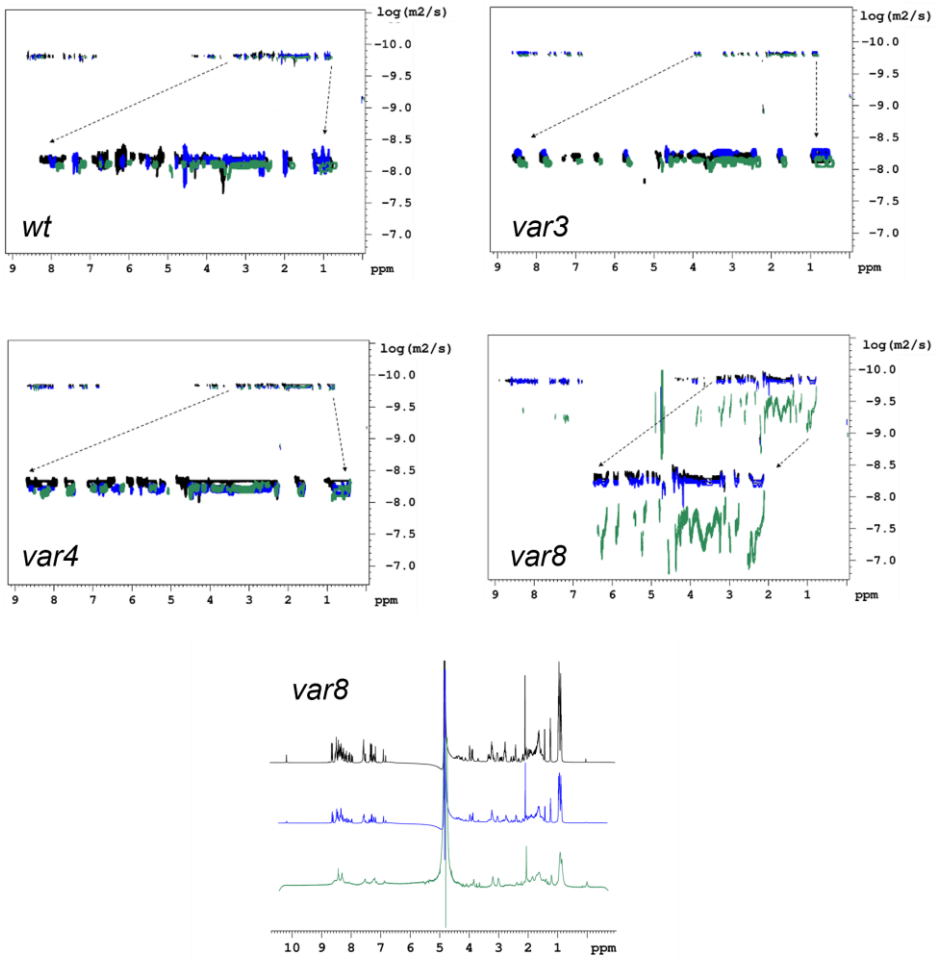


Figure 15. *Measurements with decreasing peptide concentrations are presented with colors in overlayed DOSY maps and in ^{1}H NMR spectra of var8 (2 mM: black, 1 mM: blue, 0.5 mM: green). The ^{1}H NMR shows signal broadening in 1 mM and 0.5 mM. The equilibrium between different oligomeric states of var8 is perturbed by the dilution leading to a change in the size distribution detectable in the average translational diffusion rates and peak widths by NMR.*

Interestingly, the freshly dissolved *var8* peptides with the lowest examined concentration of 0.5 mM first presented the same diffusion constant as those with higher concentrations. After a few days (usually around 6-7 days) of sample rest, the new DOSY spectra changed indicating significantly faster

diffusion, signal broadening and decreased signal-to-noise ratio. This state persisted in later measurements since the sample reached its equilibrium over time with a full or partial dissociation of oligomers but still in exchange with each other.

Sample concentrations were also modified by a serial dilution from 2 mM to 1 mM and then to 0.5 mM with the same phosphate buffer. While *wt*, *var3* and *var4* presented no or only minor spectral and diffusional changes, significant signal broadening occurred in the ^{1}H NMR spectrum of *var8*. Unfortunately, the poor spectral quality prevented the diffusion constant calculation. However, it is shown again, that different species of *var8* exist at 0.5 mM concentration, and they are exchanging with intermediate speed on the NMR time scale resulting in broad lines. The fact that the NMR observables can be modified by dilution is the evidence of oligomerization in the case of *var8*. The repeated experiments also pointed out that after dilution, a longer time (at least a day) was required to reach the equilibrium state.

Surprisingly, by replacing H_2O with D_2O in the PBS buffer, the *var8* samples remained stable over time in all concentrations (*Figure 16 A,B*). It is possible that the interaction with different solvent molecules has stabilized the oligomerization state of *var8*.

To assess the molecular compactness of the monomers, the comparison of the diffusion coefficients can be only carried out in the 0.5 mM H_2O solutions, since it was proved, that *var8* exists in a mixed oligomeric form at higher concentrations and in D_2O PBS (*Figure 16 C, D*).

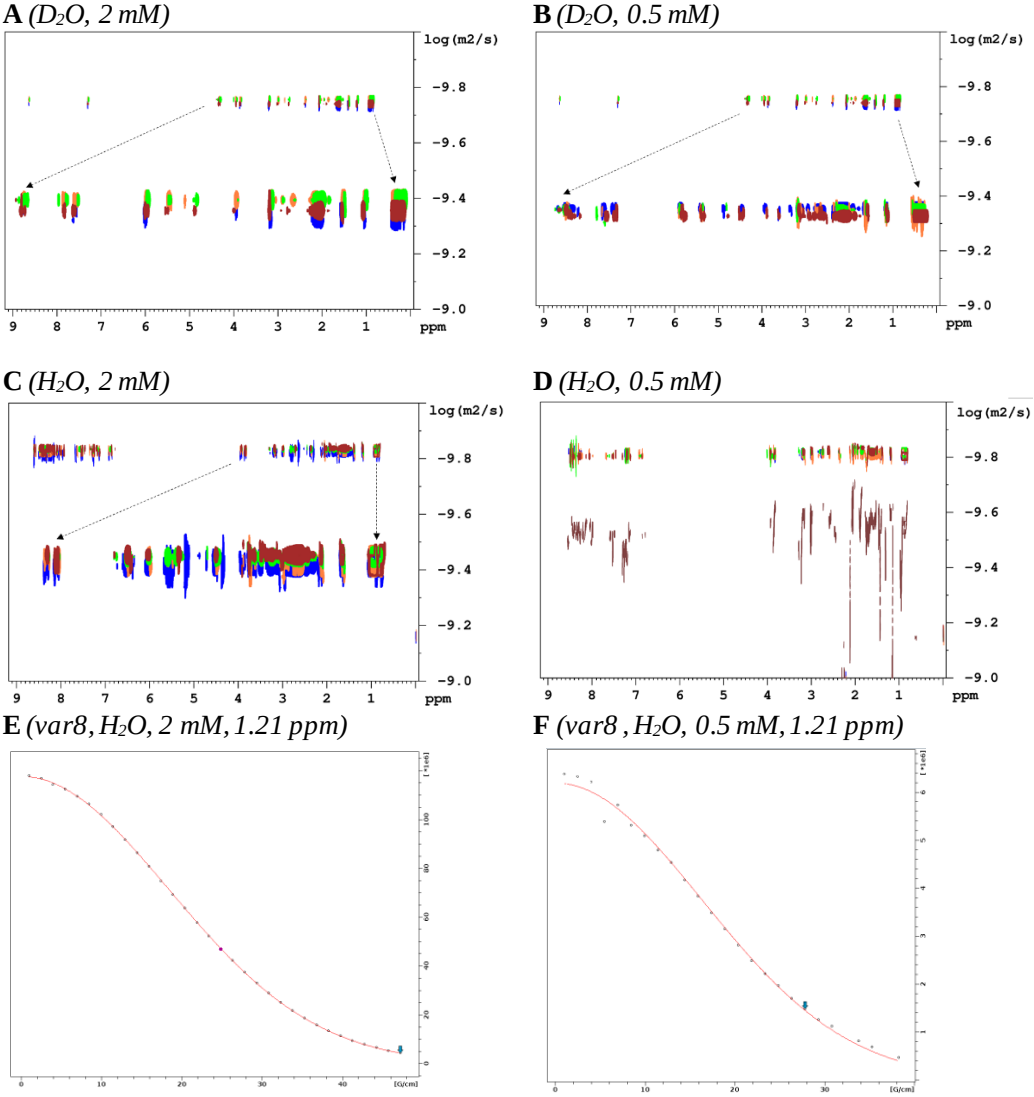


Figure 16. *Overlayed DOSY plots of 2 mM (A) and 0.5 mM (B) of peptides in D_2O and H_2O solutions (C,D). The aliphatic regions enlarged for better visualization in D shows the oligomer dissociation of var8. The diffusion constants were calculated by an exponential fit of the signal intensity decays obtained with increasing gradient strength in H_2O measurements: wt (blue, $D_{2mM} = 1.56*10^{-10}$ m^2/s, $D_{0.5mM} = 1.49*10^{-10}$ m^2/s), var3 (orange, $D_{2mM} = 1.49*10^{-10}$ m^2/s, $D_{0.5mM} = 1.56*10^{-10}$ m^2/s), var4 (green, $D_{2mM} = 1.49*10^{-10}$ m^2/s, $D_{0.5mM} = 1.50*10^{-10}$ m^2/s) var8 (brown, $D_{2mM} = 1.46*10^{-10}$ m^2/s, $D_{0.5mM} = 2.95*10^{-10}$ m^2/s). The quality of the exponential fit was good in all cases (e.g., E, var8, 2 mM) except in 0.5 mM var8 (F) where signal broadening due to chemical exchange enhanced the error of the analysis.*

There are several methods available to estimate the molecular weight of a protein compound from NMR-based diffusion coefficients. If only one oligomeric form exists in the solution, the calculated molecular weight is simply an integral multiple of the monomer. However, often several different forms exist in parallel, interconverting between each other at different time scales. If the rate of conversion is slow compared to the NMR "diffusion time scale", the forms can be separated and identified via DOSY experiments based on different diffusion coefficients. If the rate of conversion is fast, a population-weighted averaged diffusion constant is measured.

The Stokes-Einstein equation establishes a connection between the self-diffusion coefficient and the molecular mass (*equation 1*).

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

where k_B: Boltzmann-constant, T: temperature, η: viscosity of the medium, r_H: hydrodynamic radius of the molecule, F: a form factor, ρ: effective density, N_A: Avogadro number, M: molecular weight.

This model assumes a rigid spherical molecule, which is a reasonable assumption for globular proteins, but intrinsically disordered proteins have diverse extended and time-dependent molecular shape. Their behavior is closer to those of the denatured proteins or synthetic polymers.

A recent study provided an analysis tool based on an empirical formula (*equation 2*) to handle IDPs[71].

(2) logD (IDP) = -0.507logM - 8.169

Based on the measured diffusion coefficients, adjusted to the relevant experimental conditions (temperature at 298 K, buffer viscosity of $1.50*10^{-10}$ Pas) molecular weights were calculated using *equation 2* as shown in *Table 4*. The results did not yield the expected monomeric molecular mass of 4 kDa underestimating it approximately by - 1 kDa for all four compounds. Using the third formula optimized for structured proteins (*equation 3*), an overestimation around + 2 kDa was observed (*Table 4*).

(3) logD (folded) = -0.381logM - 8.499

	D *(m^2/s)*	**M_w** *(Da)*	**M_w(2)** *(Da)*	**Mw(3)** *(Da)*
wt (2.0 mM)	$1.54*10^{-10}$	4149	2979	5726
var3 (2.0 mM)	$1.48*10^{-10}$	4122	3206	6308
var4 (2.0 mM)	$1.49*10^{-10}$	4218	3191	6270
var8 (2.0 mM)	$1.49*10^{-10}$	4265	3191	6270
var8 (0.5 mM)	$2.95*10^{-10}$	4265	815	1034

Table 4. *The real (M_w) and the estimated molecular masses calculated with different formulas (M_w (2-3)) using the measured diffusion coefficients (D) of the 2 mM samples in water. For var8 data are given at 2.0 and 0.5 mM sample concentrations, respectively.*

The NMR derived diffusion coefficients of 2 mM samples did not fit the Dudás-Bodor model for IDPs (*Figure 17*, *black curve*) nor the model for folded proteins. The measured diffusion values imply that the studied proteins are not ideally intrinsically disordered, *i.e.*, they may adopt secondary structures or other forms of order. For example, oligomerization might also contribute to the deviation observed in *var8*.

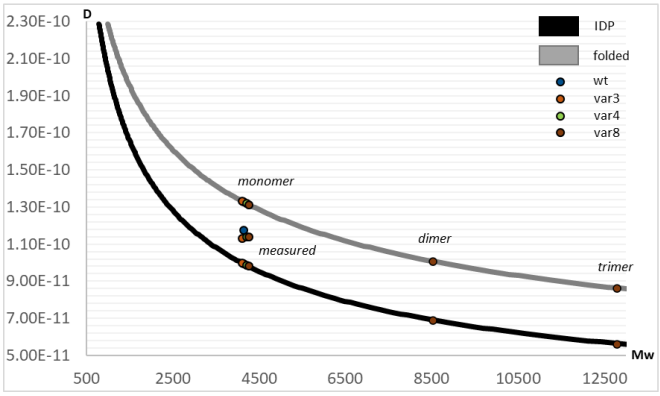


Figure 17. *The points on the curves show the diffusion coefficients calculated with the respective molecular weights of different oligomeric forms on the basis of the Dudás-Bodor model for IDPs (black curve) and for folded proteins (grey curve). The measured diffusion constants plotted against the molecular masses of the monomeric forms are located between the curves.*

An experimental technique distinct from NMR concluded a similar difference in the oligomerization behavior of the studied peptides. Nanoscale Differential Scanning Fluorometry (NanoDSF) is a tool to determine the thermal stability of globular proteins. The degree of unfolding at increasing temperatures is monitored by the intrinsic fluorescence of tryptophan and/or tyrosine residues buried in the structure, being inaccessible to the solvent molecules. The fluorescent intensity maximum of these aromatic amino acid residues is shifted from 330 nm to 350 nm during hydration. The thermal stability of a protein can be described by the thermal unfolding transition midpoint (T_m) as well as the onset temperature (T_{onset}) of the denaturation. T_m can be determined from the inflection point of the denaturation curve, which is the ratio of the tryptophan fluorescence at 350 and 330 nm plotted against the temperature or from the maximum/minimum of its first derivative with respect to the temperature. In principle, disordered proteins might be unsuitable for this experiment because of their unfolded nature. However, if the peptide includes tryptophan like in the case of *var8* and forms oligomers, the changes in tryptophan fluorescence at increasing temperature might indicate transition temperature (T_m) of the oligomeric states.

Figure 18 confirms that the tryptophan containing *var8* peptide showed concentration dependent changes in the fluorescence ratio at 350 nm and 330 nm upon increasing temperature. Evaluation of the first derivative of the changes in fluorescence clearly identified T_{onset} of 20.0 ^oC, 20.1 ^oC and 25.7 ^oC, as well as T_m values of 65.9 ^oC, 61.0 ^oC and 59.3 ^oC at 1 mM, 2 mM and 4 mM concentrations, respectively. Since the peptide is very unlikely to be in a folded state with a well-defined structure, the only explanation for this observation is the assembly of molecules into oligomers leading to changes in Trp fluorescence upon the oligomerization process. The concentration dependence of the T_m values might reflect subtle differences in the stability of oligomers at the applied concentrations. In the reference measurement on *wt*, no such effects were expected, since the peptide does not contain any aromatic residues.

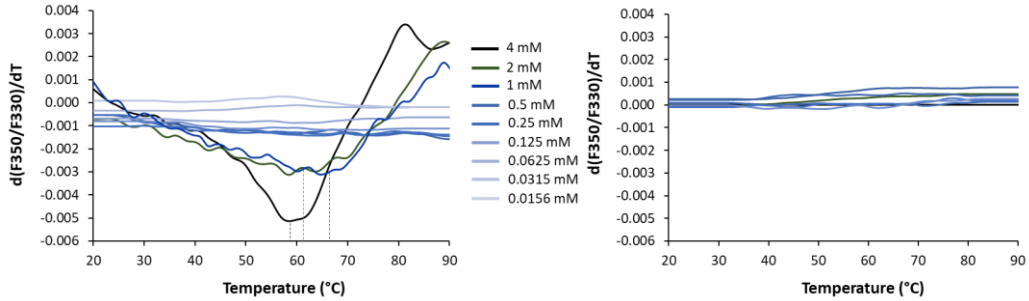


Figure 18. *First derivative of the ratio of fluorescence intensities at 350 and 330 nm over a temperature range of 20-90°C in nanoDSF measurements of var8 (left)* and *wt (right). Different colors correspond to distinct peptide concentrations, dotted lines show the T_m values calculated from the 1 mM - 4 mM datasets. The peptide samples were dissolved in PBS, identical to NMR measurements.*

4.6 MD analysis of the global effects of mutations

Molecular parameters monitoring the dynamical and structural behavior of the peptides were determined in MD simulations.

Positional fluctuations of carbonyl (CO), Cα and backbone amide *N* atoms were selected for the analysis of MEF2D variants. Based on the resulting residue-wise RMSF (Root Mean Square Fluctuations) graph (*Figure 19*), no significant changes occurred at the level of molecular motions (as the average RMSF and the deviation between the average values was 10.31 ± 0.39 Å^2 for the peptides). The mutations caused perturbations only to a small extent in the middle of the backbone and in the C-terminal. (*Table 5, column 2*).

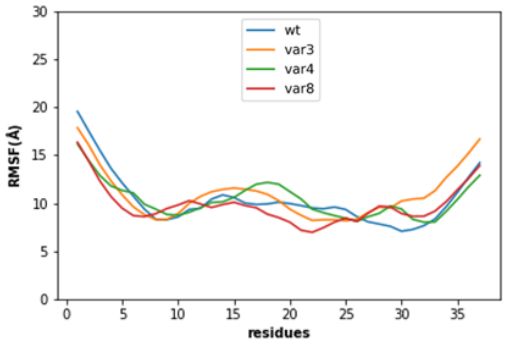


Figure 19. *Residue-wise Root Mean Square Fluctuations (RMSF) of atomic positions, derived from MD-trajectories. Averages are: 9.75 ± 1.95 Å (wt), 10.31 ± 2.70 Å (var3), 10.86 ± 2.47 Å (var4), 10.31 ± 1.83 (var8).*

The molecular compactness determines how the water (solvent) molecules access the protein surface and is described by a well-defined indicator called solvent accessible surface area (SASA). The calculation of SASA reveals buried, heavily folded sites in globular proteins and can be also used to characterize partially disordered molecules with transient secondary structures and less extended chains. No distinction could be achieved between the MEF2D peptides featuring almost identical SASA (279 Å^2) values averaged for all residues, with only minor difference being within the error limit (±1.8 Å^2) of the simulations (*Table 5, column 6*).

Although IDPs are not abundant in structure stabilizing interactions, the charged residues involved in the sequence often form transient salt bridges and hydrogen-bonding also occurs. The investigated MEF2D peptides showed no persistent electrostatic interactions with a long lifetime. But the trajectory analysis revealed, that there was a significant enhancement in the sum of individual salt bridges in *var8* (4564) compared to the *wt* peptide (2128) (*Table 5, column 3*). This can be explained by the introduction of the new, charged residue D29 in *var8* replacing N29. The increased average number of hydrogen-bonds in the simulation snapshots also indicate enhanced structural compactness in *var8* (4.9 ± 2.9), while *var3* (3.4 ± 2.2) and *var4* (3.6 ± 2.3) are less abundant in hydrogen-bonds compared to the reference *wt* (4.0 ± 2.3), suggesting higher flexibility.

The native contact analysis of MD-ensembles collects data about the spatial proximity of the residues in the protein chain at atomic level, to identify folded regions, specific motions, and interactions. The contact maps of MEF2D variants (*Figure 20*) show, that due to the mutations, *var3* and *var4* lost roughly 30% of their native contacts (991 ± 317 and 1167 ± 419) in contrast to the *wt* (1415 ± 480). Surprisingly, there is no change in the total number of native contacts in *var8* (1425 ± 433) on the total number of native contacts.

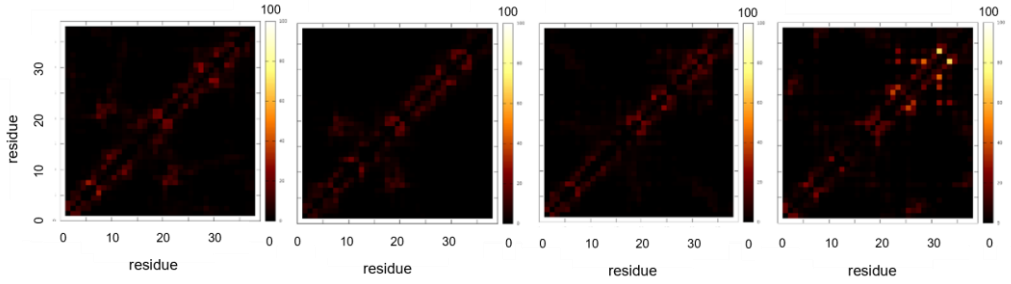


Figure 20. *Native contact heat maps of wt, var3, var4 and var8 peptides from left to right extracted from MD trajectories. Populations range between 0-100% and are labeled with a color scale at the corresponding residue pairs.*

Order parameters (S^2) were calculated from the MD simulations to assess the backbone dynamics and compared with the NMR-derived experimental values. The amplitude of the ps-ns scale motion of the amide-bond vectors was extracted from the trajectories and processed with the Isotropic Reorientational Eigenmode Dynamics (IRED) method[72]. The analysis steps were carried out in 2.5 ns fractions of the trajectories, considering the theoretical global correlation time of the molecules (2.6-2.7 ns) and the scale of the investigated motions. The pieces of S^2 data were then averaged along the whole dataset. MEF2D variants exhibit pronounced differences in backbone dynamics, altered by mutations. Calculations show high preferences for disorder and enhanced flexibility with S^2 values between 0.1-0.8 (*Figure 21*). Comparing β-domain dynamics, the peptides can be distinguished, following the order of predictions. Remarkably, the calculated order parameters are in good agreement with the NMR-derived values. *Var8* shows clear distinction in the ps-ns backbone motions from the other variants, with a + 0.07 increase compared to *wt* (0.40) in the residue-wise average S^2. *Var3* and *var4* (0.39) are similar to *wt*, but show local alterations, especially in the C-terminal of *var3*. Mutation L28K contributed by a decrease of - 0.08 and L34A by - 0.06 to the local MD order parameter.

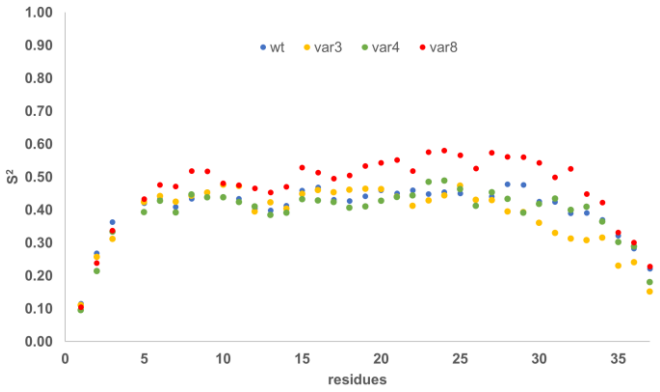


Figure 21. *Residue-wise order parameter (S^2) values from MD simulations. Averages are: wt: 0.40, var3: 0.39, var4: 0.39 and var8: 0.47.*

Mutations	**BB RMSF ±stdev(Å)**	**Salt bridges**	**H-bonds ±stdev [fraction]**	**NNC (stdev)**	**SASA avg. (±stdev) (Å²)**	**S² (NMR)**	**S² (MD)**
wt	9.75 ± 1.95	2128	3.97 ± 2.30	1415 ± 480	279 ± 42	0.43	0.40
var3	10.31 ± 2.70	3177	3.36 ± 2.17	991 ± 317	277 ± 43	0.45	0.39
L28K	-1.86	D24 [0.23]	L26 [0.03], D27 [0.03], D24 [0.03]	-	+29 (-2) 324 (8)	+0.08	-0.08
L34A	-0.53	-	-	-	-77 (+2) 218 (4)	+0.04	-0.06
var4	10.86 ± 2.47	2561	3.56 ± 2.30	1167 ± 419	280 ± 41	0.46	0.39
L28N	-0.1	-	N30 [0.04]	-	-23 (0)	+0.11	-0.05
A31H	+1.79	-	L6 [0.03]	-	+85 (+4)	+0.01	+0.01
V36T	+2.63	-	S37 [0.03]	-	+9 (0)	-0.07	0
var8	10.31 ± 1.83	4564	4.87 ± 2.87	1425 ± 433	282 ± 43	0.50	0.47
N29D	-0.10	R7 [0.05] R3 [0.05] R33 [0.15]	R33 [0.15] D27 [0.09] N30 [0.04]	-	-3 (0) 269 (6)	+0.146	+0.08
A31W	+0.32		R33 [0.13] D27 [0.09] G35 [0.03]	-	+50 (-2) 365 (9)	+0.121	+0.08

Table 5. *Statistics of experimental and calculated descriptors of structure and dynamics. Standard deviations are indicated by ± sign, values in square parenthesis denote the fraction of structures in the MD-ensemble, values after + or - signs are referenced to that of the wild-type data.*

The presence of unstructured regions strongly affects the molecular shape. For instance, molecules of p53 TAD and BPTI are similar in number of residues but have different hydrodynamic radii of 20.3 Å and 15.7 Å, respectively. BPTI is a folded protein with more compact structure than TAD, which is considered as an IDP[73].

The radius of gyration (RoG) descriptor is frequently used in MD analysis quantifying the molecular shape and compactness. The values

computed from the trajectories report differences between MEF2D peptides (*Figure 22*): *var8* had the smallest RoG (15.4 ± 3.4 Å), while *var3* (19.3 ± 3.0 Å) and *var4* (18.7 ± 3.7 Å) were 3 - 4 Å larger. The reference *wt* peptide adopted the intermediate value of 17.0 Å with a deviation of ± 3.8 Å. Although there are large-scale fluctuations in compactness, in general, there is a propensity for a more collapsed structure of *var8*, while *var3* and *var4* peptides present the largest RoG due to their flexible chains. A similar range of values was reported for other IDPs of similar size[74].

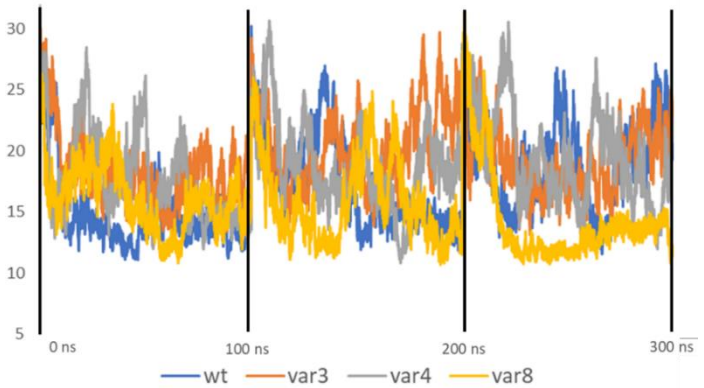


Figure 22. *The radius of gyration (RoG, Å) evolution over time in three, concatenated, parallel MD simulations.*

Cluster analysis was performed on the trajectories to extract representative conformers and to characterize the degree of disorder in the ensembles of the variants, which is reflected in the distribution of cluster populations. Ensembles with few overall number of clusters featuring only few dominant groups with sharply declining populations are more ordered. While higher number of clusters with smaller, slowly declining populations indicate more conformational diversity. The high number of structures that cannot be assigned to any clusters also indicates disorder.

The number of MEF2D conformers that could not be clustered using the same criteria was generally high (51 - 82%), but some diversity of the structures could be confirmed. In the case of *wt*, a steep decline of cluster populations (*Figure 23*) indicated a more ordered structural ensemble. Dominant *wt* conformers were identified in the first, largest cluster of the simulation with a 28 % population. More flexibility is evidenced by the moderate descend in the

population of the clusters and the high number of unclustered points for *var3*, *var4*. The ensembles contained „worm-like" peptides with enhanced dynamics, which were difficult to classify into clusters. Conversely, v*ar8* simulations produced one significant conformational group with smaller subclusters pointing towards more structural order.

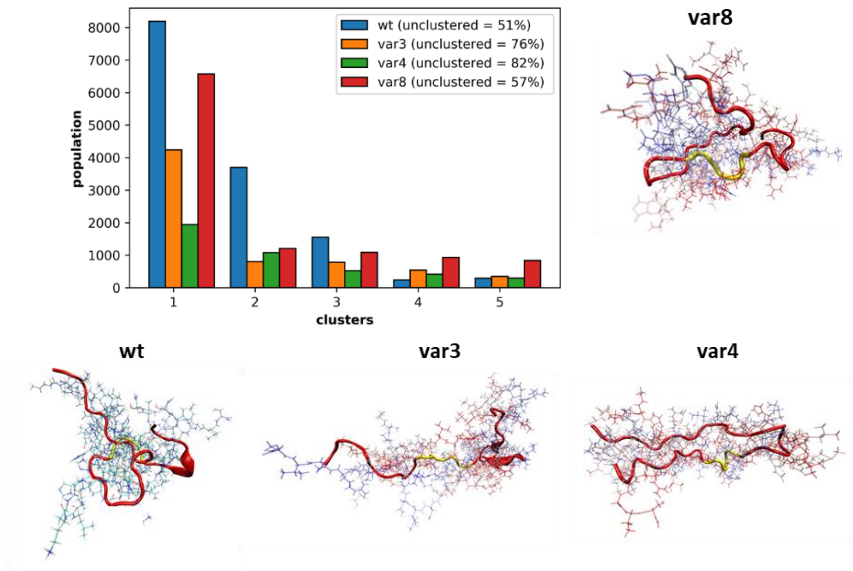


Figure 23. *Population distribution of the clustered conformations of the MEF2D MD simulations. The ensembles contain 30000 conformers for each variant. Around the graph: representative structures of the largest clusters with populations of 0.27, 0.14, 0.06, and 0.22 for wt, var3, var4, and var8, respectively. The β-domains (residues 22-29) are labeled with yellow color.*

Although IDPs often lack the well-defined structural elements, they might be partially folded, or adopt unstable, fluctuating, transient secondary structures. In MD analysis, their characterization is based on similar concepts as that of folded proteins, but the search aims to find smaller populations. To compare *MEF2D* variants, site-specific analysis was carried out with STRIDE[64] along the MD trajectories to find structured residues. The algorithm identifies the main helical and extended elements, but in the case of disordered proteins, attention should be paid to turns, loops and bent forms as well.

Simulations confirmed the lack of stable secondary structural elements in the ensembles, pointing to a preference for disorder for all peptide variants (*Table 6)*. Some propensities were observed for common transient helical components (*var8*) and small *β*-sheets (*wt*) at 1% of the ensemble, but missing

in *var3* and *var4*. Alongside these minor differences, the number of structures adopting turn elements is also decreased by one-third in the latter variants. The ratio of unstructured populations (coils) divided the studied peptides into two groups, one comprising the more structured *wt* and *var8* constructs, and the other with the marginally ordered *var3* and *var4* peptides.

	wt	**var3**	**var4**	**var8**
α-helix	-	-	-	0.01
extended	0.01	-	-	-
3_{10}-helix	0.02	-	-	-
turn	0.30	0.21	0.23	0.29
coil	0.65	0.77	0.75	0.68
isolated bridge	0.01	0.01	0.01	0.01

Table 6. *Population of conformations with transient secondary structure elements along the MD simulations.*

Following the time evolution of transient secondary structures, color coded, residue-wise plots were generated (*Figure 24*). Owing to mutations, the structured N-terminal part (V8-Q12) of *wt* (blue: 3_{10} helix and yellow: small β-sheet, first 100 ns) disappeared in *var8*. The structural order was shifted to the binding region with the appearance of α-helical stretches (magenta) at the C-terminal end (residues H25-N30) near the β-domain region in the first 100 ns. *Var3* and *var4* confirmed a rarely structured backbone adopting mainly turn elements with short lifetimes.

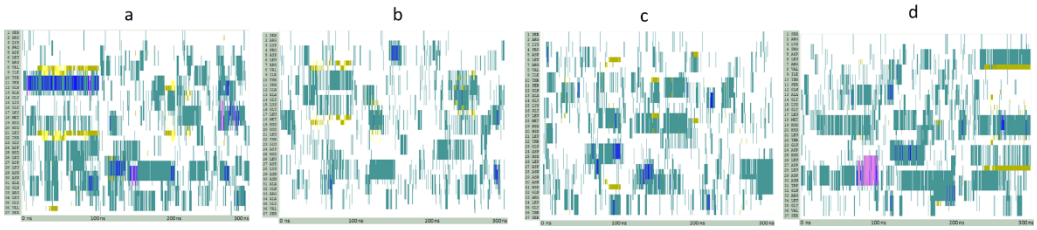


Figure 24. *Time evolution of transient secondary structures for wt (a), var3 (b), var4 (c) and var8 (d) in the concatenated MD-trajectories (3x100 ns each). The colors represent the following secondary structures: turn (green), extended (yellow), helical (magenta) and 3_{10} helix (blue).*

4.7 Site-specific MD-analysis

The structural and dynamical perturbations at the site of mutations were also analyzed (all data provided in *Table 5*). In *var3*, mutation of L28K leads to significant enhancement in local backbone dynamics shown by S^2 (-0.08) and increased SASA (+ 29 Å^2), which is a sign of diminished compactness, while the backbone RMSF also decreased (- 1.86 Å). New hydrogen-bonds were formed with the neighboring residues L26, D27, D24 in a small fraction of the conformers (0.03 each). A significant population (0.24) contains a medium-range salt bridge, established between K28-D24, but without any evidence of a contribution to structure stabilization. Regarding the L34A mutation, decreased RMSF (- 0.53 Å), more intensive ps-ns dynamics (S^2 of -0.06) and smaller SASA (- 77 Å^2) were observed without the formation of new H-bonds.

Mutation of L28N in *var4* did not significantly alter the mobility locally, on the other hand, enhanced disorder is evidenced for A31H and V36T with increased atomic fluctuations (+ 1.79 and + 2.63 Å, respectively) and SASA (+ 85 and + 9 Å^2, respectively) of the construct.

Notably, for *var8* a new network of H-bonds with longer lifetimes was established at the C-terminal owing to mutations N29D and A31W (*Figure 25A*). D29 has high promiscuity for maintaining salt bridges with the neighboring (R33) and distant (R3, R7) arginine residues, which were together present in 25% of the simulation time. The aromatic W31 residue participated in electrostatic interactions with cationic residues, as the distance analysis reports
(*Figure 25B*). Mainly the neighboring R33 sidechain is positioned near the indole-ring, in close contact (< 7 Å) in 44% of the concatenated MD-ensembles, fulfilling the geometric criteria of cation-π interactions[75]. This structure stabilization effect – which is common in IDPs abundant in charged residues – may also promote the formation of a transient helical structure in the C-terminal at the end of the first 100 ns simulation (*Figure 24*). Concerning dynamics on the ps-ns scale, both mutations effectively increased the rigidity at residue level (+ 0.08 S^2) as well as in the neighboring region (around + 0.10 S^2) (*Figure 21*).

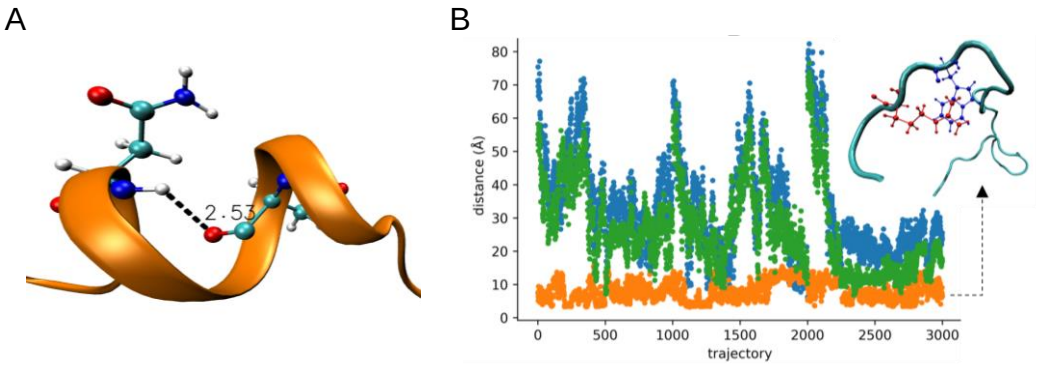


Figure 25. *A hydrogen bond between residues D29-R33 is stabilizing the helical structure in the C-terminal of var8 (A). Time evolution of the distance between var8 atoms W31(Cε2) and R2 (CZ, orange), R33(CZ, green), K15(NZ, blue) to unravel the existence of cation-π interactions between the mutant tryptophan residue and cationic sidechains (B). A representative conformation (right) shows the proximity of R33 sidechain to the W31 indole ring.*

4.8 Oligomerization characterized by simulations

The oligomerization properties of *var8* were also characterized by computations in simulation boxes containing 7 copies of the monomeric peptide. Three replicas of 100 ns classical MD simulations were performed, followed by 100 ns of scaled MD computations for better conformational sampling. As a result, molecular assemblies were formed and identified involving different number of monomeric units. The oligomerization states of *var8* were calculated with respect to the simulation time based on the atomic distances between monomers as shown in *Figure 26*. In the classical simulations, the formation of oligomers evolved and reached an equilibrium state by the end of the trajectory, which remained largely conserved in the following scaled trajectories with pentamer-hexamer transitions (d), monomer-dimer transitions (f), and a stable heptamer (e). Although the distribution of the units was different in parallel calculations, the existence of at least dimeric states was confirmed indicating a strong preference for the formation of molecular assemblies.

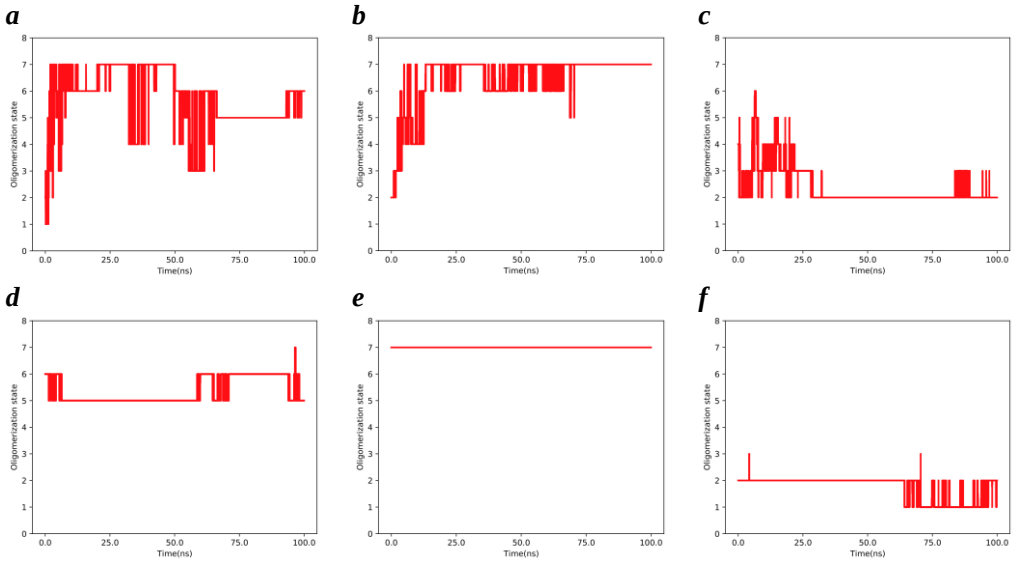


Figure 26. *Oligomerization state of peptide var8 calculated in trajectories with respect to the time based on the atomic distances between monomers for three replicas of classical MD simulations in a), b) and c) and in their follow up scaled simulations in d), e) and f).*

Since experiments indicated reversible processes, oligomer formation might be induced by non-covalent interactions. To unravel the interface points between monomers, the intermolecular contact maps of the system were extracted (*Figure 27*). The data matrix has all the possible combinations of the residues and the corresponding population of structures where the distance cut-off criteria (< 4 Å) is met. The contacts were only defined by their spatial proximity along the MD trajectories.

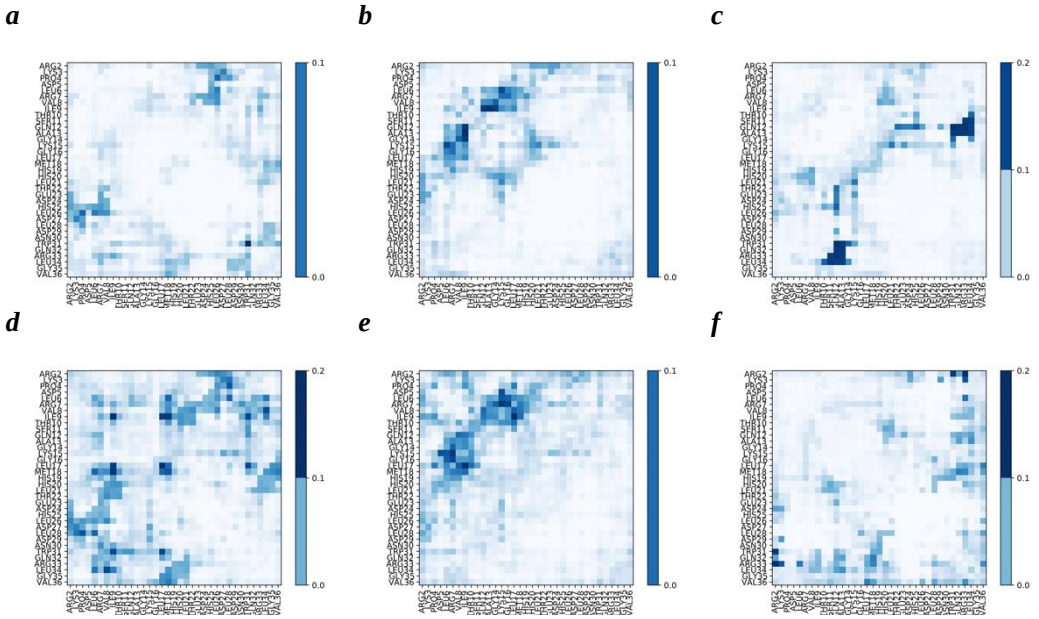


Figure 27. *Intermolecular contact maps for var8 in simulations containing 7 monomeric units for three replicas of classical simulations in a), b) and c) and in their follow up scaled simulations in d), e) and f). The color code refers to the respective populations, usually between 0.0 and 0.2 as the scales on the right of the plots show.*

The formation of hydrophobic cores is preferred in globular proteins, often driving the folding, or even aggregation in some cases. Looking at the apolar MEF2D residues, frequent contacts were identified in the maps (d, e). *Ile9*, *Leu17* and *Met18* are mainly involved in intra- and intermolecular interactions to form small cores, being likely stabilized via hydrophobic effects. Snapshots on different parts of the scaled MD trajectory highlight the spatial proximity of these residues in higher oligomeric states (*Figure 28*).

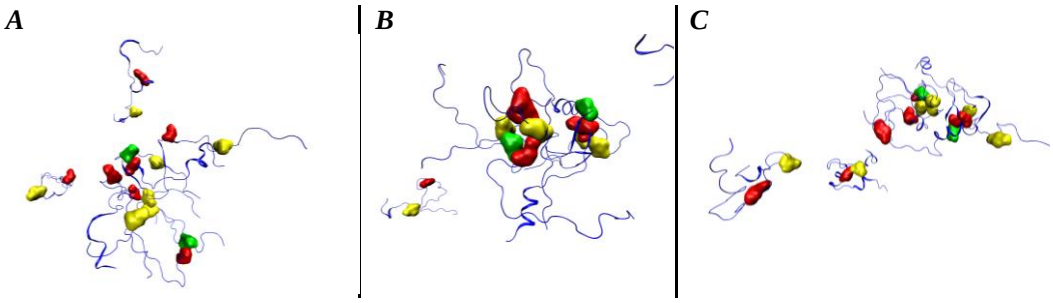


Figure 28. *Formation and dissociation of hydrophobic cores at different points of the scaled MD trajectory of var8. A: at 10 ns simulation time with oligomerization state of 5. B: at 60 ns with oligomerization state of 7. core(left): Ile231-120-83, Met129-92 core(right): Leu17, Met203-18, I9, L91. C: at 90 ns with an oligomerization state of 5. Different surfaces with colors represent the location of hydrophobic residues: Ile (yellow), Met (red), Leu17,91 (green).*

The aromatic Trp residue can only be found in the *var8* sequence. This point mutation might be important for maintaining order in the C-terminal, as well as for participating in intermolecular interactions. Aromatic-aromatic π-stacking contacts between residues can be identified in a simulation, where Trp indole rings of different protein chains are facing each other promoting a hexamer formation, visualized in *Figure 29*.

Trajectories of *var8* were scanned for identifying contacts between K15-W31 and R2-W31 (*Figure 27c,f* and *Figure 29*). Head-tail (between N and C-termini) positioning are evidenced in two of the trajectories with modest propensities (0.1-0.2 fractions) as shown in *Figure 27a,d,c,f* with combinations of classical simulations and follow up scaled MD. This molecule arrangement can be stabilized via a cation-π interaction (*Figure 29C).*

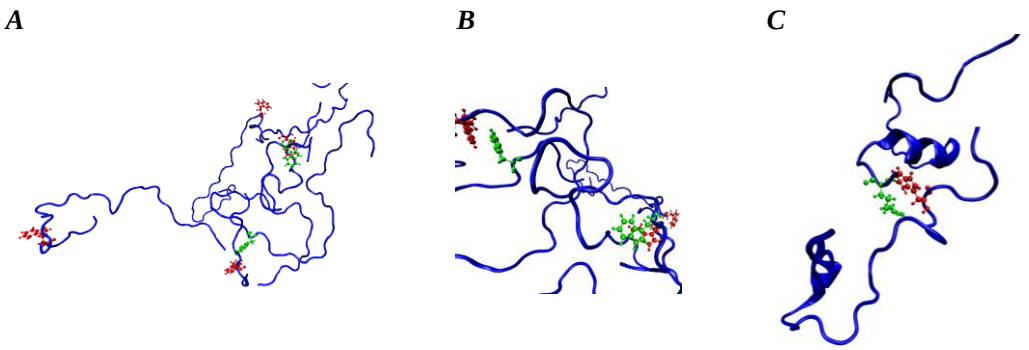


Figure 29. *Intermolecular aromatic stacking of Trp residues is visualized with a snapshot of a var8 hexamer from an MD trajectory at ca. 40 ns simulation time. A: locations of the Trp sidechains in the oligomer. B: Stacking of W105-W253 and W31-W216 in close view. Trp sidechains are labeled with green and red colors. C: Head-tail peptide connection is maintained by the cation-π interaction of N-termini Arg (green) and C-termini Trp (red) resulting in a stable dimer visualized by a snapshot of scaled MD simulation in trajectory f) at 60 ns.*

In conclusion, CHARMM36m MD simulations produced ensembles with higher propensity for disorder and low abundance of secondary structures for the studied peptides. All variants of *wt* MEF2D mini-protein showed high degree of dynamics as evidenced by intermediate values of experimental and calculated order parameters. Peptides could be separated into two groups: one with enhanced structural dynamics and tendency for disorder (D) and the other with a relatively more rigid and more compact character (O) compared to the *wt* reference compound. Variants *var3* and *var4* represent group D, with high amplitude residue motions (RMSF), larger surface area for solvent access (SASA), increased dynamics in ps-ns scale (S^2) and greater molecular radii (RoG) compared to *wt* and *var8*. The intramolecular interactions maintaining secondary elements or restrained structural mobility are fairly observable (native contacts). No significant, representative structures could be identified with cluster analysis. The observed changes of the structural and dynamical descriptors in the binding region (MEF2D β-region) verify the success of the planned mutations. *Var8* is characterized as a member of group O with different behavior. The significantly decreased ps-ns backbone dynamics demonstrated by larger S^2, the smallest RoG, the larger propensity for secondary structures and the helical preference of C-terminal are the main descriptors to describe the altered molecular nature. The rigid, more compact structure is supported by the interactions of the residues in the form of new H-bond networks, salt bridges and cation-π effects. Finally, detailed molecular simulations supported the experimental evidence for *var8* oligomerization revealing the key interaction spots and the number of monomers constituting the peptide assemblies.

5. Results part II: developing combined NMR experiments

5.1 The Ernst-angle concept

The NMR experiments consist of four main building blocks. First, the sample magnetization is excited during the preparation period. Secondly, in the evolution and mixing block of 2D/or nD experiments, the NMR observables (*e.g.*, chemical shifts, couplings) are evolved by specific spin manipulations. Thirdly, during acquisition, the free induction decay (FID, time domain data) is detected by the receiver, from which the spectrum (frequency domain data) will be generated with Fourier transformation. Finally, the longitudinal relaxation (T_1) recovers the exhausted magnetization in a delay time (called relaxation/recovery or recycle delay), to be reused in the next iteration of the experiment. To enhance the signal-to-noise ratio (S/N) of spectra, FID averaging is carried out, which requires the repetition of the whole measurement sequence.

Since NMR spectroscopy is inherently less sensitive compared to other spectroscopic/analytical techniques, it is always essential to optimize the pertinent experiment. In a basic one-pulse 1D NMR experiment, a 90° (π/2) hard pulse transfers all the available sample magnetization to the x-y plane for detection, yielding maximum peak intensity in the spectrum. Deviating from this excitation angle – particularly in 2D measurements, where the hard 90° pulse must be calibrated precisely – leads to losses in sensitivity. When T_1 is relatively short compared to the total duration of the applied acquisition time and recycle delay, the relaxation is nearly complete, yielding maximum signal intensity by the 90° excitation. However, when T_1 is longer, the relaxation becomes incomplete with the same recycle delay, therefore it must be increased for gaining sensitivity. Since the NMR experiments acquire data with the repetition of the same pulse sequence block, the duration of the relaxation delay (d_1) has significant impact on the measurement time. For example, with doubling the duration of d_1 in common measurements, the experiment time is almost doubled without any additional benefit, because during the relaxation delay no effective spin-manipulation or data acquisition is carried out. A good

compromise can be reached with reasonable optimization, where a reduced, so-called small flip angle excitation pulse is applied together with shorter d_1. This idea originated from the Nobel laureate Richard Ernst (1933-2021)[76]. Although the Ernst-angle excitation exploits less magnetization, but on the other side, the relaxation of the spins is completed in shorter time, allowing faster repetition of the measurements. Regarding systems with long T_1 relaxation (2-3 s or more), the best signal-to-noise ratio is obtained when a small flip angle excitation pulse is combined with a short relaxation delay.

The novel experiments presented in this work use the Ernst-angle concept to exploit the available sample magnetization the most efficiently. The optimized polarization share between the sequentially run experiments was carried out with partial excitation of the different magnetic reservoirs in various ways, detailed in the subsequent chapters. At the beginning of each section, the essential pulse sequence blocks applied for excitation are introduced, then the spin manipulations necessary for the different experiments are explained briefly. Finally, the optimization and application of the different combined pulse sequences are presented with various small- and medium-sized test molecules.

5.1 The BANGO excitation technique

The pulse sequence design by the NOAH (NMR by Ordered Acquisition using ^1H-detection) concept requires the independent manipulation of different magnetization pools, which is carried out by the BANGO element[77] (β-TANGO), the small-angle version of TANGO[78] (Testing for Adjacent Nuclei with a Gyration Operator) sequence in our novel NOAH-type combined NMR experiments.

With BANGO the pool of I spins not attached to S spins (where I: ^1H, S: ^{13}C) is selectively excited with a rotation angle β^I equal to (180°-2β), where pulse β is shown with the blue narrow-filled bar in *Figure 30*. Meanwhile, the magnetization of ^1H attached to ^{13}C is preserved with a rotation of angle β^{IS} (where β^{IS} is 180°(π), inversion) for the next experiment. In other words: at the beginning of the pulse sequence, BANGO is used to prepare and split the magnetization pools for the subsequent concatenated experiment modules.

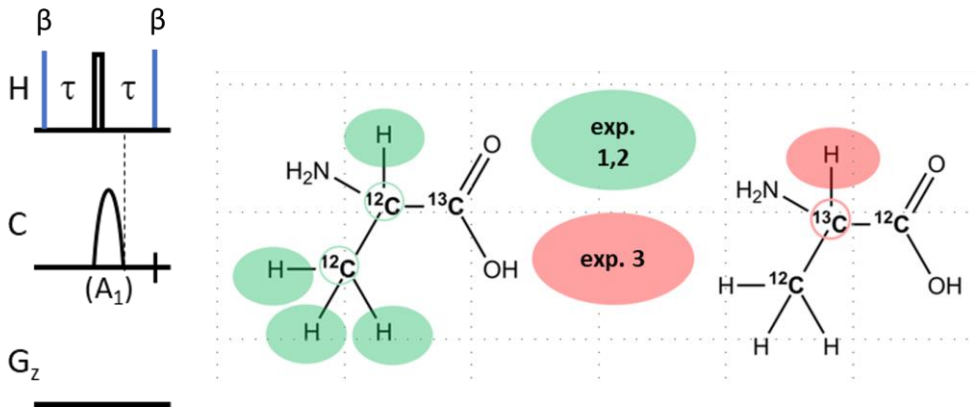


Figure 30. *The pulse scheme of the BANGO excitation element (left): H: ^{1}H, C: ^{13}C, G_z: gradient channel. A1: adiabatic 180° carbon pulse. Narrow-filled pulses: ß, Open pulses: 180°. Examples for different magnetization reservoirs (right): **^{1}H-^{12}C** (green) and **^{1}H-^{13}C** (pink).*

5.2 Concatenated experiments using a single BANGO excitation

The SEA XLOC and H2OBC are individual experiments, but in this work, they were merged into one super sequence. The resulting BANGO {SEA XLOC}-{H2OBC} concatenated experiment is a fast and efficient NOAH-type measurement, where only one recycle delay is needed for recovering the magnetization for the next experiment and data collection.

The first part of the sequence applies the BANGO element and excites the ^{1}H pool not attached to ^{13}C (*Figure 31a*). Then the magnetization is passed to the SEA XLOC module (*Figure 31b*), which is except the starting BANGO excitation block identical to the individual experiment. After the acquisition of SEA XLOC data (*Figure 31c*), H2OBC follows immediately and operates with the "untouched" pool of the ^{1}H attached to ^{13}C (*Figure 31d*). Specifically, this magnetic component is returned to the positive z-axis prior to the H2OBC experiment by the combined effect of the 180° BANGO pulse on the **^{1}H-^{13}C** magnetization (*Figure 31a*) and the last 180° ^{1}H (#) pulse of the SEA XLOC module. This foreplay takes advantage of T_1 relaxation and makes the **^{1}H-^{13}C** polarization ready for H2OBC. After completing the H2OBC module,

the acquisition follows as the last part of the sequence (*Figure 31e*). On the whole, running the two experiments in a sequence with polarization sharing, only one relaxation delay is needed, thus, a significant amount of measurement time can be saved.

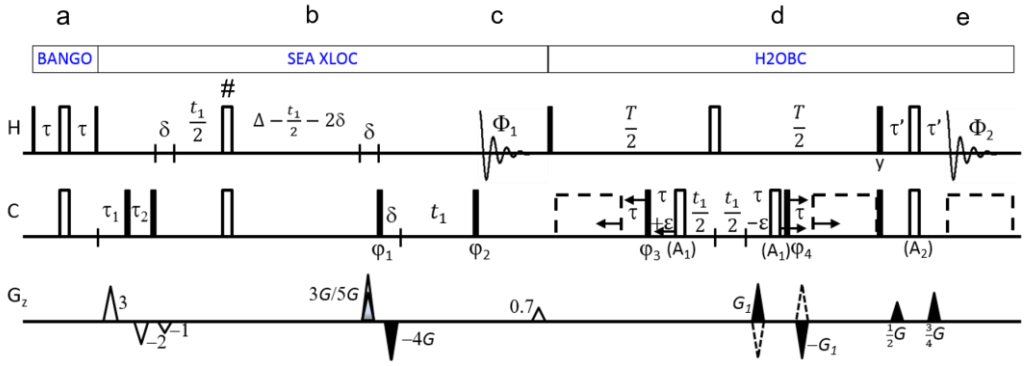


Figure 31. *The BANGO {SEA XLOC}-{H2OBC} pulse sequence. Narrow filled, wider filled and open bars refer to π/4, π/2 and π pulses, respectively. A1 and A2 are adiabatic ^{13}C inversions pulses. The delays Δ and T are for evolution under heteronuclear long-range and homonuclear proton–proton couplings, respectively. δ is a gradient delay. Further details of the sequence can be found in the first BANGO paper* [79].

The first model compound used for testing the experiment and demonstrating correlations for all carbon multiplicities, including aromatic atoms was a small drug molecule (ibuprofen). On the absolute value SEA XLOC ZQ spectrum (*Figure 32A*), two- and three-bond correlations are demonstrated. The two magnified peaks show that the distinction between two (C1-H2)- and three-bond (C1-H13) correlations can be accomplished by comparing ZQ/2Q peak widths. The red ZQ and the black 2Q spectra are placed on the same plot, but the latter was displaced to visualize the spectral differences better. The examined peaks are enlarged, and the corresponding 1D sections were also included for better comparison. This method is unique since all carbons – including quaternary carbons – can be analyzed. For instance, the C1 carbonyl resonance could be identified at 181.5 ppm (^{13}C) in correlation with the adjacent proton H2 at 3.75 ppm. Compared to the displaced 2Q counterpart, the multiplet line width of ZQ in the carbon (F1) dimension is

narrower, shown in both the contour plot and the extracted 1D sections. Consequently, it is identified as a two-bond correlation with ZQ peak width determined by $^3J_{HH}$ - $^3J_{CH}$. On the other hand, the peak at 1.55 ppm showing the opposite multiplet width pattern – namely, the line width of ZQ correlation is wider ($^3J_{HH}$ - $^2J_{CH}$, $^2J_{CH} < 0$) than that of the 2Q multiplet ($^3J_{HH} + {}^2J_{CH}$, $^2J_{CH} < 0$) – corresponds to a three-bond correlation with H13.

The high digital resolution makes the accurate measurement of the three-bond homonuclear coupling constants ($^3J_{HH}$) also feasible for protons attached to carbons. As an example, selected region of the ZQ spectrum (*Figure 32B*) presents protonated carbon resonances and focuses on the two-bond heteronuclear correlations of C10-H9 and C10-H11/H12. The $^3J_{HH}$ values of 6.7-7.2 Hz can be easily extracted from the E.COSY patterns of the cross-peaks.

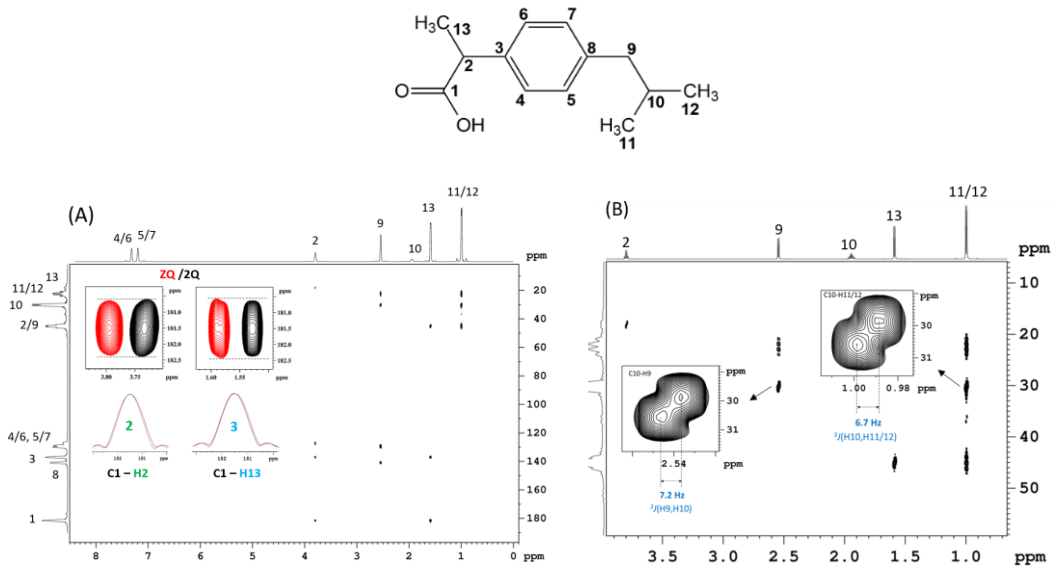


Figure 32. (A) *The structure of ibuprofen (top) and the SEA XLOC zero-quantum (ZQ) spectra extracted from BANGO {SEA XLOC}-{H2OBC}. ZQ and 2Q F1-linewidth comparison is demonstrated on two distinct, enlarged peaks. (B) Same correlation map demonstrating the measurement of two selected three-bond homonuclear coupling constants. Spectra were recorded with the parameters accessible in Table 3(14)***.**

The H2OBC spectrum recorded in the combined experiment allows an 'assignment walk' along the C9-C10-C11/C12 chain as shown in *Figure 33*. The peak phases in the H2OBC spectrum are determined by the correlation types: H2BC peaks are antiphase dispersion (COSY-type), while HSQC peaks are pure absorption. Moreover, there is one more layer of phase editing, in which ^{13}C atoms with odd and even multiplicity give rise to peaks with opposite phases in the H2BC and HSQC subspectra.

The starting point of the 'assignment walk' is the one-bond, HSQC correlation of C9 (45 ppm) with its methylene protons (~2.5 ppm), appearing with negative magnitude (green) due to the even carbon multiplicity (CH_2). The vertical line connects the neighboring carbon C10 (30 ppm) with a negative peak due to two-bond, H2BC correlation with odd (CH) multiplicity. The following dashed horizontal line points to the neighboring H10 proton, in the form of an HSQC peak with positive magnitude (red) because of odd multiplicity (CH). The assignment walk continues in this up and down fashion until the spin system endpoint, the one-bond heteronuclear correlations of C11/C12 methyl groups (22.2/1.0 ppm) with positive magnitude.

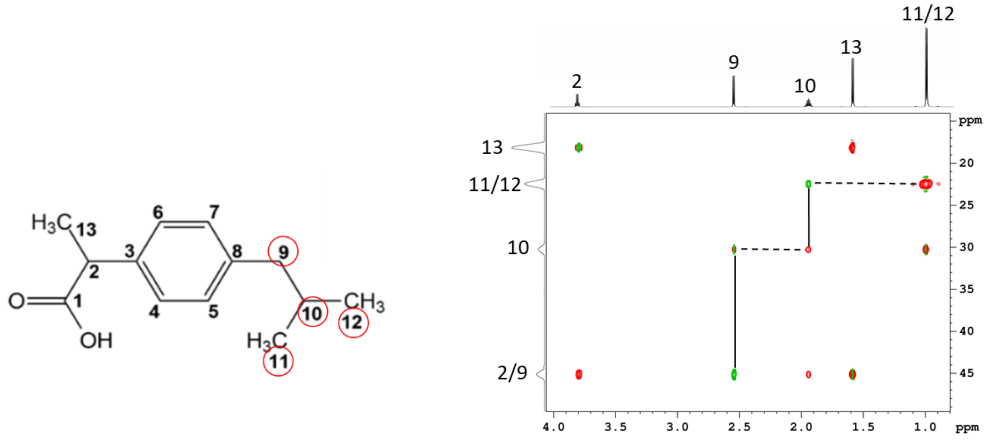


Figure 33. *An "assignment walk" in one of the ibuprofen spin systems is demonstrated with an excerpt of H2OBC spectrum recorded with the BANGO {SEA XLOC}-{H2OBC} two-module experiment. Experimental parameters can be found in Table 3(14). Peaks with even carbon multiplicity (negative magnitude) are colored green, while HSQC peaks of odd multiplicity (positive magnitude) are red.*

By replacing the SEA XLOC module with the conventional HMBC experiment, instrumental time could be reduced to 11 minutes since the latter experiment has slightly better sensitivity. Although the two- and three-bond correlations cannot be distinguished, the experiment performs well under the experimental conditions applied.

The resulting HMBC spectrum (*Figure 34a*) has all the multiple bond correlations that can be easily analyzed. The H2OBC spectrum recorded in the same BANGO {HMBC}-{H2OBC} experiment is shown (*Figure 34b*) with an extracted row of C2/C9 (45 ppm), demonstrating the phase differences in 1D view. To test the performance of the novel combined experiment, comparisons were also carried out with the conventional standalone experiments. In *Figure 34c*, a region of HMBC spectrum recorded in the standalone experiment (black) was displaced relative to the BANGO version (red) for better comparison. As expected, the spectra show no differences in the pattern of cross-peaks. This is also the case with the H2OBC spectra in *Figure 34b,d*, where the 2D H2OBC map recorded in the standalone experiment is plotted in *Figure 34d*.

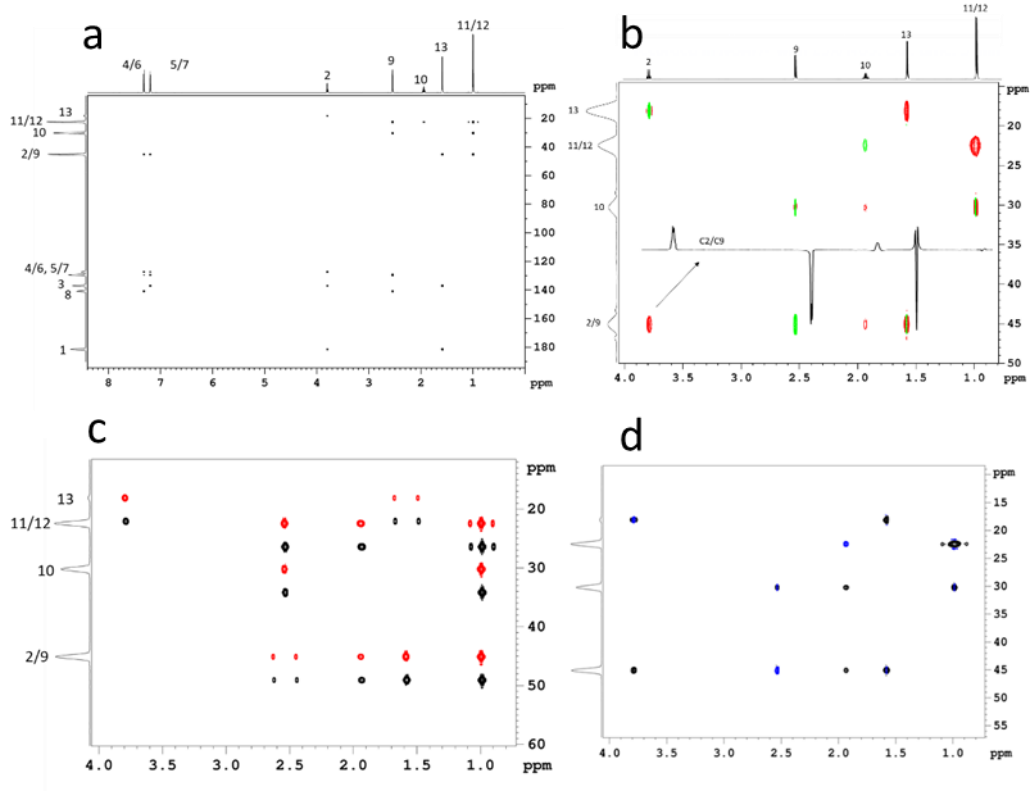


Figure 34. *Comparing different spectra of ibuprofen. a) Full HMBC. b) Expansion of H2OBC recorded in the concatenated experiment. Peaks with even carbon multiplicity (negative magnitude) are colored green, while HSQC peaks of odd multiplicity (positive magnitude) are red. c) Region of HMBC recorded in the concatenated (red) and standalone (black) experiments, d) H2OBC spectrum recorded in the standalone experiment. Data were extracted from the BANGO {HMBC}-{H2OBC} concatenated measurement with parameters accessible in Table 3(12).*

The sensitivity of the individual modules of the combined experiments was compared with that of the standalone measurements based on the extracted 1D traces of the corresponding correlation spectra (*Figure 35)*. There is an expected sensitivity loss due to the 90° BANGO excitation, which was on average 6% for H2OBC and 8% for SEA XLOC. On the other side, a 6% sensitivity gain was achieved in the HMBC spectrum.

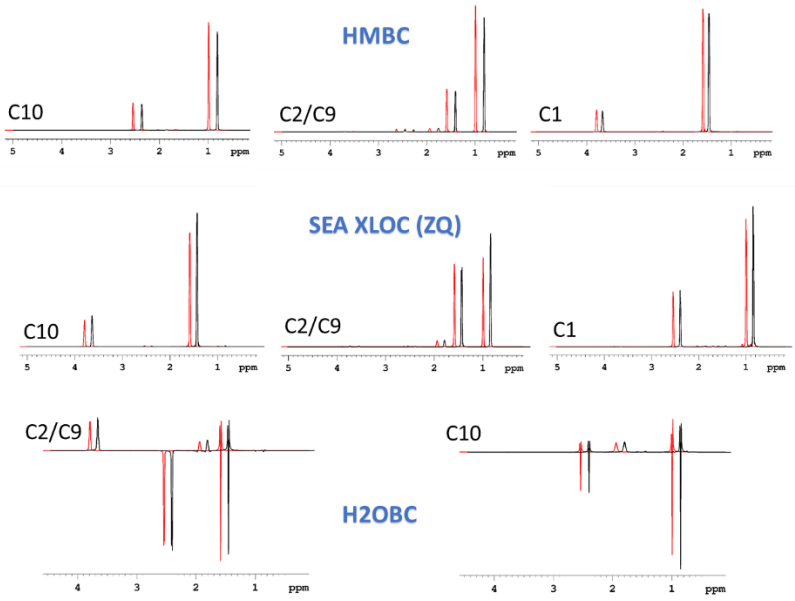


Figure 35. *Sensitivity comparison of the respective standalone (black) and concatenated (red) measurements on ibuprofen with the extracted ^{1}H rows from BANGO {HMBC}-{H2OBC} (Table 3(12)) and BANGO {SEA XLOC}-{H2OBC} (Table 3(14)).*

With regard to the invaluable instrument time, significant time saving was achieved with the novel pulse sequences. On ibuprofen, the combined experiment providing complete heteronuclear correlation maps required 42 minutes. While performing the experiments separately, SEA XLOC 2Q and ZQ were run for 18 minutes each and then followed by a 35 minutes H2OBC. In total, 71 minutes were required, thus 41% of instrument time was saved with the concatenated experiment.

The second test molecule for the same experiments was a trisaccharide. The determination of the sequential order of sugar residues is a challenging task, which was successfully achieved with the identification of the interglycosidic heteronuclear long-range connectivities together with the complete and unambiguous assignment of ^{1}H and ^{13}C resonances (*Figure 36*). The sample concentration and so the sensitivity allowed further reduction of the experimental time with the application of non-uniform sampling (NUS) for recording the F1 frequency dimension. With separate measurements, both HMBC and H2OBC experiments took 5 minutes, while the combined experiment with NUS sampling required only 5.5 minutes altogether. Thus, in this case a timesaving of 45% was achieved.

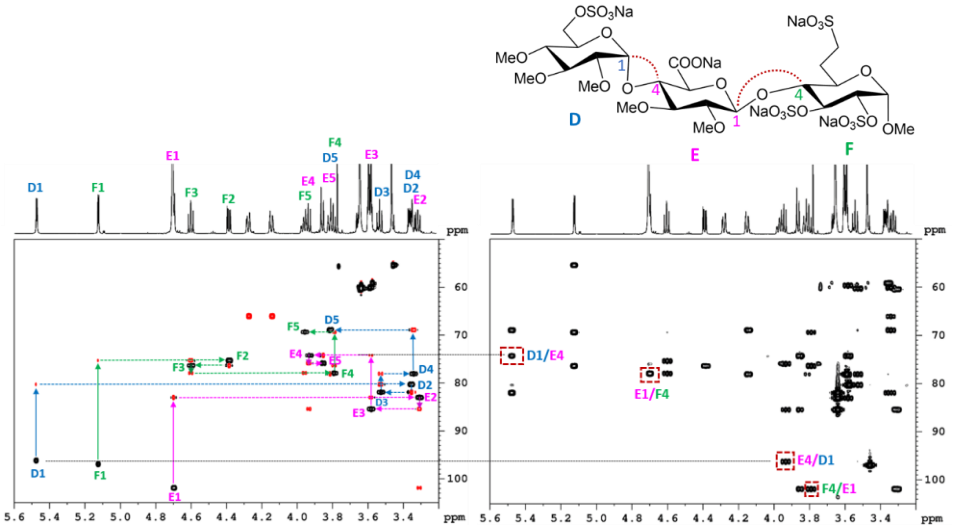


Figure 36. *H2OBC (left) and HMBC (right) spectra of a trisaccharide obtained with BANGO {HMBC}-{H2OBC} experiment using parameters in Table 3(13). The assignment of the spin systems of the three sugar residues (labeled with different colors) is demonstrated in H2OBC. HMBC shows the interresidual connections.*

5.3 Concatenated experiments using double BANGO excitation.

The following improvement in the NOAH-type NMR combined sequences was the introduction of a second BANGO excitation element in the pulse program. This allows the incorporation of an additional experimental module, sharing the pool of magnetization with the preceding modul and eliminating the need for any recovery delay between the experiments (*Figure 37)*. The sequence is constructed for time efficiency, which is achieved without significantly compromising the sensitivity of the individual measurements.

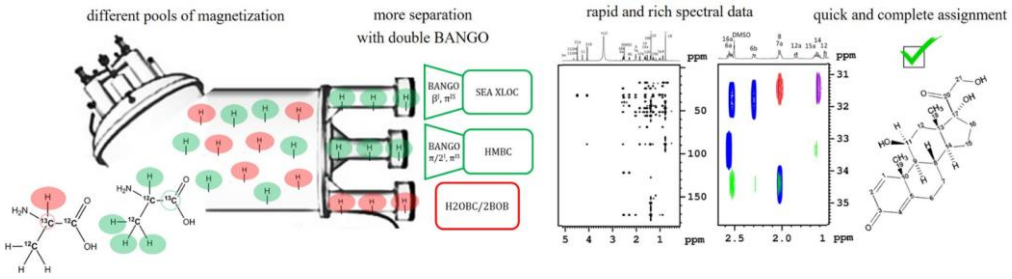


Figure 37. *The illustration demonstrates how the different pools of magnetization are exploited by three experimental modules. NMR plots show the data from the measurement of prednisolone compound discussed in detail on page 76.*

The manipulation of the two magnetization pools (I: ^{1}H-(^{12}C) and IS: ^{1}H-(^{13}C)) is carried out as before, but the protons (I) not attached to ^{13}C are tapped twice with the two BANGO elements. The β^{I} flip-angle of the first BANGO pulse is optimized between 90° ($\pi/2$) and 180° (π) to achieve only partial excitation of the ^{1}H-(^{12}C) magnetization pool for the first experiment. Thus the magnetization left can be utilized in the subsequent module of the combined experiment. To achieve the best sensitivity, the second BANGO with 90° β^{I} flip-angle excites the whole remaining magnetization. Once this pool is exhausted, the so far preserved ^{1}H-(^{13}C) proton magnetization is exploited by the last experiment.

The performance of the double BANGO measurements is demonstrated with different combinations of three experimental modules: SEA XLOC, HMBC, and H2OBC/2BOB (*Figure 38*). In the BANGO {SEA XLOC}-

{HMBC}-{H2OBC} experiment (A) all three modules are concatenated. In the second combination (B), the HMBC module is discarded and instead, the SEA XLOC ZQ/2Q subspectra are recorded separately in the first two modules for better sensitivity.

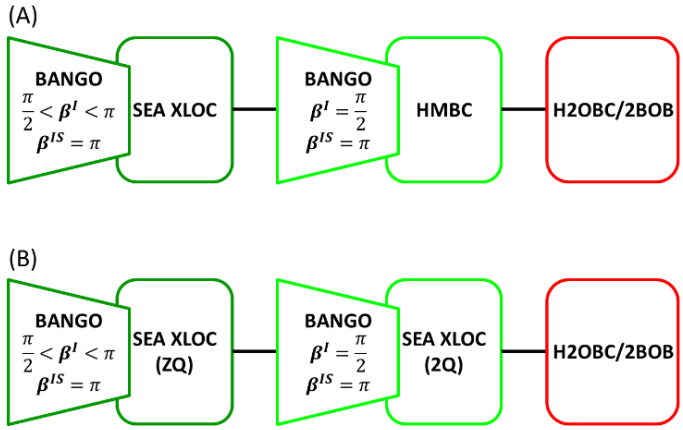


Figure 38. *Schemes of the three-module double BANGO experiments: BANGO {SEA XLOC}-{HMBC}-{H2OBC/2BOB} (A) and BANGO {SEA XLOC(ZQ)}-{SEA XLOC (2Q)}-{H2OBC/2BOB} (B). The optimized β^I flip-angle of BANGO is used for partial excitation of the ^{12}C-^{1}H magnetization reservoir in the first module, while β^I 90° excites the whole remaining magnetization for the second module. β^{IS} 180° is used to preserve the ^{1}H-(^{13}C) pool for the last module.*

The sensitivity of the individual experiments was the key point of the sequence design strategy, including the determination of the best order of the experimental modules. SEA XLOC is less sensitive than HMBC, therefore it must be run first to gain a larger share from the ^{12}C - attached ^{1}H magnetization pool. The optimization of the β^I excitation angle (*Figure 39a*) suggested a 120° rotation being optimal in the first BANGO to achieve balanced magnetization distribution between the SEA XLOC and the HMBC experiments. Then the HMBC starting with the second BANGO of 90° β^I flip angle exhausts the same magnetization pool. The third module (H2OBC or 2BOB) relying on the preserved ^{1}H - (^{13}C) magnetization is excited by a ^{1}H 90° hard pulse, just like in the standalone experiment. Regarding the second pulse sequence variation (*Figure 38b*), HMBC was omitted, and the ^{1}H - (^{12}C) magnetization pool was

shared between ZQ (first module) and 2Q (second module) SEA XLOC experiments. Since the former module is more sensitive, a β^I flip angle of 150° was chosen in the first BANGO to balance the signal-to-noise ratio of the obtained spectra (*Figure 39b*).

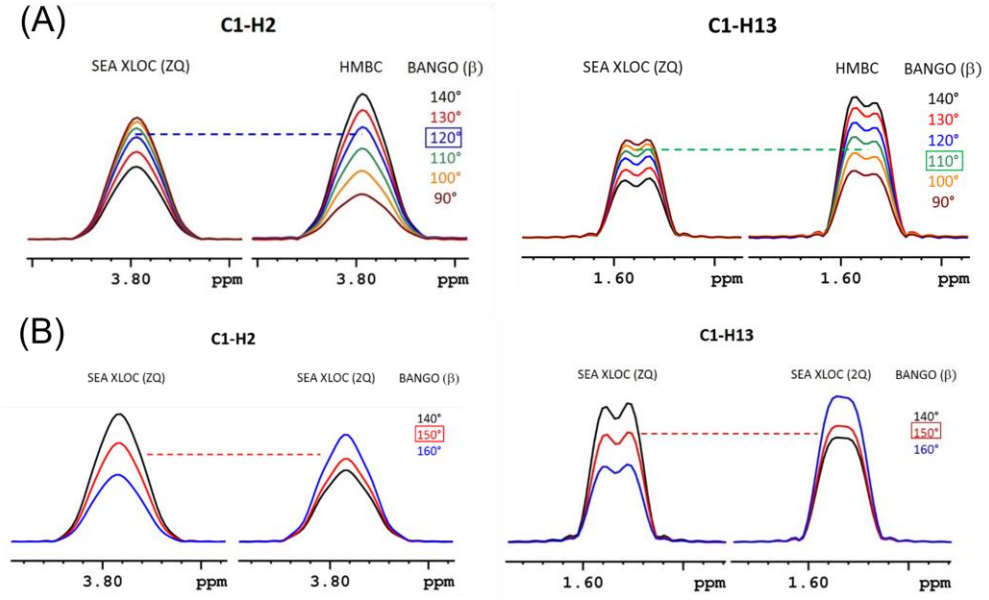


Figure 39. *β^1 angle optimization of the first BANGO element of the combined sequences (Figure 38A, B). The intensity changes of C1-H2 two- and C1-H13 three-bond correlations of ibuprofen was monitored. If the BANGO β^1 angle is around 90°, the second HMBC module will suffer from significant sensitivity loss (peak intensities decrease). To ensure that enough magnetization is left for HMBC, a β^1 angle around 110° - 120° was found to be a good compromise (A). SEA XLOC 2Q and ZQ differ in sensitivity; thus, optimization must be carried out, resulting in a β^1 angle of 150° to be optimal (B).*

To improve further the sensitivity of the experiments, adiabatic carbon inversion pulses were introduced to handle the B_1 field inhomogeneity and the frequency offset effects. Unlike amplitude modulated RF hard pulses, this particular class of pulses slowly sweeps the desired frequency range, allowing to manipulate the carbon spins more accurately[80].

In the above-mentioned BANGO experiments, all ^{13}C inversion pulses were replaced by an adiabatic counterpart to eliminate the offset effects and to

compensate for the spread of $^{1}J_{CH}$, as suggested by Kupče in an earlier study[81]. The 1D test experiments with BANGO excitation on [^{13}C-1]Me-α-D-Glcp reported promising results. Performing the experiments with the conventional hard pulses (*Figure 40*), significant intensity drops were observed, when intentional mismatches in $^{1}J_{CH}$ (left, - 40%) or in the offset (right, - 57%) were applied. However, the experiments performed with *J*-compensated adiabatic carbon inversion pulses (CA-WURST-20[82]) suffered only 5% intensity loss in the worst-case scenario. In addition, the other magnetization pool (^{1}H attached to ^{13}C) was not affected by the adiabatic inversion pulses and remained intact to be exploited by the last H2OBC module.

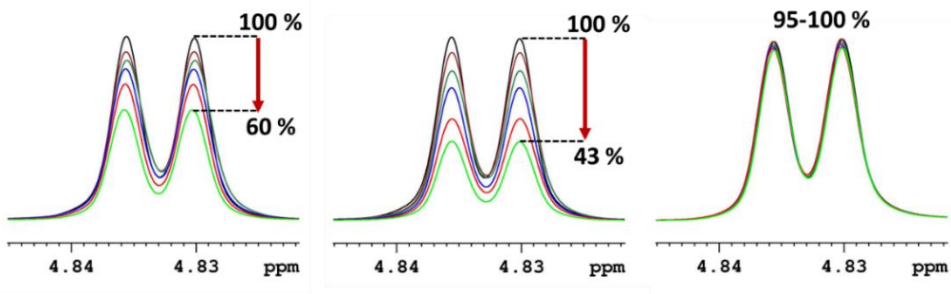


Figure 40. *The performance of adiabatic carbon inversion pulses compared to experiments ran with conventional hard pulses. The intensities of [^{13}C-1] Me-α-D-Glcp anomeric proton peaks (in D_2O) were compared. Left: $^{1}J_{CH}$ mismatch dependence of hard carbon pulse (120 - 170 Hz). Middle: ^{13}C-offset dependence of hard carbon pulse (0-100 ppm). Right: the $^{1}J_{CH}$ mismatch and the ^{13}C-offset dependence of J-compensated adiabatic (CA-WURST-20) carbon pulse in one plot.*

The performance of double BANGO experiments was evaluated on prednisolone. The spin systems of this steroid medication (see structure below) feature various carbon multiplicities and overlapping peaks in the spectra, making the resonance assignment challenging. The role of SEA XLOC was to identify the two- and three-bond correlations by comparing the linewidths in the two subspectra, as shown with the C5-H7a and C20-H21a correlations of C5 and C20 quaternary carbons, respectively (*Figure 41*).

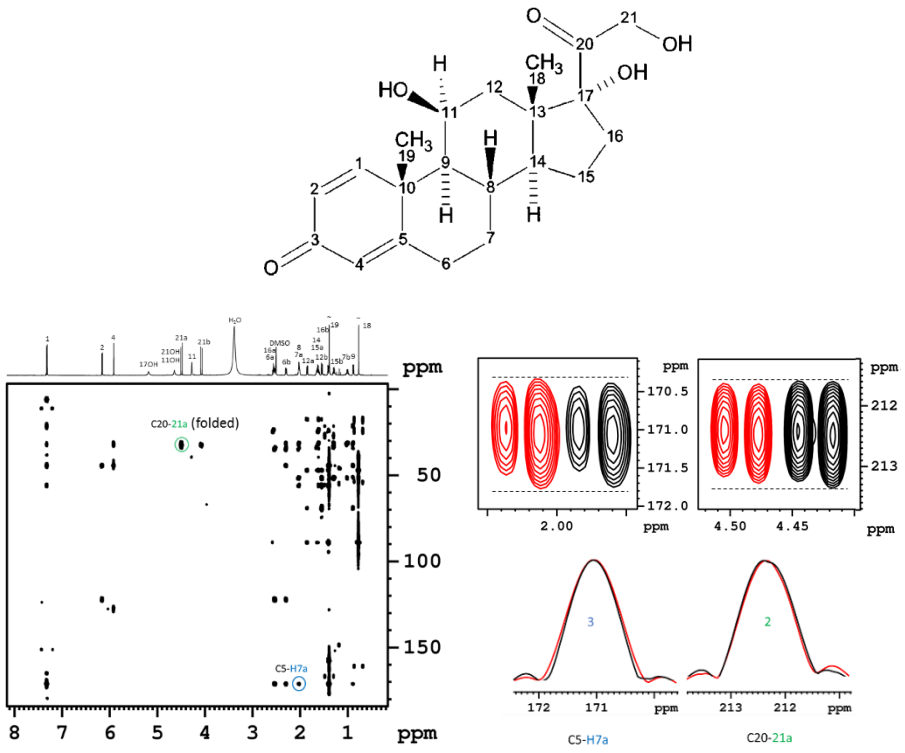


Figure 41. *Three- (C5 - H7a) and two-bond (C20 - 21a) ^{13}C-^{1}H correlations can be distinguished with the SEA XLOC experiment, by comparison of the line width of ZQ/2Q peaks in the F1 carbon dimension. Spectra were extracted from the BANGO {SEA XLOC}-{HMBC}-{H2OBC} recorded on prednisolone (Table 3(16)).*

It should be noted that the two- and three-bond correlations of the protonated carbons, like C15 and C7 can be also distinguished with the combined analysis of HMBC and H2OBC spectra (*Figure 42*). The green H2OBC spectrum presents the one- and two-bond heteronuclear correlations, while the black HMBC spectrum conveys all multiple bond correlations. Accordingly, the peaks overlapping in the two spectra correspond to two-bond correlations, while the black (non-overlapped) HMBC peaks identify the three- (or more) bond correlations.

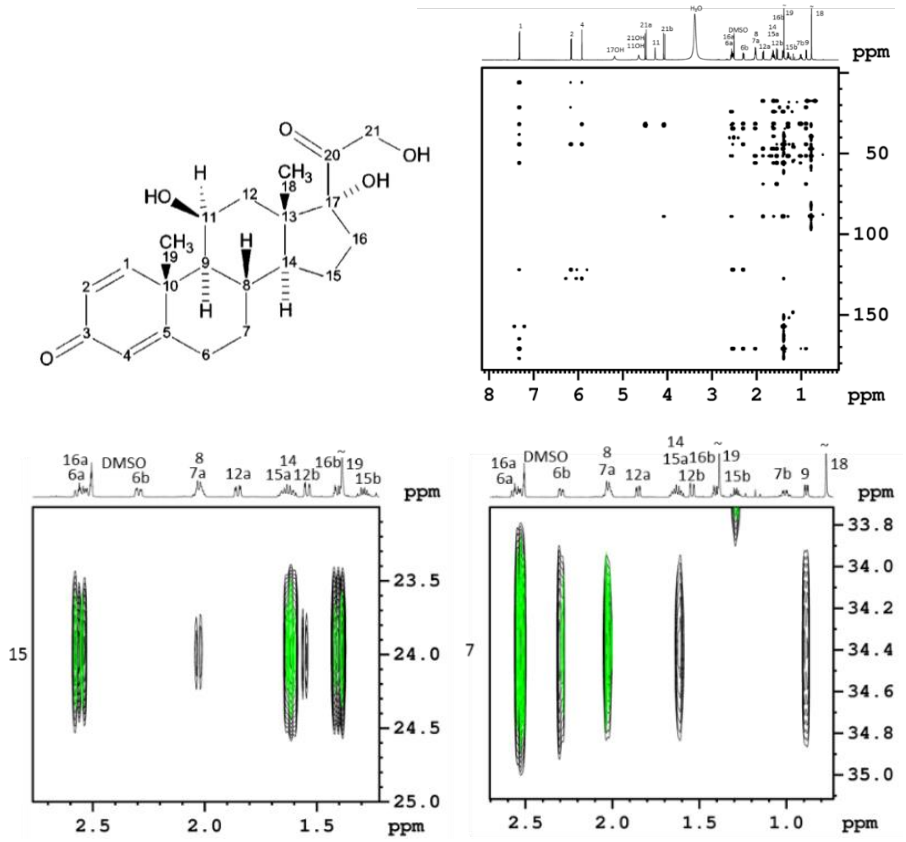


Figure 42. *HMBC (top) and excerpts from H2OBC overlapped with the corresponding HMBC spectra (bottom) of prednisolone recorded with the BANGO {SEA XLOC}-{HMBC}-{H2OBC} experiment using the parameters as given in Table 3(16).*

Phase editing of peaks was also utilized to identify the ^{13}C-^{1}H correlations in crowded, complex spectral regions. Modifications in the H2OBC pulse sequence were carried out including two or four editing cycles, providing distinct subspectra with additional information on carbon multiplicity and/or correlation types (two- vs. one-bond). *Figure 43* presents color coded plots of one-bond correlations of carbons with even *(A)* and odd multiplicity (*B*), two-bond correlations with even (*C*) and odd multiplicity (*D*). As a result, the corresponding unedited H2OBC spectrum (*E*) with all observable correlations in one plot was simplified, and peaks could be distinguished and assigned correctly in the resulting 2BOB subspectra.

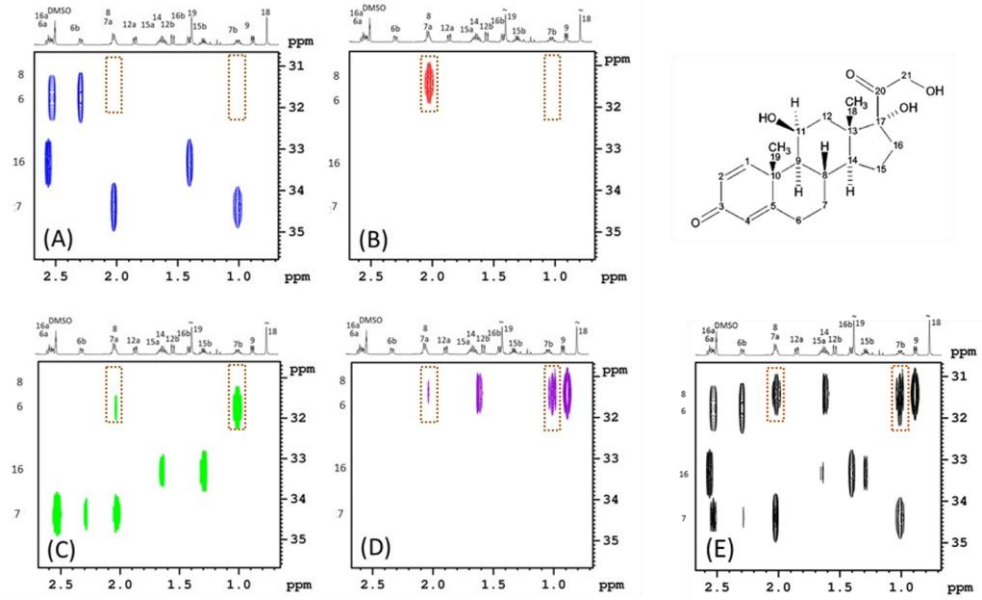
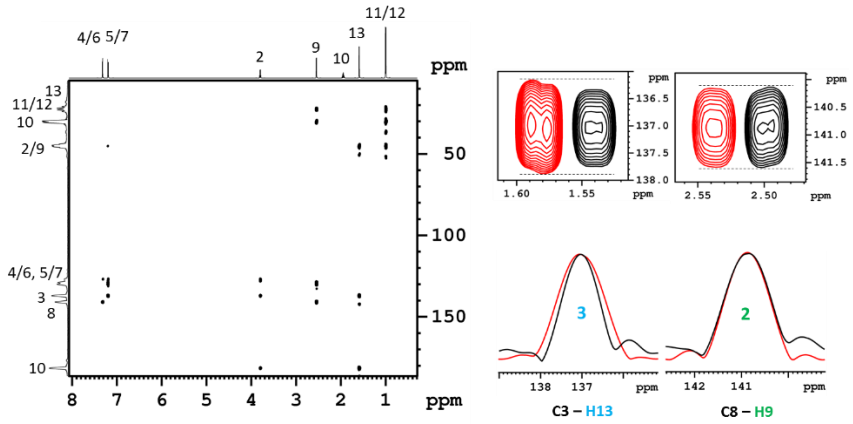


Figure 43. *Four subspectra extracted from the fully edited 2BOB module of the BANGO {SEA-XLOC}-{HMBC}-{2BOB} (Figure 38b) experiment performed on prednisolone with the setup in Table 3(16). The one- and two-bond correlations of C6, C7, C16 (A,C, even multiplicity) and the one- and two-bond correlations of C8 (B,D, odd multiplicity) are displayed. In E the same excerpt of the unedited H2OBC is shown with all correlations in one map.*

The performance of the three-module experiment was also tested on ibuprofen, the compound used in earlier single BANGO experiments. The assignment of two- and three-bond correlations are demonstrated on C8 - H9 and C3 - H13 peaks (*Figure 44A*). The 'assignment walk' along the C9 - C10 - C11/C12 carbon chain shows the connection of neighboring protons (horizontal lines) and neighboring carbons (vertical lines) in the H2OBC excerpt (*Figure 44B*).

(A)



(B)

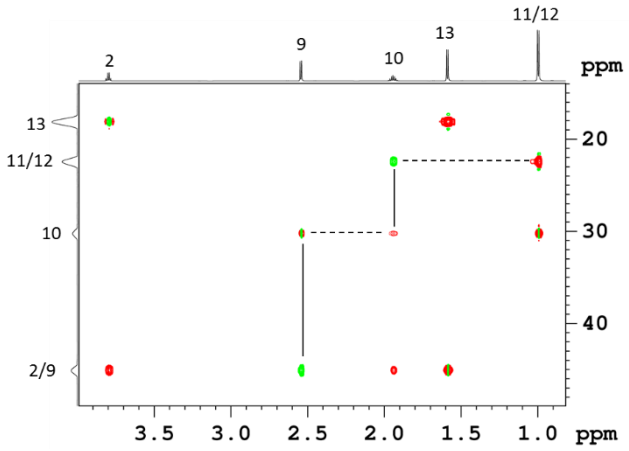


Figure 44. *SEA XLOC(ZQ) full spectrum (A) and excerpt of H2OBC (B) of ibuprofen recorded with the three-module BANGO {SEA XLOC (ZQ)}-{SEA XLOC(2Q)}-{H2OBC} experiment (Figure 38B) with the parameters in Table 3(15).*

5.4 NORD: (No Relaxation Delay) Spectroscopy

It is possible to run concatenated NMR experiments without relaxation delays if the magnetization share between the modules is well-balanced and there is enough time for the recovery of the unused reservoirs while the other experiments are running. In a two-module example, the exhausted magnetization of the first experiment relaxes back to equilibrium while the second is running and vice versa. In the previous chapters, the pool of *I* spins (^1H not attached to ^{13}C) was separated from the *IS* pool (^1H attached to ^{13}C) by the BANGO element. The latter was exploited by the last (H2OBC/2BOB) experiment, which saturates all the available magnetization (both *I* and *IS*); therefore, a recovery delay was needed to restart the sequence. The modification of the excitation module of the H2OBC/2BOB by a general BIRD (**Bi**linear **R**otation **D**ecoupling)[83] element, called BIG-BIRD (**B**iselective **I**ndependent **G**yrations BIRD)[84] leaves the *I* pool invariant and also saves some of the *IS* magnetization. In total, the BIG-BIRD acts differently on the two pools by a $(\pi/2)_y$ on the *I* and $(\pi/2)_v$ pulse on the *IS* spins, with an appropriately adjustable difference in the pulse phases. The Ernst-angle concept determines the extent of phase *v*; if it is set to 90°, corresponding to $(\pi/2)_y$, the result will be identical on both spin pools, *i.e.* no distinction is made. If $v < 90°$, then independent rotations are achieved: *I* spins remain invariant, but the *IS* pool becomes divided. One portion of the magnetization (in the x-y plane) is consumed by the experiment, while the other (first in –z, then flipped back to +z) is stored for later usage, allowing complete omission of the relaxation delay in the NORD spectroscopy. The phase shift angle *v* must be optimized prior to the experiment for the best sensitivity.

The application of NORD spectroscopy requires the replacement of the initial 90° excitation pulse of the 2BOB/H2OBC module by the BIG-BIRD element. The schematic pulse sequence (*Figure 45, left*) is selective on the *I* and *IS* pools with the incorporation of two pulses with different phases: $(\pi/4 - v/2)_{\varphi4}$ and $(\pi/2)_{\varphi5}$, respectively. Same as before in the BANGO experiments, an adiabatic pulse is applied in the ^{13}C channel for better sensitivity. The selectivity on different pools and the *v* angle dependence was demonstrated on a sample of $^{13}CH_3COONa$\$^{12}CH_3COONa$ in D_2O, where the

middle resonance of the *I* pool ($^{12}CH_3$) remains untouched, while *IS* ($^{13}CH_3$) peaks show phase shift as a function of the *v* angle (*Figure 45, right*).

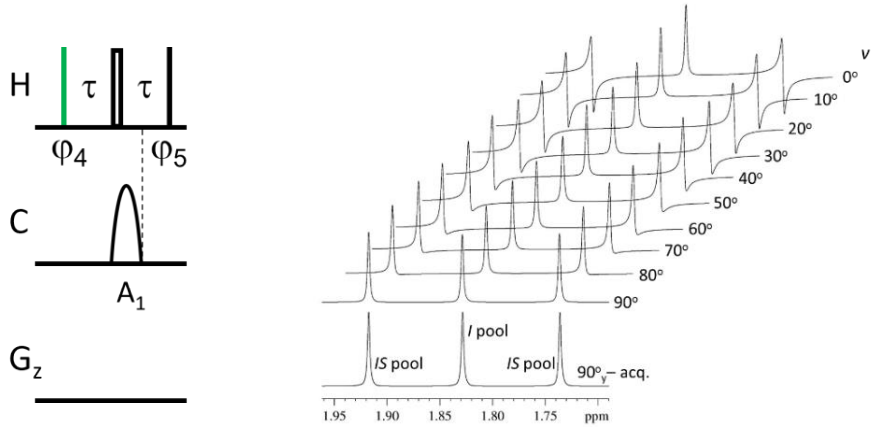
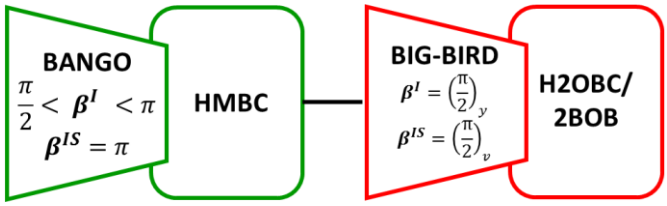


Figure 45. *The schematic pulse sequence of the BIG-BIRD element (left). H: ^{1}H, C: ^{13}C, G_z: gradient channel. A1: adiabatic carbon 180° pulse. The narrow-filled green pulse is (π/4- v/2) with phase φ4 = (π/4- v/2), while the black pulse is π/2 with phase φ5 = (5π/4+ v/2)., The open pulse is π. A series of 1D ^{1}H spectra are demonstrating how BIG-BIRD selectively handles the magnetic reservoirs of $^{12}CH_3$ and $^{13}CH_3$. The test sample was a 2:1 mixture of $^{13}CH_3COONa$ and $^{12}CH_3COONa$ dissolved in D_2O, measured on a 700 MHz Bruker spectrometer.*

In the first application of the NORD technique, H2OBC/2BOB was combined with HMBC in a two-module, and with SEA XLOC and HMBC in a three-module experiment (*Figure 46*). Every individual sequence began with a unique, isotope selective excitation instead of a non-selective 90° hard pulse. Similarly to the previous NOAH-type measurements, two BANGO elements were applied to prepare the HMBC and SEA XLOC module sharing the *I* (^{1}H not attached to ^{13}C) and the inversion of *IS* (^{1}H attached to ^{13}C) magnetization reservoirs. For optimal sensitivity, the β_1^I and β_2^I excitation angle of the BANGO elements should be optimized prior to measurements (Ernst-angle concept), just like previously (*Figure 39*). The 90° excitation pulse of H2OBC/2BOB was replaced by the novel BIG-BIRD element with the optimized *v* angle to partially exploit the so far untouched *IS* pool and save some of its magnetization for the consecutive transient. With these spin manipulations, the NORD sequences can be run without any recovery delay between the scans utilizing efficient magnetization share and relaxation, and consequently, a huge

amount of experimental time can be saved as will be demonstrated in the next section.

(A)



(B)

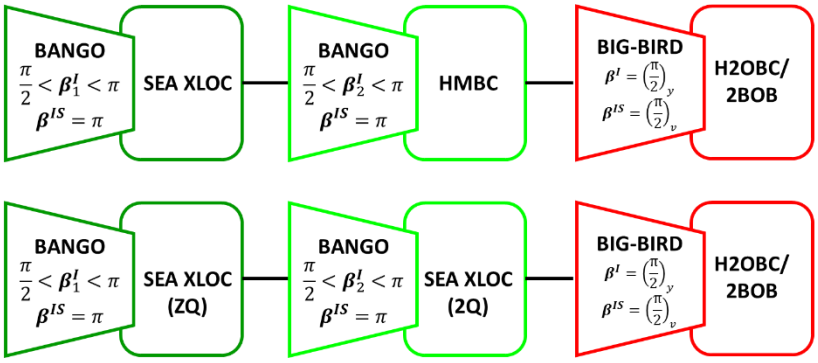


Figure 46. *Schemes of two-module (A) and three-module (B) NORD experiments. The excitation of different magnetization pools is carried out by the optimized BANGO and BIG-BIRD elements.*

Sensitivity comparisons were carried out between the individual HMBC and H2OBC versus NORD {HMBC}-{H2OBC} measurements on a trisaccharide *(Table 3[17])*. The experimental conditions were kept identical except that no relaxation delay was used in NORD. With the relative sensitivity defined as the signal-to-noise ratio per unit time, one scan per t_1 increment was set in the standalone measurements, while the NORD {HMBC}-{H2OBC} was run with 8 scans to reach nearly identical measurement times. Thus first, a 'downscaling' of NORD peak intensities was carried out by the factor of $\sqrt{\frac{1}{8}}$. Then a factor of $\sqrt{\frac{36}{43}}$ was used for correcting the difference in the measurement

times, since the separate HMBC and H2OBC ran for 36 min (18 min each), while the NORD version was 43 min. After careful corrections, an enhancement of 6 - 31% was observed in the least sensitive HMBC module, while the H2OBC showed even larger 22 - 165% enhancement, providing clear evidence for the superior performance of the NORD acquisition strategy (*Figure 47*). The sensitivity improvements for the different peaks are not identical because there is a spread in the pertinent proton T_1 relaxation times (0.9 - 2.3 s). The largest sensitivity enhancement was observed with the most slowly relaxing protons (*e.g.*, D1: 1.53 s, 2.43 and F1: 2.06 s, 2.21).

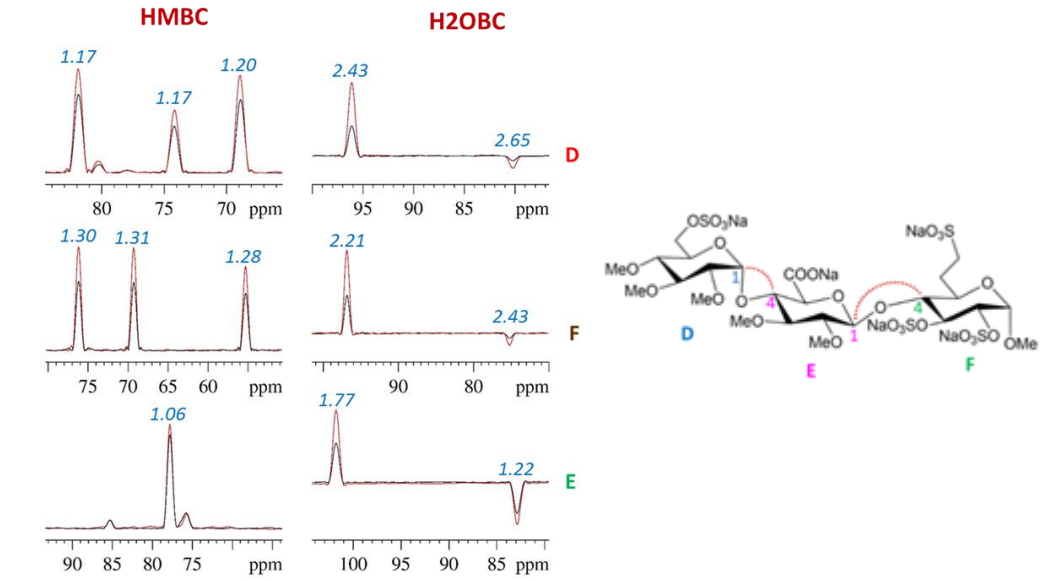


Figure 47. *Sensitivity comparison of the standalone HMBC and H2OBC measurements (black) with their implementation in NORD {HMBC}-{H2OBC} (red) recorded on a trisaccharide with the parameters as given in Table 3(17). On the 1D traces extracted from the anomeric proton region, relative intensities are shown in blue.*

The NORD {HMBC}-{H2OBC} experiment provides all long-range and one-bond correlation data, and distinction between two- and three-bond heteronuclear couplings of protonated carbons allowing the complete and unambiguous proton- and carbon resonance assignment of a saccharide with five sugar units. The measurement on the 24 mg pentasaccharide dissolved in 550 µl D_2O took only 6 minutes at 700 MHz. The 'assignment walk' in the spin systems (residues labeled with different colors in H2OBC) and the interresidual connections (HMBC) were identified and are presented in (*Figure 48*).

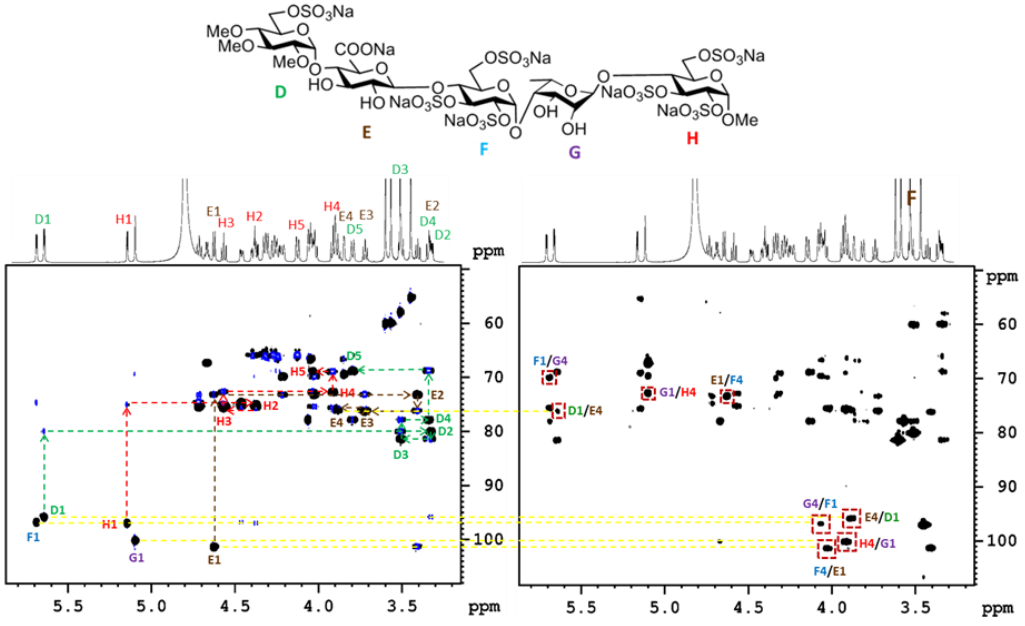


Figure 48. *H2OBC (left) and HMBC (right) spectra obtained from the two-module NORD {HMBC}-{H2OBC} measurement of a pentasaccharide in D_2O with the experimental setup in Table 3(18). Resonance assignment was carried out following the HSQC (one-bond) and COSY-type (two-bond) spectral information of H2OBC, while the interresidual connections were identified by the three-bond heteronuclear correlations detected in HMBC.*

After testing the two-module experiment, the NORD concept was introduced to more complex sequences. Concatenation of the {SEA XLOC}-{HMBC}-{H2OBC/2BOB} experiments was carried out by sharing the ^{1}H - (^{12}C) magnetization pool with BANGO excitations between the first two modules. Until the starting point of the H2OBC module, the pulse sequence is identical to the earlier NOAH-type double BANGO sequences. The difference arises with the replacement of the first proton 90° excitation pulse of the H2OBC/2BOB module with the BIG-BIRD element, allowing the experiment to run without relaxation delays. To demonstrate the performance of the three-module NORD experiments, hydrocortisone was chosen as a model compound featuring complex spectra with significant overlaps both in the proton and carbon dimensions as shown by the HMBC map from the three-module NORD {SEA XLOC}-{HMBC}-{2BOB} experiment (*Figure 49A*). All of the overlapping resonances could be completely resolved in the four-step, fully edited 2BOB experiment (*Figure 49C*). In the resulting four subspectra, the distinction of the peaks was based on the carbon multiplicity and the number of intervening bonds. The unambiguous assignment of all protonated carbons and protons was carried out by the 'assignment walk' strategy following the color codes: one-bond (1B) correlations of CH_2 and *CH* carbons with green and blue, the corresponding two-bond (2B) correlations with red and brown color label. For the assignment of the remaining quaternary carbons, the comparison of HMBC and SEA XLOC (*Figure 49A,B*) spectra was carried out. With an overall 42 minutes of measurement time, high-quality data could be obtained for the complete assignment of the molecule.

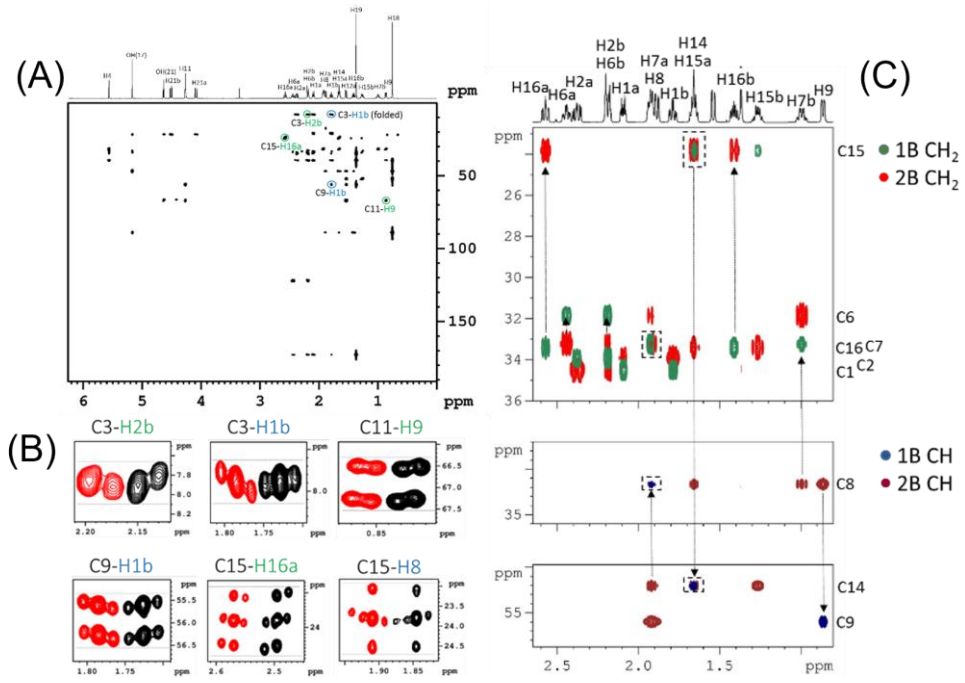


Figure 49. *HMBC (A), SEA XLOC (B) and fully edited 2BOB (C, subspectra are overlayed for demonstration of 'assignment walk') spectra of hydrocortisone obtained from the NORD {HMBC}-{SEA XLOC}-{2BOB} experiment with the parameter set in Table 3(19)*.

5.5 Unique data collection and processing

In the concatenated NMR sequences, multiple experiments were recorded sequentially, requiring a complex data collection and analysis technique. To gain the individual spectral information, the *splitx* au-program of the Bruker TopSpin software was used to separate the acquired data into blocks, corresponding to the individual experiments, which can be further processed in the conventional way. Regarding the most complex, three-module NORD {SEA XLOC}-{HMBC}-{2BOB} experiment (*Figure 50*), the processing also starts with the *splitx* algorithm to separate the three data blocks. Then the second HMBC data block can be processed in the standard way. The first SEA XLOC block integrates ZQ and 2Q data, while the last 2BOB includes complex data

due to the four-step time-shared editing. Therefore, both require further separation by the *split* au-program script. Lastly, the four 2BOB subspectra can be obtained with proper editing (adding/subtracting) of the splitted data sets. If the acquisition was carried out with non-uniform sampling (NUS) of the data points, the reconstruction had to be carried out prior to the standard processing. Momentarily, the processing of the concatenated experiments is carried out manually and requires some expertise, but in the future, it will be automatized with user-friendly scripts, integrating the different steps into one simple au-command. This will expedite the routine and widespread applications of these experiments, providing well-resolved spectra for analysis in simple steps and in a matter of minutes.

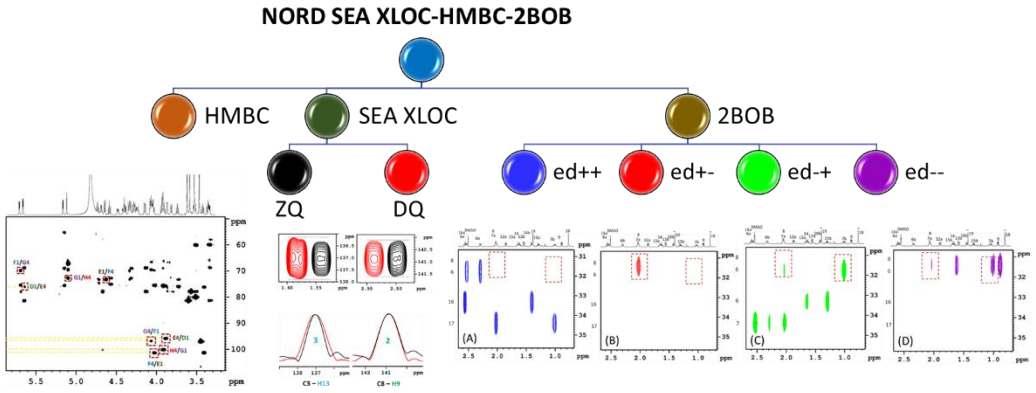


Figure 50. *The data structure of the 3-module NORD {SEA XLOC}-{HMBC}-{2BOB} experiment with the resulting spectra with color coded information. Regarding 2BOB, the edited subspectra are labeled with ± signs: ^{13}C-multiplicity editing on/off (+/-), then HMQC/COSY type editing on/off (+/-).*

6. Summary

In the first part of the dissertation, it was presented how NMR spectroscopy and *in silico* methods support the study of molecular mechanisms behind cellular processes. The MEF2D transcription factor is a protein with intrinsically disordered regions and has a pronounced role in the regulation of muscle cell differentiation (myogenesis). The Fuxreiter group showed that transcriptional activity can be fine-tuned by changing the dynamical profile of the alternatively spliced β-domain rewiring the MEF2D interaction network with partner molecules. In chapter 4, three mini-proteins were selected for detailed structural and dynamical analysis representing different MEF2D β-domain variants with the most promising biological responses.

In conclusion, mutations decreased the backbone mobility and the degree of disorder in *var8* peptide with point mutations N29D and A31W, evidenced by various parameters compared to the wild type. On the experimental side, enhanced translational diffusion and residue-wise backbone order parameters (S^2) were shown by NMR spectroscopy compared to the other variants. On the computational side, MD simulations provided further details at atomic-level and supported the experimental observations. The increased molecular compactness of *var8* was confirmed by the MD-derived order parameters as well as by the smallest radius of gyration and the larger propensity for secondary structures. Intramolecular structure stabilizing interactions were also verified, including a network of H-bonds, salt bridges, and cation-π effects. A different dynamical picture was obtained for peptides *var3* and *var4* where increased flexibility was evidenced by both experimental and theoretical data. Less compact conformations were indicated by slower translational diffusion coefficients and larger molecular radius. Furthermore, these peptides also feature a higher fraction of unclustered structures, and the populations of clustered conformations are distributed to multiple, similar-sized groups, indicating more structural diversity.

The atomic-level structural and dynamical data obtained on the mini-protein variants together with the cell culture experiments on transcriptional activity disclose the role of the MEF2D β-domain in myogenesis. This structural motif is a crucial site for transactivation, partner protein recognition

and binding. By perturbing structural dynamics, the interaction pattern of the MEF2D is rewired. In the case (*var8*), when targeted point mutations stabilize this intrinsically disordered, special motif, transcriptional activity is enhanced at the early stage of muscle cell differentiation. On the other side, myoblast fusion is slower if higher structural dynamics are introduced (*var3*, *var4*) in comparison to the wild-type MEF2D.

Additionally, diffusion and relaxation data from NMR and independent NanoDSF measurements identified an oligomerization equilibrium in *var8* solution, which was confirmed and studied in detail by MD simulations. Only this slightly more rigid variant established higher-order protein assemblies stabilized via intermolecular interactions, such as cation-π electrostatic bonding and the formation of hydrophobic cores. In transcription factor hubs, with raised local protein concentration, the transcription efficiency is facilitated[85]. This mechanism is very likely exploited by MEF2D *var8*, however, further experiments are needed to understand the biological background.

In the second part of this work, novel combined NMR pulse sequences were introduced and their applications were demonstrated in detail.

All NMR applications require the unambiguous resonance assignment of the investigated molecule, and for this purpose, a series of measurements and strategies are available for spectroscopists and organic chemists. The experimental modules built-in in the devised experiments are the advanced versions of recent HMBC and HMQC type sequences. SEA XLOC, H2OBC, and their edited versions deliver rich spectral information on one- and multiple bond heteronuclear correlations including their distinction, and on additional COSY-type (two-bond) connectivities.

In principle, the Ernst-angle concept correlates the optimum excitation angle with respect to the pertinent longitudinal relaxation times (T_1) to obtain the best sensitivity within reasonable measurement time. The concept is providing the basis for the time-efficient combined experiments, however, as simple as it looks, applying it to complex NMR pulse sequences with a series of delays and pulses is a puzzling task. During my graduate years, I have participated in method development projects aiming to find solutions for circumventing this problem by efficient magnetization sharing methods.

In the NOAH-type measurements, efficient spin manipulations were achieved by replacing the relevant hard pulses with the BANGO excitation module, splitting the available ^1H magnetization and distributing them between the experiments. With only one recovery delay at the beginning of the series of the concatenated modules, BANGO {SEA XLOC}-{H2OBC) was performed on an ibuprofen sample in 42 minutes. 59 % of instrumental time was saved compared to the standalone SEA XLOC and H2OBC measurements (102 minutes together). The same improvement was shown with BANGO {HMBC}-{H2OBC}, where correlation maps were obtained within 11 minutes, by replacing SEA XLOC with the more sensitive HMBC. 5 minutes were required to record BANGO {HMBC}-{H2OBC} on another sample (trisaccharide) with non-uniform sampling (NUS).

Further improvement was achieved with the double BANGO pulse schemes with more efficient magnetization usage, enabling the introduction of a third experimental module in the sequence. BANGO {SEA XLOC}-{HMBC}-{H2OBC} and BANGO {SEA XLOC(ZQ)}-{SEA XLOC(2Q)}-{H2OBC} utilized well-optimized isotope selective Ernst-angle excitation pulses and have been tested on the correlation maps of ibuprofen and prednisolone compounds.

The most time-efficient solution was the NORD spectroscopy technique, first published by our group. The selective handling of different magnetization components was carried out by the BANGO and the BIG-BIRD modules, where the latter one allows the desired phase manipulation of the excitation pulses. After appropriate, Ernst-concept based optimization of pulse angles, phases and sequential ordering of the individual experiments, complex and time-efficient combined pulse sequences were established without the need of magnetization recovery delays. NORD {HMBC}-{H2OBC} with NUS was measured on trisaccharide and pentasaccharide samples in less than 2 and 6 minutes, respectively. The three-module NORD {SEA XLOC}-{HMBC}-{2BOB} was acquired in 42 minutes on hydrocortisone, providing sufficient data for the complete NMR assignment of the molecule.

NORD provides a pulse sequence design framework for a novel family of efficient NMR measurements. Further work is going on in our lab to include other experiments as well, since in principle, no limitations exist. However, to efficiently tap the magnetization reservoirs and distribute polarization between

experiments, other excitation blocks must be considered beyond the BANGO and BIG-BIRD pulses. To date, the introduced NORD measurements offer one of the fastest ways to obtain ^{13}C - ^{1}H and ^{1}H - ^{1}H correlation information, but it is worth noting that this pulse sequence design strategy can be easily extended to additional, multinuclear and multidimensional NMR methods on a broader scale.

7. Összefoglalás

A doktori munkám első fejezetében megmutattam, hogyan járulhat hozzá az NMR-spektroszkópia és a számításos kémia biológiai folyamatok molekuláris szintű megértéséhez. A MEF2D transzkripciós faktor egy fehérje, amely kiterjedt rendezetlen régiókat tartalmaz és fontos szabályozó szerepet tölt be az izomsejtek differenciációs folyamataiban (miogenezis). Együttműködésben a Fuxreiter csoporttal, azt terveztük bizonyítani, hogy az alternatív hasítással szabályozott β-alegység dinamikai tulajdonságainak módosításával finomhangolható a transzkripciós aktivitás. A 4. fejezetben bemutattam három mutáns MEF2D β-alegység mini-fehérje szerkezeti és dinamikai vizsgálatát, melyeket a legígéretesebb biológiai eredmények alapján választottunk ki.

Összességében a vad-típusú modellel összehasonlítva az N29D és A31W pont-mutációk csökkentették a főlánc mobilitását és a rendezetlenség fokát a *var8* fehérje esetében. Az NMR kísérletek megnövekedett transzlációs diffúziót és rendparamétereket (S^2) jeleztek a többi variánshoz képest. Az MD számítások támogatták a kísérleti megfigyeléseket és további atomi szintű szerkezeti finomításokat biztosítottak. A *var8* fehérje megnövekedett kompaktságát megerősítették az MD trajektoriából számított rendparaméter értékek, a kisebb girációs sugár és a másodlagos szerkezeti elemek kialakítására való fokozott hajlam. Szerkezet stabilizáló kölcsönhatásokat is felderítettem, többek között H-hidak hálózatát, sóhidakat és kation-π effektusra utaló jeleket is. A *var3* és *var4* mutánsok esetében eltérő dinamikai képet kaptam, megnövekedett flexibilitásra utaltak a kísérleti és a számításos adatok is. A kevésbé kompakt konformáció nagyobb molekulasugárral és lassabb transzlációs diffúziós sebességgel társult. A fehérjék MD sokaságában a legtöbb

konformációt különbözőségük miatt nem lehetett csoportosítani, a klaszterezett szerkezetek pedig több, hasonló nagyságú csoportba kerültek, ez pedig jelentősebb szerkezeti sokszínűségre utal.

A mini-fehérjék részletes szerkezeti és dinamikai elemzése kiegészítve a sejtes vizsgálatok eredményeivel megerősíti a MEF2D β-alegységének miogenezisben betöltött szabályozó szerepét. Ez az alegység a transzkativálásért felelős a partnerfehérjék felismerésén és kötésén keresztül. Ebben a folyamatban szerepet játszik a szerkezeti flexibilitás is, melynek megváltoztatásával áthangoltuk a MEF2D kölcsönhatási mintázatait. A *var8* fehérje célzott mutációi dinamikailag stabilizálják a β-alegységet és környezetét és ezáltal fokozott transzkripciós aktivitáshoz vezetnek az izomsejtek korai differenciációs fázisában. Ezzel szemben, a sejtek mioblasztokká történő egyesülése lassul a β-alegység fokozott szerkezeti dinamikájának következtében (*var3, var4*), a vad-típusú MEF2D hatásához képest.

Az NMR diffúziós és relaxációs, valamint a NanoDSF mérések oligomerizációs egyensúlyt mutattak a *var8* fehérje oldatában. Ezt a jelenséget a molekuladinamikai számítások is megerősítették és árnyalták. Csupán ez a kompaktabb szerkezetű variáns alkotott fehérje asszociátumokat, melyek intermolekuláris kölcsönhatásokkal stabilizáltak (kation-π, hidrofób magok képződése). Az irodalomban ismert a transzkripciós faktorok magasabb rendű fehérjeszerveződése, mely fokozott transzkripciós hatékonyságot eredményez. Feltételezzük, hogy a MEF2D *var8* mutáns működése hasonló molekuláris mechanizmust követ, azonban a biológia háttér megerősítéséhez további kísérleteket szükséges elvégezni.

A disszertáció második nagy fejezetében a kutatócsoportunk által az elmúlt években kifejlesztett kombinált NMR kísérletekről és alkalmazási lehetőségeikről számoltam be.

A rutin NMR módszerek alkalmazása során szükség van a vizsgált molekula teljes jelhozzárendelésének ismeretére, ehhez számos jól bevált kísérlet és stratégia áll a spektroszkópusok vagy a szerves kémikusok rendelkezésére. A dolgozatban bemutatott új NMR módszerek a széles körben alkalmazott HMBC és HMQC kísérletek továbbfejlesztett változatai. Név szerint a SEA XLOC és a H2OBC szekvenciák, valamint ez utóbbiak

szerkesztett (editált) formái az egy- és többkötéses heteronukleáris korrelációkról és egyértelmű megkülönböztetésükről, valamint COSY-típusú proton-proton konnektivitásokról biztosítanak bőséges spektrális adatokat. Munkánk során az NMR mérések idejét döntően meghatározó szekvenciaelemet, a mágnesezettség egyensúlyi állapotba való visszatéréséhez szükséges relaxációs időt optimalizáltuk, illetve próbáltuk különböző technikákkal lecsökkenteni.

Az új kísérletek alapja az Ernst-szög elv, mely egyszerűen a gerjesztési szög optimalizálását jelenti a vizsgált molekula spin-rács relaxációs (T_1) idejének figyelembevételével, mely időtakarékos méréseket tesz lehetővé. Ez az elv egyszerű NMR kísérleteknél könnyen alkalmazható, azonban komplex mérési szekvenciák esetében kihívást jelentő feladat. Doktori munkám során több ilyen módszerfejlesztési feladatban vettem részt, melynek során hatékony mágnesezettség megosztási eljárásokat dolgoztunk ki.

A NOAH-típusú mérésekben a ‘kemény’, nem-szelektív gerjesztő impulzusokat BANGO izotóp-szelektív szekvenciaelemekre cseréltük le és ezzel a hatékony spin-manipulációval osztottuk fel a minta mágnesezettségét az egy szekvenciába felfűzött kísérletek között. Az általunk kifejlesztett BANGO {SEA XLOC}-{H2OBC} kétmodulos kísérlet egyetlen relaxációs várakozási idővel fut, melyet egy ibuprofen mintán teszteltünk le először egy gyors, 42 percet igénybe vevő méréssel. A kísérleti modulokat külön-külön használva 102 percre lett volna szükség. A kombinált szekvenciával így 59%-ot nyertünk a mérési idő tekintetében. Hasonlóan ígéretes eredményeket értünk el a BANGO {HMBC}-{H2OBC} kísérlettel ugyanezen a mintán. A SEA XLOC modult a sokkal érzékenyebb HMBC-re cserélve 11 perc alatt jutottunk korrelációs térképekhez. Nem-lineáris mintavételezési technikával (NUS) ugyanezen kísérlet lefutásához csupán 5 perc időre volt szükség egy tömény triszacharid mintán.

A bíztató eredmények után kombinált kísérleteinket három modulossá bővítettük. Ehhez a mágnesezettség takarékos használatát két különböző BANGO elem beépítésével oldottuk meg. A bemutatott BANGO {SEA XLOC}-{HMBC}-{H2OBC} és {SEA XLOC(ZQ)}-{SEA XLOC(2Q)}-{H2OBC} mérések optimalizált Ernst-szöggel gerjesztő, izotóp-szelektív

impulzusokat használ. Az új kísérletek alkalmazhatóságát ibuprofen és prednizolon mintákon mutattuk be.

Végül az általunk bevezetett NORD spektroszkópia alkalmazásával elértük a kísérletek közötti várakozási idő teljes kiküszöbölését, így a mérések idejét döntően meghatározó szekvenciaelemtől sikerült teljesen megszabadulnunk. A különböző mágnesezettségi komponensek izotóp- és egyben kívánt fázissal történő szelektív gerjesztését a BANGO és a BIG-BIRD elemek beépítése tette lehetővé. A gyors és hatékony, de ugyanakkor rendkivül komplex kísérletek tervezésének alapja egyrészt az alkalmazott kísérletek megfelelő sorrendjének meghatározása, másrészt az Ernst-szög elv alkalmazásával az optimalizált gerjesztési szögek és fázisok megtalálása. Ezzel a hatékony mágnesezettség használattal a különböző kísérleti modulok különböző mágnesezettségi tartályból táplálkoznak. Ezen felül az aktív kísérleti modul ideje alatt az éppen nem használt (passzív) mágnesezettségi komponens hatékonyan relaxálódik. A kétmodulos NORD {HMBC}-{H2OBC} kísérlet kevesebb, mint 2 illetve 6 perc alatt futott le egy triszacharid és egy pentaszacharid modellvegyületen. A hidrokortozin molekula jelhozzárendelése a bonyolult spinrendszerek és jelátfedések miatt kihívást jelent. Ezt a folyamatot könnyítette meg az általunk kifejlesztett három modulos NORD {SEA XLOC}-{HMBC}-{2BOB} kísérlet, mely segítségével 42 perc mérési idő alatt jutottunk a kiértékeléshez szükséges spektrális adatokhoz.

A NORD spektroszkópia egy új kísérlettervezési koncepciót, keretrendszert jelent, mellyel az NMR kísérletek egy egész új családját lehet létrehozni. Elviekben nem létezik limitáció az alkalmazható kísérlettípusok terén, kutatócsoportunkban folyamatosan zajlik az újabb mérési szekvenciák fejlesztése. A különböző mágnesezettségi komponensek kihasználására és a kísérletek közötti polarizáció megosztásához újabb speciális gerjesztési elemek tervezése szükséges a munkában bemutatott BANGO-n és BIG-BIRD-ön túl. Az eddig kifejlesztett NORD kísérletek ^{13}C - ^{1}H és ^{1}H - ^{1}H korrelációs adatok begyűjtésére alkalmasak, és az elérhető NMR módszerek közül a leggyorsabbak között foglalnak helyet. Az általános kísérlettervezési stratégia azonban lehetővé teszi más NMR aktív magok mérését és többdimenziós szekvenciák létrehozását is szélesítve a NORD típusú kísérletek alkalmazhatóságát.

8. Reference list

1 Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223-230, doi:10.1126/science.181.4096.223 (1973).

2 Fischer, E. Einfluss der Configuration auf die Wirkung der Enzyme. II. *Ber. Dtsch. Chem. Ges.* **27**, 3479-3483, doi:https://doi.org/10.1002/cber.189402703169 (1894).

3 Koshland, D. E., Ray, W. J. & Erwin, M. J. Protein structure and enzyme action. *Fed. Proc.* **17**, 1145-1150 (1958).

4 Agarwal, P. K., Billeter, S. R., Rajagopalan, P. T., Benkovic, S. J. & Hammes-Schiffer, S. Network of coupled promoting motions in enzyme catalysis. *Proc. Natl. Acad. Sci. USA* **99**, 2794-2799, doi:10.1073/pnas.052005999 (2002).

5 Eisenmesser, E. Z., Bosco, D. A., Akke, M. & Kern, D. Enzyme dynamics during catalysis. *Science* **295**, 1520-1523, doi:10.1126/science.1066176 (2002).

6 Tompa, P. & Fersht, A. *Structure and Function of Intrinsically Disordered Proteins*. (CRC Press, 2009).

7 Uversky, V. N. A protein-chameleon: conformational plasticity of alpha-synuclein, a disordered protein involved in neurodegenerative disorders. *J. Biomol. Struct. Dyn.* **21**, 211-234, doi:10.1080/07391102.2003.10506918 (2003).

8 Kovács, G. G. *et al.* Natively unfolded tubulin polymerization promoting protein TPPP/p25 is a common marker of alpha-synucleinopathies. *Neurobiol. Dis.* **17**, 155-162, doi:10.1016/j.nbd.2004.06.006 (2004).

9 Sedzik, J. & Kirschner, D. A. Is myelin basic protein crystallizable? *Neurochem. Res.* **17**, 157-166, doi:10.1007/BF00966794 (1992).

10 Cordero, O. J., Sarandeses, C. S., Lopez, J. L. & Nogueira, M. On the anomalous behaviour on gel-filtration and SDS-electrophoresis of prothymosin-alpha. *Biochem. Int.* **28**, 1117-1124 (1992).

11 Oldfield, C. J. *et al.* Comparing and combining predictors of mostly disordered proteins. *Biochemistry* **44**, 1989-2000, doi:10.1021/bi047993o (2005).

12 Deiana, A., Forcelloni, S., Porrello, A. & Giansanti, A. Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell. *PLoS One* **14**, e0217889, doi:10.1371/journal.pone.0217889 (2019).

13 Uversky, V. N., Oldfield, C. J. & Dunker, A. K. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit.* **18**, 343-384, doi:10.1002/jmr.747 (2005).

14 Delacourte, A. & Buée, L. Normal and pathological Tau proteins as factors for microtubule assembly. *Int. Rev. Cytol.* **171**, 167-224, doi:10.1016/s0074-7696(08)62588-7 (1997).

15 Uversky, V. N., Oldfield, C. J. & Dunker, A. K. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys.* **37**, 215-246, doi:10.1146/annurev.biophys.37.032807.125924 (2008).

16 Runnebaum, I. B., Nagarajan, M., Bowman, M., Soto, D. & Sukumar, S. Mutations in p53 as potential molecular markers for human breast cancer. *Proc. Natl. Acad. Sci. USA* **88**, 10657-10661, doi:10.1073/pnas.88.23.10657 (1991).

17 Deng, C. X. BRCA1: cell cycle checkpoint, genetic instability, DNA damage response and cancer evolution. *Nucl. Ac. Res.* **34**, 1416-1426, doi:10.1093/nar/gkl010 (2006).

18 Uyar, B., Weatheritt, R. J., Dinkel, H., Davey, N. E. & Gibson, T. J. Proteome-wide analysis of human disease mutations in short linear motifs: neglected players in cancer? *Mol. Biosyst.* **10**, 2626-2642, doi:10.1039/c4mb00290c (2014).

19 Vacic, V., Uversky, V. N., Dunker, A. K. & Lonardi, S. Composition Profiler: a tool for discovery and visualization of amino acid composition differences. *BMC Bioinformatics* **8**, 211, doi:10.1186/1471-2105-8-211 (2007).

20 Fuxreiter, M. Classifying the Binding Modes of Disordered Proteins. *Int. J. Mol. Sci.* **21**, doi:10.3390/ijms21228615 (2020).

21 Bugge, K. *et al.* Interactions by Disorder - A Matter of Context. *Front. Mol. Biosci.* **7**, 110, doi:10.3389/fmolb.2020.00110 (2020).

22 Breitbart, R. E. *et al.* A fourth human MEF2 transcription factor, hMEF2D, is an early marker of the myogenic lineage. *Development* **118**, 1095-1106 (1993).

23 Martin, J. F. *et al.* A Mef2 gene that generates a muscle-specific isoform via alternative mRNA splicing. *Mol. Cell. Biol.* **14**, 1647-1656, doi:10.1128/mcb.14.3.1647 (1994).

24 Dodou, E., Sparrow, D. B., Mohun, T. & Treisman, R. MEF2 proteins, including MEF2A, are expressed in both muscle and non-muscle cells. *Nucl. Ac. Res.* **23**, 4267-4274, doi:10.1093/nar/23.21.4267 (1995).

25 Al-Khalili, L. *et al.* MEF2 activation in differentiated primary human skeletal muscle cultures requires coordinated involvement of parallel

pathways. *Am. J. Physiol. Cell. Physiol.* **286**, C1410-1416, doi:10.1152/ajpcell.00444.2003 (2004).

26 Kim, Y. *et al.* The MEF2D transcription factor mediates stress-dependent cardiac remodeling in mice. *J. Clin. Invest.* **118**, 124-132, doi:10.1172/JCI33255 (2008).

27 Pon, J. R. & Marra, M. A. MEF2 transcription factors: developmental regulators and emerging cancer genes. *Oncotarget* **7**, 2297-2312, doi:10.18632/oncotarget.6223 (2016).

28 Molkentin, J. D., Black, B. L., Martin, J. F. & Olson, E. N. Mutational analysis of the DNA binding, dimerization, and transcriptional activation domains of MEF2C. *Mol. Cell. Biol.* **16**, 2627-2636, doi:10.1128/MCB.16.6.2627 (1996).

29 Dosztányi, Z., Csizmok, V., Tompa, P. & Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433-3434, doi:10.1093/bioinformatics/bti541 (2005).

30 Zahn-Zabal, M. *et al.* The neXtProt knowledgebase in 2020: data, tools and usability improvements. *Nucleic Acids. Res.* **48**, D328-D334, doi:10.1093/nar/gkz995 (2020).

31 Huang, K. *et al.* Solution structure of the MEF2A-DNA complex: structural basis for the modulation of DNA bending and specificity by MADS-box transcription factors. *EMBO J.* **19**, 2615-2628, doi:10.1093/emboj/19.11.2615 (2000).

32 Zhu, B., Ramachandran, B. & Gulick, T. Alternative pre-mRNA splicing governs expression of a conserved acidic transactivation domain in myocyte enhancer factor 2 factors of striated muscle and brain. *J. Biol. Chem.* **280**, 28749-28760, doi:10.1074/jbc.M502491200 (2005).

33 Lee, K. S. *et al.* RNA-binding protein Muscleblind-like 3 (MBNL3) disrupts myocyte enhancer factor 2 (Mef2) beta-exon splicing. *J. Biol. Chem.* **285**, 33779-33787, doi:10.1074/jbc.M110.124255 (2010).

34 Zhang, M. *et al.* Functional, structural and molecular characterizations of leukemogenic driver MEF2D-HNRNPUL1 fusion. *Blood*, doi: 10.1182/blood.2022016241 (2022).

35 Sprangers, R. & Kay, L. E. Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* **445**, 618-622, doi:10.1038/nature05512 (2007).

36 Wiegand, T. A solid-state NMR tool box for the investigation of ATP-fueled protein engines. *Prog. Nucl. Magn. Reson. Spectrosc.* **117**, 1-32, doi:10.1016/j.pnmrs.2020.02.001 (2020).

37 Luchinat, E. & Banci, L. In-cell NMR: a topical review. *IUCrJ* **4**, 108-118, doi:10.1107/S2052252516020625 (2017).

38 Kakita, V. M. R., Rachineni, K., Bopardikar, M. & Hosur, R. V. NMR supersequences with real-time homonuclear broadband decoupling: Sequential acquisition of protein and small molecule spectra in a single experiment. *J. Magn. Reson.* **297**, 108-112, doi:10.1016/j.jmr.2018.10.013 (2018).

39 Kupče, Ē., Kay, L. E. & Freeman, R. Detecting the "afterglow" of 13C NMR in proteins using multiple receivers. *J. Am. Chem. Soc.* **132**, 18008-18011, doi:10.1021/ja1080025 (2010).

40 Viegas, A. *et al.* UTOPIA NMR: activating unexploited magnetization using interleaved low-gamma detection. *J. Biomol. NMR* **64**, 9-15, doi:10.1007/s10858-015-0008-7 (2016).

41 Schanda, P. & Brutscher, B. Very Fast Two-Dimensional NMR Spectroscopy for Real-Time Investigation of Dynamic Events in Proteins on the Time Scale of Seconds. *J. Am. Chem. Soc.* **127**, 8014-8015, doi:10.1021/ja051306e (2005).

42 Kupce, E. & Freeman, R. Fast multidimensional NMR by polarization sharing. *Magn. Reson. Chem.* **45**, 2-4, doi:10.1002/mrc.1931 (2007).

43 Kupče, Ē. & Claridge, T. D. W. NOAH: NMR Supersequences for Small Molecule Analysis and Structure Elucidation. *Angew. Chem. Int. Ed. Engl.* **56**, 11779-11783, doi:10.1002/anie.201705506 (2017).

44 Kazimierczuk, K. & Orekhov, V. Y. Accelerated NMR spectroscopy by using compressed sensing. *Angew. Chem. Int. Ed. Engl.* **50**, 5556-5559, doi:10.1002/anie.201100370 (2011).

45 Petersen, B. O. *et al.* H2BC: a new technique for NMR analysis of complex carbohydrates. *Carbohydr. Res.* **341**, 550-556, doi:10.1016/j.carres.2005.11.020 (2006).

46 Benie, A. J. & Sørensen, O. W. Improved multiplicity-editing of HMBC NMR spectra. *Magn. Reson. Chem.* **44**, 739-743, doi:10.1002/mrc.1855 (2006).

47 Krishnamurthy, V. V., Russell, D. J., Hadden, C. E. & Martin, G. E. 2J,3J-HMBC: A New Long-Range Heteronuclear Shift Correlation Technique Capable of Differentiating 2JCH from 3JCH Correlations to Protonated Carbons. *J. Magn. Res.* **146**, 232-239, doi:https://doi.org/10.1006/jmre.2000.2141 (2000).

48 Kupče, Ē. & Sørensen, O. W. 2BOB - extracting an H2BC and an HSQC-type spectrum from the same data set, and H2OBC - a fast experiment delineating the protonated. *Magn. Reson. Chem.* **55**, 515-518, doi:10.1002/mrc.4584 (2017).

49 Oostenbrink, C., Villa, A., Mark, A. E. & van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **25**, 1656-1676, doi:10.1002/jcc.20090 (2004).

50 Ponder, J. W. & Case, D. A. Force fields for protein simulations. *Adv. Protein Chem.* **66**, 27-85, doi:10.1016/s0065-3233(03)66002-x (2003).

51 MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586-3616, doi:10.1021/jp973084f (1998).

52 Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696-3713, doi:10.1021/acs.jctc.5b00255 (2015).

53 Huang, J. & MacKerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comp. Chem.* **34**, 2135-2145, doi:10.1002/jcc.23354 (2013).

54 Sinko, W., Miao, Y., de Oliveira, C. A. & McCammon, J. A. Population based reweighting of scaled molecular dynamics. *J. Phys. Chem. B* **117**, 12759-12768, doi:10.1021/jp401587e (2013).

55 Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods.* **14**, 71-73, doi:10.1038/nmeth.4067 (2017).

56 Olson, M. A. Parallel Tempering of Dark Matter from the Ebola Virus Proteome: Comparison of CHARMM36m and CHARMM22 Force Fields with Implicit Solvent. *J. Chem. Inf. Model* **58**, 111-118, doi:10.1021/acs.jcim.7b00517 (2018).

57 Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **29**, 1859-1865, doi:10.1002/jcc.20945 (2008).

58 Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S. & Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **9**, 3878-3888, doi:10.1021/ct400314y (2013).

59 Darden, T., York, D. & Pedersen, L. Particle mesh ewald - an n.log(n) method for ewald sums in large systems. *J. Chem. Phys.* **98**, 10089-10092, doi:10.1063/1.464397 (1993).

60 Miyamoto, S. & Kollman, P. A. Settle - an analytical version of the shake and rattle algorithm for rigid water models. *J. Comp. Chem.* **13**, 952-962, doi:10.1002/jcc.540130805 (1992).

61 Roe, D. R. & Cheatham, T. E., 3rd. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory. Comput.* **9**, 3084-3095, doi:10.1021/ct400341p (2013).

62 Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph. Model.* **14**, 33-38, doi:10.1016/0263-7855(96)00018-5 (1996).

63 Ester, M. K., H.-P.; Sander, J.; Xu, X. In. A Density-Based Algorithm for Discovering Clustersin Large Spatial Databases with Noise. *KDD Proceedings*, 226-231 (1996).

64 Frishman, D. & Argos, P. Knowledge-based protein secondary structure assignment. *Proteins* **23**, 566-579, doi:10.1002/prot.340230412 (1995).

65 Samantray, S. *et al.* Molecular dynamics simulations of protein aggregation: protocols for simulation setup and analysis with Markov state models and transition networks. *bioRxiv*, 2020.2004.2025.060269, doi:10.1101/2020.04.25.060269 (2020).

66 Uversky, V. N. Unusual biophysics of intrinsically disordered proteins. *Biochim. Biophys. Acta* **1834**, 932-951, doi:10.1016/j.bbapap.2012.12.008 (2013).

67 Kjaergaard, M. & Poulsen, F. M. Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. *J. Biomol. NMR* **50**, 157-165, doi:10.1007/s10858-011-9508-2 (2011).

68 Berjanskii, M. V. & Wishart, D. S. A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc.* **127**, 14970-14971, doi:10.1021/ja054842f (2005).

69 Kaderavek, P. *et al.* Spectral density mapping protocols for analysis of molecular motions in disordered proteins. *J. Biomol. NMR* **58**, 193-207, doi:10.1007/s10858-014-9816-4 (2014).

70 Lipari, G. & Szabo, A. Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .2. Analysis of Experimental Results. *J. Am. Chem. Soc.* **104**, 4559-4570, doi:Doi 10.1021/Ja00381a010 (1982).

71 Dudás, E. F. & Bodor, A. Quantitative, Diffusion NMR Based Analytical Tool To Distinguish Folded, Disordered, and Denatured Biomolecules. *Anal. Chem.* **91**, 4929-4933, doi:10.1021/acs.analchem.8b05617 (2019).

72 Prompers, J. J. & Bruschweiler, R. General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. *J. Am. Chem. Soc.* **124**, 4522-4534, doi:10.1021/ja012750u (2002).

73 Marsh, J. A. & Forman-Kay, J. D. Sequence determinants of compaction in intrinsically disordered proteins. *Biophys. J.* **98**, 2383-2390, doi:10.1016/j.bpj.2010.02.006 (2010).

74 Rahman, M. U., Rehman, A. U., Liu, H. & Chen, H. F. Comparison and Evaluation of Force Fields for Intrinsically Disordered Proteins. *J. Chem. Inf. Model* **60**, 4912-4923, doi:10.1021/acs.jcim.0c00762 (2020).

75 Gallivan, J. P. & Dougherty, D. A. Cation-pi interactions in structural biology. *Proc. Natl. Acad. Sci. USA* **96**, 9459-9464, doi:10.1073/pnas.96.17.9459 (1999).

76 Ernst, R. R. & Anderson, W. A. Application of Fourier Transform Spectroscopy to Magnetic Resonance. *Rev. Sci. Instrum.* **37**, 93-102, doi:10.1063/1.1719961 (1966).

77 Sørensen, O. W. Selective Rotations Using Non-Selective Pulses and Heteronuclear Couplings. *Bull. Magn. Reson.* **16**, 49-53 (1994).

78 Wimperis, S. & Freeman, R. An excitation sequence which discriminates between direct and long-range CH coupling. *J. Magn. Res.* **58**, 348-353, doi:10.1016/0022-2364(84)90227-0 (1984).

79 Nagy, T. M., Gyöngyösi, T., E. Kövér, K. & Sørensen, O. W. BANGO SEA XLOC/HMBC–H2OBC: Complete heteronuclear correlation within minutes from one NMR pulse sequence. *Chem. Comm.*, doi:10.1039/c9cc06253j (2019).

80 Garwood, M. & DelaBarre, L. The return of the frequency sweep: designing adiabatic pulses for contemporary NMR. *J. Magn. Reson.* **153**, 155-177, doi:10.1006/jmre.2001.2340 (2001).

81 Kupče, Ē. & Freeman, R. Compensation for Spin–Spin Coupling Effects during Adiabatic Pulses. *J. Magn. Res.* **127**, 36-48, doi:https://doi.org/10.1006/jmre.1997.1193 (1997).

82 Kupce, E. Applications of adiabatic pulses in biomolecular nuclear magnetic resonance. *Methods. Enzymol.* **338**, 82-111, doi:10.1016/s0076-6879(02)38216-8 (2001).

83 Garbow, J., Weitekamp, D. & Pines, A. Bilinear Rotation Decoupling of Homonuclear Scalar Interactions. *Chem. Phys. Lett.* **93**, 504-509, doi:10.1016/0009-2614(82)83229-6 (1982).

84 Briand, J. & Sørensen, O. W. Simultaneous and independent rotations with arbitrary flip angles and phases for I, ISalpha, and ISbeta spin systems. *J. Magn. Reson.* **135**, 44-49, doi:10.1006/jmre.1998.1556 (1998).

85 Li, C. H. *et al.* MeCP2 links heterochromatin condensates and neurodevelopmental disease. *Nature* **586**, 440-444, doi:10.1038/s41586-020-2574-4 (2020).

9. Appendix

9.1 Repository for the in-house scripts (NMR&MD analysis) and Bruker pulse programmes for NMR measurements

GitHub repository: https://github.com/ocel0t/abyss0fPhD

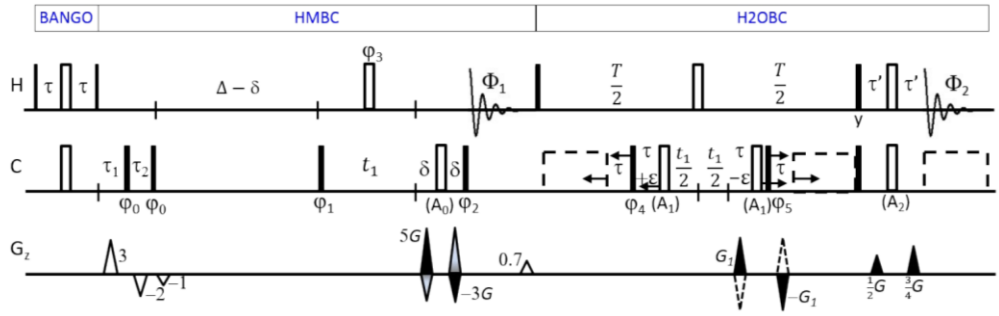
Detailed Bruker codes can be found in the supplementary materials of the following references:

1.) Nagy, T.M., Gyöngyösi, T., Kövér, K. E., Sørensen, O. W.
BANGO SEA XLOC/HMBC–H2OBC: complete heteronuclear correlation within minutes from one NMR pulse sequence
Chem. Comm. **55**, 81, doi: 10.1039/C9CC06253J (2019).

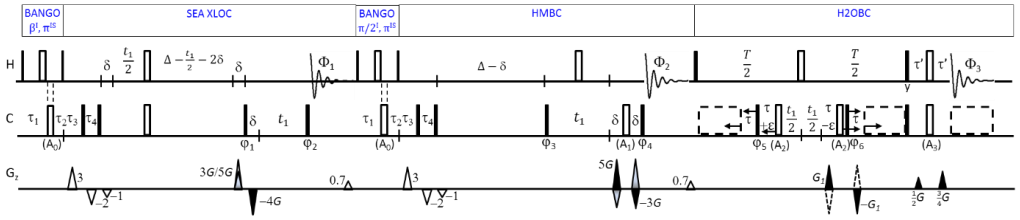
2.) Nagy, T.M., Kövér, K. E., Sørensen, O. W.
Double and adiabatic BANGO for concatenating two NMR experiments relying on the same pool of magnetization
J. Magn. Res. **316**, 106767, doi:10.1016/j.jmr.2020.106767 (2020).

3.) Nagy, T.M., Kövér, K. E., Sørensen, O. W.
NORD: NO Relaxation Delay NMR Spectroscopy
Angew. Chem. **60**, 24, doi: 10.1002/anie.202102487 (2021).

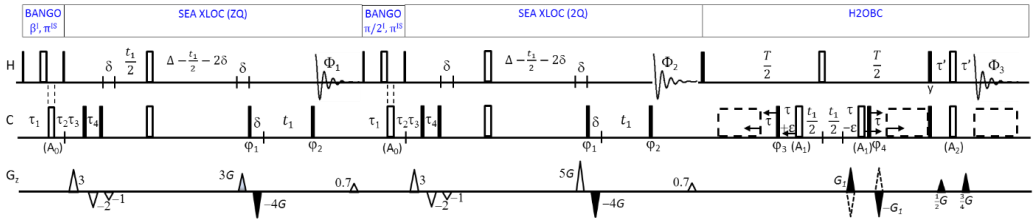
9.2 Pulse sequences of selected combined NMR experiments



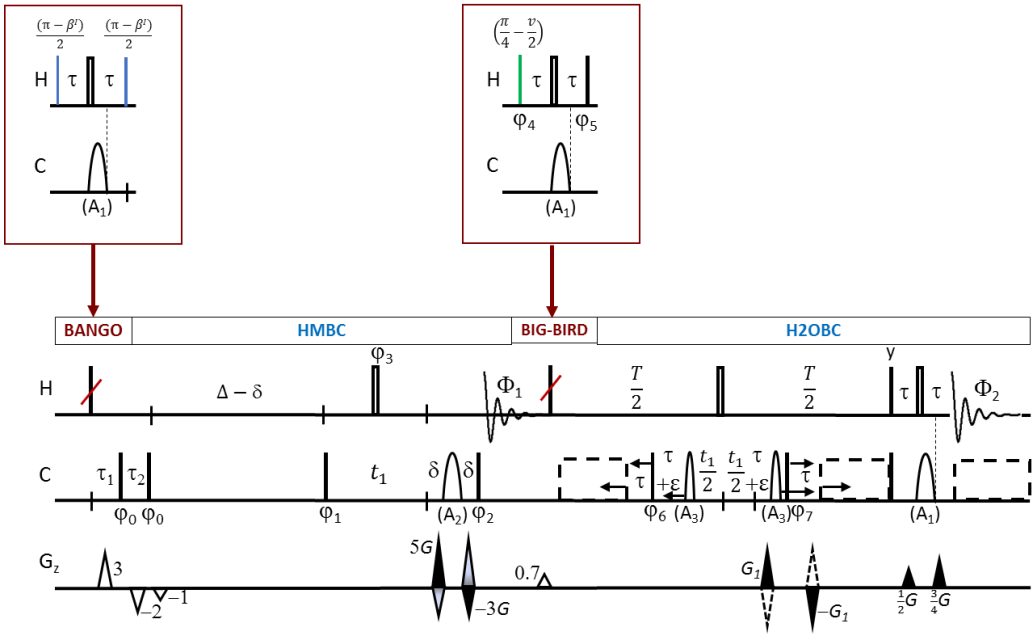
BANGO {HMBC}–{H2OBC} pulse sequence. *Bars: narrow filled: π/4; wider filled: π/2; open bars: π pulses. A_0, A_1, A_2: adiabatic refocussing and inversion pulses, composite 20% smoothed CHIRP (2 ms, 80 kHz sweep; Crp80comp.4) and CAWURST-20 (0.5 ms, 300 ppm, low-to-high frequency sweep), respectively. $\tau = ({}^1J_{max} + {}^1J_{min})^{-1}$ (BANGO delay). 2nd order low-pass J filter: $\tau 1 = 0.5({}^1J_{min})^{-1}$ and $\tau_2 = 0.5({}^1J_{max})^{-1}$. $\tau' = \tau - t(A_2)/2$. $\varepsilon = t(\pi^H)/2$. Δ: evolution delay under heteronuclear long-range couplings. T: evolution delay under homonuclear proton-proton couplings. δ: gradient delay. $\phi_0 = x$, $\phi_1/\phi_4 = \{x, -x, -x, x\}$, $\phi_2/\phi_5 = \{x, x, -x, -x\}$, $\phi_3 = x$, and receiver phases $\Phi_1/\Phi_2 = \{x,-x\}$.*



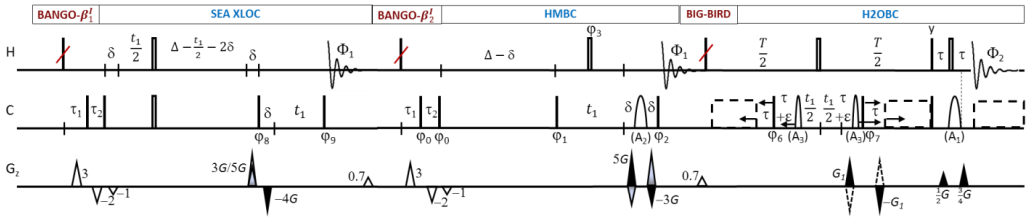
Double BANGO {SEA XLOC}–{HMBC}–{H2OBC} pulse sequence. *Bars: narrow filled: π/4; wider filled: π/2; open bars: π pulses. Narrow filled bars in the first BANGO: $(\pi - \beta')/2$ and in the second BANGO: π/4. $\tau_1 = 0.5({}^1J_{min})^{-1}$ (BANGO delay), where ${}^1J_{min}$ = 113 Hz. A_0: J-compensated adiabatic inversion pulse (CAWURST-20 of 1.92 ms with high-to-low field sweep of 240 ppm). $\tau_2 = \tau_1 - t(A_0)$. A_1, A_2 and A_3: adiabatic refocusing and inversion pulses, composite 20% smoothed CHIRP (2 ms, 80 kHz sweep; Crp80comp.4) and CAWURST-20 (0.5 ms, 300 ppm, low-to-high field sweep), respectively. A 2nd order low-pass J: $\tau_3 = 0.5[{}^1J_{min} + 0.146({}^1J_{max} - {}^1J_{min})]^{-1}$ and $\tau_4 = 0.5[{}^1J_{max} - 0.146({}^1J_{max} - {}^1J_{min})]^{-1}$. $\tau' = \tau - t(A_3)/2$. $\varepsilon = t(\pi^H)/2$. Δ: evolution delay under heteronuclear long-range couplings. T: evolution delay under homonuclear proton-proton couplings. δ: gradient delay. $\phi_1/\phi_3/\phi_5 = \{x, -x, -x, x\}$, $\phi_2/\phi_4/\phi_6 = \{x, x, -x, -x\}$, and receiver phases $\phi_1/\phi_2/\phi_3 = \{x,-x\}$.*



Double BANGO {SEA XLOC(ZQ)}–{SEA XLOC(2Q)}–{H2OBC} pulse sequence. *Bars: narrow filled: π/4; wider filled: π/2; open bars: π pulses. Narrow filled bars in the first BANGO: (π - β)/2 and in the second BANGO: π/4.* $\tau_1 = 0.5(^1J_{min})^{-1}$ *(BANGO delay), where* $^1J_{min}$ *= 113 Hz. : J-compensated adiabatic inversion pulse (CAWURST-20 of 1.92 ms with high-to-low field sweep of 240 ppm.* $\tau_2 = \tau_1 - t(A_0)$. *A_1, A_2 and A_3: adiabatic refocusing and inversion pulses, composite 20% smoothed CHIRP (2 ms, 80 kHz sweep; Crp80comp.4) and CAWURST-20 (0.5 ms, 300 ppm, low-to-high field sweep), respectively. 2nd order low-pass J:* $\tau_3 = 0.5[^1J_{min} + 0.146(^1J_{max} - {}^1J_{min})]^{-1}$ *and* $\tau_4 = 0.5[^1J_{max} - 0.146(^1J_{max} - {}^1J_{min})]^{-1}$. $\tau' = \tau - t(A_3)/2$. $\varepsilon = t(\pi^H)/2$. *Δ: evolution delay under heteronuclear long-range couplings. T: evolution delay under homonuclear proton-proton couplings.* $\phi_1/\phi_3 = \{x, -x, -x, x\}$, $\phi_2/\phi_4 = \{x, x, -x, -x\}$, *and receiver phases* $\phi_1/\phi_2/\phi_3 = \{x,-x\}$.



NORD {HMBC}–{H2OBC} pulse sequence. *Pulses with the red lines: the π/2 excitation of the standard experiments to be replaced by the element BANGO or BIG-BIRD as indicated above them. Bars: narrow filled: π/4; wider filled: π/2; open bars: π pulses. Flip angles in the BANGO and BIG-BIRD sequences are shown above the pulses. A_1: CAWURST-20 (1.92 ms, 240 ppm, low-to-high frequency sweep) adiabatic inversion pulse. A_2: composite 20% smoothed CHIRP (2 ms, 80 kHz sweep; Crp80comp.4). A_3: CAWURST-20 (0.97 ms, 240 ppm, high-to-low frequency sweep) J-compensated refocusing pulse.* $\tau_1 = 0.5({}^1J_{min})^{-1}$ *(BANGO and BIG-BIRD delay), where* ${}^1J_{min} = 113$ *Hz. J :* $\tau_3 = 0.5[{}^1J_{min} + 0.146({}^1J_{max} - {}^1J_{min})]^{-1}$ *and* $\tau_4 = 0.5[{}^1J_{max} - 0.146({}^1J_{max} - {}^1J_{min})]^{-1}$. $\tau' = \tau - t(A_3)/2$. $\varepsilon = t(\pi^H)/2$. *Δ: evolution delay under heteronuclear long-range couplings (typically 60-100 ms). T: evolution delay under homonuclear proton-proton couplings (typically 15-25 ms). δ: gradient delay.* $\varphi_0 = x$, $\varphi_1/\varphi_6/\varphi_8 = \{x, -x, -x, x\}$, $\varphi_2/\varphi_7/\varphi_9 = \{x, x, -x, -x\}$, $\varphi_3 = x$, $\varphi_4 = \left(\frac{\pi}{4} - \frac{v}{2}\right)$, $\varphi_5 = \left(\frac{5\pi}{4} + \frac{v}{2}\right)$, *and receiver phases* $\Phi_1/\Phi_2 = \{x,-x\}$.



NORD {SEA XLOC}–{HMBC}–{H2OBC} pulse sequence. *Pulses with the red lines: the π/2 excitation of the standard experiments to be replaced by the element BANGO or BIG-BIRD as indicated above them. Bars: narrow filled: π/4; wider filled: π/2; open bars: π pulses. Flip angles in the BANGO and BIG-BIRD sequences are shown above the pulses. A_1: CAWURST-20 (1.92 ms, 240 ppm, low-to-high frequency sweep) adiabatic inversion pulse. A_2: composite 20% smoothed CHIRP (2 ms, 80 kHz sweep; Crp80comp.4). A_3: CAWURST-20 (0.97 ms, 240 ppm, high-to-low frequency sweep) J-compensated refocusing pulse.* $\tau_1 = 0.5({}^1J_{min})^{-1}$ *(BANGO and BIG-BIRD delay), where* ${}^1J_{min} = 113$ *Hz. J :* $\tau_3 = 0.5[{}^1J_{min} + 0.146({}^1J_{max} - {}^1J_{min})]^{-1}$ *and* $\tau_4 = 0.5[{}^1J_{max} - 0.146({}^1J_{max} - {}^1J_{min})]^{-1}$. $\tau' = \tau - t(A_3)/2$. $\varepsilon = t(\pi^H)/2$. *Δ: evolution delay under heteronuclear long-range couplings (typically 60-100 ms). T: evolution delay under homonuclear proton-proton couplings (typically 15-25 ms). δ: gradient delay.* $\varphi_0 = x$, $\varphi_1/\varphi_6/\varphi_8 = \{x, -x, -x, x\}$, $\varphi_2/\varphi_7/\varphi_9 = \{x, x, -x, -x\}$, $\varphi_3 = x$, $\varphi_4 = \left(\frac{\pi}{4} - \frac{v}{2}\right)$, $\varphi_5 = \left(\frac{5\pi}{4} + \frac{v}{2}\right)$, *and receiver phases* $\Phi_1/\Phi_2 = \{x,-x\}$.

9.3 Publications related to the thesis:

1. Gönczi, M., Nagy, T.M., Fehér, K., Ráduly, Zs., Gregus, A. , Ambrus, V., Bécsi, B., Csernoch, L., Erdődi, F., E. Kövér, K., Fuxreiter, M.:
Alternative splicing regulates dynamics and oligomerization state of MEF2D transcription factor to program cell differentiation (under review).

2. Nagy, T.M., Gyöngyösi, T., E. Kövér, K. , Sørensen, O. W.:
NORD: NO Relaxation Delay NMR spectroscopy.
Angew. Chem.Int. Ed. 60, 24, doi: 10.1002/anie.202102487 (2021). IF: 16.823.

3. Nagy, T.M., Kövér, K. E., Sørensen, O. W.:
Double and adiabatic BANGO for concatenating two NMR experiments relying on the same pool of magnetization.
J. Magn. Res. 316, 106767, doi:10.1016/j.jmr.2020.106767 (2020). IF: 5.996.

4. Nagy, T.M., Gyöngyösi, T., Kövér, K. E., Sørensen, O. W.:
BANGO SEA XLOC/HMBC–H2OBC:
complete heteronuclear correlation within minutes from one NMR pulse sequence
Chem. Comm. 55, 81, doi: 10.1039/C9CC06253J (2019). IF: 2.229.

9.4 Other publications:

5. Timári, I., Nagy, T. M., E. Kövér, K., Sørensen O. W.:
Synergy and sensitivity-balance in concatenating experiments in NO Relaxation Delay NMR (NORD)
Chem. Comm. 58,15, doi: 10.1039/D1CC06663C (2022). IF: 6.065.

6. Bereczki I., Szűcs Zs., Batta Gy., Nagy, T. M.,
Ostorházi E., E. Kövér, K., Borbás A., Herczegh P.:
The First Dimeric Derivatives of the Glycopeptide Antibiotic Teicoplanin
Pharmaceuticals. 15, 77, doi: 10.3390/ph15010077 (2022). IF: 5.215.

7. Gyöngyösi, T., Nagy, T. M., E. Kövér, K. , Sørensen, O. W.:
Distinguishing between two- and three-bond correlations for all 13C multiplicities in heteronuclear NMR spectroscopy
Chem. Comm. 54, 70, doi: 10.1039/C8CC05156A (2018). IF: 6.164.

8. Nagy, T. M., Knapp, K., Illyés, E., Timári, I.,
Schlosser, G., Csík, G., Borics, A., Majer, Zs., E. Kövér, K.
Photochemical and Structural Studies on Cyclic Peptide Models
Molecules 23, 9, doi: 10.3390/molecules23092196 (2018). IF: 3.060.

9. Gróf, P., Knapp, K., Schlosser, G., Nagy, T. M.,
Timári, I., Borics, A., E. Kövér, K., Csík, G., Majer Zs.
Diszulfidhidat tartalmazó ciklikus peptidek UV-besugárzásának hatására keletkező szabadgyökök és szulfhidril-csoportok detektálása
Matud. 177, 1, (2016).

9.5 Conference presentations related to the dissertation

1.) Nagy, T. M., E. Kövér, K., Sørensen O. W:
NORD spektroszkópia: mérések holtidő nélkül!
NMR Working Comitee Meeting, 09.09.2021, Mád, Hungary.

2.) Nagy, T.M., Gönczi, M., Fehér, K., Bécsi, B., Erdődi, F., E. Kövér, K., Fuxreiter, M.:
Az aggregáció szerepe a rendezetlen fehérjék működésében: NMR módszerfejlesztés és számításos kémiai vizsgálatok.
New National Excellence Program conference, 14.06.2021, online.

3.) Nagy, T.M., Gönczi, M., Fehér, K., Fuxreiter, M., E. Kövér, K.:
A dinamika és az oligomerizáció szerepe a MEF2D rendezetlen fehérje működésében.
NMR Working Comitee Meeting, 13.10.2020, online.

4.) Nagy, T.M., Gönczi, M., Fehér, K., Fuxreiter, M., E. Kövér, K.:
Rendezetlen fehérjék dinamikai vizsgálata és NMR-kísérletek fejlesztése.
New National Excellence Program conference, 15.06.2020, online.

5.) Nagy, T.M., Gönczi, M., Fehér, K., Fuxreiter, M., E. Kövér, K.:
Rendezetlen fehérjemodellek NMR és számításos vizsgálata.
Applied quantum chemistry and molecular dynamics group, 12.03.2020, Debrecen, Hungary.

6.) Nagy, T.M., Gönczi, M., Fehér, K., Fuxreiter, M., E. Kövér, K.:
A MEF2D β-domén bolyhossága és szerepe a biológiai aktivitásban: NMR és számításos vizsgálatok. *I. Young Chemist's Symposium, 03.04.2019, Debrecen, Hungary.*

7.) Nagy, T.M., Gönczi, M., Fehér, K., Fuxreiter, M., E. Kövér, K.:
Fuzziness of MEF2D β-domain and its role in biological activity: NMR and in-silico studies. *NMR Working Comitee Meeting, 17.05.2019, Balatonszemes, Hungary.*

8.) Nagy, T.M., Gönczi, M., Fehér, K., Fuxreiter, M., E. Kövér, K.:
Determining fuzziness of the Mef2D β-domain by NMR experiments and MD-calculations.
9th Chemistry towards Biology Conference, 24.09.2018, Budapest, Hungary.

9.) Nagy, T.M., Gönczi, M., Fehér, K., Fuxreiter, M., E. Kövér, K.:
Rendezetlen peptidmodellek dinamikájának vizsgálata NMR-rel és számításos módszerekkel. *NMR Working Comitee Meeting, 16.05.2018, Balatonszemes, Hungary.*

9.6 Posters related to the dissertation

1.) Nagy, T.M., Gönczi, M., Fehér, K., Fuxreiter, M., E. Kövér, K.:
Dynamical studies of the Mef2D β-domain by NMR and computational methods.
European Magnetic Resonance Meeting, 26.08.2018, Nantes, France.

2.) Nagy, T.M., Gönczi, M., Fehér, K., Fuxreiter, M., E. Kövér, K.:
The fuzzy β-domain of MEF2D: NMR and computational studies.
European Magnetic Resonance Meeting, 27.08.2019, Berlin, Germany.