



HAL
open science

Structure and dynamics of intrinsically disordered regions of MAPK signalling proteins

Jaka Kragelj

► **To cite this version:**

Jaka Kragelj. Structure and dynamics of intrinsically disordered regions of MAPK signalling proteins. Structural Biology [q-bio.BM]. Université de Grenoble, 2014. English. NNT : 2014GRENV060 . tel-01314426

HAL Id: tel-01314426

<https://theses.hal.science/tel-01314426>

Submitted on 11 May 2016

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

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

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Doctorat CSV/Biologie Structurale et nanobiologie**

Arrêté ministériel : 7 août 2006

Présentée par

Jaka Kragelj

Thèse dirigée par **Malene Ringkjøbing Jensen**
codirigée par **Martin Blackledge**

préparée au sein du **Flexibilité et Dynamique des Protéines**
de l'**Institut de Biologie Structurale**
dans l'**École Doctorale de Chimie et Sciences du Vivant**

Structure and dynamics of intrinsically disordered regions of MAPK signalling proteins

Thèse soutenue publiquement le 11.12.2014,
devant le jury composé de:

Prof. Wolfgang Peti

Brown University, Providence - Rapporteur

Prof. Birthe B. Kragelund

University of Copenhagen, Copenhagen - Rapporteur

Prof. Eva Pebay-Peyroula

Institut de Biologie Structurale, Grenoble - Examinatrice

Prof. Peter Tompa

Vrije Universiteit, Brussel - Examineur

Dr. Malene Ringkjøbing Jensen

Institut de Biologie Structurale, Grenoble - Directrice de thèse

Dr. Martin Blackledge

Institut de Biologie Structurale, Grenoble - Codirecteur de thèse



Acknowledgments

First, I would like to thank the members of the jury to have responded to my invitations. I would like to thank the *rapporteurs*, Prof. Wolfgang Peti and Prof. Birthe B. Kragelund. It was a pleasure having in the jury two experts in exactly the topics that my thesis is about. It was a big honor and pleasure to have Prof. Eva Pebay-Peyroula as part of my jury and also Prof. Peter Tompa, a renowned IDP expert. Thank you for taking the time for judging my thesis.

I would like to thank Martin Blackledge for his supervision and for providing an environment where it is possible to absorb the knowledge accumulated in the lab and profit from the variety of expertise of the different lab members. I am grateful that I was able to learn directly from him how to critically address the current problems in modern structural biology and to see how to carry out research that is elegant, successful and that significantly contributes to the progress in the field. The discussions that we had were always insightful and stimulating no matter with what kind of question or idea I have approached him.

I would like to thank Malene Ringkjøbing Jensen. She has handed me a well designed project that provided challenges and rewards together with many possibilities to learn in the process. Her suggestions and guidance were crucial, especially in the times when tough problems appeared. Without those, coming as far as I have gotten would have been much harder. All the practical and much of the theoretical NMR spectroscopy that I have learned came from her and I am grateful for being passed that knowledge. It was always exciting to discuss and interpret together the outcomes of the experiments that I have done and even more to think about the next experiments to come. Good luck with MAPKs in the future!

I would like to thank all the colleagues, who have stayed in the lab during the time when I was here: Luca Mollica, Robert Schneider, Malene Ringkjøbing Jensen, Jie-Rong Huang, Paul Guerry, Guillaume Communie, Élise Delaforge, Éric Condamine, Damien Maurin, Vallery Ozenne, Guillaume Bouvignies, Loïc Salmon, Anton Abyzov, Sigrid Milles, Nicola Salvi. They contributed to the pleasant atmosphere, stimulating discussions and I have learned a lot from them. For the same reason, I would like to thank to the colleagues from the CIBB building: Filip Yabukarski, Nicolas Tarbouriech, Marlyse Buisson, Eric Thierry, Catherine Fallecker, Charlotte Sueur, Christopher Swale and Alexandre Monod. And I would like to thank Zsófia Sólyom for being the perfect IDPbyNMR colleague.

For overseeing the writing of the thesis manuscript and for corrections, that improved it significantly, I would like to thank Martin and Malene. And thank you Guillaume C., Elise and Yann H.D.K. for translating the parts of the thesis written in French language.

This thesis would not be as it is without the help of all the lab members and also, it would not be the same without the work and progress done by the past lab-members much before my arrival to the lab. The influences on my work did of course not come only from the lab, and therefore I would like to proceed thanking all the scientists who dedicated their time to research and development in the fields at whose intersection my thesis lies and to all the authors of the papers that I have read in search for knowledge, solutions, explanations or just sole amusement.

Finally, regardless of whether the personal character is mostly a result of genes, environment or education, I would like to thank my parents and my sister. Hvala tata, mama in Urša!

Last but not least, I would like to thank Emiko for the support during the PhD and during the thesis writing.

Table of contents

1. Introduction.....	7
2. Intrinsic disorder in signalling proteins.....	9
2.1. MAPK signalling pathways.....	9
2.1.1. MAPKs are involved in processes related to inflammation and cancer.....	9
2.1.2. The organization of MAPK signalling networks are conserved across eukaryotes.....	10
2.1.3. A closer look at MAPKs and MKKs: structure of eukaryotic Ser/Thr kinases.....	11
2.1.4. All substrates of MAPKs contain docking-site motifs.....	12
2.1.5. Upstream MKKs recognize MAPKs through docking-site motifs.....	14
2.2. Intrinsically disordered proteins and regions.....	17
2.2.1. Intrinsic disorder is predicted in every second protein.....	17
2.2.2. IDRs are a functional part of eukaryotic proteins.....	18
2.2.3. Motifs often appear in intrinsically disordered regions.....	18
2.2.4. Pre-structured motifs.....	19
2.3. Motifs in signalling networks.....	20
2.3.1. Structure of motifs in complexes.....	20
2.3.2. Affinity and specificity of motifs.....	22
2.3.3. Combining specificity of motifs and localization.....	24
2.3.4. Cooperativity from other regions.....	25
2.3.5. The JIP1 scaffold protein.....	26
2.4. Objectives.....	28
3. Nuclear magnetic resonance.....	31
3.1. Quantum description for a spin-half nucleus.....	31
3.1.1. Dipolar interaction.....	33
3.1.2. Chemical shift and chemical shift anisotropy.....	34
3.2. Relaxation theory.....	35
3.2.1. Sources of relaxation.....	35
3.2.2. Spectral density.....	35
3.2.3. Longitudinal relaxation.....	36
3.2.4. Heteronuclear NOEs.....	36
3.2.5. Transverse relaxation.....	37
3.2.6. Relaxation due to chemical exchange.....	37
3.2.7. Carr-Purcell Meiboom-Gill Relaxation Dispersion.....	38
3.2.8. Chemical Exchange Saturation Transfer.....	41
3.3. Characterizing residual structure in IDPs by NMR.....	42
3.3.1. Chemical shifts.....	42
3.3.2. Prediction of chemical shifts from structural coordinates.....	45
3.3.3. Residual dipolar couplings.....	46
4. Ensemble calculation for IDPs using NMR parameters.....	55
4.1. Abstract.....	56
4.2. Introduction.....	56
4.3. Local structure in IDPs can be described by the dihedral angle distributions.....	58
4.4. NMR parameters for characterizing local conformational propensities in IDPs.....	60
4.5. Sample-and-select approaches.....	61
4.6. Sampling space using molecular dynamics simulations.....	62
4.7. Sampling space using statistical coil generators.....	63
4.8. Selection of ensembles on the basis of experimental NMR data.....	65

4.9. Ensemble representations of the IDP Tau from chemical shifts and RDCs.....	66
4.10. The reference ensemble method.....	68
4.11. Taking into account cooperatively formed secondary structures in IDPs.....	70
4.12. Choosing an appropriate ensemble size.....	72
4.13. Ensemble size in relation to convergence properties of NMR parameters.....	73
4.14. Validation of ensemble descriptions.....	75
4.15. Conclusions and outlook.....	76
5. Methods.....	79
5.1. Sample preparation.....	79
5.1.1. Expression and purification of intrinsically disordered proteins.....	79
5.1.2. Expression and purification of full-length MKK7.....	79
5.1.3. Expression and purification of JNK.....	80
5.2. NMR measurements.....	81
5.2.1. Backbone assignment experiments.....	81
5.2.2. Measurement of RDCs.....	81
5.2.3. Relaxation dispersion and chemical exchange saturation transfer experiments.....	81
5.3. Other methods.....	82
5.3.1. Crystallization of JNK.....	82
5.3.2. Determination of the crystal structure of JNK.....	82
5.3.3. Thermal shift assay.....	83
5.3.4. ITC measurements.....	84
5.4. Ensemble calculations.....	84
5.4.1. Ensemble calculations for MKK7.....	84
5.4.2. Ensemble calculations for MKK4.....	85
6. Results.....	87
6.1. Characterization of the N-terminal IDR of MKK7.....	88
6.2. Characterization of the N-terminal IDR of MKK4.....	100
6.3. Interaction of MKK4 and MKK7 with JNK1.....	108
6.3.1. Stability of JNK1.....	108
6.3.2. NMR titration of MKK7_1-100 with JNK1.....	109
6.3.3. NMR titration of MKK4_12-86 with JNK1.....	112
6.4. Isothermal titration calorimetry.....	114
6.5. NMR titration of the full-length MKK7 kinase with JNK1.....	116
6.6. Crystal structure of the D2 docking-site motif of MKK7 bound to JNK1.....	119
6.7. Kinetics in the JNK-MKK7 complex.....	124
6.8. Alignment of JNK interacting motifs.....	127
6.9. The scaffold protein JIP1.....	131
7. Discussion.....	135
8. Résumé en français.....	141
8.1. Introduction.....	141
8.2. Les protéines intrinsèquement désordonnées dans les voies de signalisation.....	142
8.3. Description des structures et des interactions faisant intervenir des PID par RMN.....	142
8.4. Caractérisation des structures résiduelles des PID par RMN.....	144
8.5. Calcul d'ensemble pour les PDIs à partir de paramètres RMN.....	144
8.6. Résultats.....	145
8.7. Discussion.....	146
9. Publications.....	149
10. Bibliography.....	187

1 Introduction

Cells need to constantly sense and react to the stimuli coming from the environment and, in multicellular organisms, also to the surrounding cells. Protein signal transduction pathways allow them to transmit these signals and process the incoming information.

A group of such pathways, called mitogen-activated protein kinase (MAPK) signal transduction pathways, is well conserved in all eukaryotic cells and is involved in regulating many important cell processes. Discoveries in this field over the last two decades make MAPKs well-studied from the structural perspective but long intrinsically disordered region (IDRs), present in many MAPKs, have remained structurally uncharacterised. The IDRs of MAPKs are especially important as they contain docking-site motifs which control the interactions between MAPKs themselves and also between MAPKs and other interacting proteins containing the same motifs.

Examination of signal transduction pathways in general reveals that motif interactions are especially enriched in signalling proteins because of their characteristic abilities e.g. they allow for complex regulatory mechanisms and rewiring of signalling networks through evolution. The high percentage of predicted intrinsic disorder in signalling pathways is therefore not surprising as IDRs normally represent the structural context in which motifs appear. However the high flexibility and dynamics of the IDRs represent a challenge for biophysical and structural characterization. In order to obtain a better understanding of how cells and signal transduction pathways work, it is necessary to gain mechanistic insight into how IDRs and motifs are organized.

There are many interesting questions that need to be addressed from this point of view: what is the nature of low-affinity and transient interactions that occur through motifs? How does folding upon binding occur? Is there a relationship between the unbound and bound states of a motif? What are the kinetics and thermodynamics of the binding processes? What is the best way of identifying the critical residues in motifs that are necessary for the interaction? How do motifs discriminate between binding partners and what determines the specificity of interactions? Where do post-translational modifications occur and what are their effects?

Nuclear magnetic resonance (NMR) spectroscopy is well suited for studying the enumerated phenomena. Intrinsically disordered proteins (IDPs) have favourable NMR characteristics such as long transverse relaxation times and narrow line widths. There are many NMR observables, such as for example chemical shifts and residual dipolar couplings, which inform us about residual structure in IDRs. The last two decades have witnessed an enormous progress in understanding the

1. Introduction

relationship between chemical shift and structure and at the same time RDCs were measured for the first time in folded proteins and a framework for interpreting RDCs in IDPs has been established.

NMR further informs us about dynamics and the time scales at which motions are occurring. Nuclear spin relaxation can inform us about the fastest pico-nanosecond motions and slower motions can be probed by other methods such as rotating frame relaxation dispersion, Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion, ZZ-exchange and real-time NMR allowing dynamics of processes reaching the seconds time scale to be studied at atomic resolution. With such a rich toolbox NMR can give us a mechanistic insights into many aspects of how IDRs and motifs work. This will then hopefully help us to better understand how signal transduction pathways work in the context of cellular function.

In the following chapters I will describe the role of IDRs and motifs in signalling networks, the NMR parameters and experiments that are used to describe the structure and dynamics of IDPs and the approaches to ensemble modelling of IDPs. After a methods section, results of my work will be presented followed by a discussion.

2 Intrinsic disorder in signalling proteins

2.1 MAPK signalling pathways

2.1.1 MAPKs are involved in processes related to inflammation and cancer

The acronym MAP kinase (MAPK) was first used to name a kinase that was purified in the early '80s and which was activated in response to insulin stimulation (Avruch, 2007). When active, this kinase would phosphorylate a co-purified contaminant identified as microtubule-associated protein-2 (MAP-2). At the same time other protein kinases were being discovered that were activated upon growth factor stimulation and in response to other mitogens. Several years later, all these kinases were shown to be the same kinase, confirming that the MAP-2 kinase is a ubiquitous effector of mitogenic stimuli. The MAPK acronym was therefore designated to "mitogen-activated protein kinase" (Avruch, 2007; Rossomando et al., 1989).

Mammals possess multiple MAPK pathways which are classified into ERKs (extracellular signal-regulated kinases), JNKs (c-Jun N-terminal kinases) and the p38 family of kinases. These signalling pathways are activated in response to different stimuli, mostly by stress and inflammation rather than by mitogens which presents another twist to the story of the naming of MAPKs (Kyriakis and Avruch, 2012).

MAPK pathways are involved in numerous responses to the cell environment and have been described in many review articles (Keshet and Seger, 2010; Kyriakis and Avruch, 2012; Raman et al., 2007; Rincón and Davis, 2009; Rose et al., 2010). ERK kinases respond to insulin, cytokines of the tumor necrosis factor (TNF) family, pathogen-associated molecular patterns (PAMPs) and danger-associated molecular patterns (DAMPs). JNK kinases also react to PAMPs and DAMPs but also to proinflammatory cytokines, mitogens, growth factors, environmental stress (heat shock, ionizing radiation, oxidants, DNA damage), ischemic reperfusion injury, mechanical shear stress, vasoactive peptides, DNA and protein synthesis inhibition. Kinases from the p38 family are involved in responses to environmental stress, inflammatory cytokines, PAMPs, DAMPs, extracellular stress (ionizing radiation, osmotic shock, hypoxia), pro-inflammatory cytokines and less often by growth factors.

In response to the enumerated stimuli the MAPK pathways can regulate gene transcription, metabolism, protein biosynthesis, evoke inflammatory responses, cytokine production, control cell cycle progression and signal for apoptosis. These responses can be evoked through phosphorylation of transcription factors which are well-studied MAPK substrates. Many of the early examples of

transcription factors phosphorylated by MAPKs belonged to proto-oncogenes (cJun, cMyc, p62^{TCF}/Elk1, ATF2, MEF2C), but other MAPK substrates also exist. A well-documented class of substrates are MAPK-Activated Protein Kinases (MAPKAPKs or short MKs) which become activated upon phosphorylation and can subsequently phosphorylate substrates that are otherwise not targeted by MAPKs.

2.1.2 The organization of MAPK signalling networks are conserved across eukaryotes

The main feature of the MAPK signalling pathways is a cascade of three kinases (Figure 2.1) (Raman et al., 2007). Signals from the environment are first propagated to MAPK kinase kinases (MAPKKKs, MAP3Ks) which become activated after being phosphorylated by upstream kinases or

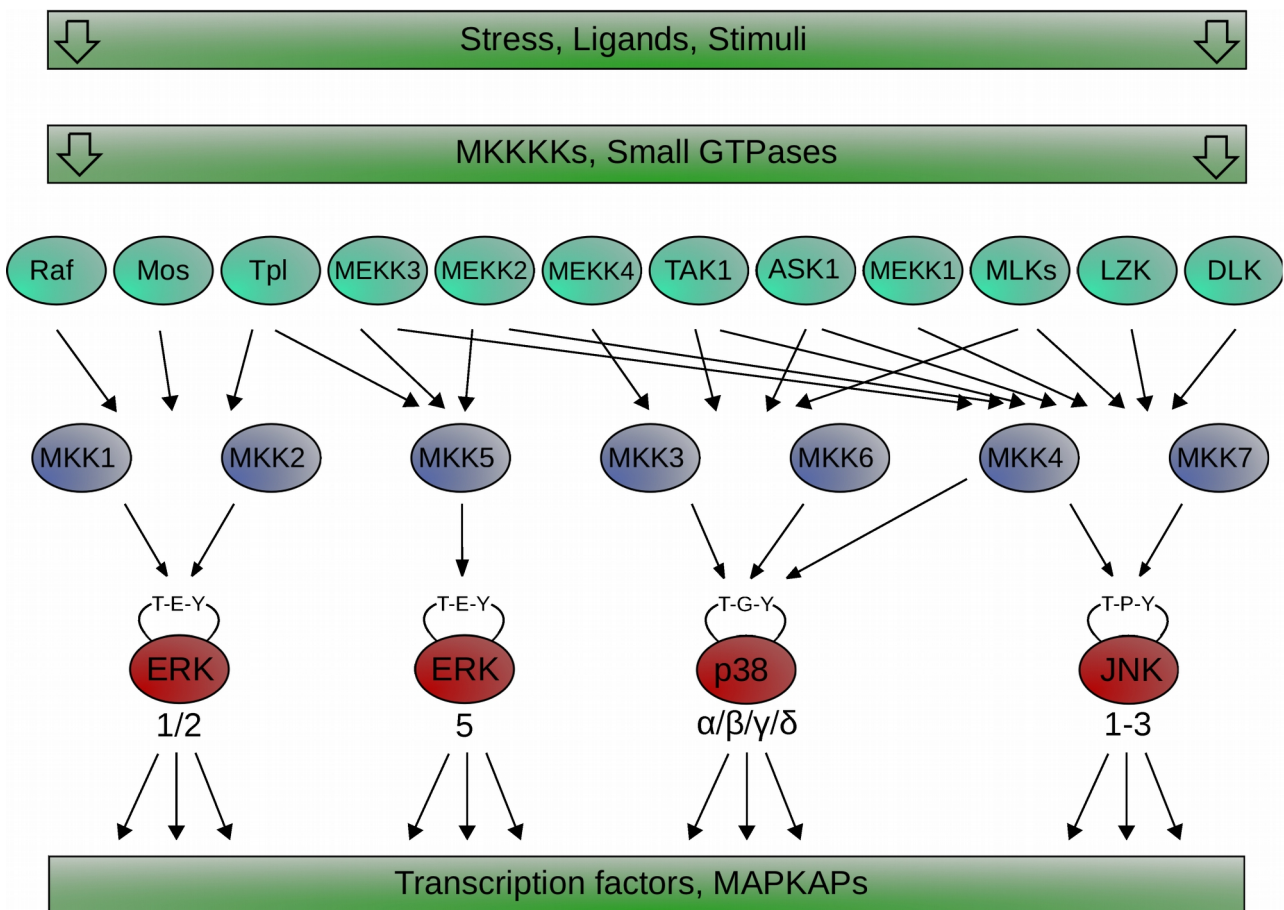


Figure 2.1: MAPK signalling pathways are organized into three tiers of MKKKs (turquoise-green), MKKs (blue) and MAPKs (red). A schematic activation loop is drawn on the MAPKs showing the three-residue motif which becomes doubly phosphorylated upon activation.

in some cases autophosphorylate after homodimerization. Activated MKKKs are able to phosphorylate and activate MAPK kinases (MKKs, MAPKK, MAP2Ks) which in turn phosphorylate the cognate MAPKs (Raman et al., 2007).

This MAPK core is conserved across all eukaryotes. One of the first MAPKs described was actually one of the Rsk family of kinases from *Xenopus* oocytes, although not yet named so at the time (Erikson and Maller, 1985, 1986; Spivack et al., 1984). MAPKs can be found even in yeast where there are four functionally separate MAPK pathways (Won et al., 2011). The best studied of the four pathways is the STE11-STE7-FUS3 which regulates mating in response to pheromones. Because of easier genetic manipulation of yeast this pathway represents a well manageable model and several studies have been performed providing important insight into the structure and mechanism of action of these kinases (Bhattacharyya et al., 2006; Bhunia et al., 2012; Good et al., 2009; Won et al., 2011).

From the three tiers of kinases the MKKs are lowest in number and most specific. MAP3Ks on the other hand are numerous and can often phosphorylate several MKKs (Figure 2.1). Upstream of MKKKs the signalling pathways show less conservation. The MKKK-activating proteins are kinases which could be called MKKKKs or equally often the activators of these pathways are small GTPases and sometimes also tyrosine kinases (Keyse, 2000; Kyriakis and Avruch, 2012).

2.1.3 A closer look at MAPKs and MKKs: structure of eukaryotic Ser/Thr kinases

The eukaryotic protein kinase fold is well-conserved. The N-terminal domain of the kinases is composed mostly of β -strands, while the C-terminal domain is α -helical. The N- and C-terminal domains are often referred to as N- and C-terminal lobes. The ATP binding site is located between the two lobes and stabilized by the flexible hinge region (Figure 2.2).

For both MAPKs and MKKs, the activation loop is located in front of the active site. MKKs are phosphorylated on serines or threonines two residues apart (Ser-Xxx-Xxx-Ser/Thr). The residues phosphorylated in MAPKs depend on the specific MAPK and are Thr-Glu-Tyr for ERK kinases, Thr-Pro-Tyr for JNK kinases and Thr-Gly-Tyr for p38 kinases (Figure 2.1). These characteristics of phosphoacceptor residues are summarized in (Kyriakis and Avruch, 2012) for both MKKs and MAPKs.

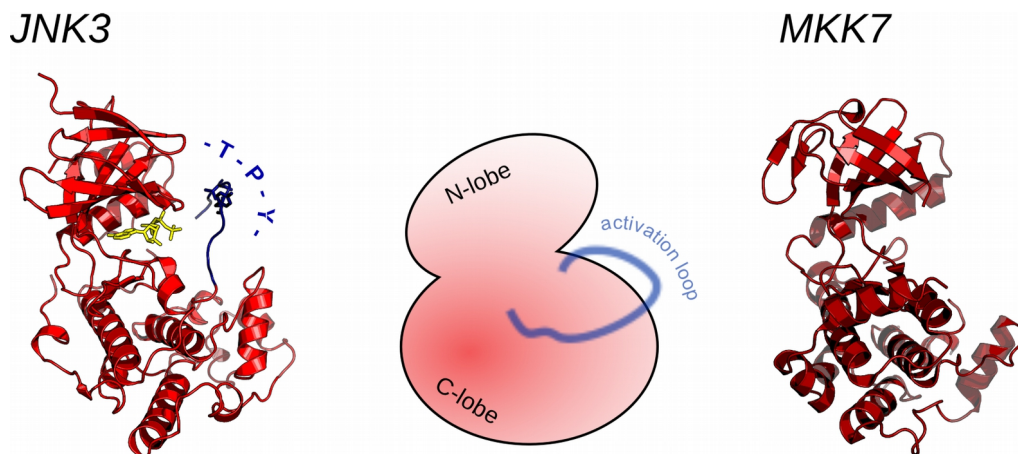


Figure 2.2: JNKs and MKKs share the same basic characteristic features of eukaryotic kinases. The ATP analog is displayed in yellow and the activation loop is blue. PDB accession codes: 1JNK (JNK3), 2DYL (MKK7)

When the MAPK is doubly phosphorylated, it adopts an active conformation involving a repositioning of the two lobes including a rearrangement of the phosphorylated activation loop and other residues around the active site. Distinguishing between active and inactive conformations of kinases is difficult as kinases are flexible molecules and can adopt two distinct active (open/closed) and multiple inactive conformations (Taylor and Kornev, 2011). A recently developed description of kinase activation seeks to simplify these conformational changes by summarizing past observations in a unified model (Kornev et al., 2006, 2008; Meharena et al., 2013; Taylor and Kornev, 2011). The conformational changes leading to activation of MAPKs have been described in studies done on p38 (Chang et al., 2002; Goldsmith, 2011; Tokunaga et al., 2014; Wang et al., 1997; Zhou et al., 2006a), ERK (Lee et al., 2004; Zhang et al., 1994; Zhou et al., 2006b) and JNK kinases (Laughlin et al., 2012).

2.1.4 All substrates of MAPKs contain docking-site motifs

Phosphorylation studies using synthetic peptides were initially used to identify sequence determinants of MAPK substrates. They concluded that a proline is required at the +1 position (Ser/Thr-Pro) (Alvarez et al., 1991; Avruch, 2007; Clark-Lewis et al., 1991; Hutti et al., 2004; Mukhopadhyay et al., 1992). Such a short recognition motif is clearly too common to allow for an efficient substrate discrimination. In addition the kinetic data indicated that the synthetic peptides containing a Ser/Thr-Pro motif were not phosphorylated as efficiently as the full-length native MAPK substrates, meaning that there are other regions that necessary for efficient phosphorylation.

It was discovered later that the JNK substrate c-Jun contained a high-affinity JNK binding site

located 26 and 36 amino acids away from the two phosphorylated Ser/Thr-Pro motifs (Avruch, 2007). Such high-affinity sites, were discovered also in other MAPK substrates (Avruch, 2007). These high-affinity sites, which are called docking-sites (D-sites), have a consensus sequence of a cluster of basic residues followed by a hydrophobic-X-hydrophobic sub-motif (R/K₂₋₃-X₂₋₄-Φ-X-Φ, where Φ denotes a hydrophobic residue and X any residue). These docking-sites bind to MAPKs in a groove that is located far from the active site (Figure 2.3a). While most of the docking-sites have a sequence conforming to the mentioned motif, there are some exceptions where docking-sites differ.

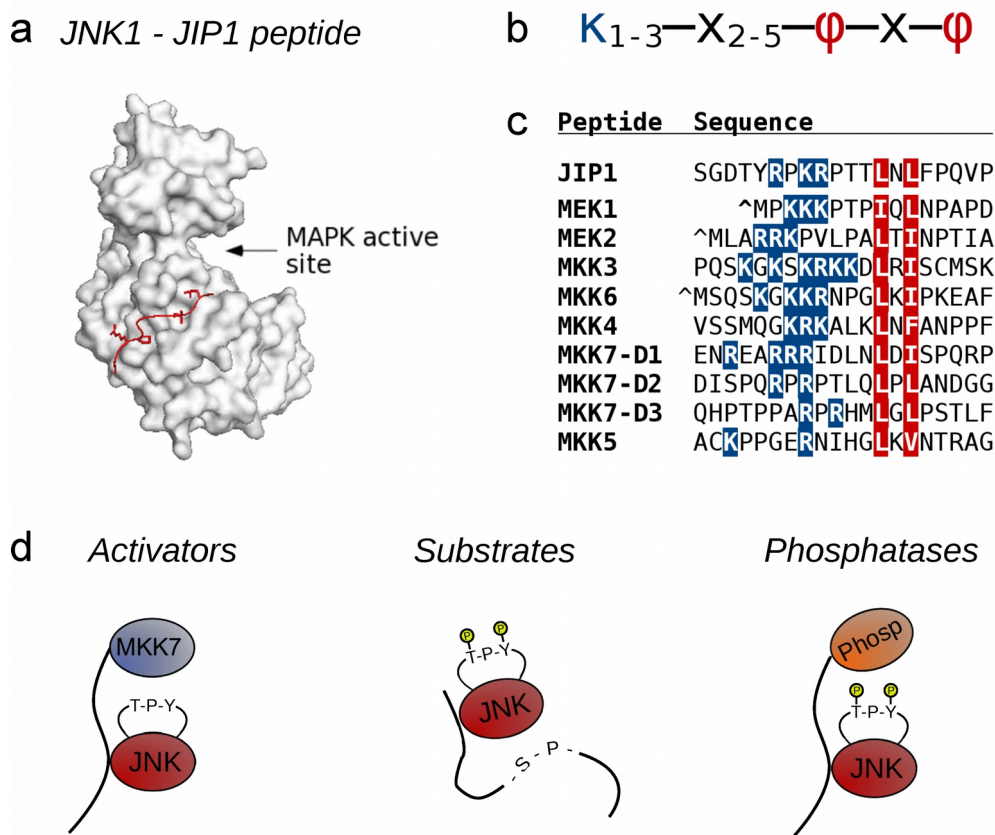


Figure 2.3: Docking motif interactions. a) Structure of JNK with a bound docking-site (1UKH). The bound peptide is shown in red and the surface of JNK is displayed in gray. b) A consensus sequence for MAPK docking motifs from JIP1 and MKKs. κ denotes basic amino acids, ϕ denotes hydrophobic amino acids. c) Alignment of docking motifs of MKKs. d) Schematic models of interactions between JNK and JNK interacting proteins with docking-site motifs.

ERK is able to bind an additional docking-site FXX(P) usually called DEF site (docking-site for ERK). A different type of motif specific to substrates from the MAPKAPK family binds in the reverse direction (from C- to N-terminus) and is therefore called reverse docking motif (Garai et al., 2012; Sharrocks et al., 2000). Docking-site interactions are essential to MAPK signalling because

they are also mediating interactions between MAPKs and their regulators (Figure 2.3d).

2.1.5 Upstream MKKs recognize MAPKs through docking-site motifs

The interactions between MKKs and MAPKs and with that the phosphorylation of MAPKs rely on the same type of docking-site interaction as the substrates. All MKKs have an N-terminal IDR in which the docking-site motif is present. The docking-site motif in MKKs is located 30 to 40 amino acids from the folded catalytic domain (Figure 2.4). MKK7 is the only exception with three docking-sites within its almost hundred amino acid long N-terminal disordered region.

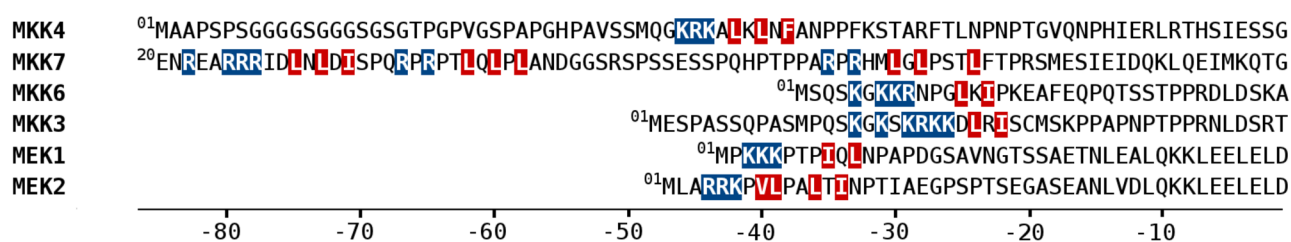


Figure 2.4: Comparison of the N-terminal IDRs of human MKKs. The sequences were aligned by the first conserved strand of the catalytic domains of the MKKs. The numbering starts with the last disordered residue before the conserved strand. The docking sites are identified by blue (basic residues) and red (hydrophobic) residues.

Sequence alignment of the catalytic domains of the human MKKs shows that they are homologous with 42% sequence identity for the least similar pairs of MKKs and 87% identity for the most similar pairs. Alignment of docking-site motifs of MKKs highlights their similarity in that they contain a cluster of basic residues and a hydrophobic-X-hydrophobic sub-motif (Figure 2.3c). Other positions within the motif do not seem to be conserved and the length of the linker between the basic and hydrophobic part varies from one MKK to the other.

Despite the similarities of docking motifs, *in vitro* phosphorylation experiments with full-length kinases have shown that MKKs display stringent specificity and phosphorylate only cognate MAPKs (Figure 2.1). The only MKK that is involved in two MAPK pathways is MKK4 which phosphorylates both the MAPKs from the JNK and p38 pathway (Kyriakis and Avruch, 2012; Raman et al., 2007).

Binding assays and competition experiments have been used in several studies to determine the affinities of peptides with sequences corresponding to docking-site motifs for cognate and non-cognate MAPKs (Bardwell et al., 2009; Garai et al., 2012). Comparing affinities of a single

docking-site motif for different MAPKs shows that each docking-site motif binds strongest to its cognate MAPK and slightly weaker to its non-cognate MAPKs. In some cases though these affinities can be quite similar, below a factor of two different, and with one exception below a factor of one (i.e. ERK2–MKK1) (Table 2.1) (Bardwell et al., 2009).

While this means that each MKK would indeed preferentially bind and presumably phosphorylate its cognate MAPK, it does not explain why an MKK would not phosphorylate other MAPKs to which it can bind with slightly lower affinity. The ERK2 kinase, for example, binds to the non-cognate docking-sites of MKK3 and MKK6 with similar affinity as to its cognate docking-sites (MKK1/MKK2) (Table 2.1). Despite this, only MKK1 and MKK2 are able to phosphorylate ERK2 *in vitro* because the amino acids around the phospho-residues (E in T-E-Y, Figure 2.1) on the activation loop of ERK2 provide an additional determinant of specificity (Garai et al., 2012). The *in vitro* observed specificity of phosphorylation reactions is therefore largely determined by the docking-sites and sometimes also by other features within the catalytic domains.

Table 2.1: Preferences of MAPKs for docking-site motifs^a

MAPK/substrate	Docking-site peptide	IC50	Median selectivity ^b
JNK1/ATF2	MKK4	2	47
JNK1/ATF2	MKK7	8	12
JNK1/ATF2	MEK2	60	
JNK1/ATF2	MKK3	90	
JNK1/ATF2	MEK1	100	
JNK1/ATF2	MKK6	100	
p38/MEF2A	MKK3	0.1	1000
p38/MEF2A	MKK4	7	14
p38/MEF2A	MKK6	20	7
p38/MEF2A	MEK2	40	
p38/MEF2A	MEK1	100	
p38/MEF2A	MKK7	> 100	
ERK2/Elk-1	MKK2	5	4
ERK2/Elk-1	MKK3	5	
ERK2/Elk-1	MKK6	17	
ERK2/Elk-1	MKK4	26	
ERK2/Elk-1	MKK1	32	< 1
ERK2/Elk-1	MKK7	85	

Lines containing cognate MAPK/MKK interaction pairs are coloured blue.

^aData in the table are adapted from (Bardwell AJ et al., 2009)

^bSelectivity is defined as IC50(MAPK with non-cognate docking-site)/IC50(MAPK with cognate docking-site peptide)

The lesson learned is that some docking-sites discriminate better than others but they all carry a

2. Intrinsic disorder in signalling proteins

certain degree of selectivity. This means that docking-sites should have some sequence characteristics that allow them to bind selectively to MAPKs. A comparison of sequences of the docking-site motifs from the six MKKs does not reveal any major differences that could be responsible for the discrimination between different MAPKs (Figure 2.3c). Going further, one can align all the known binding motifs specific to one MAPK in an attempt to identify the required residues for binding to that MAPK. Intriguingly, a careful alignment of nineteen JNK binding motifs, that have been experimentally validated so far does not reveal any obvious characteristics of JNK-specific motifs (Figure 2.5). The difficulty in delineating a consensus binding sequence is not a specific MAPK feature, but a general feature of motifs within disordered proteins. This has resulted in the development of alignment algorithms specialized for such motifs (Frith et al., 2008).

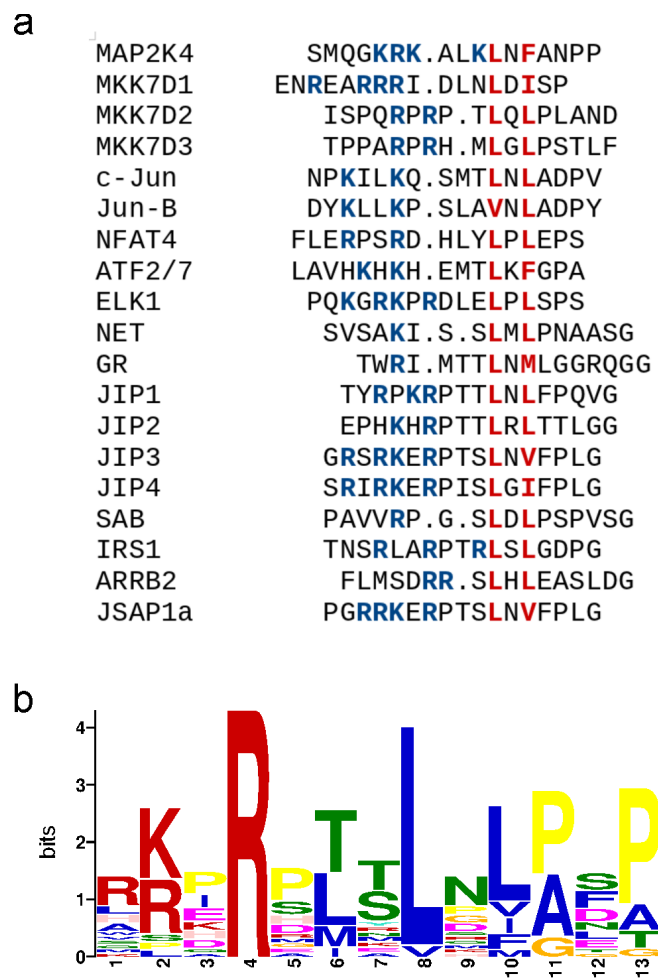


Figure 2.5: Alignment of experimentally verified JNK-specific docking-site motifs. a) GLAM2 alignment of nineteen verified JNK binding motifs. b) Sequence logo for the alignment generated by GLAM2. The sequences are taken from reference (Whisenant et al., 2010).

As it was already pointed out, the MAPK docking-site motifs reside in IDRs of MKKs. In general, binding motifs are often located in flexible and disordered regions. In the following section, I will show how this is not an exclusive characteristic of docking-site motifs from MAPK pathways, but a feature of motifs from many signalling networks.

2.2 Intrinsically disordered proteins and regions

2.2.1 Intrinsic disorder is predicted in every second protein

IDPs are not folded under physiological conditions but are still able to carry out their function despite the lack of a three-dimensional structure (Tompa, 2002). These proteins usually have low sequence complexity and an amino acid compositional bias. In comparison to folded proteins they are depleted in hydrophobic and aromatic amino acids and can be highly positively or negatively charged (Campen et al., 2008). Methods used in early characterization of IDPs were circular dichroism (CD), nuclear magnetic resonance (NMR) and X-ray crystallography (described in reviews (Dunker et al., 2001; Wright and Dyson, 1999)). The term disordered comes from the terminology used in X-ray crystallography. Electron density is missing from an X-ray structure when the atoms for that protein segment adopt different positions and do not scatter coherently and we say that these segments are disordered.

The first evidence that disorder is not an exception but present in many proteins came from bioinformatics studies. Neural network algorithms, that were trained on data sets of experimentally characterized disordered proteins, predicted that many proteins contain longer disordered regions (Romero et al., 1998). It was estimated that around 50% of proteins in eukaryotes have regions of disorder that are longer than 30 amino acids (Dunker et al., 2000). For eukaryotic signalling proteins the percentage of disordered segments longer than 30 amino acids is even higher and reaches 65% (Iakoucheva et al., 2002) and among human protein kinases (the PKinase Pfam family which includes MAPKs) more than 70% of the kinases have regions of disorder longer than 30 amino acids (Kathiriya et al., 2014).

The accumulating experimental evidence is rarely proving the predictions wrong which means that many proteins indeed contain IDRs. The proteins which are fully disordered from end-to-end comprise 5–10% of eukaryotic genomes, depending on the organism (Dunker et al., 2000). They are often used as the main argument supporting the cause for "function through disorder". These proteins are justifiably called IDPs.

2.2.2 IDRs are a functional part of eukaryotic proteins

Some eukaryotic proteins are either fully disordered or structured, but most have both types of regions and often several domains and/or several IDRs. In metazoa, 80% of the proteins are multidomain proteins (Apic et al., 2001) where the domains often perform different specific functions (SH3, SH2, kinase domains, zinc-fingers etc.), so that new combinations of such domains can give rise to new functions (Cohen et al., 1995; Lim, 2002; Pawson, 1995).

The modular organization of proteins is taken into account in protein sequence annotation algorithms. For example, the Pfam database of protein domains searches in a given sequence for domains of known families. Recently, annotation of IDRs was introduced in Pfam because, as they state, “As part of recent, focused curation efforts aimed at increasing the Pfam-A coverage of the human proteome (Mistry et al., 2013), it became apparent that many regions not covered by Pfam-A are predicted to be intrinsically disordered.” (Finn et al., 2014). The eukaryotic linear motif resource (ELM) takes the annotation a step further and also displays potential functional motifs in IDRs together with outputting Pfam or SMART annotations for folded domains (Dinkel et al., 2014).

2.2.3 Motifs often appear in intrinsically disordered regions

MKKs are modular in the sense that the N-terminal IDR which contains the docking-site motif directs the catalytic domain towards the substrate (a MAPK). The kinase domain is obviously necessary for carrying out the phosphorylation but the docking-site motifs of MKKs determine the specificity and the connections within the MAPK signalling networks. Regulatory protein interactions are often mediated by motifs which are typically located in IDRs (Davey et al., 2011; Fuxreiter et al., 2007; Ren et al., 2008). These motifs can be divided in roughly two classes: motifs that mediate binding and sites of post-translational modification (PTM). PTMs have also been shown to occur predominantly in IDRs (Collins et al., 2008; Diella et al., 2008; Iakoucheva, 2004; Xie et al., 2007).

Motifs allow for many types of regulation (Van Roey et al., 2012). Post-translational modification such as phosphorylation can turn a low-affinity binding motif into a high-affinity binding motif (Perkins et al., 2010; Van Roey et al., 2012). Nuclear import signals are often located in IDRs and conditionally accessible to regulate the import and export from the nucleus (Meng et al., 2002). Motifs can also be regulated by pretranslational modifications, such as alternative splicing and alternative promoter usage (Van Roey et al., 2013; Weatheritt and Gibson, 2012). With these and many other opportunities for regulation, motifs are an important part of signalling networks.

2.2.4 Pre-structured motifs

Many structures have been solved by both crystallography and by NMR where IDPs adopt a well-defined structure in complex with another protein. Assuming that they were fully disordered before binding one can propose the idea of folding upon binding (Dyson and Wright, 2002). An obvious advantage of such folding and binding is that the structure of a binding region can be adapted to bind several completely different interacting partners (Vacic et al., 2007). In the complex, the motifs often adopt α -helical, extended β - or polyproline type II (PPII) conformations (Figure 2.6). Turns and various coil conformations are also common which shows that there is no limitations to which secondary structure type the motifs adopt in complexes.

On the other hand, motifs are often pre-formed in the sense that residual structure corresponding to the structure in the complex can be observed in the unbound protein (Fuxreiter et al., 2004). Motifs that bind in α -helical conformations are best studied examples of pre-formed motifs, probably because they are indeed abundant but also because they are the easiest to observe (CD, NMR) and predict (Buchan et al., 2013; Lacroix et al., 1998; Muñoz and Serrano, 1994, 1995; Serrano, 1995). Development of methods that would allow us to describe β -strand and PPII residual structure is needed to fully understand the pre-formed motifs.

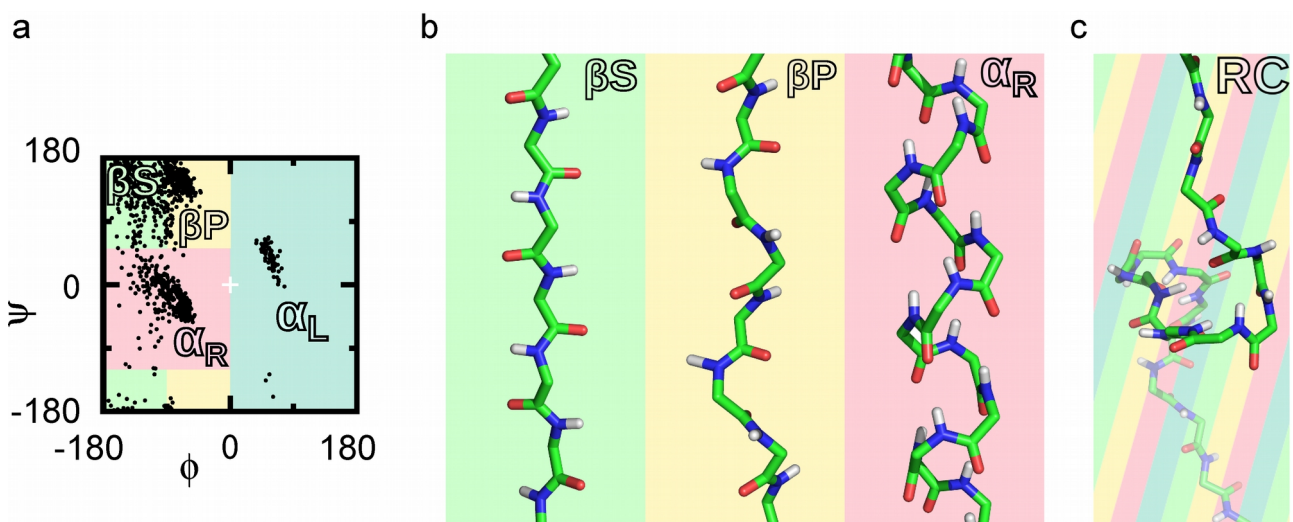


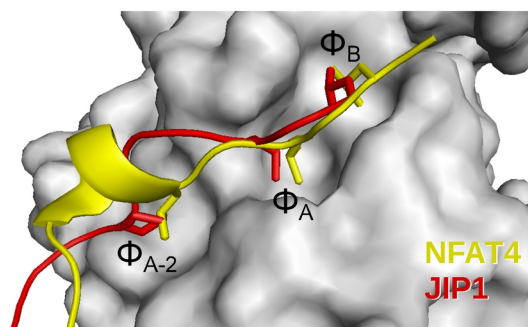
Figure 2.6: Division of Ramachandran space into four regions: α -helix (α_R) and β -sheet (β_P), poly-L-proline II (PPII) (β_P) and left-handed helix (α_L). The four regions are defined as follows: α_L $\{\varphi > 0^\circ\}$; α_R $\{\varphi < 0, -120^\circ < \psi < 50^\circ\}$; β_P $\{-100^\circ < \varphi < 0^\circ, \psi > 50^\circ \text{ or } \psi < -120^\circ\}$; β_S $\{-180^\circ < \varphi < -100^\circ, \psi > 50^\circ \text{ or } \psi < -120^\circ\}$. b) α -helix, β -sheet and poly-L-proline II (PPII) are the only three secondary structures in folded proteins that form linear groups of residues by repetition of one φ/ψ angle pairs. c) Random coil (more accurately statistical coil) samples all four regions of the Ramachandran space.

The mechanisms of binding of IDPs can be described as mostly proceeding through folding after binding or through conformational selection and for each of these two cases different advantages could exist (Kiefhaber et al., 2012). As these are descriptions of two extreme cases, further complications in analysis of binding mechanisms arise because the binding usually occurs through both mechanisms at the same time but to a different extent. Very recent studies have aimed at understanding the relationship between pre-formed structures and kinetics (Iešmantavičius et al., 2014; Rogers et al., 2014). Nevertheless this field still leaves many open questions with only few experimental data obtained at this moment about association/dissociation kinetics and binding pathway trajectories. This is therefore still a very active field of research.

2.3 Motifs in signalling networks

2.3.1 Structure of motifs in complexes

JNK1



	Φ_{A-2} Φ_A Φ_B
JIP1-DJ:	PGSGDTY R K R T T N L F P Q V P R S Q D
NFAT4 :	REFLE R P S R D H L - Y L P L E P S Y R E S S L

Figure 2.7: Comparison of docking interactions of two peptides bound to JNK1. The surface of JNK1 is displayed in grey colour, peptides are shown as ribbons with side chains of hydrophobic amino acids displayed as sticks. The hydrophobic amino acids bind to three conserved pockets (Φ_{A-2} , Φ_A and Φ_B). Alignments of longer sequences based on co-crystal structures of short peptides are shown below the kinase figure. The residues of the basic and hydrophobic part of the docking-site motif are highlighted in purple and green, respectively. Proline residues are marked in blue. Structures used for this figure can be found under the following PDB codes: 1UKH, 2XRW.

the last two pockets (Φ_A , Φ_B) seem to adopt similar conformations. The Φ_{A-2} pocket is most often

When the docking-site motifs are bound to MAPKs they adopt conformations that fall within extended β - and PPII regions. The basic cluster of the bound docking-site motifs adopts different conformations in different complexes (Figure 2.7, 2.8). The interaction surface where the basic portion of the docking-site motif binds is therefore less conserved from one MAPK to the another but also, along another line of interpretation, the basic part of the docking motif might not adopt a single conformation in the complex in solution. In support of this interpretation, the electron density for basic clusters of docking-site motifs is missing in several docking-site–MAPK complexes.

The hydrophobic part of the motifs interacts with a well-conserved surface on the MAPKs. The interface encompasses three hydrophobic pockets (Φ_{A-2} , Φ_A and Φ_B). Residues binding into

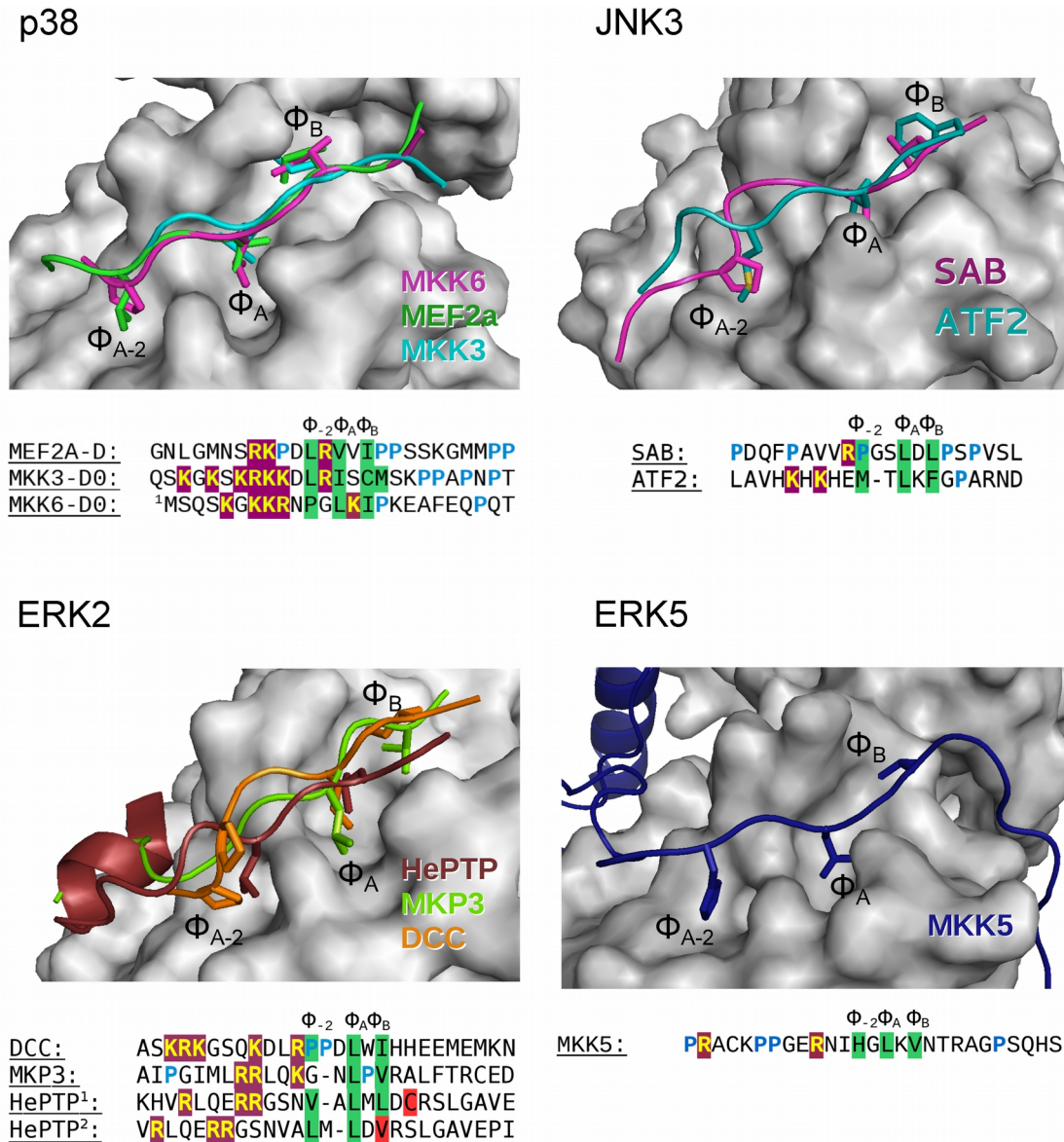


Figure 2.8: Comparison of docking interactions of MAPKs. Different MAPKs bind docking-site motifs in a similar conformation. The surface of the MAPK kinases are displayed in grey colour, peptides are shown as ribbons with side chains of hydrophobic amino acids displayed as sticks. The hydrophobic amino acids bind to three conserved pockets in all four MAPKs (Φ_{A-2} , Φ_A and Φ_B). Alignments of longer sequences based on co-crystal structures of short peptides are shown below each kinase. The residues of the basic and hydrophobic part of the docking-site motif are highlighted in purple and green, respectively. Proline residues are marked in blue. Note that the structures of the MKK3 docking-site in complex with p38 and HePTP1 in complex with ERK2 were obtained by cross-linking providing less confidence in the sequence alignments. For HePTP1 sequence in the alignment of ERK2 peptides, the residue that was mutated to cysteine to allow cross-linking is coloured in red. The alignment marked as HePTP2 is based on a study where the non-mutated peptide was used and the binding was followed by NMR (Francis DM et al., 2011). Structures used for this figure can be found under the following PDB codes: 1LEW, 1LEZ, 2Y80, 4H3B, 4H36, 2GPH, 3071, 2FYS and 4IC7.

occupied by a leucine, sometimes by a proline, in the case of ERK5 by a histidine and SAB peptide binds to Φ_{A-2} pocket of JNK1 with a methionine. From the comparison of MAPK-MKK peptide co-

crystal structures it is possible to see that the conformation of the peptides in the complex is not the factor used to distinguish between different MAPK classes as all peptides are obviously able to adopt very similar conformations. Furthermore, the motifs of all three MAPK classes (JNK, p38, ERK) bind to the three pockets which is a common structural feature of all MAPKs (Figure 2.7, 2.8).

2.3.2 Affinity and specificity of motifs

Accurate affinities of docking-site motifs for MAPKs have been obtained only recently and range from 3.5 to 18.7 μM for specific substrates (Garai et al., 2012). A survey of affinities of other signalling proteins that are not related to MAPK signalling shows that they are also in the low micromolar range. When peptide ligands of SH3 domains are synthesized based on specific motifs they usually have affinities in the 10 μM range (Mayer, 2001), although lower affinities have also been measured. For SH2 domains, affinities for specific peptide ligands range from 0.5 to 5 μM (Liu et al., 2010). But affinities of specific interactions of motifs can also be much lower. In the case of WW domains the dissociation constants for specific interactions range from 10 to 200 μM (and for some even lower) (Kato et al., 2002, 2004).

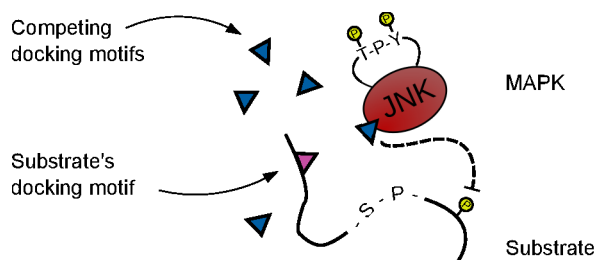


Figure 2.9: Peptides with docking-site motif sequences (blue triangle) are used to inhibit binding of the substrate and its phosphorylation.

The affinities tell us only half of the story about specificity. Specificity is the ability of the domain to bind certain motifs with higher affinity than all the other similar motifs. The first large scale study of specificity between docking-site motifs of MKKs and MAPKs was performed by Bardwell et al (Bardwell et al., 2009). To measure the strengths of peptide MAPK interactions, IC_{50} values were measured in a substrate competition enzymatic assay (Figure 2.9). IC_{50} is the

concentration of competing ligand that results in 50% reduction in substrate phosphorylation.

In the substrate competition enzymatic assays the high affinity peptides were able to out-compete the substrate of the kinase and significantly inhibit the reaction already at low concentrations (the IC_{50} was low). A peptide that is highly specific will display low IC_{50} values in an assay with a cognate kinase and high IC_{50} values in an assay with non-cognate kinases. The ratio between the IC_{50} , cognate and IC_{50} , non-cognate which is defined as the selectivity ratio and is mostly between 4

and 20 (Table 2.2). There are also a few exceptions. Interaction between MKK3 and p38 is of very high affinity, while the interactions of MKK3 with other MAPKs are of low affinity (Table 2.2). This causes the selectivity ratio of MKK3 to be close to 1000 or even higher. On the other extremity, there is ERK2 which interacts with MKK1 with an IC_{50} of 32 μ M (Table 2.2). The IC_{50} values of non-cognate docking-site peptides for ERK2 are within that range and sometimes lower which causes the selectivity ratio to be below or equal to one. Some of ERK2 non-cognate docking-sites therefore bind to ERK2 with lower affinity than the cognate docking-sites of ERK. But ERK2 is an exception and JNK1 docking-sites for which the competition assay was carried out display a reasonable degree of specificity (Table 2.2).

Table 2.2: Preference of MAPKs for docking motifs^a

MAPK/substrate	cognate peptide	IC_{50} [μ M]	non-cognate peptide	IC_{50} [μ M]	Selectivity ratio ^b
ERK2/Elk-1	MKK1	32	MKK3	5	0.2
ERK2/Elk-1	MKK1	32	MKK6	17	0.5
ERK2/Elk-1	MKK1	32	MKK4	26	0.8
ERK2/Elk-1	MKK1	32	MKK7-D2	85	3
JNK1/ c-Jun	MKK7-D2	8	MKK1	> 100	> 13
JNK1/ c-Jun	MKK7-D2	8	MKK2	60	8
JNK1/ c-Jun	MKK7-D2	8	MKK3	90	11
JNK1/ c-Jun	MKK7-D2	8	MKK6	> 100	> 13
JNK1/ c-Jun	MKK4	2	MKK1	> 100	> 50
JNK1/ c-Jun	MKK4	2	MKK2	60	30
JNK1/ c-Jun	MKK4	2	MKK3	90	45
JNK1/ c-Jun	MKK4	2	MKK6	> 100	> 50

^aData in the table is adapted from (Bardwell et al., 2009)

^bSelectivity is defined as $IC_{50}(\text{MAPK with non-cognate docking motif})/IC_{50}(\text{MAPK with cognate docking motif})$

A recent study refined the results of Bardwell et al. (Garai et al., 2012). Using fluorescence polarization anisotropy techniques Kds were measured for several MKKs and MAPK interacting peptides. Measured Kds for cognate ligands were 2 to 20 times lower than for peptides from non-cognate MKKs (Table 2.3). The high selectivity ratio of MKK3 towards p38 observed by competition assays is not supported by the measured Kds. The reason for this disagreement may be a shorter synthesized peptide (KGKSKRKKDLRIS) that is missing residues that could bind to the Φ_B pocket of p38.

It is hard to make exact comparisons between the two studies because the relation of IC_{50} with Kd is

not simple, but both studies qualitatively agree on a few points which allows us to draw the following conclusions: docking-sites of MKKs from the JNK1 pathway do not bind strongly to p38 and ERK2 with the exception of MKK4 which a well known physiological binding partner of p38; ERK2 is able to bind MKKs from both the p38 pathway and ERK pathway but only ERK specific MKK1/2 kinases phosphorylate ERK due to additional determinants around the phospho-residues in the loop of ERK (Garai et al., 2012).

Table 2.3: Binding affinities of peptides containing docking motif sequences^a

cognate peptide	JNK [uM]	p38α [uM]	ERK2 [uM]
MKK3	> 100	9.1	18.7
MKK6	> 100	7.5	9.7
MKK1	> 100	32.6	18.7
MKK4	3.5	3.7	19.1
JIP1	5.9	> 100	> 100
NFAT4	7.1	> 100	> 100
MKK7-D2	15.3	> 100	> 100

^aData in the table is adapted from (Garai A et al. 2012)

2.3.3 Combining specificity of motifs and localization

How can such small sequence modulations in motifs allow MKKs to adopt different specificities? It has been suggested that in yeast *Saccharomyces cerevisiae* the SH3 binding motifs are optimized across the entire yeast proteome such that the SH3 binding motifs of the 27 existing yeast SH3 domains do not overlap. Evidence for this comes from experiments carried out with the SH3 domains from Sho1 and its binding motif from the Pbs2 protein (Zarrinpar et al., 2003). A protein in which the Sho1 SH3 domain was replaced with any of the other SH3 domains from yeast was not able to perform its function, because the Pbs2–Sho1(mutant) interaction was lost. There was one exception in that study though. The only yeast SH3 domain that was able to bind Sho1 PXXP motifs and rescue the function has a completely different sub-cellular organization and does not interfere with Sho1 signalling *in vivo*. This points towards another principle in the organization of signalling networks: localization can govern the output of signalling. There are more examples. A GFY domain of CD2BP2 protein and an SH3 domain of Fyn kinase can both bind to a proline rich segment of CD2 *in vitro* as shown by NMR (Freund et al., 2002). *In vivo* though, these interactions do not compete for the proline rich segment of CD2 because they are spatially separated, one being associated with the detergent-soluble membrane fraction and the other being detected only in lipid rafts (Freund et al., 2002). Another study showing the importance of sub-cellular localization is a

study of specificity of mitotic kinases for their phosphorylation motifs (Alexander et al., 2011), where selectivity is achieved through selection of positive motifs (residues that contribute to higher affinity) and also negative motifs (residues that abolish binding). This suggests that a substrate of a kinase tries to avoid getting phosphorylated by other kinases by incorporating residues that reduce binding to off-pathway kinases (adopting an antimotif of the non-cognate kinases), while keeping its recognition by the cognate kinase undisturbed.

2.3.4 Cooperativity from other regions

Motifs are defined by sequence alignment and bioinformatics which search for consensus sequences. Alignments tell us about sequence conservation and conserved residues are very likely involved in binding. Nevertheless it has to be kept in mind that other, less conserved positions, might also be involved in binding and could fine tune the affinity. Biophysical methods are often performed on systems that are easy to synthesize, purify and behave well. The synthesized peptides or expressed constructs might sometimes be too short which could lead to lower measured affinities. Contributions from residues outside of the consensus sequence of the motifs may provide a way of increasing the affinities of the motifs. An example is the p67phox SH3 domain that binds to a 32-residue fragment of p47phox with a K_d of 24 nM. Only the canonical proline rich region of the p47phox fragment binds to the SH3 of p67phox with 1000-times lower affinity. The structure of the complex explains this by showing how the C-terminal part of the long 32-residue fragment makes contacts outside of the typical SH3 binding site (Dutta et al., 2004; Kami et al., 2002).

For MAPKs, additional contacts from motif proximal residues which are themselves not part of the docking-site have been identified for the MKK5-ERK5 interaction, where both the docking-site motif and an additional N-terminally located PB1 domain contribute to tighter binding (Glatz et al., 2013). As I have explained above, MKKs discern between cognate and non-cognate MAPKs by cooperation of docking-sites and residues around phospho-residues. The ERK kinase can bind the docking motifs of cognate MKK1/MKK2 and non-cognate MKK3/MKK6, but only MKK1/MKK2 are able to phosphorylate ERK because of the specific sequence (T-E-Y, Figure 2.1) around residues that need to be phosphorylated (Garai et al., 2012).

2.3.5 The JIP1 scaffold protein

It might be a general rule that in regulatory networks the signal is passed through complexes/assemblies formed by multiple weak interactions thereby increasing specificity and possibilities for regulation (Gibson, 2009; Whitty, 2008). Many examples of scaffold proteins are implicated in JNK pathways. These are proteins, that bind several components of a signalling pathway and thereby increase the efficiency of signalling, specificity, contribute to control of the signal transduction and can localize the components to different cellular compartments (Buday and Tompa, 2010; Good et al., 2011). Some of the better known JNK pathway scaffold proteins are JIP1, JIP2, JIP3, JLP, POSH, β -arrestin 2, β -arrestin 3, WDR62, GRASP-1 and Hic-5 (Cohen-Katsenelson et al., 2011; Guo and Whitmarsh, 2008; Kook et al., 2013; Lee et al., 2002; Lei et al., 2014; Wasserman et al., 2010; Whitmarsh, 2006; Xu et al., 2003).

The JNK interacting protein 1 (JIP1) was the first discovered JNK scaffold protein (Whitmarsh et al., 1998). It is a 711 amino acid long protein with an IDR at the N-terminus which is followed by two smaller folded domains at the C-terminal end: a Src Homology 3 (SH3) and a phosphotyrosine interaction domain (PID) (Figure 2.10, 2.11). It has been shown that it binds both JNK1, MKK7 and

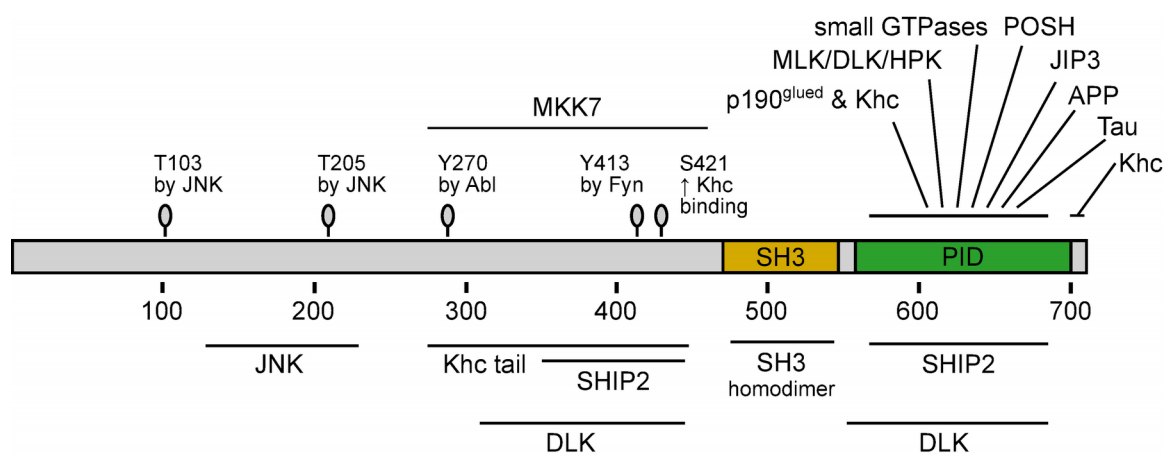


Figure 2.10: Box representation of the JIP1 protein. JIP1 is a 711 amino acid long polypeptide chain containing a JNK docking-site, SH3 domain (yellow) and a PID domain (green). Several interactions have been mapped to the JIP1 protein (lines). Phosphorylation sites are shown with lollipop marks immediately above the box representation of JIP1.

upstream MKKKs from the JNK pathway, while it does not bind the members of p38 and ERK pathways indicating an important role in signalling specificity (Heo et al., 2004; Nihalani et al., 2001; Whitmarsh et al., 1998). JNK1 binds to JIP1 through a docking-site motif located between residues 157-176 (Whitmarsh et al., 1998). Apart from the MAPKs there are other proteins with

which JIP1 interacts as shown in Figure 2.10 (Fu and Holzbaur, 2013; Ittner et al., 2009; Kukekov et al., 2006; Matsuda et al., 2001, 2003; Meyer et al., 1999; Mooney and Whitmarsh, 2004; Nihalani et al., 2001; Satake et al., 2013; Stockinger et al., 2000; Vaishnav et al., 2011; Whitmarsh et al., 1998; Xie et al., 2008; Xu et al., 2003). Colocalization of a JNK, an MKK and an MKKK with possibly other effector proteins would be a way to assure additional specificity, regulation and help the signalling to proceed most efficiently. In addition it has been shown that JIP1 can be found in different phosphorylation states, as a monomer and as a dimer and that these events control its cellular localization and function (Borsello et al., 2007; Nihalani et al., 2001, 2003).

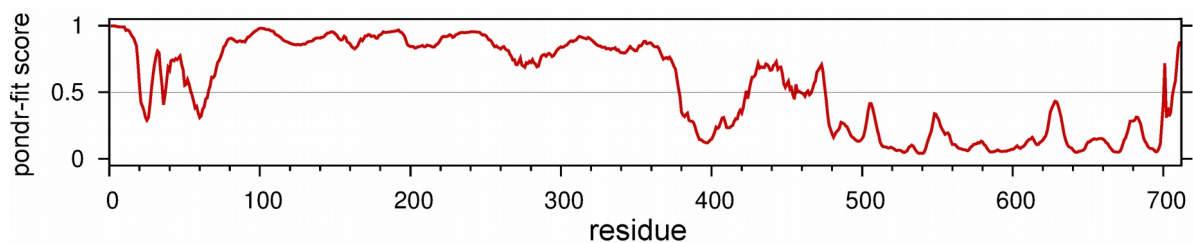


Figure 2.11: PonderFit disorder prediction for the JIP1 scaffold protein. A score above 0.5 indicates that the residue is predicted to be disordered while a score below 0.5 indicates that the residue is predicted to be ordered.

2.4 Objectives

The overall goal of this thesis is to better understand the role of docking sites of MKKs in determining signalling specificity within the MAPK pathways. In particular, I am interested in signalling specificity within the JNK pathway, which is activated by the two kinases MKK4 and MKK7. As described above, these two proteins contain large N-terminal IDRs for which no structural information is available. One of the goals of this thesis is therefore to obtain an atomic resolution structural and dynamic characterization of the IDRs of MKK4 and MKK7, with particular emphasis on obtaining the conformational sampling of their docking sites. To this end, I will apply NMR spectroscopy in combination with state-of-the-art structural ensemble selection techniques allowing me to address whether docking-site motifs are structurally pre-configured prior to binding to their cognate kinases.

A second objective of this thesis is to study the interaction between the IDRs of MKK4 and MKK7 and their cognate kinase JNK. MKK7 is the only MKK for which three putative docking-sites are present within its N-terminal IDR posing a number of questions concerning the affinity, stoichiometry, structure, dynamics and kinetics of the MKK7-JNK signalling complex. To address these questions, I will apply isothermal titration calorimetry to obtain binding affinities combined with NMR chemical shift titrations to delineate the minimal binding motifs, allowing me to address whether residues outside the canonical docking-site motifs contribute to JNK binding. In addition, I will apply NMR exchange spectroscopy, notably relaxation dispersion and chemical exchange saturation transfer experiments, to probe the kinetics of the interaction of JNK with the different docking sites. This will allow me to determine potential contributions from the pre-recognition sampling of the docking-sites to the binding affinities and kinetics.

The structural basis of the interaction between JNK and the docking sites of MKK4 and MKK7 is currently unknown. A third objective of this thesis is therefore to determine the crystal structure of JNK in complex with one or several of the docking sites from MKK4 and MKK7. This will allow me to identify key residues essential for the interaction between the docking-sites and JNK. In addition, as presented above, the consensus sequence of JNK docking-site motifs is poorly defined. The information obtained from crystal structures can be used to better align the sequences of the motifs and potentially predict how other docking-site motifs bind to JNK. In this thesis, I will address this possibility, and I will make use of extensive sequence alignments across various organisms to determine the importance of different docking-site motifs.

The final objective of this thesis is to understand the role of JIP1 in mediating signalling specificity. This involves characterizing the long N-terminal IDR of JIP1 at atomic resolution allowing me to eventually map the interacting regions with members of the JNK signalling pathway.

2. *Intrinsic disorder in signalling proteins*

3 Nuclear magnetic resonance

3.1 Quantum description for a spin-half nucleus

NMR spectroscopy characterizes transitions between magnetic energy levels of nuclei possessing an intrinsic magnetic moment and angular momentum, a physical property described as nuclear spin. Spin is defined as a product $\hbar \hat{S}$ where \hbar is the Planck constant divided by 2π and \hat{S} is the spin operator defined by spin quantum number s :

$$\hat{S} = [s(s+1)]^{1/2} \quad (3.1)$$

\hat{S}_z is the component or projection of this spin angular momentum along an arbitrary z-axis and is characterized by the quantum number m that takes values from $-s$ to $+s$ in steps of 1.

The magnetic moment $\hat{\mu}_s$ is proportional to the spin angular momentum and is given by:

$$\hat{\mu}_s = \gamma \hbar \hat{S} \quad (3.2)$$

where γ is the gyromagnetic ratio of the nucleus in question. For nuclei with positive value of γ , the magnetic moment is parallel to the spin angular momentum. For nuclei with negative value of γ , the magnetic moment is opposite in direction to the spin angular momentum.

In the presence of a static magnetic field B_0 , the Hamiltonian which describes the energy of interaction between an isolated spin and the magnetic field is:

$$\hat{H} = -\gamma B_0 \hat{S}_z \quad (3.3)$$

This is called the nuclear Zeeman interaction. \hat{S}_z has $(s+1)$ eigenfunctions each of which is characterised by the previously mentioned quantum number m . For a spin-half nucleus ($s = 1/2$) the \hat{S}_z has two eigenfunctions, characterized by $m = -1/2$ and $m = +1/2$, whose energies are given by:

$$E_\alpha = -\frac{1}{2} \hbar \gamma B_0 \quad \text{and} \quad E_\beta = \frac{1}{2} \hbar \gamma B_0 \quad (3.4)$$

3. Nuclear magnetic resonance

and with an energy difference:

$$\Delta E = \hbar \gamma B_0 = h \nu_0 \quad \Rightarrow \quad \nu_0 = \gamma B_0 / 2\pi \quad \Rightarrow \quad \omega_0 = \gamma B_0 \quad (3.5)$$

The magnetic moment of the spin precesses around the direction of the B_0 field with a so-called Larmor frequency defined in equation 3.5. The transitions between the α and β states are observable at this frequency and constitute the basis of NMR spectroscopy. The intensity of the signal observed in a spectrum depends on the population difference between the two states. From equation 3.5 it can be seen that difference in the energy is proportional to the B_0 field. For a given temperature and B_0 field the population difference between the two states can be described with the Boltzmann equation:

$$\frac{N_\alpha}{N_\beta} = \exp\left(\frac{\Delta E}{k_B T}\right) \quad \text{where} \quad \Delta E = \gamma B_0 \hbar \quad (3.6)$$

N_α and N_β are the populations of the spins in the lower and higher energy states, respectively and k_B is the Boltzmann constant. For a proton in a magnetic field B_0 of 14.1 T (600 MHz proton Larmor frequency) at $T = 298$ K, the ratio between N_α and N_β is:

$$\frac{N_\alpha}{N_\beta} = 1.000097 \quad (3.7)$$

The small excess of the state α gives rise to the NMR signal. This is why NMR is said to be an intrinsically insensitive technique. Nevertheless the excess population of nuclei in the state α is sufficient for allowing all the different types of measurements that can provide insight into the microscopic world of molecules.

In a sample where many spins precess around the B_0 field, each with its own associated magnetic moment (Eq. 3.2), the sample acquires bulk magnetization along the z-axis as a consequence of the differently populated α and β states. In the transverse xy-plane, there is no energetic preference for any particular orientation, which means that x- and y-components of the individual magnetic moments are distributed randomly and cancel each other out. The bulk magnetization can be described as:

$$M_z^0 = \gamma^2 \hbar^2 N \frac{B_0}{4k_B T} \quad (3.8)$$

where N is the number of spins in the sample. For any kind of NMR experiment, this equilibrium is perturbed by excitation into higher energy states. For example, in a simple NMR experiment composed of a 90° pulse, the magnetization precesses in the transverse plane giving rise to the well-known free induction decay (FID) signal. After such a perturbation the FID signal does not oscillate for an infinite time, but exponentially decays to zero, indicating a return to equilibrium. The process of returning to equilibrium condition is called relaxation. Before proceeding to discussing relaxation in section 3.2, I will describe the dipolar interaction and chemical shifts.

3.1.1 Dipolar interaction

Dipolar couplings are caused by magnetic dipole-dipole interactions between half-spin nuclei (^1H , ^{13}C , ^{15}N). These are large interactions caused by magnetic flux lines affecting the magnetic field at the site of the other nucleus (Fig. 3.1).

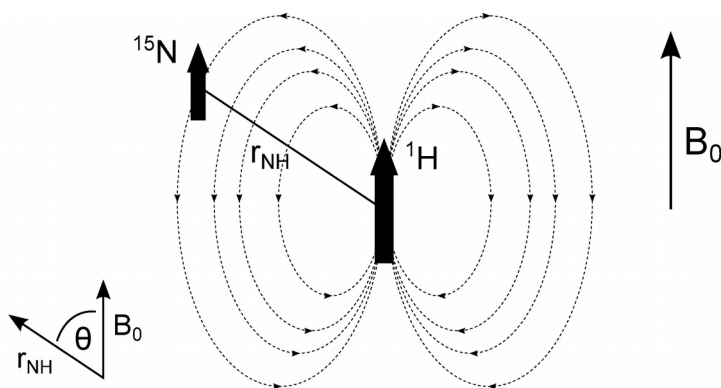


Figure 3.1: Dipole-dipole interactions. In a magnetic field B_0 , the magnetic field for the ^{15}N magnetic moment is affected by magnetic flux lines of the ^1H dipole. For a pair of selected nuclei at a fixed distance, the magnitude of the Z component of the magnetic field for the ^{15}N nucleus depends on the angle theta.

The Z component of the magnetic field around nucleus i will have an additional contribution from the Z component of the dipolar field of nucleus j . The magnitude of this contribution will depend on the internuclear distance and on the orientation of the internuclear vector relative to the static magnetic field B_0 . Because half of the j spins will be parallel to B_0 and half anti-parallel to B_0 , the

nucleus i will show two resonances (a doublet), separated in frequency by:

$$D^{ij} = -\frac{\mu_0 \hbar \gamma_i \gamma_j}{4\pi^2 r_{ij}^3} \langle (3\cos^2\theta - 1)/2 \rangle = D_{max}^{ij} \langle (3\cos^2\theta - 1)/2 \rangle = D_{max}^{ij} \langle P_2(\cos\theta(t)) \rangle \quad (3.9)$$

where θ is the angle of the internuclear vector relative to the static magnetic field, r is the internuclear distance and μ_0 is the electric permittivity of vacuum. The angular parentheses define an average over all conformations exchanging on timescales faster than the millisecond. For proteins in solution the dipolar couplings efficiently average to zero because of the isotropic rotation of the molecules in solution. When the dipolar couplings are not averaged out they can lead to very large splittings (around 11 kHz for the ^{15}N - ^1H spin pair).

3.1.2 Chemical shift and chemical shift anisotropy

Chemical shift is a change in the Larmor frequency of a proton due to other smaller contributions to the B_0 field from the local electronic environment. When an external B_0 is applied, an induced local magnetic field is created by the electrons of the molecule experiencing the external field. The effective magnetic field experienced by a nucleus is therefore

$$B_{eff} = B_0 + B_{induced} \quad (3.10)$$

With $B_{induced} = \sigma_i B_0$ and σ_i being a second-rank nuclear shielding tensor. Equation 3.10 can then be rewritten as:

$$B_{eff} = B_0 (1 - \sigma_i) \quad (3.11)$$

with 1 being the unit matrix.

In liquid samples, where the molecular tumbling is fast, the shielding tensor is isotropically averaged and the chemical shift can be represented by a scalar quantity δ_{iso} :

$$\delta_{iso} = \frac{1}{3} (\delta_{XX} + \delta_{YY} + \delta_{ZZ}) \quad (3.12)$$

where the δ_{XX} , δ_{YY} and δ_{ZZ} are the components of the chemical shift anisotropy (CSA) tensor

transformed to a frame of reference defined by axes X, Y and Z in the principal axis system. The equation 3.5 can then be modified to give:

$$\nu_0 = -\gamma B_0(1 + \delta_{iso}) / (2\pi) \quad (3.13)$$

More often the chemical shifts are expressed in units of ppm according to the definition:

$$\delta_{ppm} = 10^6 \times \frac{\nu_0 - \nu_{ref}}{\nu_{ref}} \quad (3.14)$$

where ν_{ref} is the frequency of the reference. For biomolecular NMR where proteins are dissolved in water the recommended reference is DSS (2,2-dimethyl-2-silapentane-5-sulfonate) (Wishart, 2011). To avoid mixing the reference compound with the delicate protein sample the referencing is often done through indirect referencing (Wishart et al., 1995a).

3.2 Relaxation theory

3.2.1 Sources of relaxation

Relaxation is a process that occurs due to fluctuations of magnetic fields in the vicinity of the spin. The fluctuations can induce transitions that bring the system back to equilibrium or they can vary the Larmor frequency and irreversibly dephase their coherences. In both cases the source of magnetic field fluctuations is either the dipolar interaction or the CSA. As we are only interested in ^{15}N relaxation we only consider the N-H dipole (Section 3.1.1) and the ^{15}N CSA.

3.2.2 Spectral density

The enumerated sources of differing local magnetic fields could not cause relaxation if they would not occur at the right frequencies, that is close to the Larmor frequencies. Bond vibrations that occur on the time scale of 0.1 to 10 ps (10^{-13} – 10^{-11} s) are too fast to cause relaxation. A much better candidate, which turns out to be the major cause of fluctuations at the appropriate frequency, is rotational diffusion. Rotational diffusion can be quantified using a correlation time τ_c which is the average time it takes for a molecule to change its orientation by 1 radian, where the average is taken over a large number of molecules from the sample. Correlation times in the range of 10 ns to 10 ps give frequencies in the range of 10^7 to 10^{10} Hz ($\nu = 1/(2\pi\tau_c)$). This is comparable with typical

Larmor frequencies but as we want to quantify the amount of motion present at the exact frequency it is useful to introduce the correlation function and the spectral density. The correlation function describes the change in the local magnetic field as a function of time. Its Fourier transform into the frequency domain is called the spectral density $j(\omega)$ and tells us about the amount of motion at a specific frequency.

3.2.3 Longitudinal relaxation

The z-magnetization is driven back to equilibrium by longitudinal relaxation R_1 . The information about dynamics can be obtained from longitudinal relaxation if we express it in terms of a sum of spectral densities (for an ^{15}N - $^1\text{H}^{\text{N}}$ pair):

$$R_1^{(NH)} = \frac{1}{4} d_{NH}^2 (3j(\omega_N) + j(\omega_H - \omega_N) + 6j(\omega_H + \omega_N)) + a_N^2 6j(\omega_N) \quad (3.15)$$

$$\text{with } d_{NH}^2 = \left(\frac{\gamma_H \gamma_N \mu_0 \hbar}{4\pi r_{NH}^3} \right)^2 \quad (3.16)$$

$$\text{and } a_N^2 = \frac{(\gamma_N B_0 (\sigma_{par} - \sigma_{per}))^2}{18} \quad (3.17)$$

where σ_{par} and σ_{per} are the parallel and perpendicular component of the shielding tensor, respectively. The $j(\omega_N)$, $j(\omega_H - \omega_N)$ and $j(\omega_H + \omega_N)$ in the equation 3.15 are the spectral density at the frequencies of the transitions between energy levels of the ^{15}N - $^1\text{H}^{\text{N}}$ pair: ω_N is the Larmor frequency of nitrogen and corresponds to the single-quantum transition ($\alpha\alpha \leftrightarrow \beta\beta$), $j(\omega_H - \omega_N)$ is the difference between Larmor frequencies of hydrogen and nitrogen and corresponds to the zero-quantum transition ($\alpha\beta \leftrightarrow \beta\alpha$), $j(\omega_H + \omega_N)$ is the sum of of Larmor frequencies of hydrogen and nitrogen and corresponds to the double-quantum transition ($\alpha\alpha \leftrightarrow \beta\beta$).

3.2.4 Heteronuclear NOEs

Another relaxation phenomenon that can be used for measuring dynamics is cross relaxation that causes the Nuclear Overhauser Effect (NOE). This effect is caused by dipolar relaxation without any contribution from the CSA. In contrast to longitudinal and transverse relaxation, cross

relaxation rates can be positive or negative because the cross relaxation rate is a difference of two terms:

$$\sigma_{NH} = \frac{1}{4} d_{NH}^2 [6j(\omega_H + \omega_N) - j(\omega_H - \omega_N)] \quad (3.18)$$

For the purposes of quantifying dynamics a steady state NOE experiment between a proton and heteronucleus is used. A steady state NOE depends on the ratio of the cross-relaxation rate constant and the self-relaxation rate constant R_1 . For an ^{15}N - ^1H spin system the enhancement is given by:

$$\eta_{NH} = \frac{\gamma_H \sigma_{NH}}{\gamma_N R_1^{(N)}} \quad (3.19)$$

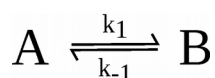
3.2.5 Transverse relaxation

Transverse relaxation is caused by two different types of magnetic field fluctuations. First, the local B_0 field can change over time which is somehow analogous to a chemical exchange event. The frequency of events that contribute most to the relaxation is $k_{\text{ex}} = 1/\tau_c$. This type of contribution to the transverse relaxation therefore largely depends on $j(0)$. The other linearly additive contribution is similar to R_1 which can also be also seen from the equation below:

$$R_2^{(NH)} = \frac{1}{8} d_{NH}^2 (4j(0) + j(\omega_H - \omega_N) + 3j(\omega_N) + 6j(\omega_H) + 6j(\omega_H + \omega_N)) + a_N^2 (4j(0) + 3J(\omega_N)) \quad (3.20)$$

3.2.6 Relaxation due to chemical exchange

A nucleus is experiencing chemical exchange when it is exposed to different chemical environments in a time dependent manner. The simplest case is when the nucleus exchanges between two states, A and B with rate constants k_1 and k_{-1}



NMR spectra are affected by chemical exchange processes and the effects of chemical exchange vary depending on the exchange rate $k_{\text{ex}} = k_1 + k_{-1}$. In the slow exchange regime when $k_{\text{ex}} \ll |\omega_A - \omega_B| = |\Delta\omega|$ two peaks separated by $\Delta\omega$ are visible in the NMR spectrum. Each has its characteristic chemical shift and the intensity of each peak is proportional to the population of each state. The transverse relaxation rate for each resonance is $R_2^{\text{obs}} = R_2^0 + k$ where R_2^0 is the relaxation rate in the absence of exchange and k is the rate constant (k_1 or k_{-1}) due to departure of spins from that state. In

the fast exchange regime when $k_{ex} \gg |\Delta\omega|$ one peak can be observed in the spectrum with a population weighted chemical shift $\omega_{obs} = P_A\omega_A + P_B\omega_B$ where P_A and P_B are the populations of states A and B. The relaxation rate in the fast exchange regime is a population weighted average between the relaxation rates of the two states ($R_2^{obs} = P_A R_{2A} + P_B R_{2B}$). In the intermediate exchange regime when $k_{ex} \approx |\Delta\omega|$ one signal is observed that is additionally broadened by R_{ex} . The relaxation rate in the intermediate exchange regime is therefore described as $R_2^{obs} = R_2^0 + R_{ex}$.

3.2.7 Carr-Purcell Meiboom-Gill Relaxation Dispersion

Line broadening caused by intermediate exchange is an obstacle when we want to obtain well-resolved spectra with narrow line widths, but can also be a source of information because it reflects the underlying dynamics. Carr-Purcell-Meibom-Gill (CPMG) relaxation dispersion is an approach that can be used to study dynamics in systems exchanging on an intermediate time scale ($k_{ex} = 100\text{--}3000\text{ s}^{-1}$) (Vallurupalli et al., 2012). The approach is based on the fact that when 180° refocusing pulses are applied at a frequency much higher than $1/k_{ex}$, the magnetization exchanging between state A and B can be completely refocused. Practically this approach is most commonly carried out in two-dimensional HSQC-type experiments where, during a fixed relaxation time T_{CPMG} , a different number of spin-echo pulses ($\tau\text{-}180^\circ\text{-}\tau$) are applied (Figure 3.2b-d). When τ is long and the CPMG frequency is low ($\nu_{CPMG} = 1/(4\tau)$) the observed transverse relaxation rates correspond to $R_2^0 + R_{ex}$, while at high CPMG frequencies when the magnetization is efficiently refocused only the R_2^0 term remains (Figure 3.2d). CPMG data are most commonly fitted to a two-site exchange model (A \leftrightarrow B) to obtain information about the chemical shifts in the B state, the kinetics and the thermodynamics of the exchange.

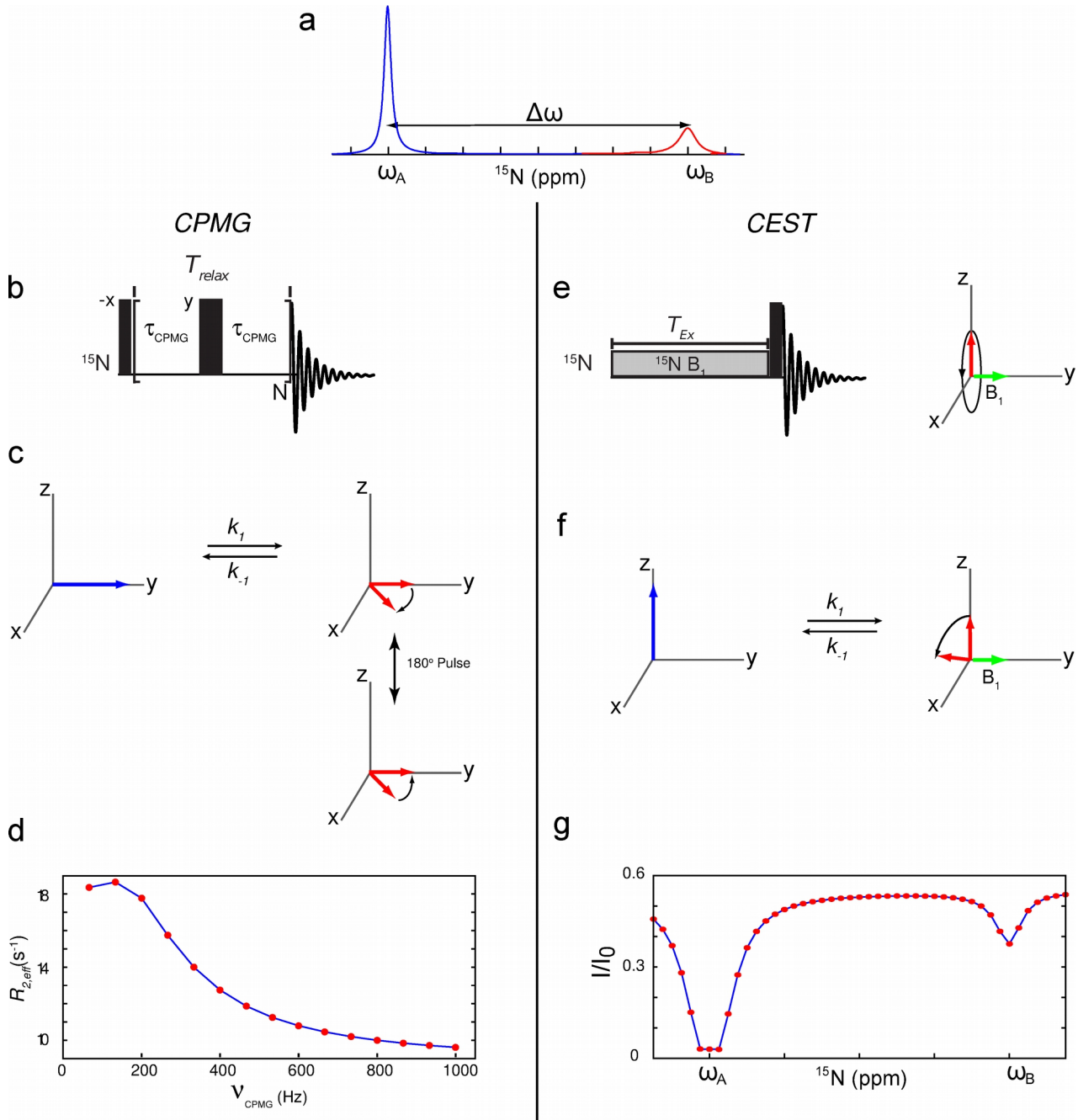


Figure 3.2: Following chemical exchange processes with CPMG and CEST experiments. a) A nucleus in exchange between two states with different chemical environments giving rise to two peaks separated by $\Delta\omega$. b) A schematic pulse sequence of the CPMG experiment. The narrow rectangle indicates a 90° pulse and the wide rectangle indicates a 180° pulse that is used to refocus the magnetization. The basic τ - 180° - τ building block is repeated N times with τ_{CPMG} adjusted to fit in the constant T_{relax} delay. c) Illustration of how the magnetization is refocused during a CPMG experiment. d) A schematic CPMG curve obtained by quantifying the peak intensities for the CPMG experiments at different CPMG frequencies. e) A schematic pulse sequence of a CEST experiment. f) Illustration of the evolution of magnetization when a nucleus in exchange between two states is affected by the weak B_1 saturating field. g) A schematic result of a CEST experiment. The result is plotted as a ratio I/I_0 where I is the intensity of the major peak after irradiation with the weak B_1 field and I_0 is the intensity of the major peak when the T_{EX} is set to zero. Adapted from (Vallurupalli et al., 2012).

3. Nuclear magnetic resonance

CPMG data can be fitted using the Carver-Richards equations that are a general solution to a two-site exchange system and valid for any exchange regime (Carver and Richards, 1972):

$$R_2^{obs} = \frac{1}{2} \left(R_{2A}^0 + R_{2B}^0 + k_{ex} - \frac{1}{2\tau} \cosh^{-1}(D_+ \cosh(\eta_+) - D_- \cos(\eta_-)) \right) \quad (3.21)$$

where τ is the half delay between the 180 degree pulses in the CPMG block (τ -180- τ). The D_{\pm} and η_{\pm} are given by

$$D_{\pm} = \frac{1}{2} \left(\pm 1 + \frac{\psi + 2\Delta\omega^2}{(\psi^2 + \zeta^2)^{1/2}} \right) \quad (3.22)$$

$$\eta_{\pm} = \left(\frac{2\tau}{\sqrt{2}} \right) \left(\pm\psi + (\psi^2 + \zeta^2)^{1/2} \right)^{1/2} \quad (3.23)$$

and ψ , ζ being

$$\psi = (R_{2A}^0 - R_{2B}^0 - P_A k_{ex} + P_B k_{ex})^2 - \Delta\omega^2 + 4P_A P_B k_{ex}^2 \quad (3.24)$$

$$\zeta = 2\Delta\omega (R_{2A}^0 - R_{2B}^0 - P_A k_{ex} + P_B k_{ex}) \quad (3.25)$$

The equations show which parameters can be obtained by fitting R_2^{obs} obtained with a relaxation dispersion experiment. The chemical shift difference between the A and B states can be obtained, but only as a magnitude ($|\Delta\omega| = |\omega_A - \omega_B|$), therefore the sign is usually obtained by other methods (Korzhnev et al 2005; Skrynnikov et al 2002). In the case where only state A is observed in the spectrum, knowing the chemical shift difference $\Delta\omega$ allows us to calculate the chemical shift of the invisible B state. The kinetic and thermodynamic information is provided by $k_{ex} = k_A + k_B$, P_A and $P_B = 1 - P_A$. R_{2A}^0 and R_{2B}^0 are the transverse relaxation rates of each state in the absence of exchange contributions. In many cases the difference between the relaxation rates can be considered negligible ($R_2^0 = R_{2A}^0 \approx R_{2B}^0$) where R_2^0 is obtained directly at high CPMG frequencies. Note that while the chemical shift difference $\Delta\omega$ is field dependent the populations P_A and P_B are not. Thus recording the experiment at two different static magnetic fields allows us, in many cases, to decorrelate the different contributions from k_{ex} , P_A and P_B and $\Delta\omega$ to the observed R_2^{obs} rates.

3.2.8 Chemical Exchange Saturation Transfer

Chemical Exchange Saturation Transfer (CEST) is a method that gives us information on a time scale close to CPMG relaxation dispersion although more applicable to slower exchange regimes (20–200/s) (Vallurupalli et al., 2012). This method is especially useful when the populations are skewed ($P_A \gg P_B$) where only the peak of the major state A is visible while the minor B state is not visible in the spectrum. To detect the minor state, a weak B_1 field is applied at the frequency $\omega_{(B_1)}$ of the invisible state to perturb the z-magnetization. The perturbation is then transferred to the major state by exchange making them detectable as changes in the intensities of the major state peaks. Practically, CEST would be most commonly performed in two-dimensional HSQC-type experiments with the weak B_1 field applied at different ^{15}N offsets sweeping through the full spectral width (Figure 3.2e). The result of a CEST experiment is then plotted as an intensity ratio I/I_0 against the B_1 offset, where I represents the irradiated spectrum and where I_0 is the reference spectrum with the evolution time T_{EX} equal to zero (Figure 3.2g). In a two-state exchange model ($A \leftrightarrow B$) the CEST experiment provides information about the chemical shift difference between A and B (including its sign), the exchange kinetics and the populations of state A and B (P_A and P_B). Additionally, it can also give us information about the difference in transverse relaxation rates between the A and B state, which the CPMG relaxation dispersion experiment is much less sensitive to.

In a fitting procedure the intensities observed in the spectrum can be simulated using the Bloch-McConnell equations for half-spin nuclei exchanging between states A and B:

$$\frac{d}{dt} \begin{bmatrix} E/2 \\ I_x^A \\ I_y^A \\ I_z^A \\ I_x^B \\ I_y^B \\ I_z^B \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -R_2^A - k_1 & -\omega_A & \omega_{(B_1)} & k_{-1} & 0 & 0 & 0 \\ 0 & \omega_A & -R_2^A - k_1 & 0 & 0 & 0 & k_{-1} & 0 \\ 2R_1^A I_{eq}^A & -\omega_{(B_1)} & 0 & -R_1^A - k_1 & 0 & 0 & 0 & k_{-1} \\ 0 & k_1 & 0 & 0 & -R_2^B - k_{-1} & -\omega_B & \omega_{(B_1)} & 0 \\ 0 & 0 & k_1 & 0 & \omega_B & -R_2^B - k_{-1} & 0 & 0 \\ 2R_1^B I_{eq}^B & 0 & 0 & k_1 & -\omega_{(B_1)} & 0 & -R_1^B - k_{-1} & 0 \end{bmatrix} \begin{bmatrix} E/2 \\ I_x^A \\ I_y^A \\ I_z^A \\ I_x^B \\ I_y^B \\ I_z^B \end{bmatrix}$$

where I_x , I_y and I_z are the magnetization along the x,y and z axes. Signal intensities can then be estimated by numerically integrating the Bloch-McConnell equations:

$$\frac{d}{dt} \hat{I}(t) = M \cdot \hat{I}(t) \quad \Rightarrow \quad \hat{I}(t) = \exp(Mt) \hat{I}(0) \quad (3.26)$$

where M is the exchange matrix of Bloch-McConnell equations. Longitudinal relaxation rates in the excited state are assumed to be equal to R_1^A ($R_1^A = R_1^B$). The variable parameters are k_{ex} , P_B , ω_B , R_2^A , R_2^B , R_1^A and I_0 . The simulated signal intensities obtained by numerically integrating the Bloch-McConnell equations are compared to the experimentally determined intensities and the variable parameters are adjusted until the best-fit model parameters are extracted by minimization of the target χ^2 function:

$$\chi^2(\xi) = \sum \left(\frac{I_i^{\text{exp}} - I_i^{\text{calc}}(\xi)}{\sigma_i^{\text{exp}}} \right)^2 \quad (3.27)$$

where the summation is over all the experimental data i , σ_i^{exp} is the estimated uncertainty of the measured peak intensity I_i^{exp} , $I_i^{\text{calc}}(\xi)$ is the calculated peak intensity, and $\xi = (x_1, \dots, x_n)$ is the set of variable parameters of the model.

3.3 Characterizing residual structure in IDPs by NMR

3.3.1 Chemical shifts

The chemical shift of a nucleus depends on the immediate chemical and electronic environment, on the medium and long-range interactions and as such provide a rich source of structural information. The many contributions to the chemical shift make this observable difficult to interpret and predict. Despite that, empirical relations between structure and chemical shifts have been successfully employed for characterization of proteins and chemical shift prediction from structure (Wishart, 2011).

In addition to exhibiting small conformational dependencies, chemical shifts depend strongly on nucleus type (primarily) and amino acid type. For interpretation of the chemical shifts in terms of secondary structure we therefore need a reference for each nucleus of the 20 amino acids. This reference is the random coil chemical shift (Table 3.1). There are two commonly used approaches for obtaining random coil chemical shifts. One way is to measure chemical shifts on short peptides representative of the unfolded state of proteins (Kjaergaard and Poulsen, 2011; Kjaergaard et al., 2011; Prestegard et al., 2013; Schwarzingner et al., 2000, 2001; Wishart et al., 1995b). This requires measurement of twenty model peptides such as for example Gly-Gly-Xxx-Gly-Gly, where Xxx each time represents one of the twenty amino acids. The other way uses databases of protein

structures with corresponding assignments (De Simone et al., 2009; Tamiola et al., 2010; Wang and Jardetzky, 2002a, 2002b; Zhang et al., 2003). It is assumed that the chemical shifts of amino acids in loop regions correspond to random coil chemical shifts because these amino acids are flexible and are potentially interchanging between different conformations. In table 3.1, I provide an overview of available random coil data sets.

Table 3.1: Overview of available random coil chemical shift datasets.¹

Origin of random coil values	Experimental conditions	Comments	Reference
Ac-G-G-X-A-G-G-NH ₂ Ac-G-G-X-P-G-G-NH ₂	pH 5.0, 1 M Urea, T=25°C	Nearest-neighbour effects are investigated, however, no systematic implementation of sequence-dependent correction factors, except for the residues followed by proline.	(Wishart et al., 1995b)
Ac-G-G-X-G-G-NH ₂	pH 2.3, 8 M Urea T=20°C	Suitable for comparisons to experimental chemical shifts in denatured proteins. Sequence-dependent correction factors are derived for all residue types.	(Schwarzinger et al., 2000, 2001)
Ac-G-G-X-G-G-NH ₂	pH 6.5, 1M Urea T=5, 15, 25, 35 and 45°C	Neutral pH matches the experimental conditions for most IDPs. Temperature coefficients of random coil chemical shifts and pH dependence of histidine shifts are reported.	(Kjaergaard et al., 2011)
Ac-Q-Q-X-Q-Q-NH ₂	pH 6.5, T=5°C	Neutral pH and non-denaturing conditions. Glutamine-based peptide allows derivation of more accurate sequence-dependent correction factors.	(Kjaergaard and Poulsen, 2011)
Ac-A-A-X-A-A-NH ₂	pH 2.5, 8M Urea T=25°C	Suitable for comparison to experimental chemical shifts in denatured proteins. Alanine-based peptide allows derivation of more accurate sequence-dependent correction factors	(Prestegard et al., 2013)
RefDB (BRMB <-> PDB)	2162 sets of PDB-CS pairs*	Amino acid specific chemical shifts are determined for the random coil state and fully populated α -helices and β -sheets.	(Zhang et al., 2003)
CamCoil	<pH>=6.5 <T>=21°C 1772 sets of PDB-CS pairs*	Random coil values are obtained from loop regions of proteins. Sequence-dependent correction factors are optimized against experimental chemical shifts from a set of denatured proteins.	(De Simone et al., 2009)
ncIDP	<pH>=6.5 <T>=14°C set of 14 IDPs	Sequence-dependent correction factors are obtained by singular value decomposition of a small IDP library.	(Tamiola et al., 2010)
PSSI and $\Delta(^X Y)_{n,s}$ $\Delta(^X Y)_{n,s}$ corrections tables	<pH>=6.0 <T>=29°C 36 sets of PDB-CS pairs*:	Sequence and structure-dependent corrections factors are determined through an iterative procedure on the entire experimental dataset.	(Wang and Jardetzky, 2002a, 2002b)

*PDB-CS pairs: PDB structures with corresponding chemical shifts

¹ Table adapted from (Kragelj et al., 2013)

The random coil chemical shifts can be subtracted from experimental chemical shifts to give secondary chemical shifts ($\Delta\delta = \delta_{\text{obs}} - \delta_{\text{rc}}$). In this way the amino acid type and nucleus identity

3. Nuclear magnetic resonance

dependent effects are taken into account such that the obtained secondary chemical shift informs us about non-covalent structural changes (e.g. $\phi/\psi/\chi$ angles, hydrogen-bonding) and potentially also about dynamics. Different chemical shifts depend differently on the ϕ/ψ angles that the amino acids adopt (Figure 3.2). For example, secondary $^{13}\text{C}^\alpha$ chemical shifts, are positive in α -helices and negative in β -strands while the secondary $^{13}\text{C}^\beta$ chemical shifts display an inverse dependence, they are negative in α -helices and positive in β -strands. Furthermore, not all nuclei are equally sensitive to all conformations. The $^{13}\text{C}^\alpha$, $^{13}\text{C}'$ and $^1\text{H}^\alpha$ are most sensitive to formation of α -helix, while β -strand conformations affect mostly $^1\text{H}^\alpha$ and $^{13}\text{C}^\beta$ nuclei (Wang and Jardetzky, 2002a) (Figure 3.3). The carbon chemical shifts are not sensitive to formation of PPII transient structure as the PPII specific chemical shifts for these nuclei are very close to the average chemical shifts over the whole Ramachandran space (Figure 3.3). Apart from that the PPII specific carbon chemical shifts and carbon chemical shift for the upper α -helical region (around $\psi = 0$) are also similar. The ^{15}N and $^1\text{H}^\text{N}$ chemical shifts display a different dependence on the ψ/ϕ angles. They mostly depend on the ψ/ϕ of the preceding residue and can in principle be used to detect an increase in PPII sampling (Figure 3.3).

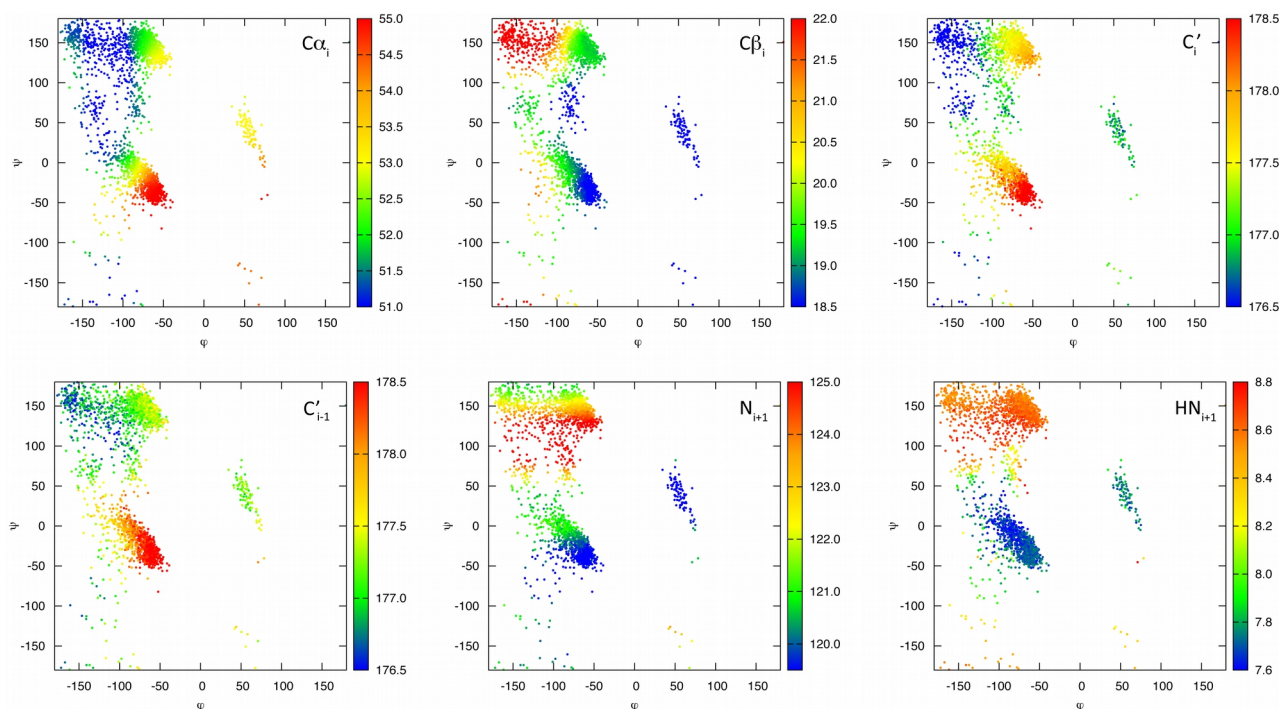


Figure 3.3: Dependence of primary experimental chemical shifts on backbone dihedral angle sampling. Distribution of predicted chemical shifts (in ppm) using SPARTA for the central residue $i = 8$ and its neighbour $i = 9$ of a poly-alanine 15-mer chain plotted against average $\{\phi, \psi\}$ of residue i . The chemical shifts were clustered into bins with a radius of 1° , averaged and plotted against the $\{\phi, \psi\}$ value of the central amino acid. Figure adapted from (Ozenne et al., 2012a).

These properties of chemical shifts are traditionally used to delineate secondary structure in assigned folded proteins. In the first algorithm that was designed to do that (the Chemical Shift Index algorithm (Wishart and Sykes, 1994; Wishart et al., 1992)), a consensus of secondary chemical shifts from several nuclei is used to digitally assign a residue either to an α -helix, β -strand or unstructured region when a certain threshold is reached. Other similar algorithms with their own specific characteristics were developed for folded proteins (Eghbalnia et al., 2005; Hung and Samudrala, 2003; Labudde et al., 2003; Mielke and Krishnan, 2009; Wang and Jardetzky, 2002a; Wang et al., 2007). In IDPs where secondary structure elements are formed only transiently, the same approach as it is used for folded proteins, cannot be applied. With this in mind a different approach was proposed called secondary structure propensity (SSP) (Marsh et al., 2006). Instead of providing a binary description of whether a residue is in an α -helix, β -sheet or in a flexible region the magnitude of the deviation from random coil is interpreted as a propensity to form an α -helix or β -extended structures. The propensity is obtained by normalizing the secondary chemical shifts with by the maximum difference between random coil shifts those corresponding to fully formed α -helix or β -sheet (Zhang et al., 2003).

3.3.2 Prediction of chemical shifts from structural coordinates

The secondary chemical shifts can be easily calculated once an assignment has been obtained and they provide a quantitative assessment of secondary structure content. However when a protein structure is being solved the secondary structure content is not an ideal constraint for structure calculation. Instead, torsion angles can describe the structure of proteins much better and can be used as restraints in structure calculations. For purposes of structure calculation of folded proteins chemical shifts can be converted to torsion angle restraints using for example the software TALOS (Cornilescu et al., 1999). The relationship can be used the other way around, chemical shifts can be calculated from a structural model and compared with the experimental target chemical shifts. This is an approach more suited to IDPs where chemical shifts are usually calculated and averaged over ensembles containing several conformations.

Going either from chemical shifts to torsion angle restraints or the other way around, the relationship between the chemical shifts and structure has to be understood. Chemical shifts depend on many factors: the already mentioned amino acid type dependence and neighbouring amino acid type influence, bond geometry from which the biggest contributions come from torsion angles ϕ , ψ ,

χ_1 and χ_2 , ring current effects, hydrogen-bonding, electric fields, solvent effects and flexibility (Wishart, 2011).

In the approach used by the SPARTA algorithm a database of tripeptides is prepared from proteins with known structure and assignments (Shen and Bax, 2007). For each overlapping tripeptide of these structures corresponding secondary chemical shifts are obtained by random coil subtraction and correction for ring-current and electric field effects.

After these initial steps of preparations the chemical shift prediction is performed as follows: for the structure in question, torsion angles are extracted for residues of each tripeptide of this query structure. The torsion angle information and residue identity are then used to search for tripeptides with similar torsion angles and amino acid types within the database. The secondary chemical shifts from twenty best matches are subsequently averaged and added to the amino acid type specific random coil value. Finally, this value is corrected for contributions from ring-current effects, hydrogen-bond geometry, electric field effects and backbone flexibility which are all calculated from the query structure.

The most accurate prediction can be made for $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ atoms. The $^{13}\text{C}'$ and ^{15}N are following in accuracy with predictions for $^1\text{H}^\text{N}$ being the most difficult. In recent years the field has constantly progressed, placing each newly released predictor in front of others in the competition for a higher accuracy (Wishart, 2011).

3.3.3 Residual dipolar couplings

In order to exploit dipolar couplings for structure determination in solution a small degree of preferential orientation of the protein in solution is required. This allows us to measure a small fraction of the total dipolar coupling. Practically, this is achieved by dissolving a protein in a medium which restrains the orientational sampling of the protein (e.g. sterically) so that not all orientations are equally likely to occur. We refer to a protein in such conditions as a protein that has been aligned. There are different media that allow us to align IDPs in the magnetic field. Most commonly used are the liquid crystal (Rückert and Otting, 2000), filamentous phages (Hansen et al., 1998), polyacrylamide gels (Sass et al., 2000), or bicelles (Tjandra and Bax, 1997).

The imposed preferential molecular orientation of an aligned protein can be described by an alignment tensor. RDCs can then be interpreted in terms of the orientation of internuclear vectors with respect to the common alignment tensor. The alignment tensor can be predicted from the three-

dimensional structure of a protein when the alignment is predominantly steric (Zweckstetter, 2008; Zweckstetter and Bax, 2000).

The original steric obstruction model for calculating the alignment tensor involves evaluation of steric clashes of the three-dimensional structure of the protein with a simplified representation of a molecule from the alignment medium. In each step of the algorithm the three-dimensional structure of the protein is rotated to sample different orientations and approached to the obstacle with a translational movement. The overall molecular alignment tensor is simply the linear average over all alignment matrices of clash-free orientations. With some modifications, the alignment tensor can also be predicted for charged alignment media, although these predictions tend to be less accurate (Zweckstetter et al., 2004).

For IDPs, which interchange between different conformations on timescales faster than the millisecond, the RDCs report on a time and ensemble average over all conformers. One way of analysing RDCs in IDPs is therefore to calculate the RDCs for each member of a given ensemble and subsequently average them over the entire ensemble. The calculated RDCs can then be compared to experimental data. This approach requires that RDCs can be accurately predicted for each member of the ensemble meaning that the prediction of the alignment tensor for each conformation should be reliable. For this reason steric alignment media are most commonly used for IDPs. RDCs have been used extensively for studying IDPs and have been used to study local flexibility (Huang et al., 2013; Salmon et al., 2010), cooperative transient structures (Jensen and Blackledge, 2008; Jensen et al., 2008, 2009), transient long-range interactions (Bernadó et al., 2005a) and in combination with ensemble description approaches they can be used to interrogate residue-specific conformational propensities of amino acids (Huang et al., 2013; Ozenne et al., 2012a).

How RDCs are influenced by the changes in conformational sampling can be studied with simulations using Flexible-Meccano (Bernadó et al., 2005b; Ozenne et al., 2012b). The Flexible-Meccano algorithm builds a large number of conformers of IDPs based on its primary sequence and a database of residue-specific ϕ/ψ angle distributions. A statistical coil model of a homopolymer (e.g. poly-alanine) generated with Flexible-Meccano exhibits featureless bell-shaped RDC curves (Figure 3.4). The $^1D_{NH}$ RDCs adopt negative values (Figure 3.4a) which can be explained as follows: the average orientation of the amide bond vector in an unfolded polypeptide chain where the protein is preferentially aligned in a direction parallel to the magnetic field (Figure 3.5a) would be expected to be approximately orthogonal to the field, and therefore negative (γ_N is negative). The

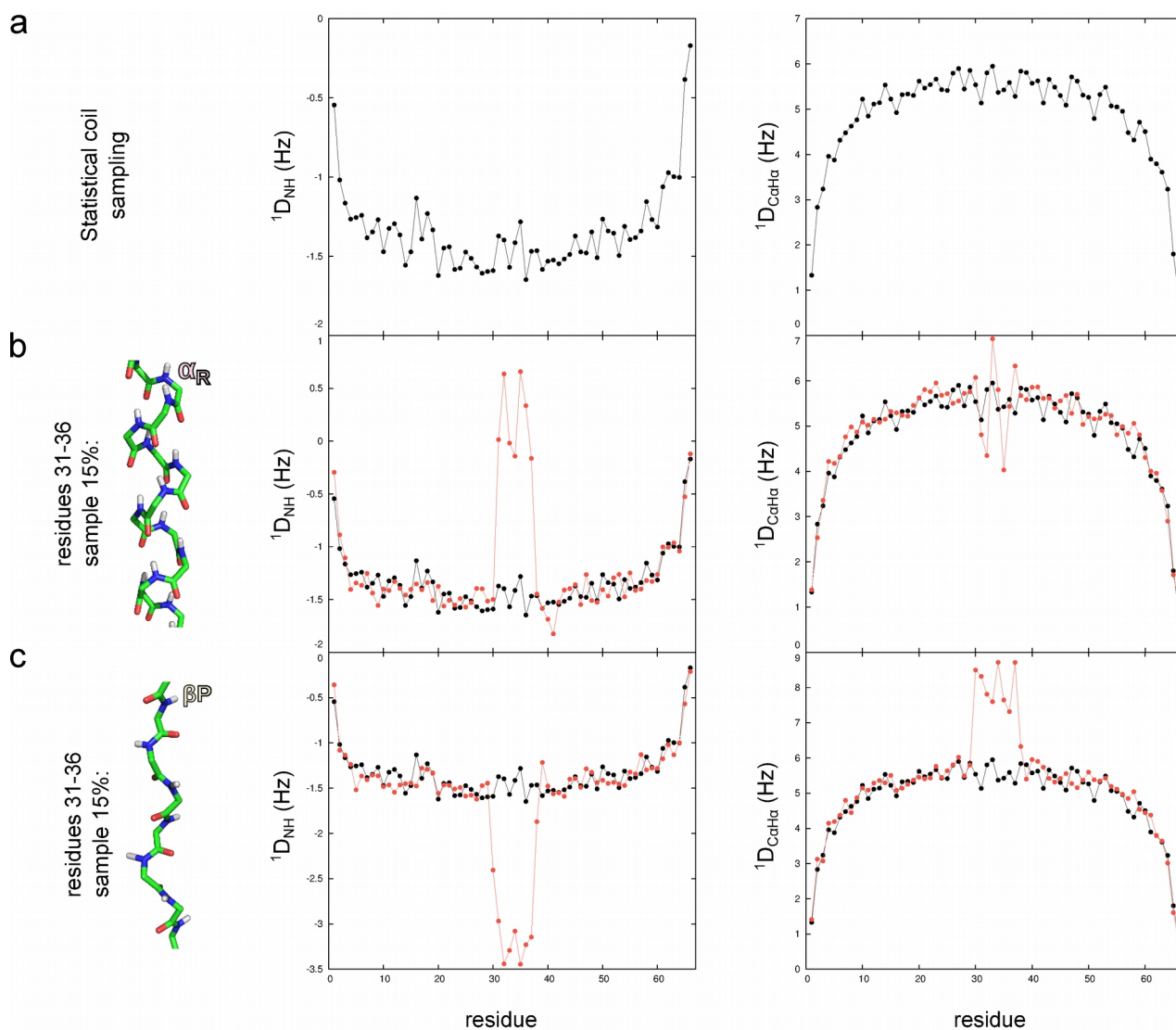


Figure 3.4: Effects of baseline and conformational sampling on RDCs. For each panel 100000 conformers were generated, alignment was calculated for each conformer and RDCs were averaged over all conformers. a) Simulated RDCs for a statistical coil display a clear baseline effect with the RDCs approaching zero at the ends of the chain. b) RDCs of a statistical coil (black) are compared to RDCs of a statistical coil in which a short region (residue 31 - 36) forms an α -helix in 15% of the conformers (red). c) RDCs of a statistical coil (black) are compared to RDCs of a statistical coil in which a short region (residue 31 - 36) forms a PPII helix in 15% of conformers (red).

bell-shaped RDC curve, that we call a baseline, reflects the flexibility of polypeptide chain which is lower in the middle of the chain and higher towards the ends. The higher flexibility at the ends of the polypeptide chain results in more comprehensive orientational sampling that is closer to isotropic orientational sampling and with that the RDC values approach zero. The described baseline shape can be observed for all types of RDCs because they depend on the chain-like nature

of IDPs and can be described relatively well with a hyperbolic cosine curve of the form (Nodet et al., 2009):

$$B(i) = 2b \cosh(a(i - d)) - c \quad (3.28)$$

where i is the residue number, d is half the number of residues and a , b and c are constants optimized for each different coupling type. For heteropolymers like IDPs the baseline is then further modulated by local conformational propensities of each residue (Figure 3.4)c.

If a stretch of residues in a chain forms an α -helix, the $^1D_{\text{NH}}$ RDCs become positive in that region (Figure 3.4c). This is because an NH bond vector in an α -helix would be aligned more or less parallel with the average chain direction and the B_0 field which would give rise to positive RDCs (Figure 3.5b). Similarly, extended β -strand and PPII structures make the NH vectors preferentially aligned orthogonal to the external magnetic field (Figure 3.5b) which causes more negative $^1D_{\text{NH}}$ RDCs in comparison to a statistical coil model. This behaviour makes the $^1D_{\text{NH}}$ RDCs able to distinguish between α -helical and extended structures.

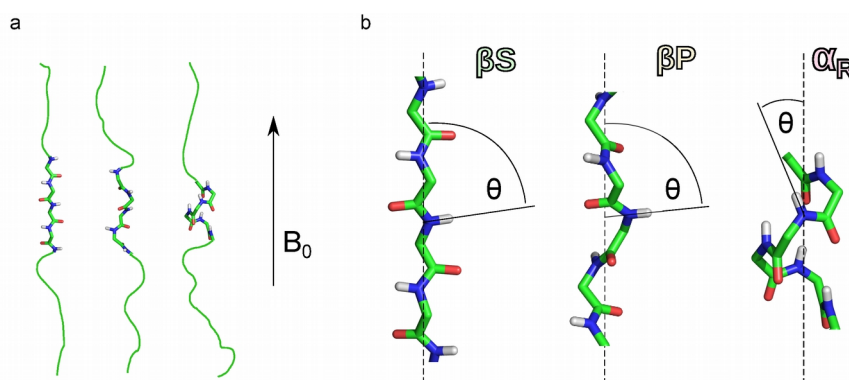


Figure 3.5: Illustration of effective angular averaging properties of ^{15}N - ^1H vectors in an unfolded protein. The direction of the weakly aligning medium is assumed to be parallel with the magnetic field B_0 . a) An extended chain will have a preferential alignment with the direction of the B_0 field. b) $^1D_{\text{NH}}$ RDCs measured for regions with extended β -strand (βS) or PPII (βP) transient structure will be more negative because the ^{15}N - ^1H vectors are orthogonal to the field ($\theta \approx 90^\circ$). In α -helices the ^{15}N - ^1H vectors are parallel with the field causing the $^1D_{\text{NH}}$ RDCs to be positive.

Another unique feature observed for regions with α -helical elements is a periodic modulation of $^1D_{\text{NH}}$ RDCs (Figure 3.6a). The origin of this periodicity within α -helical elements, which can be observed also for $^1D_{\text{C}\alpha\text{H}\alpha}$ RDCs (Figure 3.6) and to some extent for $^1D_{\text{C}\alpha\text{C}'}$ and $^2D_{\text{HNC}'}$ (not shown), is not immediately obvious. Under assumption that the α -helix is not deformed the dipolar waves can

3. Nuclear magnetic resonance

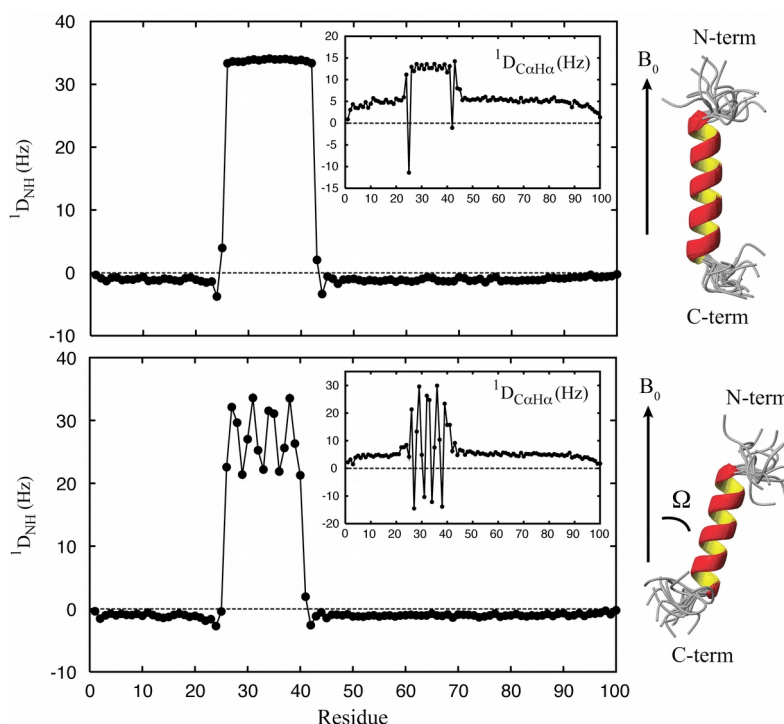


Figure 3.6: $^1D_{NH}$ and $^1D_{CaH\alpha}$ (insets) calculated for a 100 amino acid poly-Ala chain with helices at 26-41 (top) and 26-39 (bottom). RDCs were averaged over 50 000 conformers. In each case 15 superimposed conformers from Flexible-Meccano show the directionality of the unfolded chains projected from the helix caps. For clarity only the first four residues before and after the helix are shown. The effective tilt angle imposed by the helix capping is 1.3° (top) and 18° (bottom). Figure adapted from (Jensen and Blackledge, 2008).

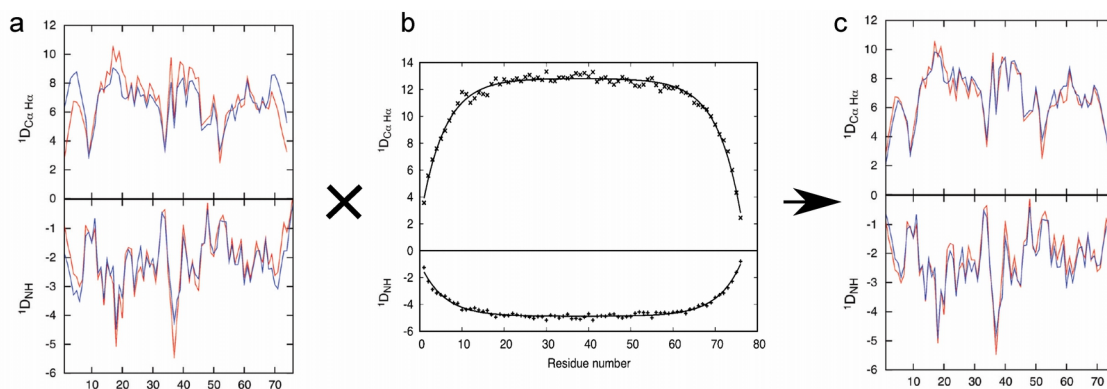


Figure 3.7: Multiplication of RDCs calculated using LAWs with RDC baselines in unfolded chains. a) $^1D_{NH}$ and $^1D_{CaH\alpha}$ RDCs calculated from the central amino acid of a 15 amino acid LAW (blue) contain no baseline information and therefore diverge from the RDCs calculated from an explicit ensemble using a global alignment tensor (red). b) Baseline effects underlying simulated ensemble-averaged RDCs from 100000 copies of a polyvaline chain of 76 amino acids in length (crosses) and predicted RDCs following a hyperbolic cosine curve of the form given in equation 3.28 (line). c) When multiplied with the baseline (b), RDCs calculated using LAW (blue) more closely resemble the RDCs calculated from the global alignment tensor (red). Figure adapted from (Nodet et al., 2009).

be explained if the α -helix exhibits a tilt of its main axis with respect to the direction of the B_0 field. In IDPs the tilt of the α -helix is determined by the directions into which the chain is projected from the termini of the α -helix. The α -helix length, which in turn determines the direction of the chains at the ends of the α -helix, has been shown to affect the amplitude and the phase of the dipolar waves in an analytical way (Jensen and Blackledge, 2008). The dipolar waves are therefore uniquely sensitive to helix termination and provides a tool for analysing experimental RDCs in terms of specific, cooperatively formed α -helices.

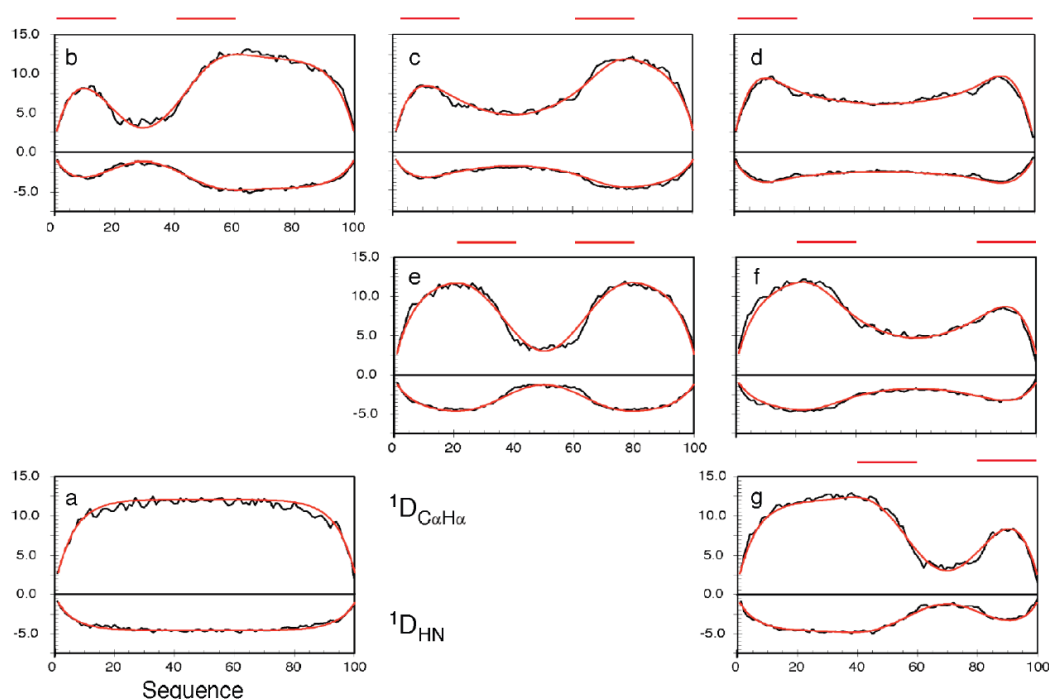


Figure 3.8: Simulation of RDC profiles for a homopolymer (polyvaline) in the presence of persistent long-range contacts between different sections of the chain. a) Profile of calculated RDCs in the absence of specific contacts. The program PALES was used to calculate RDCs from each conformer; 100 000 conformers were used in this average and the ones shown in panels (b-g). (b-g) Profiles of couplings in the presence of contacts between regions I and J: b) I) 1-20, J) 41-60; c) I) 1-20, J) 61-80; d) I) 1-20, J) 81-100; e) I) 21-40, J) 61-80; f) I) 21-40, J) 81-100; g) I) 41-60, J) 81-100. The two continuous bars above each plot indicate the positions of the contacting regions. The red curves were computed using a parametrization of a generic RDC baseline expression adapted for transiently contacting chains (Salmon et al., 2010). Figure adapted from (Salmon et al., 2010).

I have shown how RDCs depend on both local and long-range conformational degrees of freedom. This property is responsible for highly inefficient convergence statistics of ensemble-averaged RDCs calculated from statistical coil ensembles that need tens of thousands conformers before they

converge. This problem can be solved by dividing the protein into short overlapping 15 amino acid segments (local alignment windows, LAWs) and calculating the alignment tensor for each of these uncoupled segments (Marsh et al., 2008). The RDCs for such segments converge already with 200 conformers (Section 4.13) (Nodet et al., 2009). However, if only local alignment is considered all information about long-range order is lost. For proteins that do not display any long-range intrachain interactions, multiplication of the RDCs that were obtained by LAWs by the baseline profile recovers the underlying bell-shaped form of the RDC curve (Figure 3.7) (Nodet et al., 2009). Presence of long-range contacts can then further modulate the baseline. Figure 2.1 illustrates the influence of different long-range contacts on the underlying baseline calculated by numerical simulations. RDCs in the region between the two disordered segments experiencing long-range

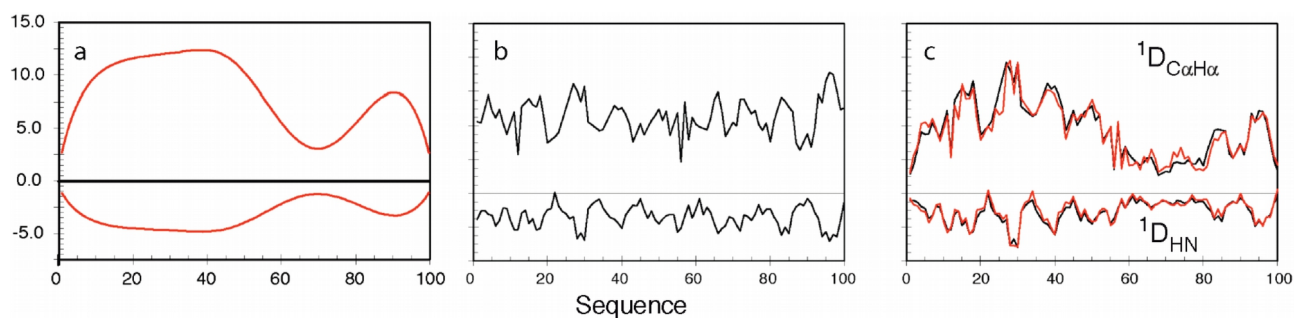


Figure 3.9: Example of the combination of baselines calculated using parametrized baseline adapted for transiently contacting chains and RDCs averaged using the local alignment window (LAW) approach. a) Baseline contribution calculated analytically for contacts between the regions centered on residues 50 and 90. b) RDCs calculated using LAW approach with windows of 15 amino acids in length; each RDC was averaged over 200 structures. c) Multiplication of the baseline from a) and the local RDCs from b) (red curves) compared to the RDCs averaged over 100000 full-length conformers in which each structure has a contact between 41-60 and 81-100 (black curves). Figure adapted from (Salmon et al., 2010).

contacts adopts values that are closer to zero. The multiplication of the modulated baseline with the RDCs calculated using LAWs can then be carried out in the same fashion as with the hyperbolic baseline (Figure 3.9). The simulations from the figure are based on ensembles for which a contact between the two regions is present for each member of the ensemble. In the presence of conformers both with and without the contact the baseline effects would be expected to be a population-weighted average between a baseline reflecting no contact (Figure 3.4a) and the baseline reflecting the specific contact (Figure 3.9a).

Ensemble-based simulations also demonstrated that a single RDC type cannot distinguish between all regions of the Ramachandran space (Figure 3.10) and that it is possible to raise some of these degeneracies when different types of RDCs are combined (Meier et al., 2007; Ozenne et al., 2012a).

Figure 3.10 also shows how varying the ϕ/ψ angles of residues $i-1$, i and $i+1$ affect the RDC of residue i . $^1D_{\text{NH}}$ exhibits the expected sensitivity for α -helical conformations and degeneracy between the extended β -strand and PPII, either for the amino acid of interest or for immediate neighbours. $^1D_{\text{CaH}\alpha}$ display a similar dependence on the sampling of residue i and on the sampling of its neighbours; the lowest values are seen in the α -helical region while the highest values are observed

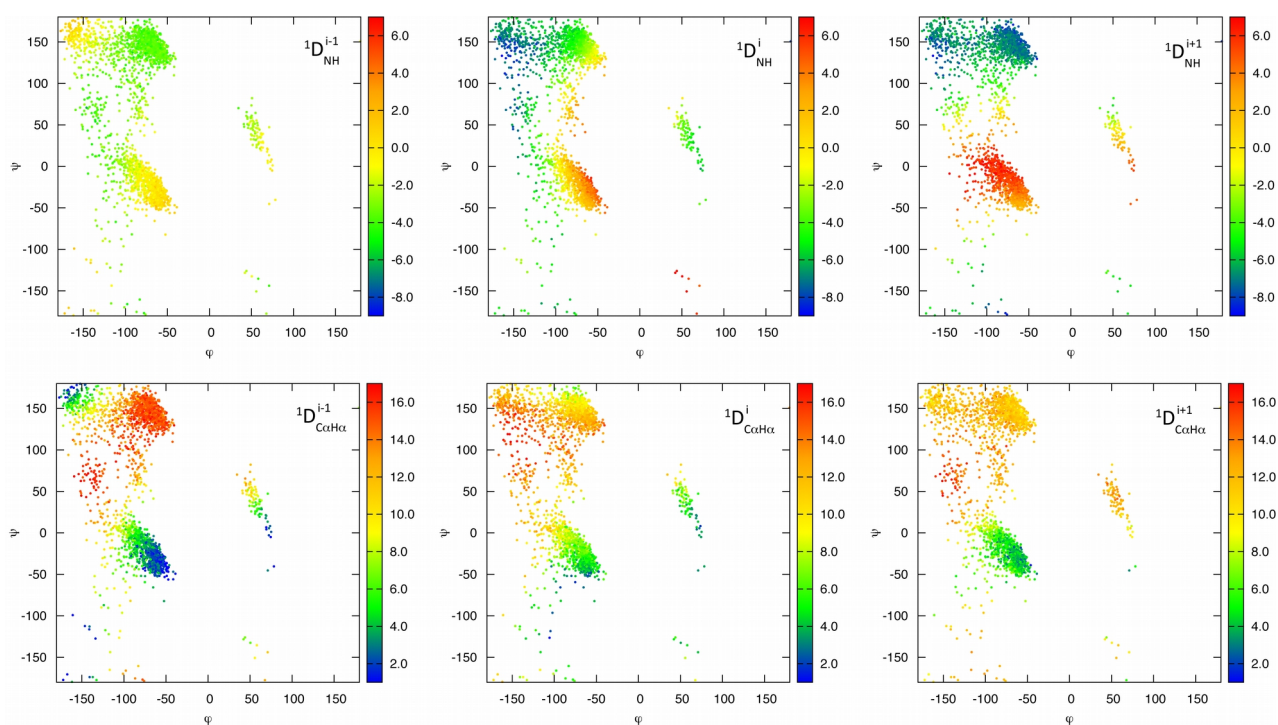


Figure 3.10: Dependence of RDCs on backbone dihedral angle sampling. Ensemble averaged backbone RDCs for the poly-alanine 15-mer chain plotted against average $\{\phi, \psi\}$ values of residue i . The RDCs were clustered into bins with a radius of 1° , averaged and plotted against the $\{\phi, \psi\}$ value of the central amino acid. Values are shown in hertz (Hz) in all cases, assuming an arbitrary level of overall alignment. Figure adapted from (Ozenne et al., 2012a).

for both extended β -strand and PPII regions. The dependence of all types of RDCs on the sampling of the residue of interest and also on the sampling of immediate neighbours complicates the interpretation in terms of local conformational sampling and clearly demonstrates why no single RDC can represent a "read-out" value for conformational propensity.

A recent study (Huang et al., 2013) has shown that the various contributions to an RDC can be deconvoluted into four factors: sampling of the amino acid i , the sampling of the immediate neighbours, sequence-dependent scaling factors taking into account six neighbours on either side (a polyglycine sequence will be more flexible and will exhibit lower degree of alignment) and length-dependent baseline that takes into account the chain-like nature of an IDP.

3. Nuclear magnetic resonance

4 Ensemble calculation for IDPs using NMR parameters

The following chapter is a manuscript for a chapter which will be published in a book entitled "Intrinsically disordered proteins studied by NMR spectroscopy (IDPs by NMR)". The book will be published as a part of a Springer's Advances in Experimental Medicine and Biology series. I am including the manuscript in the thesis because it covers the topic of ensemble selection procedures extensively. The chapter is written as a review and some of the examples are drawn from the previous studies on Tau protein, its fragments and on N_{TAIL} from Sendai virus, which were otherwise not the focus of my thesis.

In this chapter, I first mention how local structure in IDPs can be described by the dihedral angle distributions of amino acids and which NMR parameters can be used for characterizing local conformational propensities in IDPs. To some extent I have already presented this in the section about chemical shifts and RDCs (Section 3.3). Then I present the sample-and-select approach to ensemble calculation. Available conformational space has to be covered in the first sampling step. Obviously, the conformers that were not generated in the first step cannot be selected in the second step if they are part of the solution. IDPs are molecules characterised by a flat free-energy surface that spans many different conformations and covering the conformational space sufficiently represents a major challenge. The sample step is followed by a selection step in which the ensemble that is in agreement with the experimental data is selected. In this part I discuss under- and over-restraining and the ways to ensure validity of the derived structural ensembles. The reference ensemble method is presented with which a selection procedure can be validated and calibrated. The example of N_{TAIL} from Sendai virus nucleoprotein is used to illustrate how ensembles with cooperatively formed secondary structures can be calculated. Another important question related to ensemble calculations is the ensemble size. Here, I discuss how ensemble size has to be small enough to avoid overfitting and large enough to take into account the convergence properties of the NMR parameters used in the selections. The final section of the chapter is dedicated to validation of ensembles by cross-validation. In this approach we exclude some of the experimental data from the selection procedure. These data are used once the final ensemble has been calculated to assess whether the back-calculated data from the ensemble agree with excluded cross-validation data.

4.1 Abstract

Intrinsically disordered proteins (IDPs) perform their function despite their lack of well-defined tertiary structure. Residual structure has been observed in IDPs, commonly described as transient/dynamic or expressed in terms of fractional populations. In order to understand how the protein primary sequence dictates the dynamic and structural properties of IDPs and in general to understand how IDPs function, atomic-level descriptions are needed. Nuclear magnetic resonance spectroscopy provides information about local and long-range structure in IDPs at amino acid specific resolution and can be used in combination with ensemble descriptions to represent the dynamic nature of IDPs. In this chapter we describe sample-and-select approaches for ensemble modelling of local structural propensities in IDPs with specific emphasis on validation of these ensembles.

4.2 Introduction

Structural biology is an important branch of the life sciences. The number of protein structures deposited in the Protein Data Bank (PDB) is already exceeding 100000 and underlines the enormous effort that has been invested in solving ever-newer protein structures. The description of protein motion can be seen as the next logical step stemming from this wealth of structural data, strongly supported by the fact that proteins display functional dynamics occurring on a broad range of timescales (Bernadó and Blackledge, 2010; Henzler-Wildman and Kern, 2007; Kuriyan et al., 1986; Mittermaier and Kay, 2006). Nuclear magnetic resonance (NMR) spectroscopy is uniquely suited to probing protein dynamics at atomic resolution as a number of experimental parameters report on motions occurring on different time scales ranging from pico- to millisecond (Göbl et al., 2014; Mittermaier and Kay, 2006; Salmon et al., 2011).

Protein motion comes in many flavours and can span from local backbone and side chain dynamics in globular, folded proteins (Bouvignies et al., 2005; Guerry et al., 2013; Lange et al., 2008; Lindorff-Larsen et al., 2005; Salmon et al., 2009) through the concerted motion of entire domains in multi-domain proteins (Bertini et al., 2007; Deshmukh et al., 2013; Francis et al., 2011; Huang et al., 2014; Różycki et al., 2011; Yang et al., 2010) to intrinsically disordered proteins (IDPs), which represent the most extreme case of protein flexibility (Dunker et al., 2008; Dyson and Wright, 2002; Tompa, 2012). One way of representing the dynamics of a protein is to capture its characteristics – or more accurately, to explain the experimental NMR data, which depend on the underlying dynamics – with an ensemble of protein structures (Figure 4.1).

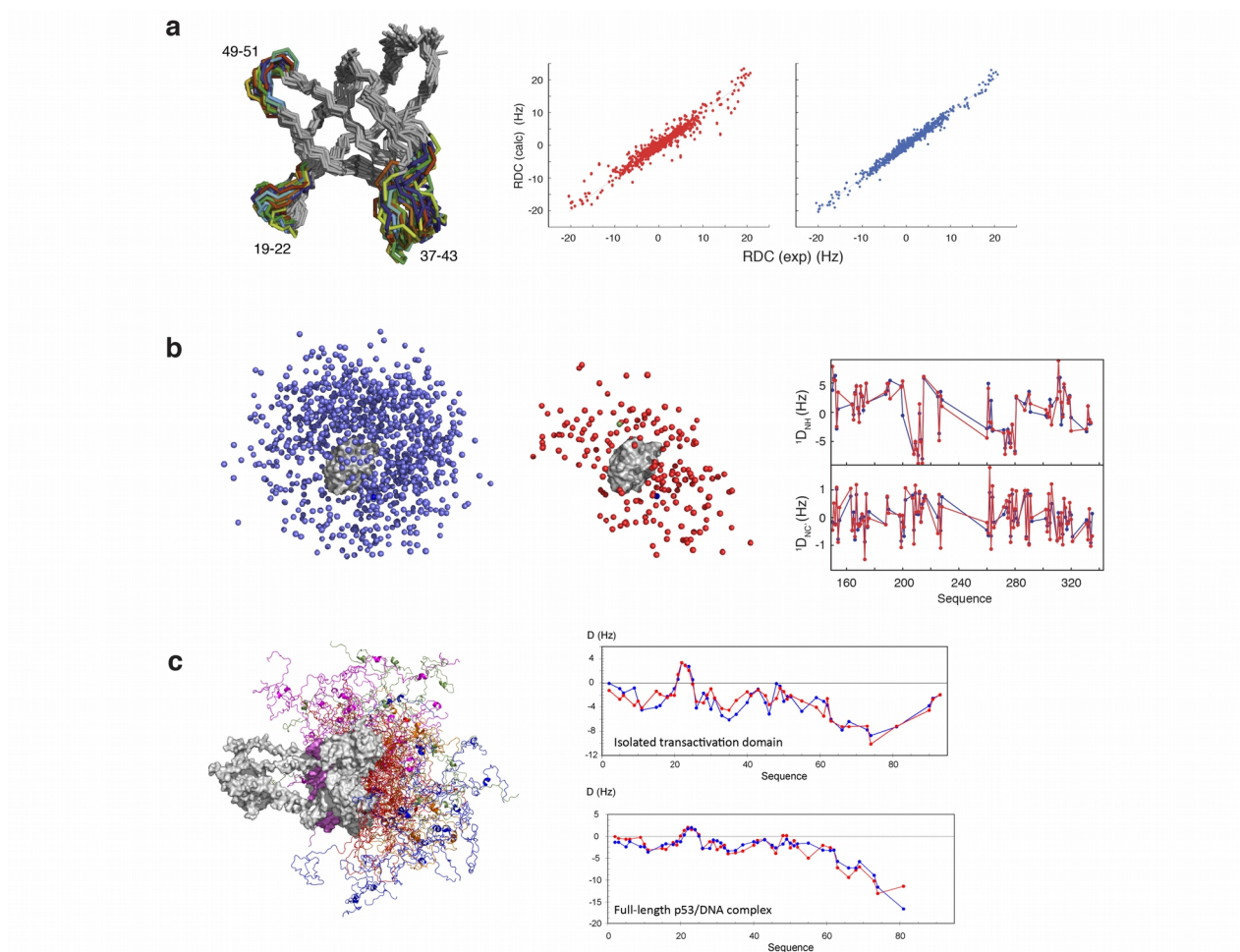


Figure 4.1: Interpreting NMR data with molecular ensembles to map conformational dynamics in proteins. (a) Dynamics of the SH3 domain from CD2AP derived from NMR residual dipolar couplings (RDCs) measured in multiple, complementary alignment media. An ensemble is shown of the SH3 domain derived from selection of conformational ensembles on the basis of experimental RDCs. The agreement between experimental and back-calculated RDCs is shown for the derived final ensemble (blue) and the starting pool of structures from which the ensemble was selected (red). Reprinted in part with permission from (Guerry et al., 2013). Copyright 2013 Wiley-VCH. (b) Dynamics of the two-domain splicing factor U2AF65 derived from RDCs and paramagnetic relaxation enhancements induced by S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonylthioate (MTSL) spin labels attached at different positions in the two-domain protein. Ensembles of the two-domain protein are shown, where the grey surface represents the location of the domain RRM1, while the location of the second domain RRM2 is shown as spheres positioned at the centre of mass of RRM2. Ensembles are shown representing the initial pool of structures sampling all conformational space (blue) and the space occupied by RRM2 after refinement against experimental data (red). The agreement between experimental RDCs (red) and those back-calculated from the derived ensemble (blue) is also shown. Reprinted in part with permission from (Huang et al., 2014). Copyright 2014 American Chemical Society. (c) Conformational ensemble of the intrinsically disordered N-terminal transactivation domain of p53 in the context of the full-length p53-DNA complex (Wells et al., 2008). The ensemble was obtained on the basis of experimental RDCs (local conformational sampling) and small angle X-ray scattering data (long-range behaviour). The agreement is shown between experimental RDCs (red) and those back-calculated from the model ensemble (blue) for both the isolated transactivation domain and in the context of the full-length p53/DNA complex. Reprinted in part with permission from (Wells et al., 2008). Copyright 2008 National Academy of Sciences, U.S.A.

In this chapter we will focus on atomic resolution ensemble descriptions of IDPs on the basis of experimental NMR data with a special emphasis on mapping local conformational propensities. The determination of a single set of three-dimensional atomic coordinates would have little meaning for these conformationally heterogeneous molecules, and ensemble descriptions are therefore necessary in order to build molecular models of IDPs that accurately capture the dynamic behaviour of the polypeptide chains. Special care has to be taken at each step of the ensemble generation protocol to ensure the validity of the obtained ensembles. The way in which ensemble generation protocols are tested and the factors that influence the modelling of the ensembles are therefore important questions that need to be addressed. In this chapter we will discuss these issues with a focus on the application of sample-and-select approaches to mapping local conformational propensities in IDPs.

4.3 Local structure in IDPs can be described by the dihedral angle distributions

It is expected that a single residue in an IDP will adopt many conformations over the time and ensemble average, and therefore undergoes exchange among many different dihedral angles. The distribution of dihedral angles sampled by a residue may at first seem like a simplistic representation of residual structure but is in reality very practical. The well-known secondary structures, the α -helix and β -sheet, are defined by hydrogen-bonding criteria and also have their own characteristic dihedral angles that are commonly used for annotating secondary structure elements in proteins (Figure 4.2a, 4.2c) (Frishman and Argos, 1995; Kabsch and Sander, 1983).

Other structural motifs can also be identified by their specific ϕ/ψ angles. Apart from the α -helix and β -sheet, poly-L-proline II (PPII) is the only secondary structure that forms linear groups of residues that all adopt the same conformation (Figure 4.2d) (Hollingsworth et al., 2009). This conformation is particularly interesting as it has been proposed to be significantly populated in IDPs and unfolded states of proteins (Schweitzer-Stenner, 2012; Shi et al., 2006). Residues within β -turns also adopt specific dihedral angles and differ for each β -turn type (I, II, I', II') (Figure 4.2e). Residues of both N-terminal and C-terminal helix capping motifs have unique dihedral angle distributions (Shen and Bax, 2012), and it has been shown that within α -helices of structured proteins, the central residues display different distributions than the C-terminal residues (Figure 4.2a, 4.2b) (Leader and Milner-White, 2011).

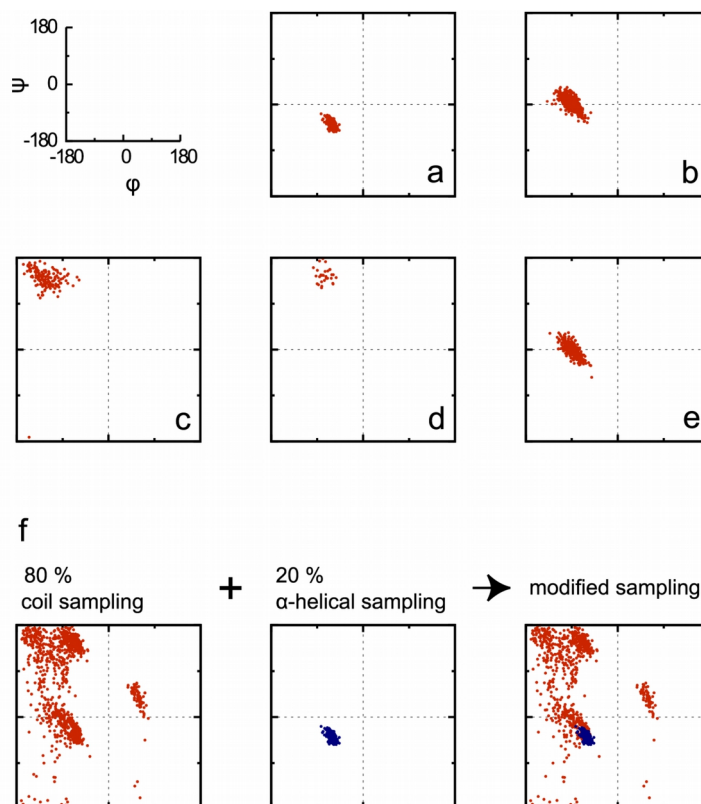


Figure 4.2: Dihedral angle distributions characteristic of different secondary structure types. (a) Central residues in α -helices. (b) Last residues in α -helices that are C-capped with Schellman loops. (c) Residues in β -strands. (d) Residues in PPII conformations. (e) Residues in type I β -turns. (f) Modification of the dihedral angle sampling of a given residue can be achieved by combining the random coil distribution with an over-sampling of other regions of Ramachandran space (in this case the α -helical region). Dihedral angles for (a), (b) and (e) were extracted from the database embedded in the Structure Motivator application (Leader and Milner-White, 2012). Dihedral angles in (c) were extracted from parallel and anti-parallel β -strands from structures with the following PDB codes: 1MLD, 1QCZ, 2CMD, 1XH3, 1OGT and 3GP6. Dihedral angles in (d) were extracted from non-proline residues of peptide ligands bound to SH3 domains in a PPII conformation (1BBZ, 1CKA, 1CKB, 1SSH, 1W70, 2DRK, 2DRM, 2O88, 2O9V, 2W0Z, 2W10, 3EG1 and 3I5R).

Since each structural motif has distinct dihedral angle distributions, we can use them to describe the conformational energy surface of each residue within the disordered protein chain and more importantly also as a metric for the presence of residual secondary structure in IDPs. An increase in sampling of dihedral angles corresponding to the α -helical region will, if sampled at a high enough propensity, give rise to transiently populated α -helices, even in the absence of cooperative effects. IDPs can therefore in general be described as random coils (i.e. a peptide chain without specific secondary or tertiary structure), with deviations from this model corresponding to the presence of residual secondary structure (Figure 4.2f). In order to map the dihedral angle distributions in an IDP we can exploit a number of different NMR parameters as described below.

4.4 NMR parameters for characterizing local conformational propensities in IDPs

NMR is a powerful technique for studying IDPs at atomic resolution and provides many experimental parameters that inform us about local conformational propensities (Jensen et al., 2014). Chemical shifts are the most readily accessible parameters and as a single NMR resonance is usually observed for each nucleus in the spectra of IDPs, the chemical shifts report on the population-weighted average over all conformations sampled in solution up to the millisecond time scale. Chemical shifts are sensitive to the backbone dihedral angle distributions and can, therefore, be interpreted in terms of local conformational propensities. A simple analysis of chemical shifts in IDPs involves the calculation of secondary structure propensities (Camilloni et al., 2012; Marsh et al., 2006; Tamiola and Mulder, 2012). This usually relies on characteristic shifts for α -helix, β -sheet and random coil derived from experimental chemical shifts of folded proteins with known three-dimensional structure or from a collection of assigned IDPs (De Simone et al., 2009; Tamiola et al., 2010; Zhang et al., 2003). When deriving conformational propensities it is important to correctly reference the experimental chemical shifts as systematic offsets may lead to erroneous estimates of the amount of secondary structure. It is possible to verify whether the chemical shift is correctly referenced using the secondary structure propensity (SSP) algorithm, which reports the potential reference offset based on the observation that C^α and C^β secondary chemical shifts are inversely correlated (Marsh et al., 2006).

Scalar couplings measured between nuclei of the protein backbone are also important structural probes in proteins and can be used to map dihedral angle distributions in IDPs. In the same way as chemical shifts, as long as the exchange rate is fast, the scalar couplings represent a population-weighted average over all conformations sampled in solution. The dependence of scalar couplings on the main chain torsion angles can be described using a so-called Karplus relationship (Karplus, 1959) that is generally parameterized against experimental scalar couplings measured in proteins of known structure (Smith et al., 1996). One of the commonly measured scalar couplings, the three-bond coupling constant $^3J_{\text{HNH}\alpha}$, depends on the backbone dihedral angle ϕ , allowing one to distinguish between α -helical ($^3J_{\text{HNH}\alpha} < 5$ Hz) and β -sheet conformations ($^3J_{\text{HNH}\alpha} > 8$ Hz) (Vuister and Bax, 1993). Other scalar couplings such as $^3J_{\text{C}\alpha\text{C}\alpha}$, $^3J_{\text{NH}\alpha}$, $^3J_{\text{NC}\alpha}$ and $^3J_{\text{NN}}$ report on the ϕ angle and in principle provide a more accurate measure of PPII conformations (Graf et al., 2007; Hagarman et al., 2010).

Residual dipolar couplings (RDCs) are obtained by partially aligning the protein molecules in the

magnetic field using, for example, a liquid crystal (Rückert and Otting, 2000), filamentous phages (Hansen et al., 1998), polyacrylamide gels (Sass et al., 2000), or bicelles (Tjandra and Bax, 1997). The inter-nuclear dipolar coupling, which is efficiently averaged to zero by the isotropic rotational tumbling of the molecules in solution, will no longer average to zero and a small part of the dipolar coupling will be measurable (Tjandra and Bax, 1997; Tolman et al., 1995). RDCs report on bond vector orientations with respect to a common reference frame and have been used extensively for structure determination of folded proteins as reporters on the relative orientations of secondary structure elements (Blackledge, 2005; Prestegard et al., 2004). Since the first measurement of RDCs in an unfolded protein (Shortle and Ackerman, 2001) we have significantly advanced in our understanding and interpretation of RDCs in IDPs (Jensen et al., 2009). It is now clear that the RDCs carry contributions from the dihedral angle distribution of the amino acid of interest as well as its nearest neighbours, and the measurement of a single RDC value does therefore not provide a direct “read-out” of residue specific sampling in the same way as chemical shifts and scalar couplings (Huang et al., 2013). In addition, a contribution from the local flexibility of the chain (bulkiness) to the RDCs should be taken into account together with a length-dependent baseline that reflects the polymeric nature of the unfolded chain (Huang et al., 2013; Salmon et al., 2010). In the case of IDPs there is a preference for alignment media that rely on steric interactions between the protein and the medium such that the alignment tensor, and thereby the RDCs, can be predicted directly from the shape of each protein conformation (Zweckstetter and Bax, 2000) and averaged over the ensemble.

4.5 Sample-and-select approaches

One way of obtaining representative ensemble descriptions of IDPs on the basis of experimental NMR data is to apply a two-step procedure involving the initial generation of a large pool of structures representing all of the conformational space available to the polypeptide chain (Figure 4.3). Experimental data are then included in the second step where a set of structures (an ensemble) that agrees with the data is selected, for example using a genetic algorithm.

Different approaches can be used to generate the initial pool of structures, but for a number of reasons it is important that the generated pool covers the entire conformational space of the molecule. Generally, a starting pool can be generated using molecular dynamics approaches or statistical coil generators.

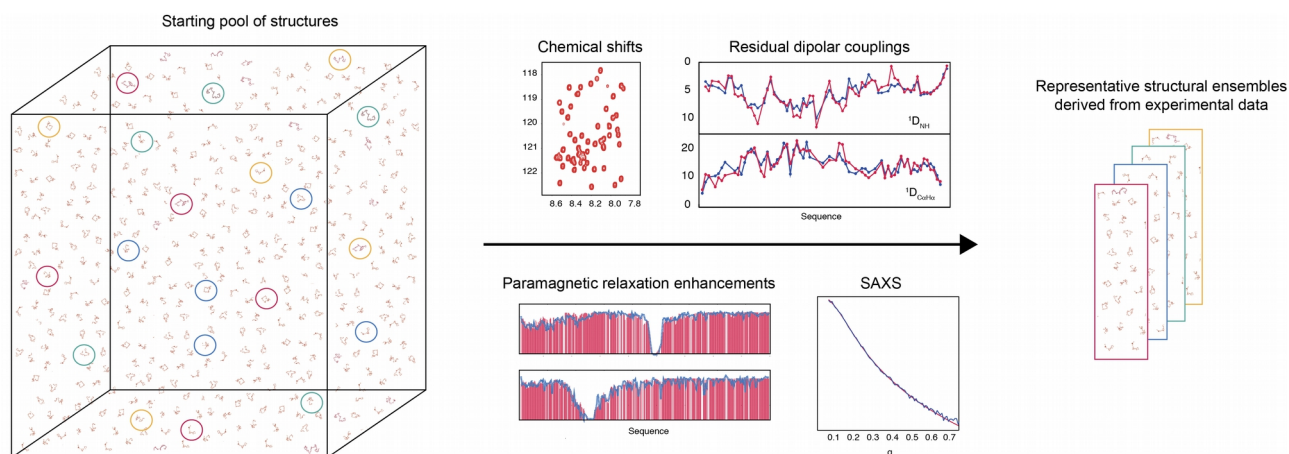


Figure 4.3: Overview of sample-and-select approaches. Initially a large pool of structures is generated that represents the entire conformational space available to the protein under investigation. Experimental data such as chemical shifts, residual dipolar couplings, paramagnetic relaxation enhancements and small angle X-ray scattering (SAXS) are exploited in a second step to refine this conformational space by selecting sub-ensembles that agree with the experimental data. Reprinted in part with permission from (Jensen et al., 2014). Copyright 2014 American Chemical Society.

4.6 Sampling space using molecular dynamics simulations

For classical molecular dynamics (MD) simulations, sufficient sampling remains a problem when studying IDPs, even when the simulations are run over long time-scales of several hundreds of microseconds (Lindorff-Larsen et al., 2012). Other types of MD simulations address this problem and provide a better sampling; one example is replica exchange molecular dynamics (REMD), which artificially enhances the sampling by exchanging copies of the simulated protein that evolve under different conditions (Hansmann, 1997; Sugita and Okamoto, 1999). In its simplest form the protein is exchanged between two different temperature reservoirs where at higher temperatures the sampling rate is faster but not physical. When the protein evolves at lower temperatures it can get trapped in a local or global minimum with the end result being an inefficient sampling of the conformational space. Exchanging the protein copy to a reservoir with higher temperature facilitates the sampling of other minima because the energy barriers between them are easier to overcome.

Other approaches include enhanced sampling techniques such as metadynamics, in which a term is added to the force field that penalizes the conformations that have already been explored by the molecule (Leone et al., 2010). The energy penalties accumulate as the protein explores an energy minimum and after some time the protein is forced to explore other minima. Metadynamics can be

combined with REMD to enhance the sampling rate even further (Piana and Laio, 2007). Accelerated molecular dynamics (AMD) is another approach for enhancing sampling that can be used for IDPs. In this method the free energy surface is modulated by a scaling factor that affects the energy barriers between minima and therefore increases the chance of barrier crossing (Hamelberg et al., 2004; Pierce et al., 2012; Voter, 1997).

4.7 Sampling space using statistical coil generators

A starting pool of conformers that is subsequently used in selections can also be produced with a statistical coil generator (Bernadó et al., 2005b; Feldman and Hogue, 2000; Jha et al., 2005a; Ozenne et al., 2012b). In this approach a protein molecule is built starting from either end of the chain by adding amino acid after amino acid with a ϕ/ψ angle that is randomly chosen from a database of dihedral angles (the statistical coil library). Each newly added amino acid is checked for steric clashes between the backbone atoms and between simplified representations of side chains. In case of steric clash the newly placed residue is rejected and rebuilt until a suitable conformation is found. Force field bond and angle potentials are not included during the generation of conformers, and the steric clash model is very simple and only defines a certain radius of exclusion for each atom. This approach allows the conformational space to be sampled roughly but efficiently and generates a pool of many different combinations of ϕ/ψ angles for consecutive amino acids.

Statistical coil libraries, which are used for generating the structures, are assembled with the help of databases of high-resolution crystal structures (Jha et al., 2005b; Serrano, 1995). The conformational preferences of amino acids in folded proteins differ from those of disordered proteins as most of the residues in folded proteins reside within secondary structure elements, while IDPs are expected to more closely resemble the loop regions. If the α -helices, β -sheets and β -turns are removed from the initial data set of high-resolution crystal structures, only motifs from non-regular loops remain in the database. These loop residues are not restrained by secondary structure hydrogen bonding criteria, unlike for example α -helices, and when a large number of the loop residues are taken into consideration, the potential contributions from the long-range tertiary contacts mostly average out. If we extract the ϕ/ψ angle distributions from the database of loop regions, we obtain a library of amino acid specific distributions of ϕ/ψ angles. These ϕ/ψ angle distributions represent a valid starting point for describing the conformational free energy surface of amino acids within IDPs and can be used in conjunction with statistical coil generators for building ensembles of IDPs. One of these statistical coil generators, Flexible-Meccano, is freely available²

4. Ensemble calculation for IDPs using NMR parameters

and is provided with a graphical interface that allows the testing of different sampling regimes by manually modifying the ϕ/ψ sampling of selected amino acids (Ozenne et al., 2012b). Flexible-Meccano calculates NMR observables such as chemical shifts, RDCs, scalar couplings and paramagnetic relaxation enhancements (PREs) from the generated ensembles that allow direct comparison with experimental data (Figure 4.4).

The statistical coil libraries still have room for improvement in terms of the inclusion of neighbour residue effects, which would be analogous to what has been carried out for random coil chemical shift tabulations (De Simone et al., 2009; Tamiola et al., 2010; Wang and Jardetzky, 2002a, 2002b; Wishart et al., 1995b). In fact the neighbour residue correction is often used in the statistical coil libraries for pre-proline residues, because the neighbour effect of prolines on the preceding residue

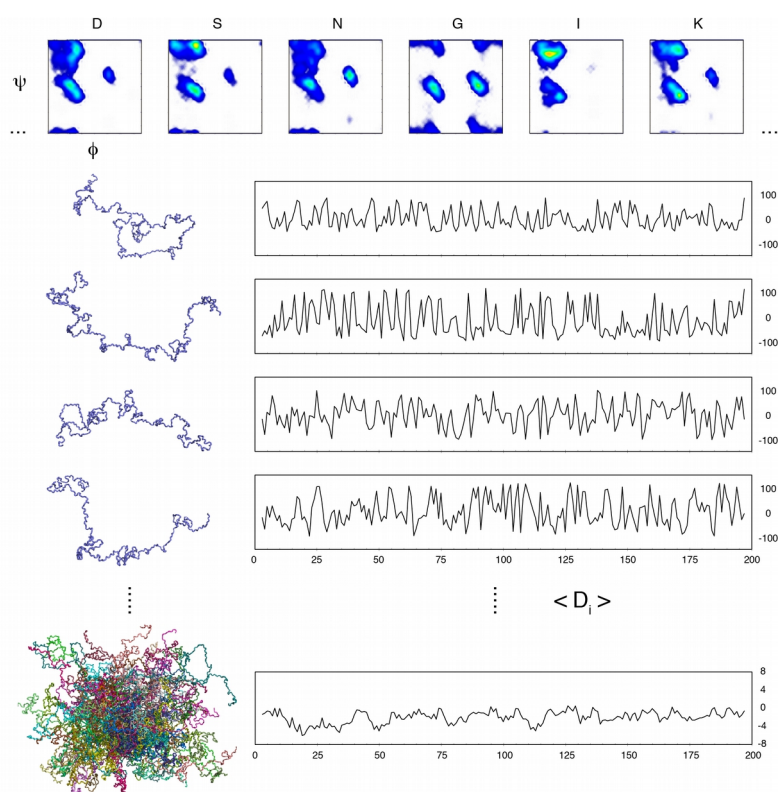


Figure 4.4: Generation of conformational ensembles of IDPs using a statistical coil generator. Ramachandran plots (ϕ/ψ distributions) are shown for the amino acids D, N, S, G, I and K as derived from loop regions of high-resolution crystal structures. These distributions are used to construct conformations of the protein for a given primary sequence by starting from either the C- or N-terminal end of the protein and building amino acid after amino acid according to randomly chosen ϕ/ψ pairs of the statistical coil library. For each copy of the molecule, experimental NMR data can be calculated and the ensemble-average over multiple copies of the protein can be compared to experimental data. Reprinted in part with permission from (Jensen et al., 2014). Copyright 2014 American Chemical Society.

is particularly pronounced. This is due to steric hindrance between the δ CH₂ side chain group of the proline and the H^N and β CH₂ atoms of the preceding residues (MacArthur and Thornton, 1991). Again, similarly to random coil chemical shifts, the statistical coil databases could be improved by refining them with the help of experimental data from IDPs themselves (Tamiola et al., 2010).

4.8 Selection of ensembles on the basis of experimental NMR data

Once the initial pool of structures has been generated, NMR parameters such as scalar couplings, chemical shifts and RDCs can be calculated for each member of the pool. The selection of sub-ensembles proceeds by calculating the averages of the NMR parameters over a given sub-ensemble and comparing them to experimental data. Different approaches have been proposed in the literature for deriving representative ensembles such as ENSEMBLE, which assigns weights to the different conformations of the pool (Krzeminski et al., 2013; Marsh and Forman-Kay, 2009), and ASTEROIDS, which relies on a genetic algorithm to select sub-ensembles (Nodet et al., 2009; Salmon et al., 2010), as well as ensemble optimization on the basis of Bayesian weighting (Fisher and Stultz, 2011; Fisher et al., 2010). It is important to note that in cases where the IDPs possess transiently populated secondary structures, it is not possible to select an ensemble that matches all the experimental data directly from a pool of statistical coil conformers. The reason for this is that the probability of finding continuous stretches of secondary structure is too low. Therefore, the sample-and-select protocol is often repeated multiple times in an iterative procedure, where the sampling pool is regenerated using the information (local conformational sampling) obtained from ensembles selected in the previous iteration. In this way, the sampling pool is enriched at each step with conformational preferences characteristic of the protein under investigation.

4.9 Ensemble representations of the IDP Tau from chemical shifts and RDCs

The combination of the statistical coil generator Flexible-Meccano (Bernadó et al., 2005b; Ozenne et al., 2012b) and the ensemble selection algorithm ASTEROIDS has allowed quantitative insight into residue-specific conformational sampling in a number of IDPs involved in neurodegenerative diseases (Bernadó et al., 2005a; Mukrasch et al., 2007; Schwalbe et al., 2014). The protein Tau is a 441 amino acid protein that is intrinsically disordered and undergoes a conformational transition to a pathological form of the same protein. The NMR spectra of Tau have been fully assigned (Narayanan et al., 2012), allowing insight into the conformational preferences of this protein at atomic resolution. A complete set of chemical shifts and $^1D_{NH}$ RDCs were obtained for the protein Tau in order to accurately map α -helical, β -strand and PPII populations. Figure 4.5a shows the agreement between experimental data and those back-calculated from selected ASTEROIDS ensembles. The ensemble selections were repeated five times and the conformational sampling of each residue along the sequence of Tau is conveniently represented by their dihedral angle distributions (Figure 4.5b).

In general, we can learn a lot from these ensembles as they provide quantitative insight into the sampling in different regions of Ramachandran space. Specifically, it is seen that the aggregation nucleation sites in Tau overpopulate the PPII region, suggesting that these conformations represent precursors of aggregation (Figure 4.5b) (Schwalbe et al., 2014). In addition to these observations, the presence of turn-like motifs can be identified in each of the Tau repeat regions (R1-R4). These turn motifs were also studied in detail previously using AMD simulations of small peptides of Tau, where it was shown that the AMD derived ϕ/ψ sampling corresponded to type I β -turns (Mukrasch et al., 2007). When this AMD sampling was incorporated into a model ensemble of the smaller K18 construct of Tau, the agreement between experimental and back-calculated $^1D_{NH}$ RDCs improved significantly, proving that these regions indeed adopt type I β -turns as predicted by AMD.

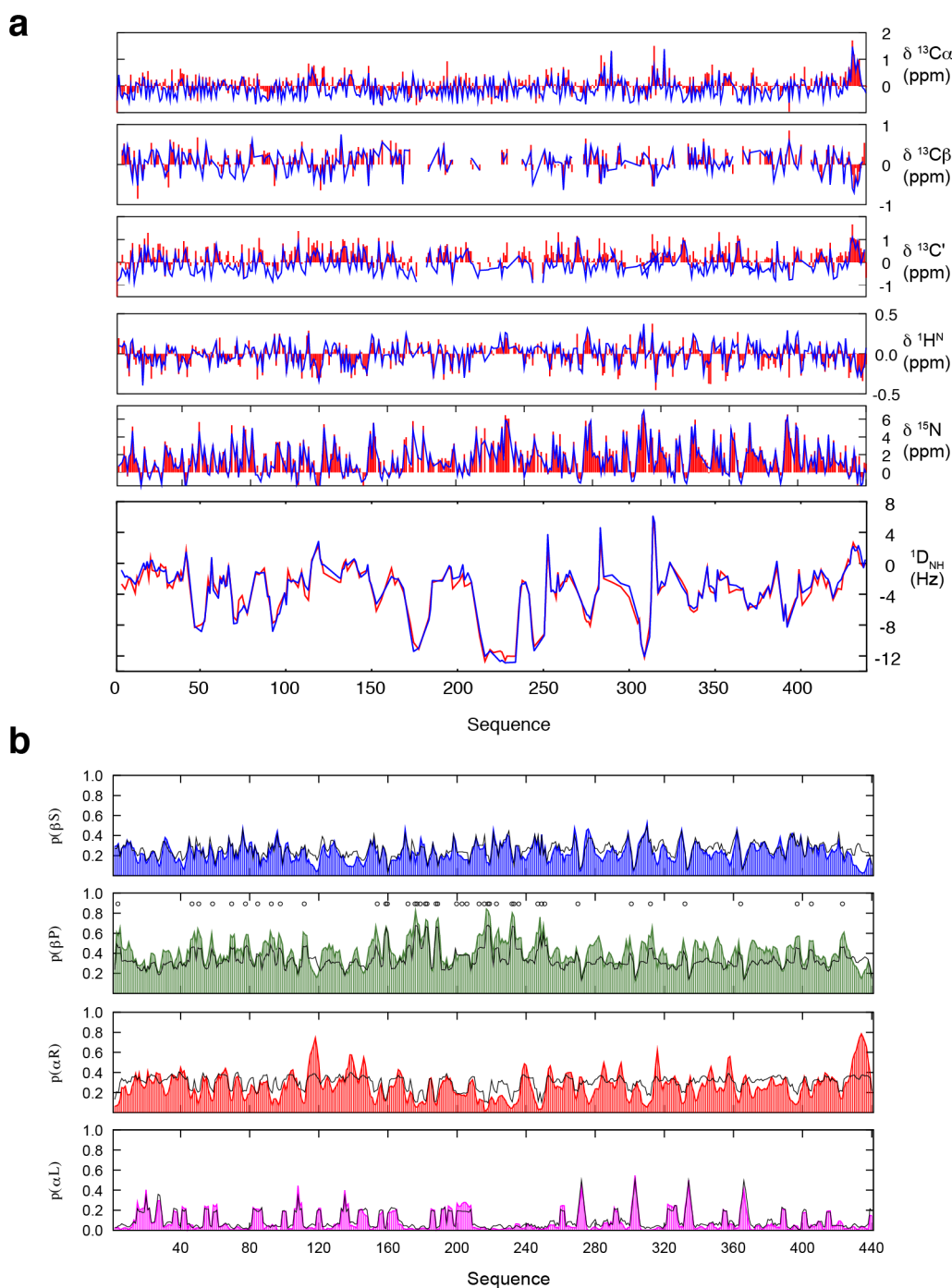


Figure 4.5: Ensemble representations of the intrinsically disordered Tau protein on the basis of chemical shifts and RDCs. (a) Agreement between experimental (red) and back-calculated secondary chemical shifts and RDCs (blue) from selected ASTEROIDS ensembles of Tau. (b) Site specific conformational sampling in Tau derived from the selected ensembles (blue, green, red, magenta) compared to standard statistical coil distributions (black). Populations are reported for four different regions of Ramachandran space corresponding to right- (αR , red) and left-handed α -helix (αL , magenta), β -strand (βS , blue) and PPII conformations (βP , green). Circles indicate the presence of proline residues. Reprinted in part with permission from (Schwalbe et al., 2014). Copyright 2014 Elsevier.

4.10 The reference ensemble method

The study described above combines different data types to map local conformational sampling. The accuracy with which conformational propensities can be determined depends on the amount of experimental data available for a given system. Assuming that we want to map the population of α -helix, β -strand and PPII conformations for each amino acid of the protein, it would be useful to determine a minimum dataset that would allow this. The α -helical and β -strand propensities can be well characterized with the help of carbon (C^α , C^β , C') chemical shifts, but a residue sampling a statistical coil distribution and a residue sampling exclusively PPII specific dihedral angles have approximately the same carbon chemical shift (Figure 4.6a). We therefore cannot use carbon chemical shifts to distinguish between the two mentioned sampling regimes. Similarly, most of the RDC types display degeneracy between β -strand and PPII conformations. The $^1D_{NH}$ RDCs are negative for both increased β -strand propensities and increased PPII propensities (Figure 4.6b).

Selection against synthetic data from a reference ensemble can help reveal such degeneracies and determine the minimum dataset necessary for accurate mapping of the conformational energy landscape. In the reference ensemble approach, an ensemble of structures is generated using either an MD simulation or a statistical coil generator. These structures constitute the target ensemble for which a synthetic dataset is calculated. If our ensemble selection protocol is working without bias and we have sufficient and complementary data types, we should be able to regenerate the local conformational sampling preferences by targeting the synthetic dataset using the sample-and-select approach. A study by Ozenne et al. demonstrated how useful this approach can be when applied to IDPs (Ozenne et al., 2012a). Initially, an ensemble of a model protein of 60 amino acids of arbitrary sequence was obtained using the statistical coil generator Flexible-Meccano, where three distinct regions of the protein over-sampled the α -helical, β -strand and PPII region of Ramachandran space (50% additional sampling in each region compared to the statistical coil). Different types of ensemble-averaged chemical shifts and RDCs were calculated for this ensemble and used as targets in a selection protocol using the genetic algorithm ASTEROIDS by starting from a statistical coil pool, i.e. a pool without any particular secondary structure preferences. After the ensemble selection using only data for carbon chemical shifts (C^α , C^β , C'), the ϕ/ψ sampling was reproduced in the regions with enhanced α -helical and β -strand sampling, but not in the region with enhanced PPII chemical shifts or $^1D_{NH}$ RDCs in the selection procedure allowed a reproduction of the PPII sampling in the third biased region of the target ensemble, while inclusion of both backbone chemical shifts and RDCs represents a robust and accurate way to map the local conformational

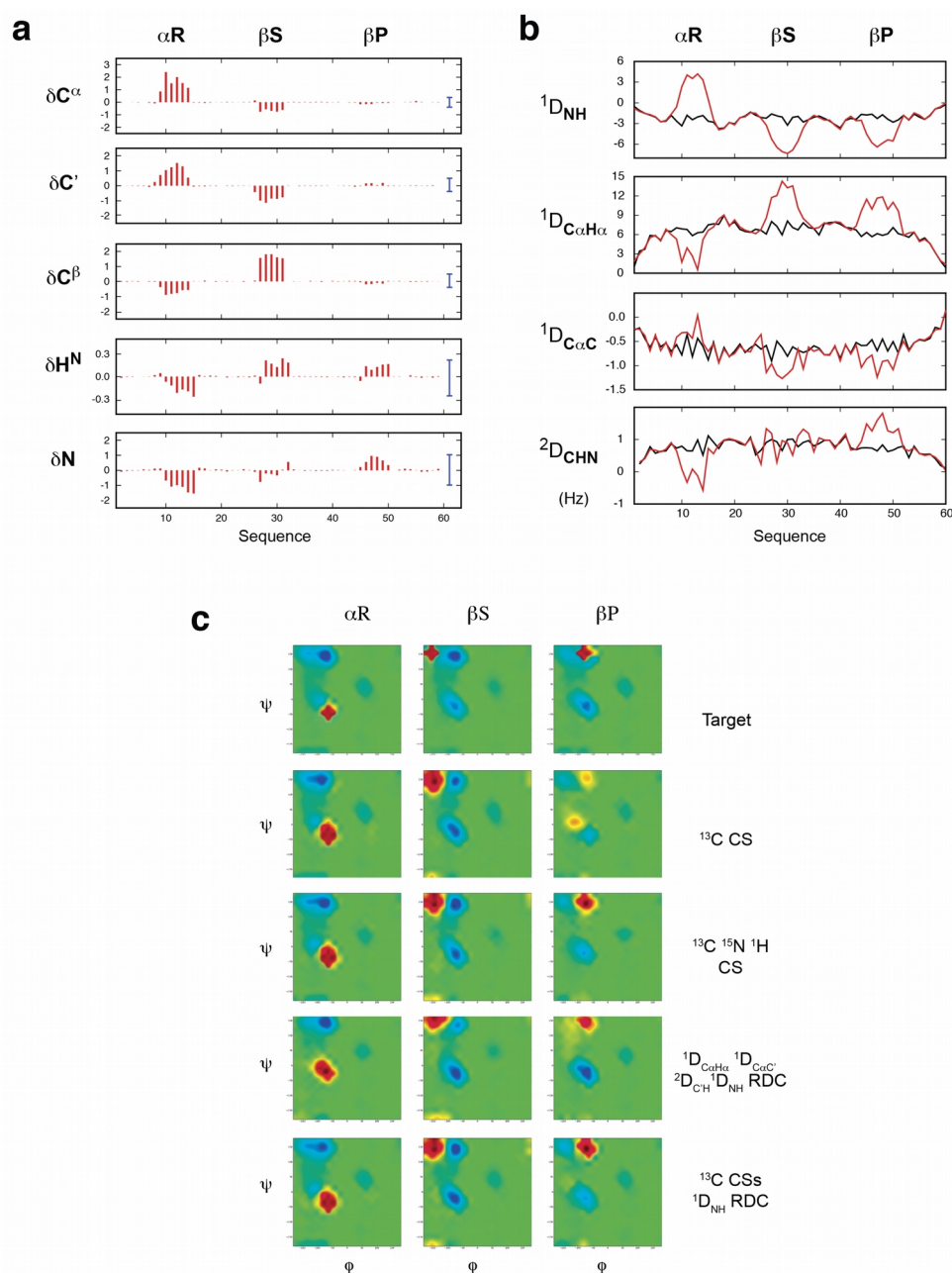


Figure 4.6: Testing the accuracy with which experimental data can map local conformational propensities in IDPs using the reference ensemble method. (a) Synthetic chemical shift dataset calculated for an ensemble of a model protein of 60 amino acids of arbitrary sequence. Three different regions of the protein over-sample the α -helical, β -strand and PPII region, while the remaining regions sample statistical coil conformations. The difference is shown between the predicted chemical shifts for this ensemble and the ensemble-averaged chemical shifts for a statistical coil ensemble. (b) Synthetic RDC dataset calculated for the model protein over-sampling the three different regions of Ramachandran space (red) compared to RDCs predicted for a statistical coil ensemble of the same protein (black). (c) Selection of sub-ensembles using ASTEROIDS on the basis of different combinations of the synthetic chemical shift and RDC datasets. Ramachandran plots of the target (top line) and the results of the selections employing different data types are shown. Reprinted in part with permission from (Ozenne et al., 2012a). Copyright 2012 American Chemical Society.

sampling of IDPs (Figure 4.6c). Calibration of ensemble generation protocols against a synthetic target can therefore tell us if we are able to reproduce the sampling of a synthetic ensemble, and consequently also a real ensemble with the same characteristics. The reference ensemble method can also be used in a quantitative way by adding Gaussian noise to the synthetic dataset to determine the accuracy with which the conformational space of IDPs can be mapped using different data types (Ozenne et al., 2012a).

4.11 Taking into account cooperatively formed secondary structures in IDPs

When an IDP contains a longer stretch of a cooperatively formed structure, such as an α -helix, the sample-and-select approach does not work as efficiently. An α -helix can be stabilized by many cooperative interactions (Doig, 2002; Muñoz and Serrano, 1995). For example, the effect of helix capping can span several amino acids further down the protein sequence and can affect the stability of the helix as a whole. Apart from capping interactions and the regular backbone-backbone i to $i + 3$ hydrogen bonding pattern, many other stabilizing interactions are present between i and $i + 3$ residues and between residues even further away. As helices in IDPs can span more than ten residues, we expect that amino acids that are far apart in the primary sequence should contribute together to the formation of the helix.

Statistical coil generators take into account amino acid type conformational preferences that are mainly local. As a consequence the sampling in the statistical coil library can correctly sample α -helical conformations in selected regions of the protein; however, the chance of building a long helix without an interruption is relatively small. For example, with a statistical coil library with 80% helical sampling, the probability of forming a helical element consisting of six consecutive amino acids is 0.86, which is around 25%. The probability of forming a longer α -helix with a high enough population to fit the data is therefore very low. Approaches using MD simulations experience a similar problem when it comes to long cooperatively folded helices (or other secondary structures), and breaks in helices are often observed throughout the simulations.

A solution to this problem is to generate many different starting ensembles where each ensemble incorporates an α -helix with a different start and end point, calculate the ensemble-averaged NMR data for each of these ensembles, and subsequently find the best combination of ensembles with corresponding populations that agree with the experimental data. Essentially this corresponds to

enriching the initial starting pool with cooperatively formed α -helices in specific regions of the protein that are known to over-sample the α -helical region of Ramachandran space.

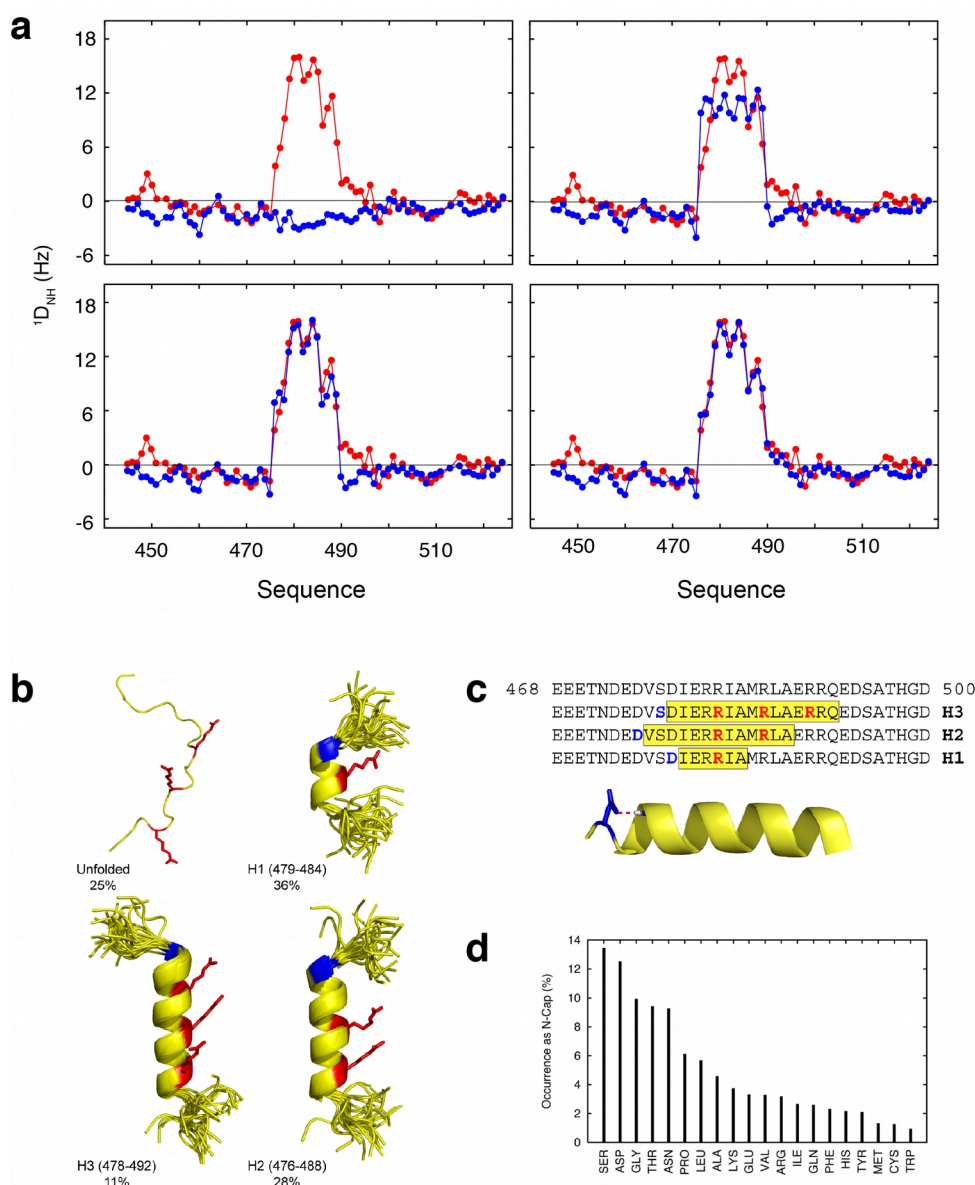


Figure 4.7: Analysis of cooperatively formed α -helices within the molecular recognition element of the C-terminal domain, N_{TAIL} , of the Sendai virus nucleoprotein. (a) Reproduction of experimental $1D_{NH}$ RDCs in N_{TAIL} for models with an increasing number, N , of helical ensembles: $N=0$ (top, left), $N=1$ (top, right), $N=2$ (bottom, left), $N=3$ (bottom, right). Experimental RDCs are shown in red, while back-calculated RDCs from the different models are shown in blue. (b) Molecular representation of the equilibrium of the molecular recognition element of N_{TAIL} in solution. The four different helical states are presented as a single structure for the completely disordered form and as twenty randomly selected conformers for the three helical states. The molecular recognition arginines are displayed in red, while N-capping residues are shown in blue. (c) The amino acid sequence of the molecular recognition element of N_{TAIL} showing that the selected helical elements are all preceded by aspartic acids or serine residues. The cartoon representation illustrates an N-capping aspartic acid side chain-backbone interaction. (d) The occurrence of different amino acid types as N-capping residues in helices of folded proteins. Reprinted in part with permission from (Jensen et al., 2008). Copyright 2008 American Chemical Society.

This approach was developed and applied to the C-terminal intrinsically disordered domain, N_{TAIL} , of Sendai virus nucleoprotein, which undergoes induced α -helical folding of its molecular recognition element upon binding to its partner protein PX (Jensen et al., 2008). The $^1D_{\text{NH}}$ RDCs measured in N_{TAIL} were positive within the molecular recognition element and showed a characteristic dipolar wave pattern consistent with the formation of cooperatively formed α -helices (Figure 4.7a) (Jensen and Blackledge, 2008). The experimental RDCs were fitted with models of increasing complexity, i.e. starting from a statistical coil model and increasing the number of helical ensembles until a satisfactory fit was obtained (Figure 4.7a). For each model the populations of the helical elements were optimized to best agree with the experimental data. Data reproduction evidently improves as the number of helical ensembles increases, and a standard F-test was therefore used to test for the statistical significance of this improvement. It was found that three helical ensembles with different populations in exchange with a disordered form of the protein are needed to describe the experimental RDCs (Figure 4.7b). Interestingly, all the selected helical ensembles are preceded by aspartic acids or serines, which are the most common N-capping residues in helices of folded proteins (Figure 4.7c, 4.7d). An N-capping residue stabilizes a helix by forming a hydrogen bond between its side chain and the backbone amides at position 2 or 3 in the helix (Figure 4.7c). Importantly, this indicates that the helices preferentially being populated in solution in N_{TAIL} are stabilized by N-capping interactions, and that the helical formation is being promoted by strategically placed aspartic acids and serines in the primary sequence. The partial pre-structuration of N_{TAIL} in its free state suggests that the interaction with PX occurs through conformational selection, where one of the helices is selected by the partner protein in order to form the complex (Hammes et al., 2009).

4.12 Choosing an appropriate ensemble size

A scoring function that measures the agreement between the experimental and simulated data for the model ensemble is applied during the ensemble selection procedure. A measure commonly used is chi square ($\chi^2 = \sum (s_i - m_i)^2 / \sigma_i$) where s_i represents the back-calculated data from the ensemble, m_i represents the measured data, and σ_i is the experimental error associated with the different NMR parameters. One has to take care in order not to over-fit the experimental data. Over-fitting happens when the difference between the simulated and experimental data is minimized during the fitting process not in order to improve the physical model describing the system, but because the model is modified to fit the random error and noise contributions. A good fit therefore always means a good

fit within the defined experimental error.

Ensemble size also influences the goodness of the fit and the ensemble should not be too small or too large. The ensemble obtained in the selection procedure is not accurate if it is composed of too few structures and therefore does not represent the conformational heterogeneity present in solution. In this case, with too few conformers in the ensemble, we say that we are over-restraining or also under-fitting. On the other hand, as we increase the ensemble size, the number of parameters (e.g. dihedral angles) that can be independently adjusted increases and the total χ^2 value will therefore decrease. The fit may improve because of an improvement in our model, but also because inaccuracies in the model are compensated by newly added structures.

There are tests that can help us decide on the ensemble size that we should choose for ensemble selection. Most commonly a plot of final χ^2 against ensemble size is used to determine the appropriate size for a given set of experimental data. The fit does not improve significantly above a certain ensemble size, and the increase in the number of degrees of freedom introduced by selecting a larger ensemble is no longer justifiable.

An alternative method, and in principle a more correct one, is to use cross-validation procedures where a part of the experimental data is left out of the ensemble selection procedure. Ensembles of different sizes are selected and the “passive” data are back-calculated from the selected ensembles and compared to the experimental data. The optimal reproduction of the passive data will normally occur for the most appropriate ensemble size. This procedure has for example been used to obtain the most appropriate ensemble size (200 structures) for describing the local conformational sampling of urea-denatured ubiquitin on the basis of multiple types of RDCs (Nodet et al., 2009).

4.13 Ensemble size in relation to convergence properties of NMR parameters

When optimizing the size of the selected ensembles, one also needs to consider the convergence characteristics of the different NMR parameters when averaged over the sub-ensembles. We say that convergence of a parameter has been reached when the addition of one more conformer to the ensemble does not perturb the calculated average parameter within a predefined limit. The convergence of parameters is particularly important as the use of too few structures in the selected ensembles will force the fitting procedure to accommodate fluctuations in the averaged NMR parameters that do not necessarily correspond to specific conformational propensities, thereby

potentially leading to incorrect residue-specific conformational sampling.

The number of conformers needed for a certain simulated parameter to converge depends on its variance. This is the reason for the different convergence properties of RDCs and chemical shifts. Chemical shifts are sensitive to the local chemical environment and are affected by main and side chain dihedral angles, amino acid identity, ring current effects and hydrogen bonding. When chemical shifts are predicted in IDPs the most important factor is the dihedral angle distribution. Carbon (C^α , C^β , C') and proton H^α chemical shifts depend mostly on the ϕ/ψ angles of the residue of interest, while the chemical shifts of the nitrogen (N) atom and the amide proton (HN) depend mostly on the ψ angle of the preceding residue. The fact that chemical shifts can be predicted from local structure only makes them a well-behaved parameter when it comes to convergence. Sufficient sampling of the ϕ/ψ space of a single amino acid can even be achieved with only a few hundred structures. As a consequence, when selecting ensembles against experimental chemical shifts, 100-200 structures are sufficient for achieving convergence of the predicted chemical shifts (Figure 4.8a-c). Scalar couplings also report on local conformational features of the polypeptide chain, and similarly to chemical shifts, a hundred conformers in the model ensemble suffice for achieving convergence.

As mentioned above, RDCs depend on both local and long-range structure, i.e. on the conformational sampling of the residue itself and immediate neighbours as well as intra-peptide long-range contacts. The large number of combinations of dihedral angle pairs that potentially all give different RDC values combined with the large range of RDCs calculated from a single structure make the convergence of the RDC average much slower. In addition, RDCs converge more slowly for longer polypeptide chains and for an IDP of one hundred amino acids, more than ten thousand structures are needed in order to achieve convergence of the RDC average (Figure 4.8d). In order to overcome this problem, we can divide the protein chain into shorter, uncoupled segments and predict the RDCs for the central amino acid of each segment (Marsh et al., 2008), thereby achieving sufficient convergence of the RDC average with only a few hundred structures (Figure 4.8d). The disadvantage of this approach is that we remove any information about long-range structure from the predicted RDCs; however, this information can be reintroduced by multiplying the predicted RDCs by a baseline that takes into account the chain-like nature of the IDP (Nodet et al., 2009; Salmon et al., 2010). Our ability to separate the contribution to the RDCs from local conformational sampling and long-range interactions allows convergence of the RDC average with an ensemble of only a few hundred structures. The use of short segments for the

calculation of RDCs therefore appears essential when using RDCs in ensemble selection procedures.

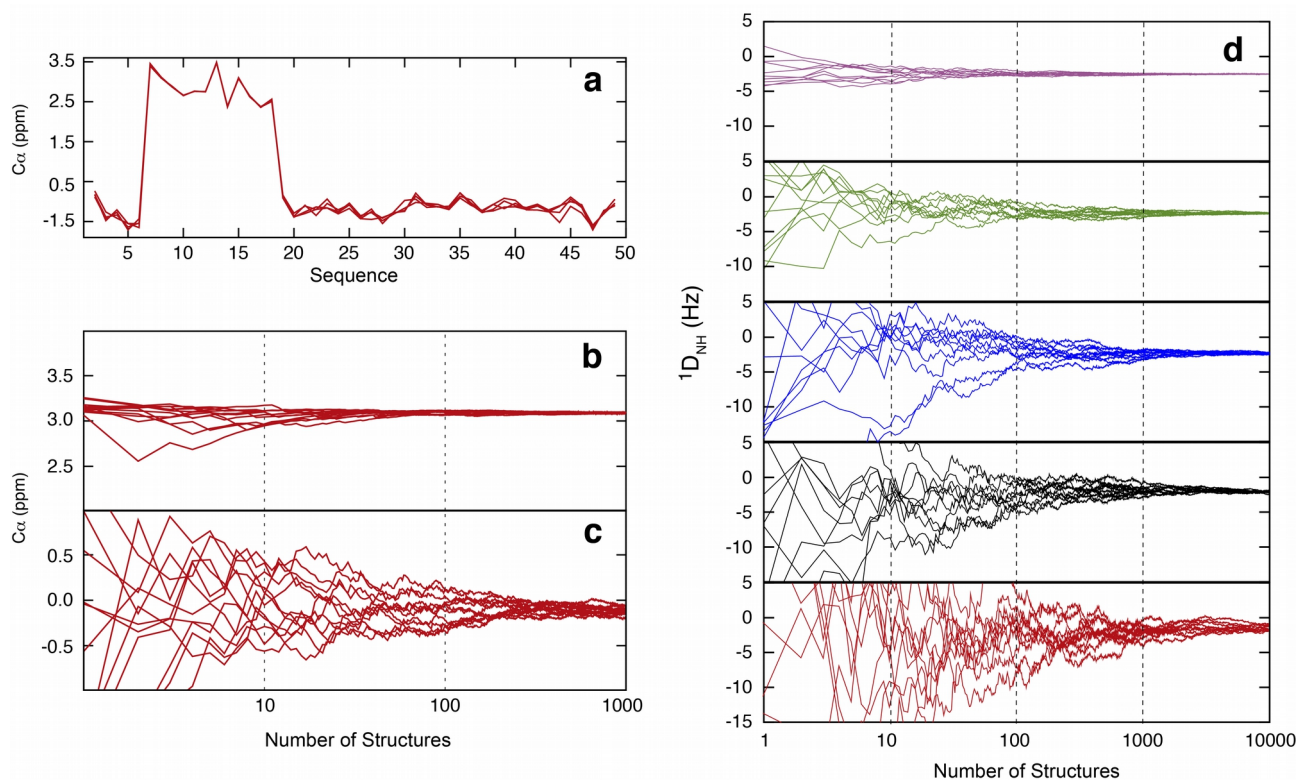


Figure 4.8: Convergence of experimental NMR parameters over structural ensembles. (a) Secondary C_{α} chemical shifts averaged over 250 conformers generated using Flexible-Meccano for a model protein of 50 amino acids of arbitrary sequence. The results for five different ensemble averages are shown. Residues 7-18 populate the α -helical region of Ramachandran space, while the remaining residues adopt random coil conformations. (b) Ensemble-averaged secondary C_{α} chemical shifts for increasing ensemble size for residue 15 of the model protein. (c) Ensemble-averaged secondary C_{α} chemical shifts for increasing ensemble size for residue 32 of the model protein. (d) Convergence of $^1D_{NH}$ RDCs over a structural ensemble with an increasing number of conformers for a model protein of 76 amino acids. Results are shown for the calculation using a global alignment tensor (red) and employing different sizes of short segments for calculating the alignment tensor: 25 (black), 15 (blue), 9 (green) and 3 (pink) amino acids. Reprinted in part with permission from (Nodet et al., 2009). Copyright 2009 American Chemical Society.

4.14 Validation of ensemble descriptions

Due to the under-determined nature of ensemble selections in general, it is useful to think about how we can potentially validate the structural ensembles that we derive from experimental NMR data. One way of doing this is to exploit the complementary nature of different data types and use cross-validation procedures where a part of the experimental data is left out of the ensemble

selection and subsequently back-calculated from the selected ensembles. If the selected ensemble correctly reproduces the local conformational sampling, the agreement between the “passive” data and that back-calculated from the selected ensemble should be good and no systematic deviations should be observed. An example of this procedure is shown in Figure 4.9 where experimental $^1D_{NH}$ RDCs measured in Tau protein are compared to the RDCs extracted from ASTEROIDS ensembles of Tau selected on the basis of chemical shift data alone. The agreement between the two sets of data is excellent and even the turn motifs in the repeat regions of Tau – where positive $^1D_{NH}$ RDCs are observed experimentally – are reproduced by the chemical shift ensemble. This type of procedure therefore validates the local conformational sampling of Tau derived from chemical shifts only.

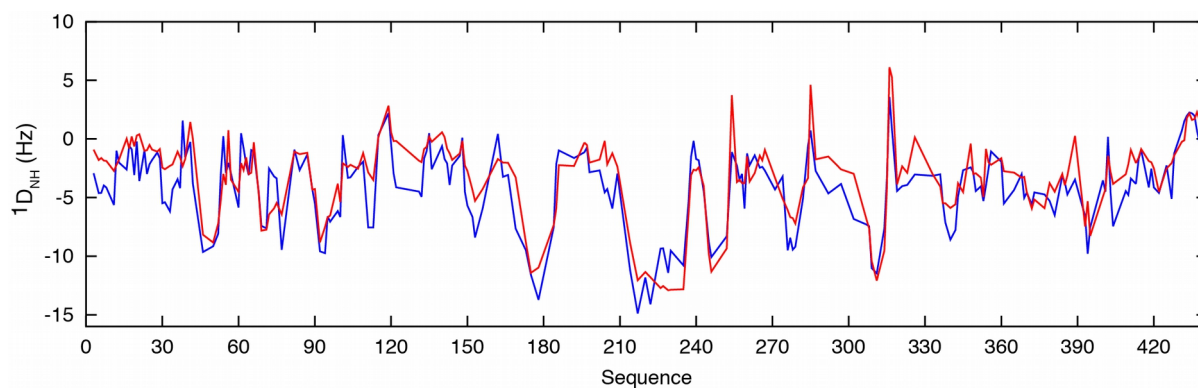


Figure 4.9: Validation of structural ensembles of IDPs derived from experimental NMR data. An example is shown of cross-validation of experimental $^1D_{NH}$ RDCs of the IDP Tau. Experimental data are shown in red, while back-calculated RDCs from an ensemble of Tau selected by the genetic algorithm ASTEROIDS on the basis of experimental chemical shifts only are shown in blue.

4.15 Conclusions and outlook

Ensemble descriptions have in recent years emerged as the preferred tool for representing the structural and dynamic properties of IDPs and their functional complexes. Within such descriptions it is assumed that the protein adopts a continuum of rapidly interconverting structures, and the determination of these representative ensembles is one of the major challenges in the studies of IDPs. In this chapter we have described how different NMR data types can be combined with sample-and-select approaches to map local conformational propensities in IDPs. In particular, we have emphasized some of the pitfalls associated with these approaches such as under- and over-restraining, and we have discussed ways to validate the derived structural ensembles. Validating

structural ensembles is particularly important if we are to use these ensembles in the future for the prediction of other, independent experimental observables or for the development of small molecules that can interfere with the biological function of IDPs.

4. Ensemble calculation for IDPs using NMR parameters

5 Methods

5.1 Sample preparation

5.1.1 Expression and purification of intrinsically disordered proteins

After initial screening of optimal expression conditions a protocol was established for expression and purification that could be used for all the IDP constructs. The IDPs expressed and purified using these procedures were: MKK4 construct (residues 12-86), the constructs of the IDR of MKK7 comprising one (residues 1-42), two (residues 1-55) or three docking-sites (residues 1-100), the JIP1 constructs termed JIP1_1 (residues 1-145), JIP1_2 (residues 115-266) and JIP1_3 (residues 238-273).

The IDP constructs were sub-cloned into a modified pET-28a vector containing an N-terminal thioredoxin and a 6xHis tag followed by a tobacco etch virus (TEV) cleavage site. *Escherichia coli* (*E. coli*) BL21(DE3) cells transformed with one of the constructs were grown in LB medium at 37°C until the optical density (OD) at 600 nm reached 0.6. Protein expression was induced by addition of isopropyl β -D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. The cultures were grown while shaking at 37°C for an additional three hours. The cells were harvested by centrifugation and frozen at -80°C. Isotopically $^{15}\text{N}/^{13}\text{C}$ and ^{15}N labelled samples were produced by growing transformed *E. coli* BL21(DE3) cells according to the protocol described by (Marley et al 2001).

All constructs were purified using Ni affinity chromatography followed by size-exclusion chromatography. For cell lysis, inhibitor cocktail (Roche) was added to the purification buffer (50 mM HEPES pH 8.0, 150 mM NaCl, 5 mM imidazole, 5 mM β -mercaptoethanol). The elution buffer was the same as the purification buffer, but with 250 mM imidazole. After removing thioredoxin and the 6xHis tag with the TEV protease, and a second Ni affinity column was performed. When purity was not sufficient another cation exchange chromatography step was performed. As the final step size-exclusion chromatography was performed on a Superdex 75 column (GE Healthcare) equilibrated with NMR buffer (50 mM HEPES pH 7.0, 150 mM NaCl, 5% glycerol, 2 mM DTT) or ITC buffer (50 mM HEPES pH 8.2, 150 mM NaCl, 10% glycerol, 0.5 mM TCEP).

5.1.2 Expression and purification of full-length MKK7

Synthetic codon-optimized full-length MKK7- β 1 gene was sub-cloned into a pET vector with an N-

5. Methods

terminal 6xHis tag followed by a TEV cleavage site. Transformed *E. coli* BL21(DE3) cells were initially grown in LB medium at 37 °C and were transferred to 18 °C when the OD at 600 nm reached 0.3. Protein expression was induced with addition of IPTG at OD_{600nm} of 0.6. Cells were harvested 12 -14 hours later by centrifugation and were frozen at -80 °C.

Harvested cells were resuspended in lysis buffer (50 mM HEPES pH 8.0, 150 mM NaCl, 10% glycerol, 5 mM imidazol, 5 mM β-mercaptoethanol, protease inhibitor cocktail (Roche)), sonicated on ice, applied to a Ni affinity chromatography column and washed with lysis buffer without protease inhibitor cocktail. Protein was eluted with elution buffer (50 mM HEPES pH 8.0, 150 mM NaCl, 5 mM imidazol, 5 mM β-mercaptoethanol) using a gradient elution 1 column volume long.

The protein was diluted to 1 mg/ml and dialysed against NMR buffer. The protein was concentrated and purified on a size-exclusion chromatography column equilibrated with NMR buffer. Under these conditions the protein concentrates to 60 μM concentration. All purification steps were performed at 4 °C.

5.1.3 Expression and purification of JNK

Synthetic codon-optimized JNK1α1 gene corresponding to residues 1-364 was sub-cloned into a pET vector with a C-terminal 6xHis tag. Transformed *Escherichia coli* BL21(DE3) cells were initially grown in LB medium at 37 °C and were transferred to 18 °C when the OD at 600 nm reached 0.3. Protein expression was induced with addition of IPTG at OD_{600nm} of 0.6. Cells were harvested 12 -14 hours later by centrifugation and were frozen at -80 °C.

Harvested cells were resuspended in lysis buffer (50 mM HEPES pH 8.0, 150 mM NaCl, 10% glycerol, 5 mM imidazole, 5 mM β-mercaptoethanol, protease inhibitor cocktail (Roche)), sonicated on ice, applied to a Ni affinity chromatography column and washed with lysis buffer without protease inhibitor cocktail. Protein was eluted with elution buffer (50 mM HEPES pH 8.0, 150 mM NaCl, 5 mM imidazole, 5 mM β-mercaptoethanol) using a gradient elution 1 column volume long.

For ITC measurements, the protein was loaded on a size-exclusion chromatography column equilibrated with ITC buffer. Protein concentrated to 100 μM under these conditions. For NMR measurements, the protein was diluted to 0.5 mg/ml and dialysed against NMR buffer. The protein was concentrated and subsequently purified on a size-exclusion chromatography column equilibrated with NMR buffer. Under these conditions, the protein concentrates to approximately 60 μM. All purification steps were performed at 4 °C.

5.2 NMR measurements

5.2.1 Backbone assignment experiments

Spectral assignments of the regulatory domain of MKK7 (residues 1-100) and MKK4 (12-86) were obtained at 5 °C, 150 mM NaCl, 50 mM HEPES, pH 7.0 using a set of BEST-type triple resonance spectra: HNCO, intra-residue HN(CA)CO, HN(CO)CA, intra-residue HNCA, HN(COCA)CB and intra-residue HN(CA)CB (Lescop et al., 2007). All spectra were processed in NMRPipe (Delaglio et al., 1995) and the program MARS (Jung and Zweckstetter, 2004) was used for automatic assignment of spin systems followed by manual verification. Random coil values for the calculation of secondary chemical shifts were obtained from the neighbour corrected intrinsically disordered protein library (Tamiola et al., 2010).

5.2.2 Measurement of RDCs

RDCs of MKK7₁₋₁₀₀ and MKK4₁₂₋₈₆ RDCs were obtained at 5 °C by aligning the proteins in a liquid crystal composed of poly-ethylene glycol (PEG C8E5, Sigma) and 1-octanol. The protein sample of (220 µl) was first mixed with 12.5 µl PEG. Octanol was titrated into this solution in steps of 0.25 µl. After ~ 1.5 µl of added octanol the solution became turbid and cleared again at ~ 2.5 µl of added octanol at room temperature and > 3 µl at 5 °C. The total amount of added octanol was 3.7 µl for MKK7₁₋₁₀₀ giving rise to a ²H quadrupole splitting of 31 Hz (Rückert and Otting, 2000). For MKK4₁₂₋₈₆ the amount of added octanol was 3.25 µl which gave rise to a ²H quadrupole splitting of 32 Hz. The RDCs were measured using BEST-type 3D HNCO- and HN(CO)CA experiments modified to allow for spin-coupling measurements in the ¹³C dimension (Rasia et al., 2011).

5.2.3 Relaxation dispersion and chemical exchange saturation transfer experiments

The ¹⁵N relaxation dispersion (Hansen et al., 1998) and CEST experiments (Vallurupalli et al., 2012) were carried out on a ¹³C, ¹⁵N labelled sample of the regulatory domain of MKK7 (200 µM) containing 10% molar ratio of JNK1. All experiments were recorded at 5 °C in a buffer consisting of 150 mM NaCl, 50 mM HEPES, pH 7.0 containing 5% glycerol. The relaxation dispersion experiments were carried out at ¹H frequencies of 600 and 800 MHz using a constant-time relaxation delay of 32 ms. Fourteen points (including two duplicates) were recorded for each dispersion curve corresponding to CPMG frequencies between 31.25 and 1000 Hz.

The CEST experiments were carried out at a ^1H frequency of 700 MHz with two different ^{15}N B_1 field strengths of 22 and 44.5 Hz, where the B_1 field was applied during a constant period of 0.3 s. The experiments were recorded with 78 (B_1 field of 44.5 Hz) or 94 (B_1 field of 22 Hz) two-dimensional planes with the position of the ^{15}N B_1 field ranging from 101.2 to 130.6 ppm in steps of 30 Hz (B_1 field of 44.5 Hz) or 25 Hz (B_1 field of 25 Hz). The relaxation dispersion and CEST data were fitted using ChemEx (Vallurupalli et al., 2012) assuming a two-site exchange model. For the D2 docking-site, the residues were fitted individually, while for the D3 docking-site all residues were fitted simultaneously to a single exchange rate and population as described in the text. For the D3 docking-site, it was not necessary to evoke a ΔR_2 contribution in order to fit the experimental CEST profiles and they were therefore fixed to 0 for all residues (See Results).

5.3 Other methods

5.3.1 Crystallization of JNK

JNK1 sample (2 mg/ml) was mixed with peptide (QRPRPTLQLPLA) solution to yield an approximate fourfold peptide excess, and the solution was supplemented with 2 mM AMP-PNP and 2 mM MgCl_2 and concentrated to 7.5 mg/ml. Crystallization conditions were initially found in an HTX platform screen (EMBL, Grenoble). Cartesian PixSys 4200 crystallization robot (Genomic Solutions, U.K.) was used to set up 200 nl drops in CrystalQuick plates (Greiner Bio-One North America, Inc.) in vapor diffusion sitting drop setup. 100 nl of protein solution was mixed with 100 nl of precipitant. The precipitant was 3 M NH_4SO_4 with 100 mM Tris pH 8.0 or 100 mM HEPES pH 7.0. After the drops were set up the plates were kept at 4 °C. The crystals have become visible after 7 days. A manual screening in 24 well vapour diffusion hanging drop setup was performed in order to reproduce the crystals on a larger scale. 1 μl of protein was mixed with 1 μl of precipitant. Crystals eventually grew in a narrow range of precipitant concentration: 2.4 M NH_4SO_4 , pH \sim 7.4 at 4 °C. First crystals became visible after 4 days.

5.3.2 Determination of the crystal structure of JNK

The complex between JNK1 and the peptide (QRPRPTLQLPLA) corresponding to the second docking-site of MKK7 crystallized in space group C2 with unit cell parameters of $a = 108.4$, $b = 179.6$, $c = 100.8$, $\beta = 110.2$ with four molecules of JNK1 and four peptides in the asymmetric unit. Oscillation data were collected on the ESRF beamline ID29 (de Sanctis et al., 2012) at 0.9763 Å

wavelength on a Pilatus 6M-F detector with 37 ms exposures and 0.15 degrees per exposure and 1.1×10^{12} photons/exposure. Data were processed with XDS/XSCALE (Kabsch, 2010), and a Free set of reflections was selected in thin shells and set aside. Phases were obtained by molecular replacement with PHASER (McCoy et al., 2007), splitting the search model (3O2M) into two sub-domains consisting of residues 7-110+334-364, and 111-333. Solutions were found for the first sub-domain with Translational Z-scores of 34.8, 34.2, 28.8 and 28.8 and for the second domain with Translational Z-scores of 26.4, 46.1, 54.8 and 50.1. The structures were refined by multiple rounds of manual rebuilding in COOT (Emsley and Cowtan, 2004) followed by refinement in BUSTER (Bricogne et al., 2011). Several regions of poor electron density were left unmodeled, specifically the loops in regions K30-I39, G177-P184, P281-L289 and the linker region between P338 and the C-terminal helix. The latter region is especially poorly ordered in subunit B. The structure was refined to acceptable geometry and R-values, but the overall Wilson B-factor and average B-values are higher than other structures at the same resolution. The structure was deposited in the PDB database under accession number 4UX9.

5.3.3 Thermal shift assay

The fluorescent dye Sypro orange was used to monitor protein unfolding with increasing temperatures (25 °C - 75 °C). Sypro orange is an environmentally sensitive dye which non-specifically binds to hydrophobic surfaces. The unfolding of the protein exposes the hydrophobic regions of proteins, and binding of the dye to these surfaces provides shielding from water and results in an increase in fluorescence.

A mixture of 1 µl of 5000x diluted Sypro, 2 µl of JNK1 (40 µM), 4 µl 5x concentrated buffer and an additive was prepared in a ELISA 96 wells PCR plate. Mixtures in all of the wells were diluted to 20 µl final volume.

The thermal shift was conducted in the Mx3005P qPCR (real time PCR) system (Agilent technologies, CA) originally developed for qPCR reactions. The MX3005P qPCR machine uses a quartz tungsten halogen lamp in combination with one of five selectable bandpass filters for excitation and a single scanning photomultiplier tube with one of five bandpass filters for fluorescence detection.

The temperature step was 1 degree for each minute from 25°C to 75°C. Fluorescence readings were performed in triplicate for each step. The excitation and emission bandpass filters used for the dyes

5. Methods

were 492 nm and 585 nm, respectively. The melting temperature, T_m , value was calculated as an average of the highest and lowest value of fluorescence.

5.3.4 ITC measurements

ITC measurements were performed on the MicroCal iTC200 (GE Healthcare, PA) at 20°C. Injection parameters were identical for all experiments (1.5 μ l injections every 180 seconds, 26 in total at a stirring speed of 800 rpm). Prior to the experiments, a size-exclusion chromatography was performed for all proteins using the ITC buffer (50 mM HEPES pH 8.2, 150 mM NaCl, 10% glycerol, 0.5 mM TCEP). Constructs of the regulatory domain of MKK7 were titrated into a solution of JNK1 with concentrations between 45 μ M and 90 μ M. For the constructs of the intrinsically disordered domain of MKK7, the Trx 6xHis tag was not removed by the TEV protease allowing accurate concentrations to be determined from UV absorbance and calculated molar extinction coefficients. A control experiment was carried out showing that no interaction occurs between isolated Trx and JNK1. The analysis of the ITC titration data reveals that JNK1 is only 70% binding competent (or the concentration of the protein is not accurately determined) and we, therefore, corrected all the JNK1 concentrations by the same factor prior to the fitting of the experimental data.

Peptides (>98% purity) corresponding to the three docking-sites of MKK7 (ARRRIDLNLDIS, QRPRPTLQLPLA and ARPRHMLGLP) were obtained from CASLO Laboratory ApS (Denmark). The peptide corresponding to the first docking-site D1 was not soluble and could, therefore, not be used for affinity measurements. The concentrations of the peptides corresponding to the docking-sites D2 and D3 were estimated by amino acid analysis.

5.4 Ensemble calculations

5.4.1 Ensemble calculations for MKK7

Ensembles of MKK7₁₋₁₀₀ were constructed using Flexible-Meccano (Bernadó et al., 2005b; Ozenne et al., 2012b) imposing α -helices of different lengths and different positions within the region comprising residues 3-32 (minimum helix length was four amino acids). A total of 378 ensembles comprising 10000 conformers each were created and ensemble-averaged chemical shifts and RDCs were calculated using SPARTA (Shen and Bax, 2007) and PALES (Zweckstetter and Bax, 2000), respectively. In addition, a statistical coil ensemble comprising 50000 conformers was

used to obtain expected chemical shifts and RDCs for MKK7 in the absence of helical elements. The genetic algorithm ASTEROIDS was used to select combinations of ensembles that best fit the experimental data i.e. a limited number of helical ensembles in exchange with the unfolded form as described previously (Jensen et al., 2008, 2011). Three different types of RDCs, $^1D_{NH}$, $^1D_{CaH\alpha}$ and $^1D_{CaC'}$, were used in the optimization procedure together with C^α chemical shifts for the residues 1-35. The data were included with experimental uncertainties of 1.0 Hz ($^1D_{NH}$), 2.0 Hz ($^1D_{CaH\alpha}$) and 0.3 Hz ($^1D_{CaC'}$) and 0.15 ppm (C^α chemical shifts) ensuring an almost equal contribution to the total target function. The number of helical ensembles, N , was gradually increased (from $N = 1$ to $N = 4$) and a standard F-test was applied to the reproduction of the RDCs at each step to test for the statistical significance of adding additional helical ensembles to the fitting procedure. The analysis shows that residues 1-35 of MKK7 is best described by an ensemble of three specific helical conformers in exchange with an unfolded form.

In a second step, we carried out an ensemble selection of the entire regulatory domain on the basis of the experimental chemical shifts and RDCs. Five ensembles comprising 200 conformers each were selected from a large pool of structures containing the already described conformational helical equilibrium at the N-terminus. Chemical shifts ($^{13}C^\alpha$, $^{13}C^\beta$, $^{13}C'$, ^{15}N and $^1H^N$) and RDCs ($^1D_{NH}$, $^1D_{CaH\alpha}$ and $^1D_{CaC'}$) were used in the ensemble selection employing the following experimental uncertainties: 0.5 Hz ($^1D_{NH}$), 1.0 Hz ($^1D_{CaH\alpha}$), 0.15 Hz ($^1D_{CaC'}$), 0.1 ppm ($^{13}C^\alpha$, $^{13}C^\beta$, $^{13}C'$), 0.2 ppm (^{15}N) and 0.04 ppm ($^1H^N$). In general, the protocol for the ensemble selections employed five iteration steps (regeneration of pool and subsequent selections) as described previously (Jensen et al., 2010; Ozenne et al., 2012a).

5.4.2 Ensemble calculations for MKK4

Ensembles of MKK4_12-86 were constructed using the algorithms Flexible-Meccano (Bernadó et al., 2005b; Ozenne et al., 2012b) and ASTEROIDS (Nodet et al., 2009). The selection was performed through a previously described protocol involving iterative Flexible-Meccano/ASTEROIDS selection (Jensen et al., 2010). First, 10000 conformers were created and ensemble-averaged chemical shifts and RDCs were calculated using SPARTA (Shen and Bax, 2007) and an in-house RDC calculation algorithm (Ozenne et al., 2012b). The genetic algorithm ASTEROIDS was used five times to select sub-ensembles that fit the experimental data best and each ensemble contains 200 structures. A new residue-specific statistical coil library was generated using the 5x200 structures. The new statistical coil library containing 1000 ϕ/ψ angles per residue

5. *Methods*

was used to generate another starting pool containing 10000 conformers from which five 200-residue ensembles were again selected with ASTEROIDS. This procedure was repeated until the back-calculated data fitted the experimental data.

6 Results

Despite their important role in directing the specificity in signalling networks of MAPKs, the N-terminal IDRs of MKKs have not yet been studied by atomic-level resolution techniques. In this chapter I present the results of the characterization of the N-terminal IDRs of MKK7 and MKK4 and of the studies of their interaction with JNK1.

MKK7 is the only human MKK for which more than one putative docking site is present within the N-terminal IDR. Structural insight into the conformational sampling of the N-terminal IDR was obtained using chemical shifts, RDCs and relaxation. Ensemble selection approaches were used to determine the conformational sampling of the three docking sites of MKK7 in their pre-recognition states, identifying distinct intrinsic conformational sampling in each of the three docking-sites. Preceding the three sites an α -helical transient structure was observed at the N-terminus of the IDR.

Chemical shift, RDC and relaxation data were also obtained for the N-terminal IDR of MKK4. First examination of the data showed that this IDR can be described with a statistical coil ensemble without regions of transient structure. The selection approach was used to determine the conformational sampling of the docking-site of MKK4, revealing elevated PPII sampling in the docking-site region.

To get more insight into the mechanism of signal transduction between MKKs and JNK, interaction of the N-terminal IDRs of MKK7 and MKK4 with JNK1 was further studied by NMR chemical shift titrations. These experiment have helped delineating the residues of the docking-sites from MKK7 and MKK4 that are required for binding to JNK. The interaction of the IDR of MKK7 with JNK1 was also studied by relaxation dispersion and CEST experiments, providing a comprehensive picture of the kinetics of the JNK1-MKK7 complex.

Affinity measurements for the docking-sites from MKK7 and MKK4 towards JNK1 were obtained using isothermal titration calorimetry (ITC) showing that all docking sites are able to bind JNK1 with low micromolar Kds.

Last but not least, we have determined the crystal structure of JNK1 in complex with a peptide of the second docking-site of MKK7. The bound peptide displays flexibility in the complex and adopts different conformations in each of the four JNKs from the unit cell. This agrees well with results from the CEST experiments which show that the second docking-site of MKK7 remains flexible in the complex forming a so-called fuzzy complex.

6.1 Characterization of the N-terminal IDR of MKK7

MKK7 is a 419 amino acid long polypeptide chain (Figure 6.1a). Residues 1–96 are predicted to be disordered (Figure 6.1b), residues 100–380 fold into a protein kinase domain (Figure 2.2) and the last 19 residues are also predicted to be disordered. MKK7 is unique among the MKKs in that it contains three docking-site motifs that include residues 25–34 (D1), 38–47 (D2) and 70–77 (D3)

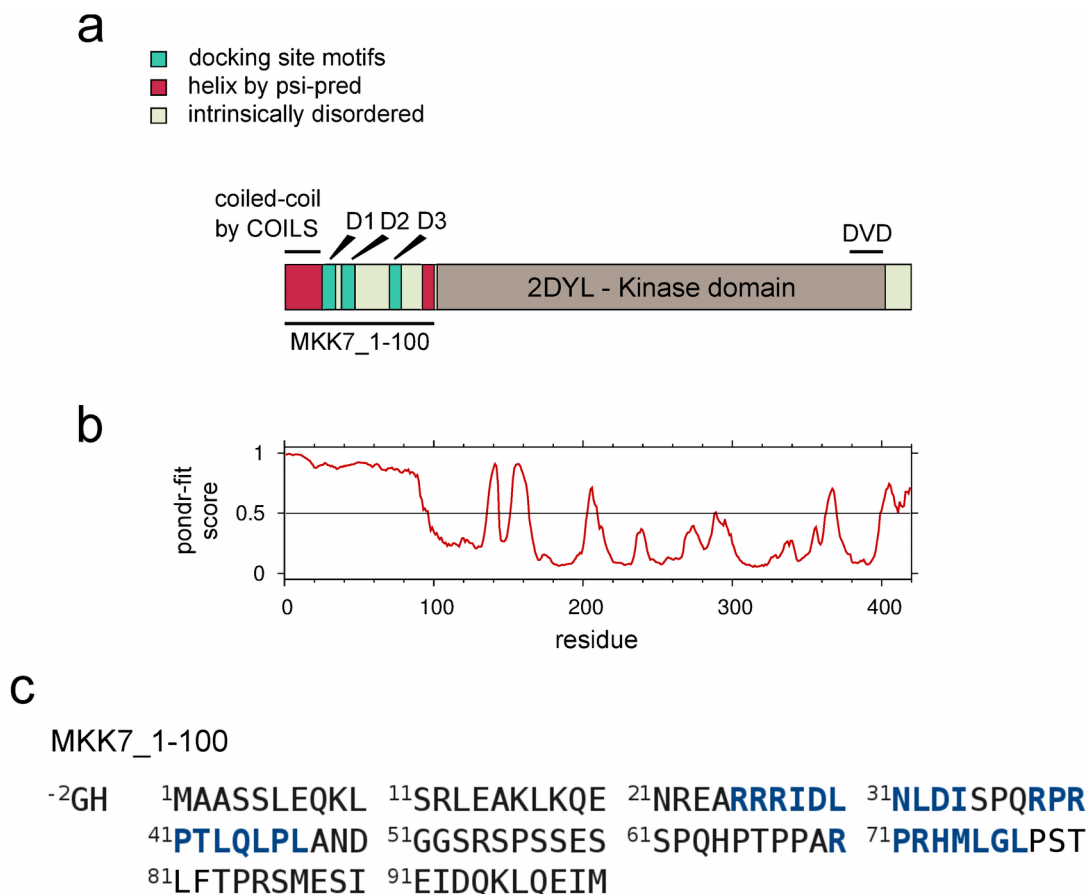


Figure 6.1: Domain organization of MKK7. a) The structure of the catalytic kinase domain (brown) has been solved previously. For regions with unknown structure, predictions from the PSIPRED (Buchan et al., 2013) and COILS (Lupas et al., 1991) programs are shown. The locations of docking-site motifs and of the DVD site are marked on the figure. b) Disorder prediction for MKK7 by the POND-FIT program (Xue et al., 2010). A POND-FIT score above 0.5 indicates that a region is predicted to be disordered. The first 96 residues of MKK7 are predicted to be disordered. c) Sequence of the studied MKK7_1-100 construct. Preceding glycine and histidine are part of the TEV site remaining after the purification tag was removed by TEV protease. Blue residues mark the three putative docking-sites.

(Figure 6.1a, 6.1c). Residues 377–400 are part of a DVD (domain for versatile docking) site (Figure 6.1a). A DVD site is used by MKKs to bind to their upstream MKKK activators (Takekawa et al., 2005).

Secondary structure prediction by the PSIPRED program predicts an additional α -helix spanning residues 90 to 102 that is not included in the MKK7 construct crystallized by Kukimoto-Niino, M. et al. (RIKEN Structural Genomics/Proteomics Initiative). Interestingly, an α -helix is predicted for residues 5–26 in the region predicted to be intrinsically disordered (Figure 6.1, 6.2a). In approximately the same region (residues 1 to 30, with scanning window of 21) the COILS program (Lupas et al., 1991) predicts a coiled-coil with high confidence (Figure 6.2b).

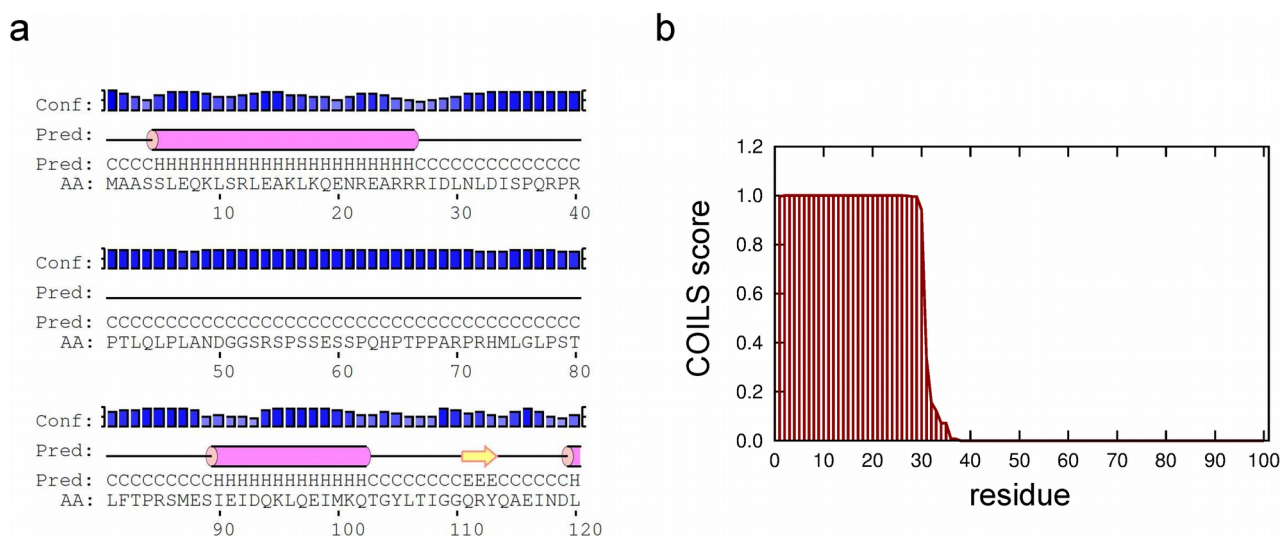


Figure 6.2: Structure predictions for MKK7. a) A PSIPRED prediction for the MKK7 kinase (shown only for the first 120 amino acids). An α -helix is predicted at the beginning of the N-terminal IDR. Additionally, another α -helix is predicted for residues 90 to 102 which were not included in the construct for which the crystal structure was solved. PSIPRED incorrectly assigns residues 106 to 109 to coil conformation while in the crystal structure these adopt a three-residue β -strand. H – helix, C – coil, E – strand b) COILS program prediction for MKK7. The COILS program calculates the potential of a sequence to form a coiled-coil. The first 30 amino acids of MKK7 have a high potential to form coiled-coil.

With the goal of characterizing the N-terminal IDR of MKK7 and of studying the interaction of the docking-site motifs with JNK1, residues 1–100 (MKK7_1-100) were cloned into several vectors with different tags. N-terminally or C-terminally His-tagged protein did not express well and degraded and a GST-tag enhanced the degradation even more. Trx and GB1-tags enhanced the expression by more than 10 times which allowed purification of sufficient amounts of non-degraded protein using Ni^{2+} affinity column and size-exclusion chromatography (Methods, Figure 6.3).

6. Results

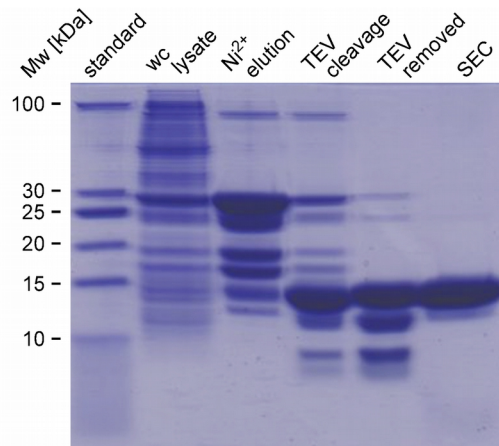


Figure 6.3: SDS-PAGE gel analysis of MKK7_1-100 purification fractions. MKK7_1-100 is degraded during expression as can be seen by three bands in the "TEV removed" fraction, but the longest intact product can be separated from the two degradation products using a size-exclusion chromatography (SEC). wc lysate – whole cell lysate

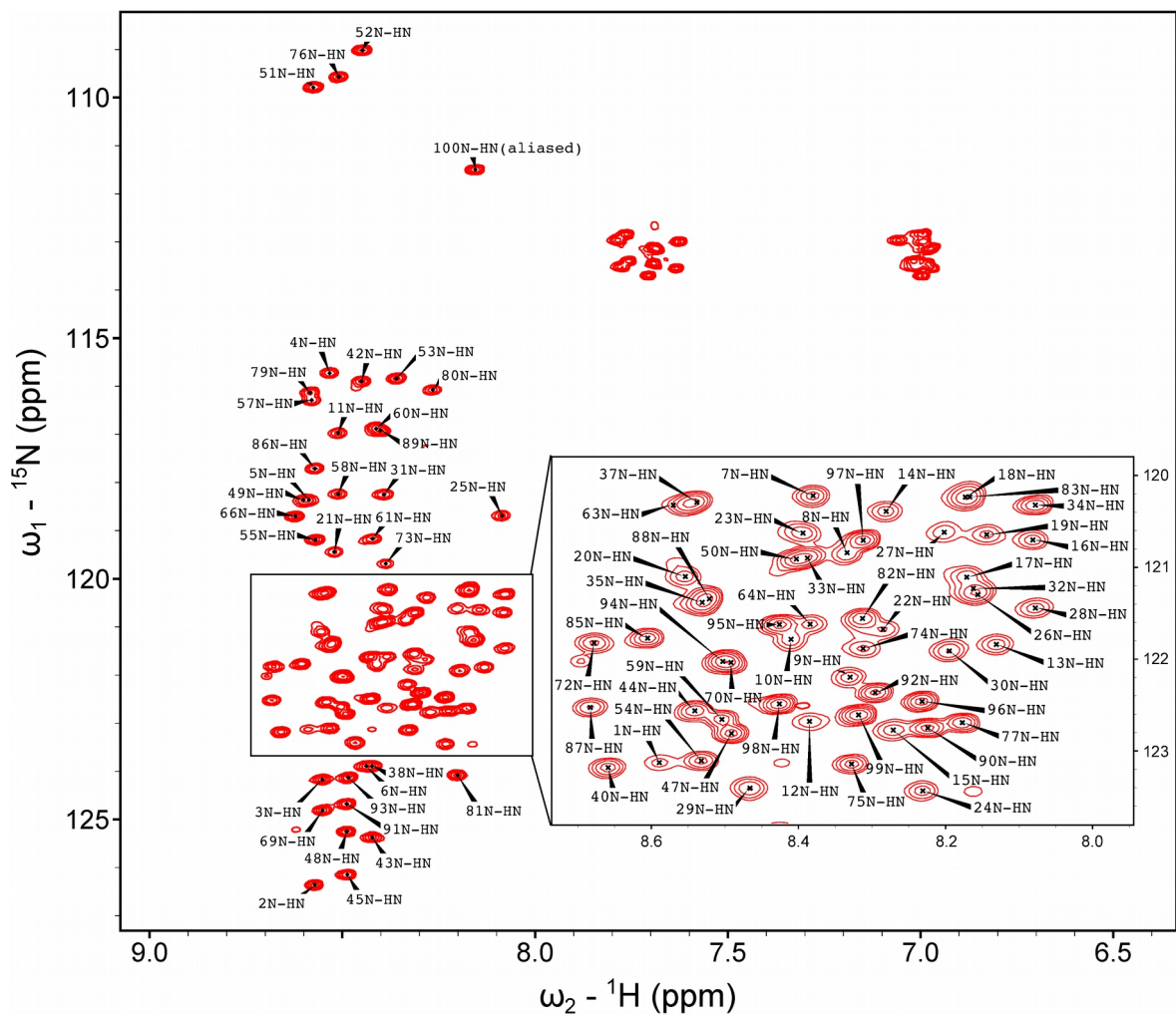


Figure 6.4: Spectral assignment of the ^1H - ^{15}N HSQC spectrum of MKK7_1-100 at 278 K in 50 mM HEPES, 150 mM NaCl at pH 7.0.

The purified MKK7_1-100 construct was assigned using standard triple-resonance assignment experiments (Figure 6.4). BEST-type HNCO, HN(CO)CA, HN(COCA)CB experiments were used to obtain inter-residue chemical shifts and intra-residue HN(CA)CO, HNCA, HN(CA)CB type experiments were used to obtain intra-residue chemical shifts (Lescop et al., 2007). A temperature and pH titration resolved remaining ambiguities for five residues.

Secondary chemical shifts were calculated from the chemical shifts obtained from the assignment experiments. The temperature at which chemical shifts are obtained must match the temperature at which experimental random coil values were obtained, otherwise the temperature dependent chemical shift deviations may be interpreted as secondary structure content. Adjustment for non-standard temperature was carried out before the secondary chemical shift calculation (Kjaergaard et al., 2011). The random coil values were calculated using the ncIDP web-server (nmr.chem.rug.nl/ncIDP) (Tamiola et al., 2010). Secondary chemical shifts reveal that an α -helical element is present at the N-terminus, spanning residues 5–29 (Figure 6.5). The presence of the transient α -helix means that D1 partially adopts α -helical conformations. This is an interesting observation as it is known that docking-sites bind in extended conformations (Section 2.3.1., Figure 2.8).

Examination of the rest of the protein sequence shows that it adopts extended statistical coil conformations. The secondary ^{13}C chemical shifts indicate that D2 and D3 could be equal as they can both be described by the random coil model of chemical shifts. Nevertheless, characteristic strongly positive ^{15}N secondary chemical shifts in the D2 site show a possible residual structure that the ^{13}C chemical shifts do not report on in an unambiguous manner. Positive ^{15}N secondary chemical shifts have been shown to correlate with increased PPII structure (Ozenne et al. 2012) which means that the D2 could have a higher PPII content compared to D3 (Figure 4.6).

The presence of a transient α -helix as observed in the secondary chemical shifts is confirmed by the $\{^1\text{H}\}$ - ^{15}N heteronuclear NOEs. These are higher for residues 5–32 compared to the remainder of the chain indicating a reduction in the flexibility on the pico to nanosecond time scale caused by the presence of residual α -helical propensity (Figure 6.6a). The rest of the protein sequence displays low hetNOE values that are in agreement with the high degree of flexibility expected for IDPs.

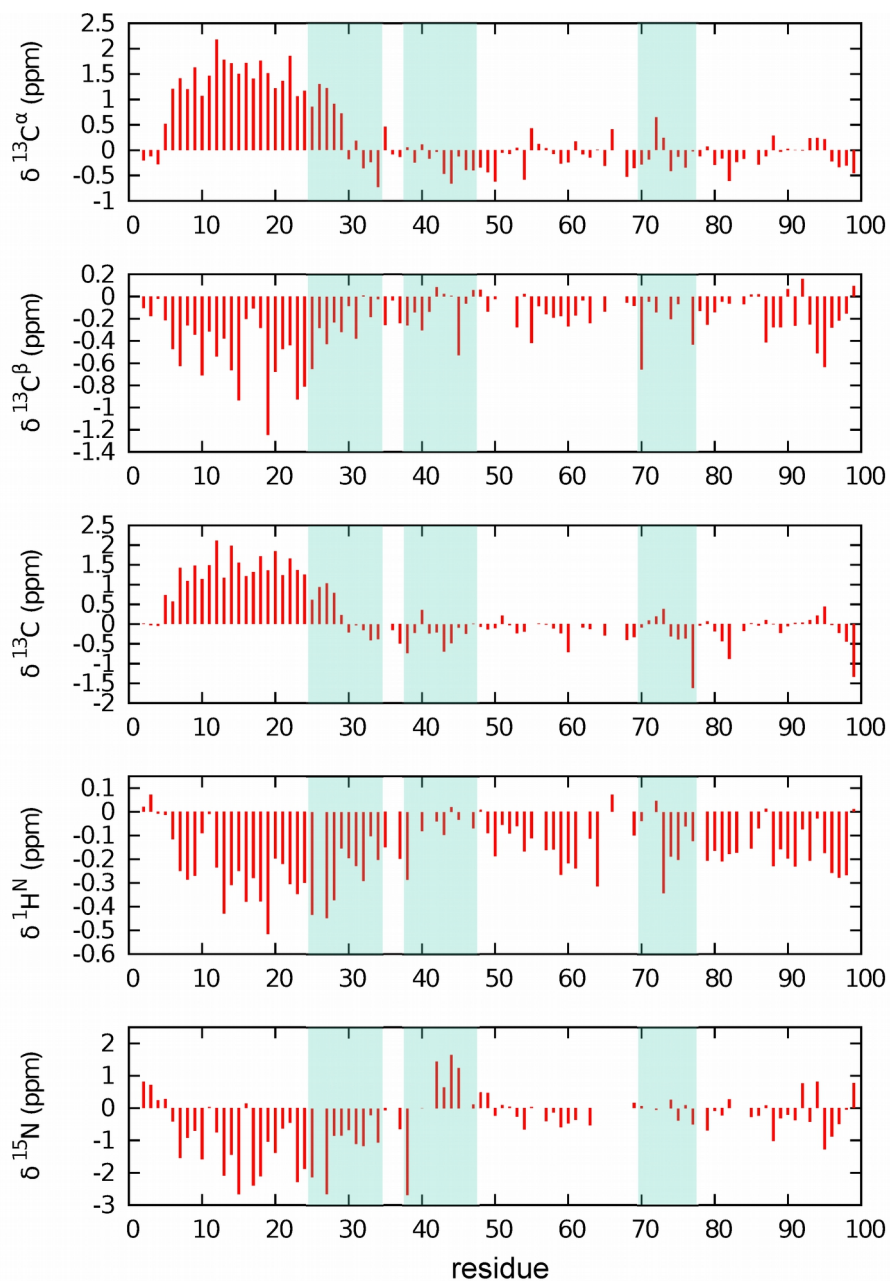


Figure 6.5: Analysis of secondary carbon, nitrogen and amide proton chemical shifts confirms that MKK7-1-100 is disordered and reveals the presence of a transient α -helix at the N-terminus of the construct. ncIDP random coil values were used for secondary chemical shift calculation.

HetNOEs are slightly below average for residues 44–64 where the sequence is rich in glycines, serines, prolines and a few charged amino acids. Measured longitudinal R_1 and transverse R_2 relaxation rates further confirm that the α -helical region (5–32) and the fully disordered region (33–100) display differential dynamics on the ps- to ns- timescale (Figure 6.6b). The R_1 relaxation rates are similar throughout the sequence (Figure 6.6c).

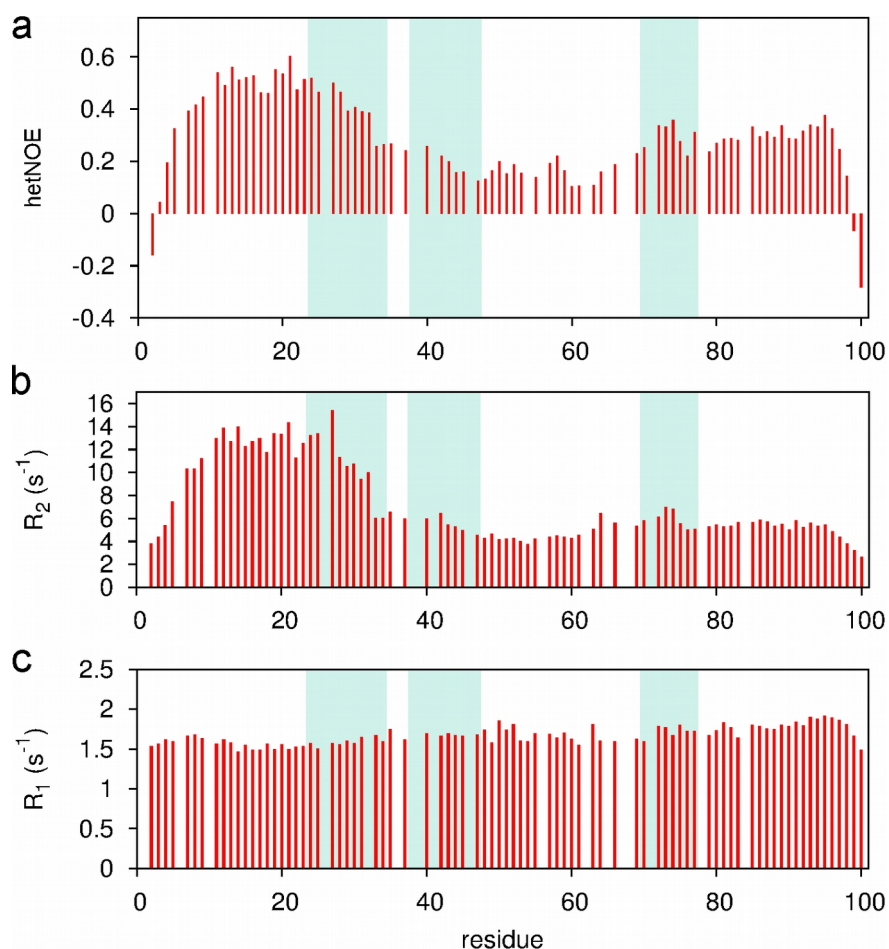


Figure 6.6: $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE, ^{15}N R_1 and R_2 relaxation rates (600 MHz, 5 °C) for MKK7_1-100 agree with secondary chemical shift analysis showing the presence of a transiently populated α -helix within the first 32 residues of the construct. a) $\{^1\text{H}\}$ - ^{15}N heteronuclear NOEs b) ^{15}N R_2 c) ^{15}N R_1 .

In order to structurally characterize MKK7_1-100 in more detail, five types of RDCs were measured and compared to a prediction of RDCs from a statistical coil model obtained by Flexible-Meccano (Bernadó et al., 2005b; Ozenne et al., 2012b) (Figure 6.7). Again, the region including the residues 5–32 displays the largest deviations from the random coil model where consistently positive $^1\text{D}_{\text{NH}}$ RDCs can be observed and where dipolar waves are present in all RDC datasets. Two other prominent features can be seen spanning residues 40–50 and surrounding residue 65. In these two regions the $^1\text{D}_{\text{NH}}$ and $^4\text{D}_{\text{H}\alpha\text{HN}}$ are on average lower, while $^1\text{D}_{\text{C}\alpha\text{H}\alpha}$ are higher. This is possibly a consequence of a high content of proline residues in those two regions which can cause the polypeptide chain to be more extended. The Flexible-Meccano algorithm is partially able to predict these deviations because the prolines affect the sampling through steric effects. A quenching of the RDCs between the two proline rich regions can be observed for residues 44–64 due to higher

flexibility of this region (the previously mentioned glycine-serine-proline rich region).

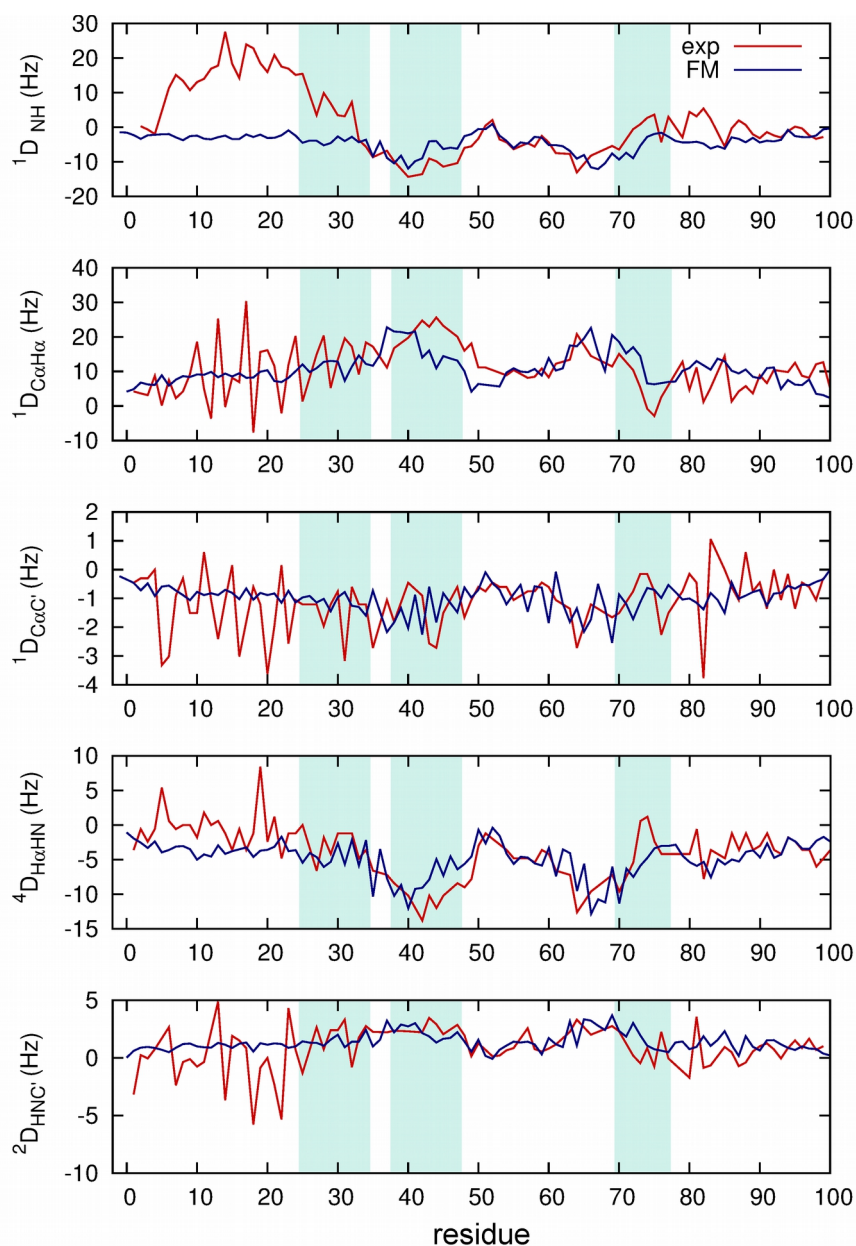


Figure 6.7: $^1D_{NH}$, $^1D_{C\alpha H\alpha}$, $^1D_{C\alpha C'}$, $^4D_{H\alpha HN}$ and $^2D_{HNC'}$ RDCs (red) measured for MKK7_1-100 aligned in a PEG-octanol liquid crystal at 5 °C. The residual 2H quadrupole splitting was 31 Hz. The RDCs are compared to a prediction of RDCs using a statistical coil model (blue, Flexible-Meccano (FM)).

We can therefore see that secondary chemical shifts and RDCs are sometimes sensitive to the same features in the MKK7_1-100 sequence (the α -helix), while other features are detected only by a few chemical shift types or by a few sets of RDCs (region 40–50) demonstrating the complementarity of

the two experimental parameters. The ensemble selection methods can be used to integrate both of these data types and to determine the conformational sampling of MKK7_1-100 more accurately. Furthermore, both chemical shift data and RDCs show that D1 is partially located in the α -helix but the exact end point of the helix within the D1 site still needs to be determined. The analysis of dipolar waves observed in several sets of RDCs using ensemble selection methods allows us to accurately determine the start and end points of the α -helices within the MKK7_1-100 ensemble.

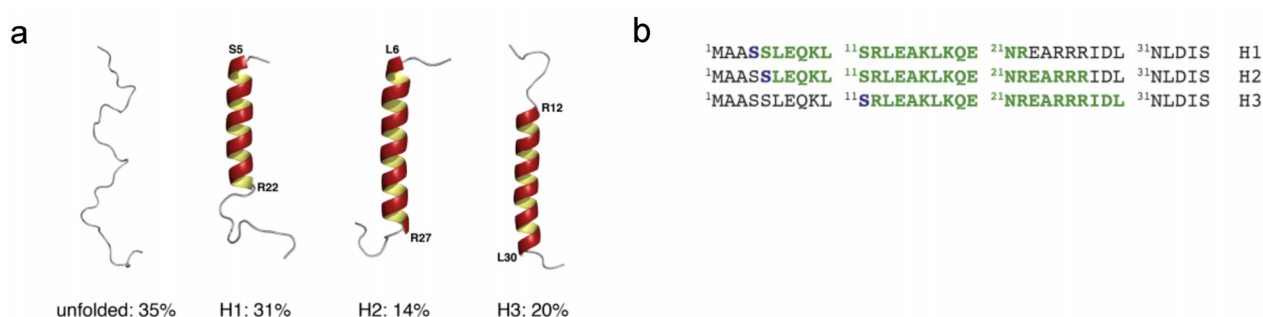


Figure 6.8: Analysis of experimental $^{13}\text{C}^\alpha$ chemical shifts and $^1\text{D}_{\text{NH}}$, $^1\text{D}_{\text{CaHa}}$ and $^1\text{D}_{\text{CaC}'}$ RDCs using the minimal ensemble approach. a) The analysis revealed a minimal ensemble consisting of three helices with different start and end points (marked on the figure) and a disordered form. The populations of each sub-ensemble are indicated below the representative α -helices. b) The three α -helices (green residues) are preceded by serine residues that stabilize the α -helices through capping interactions.

With these questions in mind, the experimental $^1\text{D}_{\text{NH}}$, $^1\text{D}_{\text{CaHa}}$ and $^1\text{D}_{\text{CaC}'}$ RDCs and $^{13}\text{C}^\alpha$ chemical shifts were used in a minimal ensemble selection approach (Jensen et al., 2008) (Methods 5.4.1, Table 6.1) to analyse the region containing residues 5–32 where high α -helix propensity is observed. The analysis revealed that this region can be described as an equilibrium between three α -helices, each with specific start and end points, and a disordered form of the protein (Figure 6.8a). All three α -helices within the calculated ensemble are preceded by serine residues that may stabilize the α -helices through N-capping interactions (Figure 6.8b). Residue 30 is the last α -helical residue in the docking-site D1 meaning that the basic part of the docking-site is largely helical.

After the initial characterization of the N-terminal α -helix, the sampling in the full MKK7_1-100 construct was determined using ASTEROIDS. Ensembles comprising 200 conformers each were selected from a large pool of structures containing the already described conformational helical equilibrium at the N-terminus. Chemical shifts ($^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, $^{13}\text{C}'$, ^{15}N , $^1\text{H}^\text{N}$) and $^1\text{D}_{\text{NH}}$, $^1\text{D}_{\text{CaHa}}$ and $^1\text{D}_{\text{CaC}'}$ RDCs were used in the selection procedure giving excellent agreement with values calculated from the selected ensembles (Figure 6.9).

Table 6.1: Modeling the N-terminus of MKK7 using ensembles with increasing number of helical states

Number of helical conformers	χ^2 ^a	Number of optimized parameters ^b	Helical conformers ^c	Population (%) ^d	Significance ^e
1	1445	4	5-26	52	
2	740	7	5-23 12-30	40 24	P<0.0001
3	605	10	5-22 6-27 12-30	31 14 20	P=0.0005
4	577	13	5-22 6-13 6-27 12-30	28 15 10 22	P=0.2610

^aThe target function for the χ^2 included 97 experimental RDC data points.

^bOne helix implies the optimization of three parameters: starting amino acid, final amino acid and the population. In addition a scaling factor is optimized to take into account the absolute level of alignment for the RDCs.

^cRange of the invoked helices.

^dThe population of the invoked helices. The remaining conformers are completely unfolded.

^eSignificance of the improvement of this model as compared to the simpler model calculated using a standard F-test.

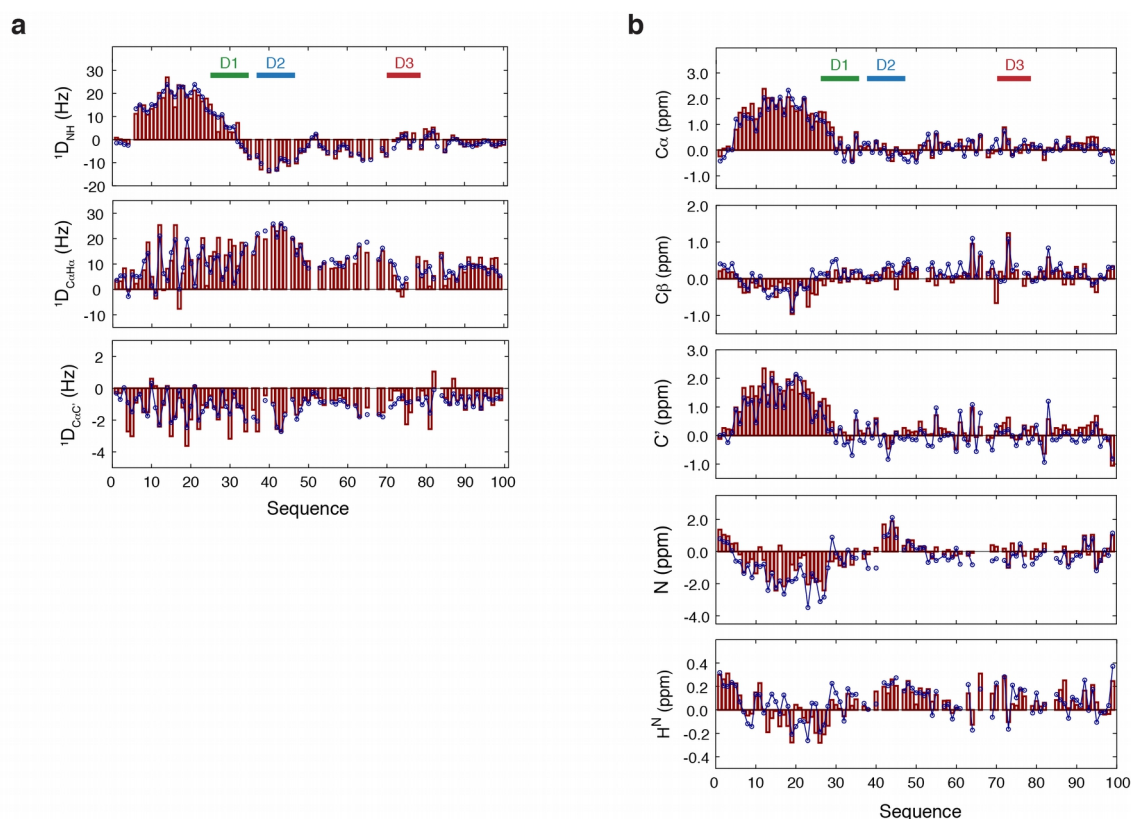


Figure 6.9: Comparison of experimental RDCs (red) with values back-calculated (blue) from a selected ASTEROIDS ensemble comprising 200 conformers. (b) Comparison of experimental chemical shifts (red) with values back-calculated (blue) from a selected ASTEROIDS ensemble comprising 200 conformers.

The residue specific conformational sampling in the selected MKK7_1-100 ensembles can be described by dividing the Ramachandran space into four regions termed β -strand (β S), PPII (β P), α -helix (α R) and left-handed helix (α L) (Figure 6.10a). The populations in each of these regions can then be plotted separately as a function of sequence (Figure 6.10b). Compared to the Flexible-Meccano model the selected ensembles sample β -strand conformations less frequently and PPII conformations are overall slightly enhanced. One can also look closer, at the sampling in the docking-sites of MKK7 by plotting the φ/ψ distributions for each residue separately as Ramachandran plots (Figure 6.10c). The differences in the sampling between the three docking-sites of MKK7 are apparent. D1 displays increased sampling in the α -helical region compared to the other two docking-sites. D2 samples mostly PPII conformations and very little α -helix which is in contrast to D1. D3 samples extended conformations (β -strand and PPII) equally and its sampling remains closest to the starting statistical coil library model. The obtained distributions can be compared to other conformations of the docking-sites when bound to JNK. To this end, the dihedral angles were extracted for all the docking-site motifs that were co-crystallized with JNK (Figure 6.10d). The JNK-bound docking-site motifs extensively sample extended conformations with most dihedral angles located between the β S and β P regions. The dihedral angles before the Φ_{A-2} position, corresponding to the residues of the basic cluster, are more scattered reflecting the different conformations that this part of the peptides adopts in the crystal structures (Section 2.3.1, Figure 2.8). The sampling in the hydrophobic part of the motif (Φ_{A-2} , Φ_A and Φ_B) is more consistent between different co-crystallized peptides and can be interpreted with higher certitude. Of the three docking-sites of MKK7 the docking-site D2 adopts conformations that most closely resemble the sampling in the crystal structures. As expected, residues that are followed by prolines over-populate PPII but interestingly, the enhanced sampling of PPII is observed also for residues that are not followed by prolines (Thr42, Leu43 and Gln44). The different sampling regimes observed for the three docking-sites of MKK7 pose several questions: Does the intrinsic conformational sampling in the pre-recognition state of MKK7 affect the affinities of the docking-sites for JNK (Section 6.4)? Does the different intrinsic conformational sampling also affect the kinetics of binding (Section 6.7)?

6. Results

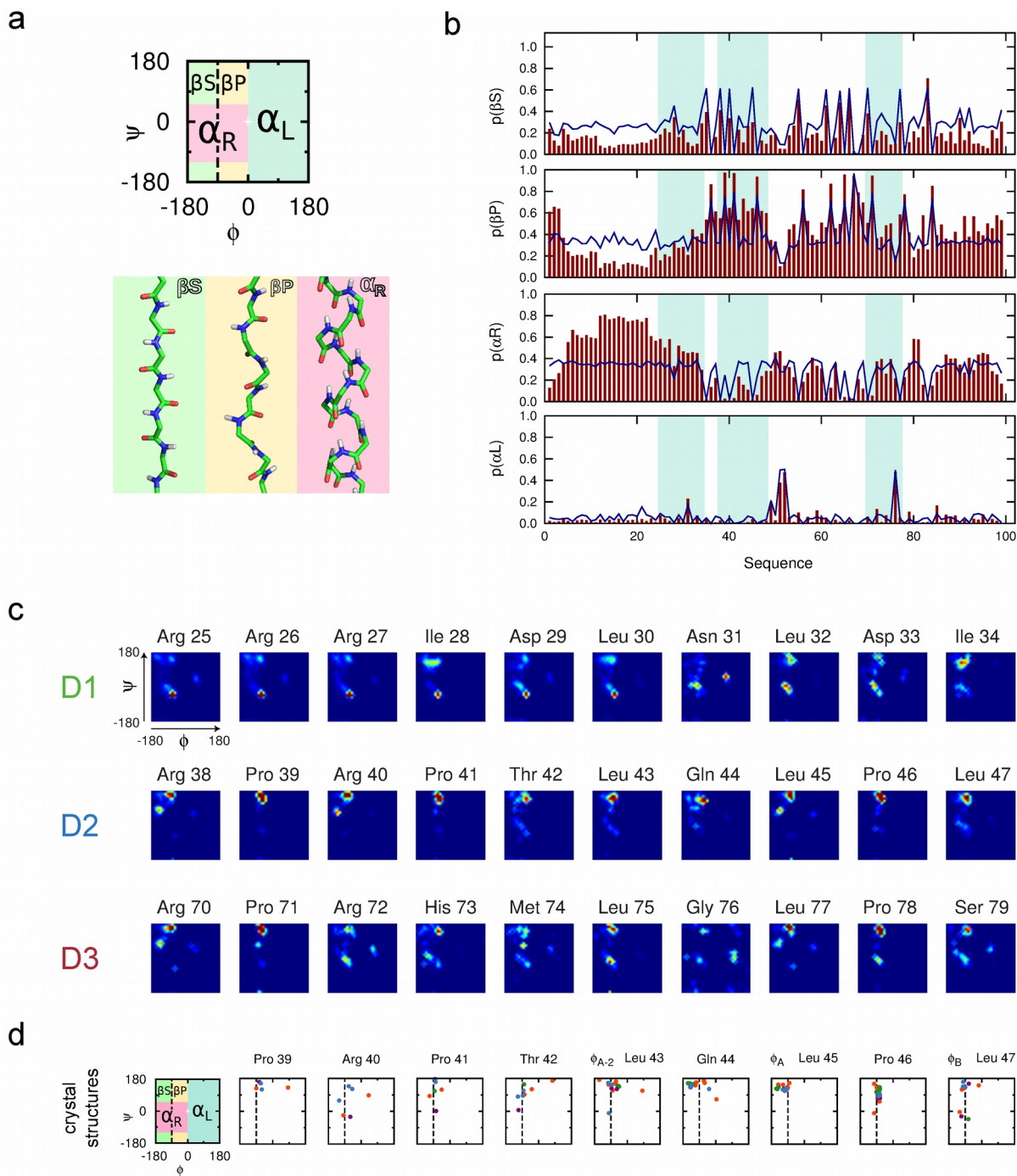


Figure 6.10: Conformational sampling within the docking-sites in the unbound and bound forms. a) Division of Ramachandran space into four regions α_L , α_R , β_S and β_P . b) Per-residue plot of populations belonging to each of the four categories defined in (a) derived from ASTEROIDS ensembles of MKK7_1-100 selected on the basis of experimental chemical shifts and RDCs. Red bars correspond to the sampling derived from the ensembles, while blue lines correspond to populations of the statistical coil library c) Conformational sampling of residues from docking-sites of MKK7_1-100 ensembles. d) Analysis of dihedral angles of docking-site peptides co-crystallized with MAPKs (purple - ERK and p38-bound peptides, red - JNK-bound NFAT4, green - JNK-bound JIP1, blue - JNK-bound D2 (see Results section 6.6))

Another interesting question concerns the described N-terminal α -helix which is predicted to form a coiled-coil as predicted by the COILS algorithm (Figure 6.2b). One possibility is that the coiled-coil is formed through homodimerization of the N-terminal α -helix. The relaxation experiments argue against this possibility, as a homodimeric coiled-coil would have a higher molecular weight and slower tumbling which would give rise to higher R_2 rates and lower R_1 rates. To confirm this hypothesis Multi Angle Laser Light Scattering (MALLS) was used to examine the oligomeric state of MKK7_1-100 (Figure 6.11). The MKK7_1-100 sample at a concentration of 28 mg/ml was injected on a size-exclusion column coupled to an online MALLS system. MKK7_1-100 eluted as a single peak and the obtained molecular weight of 12.1 ± 0.5 kDa corresponds to a monomer (expected mass: 11.5 kDa). The N-terminal α -helix is therefore most probably forming a heterodimeric coiled-coil with an MKK7 interacting protein (see Discussion section).

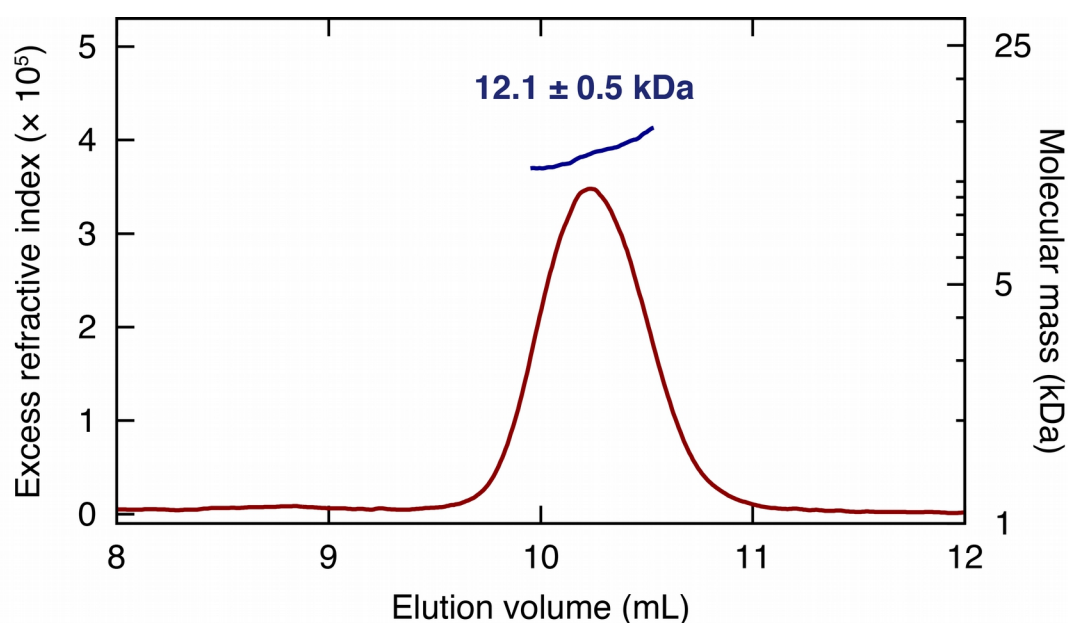


Figure 6.11: MALLS confirms that MKK7_1-100 exists as a monomeric species in solution. The red line shows the SEC elution profile as monitored by refractometry (left axis), while the blue line shows the molecular mass calculated from light scattering and refractometry data across the elution peak (right axis). The data show that the regulatory domain of MKK7 is monomeric in solution (molecular weight calculated from sequence is 11458 Da). The protein was injected into the SEC column at a concentration of 28 mg/mL.

6.2 Characterization of the N-terminal IDR of MKK4

MKK4 is a 399 amino acid long polypeptide chain. Residues 1–73 are predicted to be disordered (Figure 6.12), residues 102–367 fold into a protein kinase domain whose crystal structure was solved previously (Matsumoto et al. 2010, PDB accession code: 3ALN). The C-terminal residues 387–399 are also predicted to be disordered. The canonical docking-site motif spans residues 40–48. Upstream MKKKs bind to MKK4 through the DVD site which is annotated at residues 364–387.

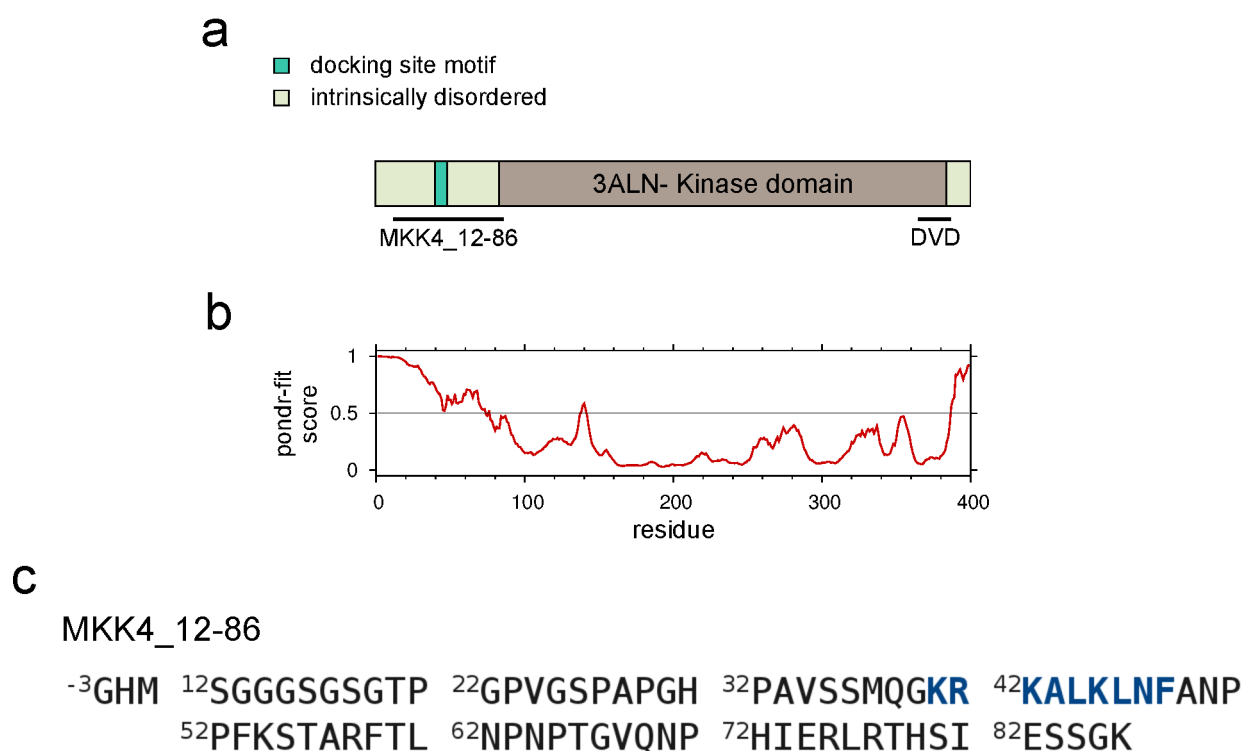


Figure 6.12: Domain organization of MKK4 kinase. a) The structure of catalytic kinase (brown, pdb accession code: 3ALN) domain was solved previously. The location of the DVD site is experimentally confirmed (Ho et al., 2003). b) Disorder prediction for MKK4 obtained by the POND-R-FIT program. A POND-R-FIT score above 0.5 indicates that a region is predicted to be disordered. The first 73 residues are predicted to be disordered. c) Sequence of the MKK4_12-86 construct. N-terminal glycine and histidine are part of the TEV site remaining after the purification tag was removed BY TEV protease. Blue residues mark the putative docking site of MKK4.

To characterize the N-terminal IDR of MKK4 and for the studies of the interaction of the MKK4 docking-site with JNK1, residues 12–86 (MKK4_12-86) were cloned into a pET series vector with an N-terminal Trx-6His-TEV sequence. The prepared construct starts at residue 12 because of the high glycine content in the region 1–22. Lowering the number of glycines in the studied construct

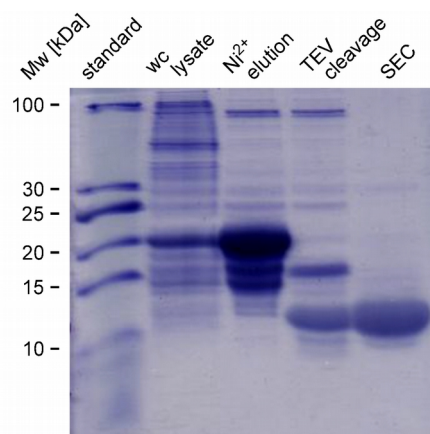


Figure 6.13: SDS-PAGE gel analysis of MKK4_12-86 purification fractions. MKK4_12-86 is slightly degraded during expression as it can be seen by three bands in the "Ni²⁺ elution" fraction, but the intact product can be separated from the other two degradation products using a size-exclusion chromatography (SEC). wc lysate – whole cell lysate.

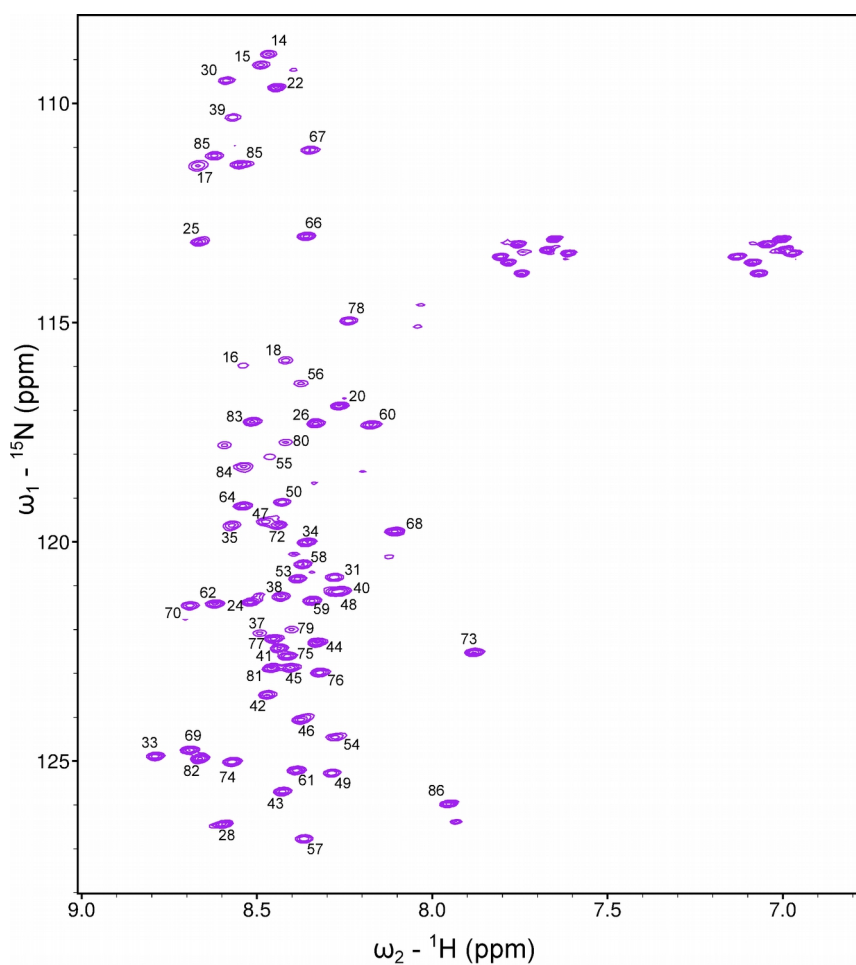


Figure 6.14: Fully assigned ¹H-¹⁵N HSQC spectrum of MKK4_12-86 at 278 K in 50 mM HEPES, 150 mM NaCl at pH 7.0.

has reduced the spectral overlap. The expression of the His-tagged construct was too low for performing NMR studies. Expression was improved to give sufficient amount of protein when MKK4_12-86 was expressed from a pET-28 derived vector with either a Trx-tag or a GB1-tag. The protein was purified using standard chromatographic procedures (Methods, Figure 6.13). The purified MKK4_12-86 construct was assigned using standard triple-resonance assignment experiments (Figure 6.14). HNCO, HN(CO)CA, HN(COCA)CB type experiments were used to obtain inter-residue chemical shifts and intra-residue HN(CA)CO, HNCA, HN(CA)CB type experiments were used to obtain intra-residue chemical shifts.

Analysis of the secondary chemical shifts shows that MKK4_12-86 does not have any regions with pronounced structural propensity (Figure 6.15). The largest deviations, observed for residues His31, His72 and His79, possibly occur because chemical shifts of histidines are most affected by pH changes at the current experimental conditions (pH 7.0). Apart from that, the secondary C α and C β chemical shifts do not show any systematic deviations indicative of cooperative α -helical or β -strand residual structure. The ^{13}C secondary chemical shifts are generally not larger than 0.5 ppm (Figure 6.15). This is a small difference compared to the secondary C α chemical shifts within the α -helix of the MKK7_1-100 construct which show values up to 1.5 ppm. Amide nitrogen ^{15}N secondary chemical shifts, which are sensitive to the presence of α -helices and PPII residual structure, are also quite small with the exception of a region within the docking-site where they are slightly positive values are observed.

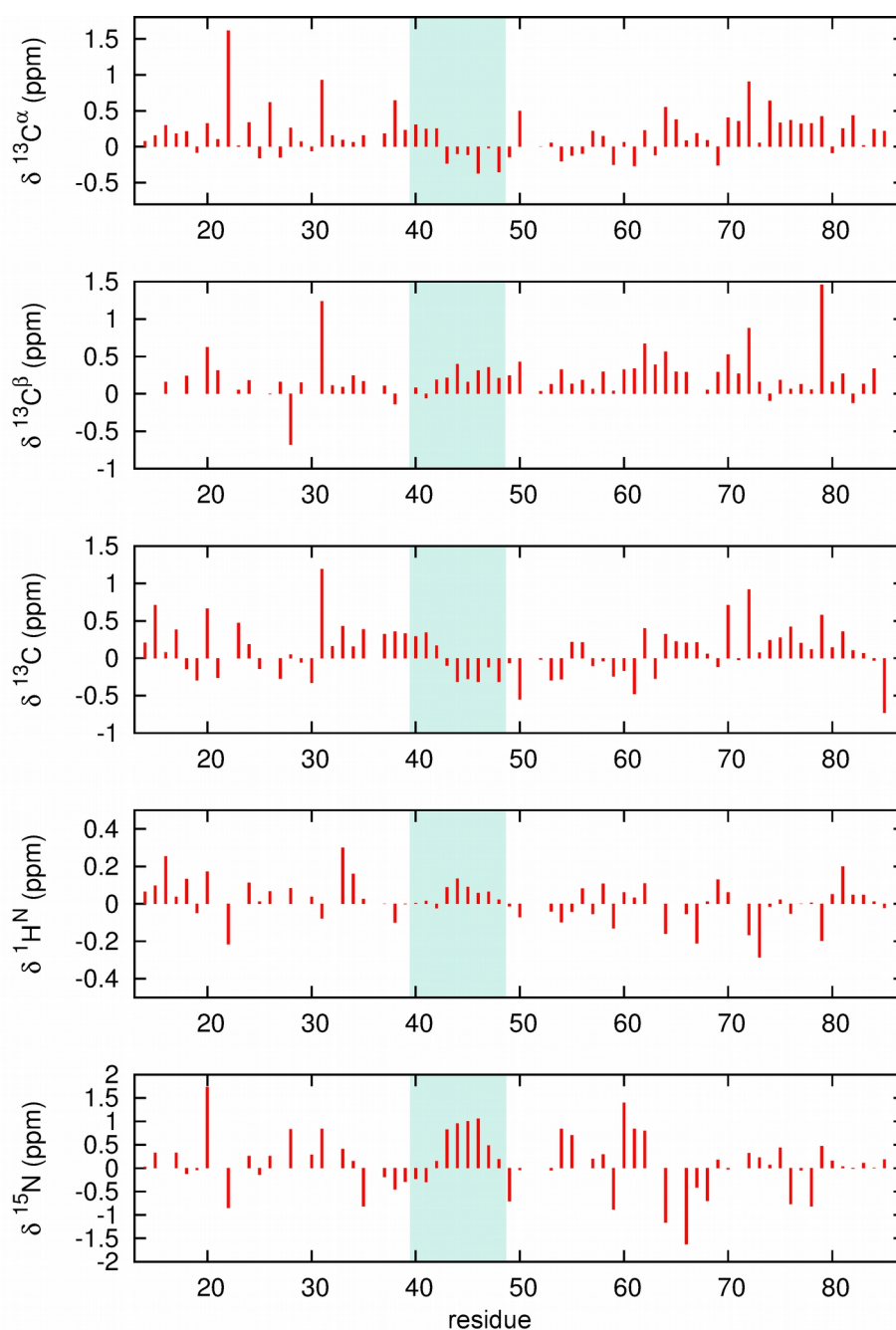


Figure 6.15: Secondary chemical shifts of MKK4_12-86 showing that the protein is disordered and that it does not contain transiently populated secondary structures.

The measured $\{^1\text{H}\}$ - ^{15}N heteronuclear NOEs, longitudinal R_1 and transverse R_2 relaxation rates confirm that MKK4_12-86 is a disordered protein (Figure 6.16). The measured hetNOEs are negative or close to zero for the first 17 residues (uniprot numbering 14–30). This is due to the high content of glycines and serines up to residue 19 and the presence of a few prolines in the following region between residues 20–30.

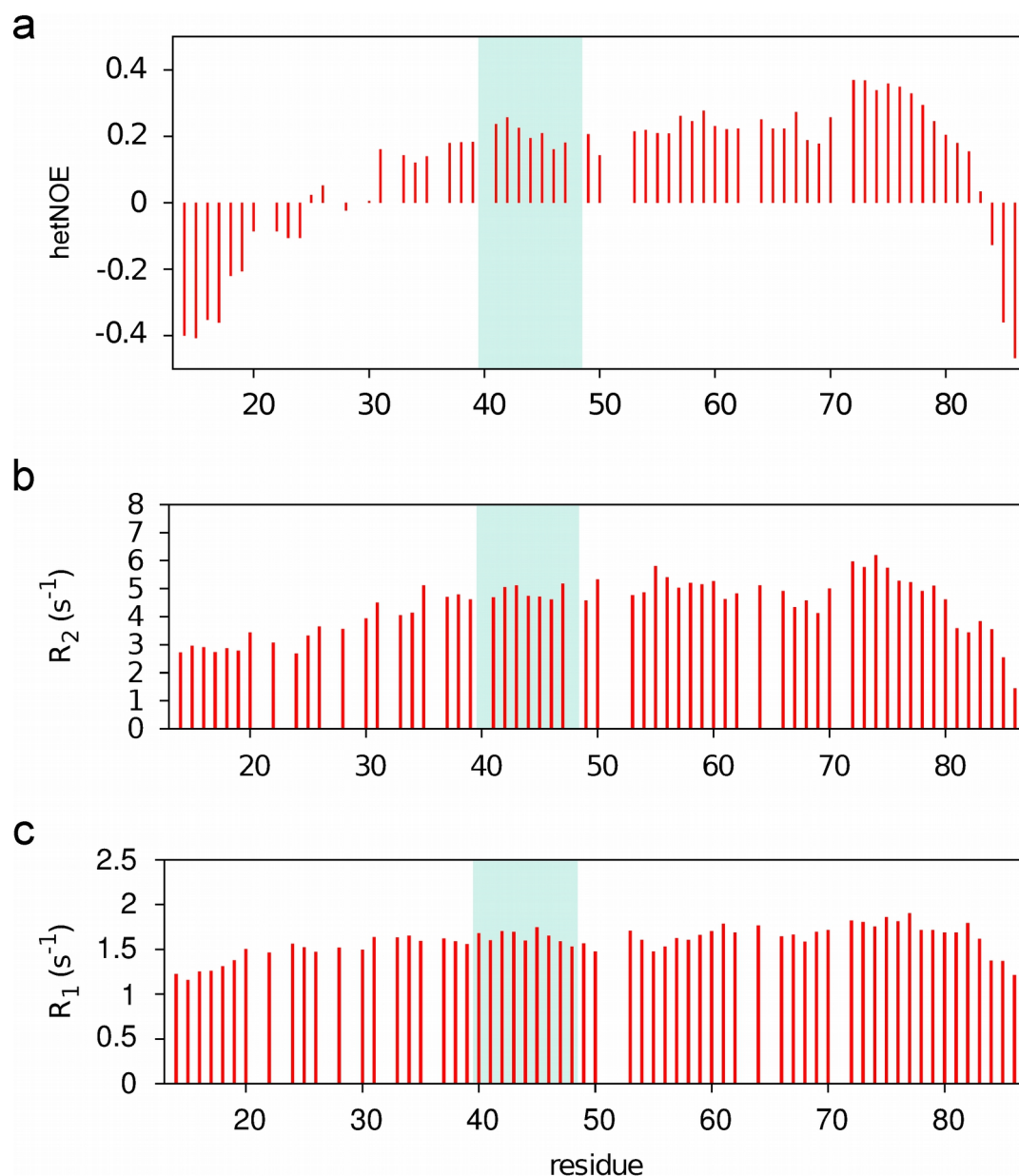


Figure 6.16: $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE, ^{15}N R_1 and R_2 relaxation rates (600 MHz, 5 °C) for MKK4_12-86. a) $\{^1\text{H}\}$ - ^{15}N hetNOE b) ^{15}N R_2 c) ^{15}N R_1 .

As in the case of the IDR of MKK7, five types of RDCs were measured to characterize the MKK4_12-86 further. In general, the RDCs do not show large deviations from a Flexible-Meccano statistical coil model (Figure 6.17). This is in agreement with the secondary chemical shifts. Nevertheless small deviations do exist in a few regions. Especially the $^1\text{D}_{\text{CaH}\alpha}$ and $^4\text{D}_{\text{HaHN}}$ RDCs, are not predicted well in the docking-site and for residues between 60 and 70. Interestingly, RDCs and chemical shifts agree that there is a deviation from random/statistical coil models in the region of the docking-site.

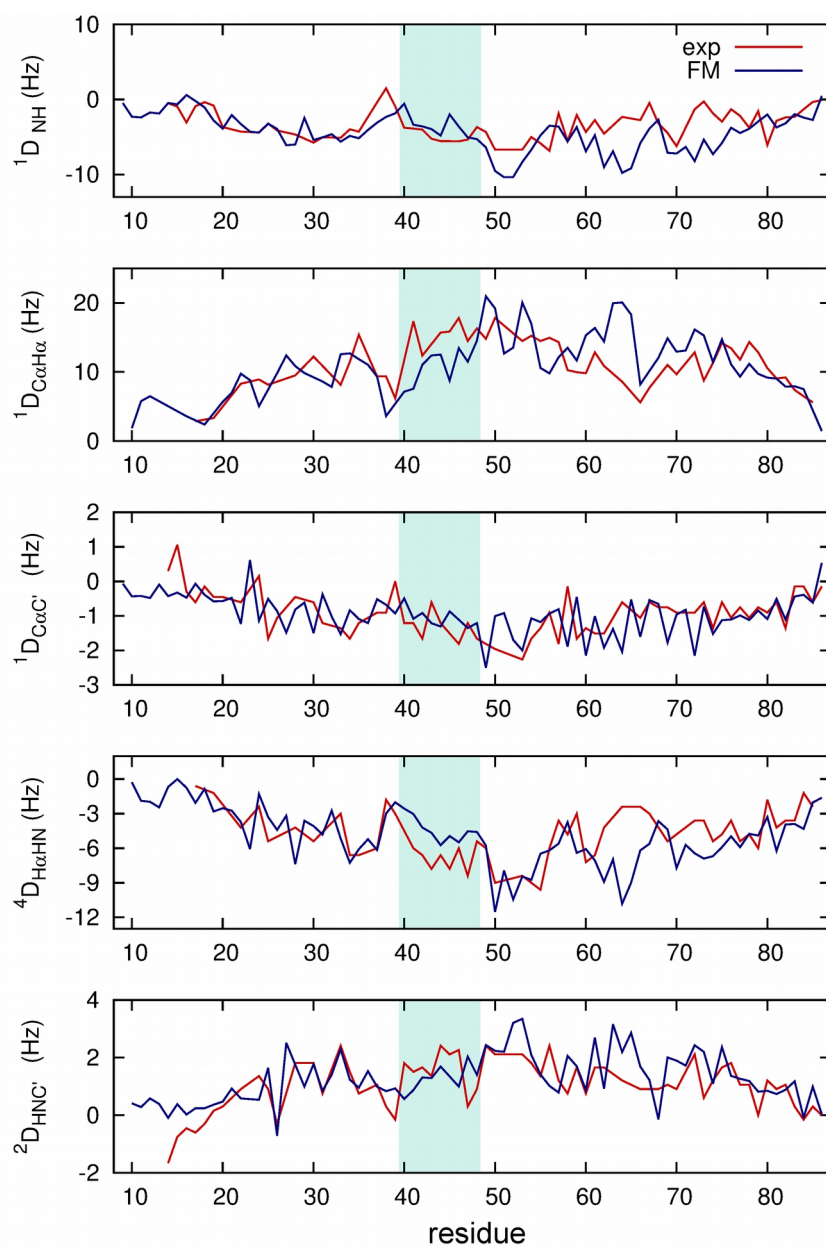


Figure 6.17: $^1D_{NH}$, $^1D_{C\alpha H\alpha}$, $^1D_{C\alpha C'}$, $^4D_{H\alpha HN}$ and $^2D_{HNC'}$ RDCs measured for MKK4_12-86 aligned in a PEG-octanol liquid crystal at 5 °C. The residual 2H quadrupole splitting was 32 Hz. The RDCs are compared to a prediction of RDCs using a statistical coil model (blue, Flexible-Meccano (FM)).

Chemical shifts and RDCs can be used in an ensemble selection to determine the sampling of MKK4_12-86. Ensembles comprising 200 conformers each were selected from a large pool of structures containing. All available chemical shifts ($^{13}C^\alpha$, $^{13}C^\beta$, $^{13}C'$, ^{15}N , $^1H^N$) and $^1D_{NH}$, $^1D_{C\alpha H\alpha}$ and $^4D_{H\alpha HN}$ RDCs were used in the selection procedure giving excellent agreement with values calculated from the selected ensembles (Figure 6.18a,b). Furthermore, a cross-validation can be

6. Results

made by comparing the back-calculated $^2D_{\text{HNC}}$ RDCs which were not fitted during the selection procedure with the experimentally measured values. If the selected ensemble correctly reproduces the sampling of MKK4_12-86, the agreement between the experimental RDCs and the back-calculated RDCs from the selected ensemble should be good and without any systematic deviations. The agreement between experimental and back-calculated RDCs is in general good (Figure 6.18c).

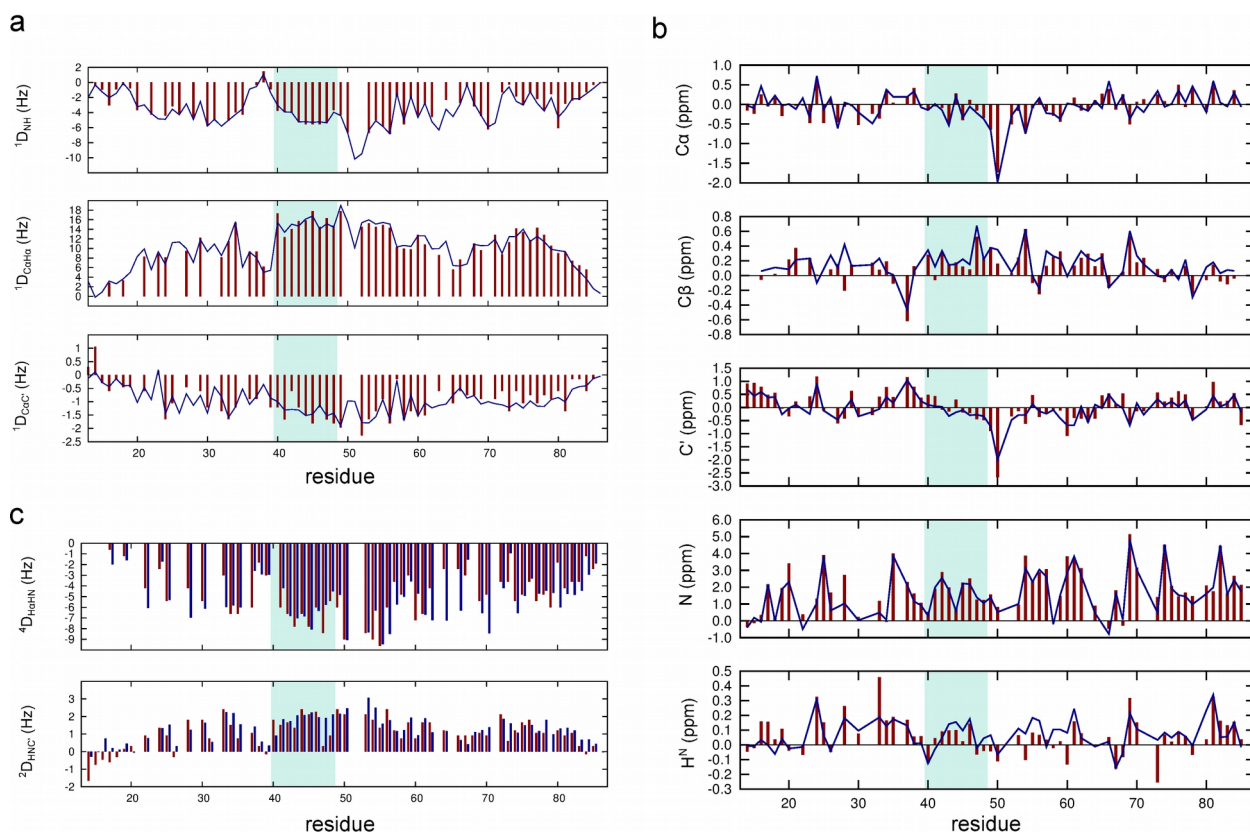


Figure 6.18: Comparison of experimental RDCs (red) with values back-calculated (blue) from a selected ASTEROIDS ensemble comprising 200 conformers. (b) Comparison of experimental chemical shifts (red) with values back-calculated (blue) from a selected ASTEROIDS ensemble comprising 200 conformers. (c) Comparison of $^4D_{\text{H}\alpha\text{HN}}$ and $^2D_{\text{HNC}}$ RDCs backcalculated from the final ensemble with the experimental RDCs that were not fitted in the selection.

The residue specific conformational sampling in the selected MKK4_12-86 ensembles can be described by dividing the Ramachandran space into four regions termed β -sheet (βS), PPII (βP), α -helical (αR) and left-handed helix (αL) (Figure 6.19a). The populations in each of these regions can then be plotted separately as a function of sequence (Figure 6.19b). Compared to the Flexible-Meccano model the selected ensembles sample β -sheet conformations less frequently and PPII conformations are overall slightly enhanced, for some residues more than for the others. The docking-site region shows a consistently higher sampling of PPII conformations with the exception

of the last two residues.

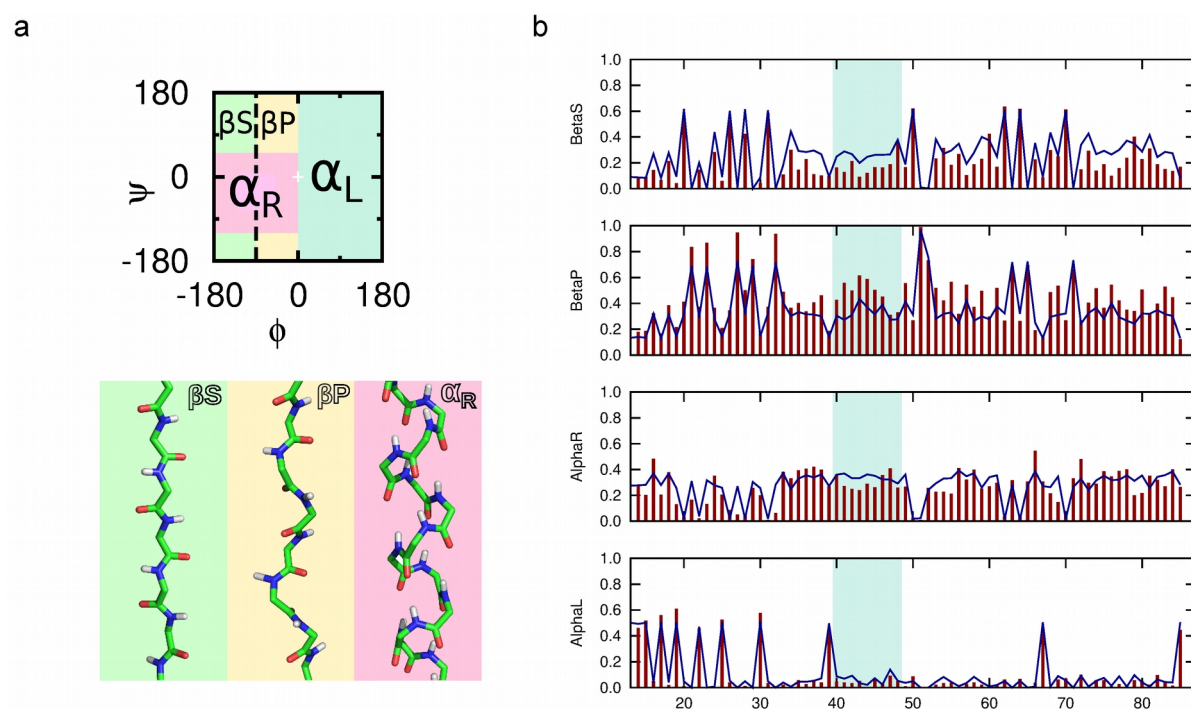


Figure 6.19: Conformational sampling within the docking-sites of MKK4_12-86. a) Division of Ramachandran space into for α_L , α_R , β_S and β_P regions. b) Per-residue plot of populations belonging to each of the four regions defined in (a) derived from ASTEROIDS ensembles of MKK4_12-86 selected on the basis of experimental chemical shifts and RDCs. Red bars correspond to the sampling derived from the ensembles while blue lines correspond to populations of the statistical coil library.

6.3 Interaction of MKK4 and MKK7 with JNK1

6.3.1 Stability of JNK1

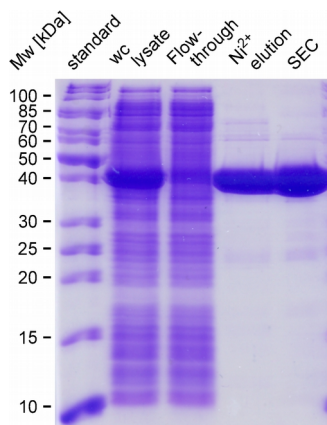


Figure 6.20: SDS-PAGE gel analysis of JNK1 purification fractions. wc lysate – whole cell lysate.

In order to study interactions between the MKKs and JNK1, the sequence corresponding to residues 1-365 of JNK1 was cloned into a pET-41c vector with a C-terminal His-tag. JNK1 was purified using standard chromatographic procedures with a nickel affinity column followed by size-exclusion chromatography (Methods 5.1.3). The solubility of JNK was assessed by concentration in centrifugal concentrators under different buffer conditions. JNK1 appeared to be more stable at higher pH, higher concentration of NaCl and higher glycerol concentration (Table 6.2).

Thermal shift assay which is used to assess the melting temperature and with that the stability of a protein was performed with JNK1 to screen for different stabilizing factors and assess their efficiency (Methods 5.3.3, Table 6.3). High concentrations of NaCl and glycerol were identified as efficient stabilizing factors, while the D2 peptide of MKK7 and ATP with $MgSO_4$ raise the melting temperature of JNK1 significantly. Apart from PEG 400 and 1 mM Glu/Arg buffer which also stabilized JNK1, other additives from the screen (Boivin et al., 2013) did not have any stronger influence on the stability.

Table 6.2: Dependence of JNK1 solubility on buffer conditions

pH = 6.5	5% glycerol	50 mM Glu + 50 mM Arg	no additive
150 mM NaCl	1.6 mg/ml	1.3 mg/ml	0.66 mg/ml
300 - 500 mM NaCl	4.0 - 5.0 mg/ml	2.3 mg/ml	
pH = 7.0	5% glycerol		
150 mM NaCl, 5% Glycerol	2.6 mg/ml		
pH = 8.0	5% Glycerol		
150 mM NaCl	4.0 mg/ml		

Table 6.3: Conditions that improve the T_m of JNK in thermal shift assay

Reference conditions (T_A)	Additive (T_B)	T_A [°C]	T_B [°C]	ΔT [°C]*
AMP-PNP, 5% glycerol	10% glycerol	41	44	3
AMP-PNP, 5% glycerol	20% glycerol	41	45	4
AMP-PNP, 5% glycerol	500 mM NaCl	41	44	3
AMP-PNP, 5% glycerol	20% glycerol + 500 mM NaCl	41	44	3
AMP-PNP, 5% glycerol	D2 peptide	41	48	7
AMP-PNP, D2 peptide, 5% glycerol	20% glycerol + 500 mM NaCl	48	49	1
Apo	5 mM MgSO ₄	34	34	0
Apo	2 mM ATP, 5 mM MgSO ₄	34	44	10
Apo	5% glycerol	34	35	1
Apo	10% glycerol	34	37	3
Apo	15% glycerol	34	39	5
Apo	PEG 400	34	43	9
Apo	50 mM Glu + 50 mM Arg	34	34	0
Apo	500 mM Glu + 500 mM Arg	34	41	7

*Significant changes (>2 °C) are marked with grey background colour.

6.3.2 NMR titration of MKK7_1-100 with JNK1

The interaction between ¹⁵N labelled MKK7_1-100 and JNK1 was studied by NMR. A series of ¹H-¹⁵N HSQC spectra was recorded at different MKK7_1-100 to JNK1 ratios. Upon addition of JNK1 to MKK7_1-100 only small chemical shift changes occurred (Figure 6.21a, 6.22). On the other hand, peak intensities decreased significantly with higher JNK/MKK7_1-100 ratios (Figure 2.1b). The peaks disappear because a large molecular weight complex is formed upon JNK1 binding (~55 kDa) where the higher relaxation rates cause the peaks to broaden beyond detection. In addition, exchange between free and bound forms occurring on the μ s-ms time scale could be another source of the observed line broadening. Only small chemical shift perturbations occur probably because the exchange between the free and bound form is slow to intermediate on the chemical shift time scale. The titration data were analysed by plotting the peak intensities normalized by the intensity of the same peak when MKK7_1-100 is unbound (0%) revealing which resonances experience the strongest decrease in intensity (Figure 6.21b). The D2 and D3 docking-sites show the clearest signs of binding, while the D1 site exhibits a much lower decrease in peak intensity for which there could be two non-exclusive explanations. The simpler explanation is that the decrease in intensity is

6. Results

observed because JNK1 is binding to that region. The other explanation is that the binding of JNK1 to D2 affects the motion of the α -helix due to a longer persistence length of the α -helix. Therefore it cannot be concluded whether D1 binds to JNK1 because of the different contributing factors to the I/I_0 profile. The most affected peaks around the D3 site seem to extend past the canonical D3 motif. Especially residues Thr80-Leu81-Phe82, which previously were not annotated as a part of D3, seem to be strongly affected by the binding.

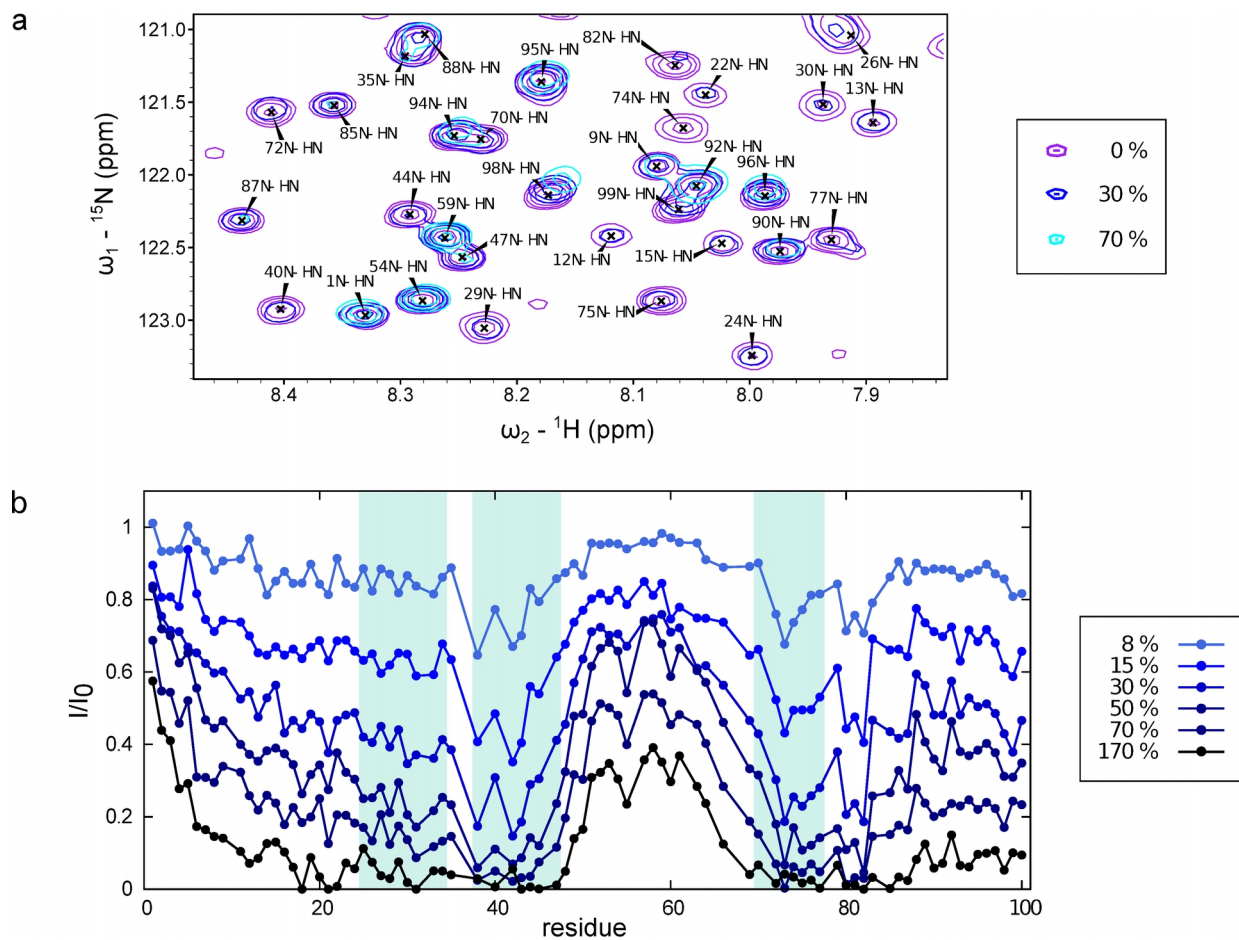


Figure 6.21: Interaction of MKK7_1-100 with JNK1 monitored by NMR. a) A selected region of the ^1H - ^{15}N HSQC spectrum of MKK7_1-100 in the absence and presence of different ratios of JNK1. b) Intensity profile for the resonances of MKK7_1-100 at different ratios of JNK1.

We can further learn about the binding kinetics of the docking-sites to JNK1 by noticing that the peaks corresponding to residues of the docking-sites disappear already at sub-stoichiometric ratios of JNK compared to MKK7 (0.7:1). In the case where each docking-site binds a JNK1 molecule the ratio at which saturation is reached would be 3:1 meaning that the intensities in the D2 and D3 decrease approximately four times faster than expected. This indicates, that rather than being in

slow exchange, the binding occurs on the intermediate-to-slow chemical shift time scale in the range of tens to hundreds of milliseconds.

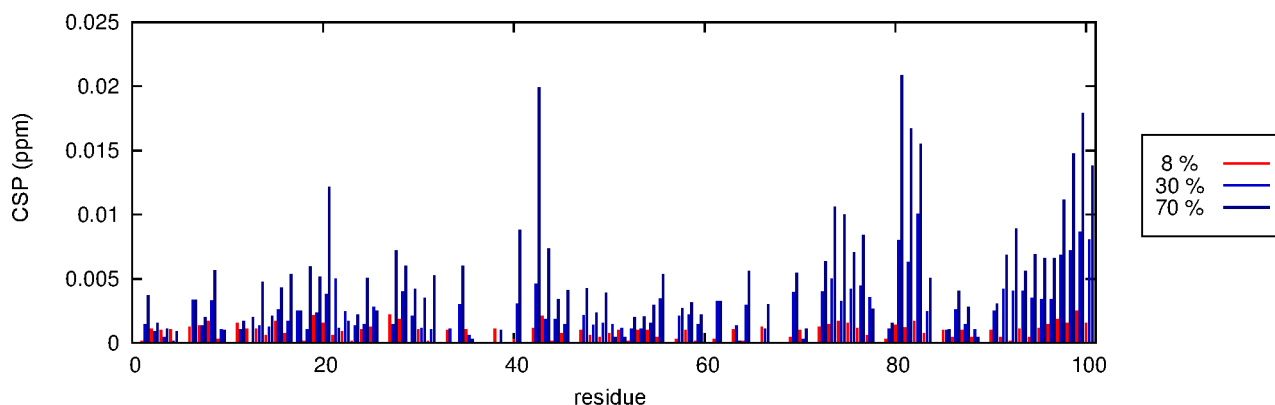


Figure 6.22: Chemical shift perturbations (combined ^1H and ^{15}N perturbations) of MKK7_1-100 in the presence of different ratios of JNK1 compared to MKK7.

An analysis of chemical shift changes in terms of combined ^1H and ^{15}N chemical shift perturbations shows that they mostly occur for the residues of D3 and at the C-terminus of MKK7_1-100 indicating that these two regions bind with faster exchange rates (Figure 6.22). The C-terminus of MKK7_1-100 contains a number of hydrophobic residues that may contact the surface of JNK1, when JNK1 is bound to D3.

Transverse R_2 relaxation rates measured at selected titration points show modest increases in relaxation rates in the N-terminal α -helix, as well as the D2 and D3 docking-sites (Figure 6.23). On the contrary, in the glycine-serine-proline rich region between the D2 and D3 docking-sites the relaxation rates stay almost the same compared to when JNK1 is not bound (Figure 6.23). The R_2 relaxation rates can also inform us about the kinetics of complex formation. In the fast exchange regime the R_2 relaxation rates would be a population-weighted average between the relaxation rates in the complex (high R_2) and the R_2 relaxation rates of the unbound form. This is probably not our case as the increase in the R_2 relaxation rates during the course of the titration is not so big. In the slow exchange regime a measurement of the R_2 relaxation rates of the peaks of the unbound form would inform us only about the sum of the R_2 relaxation rates of the unbound MKK7_1-100 and the on-rate for the complex formation. This is probably not the case as non-uniform R_2 values are measured for residues within the docking sites for sub-stoichiometric amounts of JNK (Figure 6.23). The observed increase in the R_2 relaxation rates is therefore most probably an exchange contribution caused by the binding occurring on the slow-to-intermediate time scale.

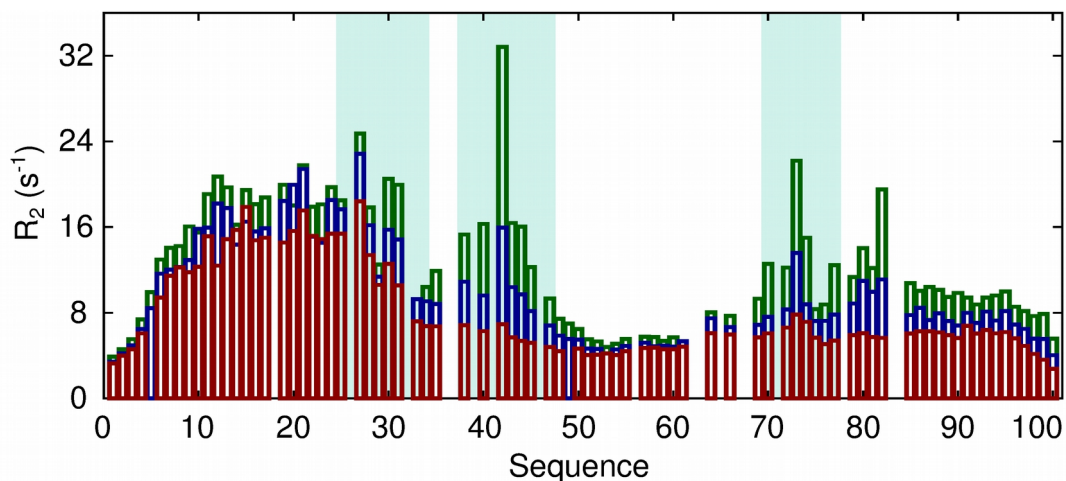


Figure 6.23: R_2 relaxation rates (5 °C) for MKK7_1-100 in complex with JNK1. The R_2 rates for 14% saturated MKK7_1-100 (blue) and 28% saturated MKK7_1-100 (green) increase compared to the R_2 rates for unbound MKK7_1-100 (red) in particular within the docking-sites.

6.3.3 NMR titration of MKK4_12-86 with JNK1

In order to monitor the interaction between MKK4_12-86 and JNK1, a chemical shift titration was carried out. A series of ^1H - ^{15}N HSQC spectra was recorded at different MKK4_12-86 to JNK1 ratios. Only small chemical shift changes were observed during the titration similarly to the MKK7_1-100 titrations (Figure 6.24a and 6.25). The biggest changes occur in form of intensity loss of certain peaks principally located in the docking-site motif and for residues in the immediate vicinity (38 - 53), however interestingly there are other affected residues outside the canonical motif (Figure 6.24b). The two short regions around residue 60 ($^{59}\text{FTL}^{61}$) and around residue 75 ($^{73}\text{IERLR}^{77}$), that have a hydrophobic character, show clear decreases in intensities for the last two titration points. According to disorder prediction the region around residue 75 is already a part of the catalytic folded domain in the full-length protein and thus this region may not interact in the full-length protein.

In comparison to the titration of MKK7_1-100, the intensities of the resonances of MKK4_12-86 decrease gradually and at near stoichiometric ratios (0.9:1) the peaks corresponding to the residues of the docking-site are still visible. This indicates an exchange between free and bound form on a rather slow time scale.

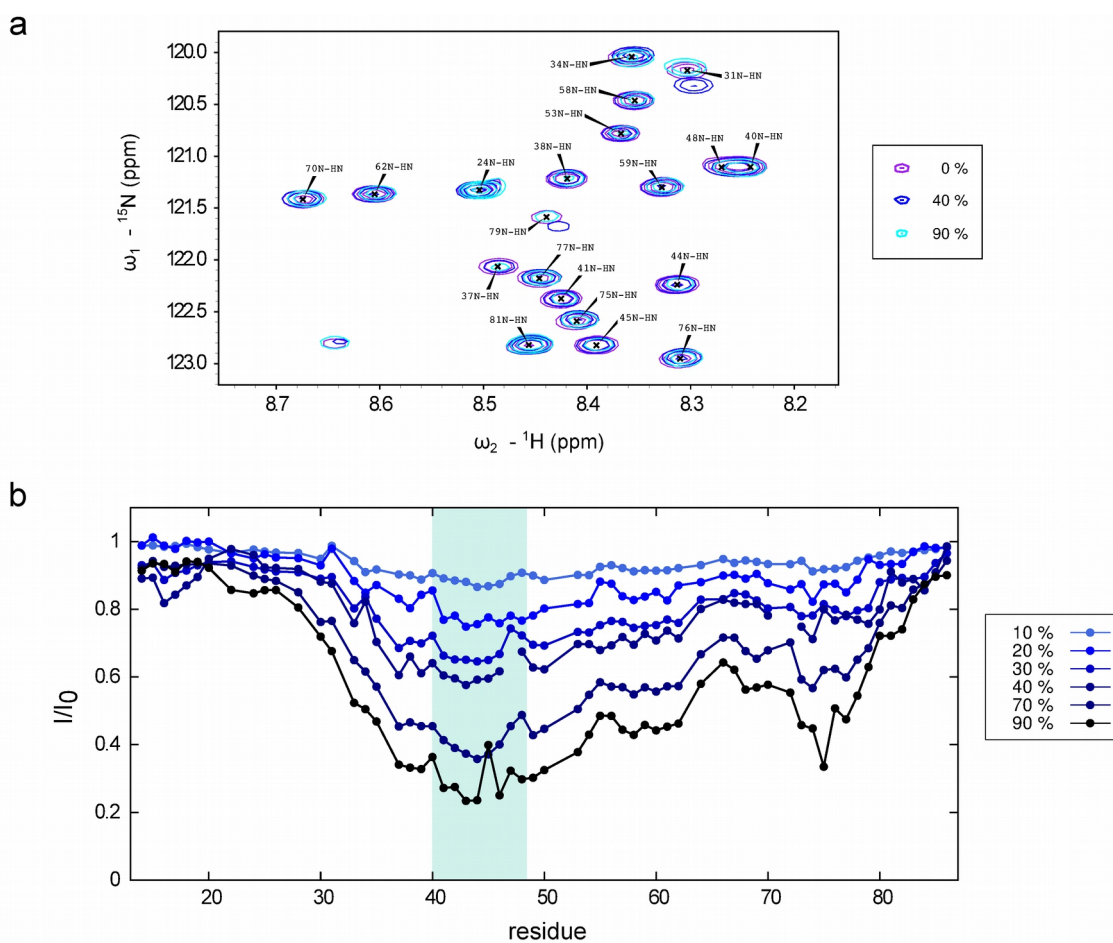


Figure 6.24: Interaction of MKK4_12-86 with JNK1 monitored by NMR. a) A selected region of ^1H - ^{15}N HSQC spectrum of the MKK4_12-86 in the absence and presence of JNK1. b) Intensity profile for the resonances of MKK4_12-86 at different ratios of JNK1 compared to unbound MKK4_12-86.

An analysis of chemical shift changes as combined ^1H and ^{15}N chemical shift perturbations shows that there are no significant chemical shift changes upon addition of JNK1 (Figure 6.25). This is in agreement with the notion made above, that the binding occurs on a slow time scale.

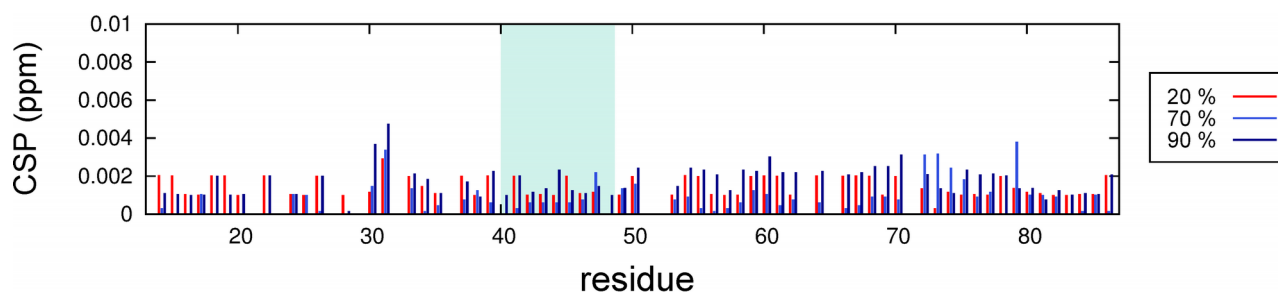


Figure 6.25: Chemical shift perturbations for MKK4_12-86 in the presence of different JNK1 to MKK4_12-86 molar ratios.

6.4 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) was used to determine the affinities and stoichiometry of binding of MKK7 and MKK4 to JNK1. Constructs of MKK7_1-100 with a Trx-6His-TEV tag on the N-terminus containing one, two and three docking motifs were prepared (Figure 6.26a, 6.26b). MKK7_1-100 does not contain tryptophanes or tyrosines which contribute to the absorbance at 280 nm. Trx tag has an extinction coefficient of $15595 \text{ M}^{-1} \text{ cm}^{-1}$ and was not removed during the purification process, enabling us to determine precise protein concentrations. A sample containing only Trx-6His-TEV was used as a negative control to confirm that the Trx tag is not binding to JNK1. For each single site a synthetic peptide was purchased (Figure 6.26c). The D1 peptide was insoluble, preventing further experiments, and the peptide corresponding to the MKK4 docking-site sequence did not bind with high affinity to JNK1, probably due to the lack of a third hydrophobic residue (Phe48) within the peptide sequence.

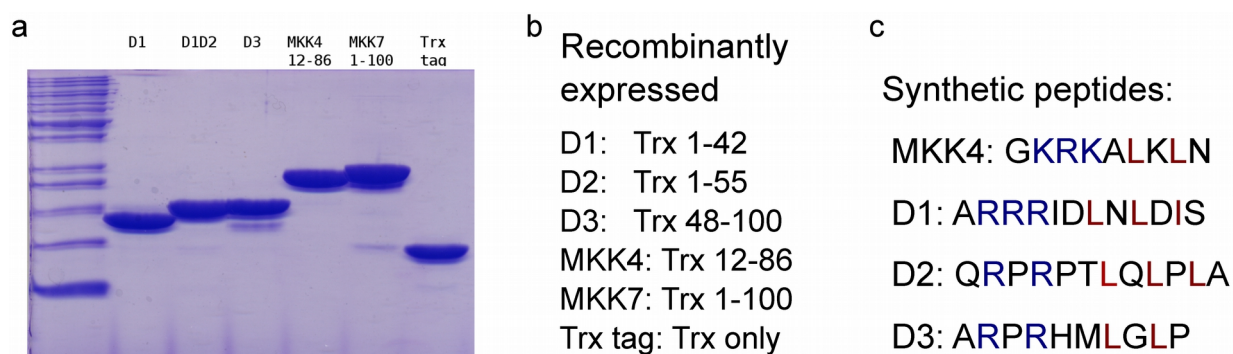









Figure 6.26: Samples used for ITC measurements. a) SDS-PAGE of purified Trx-constructs used for ITC experiments. b) Legend explaining the abbreviations above the gel in (a). c) Sequences of the synthetic peptides.

The affinities for all the Trx constructs and the peptides obtained by ITC are in the low micromolar range (Figure 6.27). D2 docking motif displays the highest affinity of all motifs ($7.9 \pm 1.3 \mu\text{M}$) and is followed by D3 and D1 which display only slightly lower affinity ($11 \pm 1 \mu\text{M}$ and $12 \pm 5 \mu\text{M}$, respectively). Importantly, as it was unclear from chemical shift titrations of MKK7_1-100 with JNK whether D1 is binding, the ITC measurements confirmed the binding of D1 to JNK. The peptide containing the canonical D3 docking motif sequence does not bind as strongly as the Trx construct containing the D3 docking-site ($49 \pm 12 \mu\text{M}$), indicating that residues outside the canonical site contribute significantly to the affinity. The three residues (Thr80-Leu81-Phe82)

outside the canonical D3 motif that were identified using chemical shift titrations might be the ones contributing to the higher affinity of the Trx-construct compared to the D3 peptide as the synthetic

a

Ligand	Sites	Length	N [sites]	K _d [μ M]	Δ H [kcal mol ⁻¹]	
Trx tagged	D-site 1	1 	42	0.95 \pm 0.05	12 \pm 5	-3400 \pm 400
Peptide	D-site 2	37 	48	0.68 \pm 0.05	7.9 \pm 1.3	-7500 \pm 3300
Trx tagged	D-site 3	48 	100	0.89 \pm 0.18	11 \pm 1	-2700 \pm 400
Peptide	D-site 3	69 	78	1.25 \pm 0.49	49 \pm 12	-1600 \pm 100
Trx tagged	D-site 1, 2	1 	55	0.54 \pm 0.01	4.5 \pm 0.4	-7100 \pm 300
Trx tagged	D-site 1, 2, 3	1 	100	0.31 \pm 0.01	4.4 \pm 0.3	-15000 \pm 20
Trx tagged	D-site MKK4	1 	86	0.989	1.333	-9974

b

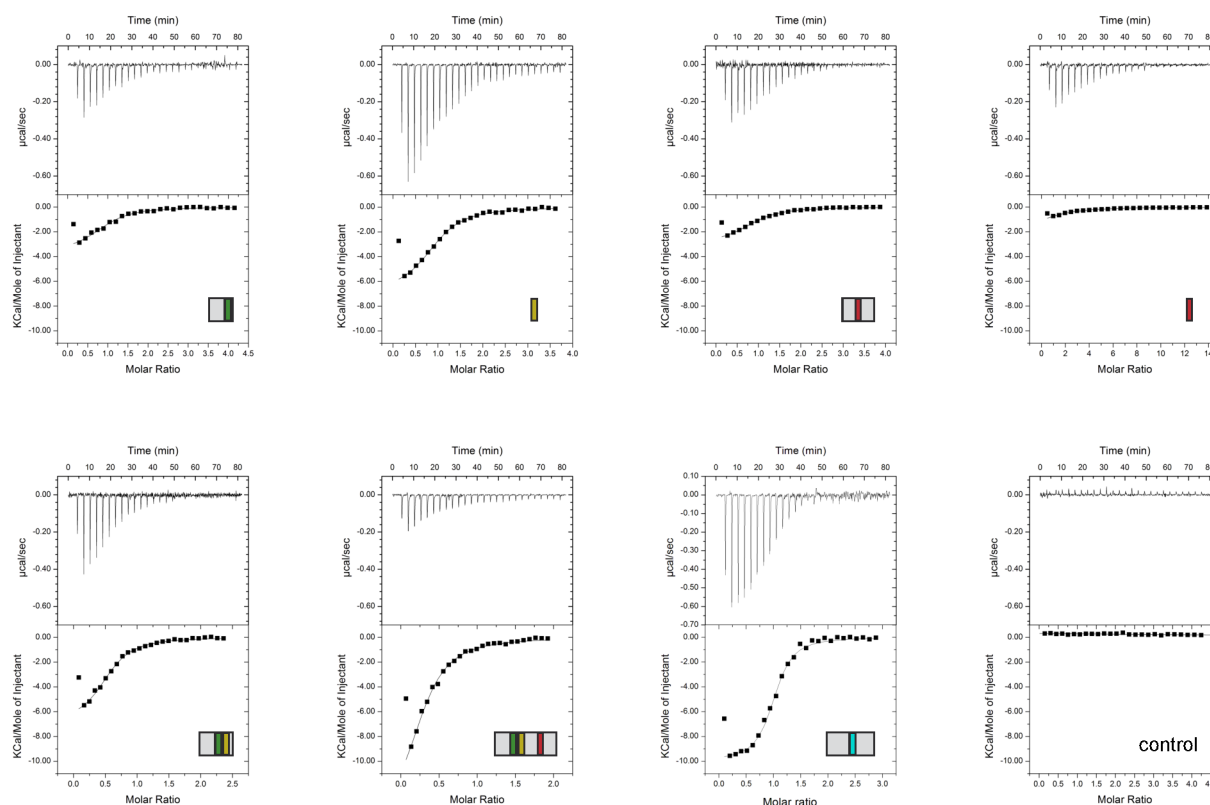


Figure 6.27: Affinity and thermodynamic measurements of MKK7 and MKK4 binding to JNK1 by ITC. a) Summary table for different Trx-constructs and peptides with the averaged parameters for two or three independent experiments (except for MKK4_1-86, for which only one experiment was performed). b) Representative ITC data are shown with raw injection heats (top) and the corresponding specific binding isotherms (bottom) obtained for each Trx-construct and peptide.

D3 peptide does not include those residues. The docking motif of MKK4 binds JNK1 with higher affinity than any of the MKK7 binding sites (1.3 μM). The stoichiometry obtained with the construct containing both the D1 and D2 site is very close to 0.5 (or 1:2) and the stoichiometry for the construct containing D1, D2 and D3 is 0.3 (or 1:3). The N-terminal IDR of MKK7 is therefore able to bind three molecules of JNK1. In agreement with this the enthalpy change is increasingly negative for constructs with more docking motifs (- 3400 kcal mol⁻¹ for Trx-D1, -2700 kcal mol⁻¹ for Trx-D3, -7100 kcal mol⁻¹ for Trx-D1D2 and -15000 kcal mol⁻¹ for Trx-MKK7_1-100).

6.5 NMR titration of the full-length MKK7 kinase with JNK1

To investigate whether the N-terminal IDR of MKK7 is disordered in the context of full-length MKK7 and to know whether all the three docking motifs of MKK7 are accessible within in the full-length protein, full-length ¹⁵N labelled MKK7 was studied by NMR. The ¹H-¹⁵N HSQC spectrum of the full-length MKK7 kinase superimposes well on the ¹H-¹⁵N HSQC spectrum of MKK7_1-100 (Figure 6.28a). Resonances corresponding to residues of the catalytic kinase domain cannot be observed in the spectrum because the slow correlational time of the 43 kDa folded domain causes the peaks to broaden beyond detection. On the other hand, the peaks of the N-terminal IDR can be observed, because this region stays disordered and flexible in the context of full-length MKK7. Residues in the region 85–100 are also missing from the full-length MKK7 ¹H-¹⁵N HSQC as they are very close to the catalytic domain (Figure 6.28a). A plot of full-length MKK7 peak intensities shows how the intensities slowly decrease towards the catalytic domain (Figure 6.28b). Combined hydrogen-nitrogen chemical shift perturbations show that some residues whose peaks are still visible could be transiently interacting with the catalytic domain in a fast exchange regime (Figure 6.28c). Because the chemical shifts of MKK7_1-100 and full-length MKK7 are very similar we can conclude that their conformational behaviour is the same, which means that the conformational sampling obtained by ensemble selections with the MKK7_1-100 data is also valid for full-length MKK7. Finally, titration with JNK1 at ratios 1:0.2 and 1:0.4 demonstrates that all three sites bind to JNK1 as the intensity profile is very similar to the one obtained in the MKK7_1-100 titration with JNK1 (Figure 6.28d).

As the conformational behaviour of MKK7_1-100 and full-length MKK7 is similar, the conformational sampling obtained with the ensemble selections can be used to build the full-length kinase model (Figure 6.29). The model allows us to represent the volume space accessible to the docking-sites. As no long-range distance information (paramagnetic relaxation enhancements, small

angle X-ray scattering) have been collected, the generated model assumes that there are no weak long-range interactions between the IDR and the catalytic domain.

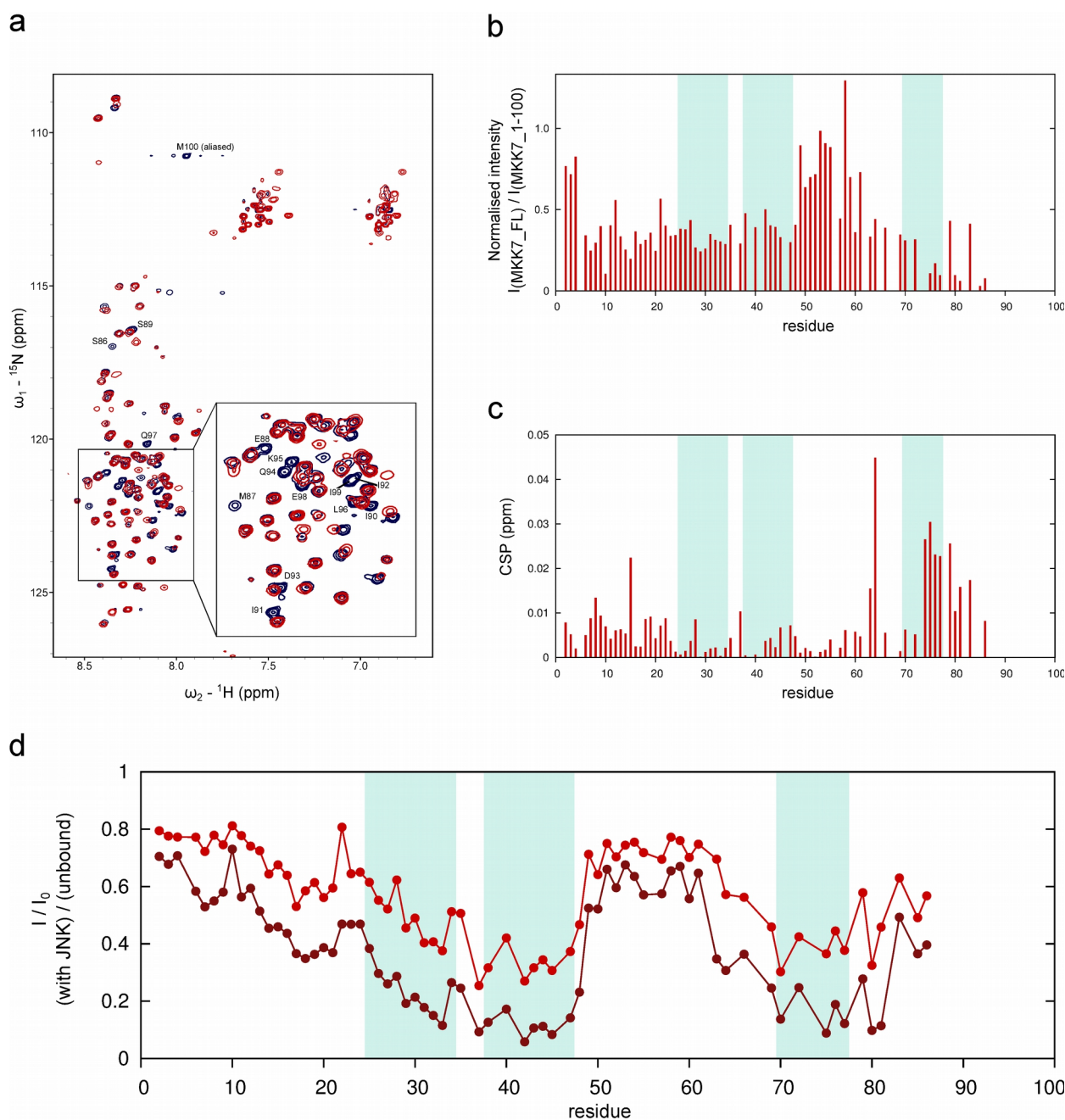


Figure 6.28: Comparison of MKK7_1-100 and full-length MKK7. a) Superposition of ^1H - ^{15}N HSQC spectra of MKK7_1-100 (black) and full-length MKK7 (red). Peaks of residues that are close to the folded catalytic domain are marked with one-letter amino acid code and residue number. b) Plot of peak intensities in the HSQC spectrum of full-length MKK7 normalized by the peak intensities of MKK7_1-100. c) CSP for residues of full-length MKK7 in comparison to MKK7_1-100. d) NMR titration of full-length MKK7 with JNK1 at two different ratios (light red - 20%, dark red - 40%). Intensities of the peaks from a titration point were normalized by the peak intensities of the unbound full-length MKK7.

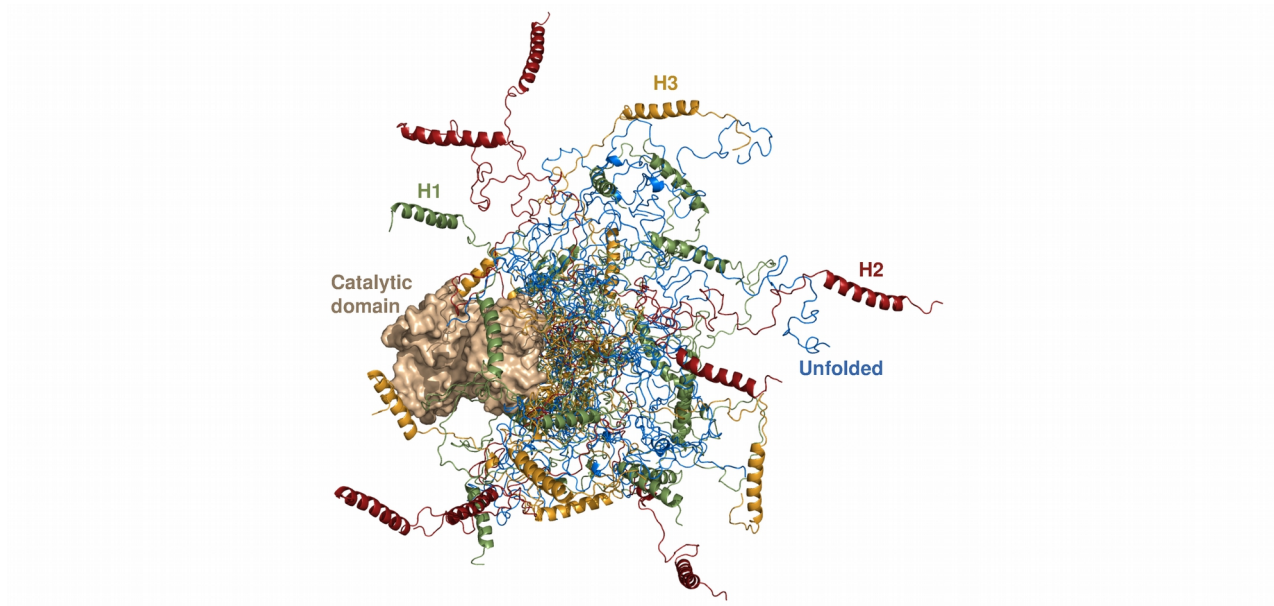


Figure 6.29: Model of the full-length MKK7. The kinase domain is shown as a surface representation (beige), while the disordered regulatory domain is represented in four different colours (blue: unfolded, green: helix H1, red: helix H2, yellow: helix H3) corresponding to the conformational equilibrium presented in Figure 6.8.

6.6 Crystal structure of the D2 docking-site motif of MKK7 bound to JNK1

To obtain information about the structural basis of the interaction between JNK1 and MKK7, crystallization was attempted using a ternary complex of JNK1, the ATP analog AMP-PNP and peptides corresponding to the three docking motifs of MKK7. The first and third docking motifs of MKK7 were produced as Trx fusion constructs, while the peptide of the D2 site was synthetic. Of the three only the JNK1–AMP-PNP–D2 complex yielded crystals. The structure of D2 bound to JNK1 was determined to 2.4 Å resolution (PDB number 4UX9). The refinement statistics are shown in Table 6.4.

Table 6.4: Data collection and refinement statistics for the complex of JNK1 and the D2 peptide.

	4UX9
Wavelength (Å)	0.9
Resolution range (Å)	49.21 – 2.34 (2.424 – 2.340)
Space group	C 1 2 1
Unit cell	
a, b, c (Å)	108.666, 180.159, 101.144
α , β , γ (°)	90.0, 110.3, 90.0
Total reflections	259709 (20637)
Unique reflections	75430 (6959)
Multiplicity	3.4 (3.0)
Completeness (%)	98.42 (91.22)
Mean I/ σ (I)	15.74 (1.53)
Wilson B-factor	54.67
R-merge	0.05315 (0.7338)
R-meas	0.06293
CC1/2	0.999 (0.576)
R-work	0.1912 (0.3219)
R-free	0.2375 (0.3605)
Number of non-hydrogen atoms	11678
macromolecules	11189
water	275
Number of protein residues	1386
RMS deviations	
bond lengths (Å)	0.014
bond angles (°)	1.76
Ramachandran favored (%)	97
Ramachandran outliers (%)	0
Clash score	2.46

Statistics for the highest-resolution shell are shown in parentheses.

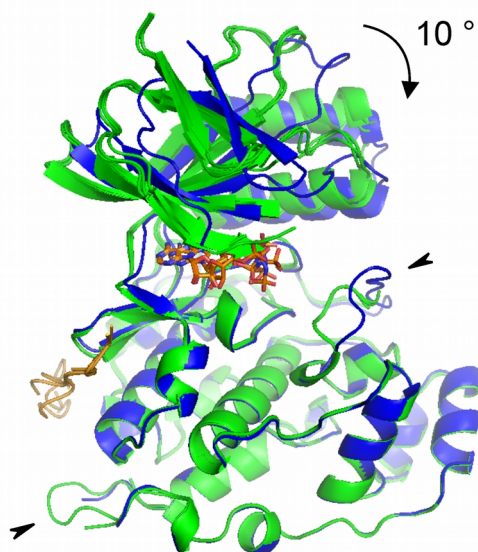


Figure 6.30: The crystal of JNK1–AMP–PNP–D2 peptide contains four JNK1 molecules of which one adopts a slightly different conformation. The three molecules with similar conformation are coloured in green, the molecule with the N-lobe twist is displayed in blue. AMP–PNP is displayed as orange sticks. The bound D2 peptide is displayed with light orange colour. The upper arrowhead indicates the activation loop which is structured differently in one of the molecules. The bottom arrow indicates a loop region whose conformation differs from the loop conformation of other published JNK1 structures.

The crystals contained four protein chains in the asymmetric unit. Three JNK1 molecules from the asymmetric unit superimpose well on each other, while one JNK1 molecule adopts a slightly different conformation with a 10° twist between the C- and N-lobes (Figure 6.30). When compared to the previously solved structure of JNK1 in complex with a peptide from the scaffold protein JIP1 (1UKH) the RMSD is $0.75 \pm 0.10 \text{ \AA}$ and $0.98 \pm 0.02 \text{ \AA}$ for the N- and C-lobes, respectively.

A closer look at the bound D2 peptides reveals that the residues binding the Φ_A and Φ_B hydrophobic pockets adopt the same conformations in all four molecules, while the basic part adopts a different conformation in each peptide (Figure 6.31b). Residue Thr42, located between the basic and hydrophobic part of the peptide, displays a high degree of flexibility and in one of the structures the electron density is missing for this residue and also for the following Leu43 (red in Figure 6.31a-c). The JNK1 molecule to which the mentioned interrupted peptide is bound is the unique molecule from the asymmetric unit which adopts a different C- to N-lobe angle.

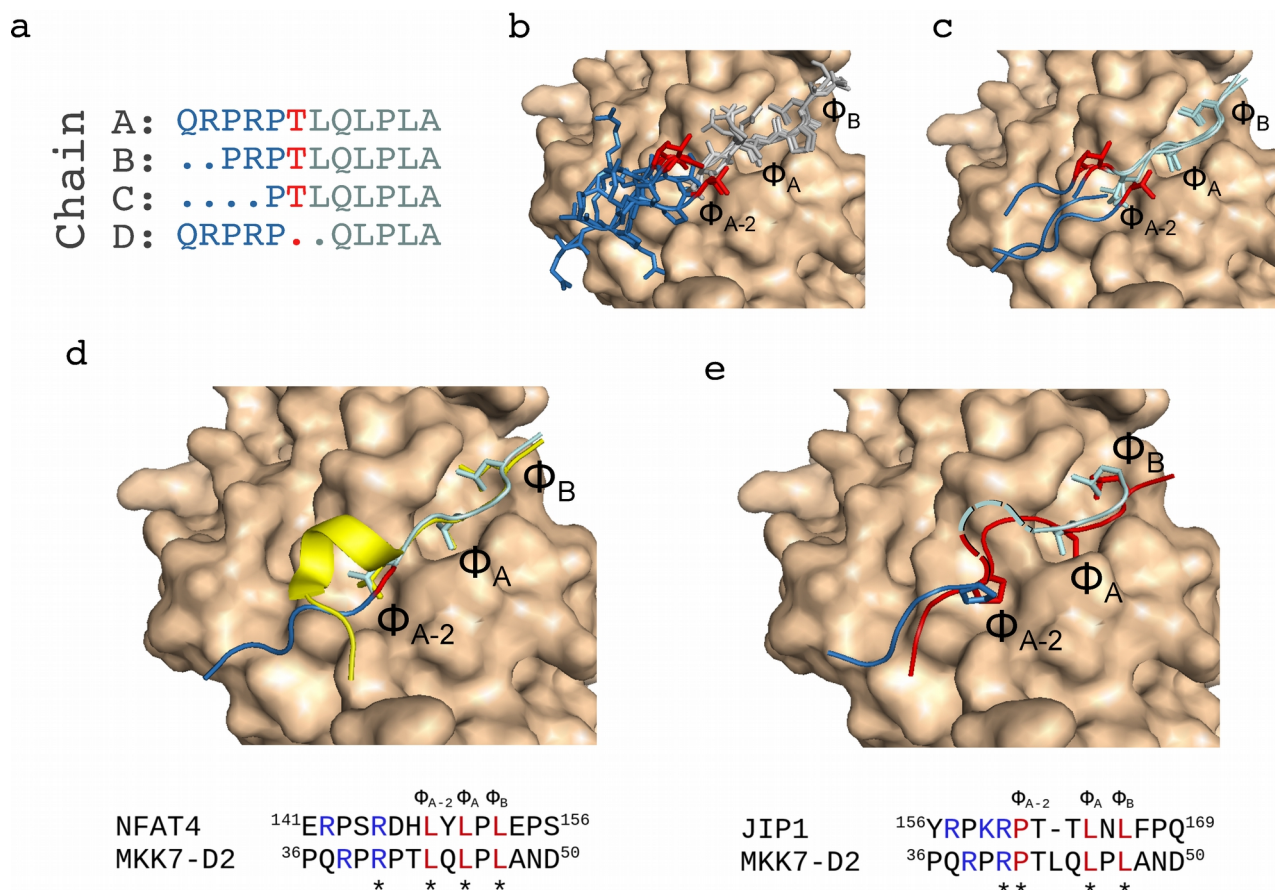


Figure 6.31: The D2 peptide adopts different conformations in complex with JNK1. a) Sequences of bound peptides with points denoting residues for which the electron density is missing. b) The four molecules from the unit cell were superimposed by their C-lobes. Peptides are displayed as sticks on the surface of JNK1 (wheat colour). c) Same as in (b) with cartoon representation of peptides which better illustrates the backbone trace of the peptides. d) The peptide with the best defined electron density in main conformation superimposed with the NFAT4-JNK1 crystal structure (yellow). The Φ_{A-2} pocket is occupied by Leu43 of MKK7. e) The peptide in the alternative conformation. The Φ_{A-2} pocket is occupied by Pro41 of MKK7. The hypothetical loop for which the electron density is missing is represented with a dashed line.

We can compare the conformation of the D2 peptide to the previously solved structures of JNK1 in complex with docking-site peptides notably those of the JNK1-NFAT4 and JNK1-JIP1 complexes. The co-crystallized peptides adopt an NFAT4-like conformation occupying the three pockets (Φ_{A-2} , Φ_A and Φ_B) with leucines (Figure 6.31d). The peptide in which Thr42 and Leu43 are missing contacts the pockets differently, with a proline in Φ_{A-2} position and the last two leucines in the Φ_A and Φ_B pockets (Figure 6.31e). This adopted conformation resembles the JIP1 binding mode, except that the linker region between the two pocket-binding residues is one residue longer and not observed in the electron density. Because the NFAT4-like conformation is observed three times and the JIP1-like conformation is observed only once, they are named main and alternative binding modes, respectively.

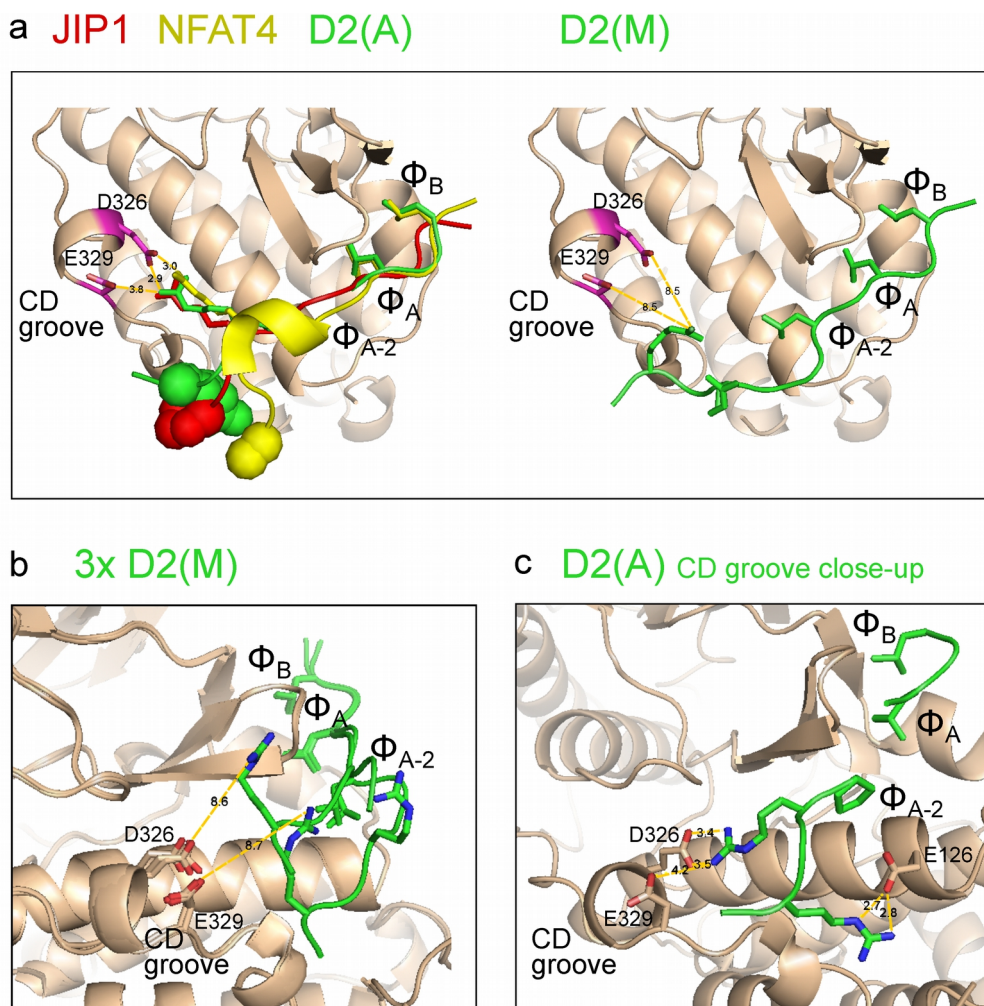


Figure 6.32: Contacts between arginines of the docking-site motif and the CD groove of JNK1. a, left) Similarities in the CD groove (purple) contacts between JIP1 (red), NFAT4 (yellow) and D2 in the alternative conformation (green). The three peptides occupy three hydrophobic pockets (Φ_{-2} , Φ_A and Φ_B) and contact the CD groove through one of its arginines. The residues represented by spheres are the first arginine/lysine of the docking-site motif (for NFAT4 and D2(A)). As the JIP1 peptide is an amino acid too short to include the first arginine of the JIP1 docking-site motif a following proline (RP) was displayed as spheres instead. It is possible to notice that they adopt a similar location on the surface of the kinase. a, right) The D2 peptide in the main conformation also occupies all three hydrophobic pockets (Φ_{-2} , Φ_A and Φ_B), but does not contact the CD groove. b) Examination of all three peptides in the main conformation shows that none of the peptides in the main conformation makes contacts to the CD groove in the crystal structure. c) The D2 peptide in the alternative conformation makes an additional contact to Glu126 outside the CD groove. This contact is not observed in crystal structures of JIP1 and NFAT4.

Interestingly, the alternative binding mode makes significant contacts with the JNK1 molecule, while the three peptides in the main binding mode do not form any contacts via their arginines (Figure 6.32). Figure 6.32a shows how the second arginine (Arg39) of D2 forms a salt bridge with Asp326 and Glu329 located in a region for which it was previously shown that it binds positively

charged residues from the basic part of the docking-site motifs (Tanoue et al., 2000). This region is called common docking (CD) groove. The JIP1 and NFAT4 peptides contact the CD groove in the same way as D2. From the same view angle it is possible to see that the longest of the peptides in the main conformation does not make these contacts (Figure 6.32a, right). The distance between the CD groove and the other two peptides in the main conformation seems even larger (Figure 6.32b). Another contact which contributes to the interaction of the peptide in the alternative conformation is the salt bridge between Arg37 and Glu126 (Figure 6.32c).

The observation of different binding modes of the D2 peptide and the fact that the kinase molecule with the bound peptide in the JIP1-like conformation displays a 10° rotation between the N- and C-lobes raises the question whether these changes in structure are somehow coupled and related to the function of the kinase. Comparison to other structures of JNKs can help us address this question. Among the previously solved JNK structures, the structure of JNK3 with the ATP analogue AMP-

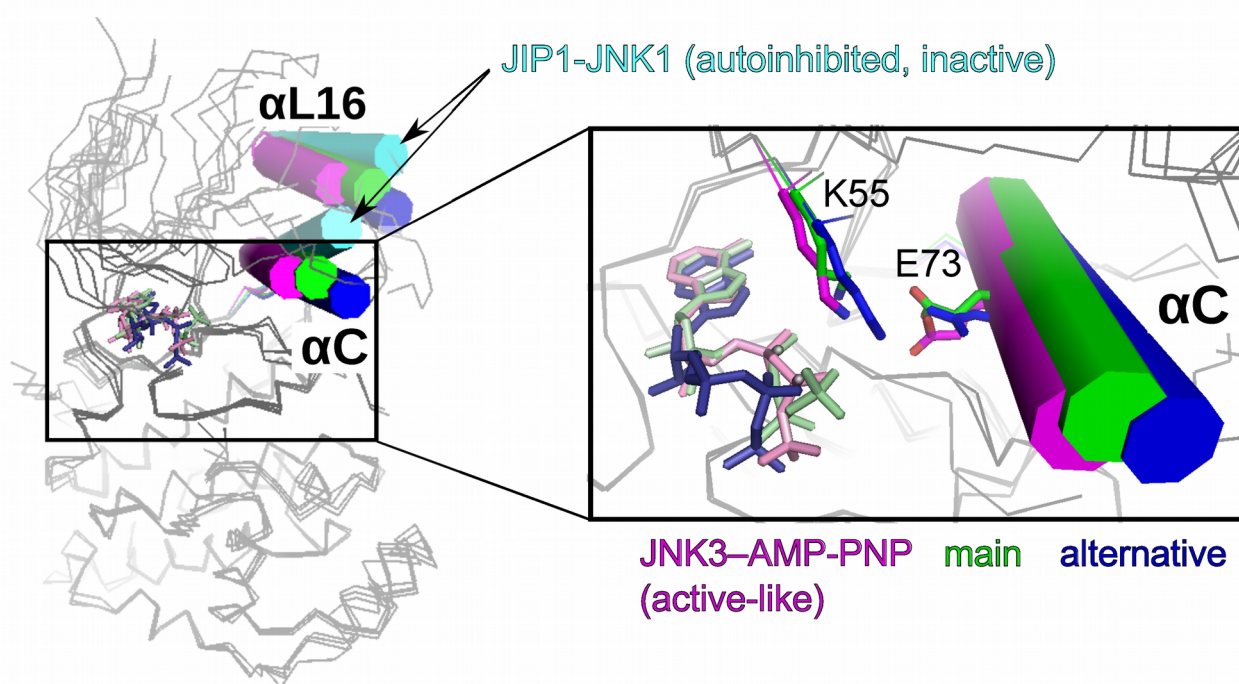


Figure 6.33: Comparison of the kinase domains with the previously solved structures. The twist of the N-terminal lobe can be followed by observing the αC and $\alpha L16$ helices (cylinders). The structure of JNK3 (magenta, PDB accession code: 1JNK) adopts a conformation similar to active conformation. The conformation of JNK1 with bound JIP1 peptide and an inhibitor in the ATP binding pocket (cyan, PDB accession code 1UKH) displays a 15° interlobe twist. In the JNK-main (green) and JNK1-alternative (blue) structures αC and $\alpha L16$ helices are also displaced compared to the JNK3 structure. Only the JNK1-alternative structure shows changes in the ATP binding pocket that indicate that JNK1-alternative adopts an inactive conformation.

PNP is considered to be closest to the active state conformation, despite the lack of phosphorylation of the activation loop (Laughlin et al., 2012; Xie et al., 1998). The structure of JNK1 in complex with the JIP1 peptide and with an inhibitor bound to the ATP binding site displays a 15° degree interlobe twist and has been proposed to represent an autoinhibited conformation that has a distorted ATP binding pocket and a decreased ATP binding affinity (Ember and LoGrasso, 2008; Laughlin et al., 2012). A comparison of these two structures with the two structures obtained here (JNK1 with the D2 peptide in the main and alternative conformation) can be made by comparing the tilt between the α C and α L16 helices (Figure 6.33). Both structures that have the peptide bound in the main and alternative conformations (JNK1-main and JNK1-alternative, respectively) display a tilt of the α C and α L16 helices. The JNK1-main conformation is more similar to the structure of the active-like JNK3. The position of the ATP analogues in the JNK3 and JNK1-main are overlapping and the catalytic lysine K55 is coordinated between the ATP analogue and the glutamate from the α C helix (E73) (Figure 6.33, magnification). The α C helix in the JNK1-alternative conformation moves even further away from the ATP analogue and in this JNK1 structure the triphosphate moiety of ATP-PNP has moved out of the ATP binding pocket. Consequently the coordination of the catalytic lysine K55 by the ATP-PNP and Glu73 is lost. It can therefore be concluded that the JNK1-alternative conformation represents an inactive conformation in comparison with the more active-like conformation of the JNK1-main structures.

6.7 Kinetics in the JNK-MKK7 complex

The binding kinetics of docking-sites from MKK7_1-100 were studied using a combination of ^{15}N CPMG and CEST experiments (Vallurupalli et al., 2012) using a sample with 10% molar ratio of JNK1. The ^{15}N CPMG experiment was measured at two different magnetic field strengths (600 and 800 MHz) and the CEST experiment was measured at two different saturating B_1 fields (22 Hz and 44.5 Hz) at 700 MHz. The three docking-sites of MKK7 display different binding kinetics. The docking-site D1 does not show any clear sign of conformational exchange. It could be that the exchange time scale is inaccessible to the two methods used or because the population of the bound state is too low. The second docking motif of MKK7 displays clear exchange contribution to the relaxation and resolved CEST profiles which can be simultaneously analysed in terms of a two-state model on a per-residue basis (Figure 6.34a, 6.34b). The obtained populations for the bound state are around 3% and the exchange rates vary from 275 to 70 s^{-1} over the D2 site. Slower exchange rates are observed for residues of the hydrophobic part of the motif and faster exchange rates for the

basic cluster, suggesting higher mobility in the basic cluster. This agrees with our crystal structure of the JNK1-D2 complex and also with the previously determined crystal structures where the basic part of the motif is seen to adopt many conformations. The difference in relaxation rates between the free and bound form (ΔR_2) derived from CEST experiments is high for the D2 site of MKK7 and reaches rates around 800 s^{-1} for some residues, which indicates that the docking motif is exchanging between different conformations in the bound state and forms a so-called fuzzy complex (Tompa and Fuxreiter, 2008).

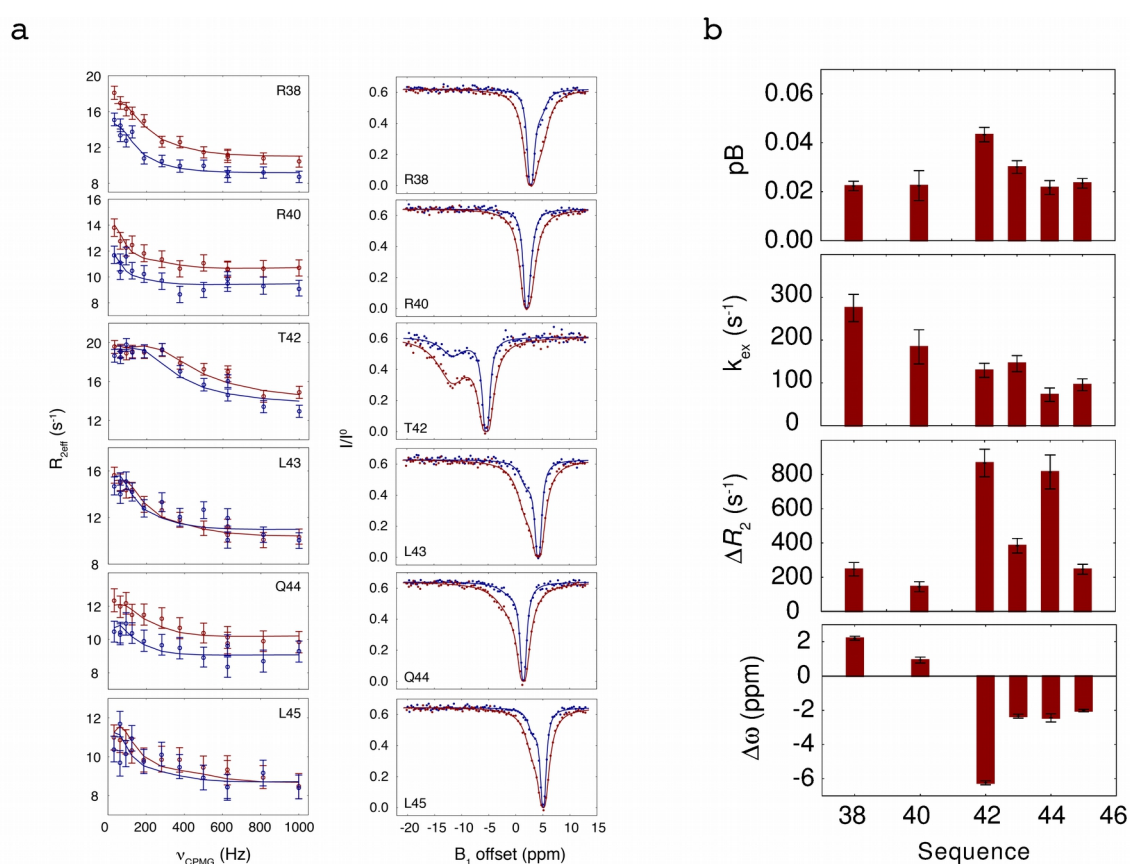


Figure 6.34: CEST and CPMG relaxation dispersion data for the second docking-site motif of MKK7. CEST and CPMG data were measured on MKK7-JNK1 complex (10% saturation). The relaxation dispersion experiments were recorded at 600 MHz (blue) and 800 MHz (red) spectrometers. CEST experiments were recorded on a 700 MHz spectrometer with two B_1 saturating fields of 22 Hz (blue) and 44.5 Hz (red). a) Relaxation dispersion (left) and CEST (right) data with displayed fits to a two state model. b) Parameters obtained from the simultaneous fits of the CPMG/CEST data on a per-residue basis: population of bound second docking-site of MKK7 (pB), the exchange rate (k_{ex}), the difference in the effective R_2 relaxation rates (ΔR_2) and the difference in the ^{15}N chemical shifts between the two states ($\Delta\omega$).

6. Results

The D3 docking-site displays less pronounced exchange contributions to relaxation and less pronounced CEST profiles. Despite that, residues contributing to binding of the docking motif can be clearly identified, including three residues (T80, L81 and F82) outside the canonical docking motif which also contribute to the binding (Figure 6.35a). The analysis of the data was performed

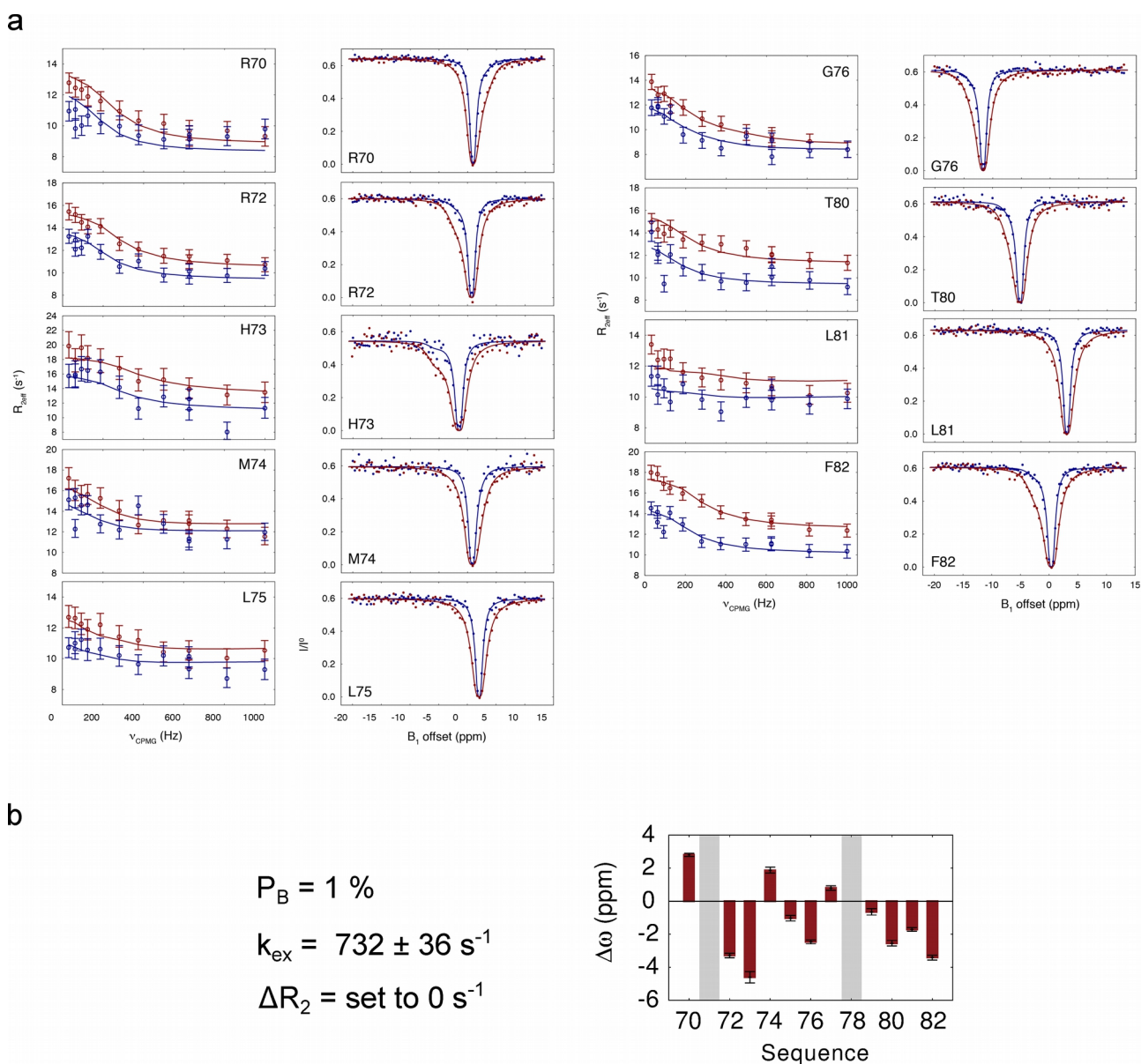


Figure 6.35: CEST and CPMG relaxation dispersion data for the third docking site of MKK7. CEST and CPMG data were measured on the MKK7-JNK1 complex (10% saturation). The relaxation dispersion experiments were recorded at 600 MHz (blue) and 800 MHz (red) spectrometers. CEST experiments were recorded on a 700 MHz spectrometer with two B_1 saturating fields of 22Hz (blue) and 44.5 Hz (red). a) Relaxation dispersion (left) and CEST (right) data with displayed fits to a two-state model. b) Parameters obtained from the simultaneous fits of the CPMG/CEST data. The population of bound D3 (p_B) and the exchange rate (k_{ex}) were fitted as common parameter for all residues of the docking-site motif while the difference in the effective R_2 relaxation rates (ΔR_2) was fixed to zero. The difference in the ^{15}N chemical shifts between the two states ($\Delta\omega$) was allowed to vary on a per-residue basis.

by fitting a single exchange rate for all residues displaying exchange, while the ΔR_2 was fixed to zero for all residues as this did not affect the fitting of the CEST data. In reality the transverse relaxation rates are expected to be higher for the complex than for the unbound form making ΔR_2 larger than zero. The exchange rate obtained by the analysis is $723 \pm 36 \text{ s}^{-1}$ which is significantly faster than the exchange rates obtained for D2 (Figure 6.35b).

6.8 Alignment of JNK interacting motifs

The structure of the D2 peptide in complex with JNK1 can be used to inform the alignment of the other docking-site motifs of MKK7. Sequence alignments suggest that all three docking motifs of MKK7 could adopt two binding modes (Figure 6.36). All docking-site motifs of MKK7 are potentially able to adopt a JIP1 conformation because of the nonpolar/aromatic residue directly following the RPR residues. A conformation similar to the NFAT4 peptide could also be adopted by the docking-site motifs of MKK7 because they all have a two residue linker between the basic cluster and the hydrophobic LXL motif. On the other hand JIP1 would not be able to adopt the NFAT4 conformation as it has a three-residue linker before the first leucine and because it lacks a hydrophobic residue following the LXL motif. Similarly, we can exclude the possibility that the NFAT4 peptide can adopt a JIP1-like conformation because it lacks an nonpolar/aromatic residue following the RPR residues.

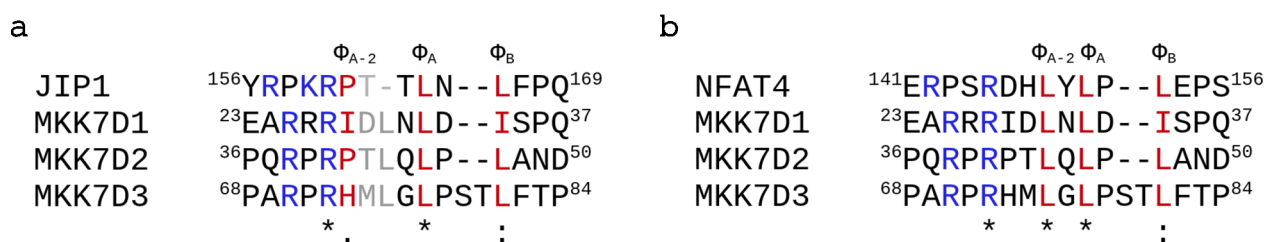


Figure 6.36: Alignment of the docking-site motifs from MKK7 to JIP1 (a) and NFAT4 (b) docking-site motifs. An * (asterisk) marks positions with identical residues, a : (colon) marks conservation between groups of strongly similar properties, a . (period) marks conservation between groups of weakly similar properties.

The observation of two alternative docking-site conformations in the crystal structure poses a question whether these two conformations could account for most of the known JNK1 binding peptides and if the separation into these two groups could facilitate and improve sequence

6. Results

alignments. Manual alignment of experimentally confirmed JNK1-binding motifs shows that they can be divided into three groups, depending on the linker length between the basic and hydrophobic part of the motifs (Figure 6.37). Two of these groups contain JIP1-like peptides and NFAT4-like peptides. The third group contains peptides with a linker length of one amino acid and includes the docking-site motif of MKK4. None of the peptides from this group have been crystallized in complex with JNK.

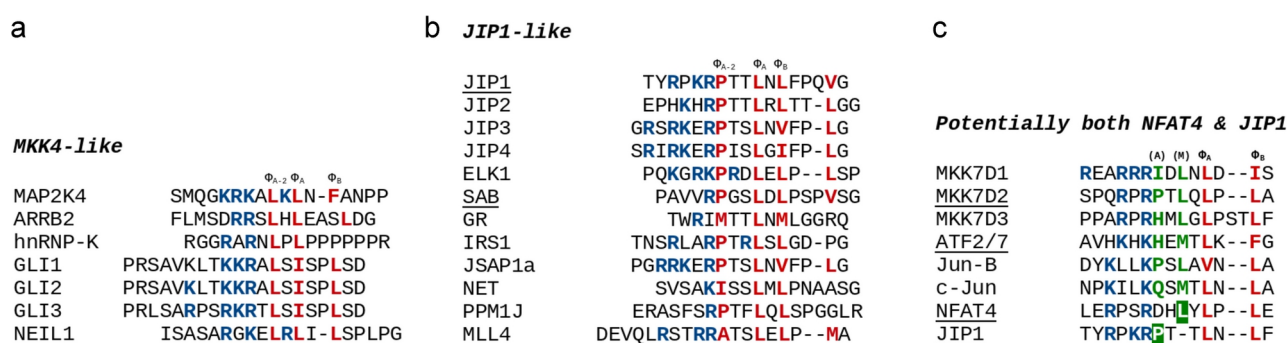


Figure 6.37: Grouped and manually aligned sequences of verified JNK1-binding motifs. a) There are several motifs with a one-residue linker between the basic and hydrophobic part of the motif. The docking-site motif of MKK4 belongs to this group. b) Most of the peptides with a three-residue linker between the basic and hydrophobic part contain proline immediately following the basic part. It is therefore highly likely that these peptides bind in the same conformation as JIP1. c) There are several docking-site motifs that potentially bind in both NFAT4 and JIP1 binding mode, by adopting a main and alternative binding mode just like D2 of MKK7. The residues from the two positions potentially exchanging in the Φ_{A-2} pocket are coloured green. Note the NFAT4 and JIP1 peptide which only have one hydrophobic residue highlighted in green. The peptides whose name is underlined have already been co-crystallized with JNK. ATF2/7 was co-crystallized with JNK3 in a conformation resembling the main conformation, but contacting the CD groove with the histidine. The reason could be that the ATF2/7 peptide is short encompassing only a part of the basic cluster (⁴⁸KHEMTLKF⁵⁵).

Sequence comparisons and alignments of distant homologues can provide us with information about conservation and therefore importance of the different docking-sites. To assess the conservation of the three docking-site motifs of MKK7 multiple sequences were retrieved using a BLAST search. The search was performed with the sequence of the catalytic domain of human MKK7 which displays higher conservation than the IDR. Sequences of full-length proteins including the IDRs were retrieved and analysed. A comparison of the MKK7 sequences shows that not all MKK7 kinases have three docking-site motifs (Figure 6.38). The three docking-sites first appear in deuterosomes. Alignment of the docking-site motifs that are located closest in sequence to the catalytic kinase domain reveals that the spacing between the third docking-site and the catalytic

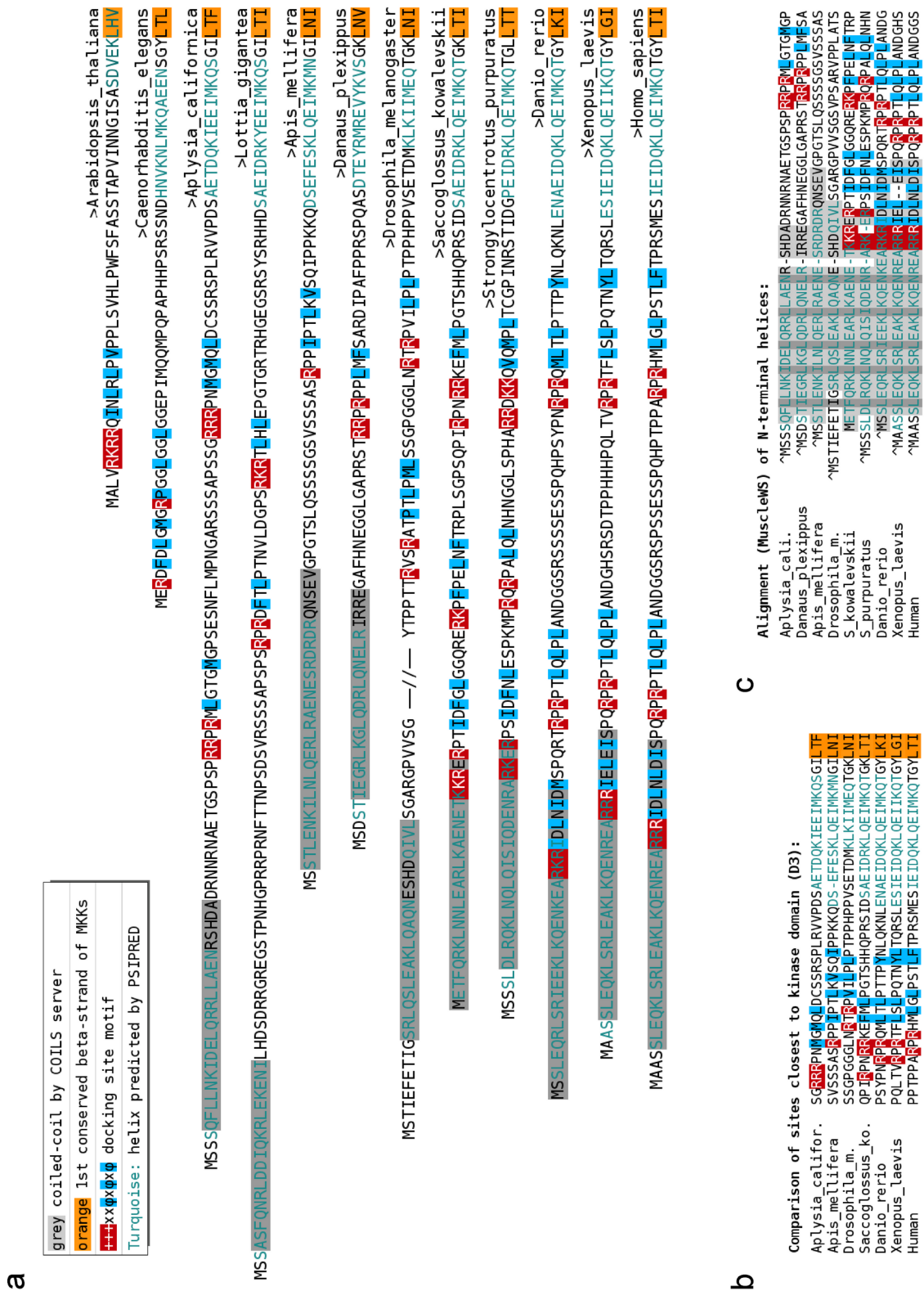


Figure 6.38: Comparison of distal MKK7 homologues. a) Sequences of IDRs of MKK7 aligned by the first conserved beta sheet of the catalytic domain of the MKK7 kinase. For each sequence coiled-coil (grey) and α -helix propensity (turquoise) were predicted with COILS and PSIPRED, respectively. b) Alignments of the docking-site motifs that are closest to the catalytic domain. c) Alignment of the N-terminal α -helices from distal MKK7 homologues.

6. Results

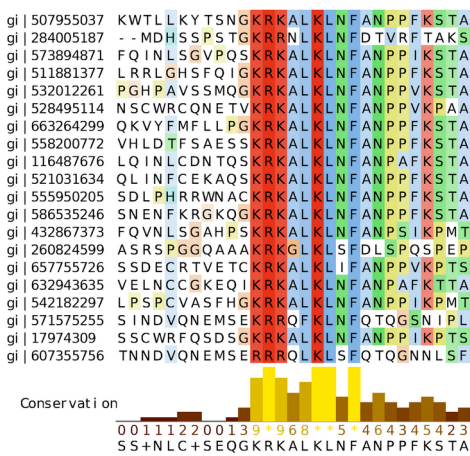


Figure 6.39: Alignment of MKK4 docking-site motif from vertebrates.

The docking-site motif from evolutionary distant species cannot be easily aligned. Instead the docking-site motifs from vertebrates can be aligned well and are enough diverse to provide information about the conservation of different positions of the docking-site motifs. The docking-site motif of MKK4 is well-conserved. Among the docking-sites of MKK7 the D2 docking-site motif is most conserved. The alignment of less conserved D3 docking-site motif reveals an additional hydrophobic position that is part of the Thr80-Leu81-Phe82 cluster, which was identified by NMR and ITC as necessary for high-affinity binding.

domain is conserved (Figure 6.38b, left). Alignment of the N-terminal regions of the MKK7 kinases reveals that the characterised N-terminal α -helix is an evolutionary well conserved feature across species as distant as insects and molluscs (Figure 6.38c).

The docking-site motifs from evolutionary distant species cannot be easily aligned. Instead the docking-site motifs from vertebrates can be aligned well and are enough diverse to provide information about the conservation of different positions of the docking-site motifs. The docking-site motif of MKK4 is well-conserved. Among the

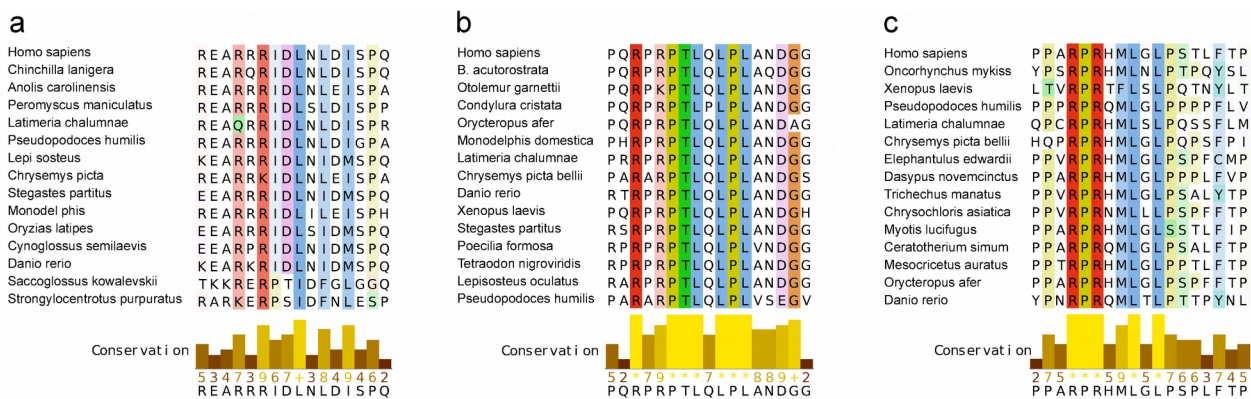


Figure 6.40: Alignment of MKK7 docking-site motifs from vertebrates. a) The D1 docking-site motif. b) The D2 docking-site motifs. c) The D3 docking-site motifs. Alignment of the D1 docking-site motif contains two sequences from non-vertebrates (deuterosomes) (*Saccoglossus kowalevskii*, *Strongylocentrotus purpuratus*).

6.9 The scaffold protein JIP1

In order to structurally characterize the JIP1 scaffold protein and its interactions, expression tests were performed for several different constructs of JIP1. Trx-tagged constructs were prepared to cover the whole sequence of JIP1 until the end of the SH3 domain. Most of these constructs expressed well with the exception of the longer constructs (Figure 6.41). The construct number 4 to which the interaction with MKK7 was mapped expressed in inclusion bodies. In the region where MKK7 is predicted to interact, the disorder prediction (Figure 6.41) indicates the presence of a highly hydrophobic region characterised by a score below the 0.5 threshold value. The region predicted to be ordered spans 45 amino acids which is too few to form an independently folded domain. I therefore propose that this region interacts with MKK7 in which case co-expression with MKK7 would solve the problem of inclusion body formation during expression of construct number 4.

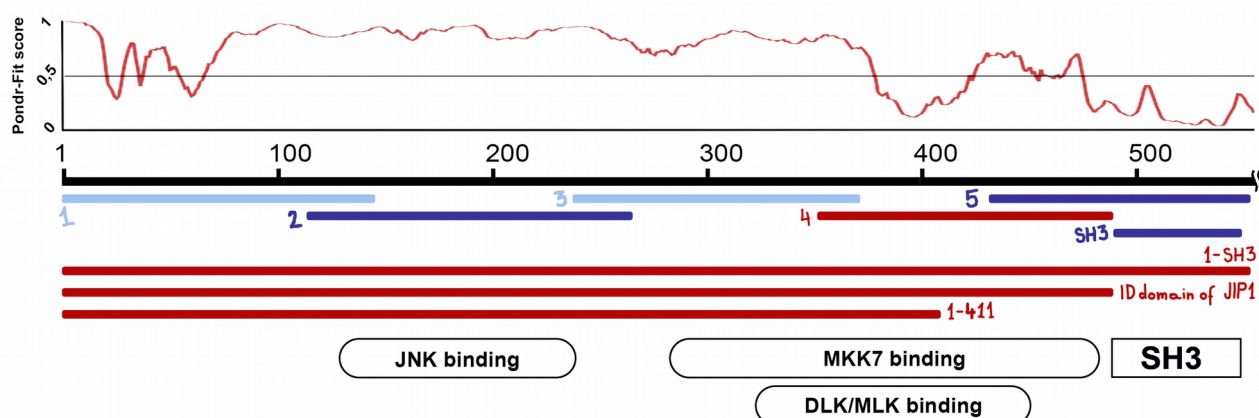


Figure 6.41: Test expression of JIP1 constructs. The blue bars denote constructs that were expressed. Light blue: 0.5 mg/L M9 (1 NMR sample). Dark blue: > 2 mg/L M9 (4 NMR samples). Red bars indicate no expression and in the case of construct number 4 expression yielded inclusion bodies.

JIP1 constructs denoted 1, 2 and 3 were expressed ^{13}C , ^{15}N labelled for assignment experiments. The ^1H - ^{15}N HSQC spectra recorded on these three constructs covering the first 372 amino acids of JIP1 show that this region of JIP1 is indeed disordered (Figure 6.42). Standard triple-resonance assignment experiments were recorded for the JIP1 constructs 1, 2 and 3. An assignment strategy using ^{13}C detected experiments was also undertaken. Apart from better signal dispersion such a strategy also allows assignment of proline resonances. Four experiments were recorded with the JIP1 construct number 2: 4D HCBCACON, 4D HCBCANCO, 4D ($\text{H}^{\text{N-flip}}\text{N}$)CON(CA)CON and

6. Results

(HCA)CON(CA)CON (Bermel et al., 2012, 2013). Two experiments of the type CON(CA)CON that start with excitation of different types protons were used because the experiment with HCA start allows detection of the proline crosspeaks while the $H^{N\text{-flip}}N$ start is a slightly more sensitive experiment. The assignment of JIP1 resonances is ongoing.

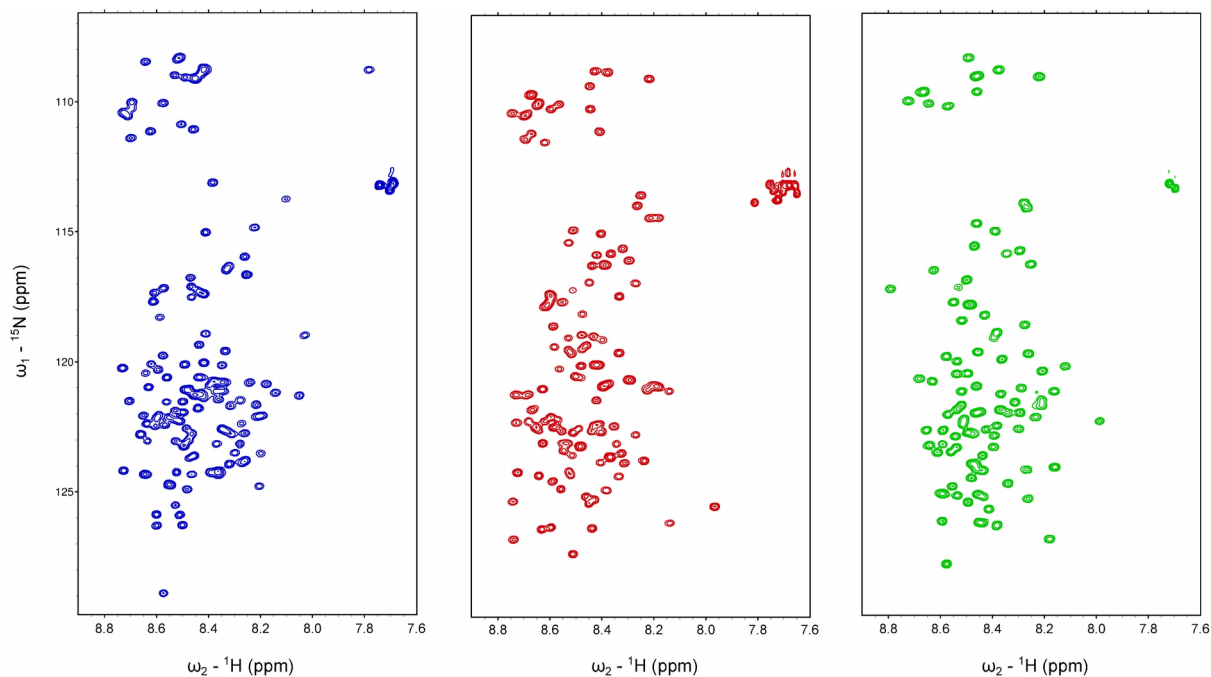


Figure 6.42: 1H - ^{15}N HSQC spectra of JIP1 constructs JIP1_1, JIP1_2 and JIP1_3 (pH 7.0, 5 °C).

The 2D CON spectrum is used for initial picking of peaks analogously to the first peak list created from an HSQC when triple-resonance experiments are used for the assignment. In the Figure 6.43 it is possible to see the superior dispersion of CON, recorded on the JIP1_2 construct, compared to the 1H - ^{15}N HSQC spectrum. The CON peak list is then used for Sparse Multidimensional Fourier Transform (SMFT) (Kazimierczuk et al., 2009) of the 4D spectra to obtain 2D slices corresponding to each CON cross-peak.

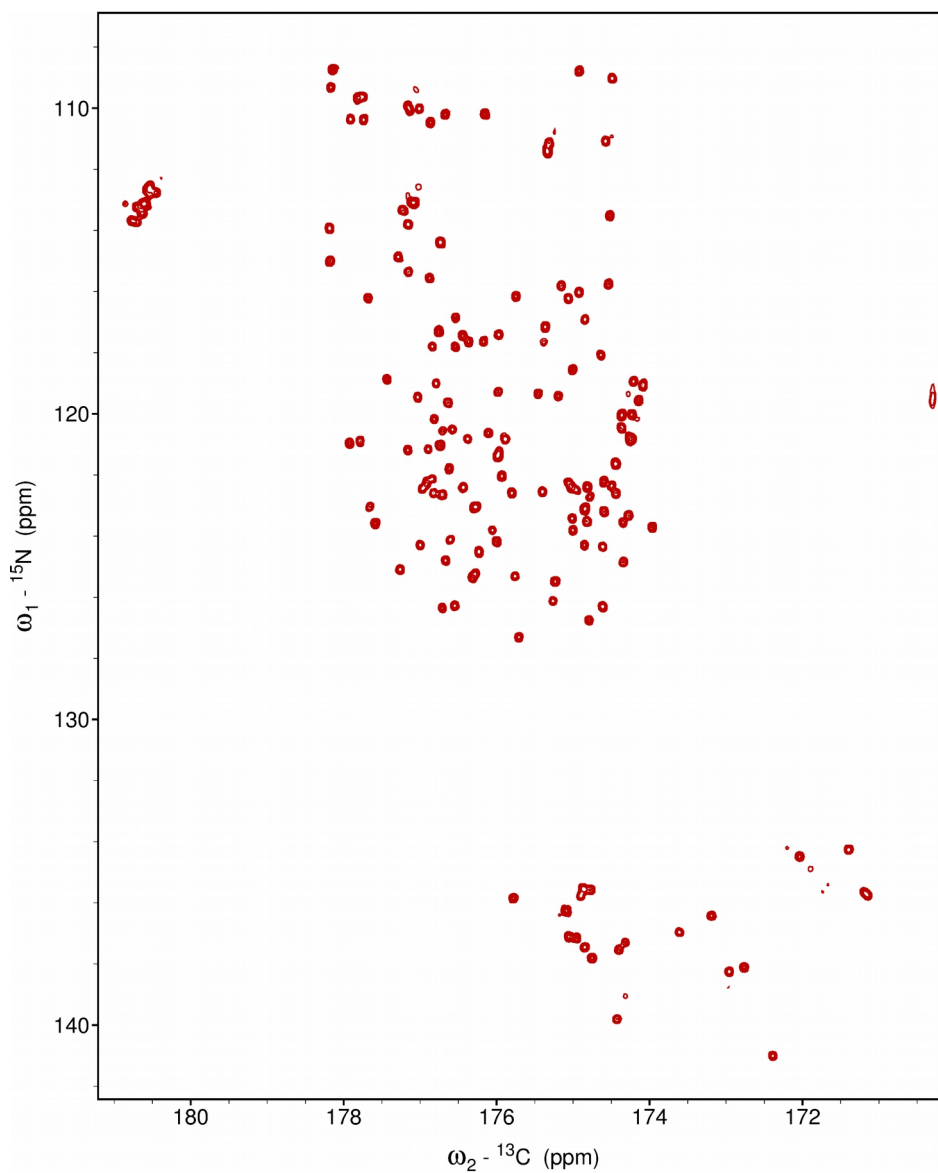


Figure 6.43: ^{13}C - ^{15}N CON spectrum of JIP1_2 recorded on 700 MHz Bruker Avance Spectrometer equipped with a cryogenically cooled probehead optimized for ^{13}C -direct detection. Sample contained 500 mM JIP1_2, 50 mM HEPES, 150 mM NaCl, pH 7.0 at 5 °C. Resonances corresponding to proline cross-peaks are located around 136 ppm in the indirect nitrogen dimension.

7 Discussion

In recent years it has become apparent that signalling proteins are enriched in intrinsic disorder. In MAPK signalling pathways the N-terminal IDRs of MKKs determine specificity by targeting the kinase domain to the cognate substrates. Characterization of these highly flexible domains and their functional complexes is therefore required to gain insight into the molecular basis of MAPK signalling. During my thesis I have provided novel mechanistic insights into how MKKs and MAPKs interact using a multidisciplinary approach, combining NMR, computational NMR techniques, X-ray crystallography and ITC. I have characterised the IDRs of MKK4 and MKK7 at atomic level resolution by NMR, I have studied the interaction of MKK7 and MKK4 with JNK1 and I have determined the affinities, stoichiometries and kinetics of complex formation. During my thesis I have also made the first steps towards characterising the JNK scaffold protein JIP1 that contains a high percentage of intrinsic disorder.

Analysis of experimental NMR data for the N-terminal IDR of MKK7 (MKK7_1-100) in terms of ensembles reveals several regions with transient structure and different characteristics. The most striking feature is the presence of a transient α -helix populated to around 65% that can be described with an ensemble of interconverting α -helices of different lengths (Figure 6.8). The region that contains the transient α -helix has previously been reported to interact with the upstream dual leucine zipper kinase DLK and with MEKK2 (Nakamura and Johnson, 2007). The transient α -helix could therefore be a molecular recognition element for the two upstream kinases. The transient α -helix is predicted to form a coiled-coil (Figure 6.2) but does not homodimerize as we have shown by MALLS (Figure 6.11). As the interaction between MKK7 and DLK was mapped to the leucine zipper region of DLK (Mooney and Whitmarsh, 2004), I suggest that it forms a coiled-coil with DLK. The transient α -helix could also be involved in the interaction with another upstream kinase (MEKK2). Interaction of MEKK2 PB1 domain was mapped to the N-terminal IDR (residues 1-114) and to the kinase domain of MKK7 (Nakamura and Johnson, 2007).

At the beginning of my thesis one of the main questions concerned the nature of the conformational sampling in the docking-sites. In the crystal structures of docking-sites bound to MAPKs, the docking-sites adopt extended conformations (Figure 2.7, 2.8 and 6.10d). Could it be that they form pre-structured elements and oversample the extended conformations when unbound? The analysis of NMR data in terms of ensembles reveals that each of the studied docking-sites samples different conformational regimes in the unbound state (Figure 6.10b, 6.10c). The D1 docking-site of MKK7

over-samples α -helical conformations because it partially overlaps with the N-terminal transient α -helix of MKK7. The D2 docking-site of MKK7 exhibits elevated propensity to populate extended (PPII) conformations and most closely resembles the bound-state of docking-site motifs (Figure 6.10d). The D3 docking-site of MKK7 samples conformations that are similar to the statistical coil library and does not display any stronger deviations towards β -strand, PPII or α -helical conformations. Last but not least, the conformational sampling of the docking-site of MKK4 shows an elevated propensity to sample PPII conformations (Figure 6.19) and could be classified between the sampling regimes of the D2 and D3 docking-sites.

In the light of these differences it is interesting to examine whether the different conformational behaviour of the docking-sites affects the affinities or kinetics of binding to JNK1. ITC studies show that all three docking-sites of MKK7 and the docking-site of MKK4 are able to bind to JNK1 with low micromolar affinities (Figure 6.27). The docking-site of MKK4 binds with the highest affinity (1.5 μ M) of all docking-sites. The affinities of the docking-sites of MKK7 are 12 ± 5 μ M, 7.9 ± 1.3 μ M and 11 ± 1 μ M for D1, D2 and D3, respectively. The docking-sites of MKK7 have similar affinities despite having different conformational propensities in the unbound state. D2 which displays an increased PPII sampling has perhaps a slightly higher affinity than the other two docking-sites of MKK7. The docking-site of MKK4, which displays conformational propensities similar to D2, binds with the highest affinity of all measured docking-sites. It is interesting to note that the two strongest binding sites (D2, and docking-site of MKK4) display a higher propensity to sample PPII conformations, although highest PPII propensity (i.e. D2) does not imply the highest affinity. With the available data, we can say that the amount of pre-structuration does not seem to have a clear correlation with the binding affinities. While the conformational sampling in the unbound state definitely influences kinetics and thermodynamics of complex formation, the influence may be smaller in comparison with other determinants such as different contacts formed in the complex due to sequence differences. Apart from providing affinities, the ITC experiments were additionally insightful as they showed that the N-terminal IDR of MKK7 containing three docking-sites can bind three molecules of JNK1 forming a 3:1 JNK:MKK7 complex. Both stoichiometry of the titrations and additive enthalpies confirm this (Figure 6.27). The fact that the D1 and D2 docking-sites are able to simultaneously bind JNK1 is interesting considering that these two binding sites are separated by only 5 residues and are still able to accommodate two large JNK1 molecules. The physiological relevance of the 3:1 stoichiometry is not clear, but it may be related to occlusion of individual binding sites upon interaction with additional binding partners. An example of such event could be the binding of the upstream kinases DLK and MEKK2 to the identified of a

putative α -helical molecular recognition element. In the case of such a binding event, the D1 and D2 docking-sites could become inaccessible leaving D3 accessible for interaction. Another function for the multiple docking-sites could be binding to JNKs in the context of macromolecular complexes where kinases are often found in dimeric arrangements.

The different conformational propensities of the docking-sites could influence kinetics of binding to JNK1. To examine the kinetics of binding, CEST and relaxation dispersion experiments were recorded at a low molar fraction of JNK1 compared to MKK7 meaning that the exchange rate obtained by these experiments is dominated by the rate of dissociation k_{off} . The obtained off-rate for D3 is $723 \pm 36 \text{ s}^{-1}$, while the D2 exhibits an approximately five-times slower off-rate of $137 \pm 8 \text{ s}^{-1}$. The D1 site does not display neither CEST nor relaxation dispersion, clearly indicating a very different kinetic regime for this interaction. In conclusion, the three docking-sites of MKK7 display different conformational propensities, similar affinities and different kinetics of complex formation. Additionally, a qualitative analysis of the MKK4 titration data shows that the docking-site of MKK4 probably binds JNK1 in a slow exchange regime, while the ITC studies show that the docking-site of MKK4 binds with a higher affinity compared to the docking-sites of MKK7. Although a detailed analysis of how pre-formed motifs affect binding affinities and kinetics is not possible due to the large sequence differences between the docking-sites of MKKs, pre-structuration probably has an effect on binding affinities and kinetics of binding. The characterization of structure of docking-sites and affinities and kinetics of binding paves the ground for more detailed studies in the future.

An interesting detail is how we were able to identify the minimal docking-site motif for D3 of MKK7 by combining the chemical shift titration, CEST, relaxation dispersion and ITC. The previously defined D3 motif was too short and the affinity of D3 towards JNK1 was reported to be much lower than the affinities of the D1 and D2 docking-sites for JNK1 (Ho et al., 2006). Our measurements show that the affinity of D3 for JNK is comparable to the affinities of the other two sites, when the synthesized/expressed peptides used in the study contained additional residues Thr80-Leu81-Phe82. Evidence for that comes from several methods. ITC shows that when experiments are performed on a synthetic peptide with the canonical motif sequence, the measured affinities are much lower ($49 \pm 12 \mu\text{M}$) compared to when the longer construct is used ($11 \pm 1 \mu\text{M}$) (D3-pept and Trx-D3 measurements in Figure 6.27). In the chemical shift titration and CEST/CPMG experiments the residues of D3 display clear signs of binding and include the additional Thr80-Leu81-Phe82 (Figure 6.21, 6.35). Furthermore, alignment of D3 sequences from

vertebrates displays a conserved hydrophobic position corresponding to the newly identified Thr80-Leu81-Phe82 cluster (Figure 6.40c) and comparison of N-terminal IDRs of distant MKK7 homologues shows that the distance between the putative D3 sites and the kinase domain is conserved through evolution, adding weight to the argument that D3 is indeed a functional and important site. Delineating the shortest motif that displays the affinities which are as close as possible to the affinities of the full-length proteins represents a general problem in motif discovery. If the affinity measurement is made with a peptide with incomplete motif sequence the peptide will still display binding but with lower affinity. For a motif displaying an affinity in the nanomolar range it is possible that a measurement with a truncated motif would give micromolar affinities which may still allow crystallization and analysis of binding determinants, but would not be representative in terms of affinities and structure compared to how the full-length proteins interact. Such mistakes easily happen (Hao et al., 2007; Orlicky et al., 2003), potentially leading to proposed mechanisms of cellular processes (Mittag et al., 2008; Tang et al., 2012) that eventually need to be corrected (Kõivomägi et al., 2011, 2013; McGrath et al., 2013).

The CEST experiment provided another important insight into MKK7-JNK1 complex formation. Analysis of the data displays large ΔR_2 values between the free and bound states of MKK7 for residues within the D2 docking site (Figure 6.34). This indicates that MKK7 undergoes conformational exchange within the complex and forms a so-called ‘fuzzy complex’ (Tompa and Fuxreiter, 2008) characterized by exchange between different bound states. The crystal structure that I obtained agrees with these observations. There are four molecules of JNK1 in the unit cell and for each molecule the bound D2 peptide adopts a slightly different conformation (Figure 6.31). In the crystal structure, the basic part of the docking-site peptide is disordered in all peptides, while the hydrophobic residues occupying the two pockets Φ_A and Φ_B seem to adopt a better defined single conformation. Interestingly, the intervening segment, occupying the Φ_{A-2} pocket, displays two clearly different conformations. In three out of four peptides, the Φ_{A-2} pocket is occupied by Leu43 making the peptide conformation similar to the NFAT4 peptide previously co-crystallized with JNK1 (Figure 6.31d). One of the peptides adopts a conformation that is similar to the conformation of the JIP1 peptide previously co-crystallized with JNK1 (Figure 6.31e). I propose that the exchange contribution to the relaxation arises due to the switching between these two conformations and is responsible for the large ΔR_2 values observed in the CEST experiments.

One may ask oneself if the observed dynamics in the complex has some functional role. One theory is that the fuzziness of the docking-sites in the bound state would allow for transient interactions

formed during signalling events and fast displacement of MKKs by the substrates of JNK1 once JNK1 is phosphorylated and activated. Another explanation could be that the conformation of the D2 peptide is coupled to conformational changes in the JNK1 kinase because the examination of the JNKs shows that in the structure with the D2 peptide bound in the JIP1-like conformation shows a 10° interlobe twist compared to the structure with the D2 peptide in the NFAT4-like conformation (Figure 6.30). A comparison of the two conformations with the previously crystallized JNKs (Figure 6.33) shows that the different peptide conformation in the complex could indeed be coupled to the change of the interlobe orientation as it was suggested previously based on studies of other peptides (Laughlin et al., 2012). According to this explanation, the JNKs with the peptide in the NFAT-like conformation adopt a structure that is more similar to an active conformation of the kinase which was observed for JNK3-AMP-PNP complex (Figure 6.33). The JNK with the peptide bound in the alternative JIP1-like conformation more closely resembles an autoinhibited form of the kinase (Figure 6.33). The previous structures of JNKs co-crystallized with the JIP1 peptide display only the autoinhibited conformation and the structures of JNK with the substrate peptides (NFAT4, ATF2) display an interlobe twist similar to the JNK1 with the D2 peptide in NFAT4-like conformation. The ability of the D2 peptide of MKK7 to bind in two binding modes was observed here for the first time and the functional significance of this ability still remains to be elucidated.

7. Discussion

8 Résumé en français

8.1 Introduction

Les cellules ont constamment besoins de sentir et de réagir aux stimuli provenant de leur environnement mais aussi des autres cellules voisines dans le cas d'organismes pluricellulaires. Les voies de transduction du signal leur permettent de transmettre ces signaux et ainsi de traiter l'information qui en découle.

Certaines de ces voies, appelées voies de transduction du signal par protéines kinases activées par des agents mitogènes (MAPK, Mitogen-Activated Protein Kinase), sont très conservées chez les eucaryotes et sont impliquées dans la régulation de nombreux processus cellulaires importants. De longues Régions Intrinsèquement Désordonnées (RID), présentes dans les MAPK, restent peu caractérisées au niveau structural. Les RID des MAPK sont particulièrement importantes car elles contiennent des motifs de reconnaissance qui contrôlent les interactions entre les MAPK directement mais aussi entre les MAPK et d'autres protéines contenant les mêmes motifs. Dans le but d'acquérir une meilleure compréhension du fonctionnement des voies de transduction du signal par les MAPK, il est nécessaire d'obtenir de nouvelles données relatives à l'organisation des RID ainsi qu'aux motifs de reconnaissance.

La spectroscopie par Résonance Magnétique Nucléaire (RMN) est très bien adaptée à l'étude des protéines intrinsèquement désordonnées (PID) à l'échelle atomique. Il existe de nombreuses observables RMN, comme par exemple les déplacements chimiques ou les couplages dipolaires résiduels, qui nous informent sur les structures résiduelles des RID. La RMN permet également d'étudier, à l'échelle atomique, la dynamique et les échelles de temps des mouvements qui se produisent dans les protéines ; là où la relaxation des spins nucléaires nous informe sur la dynamique se produisant à l'échelle de la pico à la nanoseconde, d'autres méthodes comme la dispersion de la relaxation ou encore la RMN en temps réelle, nous informent sur des processus plus lents. Avec autant d'expériences applicables, la RMN peut nous permettre d'obtenir une vision du fonctionnement des RID et des motifs de reconnaissance. Cela devrait nous amener, à terme, à mieux comprendre les voies de transduction du signal.

8.2 Les protéines intrinsèquement désordonnées dans les voies de signalisation.

Les mammifères possèdent de multiples voies de signalisation MAPK réparties en ERK (kinases régulées par un signal extracellulaire, extracellular signal-regulated kinases), JNK (c-Jun N-terminal Kinases) et kinases de la famille de p38. Ces voies de signalisation sont activées en réponse à différents stimuli, principalement le stress et l'inflammation, mais aussi en réponse aux agents mitogènes.

La principale composante des voies de signalisation à MAPK est une cascade de 3 kinases se phosphorylant mutuellement de façon séquentielle. Des signaux de l'environnement sont tout d'abord transmis à des MAPK kinase kinase (MKKK) qui s'activent par phosphorylation. Les MKKK activées sont capable de phosphoryler et d'activer des MAPK kinases (MKK) qui, à leur tour, vont phosphoryler et activer des MAPK. Les MAPK activées peuvent finalement phosphoryler différents substrats.

Toutes les MKK ont une RID en N-terminal contenant un motif nommé motif de liaison. Les motifs de liaison contiennent une partie basique et une partie hydrophobe ($R/K_{2-3}-X_{2-4}-\Phi-X-\Phi$, où Φ représente un acide aminé hydrophobe et X n'importe quels résidus). La liaison des MKK aux MAPK se produit grâce à ces motifs et ces liaisons sont en grande partie responsables de la haute spécificité des réactions de phosphorylation observées *in vivo*.

Une description détaillée de l'affinité, de la stœchiométrie et de la cinétique des interactions avec les motifs de liaisons est nécessaire à la compréhension des connections au sein des réseaux de MAPK. La détermination des structures transitoires des motifs permettrait de savoir s'ils correspondent aux structures observées en complexe. Les motifs de liaison adoptent des conformations étendues dans leur état lié mais aucune description de ces mêmes sites dans leur état lié n'était disponible au moment où j'ai commencé ma thèse.

8.3 Description des structures et des interactions faisant intervenir des PID par RMN

La RMN nous permet d'étudier la dynamique et les échelles de temps au cours desquelles les mouvements ont lieu. La relaxation des spins nucléaires peut nous informer sur les mouvements les plus rapides, dont la gamme de temps est de l'ordre de la pico à la nanoseconde. Elle se caractérise par le processus de retour à l'équilibre de la magnétisation, précédemment excitée vers des niveaux

de plus haute énergie au début de l'expérience. La relaxation est due aux fluctuations des champs magnétiques à proximité du spin qui doivent se produire aux fréquences des transitions entre les niveaux énergétiques du système. Ces fréquences de fluctuations peuvent être décrites via la fonction de corrélation et sa transformée de Fourier, la fonction de densité spectrale, qui nous informe sur la quantité de mouvements à une fréquence précise. Les fluctuations sont la conséquence de l'interaction dipolaire ou de l'anisotropie de déplacement chimique et peuvent soit induire des transitions qui vont ramener le système à l'état d'équilibre, soit faire varier la fréquence de Larmor et donc déphaser irréversiblement leurs cohérences conduisant respectivement à la relaxation longitudinale et transversale.

L'échange chimique est une autre source de relaxation, sa contribution est décrite différemment en fonction du régime auquel il se produit. Pour l'échange en régime intermédiaire, quand $k_{\text{ex}} \approx |\Delta\omega|$, un seul signal est observé qui est élargi par R_{ex} . La vitesse de relaxation en régime intermédiaire s'exprime donc $R_2^{\text{obs}} = R_2^0 + R_{\text{ex}}$. L'élargissement du signal causé par l'échange chimique reflète la dynamique sous-jacente et les techniques de dispersion de la relaxation CPMG (Carr-Purcell-Meibom-Gill) se base sur ce phénomène pour pouvoir étudier la dynamique dans des systèmes subissant de l'échange en régime intermédiaire. Des impulsions de relocalisation à 180° sont envoyées à une fréquence beaucoup plus grande que $1/k_{\text{ex}}$, ainsi l'aimantation qui échange entre les états A et B peut être complètement refocalisée. Les données de dispersion de la relaxation peuvent le plus souvent être modélisées par un échange à deux sites selon les équations de Carver-Richards et ainsi permettre d'obtenir les déplacements chimiques de l'état B ou encore les constantes cinétiques et la thermodynamique de l'échange.

Une autre technique utilisée durant cette thèse est le transfert de saturation par échange chimique (CEST, Chemical Exchange Saturation Transfer). Elle permet d'obtenir des informations sur les échanges dont l'échelle de temps est proche de celle des expériences de dispersion de la relaxation CPMG tout en étant plus applicable aux régimes d'échange plus lents. Lors de cette expérience, un faible champ B_1 est appliqué à la fréquence $\omega(B_1)$ pour perturber la magnétisation selon z de l'état B. La perturbation est ensuite transférée à l'état A par l'échange ce qui rend celle-ci détectable dans l'intensité du signal de A. Lors d'un échange à deux états ($A \leftrightarrow B$), l'expérience CEST permet d'obtenir la différence de déplacement chimique entre A et B, la cinétique de l'échange, les populations des deux états ainsi que la différence de relaxation transversale entre les états A et B.

8.4 Caractérisation des structures résiduelles des PID par RMN

Le déplacement chimique d'un noyau dépend de son environnement chimique et électronique immédiat, des interactions à moyenne et longue distance, et fournit donc une riche source d'informations structurales. Une procédure simplifiée pour analyser les déplacements chimiques des PDIs en terme de structure secondaire est le calcul des déplacements chimiques secondaires. Dans cette approche, le déplacement chimique *random coil* est soustrait des déplacements chimiques expérimentaux. Deux approches sont communément utilisées pour obtenir les déplacements chimiques *random coil*: mesurer les déplacements chimiques de courts peptides, ou utiliser des bases de données de structures de protéines attribuées. Les déplacements chimiques secondaires ainsi obtenus sont analysés en détails et peuvent être interprétés en termes d'angles ϕ et ψ . La relation entre les déplacements chimiques et les angles ϕ et ψ est établie en utilisant des bases de données de structures de protéines avec leur attribution. Dans cette partie, je présente donc une de ces approches: SPARTA.

Les couplages dipolaires résiduels (RDCs) peuvent être utilisés pour étudier l'échantillonnage conformationnel des PDIs. Pour mesurer les RDCs, une légère orientation préférentielle de la protéine en solution est nécessaire. Pour cela, la protéine est placée dans un milieu qui contraint l'échantillonnage conformationnel de la protéine (par exemple stériquement), de façon à ce que toutes les orientations n'aient pas la même probabilité d'exister. L'orientation moléculaire préférentielle imposée d'une protéine alignée peut être décrite par un tenseur d'alignement. Les RDCs peuvent ensuite être interprétés en termes d'orientation des vecteurs internucléaires par rapport au tenseur d'alignement général. Dans cette partie, je présente la façon de calculer le tenseur d'alignement, les contributions aux motifs de RDCs particuliers observés dans les protéines désordonnées, et la façon dont les structures locales et longues distances affectent les RDCs.

8.5 Calcul d'ensemble pour les PDIs à partir de paramètres RMN

Dans ce chapitre, j'expose tout d'abord comment les structures locales des PDIs peuvent être décrites par la distribution des angles diédraux de leurs acides aminés et quels paramètres RMN peuvent être utilisés pour caractériser les propensions conformationnelles locales des PDIs. Dans une certaine mesure, j'ai déjà présenté ces notions dans la partie à propos des déplacements chimiques et des RDCs. Ensuite, je présente l'approche «sample and select» pour le calcul

d'ensemble. L'espace conformationnel disponible doit être entièrement couvert dans la première étape d'échantillonnage. De façon évidente, les conformères qui n'ont pas été générés dans la première étape ne pourront être sélectionnés pour la seconde, même s'ils font partie de la solution. Les PDIs sont des molécules caractérisées par une surface d'énergie libre plate, qui couvre beaucoup de conformations différentes. Échantillonner suffisamment l'espace conformationnel est donc un défi majeur. L'étape d'échantillonnage est suivie d'une étape de sélection, durant laquelle l'ensemble qui est en accord avec les données expérimentales est sélectionné. Dans cette partie, je discute l'*over-* et l'*under-restraining* et les façons de s'assurer de la validité de l'ensemble structural obtenu. La méthode d'ensemble de référence, avec laquelle une procédure de sélection peut être validée et calibrée, sera présentée. L'exemple de N_{TAIL} de la nucléoprotéine du virus de Sendai est utilisé pour illustrer la façon dont les ensembles avec des structures secondaires se formant avec coopérativité peuvent être calculés. Une autre question importante concernant le calcul d'ensemble est la taille de l'ensemble. Ici, je discute comment un ensemble doit être suffisamment petit pour éviter l'*overfitting*, mais suffisamment grand pour prendre en compte les propriétés de convergence des paramètres RMN utilisés dans les sélections. La dernière partie de ce chapitre est dédiée à la validation des ensembles par *cross-validation*. Dans cette approche, on exclut certaines des données expérimentales de la procédure de sélection. Ces données sont utilisées une fois que l'ensemble final a été calculé pour évaluer si les données prédites à partir de l'ensemble sont en accord avec les données exclues de la sélection.

8.6 Résultats

Dans ce chapitre, je présente les résultats de la caractérisation des régions intrinsèquement désordonnées (IDRs) en N-terminal des protéines MKK7 et MKK4, ainsi que l'étude de leur interaction avec JNK1.

MKK7 est la seule MKK humaine pour laquelle plus d'un site de liaison putatif est présent dans l'IDR en N-terminal. Un aperçu structural de l'échantillonnage conformationnel de l'IDR N-terminale a été obtenu en utilisant les déplacements chimiques, les RDCs et la relaxation. L'approche de sélection d'ensemble a été utilisée pour déterminer l'échantillonnage conformationnel des trois sites d'interaction de MKK7 dans leur état pré-reconnaissance, identifiant un échantillonnage conformationnel distinct dans le cas des trois sites. Une structure alpha hélicoïdale transitoire précédant les trois sites a été observée en N-terminal de l'IDR. Les données de déplacements chimiques, RDCs et relaxation ont également été obtenues pour l'IDR N-terminale

de MKK4. L'approche de sélection a été utilisée pour déterminer l'échantillonnage conformationnel du site d'interaction de MKK4, révélant un échantillonnage élevé de PPII (polyproline II) dans la région du site d'interaction.

Pour avoir un meilleur aperçu du mécanisme de transduction du signal entre les MKKs et JNK, l'interaction des IDRs N-terminales de MKK7 et MKK4 avec JNK1 a été étudiée par titration en RMN. Ces expériences ont aidé à délimiter les résidus des sites d'interactions de MKK7 et MKK4 qui sont requis pour la liaison à JNK. L'interaction de l'IDR de MKK7 avec JNK1 a aussi été étudiée par dispersion de relaxation et CEST, fournissant une image complète des cinétiques du complexe JNK1-MKK7.

Des mesures d'affinité pour les site d'interaction de MKK7 et MKK4 envers JNK1 ont été obtenues par titration calorimétrique isotherme (ITC), montrant que tous les sites d'interactions sont capables de se lier à JNK1 avec un K_d de l'ordre du micromolaire.

Enfin, nous avons déterminé la structure cristallographique de JNK1 en complexe avec un peptide du deuxième site d'interaction de MKK7. Le peptide lié montre une certaine flexibilité dans le complexe et adopte différentes conformations dans chacune des quatre JNKs présentes dans l'unité asymétrique. Ces données sont en accord avec les résultats de CEST, qui montrent que le second site d'interaction de MKK7 reste flexible dans le complexe, formant un complexe dit « fuzzy ».

8.7 Discussion

Dans cette partie, je résume les résultats et établis les connections entre mon travail de thèse et la recherche effectuée précédemment dans ce domaine par d'autres méthodes.

La région qui contient l'hélice alpha transitoire a précédemment été montrée comme interagissant avec la kinase à deux leucine zipper située en amont, DLK, et avec MEKK2. Je discute la possibilité que l'hélice alpha transitoire soit un élément de reconnaissance moléculaire pour les deux kinases en amont. L'autre question par rapport à la structure transitoire de l'IDR N-terminal des MKKs concerne la nature de l'échantillonnage conformationnel des sites d'interaction. Les quatre sites d'interaction des MKKs que j'ai examinés échantillonnent différents régimes conformationnels. Je compare les propensités conformationnelles des sites d'interaction à l'échantillonnage dérivé de structures cristallographiques de sites d'interactions liés à des MAPKs.

L'ITC, le CEST et la dispersion de relaxation ont été utilisés pour déterminer les affinités des

interactions et étudier les cinétiques de formation des complexes. Je discute les effets possibles des similarités et différences dans l'échantillonnage de l'état lié et non-lié sur l'affinité et la cinétique de liaison. La structure cristallographique de JNK1 en complexe avec un peptide du second site d'interaction de MKK7 a montré que le peptide est flexible dans le complexe et adopte différentes conformations, formant un complexe «fuzzy». La comparaison de la structure résolue ici avec d'autres structures cristallographiques de JNK résolues précédemment montre comment la flexibilité dans le complexe pourrait être liée à la fonction, en régulant le passage de la forme active à la forme auto-inhibée de la kinase.

8. *Résumé en français*

9 Publications

Kragelj J, Ozenne V, Blackledge M, Jensen MR. Conformational propensities of intrinsically disordered proteins from NMR chemical shifts. *Chemphyschem*. 2013;14(13):3034-45.

Sterckx YG, Volkov AN, Vranken WF, Kragelj J, Jensen MR, Buts L, Garcia-Pino A, Jové T, Van Melderen L, Blackledge M, van Nuland NA, Loris R. Small-angle X-ray scattering- and nuclear magnetic resonance-derived conformational ensemble of the highly flexible antitoxin PaaA2. *Structure*. 2014 Jun 10;22(6):854-65

Krishnan N, Koveal D, Miller DH, Xue B, Akshinthala SD, Kragelj J, Jensen MR, Gauss CM, Page R, Blackledge M, Muthuswamy SK, Peti W, Tonks NK. Targeting the disordered C terminus of PTP1B with an allosteric inhibitor. *Nat Chem Biol*. 2014 Jul;10(7):558-66

Schneider R, Maurin D, Communie G, Kragelj J, Hansen DF, Ruigrok R, Jensen MR, Blackledge M. Visualizing the molecular recognition trajectory of an intrinsically disordered protein. submitted to *J Am Chem Soc*

Kragelj J, Palencia A, Nanao M, Damien M, Bouvignies G, Blackledge M, Jensen MR. Structure and dynamics of the MKK7-JNK signalling complex. submitted to *Proc Natl Acad Sci U S A*

9. Publications

DOI: 10.1002/cphc.201300387

Conformational Propensities of Intrinsically Disordered Proteins from NMR Chemical Shifts

Jaka Kragelj, Valéry Ozenne, Martin Blackledge, and Malene Ringkjøbing Jensen^{*[a]}

The realization that a protein can be fully functional even in the absence of a stable three-dimensional structure has motivated a large number of studies describing the conformational behaviour of these proteins at atomic resolution. Here, we review recent advances in the determination of local structural propensities of intrinsically disordered proteins (IDPs) from experimental NMR chemical shifts. A mapping of the local structure in IDPs is of paramount importance in order to understand the molecular details of complex formation, in particular,

for IDPs that fold upon binding or undergo structural transitions to pathological forms of the same protein. We discuss experimental strategies for the spectral assignment of IDPs, chemical shift prediction algorithms and the generation of representative structural ensembles of IDPs on the basis of chemical shifts. Additionally, we highlight the inherent degeneracies associated with the determination of IDP sub-state populations from NMR chemical shifts alone.

1. Introduction

Over the last decade, classical structural biology has experienced a shift towards a more dynamic paradigm with the realization that a protein can be fully functional even in the absence of a stable, folded structure.^[1–7] The discovery of these so-called intrinsically disordered proteins (IDPs) initiated a large number of bioinformatics studies aiming at defining specific primary sequence characteristics. IDPs are in general predicted to contain a higher number of disorder-promoting amino acids of low hydrophobicity and high net charge compared to globular, folded proteins.^[8–10] These identified IDP characteristics now form the basis of a number of available web-servers, which offer the prediction of the disorder tendency for a given protein primary sequence.^[11–16] This has resulted in the estimation that around 40% of the human proteome is intrinsically disordered or contains disordered regions of significant length (> 50 consecutive amino acids). Interestingly, intrinsic disorder has been shown to be particularly abundant in proteins implicated in a number of human diseases such as cancer, neurodegenerative diseases and diabetes,^[17,18] underlining the compelling need for the development of robust methods for characterizing these proteins and their physiological interactions at atomic resolution.

IDPs apply potentially very different molecular recognition mechanisms and functional modes compared to folded, globular proteins.^[19,20] These include the folding-upon-binding mechanism where the IDP undergoes a structural transition from a partially disordered state into a specific conformation in the complex.^[21–24] Other IDPs are capable of folding into different

conformations depending on the partner protein, illustrating the complexity of the different binding modes.^[25] Finally, some IDPs engage in complexes that are much more dynamic, in which the IDPs do not necessarily adopt a specific conformation in the complex, but rather sample various states on the surface of the partner (fuzziness).^[26,27] Although the importance of specific pre-recognition states in IDPs have been highlighted,^[28–31] the link between the intrinsic structural propensities of the protein primary sequence, the kinetics of the complex formation and the final conformation(s) adopted in the complex are not well understood.^[32] For example, transiently populated secondary structures have been shown to be particularly abundant in the interaction sites of IDPs; they may contribute to complex formation through conformational selection. In order to fully understand how IDPs carry out their functions, atomic resolution models of these proteins are necessary both in their isolated states as well as in complexes with their partner proteins.

While the conformational energy landscape of a globular, folded protein contains a well-defined energy minimum corresponding to the three-dimensional structure, which would be determined by, for example, X-ray crystallography, the situation is quite different for IDPs (Figure 1). Their energy landscape is



Figure 1. Hypothetical conformational energy landscape for a globular, folded protein (left) and an intrinsically disordered protein (right).

[a] J. Kragelj, Dr. V. Ozenne, Dr. M. Blackledge, Dr. M. R. Jensen
Protein Dynamics and Flexibility
Institut de Biologie Structurale CEA-CNRS-UJF UMR 5075
41, rue Jules Horowitz
38027 Grenoble (France)
E-mail: malene.ringkjobering-jensen@ibs.fr

relatively flat with a number of local minima characterized by comparable energies, meaning that an ensemble of inter-converting conformers is required to describe their conformational behaviour. The determination of representative ensembles that accurately capture both the local and long-range behaviour of a poly-peptide chain is one of the major challenges of contemporary structural biology. Thus, the inherent flexibility of IDPs calls upon new experimental approaches as well as computational strategies for studying these proteins.

Nuclear magnetic resonance (NMR) spectroscopy has evolved into one of the most powerful techniques for studying IDPs at atomic resolution.^[33–40] In this context, chemical shifts are the most readily accessible NMR parameters and provide ensemble-averaged local structural information about all conformations sampled by the protein up to the millisecond time scale. Here, we review recent advances in the determination of local structural propensities of IDPs from experimental NMR chemical shifts. A mapping of local structure in IDPs is of paramount importance in order to understand the molecular details of complex formation, in particular in IDPs that fold upon binding or undergo structural transitions to pathological forms of the same proteins. We discuss experimental strategies for the spectral assignment of IDPs, chemical shift prediction algorithms and the generation of representative structural ensembles of IDPs on the basis of chemical shifts. Additionally, we highlight the inherent degeneracies associated with the determination of IDP sub-state populations from NMR chemical shifts alone.

2. NMR Spectral Assignment of IDPs

NMR spectra of IDPs are characterized by a low ^1H signal dispersion reflecting the similar electronic environment experienced by the different protons (Figure 2). This would normally result in an increased spectral overlap; however, the low signal dispersion is to some extent counterbalanced by the narrower line widths of the NMR signals of IDPs due to their inherent dynamic nature. Typical local correlation times of IDPs derived from nuclear spin relaxation rates range from approximately two to six nanoseconds at ambient temperature giving them the favourable NMR characteristics of small molecules.^[41–43] The ^{15}N and ^{13}C nuclei display a much larger chemical shift disper-

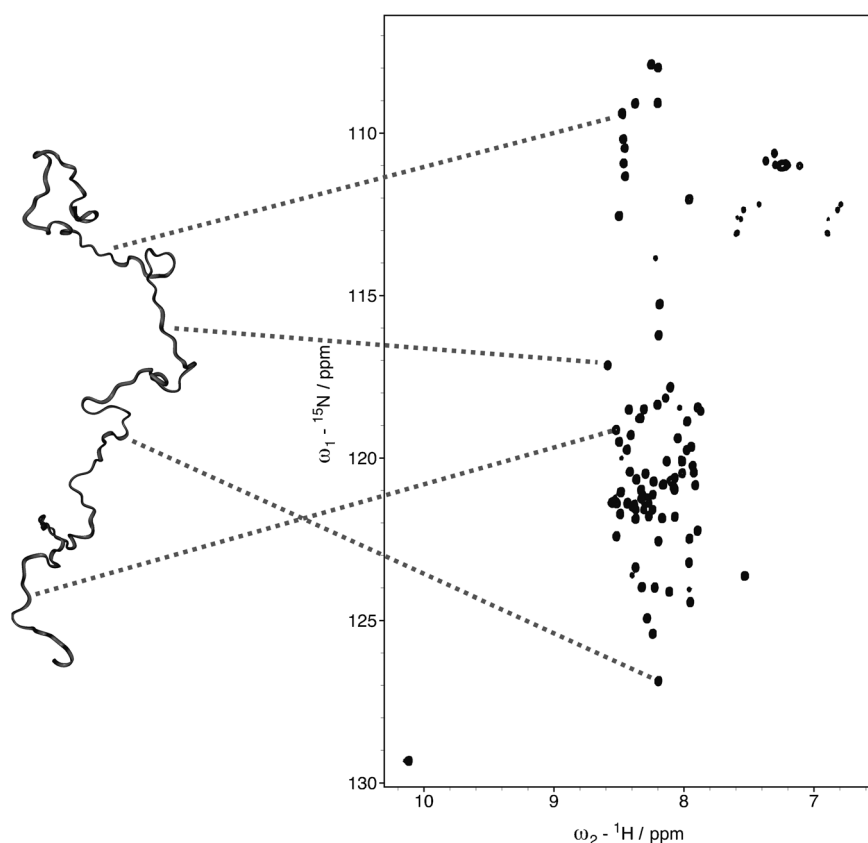


Figure 2. ^1H - ^{15}N heteronuclear single quantum coherence (HSQC) spectrum of the intrinsically disordered C-terminal domain of the Sendai virus nucleoprotein showing the low signal dispersion in the ^1H dimension characteristic of an IDP. NMR spectroscopy provides site-specific information as each resonance in the HSQC spectrum corresponds to a specific amide group in the protein.

sion compared to ^1H nuclei due to their stronger dependence on the amino acid type.

Assignment of the NMR spectra is a pre-requisite for any further study of the residual structure and dynamics of IDPs at atomic resolution. In recent years, a number of NMR experiments specifically adapted to IDPs have been proposed with the aim of maximizing the spectral resolution and/or the signal-to-noise ratio per unit time. In this context, multi-dimensional experiments based on the band-selective excitation of $^1\text{H}^{\text{N}}$ nuclei are promising. Selective excitation enhances the longitudinal relaxation rates of the $^1\text{H}^{\text{N}}$ nuclei through dipole-dipole interactions with the aliphatic protons, thereby allowing the use of short inter-scan delays.^[44–48] Efficient acquisition of NMR data can also be achieved by using sparse sampling techniques allowing multiple dimensions to be recorded in a short amount of time.^[49–54]

The large solvent accessibility of IDPs often diminishes the signal intensities of $^1\text{H}^{\text{N}}$ nuclei, and direct detection of $^1\text{H}\alpha$ nuclei can alleviate this problem. In addition, IDPs are often rich in proline residues, which render the straightforward assignment of the poly-peptide chain difficult by using standard triple-resonance experiments based on $^1\text{H}^{\text{N}}$ detection. A range of experiments that detect $^1\text{H}\alpha$ protons for the assignment of IDPs have been proposed, which allow sequential connectives to be obtained across proline residues.^[55–57] Similar advantages

are obtained from experiments that directly detect ^{13}C , which in addition significantly enhance the spectral resolution by relying only on the resonance frequencies of heteronuclei.^[58–64]

3. Chemical Shifts and Their Dependence on Backbone Conformation

Chemical shifts are exquisitely sensitive to the backbone conformation^[65–68] and the possibility of exploiting this sensitivity to determine conformational propensities of IDPs is particularly attractive.^[69] Remarkable progress has been made in recent years in the prediction of chemical shifts from protein conformations using a number of different approaches. The predictor SHIFTS^[70] relies on an assembled database of chemical shifts obtained by using density functional theory calculations of peptides with backbone dihedral angles covering different regions of the Ramachandran space. SHIFTY^[71] exploits sequence homology with proteins in the BioMagResBank (BMRB), which contains sequences and NMR chemical shifts of a large number of assigned proteins.^[72] SHIFTX^[73] relies on empirical chemical shift hypersurfaces derived from high-resolution crystal structures for which chemical shift assignments are available, while the more recent version SHIFTX2^[74] employs the concepts of both SHIFTX and SHIFTY. SPARTA^[75] and its successor SPARTA+^[76] as well as CamShift^[77] also rely on a database of crystal structures with known NMR chemical shifts. While SPARTA uses the database as a lookup table by comparing amino acid type, backbone and side-chain dihedral angles of tri-peptides in the target protein to entries in the table, CamShift uses the database to parameterize a set of equations linking chemical shift values to inter-atomic distances. CamShift has the advantage of providing a differentiable function necessary for the inclusion of chemical shift restraints in molecular dynamics (MD) simulations. It was recently noted that predicted chemical shifts of folded proteins averaged over long MD simulations reproduce experimental chemical shifts significantly better than a single calculation of chemical shifts from a static structure.^[78–80] In this context, the chemical shift predictor PPM^[81] was recently proposed, which explicitly accounts for the effect of protein dynamics on chemical shifts by refining the predictor against long MD simulations of a set of proteins with known chemical shifts.

The advances made in chemical shift prediction from protein structures have culminated in the successful structure determination of small folded proteins (<150 amino acids) by using only chemical shifts in combination with molecular mechanics force fields.^[82,83]

The advantage of chemical shifts as structural probes is that they depend in a different, complementary manner on the backbone dihedral angles and therefore, in principle, allow a site-specific mapping of the conformational sampling of IDPs. This is illustrated in Figure 3, which shows predicted chemical shifts using SPARTA for a poly-alanine chain sampling various regions of the Ramachandran space. It is seen that $^{13}\text{C}\alpha$ and C' chemical shifts exhibit the largest and smallest values in α -helical and β -strand conformations, respectively, while $^{13}\text{C}\beta$ and $^1\text{H}\alpha$ chemical shifts display an almost inverse

dependence on the ϕ/ψ distribution. Interestingly, both the poly-proline II (PPII) region and the upper-left half of the α -helical region are characterized by intermediate $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$, $^{13}\text{C}'$ and $^1\text{H}\alpha$ chemical shifts, which become essentially indistinguishable from an average over the entire Ramachandran space (i.e. the random coil state).

We also note that while $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$, $^{13}\text{C}'$ and $^1\text{H}\alpha$ chemical shifts depend strongly on ϕ and ψ of the same residue, ^{15}N and $^1\text{H}^{\text{N}}$ chemical shifts show a more or less uniform dependence on the two dihedral angles. Only ^{15}N and $^1\text{H}^{\text{N}}$ chemical shifts of the following residue ($i+1$) report on the ϕ/ψ sampling of residue i . On the basis of the SPARTA prediction (Figure 3), experimental ^{15}N and $^1\text{H}^{\text{N}}$ chemical shifts are necessary in order to detect enhanced PPII conformations in proteins. Some variation of this observation between different predictors can, however, be expected.

4. Random Coil Chemical Shifts

In the simplest application of chemical shifts to reveal structural propensities in IDPs (or globular, folded proteins), experimental chemical shifts (δ_{EXP}) are directly compared to a set of so-called random coil shifts (δ_{RC}).^[84,85]

$$\delta_{\text{SCS}} = \delta_{\text{EXP}} - \delta_{\text{RC}}$$

The random coil shifts are amino acid specific and represent the expected chemical shifts for a protein devoid of secondary and tertiary structure. The resulting secondary chemical shifts (SCS) are positive or negative depending on the nucleus type and the region of the Ramachandran space preferentially being populated (Figure 3). The set of random coil values employed for comparison with the experimental data should somehow match the conditions under which the experimental data were acquired. Therefore, different sets of random coil values have been proposed and are summarized in Table 1.

One common approach for obtaining random coil values is to measure the chemical shifts of each of the twenty amino acids within small peptides under non-denaturing,^[86] mildly denaturing^[87,88] or denaturing conditions.^[89–91] Another approach is to use experimentally determined chemical shifts from proteins of known three-dimensional structures and derive average, characteristic chemical shifts corresponding to loop regions (random coil), α -helix and β -sheet.^[92–95] In both approaches sequence-dependent correction factors have been derived and nearest neighbour effects have been studied (Table 1).

Recently, random coil shifts have been derived specifically for IDPs using experimentally determined chemical shifts for fourteen IDPs.^[96] Using singular value decomposition, random coil shifts as well as sequence-dependent left and right neighbour correction factors were derived from the experimental chemical shifts.

A complementary approach for obtaining random coil values is to denature the protein under investigation, for example, by using high concentrations of urea.^[97] This provides a set of coil chemical shifts that is intrinsic to the protein stud-

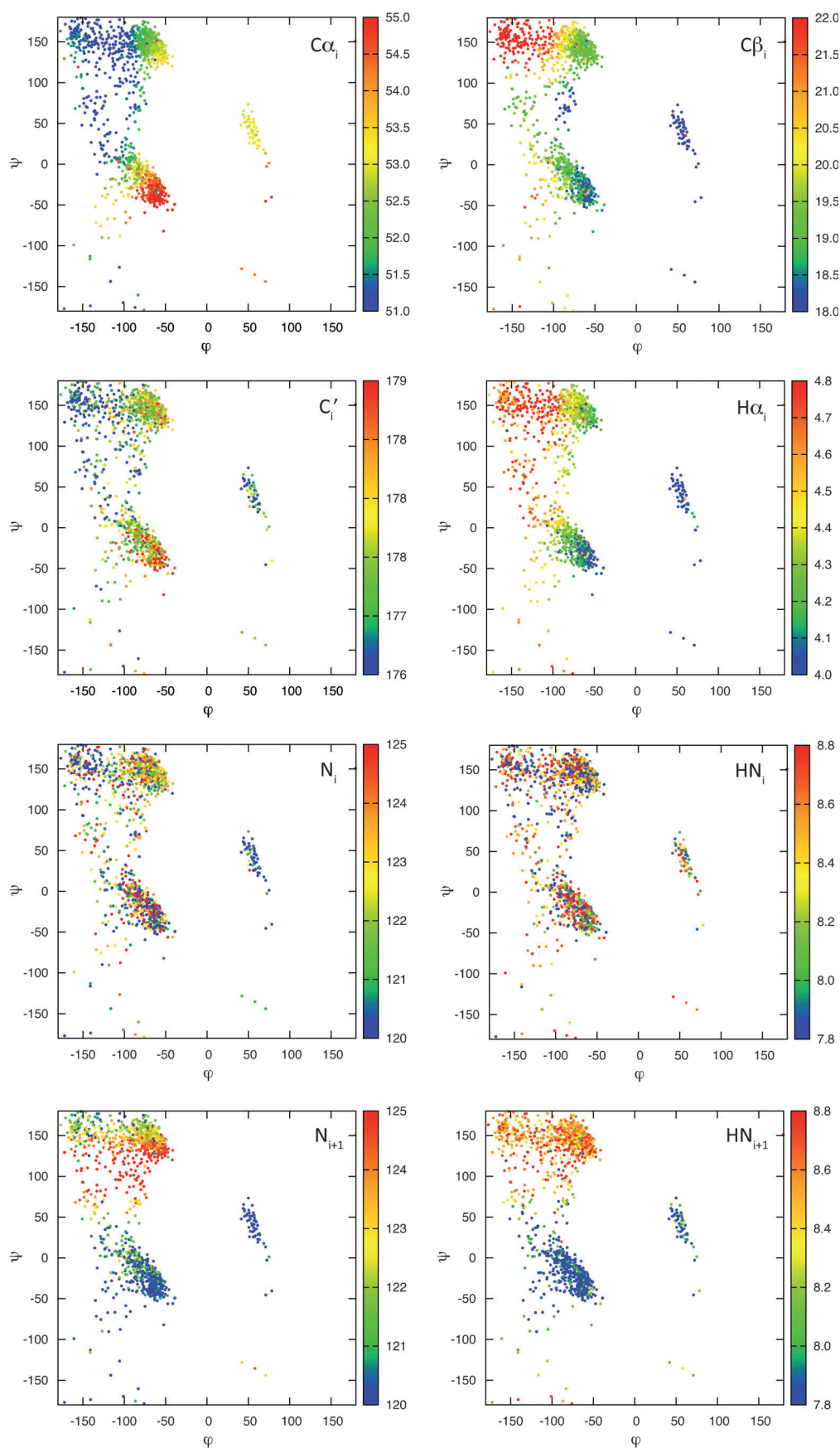


Figure 3. Distribution of predicted chemical shifts using SPARTA of an Ala–Ala_{*i*-1}–Ala_{*i*}–Ala_{*i*+1}–Ala peptide as a function of the backbone dihedral angles of residue *i*. Chemical shift predictions are shown for residue *i* (C α , C β , C', H α , N and H^N) and residue *i* + 1 (N and H^N). Results are shown for 2000 statistical coil conformers of the peptide generated using Flexible-Meccano. Adapted with permission from ref. [139]. Copyright (2012) American Chemical Society.

Table 1. Overview of available random coil chemical shift datasets.

Origin of random coil values	Experimental conditions	Comments	Reference
Ac-G-G-X-A-G-G-NH ₂ Ac-G-G-X-P-G-G-NH ₂	pH 5.0, 1 M urea T = 25 °C	Nearest neighbour effects are investigated, however, no systematic implementation of sequence-dependent correction factors, except for residues followed by proline.	[87]
Ac-G-G-X-G-G-NH ₂	pH 2.3, 8 M urea T = 20 °C	Suitable for comparison to experimental chemical shifts in denatured proteins. Sequence-dependent correction factors are derived for all residue types.	[89, 90]
PSSI: Derived from three-dimensional (3D) structures and their corresponding chemical shifts	Various conditions: pH = 6.0 (T) = 29 °C	Sequence and structure-dependent corrections factors are determined through an iterative procedure on the entire experimental dataset.	[92, 93]
RefDB: Derived from 3D structures and their corresponding chemical shifts	Various conditions corresponding to the entries in the BMRB	Amino acid specific chemical shifts are determined for the random coil state and fully populated α -helices and β -sheets.	[94]
CamCoil: Derived from loop regions of 3D structures and their corresponding chemical shifts	Various conditions: pH = 6.5 (T) = 21 °C	Sequence-dependent correction factors are optimized against experimental chemical shifts from a set of denatured proteins.	[95]
ncIDP: Derived from experimental chemical shifts of fourteen IDPs	Various conditions: pH = 6.5 (T) = 14 °C	Sequence-dependent correction factors are obtained by singular value decomposition of a small IDP library.	[96]
Ac-G-G-X-G-G-NH ₂	pH 6.5, 1 M urea T = 5, 15, 25, 35 and 45 °C	Neutral pH matches the experimental conditions for most IDPs. Temperature coefficients of random coil chemical shifts and pH dependence of histidine shifts are reported.	[88]
Ac-Q-Q-X-Q-Q-NH ₂	pH 6.5 T = 5 °C	Neutral pH and non-denaturing conditions. Glutamine-based peptide allows derivation of more accurate sequence-dependent correction factors.	[86]
Ac-A-A-X-A-A-NH ₂	pH 2.5, 8 M urea T = 25 °C	Suitable for comparison to experimental chemical shifts in denatured proteins. Alanine-based peptide allows derivation of more accurate sequence-dependent correction factors.	[91]

ied, however, does not take into account the effect that the presence of urea may have on chemical shifts, for example, from hydrogen bonding to the backbone of the protein.^[98, 99]

5. Mapping of Conformational Propensities from Experimental Chemical Shifts

Different approaches have been developed for the determination of conformational propensities in IDPs directly from experimental chemical shifts. One approach relies on the calculation of expected chemical shifts from defined regions of the Ramachandran space. For example, RefDB^[94] provides amino acid specific chemical shifts for the random coil state and fully populated α -helices and β -sheets. The experimental chemical shifts are then used to solve for populations of these regions.

In the earliest application of this approach, Forman-Kay and co-workers proposed the secondary structure propensity algorithm (SSP), which provides populations of α -helices and β -strands (or extended conformations) along the primary sequence of IDPs.^[100] SSP is most accurate if more than one type of chemical shift is included in the calculation. The combination of C α and C β chemical shifts is particularly powerful due

to their inverse dependence on α -helix and β -strand populations, and the SSP algorithm allows a simultaneous determination of potential reference offsets when these two types of chemical shifts are included. The SSP algorithm was recently refined in order to explicitly account for the differential sensitivity of different types of chemical shifts to secondary structures.^[101] The algorithm, ncSPC (neighbour-corrected structural propensity calculator), uses the random coil database derived from experimental chemical shifts of IDPs and the corresponding neighbour corrections^[96] to obtain populations of α -helix and β -extended conformations.

From complementary experimental techniques such as vibrational spectroscopy and circular dichroism, it has been proposed that PPII is one of the dominant conformations in IDPs.^[102–104] A method that unequivocally maps the population of PPII conformations in disordered proteins on a residue-specific basis is thus of great interest. Therefore, the basic ideas of SSP and ncSPC was recently extended with the aim of mapping in addition the population of PPII conformations.^[105] The algorithm, δ 2D, uses derived chemical shifts characteristic of random coil, α -helix, β -strand and PPII conformations and solves for their populations using experimental chemical

shifts.^[105] The populations of the different regions of the Ramachandran space can only be accurately determined if the chemical shifts associated with each region are distinct at least for one, but preferably for a number of nuclei. For this reason the populations obtained using $\delta 2D$ are necessarily more accurate for α -helices and β -strands, while some degeneracy is expected between the determined populations of random coil and PPII (Figure 3). In general, the approaches SSP, ncSPC and $\delta 2D$ offer very fast determination of structural propensities from experimental chemical shifts, but strongly depend on which chemical shift values are considered characteristic for the different regions of the Ramachandran space and how precisely one defines these distinct regions. Optimally the entire Ramachandran space would be filled with regions with attributed characteristic shifts, obviating the need for an additional residual random coil term to account for sampling outside these regions. This is in general not the case, and an additional evocation of the population of this poorly defined average value is still required in order to quantify the population density of the Ramachandran space for each amino acid.

6. Ensemble Descriptions of IDPs from Experimental Chemical Shifts

6.1. Ensemble Selection Using TraDES and ENSEMBLE

A more elaborate way of analysing chemical shifts in terms of conformational propensities of IDPs is to invoke a structural ensemble description. Here, the basic idea is to generate a representative ensemble of the IDP on the basis of the experimental chemical shifts and subsequently analyse the sub-state populations within this ensemble. One common way of achieving this is to select sub-ensembles in agreement with experimental data from a large pool of conformers that is assumed to efficiently sample the conformational space available to the protein under investigation. Statistical coil models have been shown to be predictive of experimental NMR data acquired on IDPs and chemically denatured proteins and are therefore considered as a good starting point for the generation of the initial pool of structures.^[106–111]

A number of groups have contributed to the ensemble description of IDPs on the basis of experimental chemical shifts and other NMR parameters. Forman-Kay and co-workers have developed the algorithm ENSEMBLE^[112–114] that assigns weights to a pool of conformers obtained by using the statistical coil generator TraDES^[115] in order to maximize the agreement with experimental data. ENSEMBLE allows the incorporation of several different NMR parameters in the ensemble selection such as chemical shifts, residual dipolar couplings (RDCs) and paramagnetic relaxation enhancements (PREs), as well as data from complementary biophysical techniques such as small angle X-ray scattering (SAXS).^[116] ENSEMBLE has been used to characterize the conformational sampling of IDPs both in their isolated state as well as in complex with their physiological partners.^[117–119]

6.2. Ensemble Selection Using Bayesian Weighting

Another approach for obtaining ensemble descriptions of IDPs is based on the Bayesian weighting.^[120,121] In this approach, Stultz and co-workers propose to generate an initial pool of structures using replica exchange MD simulations on extended copies of the protein and/or on smaller fragments of the protein that are subsequently assembled in order to enhance the sampling diversity. The main limitation of the Bayesian weighting approach is that it is not feasible to assign weights to a large number of conformers, as a probability distribution over all possible ways of weighting the conformations has to be calculated. Therefore, it is necessary to reduce the starting pool to a small conformational library (typically containing 300–600 conformers) that is assumed to be representative of the protein. Structures for the conformational library are selected based on energy and structural diversity ensuring the presence of different secondary structures. This pruning step is probably critical as no conformations other than those in the library can contribute to the final ensemble. The results will inevitably depend on how the pruning is carried out. Once the conformational library has been chosen, Bayesian weights are assigned to each conformation of the library on the basis of the experimental NMR data. We note that in practice the Bayesian approach allows for a uniform shift of each chemical shift type to best fit the experimental data, which may result in inaccuracies in the local conformational sampling in the final ensemble, as this procedure can mimic systematic population shifts in the Ramachandran space (Figure 3).

Several IDPs involved in neurodegenerative diseases such as Tau,^[120] A β 40/42^[122] and α -synuclein^[123] have been analysed using the Bayesian approach on the basis of experimental chemical shifts and RDCs. Interestingly, the disordered nature of α -synuclein was recently questioned, as it was proposed to exist predominately as a tetramer in vivo with a substantial amount of helical structure.^[124] These observations were followed by the presentation of a model of a helical, tetrameric form of α -synuclein on the basis of sparse NMR data.^[125] These results have stimulated a discussion in the literature concerning the native state of α -synuclein, for which a number of groups recently confirmed the disordered nature of the protein as evidenced by comprehensive biochemistry and in-cell NMR.^[126–128] In this context, chemical shifts^[125] of α -synuclein were recently analysed using the Bayesian weighting approach.^[129] In this case, the initial structural library comprised both monomeric α -synuclein as well as trimeric and tetrameric structures obtained either from the NMR data or from threading a helix between residues 52 and 64 of α -synuclein to the crystal structure of myosin (trimer) or ferritin (tetramer). According to the Bayesian analysis, the trimeric and tetrameric structures occur with probabilities of 30–42% compared to 58–70% for the monomeric species. The analysis implicitly assumes rapid inter-conversion between conformers, implying unfolding and refolding of monomeric and multimeric species on timescales equal to or faster than tens of microseconds. The results are surprising as the NMR spectrum would normally be dominated by the monomeric forms due to their enhanced

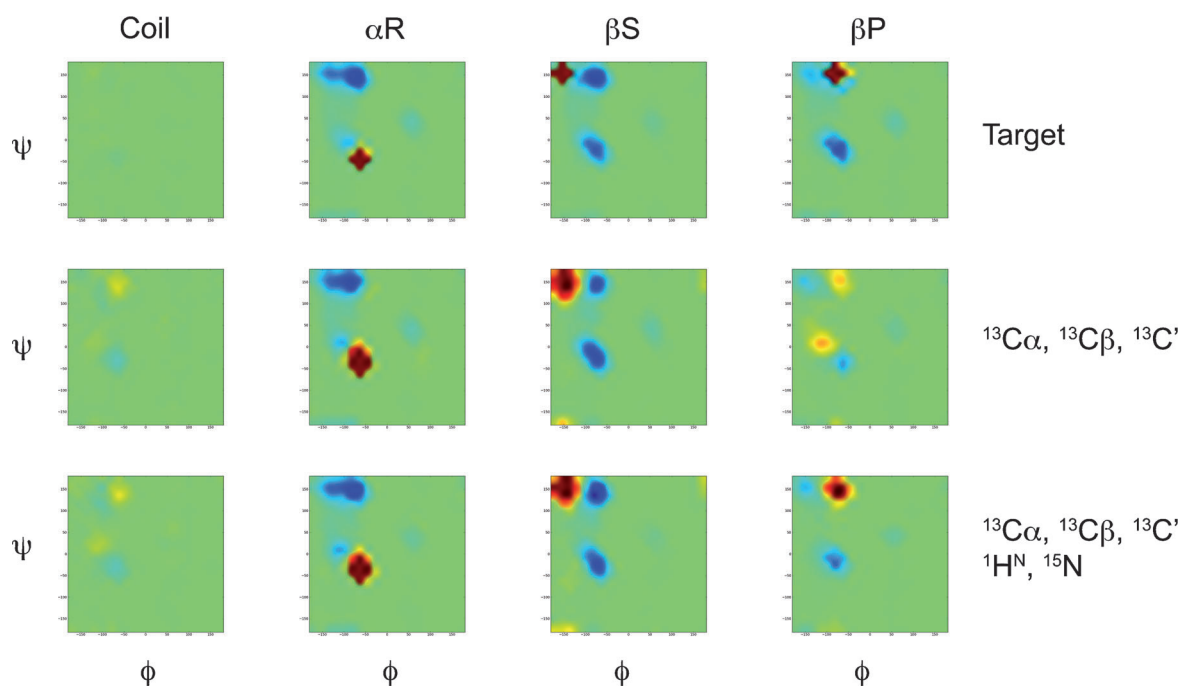


Figure 4. Mapping of local conformational sampling of IDPs based on experimental chemical shifts. An ensemble of a model protein of 100 amino acids was constructed using Flexible-Meccano with three regions of five amino acids each over-sampling the α -helical (α R), β -strand (β S) and PPII (β P) regions. Ensemble-averaged chemical shifts were calculated and targeted by the ensemble selection algorithm ASTEROIDS. Average Ramachandran plots are shown for the target sampling in the different regions of the protein (top line), resulting sampling from ASTEROIDS selections targeting ^{13}C chemical shifts only (middle line) and all chemical shifts (bottom line). Red and blue colours correspond to oversampling and undersampling, respectively, compared to the statistical coil. Adapted with permission from ref. [139]. Copyright (2012) American Chemical Society.

flexibility compared to their multimeric counterparts. In support of this, it was noted recently that the ^1H - ^{15}N heteronuclear single quantum coherence (HSQC) spectrum of the proposed tetrameric species superimposes within the experimental error on the spectrum of monomeric α -synuclein from *E. coli*.^[126]

6.3. Ensemble Selection Using Flexible-Meccano and ASTEROIDS

Another method for obtaining representative structural ensembles of IDPs has been proposed by Blackledge and co-workers. Their approach relies on the initial creation of a large pool of statistical coil conformers using the algorithm Flexible-Meccano.^[109,130,131] The approach is conceptually simple, creating conformers using an amino acid specific random coil backbone dihedral angle database derived from loop regions of high resolution crystal structures. Pairs of ϕ/ψ angles are selected randomly in the database and peptide planes are constructed accordingly while, at the same time, avoiding steric clashes with previously built planes. Typically, 50 000 structures (PDB files) can be generated in a few hours on a single processor. For each conformer in the pool, NMR parameters such as chemical shifts, RDCs and PREs are calculated.

Representative ensembles of a given IDP are then selected from the pool of statistical coil conformers using the algorithm ASTEROIDS (a selection tool for ensemble representations of intrinsically disordered states) targeting experimental NMR

data.^[132] The selection of conformers follows the evolution scheme of a genetic algorithm, in which generations are obtained by random selection, mutations and crossings of conformers. ASTEROIDS has been used to generate an ensemble of urea-denatured ubiquitin on the basis of multiple RDCs^[132,133] and ensembles of the IDPs α -synuclein and Tau on the basis of PREs in order to map long-range interactions within the proteins.^[134–136]

More recently, chemical shifts were included allowing a detailed mapping of the local conformational sampling in IDPs.^[137,138] In this context, we note that it is practically impossible to select a representative sub-ensemble of an IDP containing significant amounts of secondary structure directly from a pool of statistical coil conformers, as the probability of finding continuous stretches of secondary structure is too low. For this reason, an iterative procedure was implemented, in which several ensembles (typically five ensembles of 200 conformers each) are initially selected using ASTEROIDS from the statistical coil pool. The backbone dihedral angles are then extracted for each residue from the selected conformers and used to create a new pool of conformers using Flexible-Meccano. Each time this procedure is repeated the amount of secondary structure in the re-generated pool increases until convergence is achieved at the point for which the agreement between experimental and back-calculated chemical shifts is optimal.

Using this approach, the accuracy with which chemical shifts alone can map local conformational sampling in IDPs was re-

cently investigated using extensive simulations.^[139] A coil ensemble of a 100 amino acid protein of arbitrary sequence was generated using Flexible-Meccano with three specific regions of five amino acids each over-sampling the α -helical, β -sheet and PPII regions (50% enhanced sampling compared to the statistical coil). Chemical shifts were calculated for each conformer using SPARTA and subsequently averaged over the entire ensemble. This synthetic chemical shift data set was then used as target for ASTEROIDS starting from a pool of statistical coil conformers. The simulations show that if only ^{13}C chemical shifts ($^{13}\text{C}\alpha$, $^{13}\text{C}\beta$ and $^{13}\text{C}'$) are included in the ensemble selection, the populations in the α -helical and β -strand regions are correctly determined. The enhanced population in the PPII region of the Ramachandran space is, however, not well determined and an inherent degeneracy exists between the upper half of the α -helical region and the PPII region (Figure 4). The addition of $^1\text{H}^{\text{N}}$ and ^{15}N in the ensemble selection improves

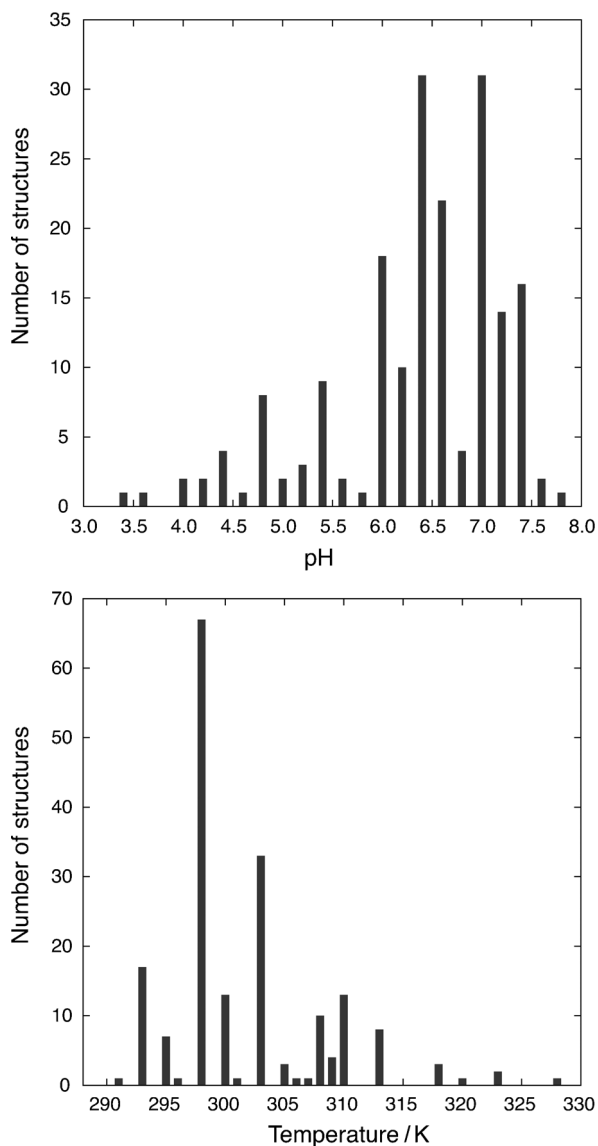


Figure 5. Distribution of pH and temperature in the NMR assignment experiments of the proteins constituting the SPARTA database.

this situation considerably and allows a mapping of all three enhanced propensities (Figure 4). However, it should be mentioned that $^1\text{H}^{\text{N}}$ and ^{15}N chemical shifts depend more on exper-

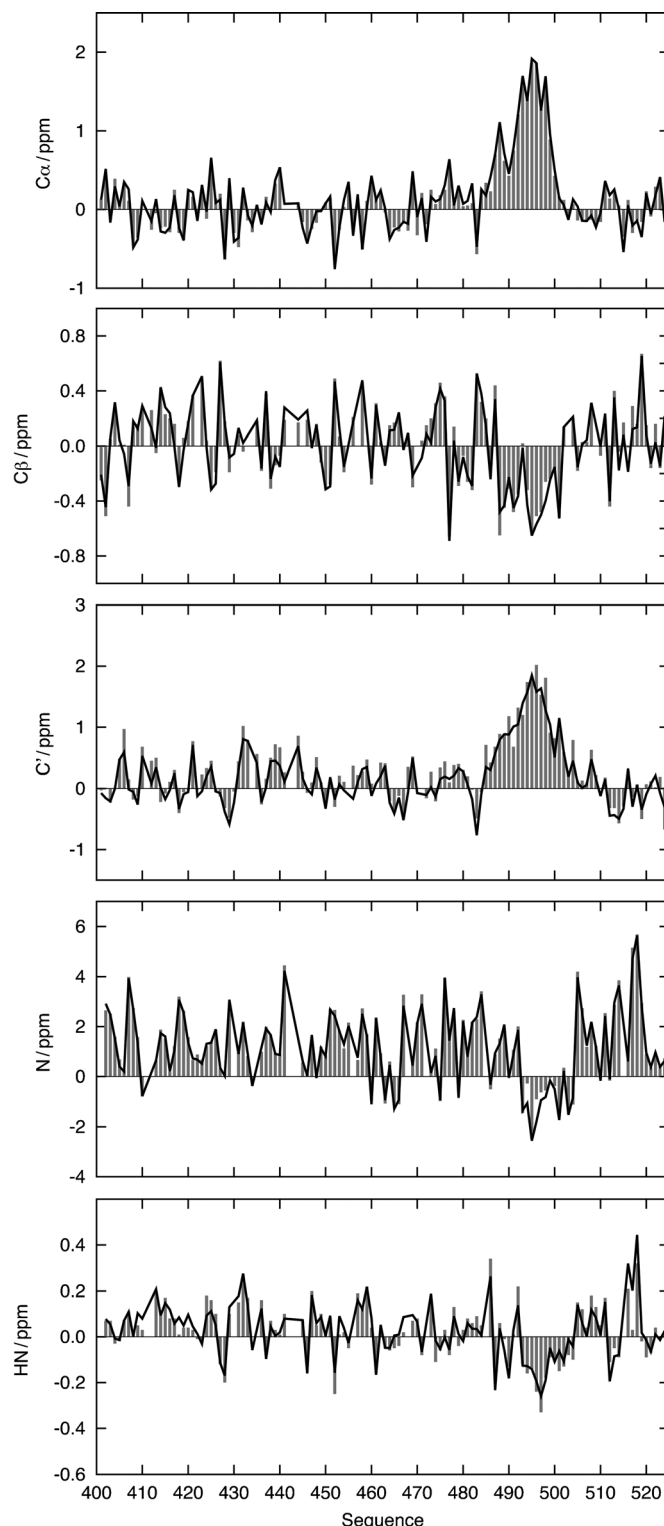


Figure 6. Experimental (gray bars) and back-calculated (black line) secondary chemical shifts from a selected ASTEROIDS ensemble of the intrinsically disordered C-terminal domain, N_{TAIL} , of the measles virus nucleoprotein. Random coil values from RefDB were used for calculating the secondary chemical shifts.

imental conditions such as temperature, ionic strength and pH compared to ^{13}C chemical shifts rendering them potentially volatile in terms of conformational mapping.^[88] Figure 5 shows the distribution of pH and temperature used in the assignment experiments for the proteins included in the SPARTA database. Mapping of conformational sampling using SPARTA as a predictor will, therefore, be most accurate for experimental chemical shifts obtained close to neutral pH and around room temperature.

As an example, ASTEROIDS was used to target experimental chemical shifts obtained at 25 °C and pH 6.5 of the intrinsically disordered C-terminal domain, N_{TAIL} , of the measles virus nucleoprotein.^[140] Ensembles comprising 200 conformers each were obtained and the ensemble-averaged chemical shifts

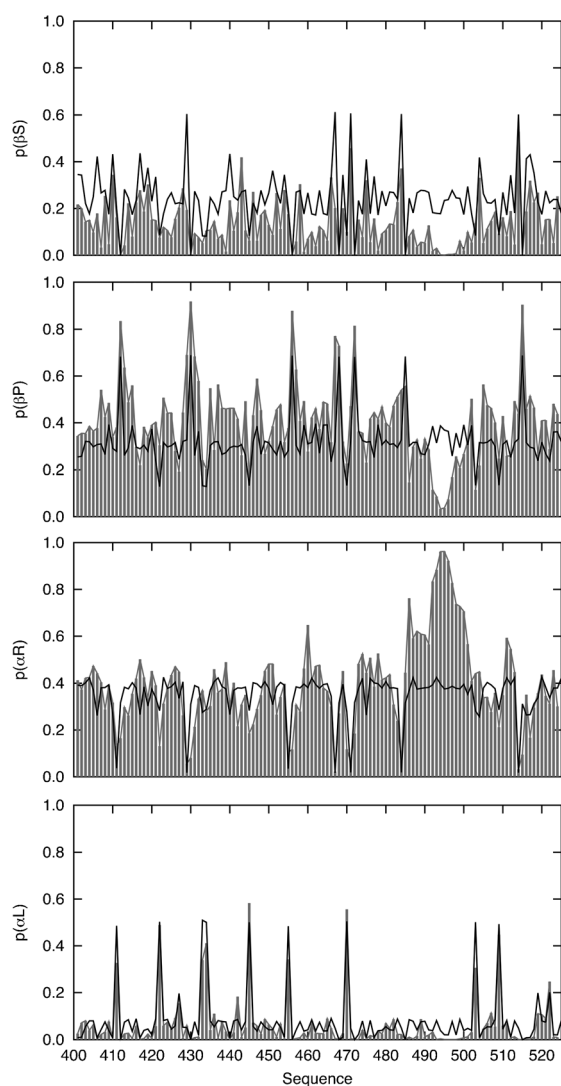


Figure 7. Conformational sampling of the intrinsically disordered C-terminal domain of the measles virus nucleoprotein derived from selected ASTEROIDS ensembles on the basis of experimental chemical shifts (gray bars) compared to the standard statistical coil sampling (black lines). The Ramachandran space is divided into four regions defined as follows: αL $\{\phi > 0^\circ\}$; αR $\{\phi < 0, -120^\circ < \psi < 50^\circ\}$; βP $\{-100^\circ < \phi < 0^\circ, \psi > 50^\circ \text{ or } \psi < -120^\circ\}$; βS $\{-180^\circ < \phi < -100^\circ, \psi > 50^\circ \text{ or } \psi < -120^\circ\}$. The populations of these quadrants are denoted as $p(\alpha\text{L})$, $p(\alpha\text{R})$, $p(\beta\text{P})$ and $p(\beta\text{S})$, respectively.

agree well with the experimental data (Figure 6). The local conformational sampling derived from the selected ensembles display an overall depletion of the β -strand propensity compared to the statistical coil library, as well as local enhancements of PPII conformations (Figure 7). In addition, residues 485–505 heavily sample the α -helical region of the Ramachandran space. This part of the protein constitutes the interaction site with the cofactor P of the viral polymerase. It has been shown that N_{TAIL} undergoes folding into a complete α -helix upon binding to the cofactor, revealing that N_{TAIL} is to a large extent structurally pre-configured before the interaction.^[141–143]

The representation of IDPs by ensembles containing a restricted number of conformers is a classical ill-posed problem that has no unique solution. In the face of such potential for under-determination, and hence over-fitting, it would appear essential that the validity of the ensembles be independently verified. This can be achieved, for example, by using cross-validation procedures (consistency checks), where a fraction of the experimental data is left out of the selection and subsequently compared to the same data back-calculated from the selected ensemble. In the case of N_{TAIL} from the measles virus, independent measurements of RDCs were performed providing means of cross-validating the ASTEROIDS ensembles obtained on the basis of experimental chemical shifts. RDCs provide a useful consistency check as they depend in a different, complementary manner on local conformational sampling of IDPs.^[36, 139, 144, 145] Figure 8 shows the comparison of the experimental $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$ RDCs in N_{TAIL} and those back-calculated from the ASTEROIDS ensembles. The good agreement between the two sets of RDCs validates the chemical shift-derived ensembles and shows that the ASTEROIDS approach is not noticeably prone to over-fitting.

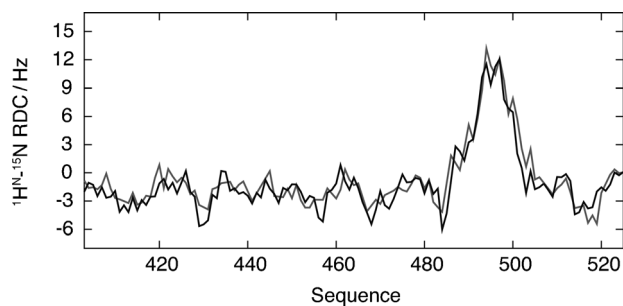


Figure 8. Experimental $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$ RDCs measured in the C-terminal domain of the measles virus nucleoprotein (gray) compared to those back-calculated from ASTEROIDS ensembles obtained on the basis of experimental chemical shifts alone (black). Adapted with permission from ref. [139]. Copyright (2012) American Chemical Society.

7. Outlook

The recent discovery of IDPs has resulted in a large number of studies aiming at the development of robust experimental and computational methods for studying the conformational behaviour of IDPs. In this context, NMR has contributed tremendously allowing structural and dynamic characterization of

IDPs at atomic resolution. Owing to these developments, we are now capable of studying a whole range of proteins with potential links to human diseases.

Despite numerous publications concerning IDPs, only few provide atomic resolution models of the complexes between the IDPs and their physiological partners. One reason for this is that IDPs often engage in dynamic, so-called fuzzy, complexes that pose inherent experimental difficulties in NMR, most likely because the inter-conversion between different complex sub-states occurs on the micro- to millisecond time scale. We are now facing the challenge of characterizing these complexes in terms of structure, dynamics and kinetics, and relaxation dispersion measurements are potentially very powerful in this context. Hopefully, detailed studies of IDP complexes will in the future lead to the discovery of more efficient drugs targeting this type of protein–protein interactions.

Acknowledgements

This work was carried out with financial support from the French Agence Nationale de la Recherche through ANR JCJC ProteinDisorder (to M.R.J.) and ANR MALZ TAUSTRUCT (to M.B.)

Keywords: chemical shifts · conformational propensities · ensemble description · intrinsically disordered proteins · NMR

- [1] V. N. Uversky, *Protein Sci.* **2002**, *11*, 739–756.
- [2] P. Tompa, *Trends Biochem. Sci.* **2002**, *27*, 527–533.
- [3] H. J. Dyson, P. E. Wright, *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197–208.
- [4] A. K. Dunker, I. Silman, V. N. Uversky, J. L. Sussman, *Curr. Opin. Struct. Biol.* **2008**, *18*, 756–764.
- [5] H. J. Dyson, *Q. Rev. Biophys.* **2011**, *44*, 467–518.
- [6] P. Tompa, *Curr. Opin. Struct. Biol.* **2011**, *21*, 419–425.
- [7] P. Tompa, *Trends Biochem. Sci.* **2012**, *37*, 509–516.
- [8] V. N. Uversky, J. R. Gillespie, A. L. Fink, *Proteins* **2000**, *41*, 415–427.
- [9] R. M. Williams, Z. Obradovic, V. Mathura, W. Braun, E. C. Garner, J. Young, S. Takayama, C. J. Brown, A. K. Dunker, *Pac. Symp. Biocomput.* **2001**, 89–100.
- [10] A. K. Dunker, C. J. Oldfield, J. Meng, P. Romero, J. Y. Yang, J. W. Chen, V. Vacic, Z. Obradovic, V. N. Uversky, *BMC Genomics* **2008**, *9*, S1.
- [11] R. Linding, L. J. Jensen, F. Diella, P. Bork, T. J. Gibson, R. B. Russell, *Structure* **2003**, *11*, 1453–1459.
- [12] R. Linding, R. B. Russell, V. Neduvu, T. J. Gibson, *Nucleic Acids Res.* **2003**, *31*, 3701–3708.
- [13] J. J. Ward, L. J. McGuffin, K. Bryson, B. F. Buxton, D. T. Jones, *Bioinformatics* **2004**, *20*, 2138–2139.
- [14] Z. Dosztányi, V. Csizmok, P. Tompa, I. Simon, *Bioinformatics* **2005**, *21*, 3433–3434.
- [15] P. Lieutaud, B. Canard, S. Longhi, *BMC Genomics* **2008**, *9*, S25.
- [16] L. P. Kozlowski, J. M. Bujnicki, *BMC Bioinformatics* **2012**, *13*, 111.
- [17] V. N. Uversky, C. J. Oldfield, U. Midic, H. Xie, B. Xue, S. Vucetic, L. M. Iakoucheva, Z. Obradovic, A. K. Dunker, *BMC Genomics* **2009**, *10*, S7.
- [18] M. M. Babu, R. van der Lee, N. S. de Groot, J. Gsponer, *Curr. Opin. Struct. Biol.* **2011**, *21*, 432–440.
- [19] V. N. Uversky, *Chem. Soc. Rev.* **2011**, *40*, 1623–1634.
- [20] V. N. Uversky, *Biochim. Biophys. Acta* **2013**, *1834*, 932–951.
- [21] A. P. Demchenko, *J. Mol. Recognit.* **2001**, *14*, 42–61.
- [22] H. J. Dyson, P. E. Wright, *Curr. Opin. Struct. Biol.* **2002**, *12*, 54–60.
- [23] K. Sugase, H. J. Dyson, P. E. Wright, *Nature* **2007**, *447*, 1021–1025.
- [24] J. M. Rogers, A. Steward, J. Clarke, *J. Am. Chem. Soc.* **2013**, *135*, 1415–1422.
- [25] C. J. Oldfield, J. Meng, J. Y. Yang, M. Q. Yang, V. N. Uversky, A. K. Dunker, *BMC Genomics* **2008**, *9*, S1.
- [26] P. Tompa, M. Fuxreiter, *Trends Biochem. Sci.* **2008**, *33*, 2–8.
- [27] M. Fuxreiter, *Mol. Biosyst.* **2012**, *8*, 168–177.
- [28] S. Kumar, B. Ma, C.-J. Tsai, N. Sinha, R. Nussinov, *Protein Sci.* **2000**, *9*, 10–19.
- [29] C. J. Tsai, B. Ma, Y. Y. Sham, S. Kumar, R. Nussinov, *Proteins* **2001**, *44*, 418–427.
- [30] M. Fuxreiter, I. Simon, P. Friedrich, P. Tompa, *J. Mol. Biol.* **2004**, *338*, 1015–1026.
- [31] M. R. Jensen, K. Houben, E. Lescop, L. Blanchard, R. W. H. Ruigrok, M. Blackledge, *J. Am. Chem. Soc.* **2008**, *130*, 8055–8061.
- [32] A. H. Mao, N. Lyle, R. V. Pappu, *Biochem. J.* **2013**, *449*, 307–318.
- [33] H. J. Dyson, P. E. Wright, *Chem. Rev.* **2004**, *104*, 3607–3622.
- [34] T. Mittag, J. D. Forman-Kay, *Curr. Opin. Struct. Biol.* **2007**, *17*, 3–14.
- [35] D. Eliezer, *Curr. Opin. Struct. Biol.* **2009**, *19*, 23–30.
- [36] M. R. Jensen, P. R. L. Markwick, S. Meier, C. Griesinger, M. Zweckstetter, S. Grzesiek, P. Bernadó, M. Blackledge, *Structure* **2009**, *17*, 1169–1185.
- [37] R. Schneider, J. Huang, M. Yao, G. Communie, V. Ozenne, L. Mollica, L. Salmon, M. R. Jensen, M. Blackledge, *Mol. Biosyst.* **2012**, *8*, 58–68.
- [38] N. Rezaei-Ghaleh, M. Blackledge, M. Zweckstetter, *ChemBioChem* **2012**, *13*, 930–950.
- [39] J. A. Marsh, S. A. Teichmann, J. D. Forman-Kay, *Curr. Opin. Struct. Biol.* **2012**, *22*, 643–650.
- [40] M. R. Jensen, R. W. H. Ruigrok, M. Blackledge, *Curr. Opin. Struct. Biol.* **2013**, *23*, 426–435.
- [41] B. Brutscher, R. Brüschweiler, R. R. Ernst, *Biochemistry* **1997**, *36*, 13043–13053.
- [42] K. Houben, L. Blanchard, M. Blackledge, D. Marion, *Biophys. J.* **2007**, *93*, 2830–2844.
- [43] K. Modig, F. M. Poulsen, *J. Biomol. NMR* **2008**, *42*, 163–177.
- [44] K. Pervushin, B. Vögeli, A. Eletsy, *J. Am. Chem. Soc.* **2002**, *124*, 12898–12902.
- [45] P. Schanda, H. Van Melckebeke, B. Brutscher, *J. Am. Chem. Soc.* **2006**, *128*, 9042–9043.
- [46] M. Deschamps, I. D. Campbell, *J. Magn. Reson.* **2006**, *178*, 206–211.
- [47] E. Lescop, P. Schanda, B. Brutscher, *J. Magn. Reson.* **2007**, *187*, 163–169.
- [48] Z. Solyom, M. Schwarten, L. Geist, R. Konrat, D. Willbold, B. Brutscher, *J. Biomol. NMR* **2013**, *55*, 311–321.
- [49] R. L. Narayanan, U. H. N. Dürr, S. Bibow, J. Biernat, E. Mandelkow, M. Zweckstetter, *J. Am. Chem. Soc.* **2010**, *132*, 11906–11907.
- [50] J. Wen, J. Wu, P. Zhou, *J. Magn. Reson.* **2011**, *209*, 94–100.
- [51] A. Zawadzka-Kazimierzczuk, W. Koźmiński, H. Sanderová, L. Krásný, *J. Biomol. NMR* **2012**, *52*, 329–337.
- [52] J. Nováček, A. Zawadzka-Kazimierzczuk, V. Papoušková, L. Zídek, H. Sanderová, L. Krásný, W. Koźmiński, V. Sklenář, *J. Biomol. NMR* **2011**, *50*, 1–11.
- [53] J. Nováček, N. Y. Haba, J. H. Chill, L. Zídek, V. Sklenář, *J. Biomol. NMR* **2012**, *53*, 139–148.
- [54] W. Bermel, I. Bertini, I. C. Felli, L. Gonnelli, W. Koźmiński, A. Piai, R. Pierattelli, J. Stanek, *J. Biomol. NMR* **2012**, *53*, 293–301.
- [55] S. Mäntylähti, O. Aitio, M. Hellman, P. Permi, *J. Biomol. NMR* **2010**, *47*, 171–181.
- [56] S. Mäntylähti, M. Hellman, P. Permi, *J. Biomol. NMR* **2011**, *49*, 99–109.
- [57] P. Permi, M. Hellman, *Methods Mol. Biol.* **2012**, *895*, 211–226.
- [58] W. Bermel, I. Bertini, I. C. Felli, R. Kümmerle, R. Pierattelli, *J. Magn. Reson.* **2006**, *178*, 56–64.
- [59] W. Bermel, I. Bertini, I. C. Felli, Y.-M. Lee, C. Luchinat, R. Pierattelli, *J. Am. Chem. Soc.* **2006**, *128*, 3918–3919.
- [60] W. Bermel, I. Bertini, I. C. Felli, M. Piccioli, R. Pierattelli, *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *48*, 25–45.
- [61] V. Csizmok, I. C. Felli, P. Tompa, L. Banci, I. Bertini, *J. Am. Chem. Soc.* **2008**, *130*, 16873–16879.
- [62] W. Bermel, I. Bertini, V. Csizmok, I. C. Felli, R. Pierattelli, P. Tompa, *J. Magn. Reson.* **2009**, *198*, 275–281.
- [63] Y. Pérez, M. Gairí, M. Pons, P. Bernadó, *J. Mol. Biol.* **2009**, *391*, 136–148.
- [64] W. Bermel, M. Bruix, I. C. Felli, M. V. Vasanth Kumar, R. Pierattelli, S. Serrano, *J. Biomol. NMR* **2013**, *55*, 231–237.
- [65] D. S. Wishart, B. D. Sykes, *Meth. Enzymol.* **1994**, *239*, 363–392.
- [66] E. Oldfield, *J. Biomol. NMR* **1995**, *5*, 217–225.
- [67] F. A. A. Mulder, M. Filatov, *Chem. Soc. Rev.* **2010**, *39*, 578–590.
- [68] D. A. Case, *Curr. Opin. Struct. Biol.* **2013**, *23*, 172–176.

- [69] M. Kjaergaard, F. M. Poulsen, *Prog. Nucl. Magn. Reson. Spectrosc.* **2012**, *60*, 42–51.
- [70] X. P. Xu, D. A. Case, *J. Biomol. NMR* **2001**, *21*, 321–333.
- [71] D. S. Wishart, M. S. Watson, R. F. Boyko, B. D. Sykes, *J. Biomol. NMR* **1997**, *10*, 329–336.
- [72] E. L. Ulrich, H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C. F. Schulte, D. E. Tolmie, R. K. Wenger, H. Yao, J. L. Markley, *Nucleic Acids Res.* **2008**, *36*, D402–408.
- [73] S. Neal, A. M. Nip, H. Zhang, D. S. Wishart, *J. Biomol. NMR* **2003**, *26*, 215–240.
- [74] B. Han, Y. Liu, S. W. Ginzinger, D. S. Wishart, *J. Biomol. NMR* **2011**, *50*, 43–57.
- [75] Y. Shen, A. Bax, *J. Biomol. NMR* **2007**, *38*, 289–302.
- [76] Y. Shen, A. Bax, *J. Biomol. NMR* **2010**, *48*, 13–22.
- [77] K. J. Kohlhoff, P. Robustelli, A. Cavalli, X. Salvatella, M. Vendruscolo, *J. Am. Chem. Soc.* **2009**, *131*, 13894–13895.
- [78] P. R. L. Markwick, C. F. Cervantes, B. L. Abel, E. A. Komives, M. Blackledge, J. A. McCammon, *J. Am. Chem. Soc.* **2010**, *132*, 1220–1221.
- [79] D.-W. Li, R. Brüschweiler, *J. Phys. Chem. Lett.* **2010**, *1*, 246–248.
- [80] P. Robustelli, K. A. Stafford, A. G. Palmer III, *J. Am. Chem. Soc.* **2012**, *134*, 6365–6374.
- [81] D.-W. Li, R. Brüschweiler, *J. Biomol. NMR* **2012**, *54*, 257–265.
- [82] A. Cavalli, X. Salvatella, C. M. Dobson, M. Vendruscolo, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 9615–9620.
- [83] Y. Shen, O. Lange, F. Delaglio, P. Rossi, J. M. Aramini, G. Liu, A. Eletsky, Y. Wu, K. K. Singarapu, A. Lemak, A. Ignatchenko, C. H. Arrowsmith, T. Szyperski, G. T. Montelione, D. Baker, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 4685–4690.
- [84] D. S. Wishart, B. D. Sykes, F. M. Richards, *Biochemistry* **1992**, *31*, 1647–1651.
- [85] D. S. Wishart, B. D. Sykes, *J. Biomol. NMR* **1994**, *4*, 171–180.
- [86] M. Kjaergaard, F. M. Poulsen, *J. Biomol. NMR* **2011**, *50*, 157–165.
- [87] D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges, B. D. Sykes, *J. Biomol. NMR* **1995**, *5*, 67–81.
- [88] M. Kjaergaard, S. Brander, F. M. Poulsen, *J. Biomol. NMR* **2011**, *49*, 139–149.
- [89] S. Schwarzingler, G. J. Kroon, T. R. Foss, P. E. Wright, H. J. Dyson, *J. Biomol. NMR* **2000**, *18*, 43–48.
- [90] S. Schwarzingler, G. J. Kroon, T. R. Foss, J. Chung, P. E. Wright, H. J. Dyson, *J. Am. Chem. Soc.* **2001**, *123*, 2970–2978.
- [91] J. H. Prestegard, S. C. Sahu, W. K. Nkari, L. C. Morris, D. Live, C. Gruta, *J. Biomol. NMR* **2013**, *55*, 201–209.
- [92] Y. Wang, O. Jardetzky, *J. Am. Chem. Soc.* **2002**, *124*, 14075–14084.
- [93] Y. Wang, O. Jardetzky, *Protein Sci.* **2002**, *11*, 852–861.
- [94] H. Zhang, S. Neal, D. S. Wishart, *J. Biomol. NMR* **2003**, *25*, 173–195.
- [95] A. De Simone, A. Cavalli, S.-T. D. Hsu, W. Vranken, M. Vendruscolo, *J. Am. Chem. Soc.* **2009**, *131*, 16332–16333.
- [96] K. Tamiola, B. Acar, F. A. A. Mulder, *J. Am. Chem. Soc.* **2010**, *132*, 18000–18003.
- [97] K. Modig, V. W. Jürgensen, K. Lindorff-Larsen, W. Fieber, H. G. Bohr, F. M. Poulsen, *FEBS Lett.* **2007**, *581*, 4965–4971.
- [98] F. Gabel, M. R. Jensen, G. Zaccari, M. Blackledge, *J. Am. Chem. Soc.* **2009**, *131*, 8769–8771.
- [99] J.-R. Huang, F. Gabel, M. R. Jensen, S. Grzesiek, M. Blackledge, *J. Am. Chem. Soc.* **2012**, *134*, 4429–4436.
- [100] J. A. Marsh, V. K. Singh, Z. Jia, J. D. Forman-Kay, *Protein Sci.* **2006**, *15*, 2795–2804.
- [101] K. Tamiola, F. A. A. Mulder, *Biochem. Soc. Trans.* **2012**, *40*, 1014–1020.
- [102] N. C. Maiti, M. M. Apetri, M. G. Zagorski, P. R. Carey, V. E. Anderson, *J. Am. Chem. Soc.* **2004**, *126*, 2399–2408.
- [103] Z. Shi, K. Chen, Z. Liu, N. R. Kallenbach, *Chem. Rev.* **2006**, *106*, 1877–1897.
- [104] R. W. Woody, *J. Am. Chem. Soc.* **2009**, *131*, 8234–8245.
- [105] C. Camilloni, A. De Simone, W. F. Vranken, M. Vendruscolo, *Biochemistry* **2012**, *51*, 2224–2231.
- [106] L. Serrano, *J. Mol. Biol.* **1995**, *254*, 322–333.
- [107] L. J. Smith, K. A. Bolin, H. Schwalbe, M. W. MacArthur, J. M. Thornton, C. M. Dobson, *J. Mol. Biol.* **1996**, *255*, 494–506.
- [108] H. Schwalbe, K. M. Fiebig, M. Buck, J. A. Jones, S. B. Grimshaw, A. Spencer, S. J. Glaser, L. J. Smith, C. M. Dobson, *Biochemistry* **1997**, *36*, 8977–8991.
- [109] P. Bernadó, L. Blanchard, P. Timmins, D. Marion, R. W. H. Ruigrok, M. Blackledge, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 17002–17007.
- [110] A. K. Jha, A. Colubri, K. F. Freed, T. R. Sosnick, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13099–13104.
- [111] S. Meier, M. Blackledge, S. Grzesiek, *J. Chem. Phys.* **2008**, *128*, 052204.
- [112] W. Y. Choy, J. D. Forman-Kay, *J. Mol. Biol.* **2001**, *308*, 1011–1032.
- [113] J. A. Marsh, J. D. Forman-Kay, *Proteins* **2012**, *80*, 556–572.
- [114] M. Krzeminski, J. A. Marsh, C. Neale, W.-Y. Choy, J. D. Forman-Kay, *Bioinformatics* **2013**, *29*, 398–399.
- [115] H. J. Feldman, C. W. Hogue, *Proteins* **2000**, *39*, 112–131.
- [116] J. A. Marsh, J. D. Forman-Kay, *J. Mol. Biol.* **2009**, *391*, 359–374.
- [117] J. A. Marsh, B. Dancheck, M. J. Ragusa, M. Allaire, J. D. Forman-Kay, W. Peti, *Structure* **2010**, *18*, 1094–1103.
- [118] T. Mittag, J. Marsh, A. Grishaev, S. Orlicky, H. Lin, F. Sicheri, M. Tyers, J. D. Forman-Kay, *Structure* **2010**, *18*, 494–506.
- [119] A. S. Pinheiro, J. A. Marsh, J. D. Forman-Kay, W. Peti, *J. Am. Chem. Soc.* **2011**, *133*, 73–80.
- [120] C. K. Fisher, A. Huang, C. M. Stultz, *J. Am. Chem. Soc.* **2010**, *132*, 14919–14927.
- [121] C. K. Fisher, C. M. Stultz, *Curr. Opin. Struct. Biol.* **2011**, *21*, 426–431.
- [122] C. K. Fisher, O. Ullman, C. M. Stultz, *Biophys. J.* **2013**, *104*, 1546–1555.
- [123] O. Ullman, C. K. Fisher, C. M. Stultz, *J. Am. Chem. Soc.* **2011**, *133*, 19536–19546.
- [124] T. Bartels, J. G. Choi, D. J. Selkoe, *Nature* **2011**, *477*, 107–110.
- [125] W. Wang, I. Perovic, J. Chittuluru, A. Kaganovich, L. T. T. Nguyen, J. Liao, J. R. Auclair, D. Johnson, A. Landeru, A. K. Simorellis, S. L. Ju, M. R. Cookson, F. J. Asturias, J. N. Agar, B. N. Webb, C. H. Kang, D. Ringe, G. A. Petsko, T. C. Pochapsky, Q. Q. Hoang, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 17797–17802.
- [126] A. Binolfi, F.-X. Theillet, P. Selenko, *Biochem. Soc. Trans.* **2012**, *40*, 950–954.
- [127] B. Fauvet, M. K. Mbefo, M.-B. Fares, C. Desobry, S. Michael, M. T. Ardah, E. Tsika, P. Coune, M. Prudent, N. Lion, D. Eliezer, D. J. Moore, B. Schneider, P. Aebischer, O. M. El-Agnaf, E. Masliah, H. A. Lashuel, *J. Biol. Chem.* **2012**, *287*, 15345–15364.
- [128] A. S. Maltsev, J. Ying, A. Bax, *Biochemistry* **2012**, *51*, 5004–5013.
- [129] T. Gurry, O. Ullman, C. K. Fisher, I. Perovic, T. Pochapsky, C. M. Stultz, *J. Am. Chem. Soc.* **2013**, *135*, 3865–3872.
- [130] V. Ozenne, F. Bauer, L. Salmon, J.-R. Huang, M. R. Jensen, S. Segard, P. Bernadó, C. Charavay, M. Blackledge, *Bioinformatics* **2012**, *28*, 1463–1470.
- [131] J. Huang, M. Gentner, N. Vajpai, S. Grzesiek, M. Blackledge, *Biochem. Soc. Trans.* **2012**, *40*, 989–994.
- [132] G. Nodet, L. Salmon, V. Ozenne, S. Meier, M. R. Jensen, M. Blackledge, *J. Am. Chem. Soc.* **2009**, *131*, 17908–17918.
- [133] S. Meier, S. Grzesiek, M. Blackledge, *J. Am. Chem. Soc.* **2007**, *129*, 9799–9807.
- [134] M.-K. Cho, G. Nodet, H.-Y. Kim, M. R. Jensen, P. Bernadó, C. O. Fernandez, S. Becker, M. Blackledge, M. Zweckstetter, *Protein Sci.* **2009**, *18*, 1840–1846.
- [135] L. Salmon, G. Nodet, V. Ozenne, G. Yin, M. R. Jensen, M. Zweckstetter, M. Blackledge, *J. Am. Chem. Soc.* **2010**, *132*, 8407–8418.
- [136] S. Bibow, V. Ozenne, J. Biernat, M. Blackledge, E. Mandelkow, M. Zweckstetter, *J. Am. Chem. Soc.* **2011**, *133*, 15842–15845.
- [137] M. R. Jensen, L. Salmon, G. Nodet, M. Blackledge, *J. Am. Chem. Soc.* **2010**, *132*, 1270–1272.
- [138] F. Szigat, R. Silvers, M. Hähnke, M. R. Jensen, M. Blackledge, J. Wirmer-Bartoschek, H. Schwalbe, *Biochemistry* **2012**, *51*, 3361–3372.
- [139] V. Ozenne, R. Schneider, M. Yao, J. Huang, L. Salmon, M. Zweckstetter, M. R. Jensen, M. Blackledge, *J. Am. Chem. Soc.* **2012**, *134*, 15138–15148.
- [140] S. Gely, D. F. Lowry, C. Bernard, M. R. Jensen, M. Blackledge, S. Costanzo, J.-M. Bourhis, H. Darbon, G. Daughdrill, S. Longhi, *J. Mol. Recognit.* **2010**, *23*, 435–447.
- [141] S. Longhi, V. Receveur-Bréchet, D. Karlin, K. Johansson, H. Darbon, D. Bhella, R. Yeo, S. Finet, B. Canard, *J. Biol. Chem.* **2003**, *278*, 18638–18648.

- [142] R. L. Kingston, D. J. Hamel, L. S. Gay, F. W. Dahlquist, B. W. Matthews, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 8301–8306.
- [143] M. R. Jensen, G. Communie, E. A. Ribeiro, Jr., N. Martinez, A. Desfosses, L. Salmon, L. Mollica, F. Gabel, M. Jamin, S. Longhi, R. W. H. Ruigrok, M. Blackledge, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 9839–9844.
- [144] M. R. Jensen, M. Blackledge, *J. Am. Chem. Soc.* **2008**, *130*, 11266–11267.
- [145] J. Huang, V. Ozenne, M. R. Jensen, M. Blackledge, *Angew. Chem.* **2013**, *125*, 715–718; *Angew. Chem. Int. Ed.* **2013**, *52*, 687–690.

Received: April 16, 2013

Published online on June 21, 2013

Small-Angle X-Ray Scattering- and Nuclear Magnetic Resonance-Derived Conformational Ensemble of the Highly Flexible Antitoxin PaaA2

Yann G.J. Sterckx,^{1,2} Alexander N. Volkov,^{1,2} Wim F. Vranken,^{1,2} Jaka Kragelj,⁴ Malene Ringkjøbing Jensen,⁴ Lieven Buts,^{1,2} Abel Garcia-Pino,^{1,2} Thomas Jové,³ Laurence Van Melderen,³ Martin Blackledge,⁴ Nico A.J. van Nuland,^{1,2,*} and Remy Loris^{1,2,*}

¹Structural Biology Brussels, Department of Biotechnology, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium

²Molecular Recognition Unit and Jean Jeener NMR Centre, Structural Biology Research Center, VIB, Pleinlaan 2, B-1050 Brussels, Belgium

³Laboratoire de Génétique et Physiologie Bactérienne, Institut de Biologie et de Médecine Moléculaires Faculté des Sciences, Université Libre de Bruxelles, 12 Rue des Professeurs Jeener et Brachet, B-6041 Gosselies, Belgium

⁴Protein Dynamics and Flexibility, Institut de Biologie Structurale Jean-Pierre Ebel CNRS-CEA-UJF UMR 5075, 41 Rue Jules Horowitz, 38027 Grenoble Cedex, France

*Correspondence: nico.vannuland@vib-vub.be (N.A.J.v.N.), remy.loris@vib-vub.be (R.L.)

<http://dx.doi.org/10.1016/j.str.2014.03.012>

SUMMARY

Antitoxins from prokaryotic type II toxin-antitoxin modules are characterized by a high degree of intrinsic disorder. The description of such highly flexible proteins is challenging because they cannot be represented by a single structure. Here, we present a combination of SAXS and NMR data to describe the conformational ensemble of the PaaA2 antitoxin from the human pathogen *E. coli* O157. The method encompasses the use of SAXS data to filter ensembles out of a pool of conformers generated by a custom NMR structure calculation protocol and the subsequent refinement by a block jackknife procedure. The final ensemble obtained through the method is validated by an established residual dipolar coupling analysis. We show that the conformational ensemble of PaaA2 is highly compact and that the protein exists in solution as two preformed helices, connected by a flexible linker, that probably act as molecular recognition elements for toxin inhibition.

INTRODUCTION

The occurrence and biological importance of intrinsic disorder in individual proteins (so-called intrinsically disordered proteins or IDPs) and macromolecular complexes have been recognized for several decades (Wright and Dyson, 1999; Tompa, 2012). Intrinsic disorder plays an important role in many cellular processes that require rapid responses. Because of their unstructured nature in the unbound state, the rapid turnover of IDPs provides a way for the cell to almost instantly adapt to changing external circumstances (Fink, 2005; Dyson and Wright, 2005). From the point of view of the structural biologist, it has become clear that these systems explore a vast conformational space. Hence, their structure cannot be represented by a single

conformation (Boehr et al., 2009; Tompa, 2011). Instead, an ensemble of conformers can convincingly describe the structural properties of a macromolecular entity with intrinsically disordered segments (Wells et al., 2008; Bertini et al., 2009; Huang et al., 2009; Bernadó et al., 2010; Marsh et al., 2010; Mittag et al., 2010; Leyrat et al., 2012). Solution structure techniques, such as small-angle X-ray scattering (SAXS) and nuclear magnetic resonance (NMR) spectroscopy, are well suited to achieve this goal. Although these techniques independently provide valuable information, it is increasingly recognized that a correct description of the conformational ensemble of an IDP often requires a combination of both (Mittag and Forman-Kay, 2007; Madl et al., 2011). Different approaches combining SAXS and NMR data have been devised to quantitatively describe the behavior of intrinsically disordered systems. Most methods employ a similar rationale. In a first step, a statistical coil generator is used to sample the available conformational space. This leads to the generation of a large pool of structures from which a subset of conformers is selected by comparing back-predicted properties to experimental data sets. By definition, the final ensemble is the collection of conformers that best agrees with the experimental data sets and thus best describes the properties of the flexible system of interest.

In the ASTEROIDS approach, the initial pool is produced by the statistical coil generator *Flexible-meccano* (Ozenne et al., 2012). The selection of the final ensemble from this pool occurs through recursive application of a genetic algorithm that uses either experimental SAXS data (Bernadó et al., 2007), experimental residual dipolar couplings (RDCs), chemical shifts (CS), and/or paramagnetic relaxation enhancements as restraints (Nodet et al., 2009; Jensen et al. 2010; Salmon et al., 2010). The maximum occurrence (MO) method (Bertini et al., 2010) employs RanCh, an offspring of the SAXS-based ensemble optimization method (EOM; Bernadó et al., 2007), to create the initial pool of conformers. In this approach, the SAXS data are combined with pseudocontact shifts and RDCs arising from the presence of paramagnetic lanthanide ions introduced into the biomolecular frame. Subsets of conformers are then picked from the initial pool by a stochastic selection procedure and optimized through a simulated annealing approach until the final

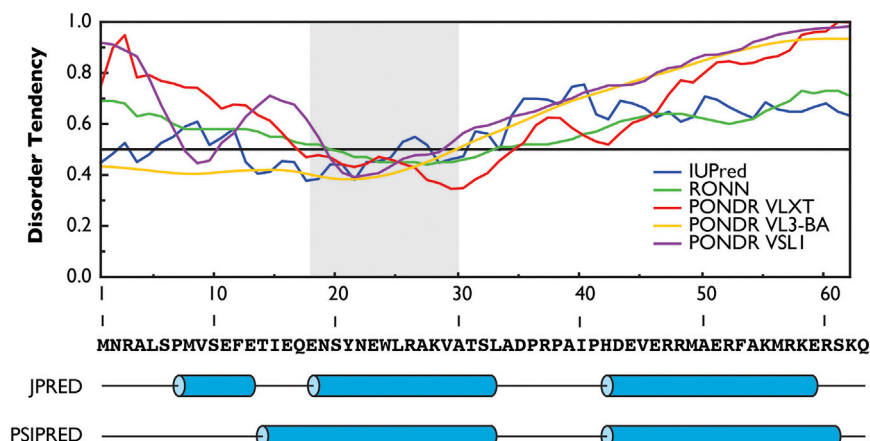


Figure 1. Sequence-Based Structural Analysis of PaaA2

The prediction of PaaA2 disorder tendency using different prediction algorithms (annotated in the plot and shown in different colors) is shown at the top. The region Glu18 to Ala30 with a slight tendency for structure is highlighted by the gray area. The predicted secondary structure of PaaA2 based on in silico analysis through JPRED and PSIPRED is shown below. Cylinders indicate α helices.

See also Figure S1.

ensemble has been obtained. Another approach, applied in the program ENSEMBLE (Marsh and Forman-Kay, 2011), may use either TraDES (Feldman and Hogue, 2000) or a set of user-defined structures to generate the initial pool. Like in the MO method, the selection procedure for subsets of conformers is stochastic. The back-calculated properties of the selected structures may be fitted against a wide variety of experimental restraints (Marsh and Forman-Kay, 2011). The process is repeated until a final ensemble of desired size has been obtained.

Most of these approaches involve the use of RDCs as restraints in the ensemble calculation because they report not only on the structural propensities of IDPs, but also on their dynamics on various timescales. Therefore, RDCs are powerful tools in the quantitative description of IDPs (Jensen et al., 2009). Although these proteins display high degrees of flexibility, many IDPs contain secondary structure elements (predominantly α helices) in their unbound state and are thus not entirely disordered (Fuxreiter et al., 2004). Although RDCs are sensitive probes for the presence of local (and long-range) structure, nuclear Overhauser effects (NOEs) are equally powerful descriptors of secondary structure elements. In the case of IDPs containing preformed α helices, we therefore reasoned that NOE analysis may serve a role for establishing the local structure in the ensemble calculation. In this paper, we present a combination of SAXS and NMR to describe the model ensemble of a highly flexible protein encoded by a toxin-antitoxin (TA) module, which contains preformed α helices in the absence of its binding partner.

TA modules are stress-response elements (Christensen et al., 2003; Cooper et al., 2009; Christensen-Dalsgaard et al., 2010) implicated in the establishment of persisters (Shah et al., 2006; Maisonneuve et al., 2011). Their behavior and biological functions are mediated by intrinsic disorder (Lah et al., 2005; De Jonge et al., 2009; Garcia-Pino et al., 2010). Many different TA families exist that differ in the targets of the toxins and the nature of the DNA-binding domain of the antitoxins (Buts et al., 2005; Gerdes et al., 2005; Leplae et al., 2011; Blower et al., 2011; Hayes and Van Melderen, 2011). The *Escherichia coli* O157 three-component *paaR2-paaA2-parE2* TA module encodes the PaaA2 antitoxin and the ParE2 toxin. Although this module is

related to the *relBE* and *parDE* families, the target of the ParE2 toxin remains unknown (Hallez et al., 2010). The model conformational ensemble of the antitoxin

PaaA2 is determined by a structure calculation protocol combining SAXS and NMR. In our approach, we generate a large pool of structures derived from experimental CS and NOE data and select ensembles of varying sizes using SAXS data through EOM, followed by a final block jackknife refinement (Efron, 1982). Validated by an established RDC analysis, our results show that this antitoxin has all the hallmarks of an IDP, yet its final model conformational ensemble is rather compact, which may have important biological implications.

RESULTS

PaaA2 Is an Atypical Nonglobular Member of the RelB-Antitoxin Family

Recently, it was shown that PaaA2 belongs to the RelB-antitoxin family (Leplae et al., 2011). Compared with the classic RelB antitoxins (typically 80–100 amino acids), the *E. coli* O157 PaaA2 antitoxin is significantly shorter (63 residues). Sequence analysis reveals that, in contrast to canonical RelB antitoxins such as *E. coli* RelB1, PaaA2 lacks a DNA-binding domain (Figure S1 available online). This raises the question whether PaaA2 is merely an intrinsically disordered toxin-binding domain rather than a classic antitoxin with dual binding functionalities (i.e., DNA and toxin binding). Different intrinsic disorder predictors indeed indicate that PaaA2 has a large tendency to be unstructured. However, a stretch of residues (Glu18 to Ala30) has a slight propensity for structure, suggesting that the antitoxin is not fully disordered (Figure 1). Secondary structure prediction suggests the presence of two α helices, one of which corresponds to the aforementioned segment (Glu18 to Ala30). Interestingly, in our sequence alignment, the two predicted α helices coincide with the C-terminal toxin-neutralizing segments of the classic RelB antitoxins (Figure S1). Taken together, these results suggest that PaaA2 is an IDP-like protein possessing secondary structure elements in the absence of its cognate binding partner ParE2.

To confirm the above-mentioned hypothesis experimentally, a combination of dynamic light scattering (DLS), analytical size exclusion chromatography (SEC), and circular dichroism (CD) spectroscopy was used. Estimation of the antitoxin's secondary structure content based on the analysis of the CD spectrum

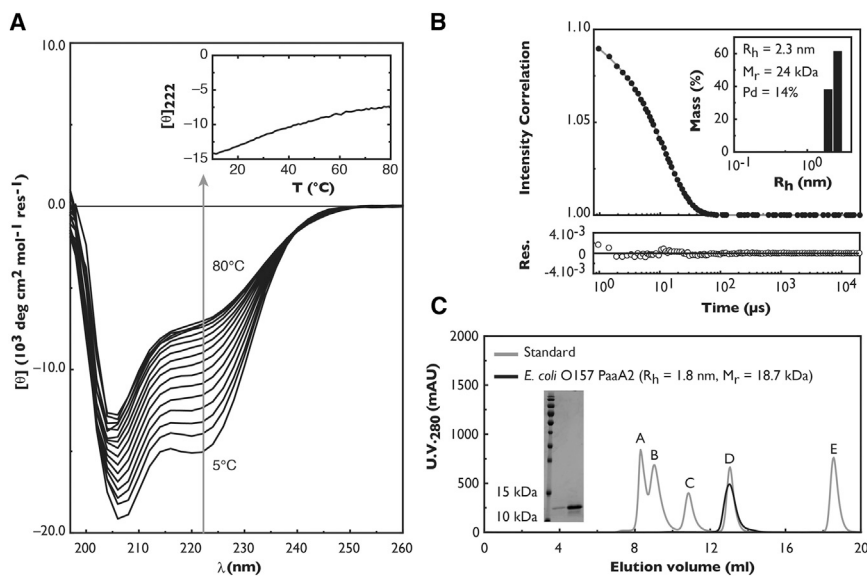


Figure 2. Biophysical Characterization of PaaA2 in Solution

(A) CD spectra of PaaA2 collected at different temperatures. The gray arrow represents the effect of the increasing temperature on the mean residue ellipticity measured at 222 nm, plotted for convenience in the inset.

(B) DLS. The black dots and gray line represent the experimental data and fit, respectively. The residuals (Res.) are shown. The percentage of mass as a function of the R_h distribution is given in the inset.

(C) Analytical SEC. The elution peaks of the standard proteins (in gray) are indicated by letters: bovine thyroglobulin (peak A, $M_r = 670$ kDa, $R_h = 8.6$ nm), bovine γ -globulin (peak B, $M_r = 158$ kDa, $R_h = 5.1$ nm), chicken ovalbumin (peak C, $M_r = 44$ kDa, $R_h = 2.8$ nm), horse myoglobin (peak D, $M_r = 17$ kDa, $R_h = 1.9$ nm), and vitamin B₁₂ (peak E, $M_r = 1.35$ kDa, $R_h = \lambda$). The inset shows an SDS-PAGE (15% gel) analysis of the PaaA2 elution peak.

See also Figure S2 and Table S1.

reveals that PaaA2 consists of random coil with a significant amount of α helices ($\sim 35\%$ α helicity; Figure 2A and Table S1). Thermal denaturation of PaaA2, monitored at 222 nm in CD spectra (Figure 2A, inset), shows that the α -helical content diminishes with increasing temperature. It is noteworthy that the denaturation is fully reversible (Figure S2) and that the thermal profile is not sigmoidal, suggesting that the unfolding event occurs through a noncooperative process. The analytical SEC and DLS experiments indicate that PaaA2 displays IDP-like features and adopts an extended, nonglobular conformation (Figures 2B and 2C). The ratios between the apparent and theoretical values for molecular mass ($M_{r,app}/M_{r,theor}$) and hydrodynamic radius ($R_{h,app}/R_{h,theor}$) are 2.2–2.8 and 1.3–1.6, respectively, which is a typical trait for IDPs (Csizmók et al., 2006). Furthermore, the IDP-like character of PaaA2 is also in agreement with an aberrant mobility on SDS-PAGE: again, IDPs are notorious for migrating with a larger apparent molecular mass than the one based on their amino acid sequence (Tompa, 2002). In the case of PaaA2, $M_{r,theor}$ is 8.5 kDa. These findings provide strong evidence that the PaaA2 antitoxin adopts a nonglobular conformation in solution.

SAXS Reveals the Highly Flexible Nature of PaaA2

The structural properties of PaaA2 in solution were further investigated by SAXS measurements (Figure 3A and Table S2). The SAXS data also indicate that PaaA2 is an IDP-like protein as the normalized Kratky plot displays the behavior of a highly flexible biomolecule (Figure 3A, inset). Furthermore, comparison of the experimentally determined R_g of PaaA2 with the values expected for an IDP of similar length (63 amino acids) or a globular protein with a similar molecular mass (~ 8.5 kDa) suggests that the behavior of PaaA2 in solution most resembles that of an IDP and not of a globular protein (Figure 3B). It is, however, noteworthy that the R_g of PaaA2 lies just beneath the Flory region (Bernadó and Svergun, 2012), indicating that the antitoxin does not behave as a random-coil IDP in solution. For such completely unfolded proteins, the normalized Kratky plot reaches a plateau

~ 2.0 (Durand et al., 2010). This is clearly not the case for PaaA2, where the normalized Kratky plot levels off at 1.5. In conclusion, the SAXS data indicate that PaaA2 displays a highly flexible character.

PaaA2 Harbors Two α Helices Connected by a Flexible Linker

The fully assigned ^1H - ^{15}N HSQC spectrum of PaaA2 shows that most of the backbone amide resonances are confined to a narrow spectral region (Figure 4A), typical for IDPs or proteins with a high α -helical content (Wüthrich, 1986; Wishart et al., 1991). The few NH peaks outside of this range (dashed boxed peaks in Figure 4A) originate from the side chain amide groups of Gln and Asn residues and from the indole group of Trp24. Chemical shift-based secondary structure predictions with DANGLE and TALOS+ suggest the presence of two α helices (residues Glu16 to Lys28 and His42 to Arg57), connected by a linker segment (Figure 4B), which is in agreement with our bioinformatics analysis. On the other hand, $\delta 2\text{D}$ (Camilloni et al., 2012), which is benchmarked for IDPs, identifies two stable cores existing in an α -helical conformation for more than 80% of the time: residues Asn22 to Arg26 and Glu46 to Glu51 for helices 1 and 2, respectively (Figure 4C). These core regions coincide with the predicted α helices and contain flanking regions with percentages of α helicity between 15% and 80% (Figure 4C), which is overall in agreement with the CD data (Table S1). All of the other residues are predicted to adopt a coil-like conformation. In agreement with these predictions, the NOE spectroscopy (NOESY) spectra indicate the absence of β sheet signatures and contain peaks characteristic of α helices.

In order to address the PaaA2 dynamics on the pico- to nanosecond timescale, we studied the NMR relaxation properties of the PaaA2 backbone amides (Figure S3). The generalized order parameters (S^2) and local rotational correlation times (τ_l) indicate that PaaA2 contains two rigid regions (Thr14 to Leu33 and His42 to Arg60) flanked by highly flexible

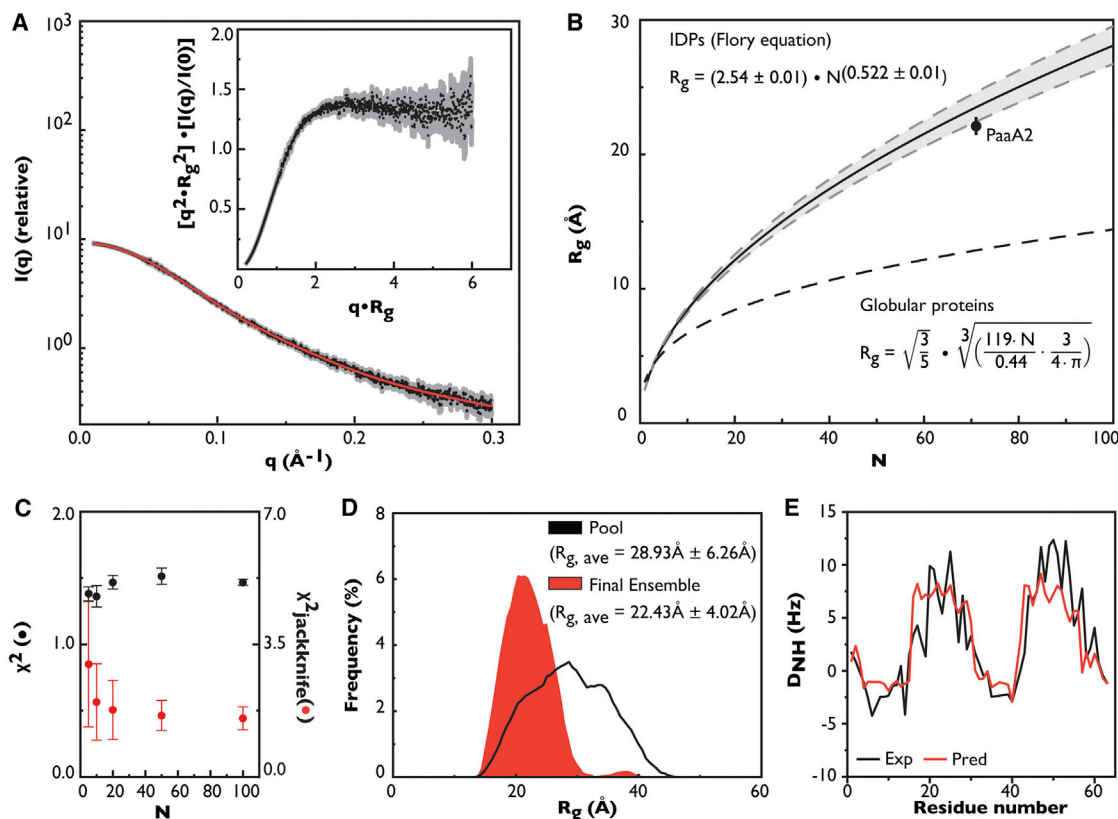


Figure 3. Determination of the Model Conformational Ensemble of PaaA2 through a Combination of SAXS and NMR

(A) The scattering curve of PaaA2 obtained by extrapolating all data at different concentrations to the zero concentration. The experimental data are shown in black, the error margins are shown in gray, and the fit of the optimal model ensemble is shown in red. The inset shows the normalized Kratky plot. (B) Comparison of the experimentally determined R_g of PaaA2 with the R_g of an IDP with the same length (full black line; the gray area represents the confidence interval of the Flory equation) and the R_g of a globular protein with the same molecular mass (dashed black line). (C) χ^2 analysis of the EOM-derived model ensembles before (black dots) and after (red dots) application of the block jackknife. (D) Comparison of the R_g distribution of the initial selection pool (black line) and the selected model ensemble (filled red lines). (E) Comparison of the experimentally determined ^1H - ^{15}N RDC values (black line) and the ensemble-averaged RDCs of the final PaaA2 ensemble back-calculated using PALES (red line). See also Figure S4 and Table S2.

polypeptide stretches (Figure 4B). This finding is in excellent agreement with the PaaA2 secondary structure as seen in the NOE-based solution structure, with rigid regions mapping out onto the two α helices. The relaxation data did not fit well to an isotropic diffusion model assuming a single, global rotational correlation time (τ_M), suggesting that different parts of the protein may sample different motional regimes. Indeed, inclusion of the individual correlation times for each residue (τ_i) significantly improves the fit, with the optimized τ_i values mirroring the S^2 trend outlining the rigid α helices (Figure 4B). Interestingly, only four PaaA2 residues needed an exchange contribution ($R_{2,ex} < 14 \text{ s}^{-1}$) to the transverse relaxation rate (R_2) for a good fit (Table S3), which agrees well with the experimentally measured $R_{2,ex}$ data showing flat relaxation dispersion profiles (data not shown). Considering that the model-free analysis is inappropriate for highly flexible proteins (Metcalfe et al., 2004), the experimental $R_{2,ex} \approx 0$ contributions match well the relaxation model prediction. Altogether, these findings reveal that, in the absence of its cognate binding partner

ParE2, the PaaA2 antitoxin in solution harbors two α helices connected by a flexible linker.

NMR and SAXS Provide an Adequate Description of the PaaA2 Model Conformational Ensemble

The results obtained from the NMR and SAXS experiments were combined in search of the model conformational ensemble that describes well the solution behavior of PaaA2. In the first step, a pool of 5,000 structures was generated based solely on the NMR data (CS and NOEs; Table 1) through a modified NMR structure calculation protocol, in an attempt to deal with the conformational heterogeneity of the protein.

Based on the two sets of dihedral angles estimated from the experimental CS data by use of DANGLE and TALOS+, a final set of dihedral angle restraints was obtained as follows: a restraint was only used if predicted by both programs, in which case the broadest combined angle range was employed. For the NOE data, the full ^{13}C and ^{15}N NOESY peak lists with partial manual assignments were directly used as input. The peak

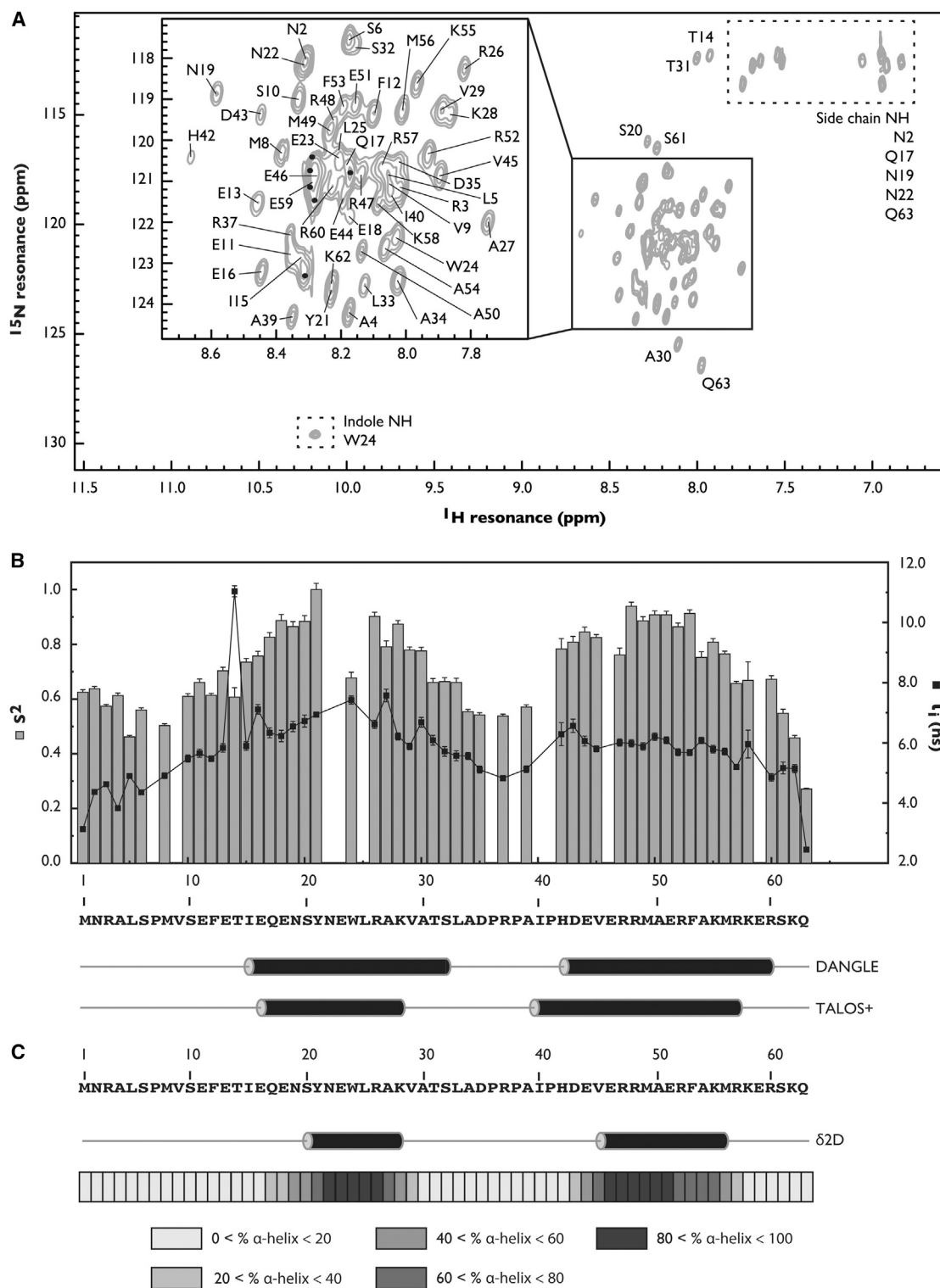


Figure 4. NMR Experiments

(A) The ^1H - ^{15}N HSQC spectrum of PaaA2. The inset represents the backbone NH peaks: each peak has been assigned with its corresponding amino acid, and the black dots represent the peaks stemming from the FLAG tag amino acids. The peaks in the dashed boxes represent signals from the NH groups of Asn, Gln, and Trp residues.

(legend continued on next page)

intensities were converted into distances with the standard ARIA spin diffusion correction, whereas partially assigned or unassigned peaks were assigned by ARIA within the modified structure calculation protocol. Finally, to treat the conformationally averaged restraints more appropriately during the final stages of the structure calculation, a log-harmonic potential was employed. It is important to emphasize that, by utilizing this approach, all structures in the starting pool individually satisfy the NOE-derived restraints and CS-estimated dihedral angles.

Model ensembles of varying size (N) were selected from the 5,000-member starting pool using the recorded SAXS data through EOM. Interestingly, all EOM runs provide model ensembles with a similar distribution for the radius of gyration R_g and the maximum intraparticle distance D_{max} . The reported χ^2 values for the runs with different N are very similar (1.43 ± 0.08) and thus cannot be used on their own to select a model ensemble that best fits the experimental data (Figure 3C, black dots). To pursue our endeavor, we refined the search protocol by applying a block jackknife procedure on the EOM-derived SAXS ensembles. The resulting χ^2 values for the different EOM runs treated with and without the block jackknife approach are plotted as a function of the ensemble size (Figure 3C). Upon implementation of the block jackknife protocol, it becomes obvious that the effect of randomly removing conformers from the population decreases with increasing population size. This trend levels off at around $N = 50$ and remains similar for larger ensemble sizes (Figure 3C, red dots), thus suggesting that a 50-member model ensemble best describes the structural properties of PaaA2 in solution. Interestingly, the final model ensemble displays an anisotropy in its coverage of the available conformational space compared with the initial selection pool. As can be seen from Figures 3D and 5, the ensemble members adopt a rather compact conformation. To validate the described calculation method, ^1H - ^{15}N RDCs were measured in a suspension of filamentous phages. The measured RDCs are negative except in the two helical regions of the protein, where positive RDCs with characteristic dipolar wave signatures are observed (Jensen and Blackledge, 2008). We back-calculated the RDCs from the 5,000-member starting pool and the final PaaA2 model ensemble using PALES (Zweckstetter and Bax, 2000) assuming steric alignment, which is a good approximation because all experiments were carried out at high ionic strength. Both back-calculated RDC data sets are compared with the experimental RDCs. As can be seen from Figures 3E and S4, the final PaaA2 model ensemble, obtained by the combination of NMR and SAXS described above, provides a better description of the experimentally determined ^1H - ^{15}N RDCs compared with the starting pool ($\chi^2 = 7.61$ versus $\chi^2 = 8.62$; $Q = 0.48$ versus $Q = 0.51$).

Finally, the influence of the starting pool of structures on the properties of the final model ensemble was also investigated. Using *Flexible-meccano*, two different 5,000-member starting pools were generated: a random coil pool containing no α helices

and a pool with α -helical populations based on the $\delta 2\text{D}$ output. Hence, both initial pools reflect the disordered nature of PaaA2 but do not represent the experimental NOE data. Applying the approach described above shows that ensembles selected from both pools can satisfy the SAXS data but that only ensembles selected from the pool containing the preformed α helices can adequately reproduce the NOE and RDC data (Figure S5).

DISCUSSION

PaaA2 is an atypical member of the RelB-antitoxin family (Lepplae et al., 2011). Unlike the archetypal members of this antitoxin family, which have a well-structured N-terminal DNA-binding domain and a disordered C-terminal toxin-binding segment, the DNA-binding domain is absent in PaaA2. This implies that this antitoxin would occur as an intrinsically disordered toxin-inhibiting protein without DNA-binding capacities. Although it has been demonstrated that PaaA2 is unable to bind the promoter region of the *paaR2-paaA2-parE2* module by itself and that PaaA2 is extremely efficient at alleviating the toxic effect of the ParE2 toxin in vivo (Hallez et al., 2010), our in vitro results (SEC, DLS, NMR, and SAXS) provide experimental evidence for the highly flexible nature of the PaaA2 antitoxin.

To obtain a more accurate description of the model conformational ensemble of PaaA2, we have employed a combination of SAXS and NMR. The method utilizes a rationale similar to that of the existing approaches combining SAXS and NMR designed to describe the properties of IDPs in solution (Bertini et al., 2010; Marsh and Forman-Kay, 2011; Nodet et al., 2009; Ozenne et al., 2012) but differs in its execution. Like in the other approaches, the first step consists of generating a large pool of conformers. However, unlike the other methods, the initial pool is not constructed by use of a statistical coil generator, but through a modified NMR structure calculation protocol given the presence of preformed helices. This protocol directly incorporates experimental NOE data and CS-derived dihedral angles but attempts to compensate for the conformational heterogeneity experienced by PaaA2 in two key ways. First, it employs a log-harmonic potential to restrain the NOE-derived interatomic distances in order to better deal with NOE inconsistencies introduced by conformational averaging. Second, it combines different dihedral angle estimation procedures from the chemical shifts, essentially resulting in a more dependable consensus prediction.

Our approach makes the assumption that the rotational correlation times affecting different interproton vectors are homogeneous throughout the protein (which is rather unlikely in such an anisotropic protein), that the helices are populated to the same extent in each conformer, and that the exchange between the conformers is slow with respect to relaxation but fast with respect to chemical shift averaging. In view of the high helical content on the basis of chemical shift analysis, these

(B) The general order parameter S^2 and local correlation time τ_i for each residue. The predicted secondary structure of PaaA2 based on the analysis of the backbone chemical shift data by DANGLE and TALOS+ is shown below.

(C) The predicted secondary structure of PaaA2 based on the analysis of the backbone chemical shift data by $\delta 2\text{D}$. The boxes under the prediction display the percentage of α helicity for every residue. The helices represent the residues with a percentage higher than 50%. See also Figure S3 and Table S3.

Table 1. Summary of Structural Statistics for the 50 Selected PaaA2 Structures

Experimental Data		
Resonance assignment completeness (%)		
Backbone H	94.26	
Backbone non-H	93.43	
Side-chain H	94.71	
Side-chain non-H	69.88	
Distance restraints		
Intraresidue ($i - j = 0$)	389	
Sequential ($ i - j = 1$)	176	
Medium range ($1 \leq i - j \leq 4$)	35	
Long range ($ i - j \geq 5$)	0	
Total	600	
Restraints per residue	8.45	
Dihedral restraints		
Phi angles	34	
Psi angles	35	
Validation Statistics		
Restraints statistics, average per structure		
NOE distance restraint violations ^a		
>0.1 Å	20.00	
>0.5 Å	4.84	
NOE energy per restraint (min/ mean/max) ^b	0.29/0.32/0.35	
NOE restraint weight (min/mean/ max) ^b	14.73/17.21/20.01	
Dihedral violations 1°–5°	1.6	
	All	Helix
CING scores ^c		
Red/orange/green (in % residues)	11/17/72	6/16/78
WHATIF first generation packing	0.35 ± 1.04	2.25 ± 1.30
WHATIF second generation packing	1.47 ± 1.50	2.72 ± 1.63
Ramachandran plot appearance	−4.25 ± 0.82	−3.34 ± 1.11
χ^1/χ^2 rotamer normality	−1.43 ± 0.94	−1.46 ± 1.17
Backbone conformation	−2.94 ± 0.87	−2.36 ± 1.41
Ramachandran plot quality (Procheck) ^d		
Most favored regions (%)	80.5	74.8
Additional allowed regions (%)	16.6	21.4
Generously allowed regions (%)	1.0	1.5
Disallowed regions (%)	1.9	2.3

For the structure validation statistics, values are given for the whole sequence (left) and only the helical parts (right).

^aThe violations listed here are based on the lower/upper distance restraints and flat-bottomed harmonic wall potentials, even though boundary-free log-harmonic potentials were used in the final stages of the structure calculation.

^bThe average NOE energy per restraint and the average automatic NOE weight are given as more meaningful measures to assess the log-harmonic restraints.

^cDoreleijers et al. (2012).

^dLaskowski et al. (1996).

approximations were considered acceptable for the generation of the initial ensemble. The second step, which involves the selection of ensembles of varying sizes from the starting pool, strongly resembles the ASTEROIDS approach. That is, the ensembles are selected using experimental SAXS data through a genetic algorithm (EOM). However, because the selection of the optimal PaaA2 ensemble size solely based on the output of the EOM analysis proves to be difficult, we have applied an additional block jackknife procedure on the EOM-derived SAXS data. This seems to provide an effective means of identifying the structural ensemble that best represents the experimental NMR and SAXS data. Furthermore, the validity of the final model ensemble is supported by a comparison of ensemble-averaged, back-calculated RDCs to an experimental RDC data set, which was not employed as a restraint during the ensemble calculation. Although the back-calculated RDCs of both the starting pool and the final model ensemble nicely reproduce the local structure of the α helices, the final PaaA2 model ensemble provides a better description of the experimentally determined ^1H - ^{15}N RDCs. The improvement in the reproduction of the RDCs by the final model ensemble is a result of the compactness that is displayed by PaaA2, a feature that was revealed by our SAXS experiments. Thus, the combination of both SAXS and NMR appears to be necessary for the description of the solution structure of IDPs because some of the ensemble features of PaaA2 are detected by one technique but not by the other. For PaaA2 in particular, an ensemble calculated based on the NMR data alone would yield an isotropic ensemble containing α helices, whereas the sole employment of SAXS would generate a compact ensemble lacking the preformed α helices. Overall, the strategy provided in this paper can be considered as an extra utensil in the structural biologist's toolbox in an attempt to quantitatively describe the solution properties of IDPs.

Another interesting issue is raised upon the application of our selection procedure on a starting pool naive to the experimental NOE data (i.e., the structures do not necessarily satisfy the NOE data independently). Selection from a starting pool generated by *Flexible-meccano*, which contains α -helical content based on the $\delta^2\text{D}$ output, provides a final model ensemble that reproduces the experimental NOE and RDC data fairly well. Interestingly, a similar trend is observed for the ensemble-averaged back-predicted NOEs as for the SAXS analysis: (1) the discrepancies between the experimental and the ensemble-averaged back-calculated data sets decrease with increasing ensemble size, and (2) some conformers individually display large violations with respect to the experimental data but are acceptable within the final model ensemble. This might indicate that NOE analysis is not only useful for establishing the local structure of secondary structure elements, but that, in the case of PaaA2 and IDPs with similar features, they may also potentially play a significant role as a filter in refining the composition of the IDP's conformational ensemble. NOEs are not conventionally used for the quantitative description of IDPs because of the sharp dependence on the exact value of the rotational correlation time, τ_c , for each particular internuclear vector (Jensen et al., 2009) and probable confusion of exchange rates between conformers and rotational correlation times. In this case, where preformed α helices appear to behave as separate structural units with respect to the rest of the polypeptide chain and each unit can be characterized by its

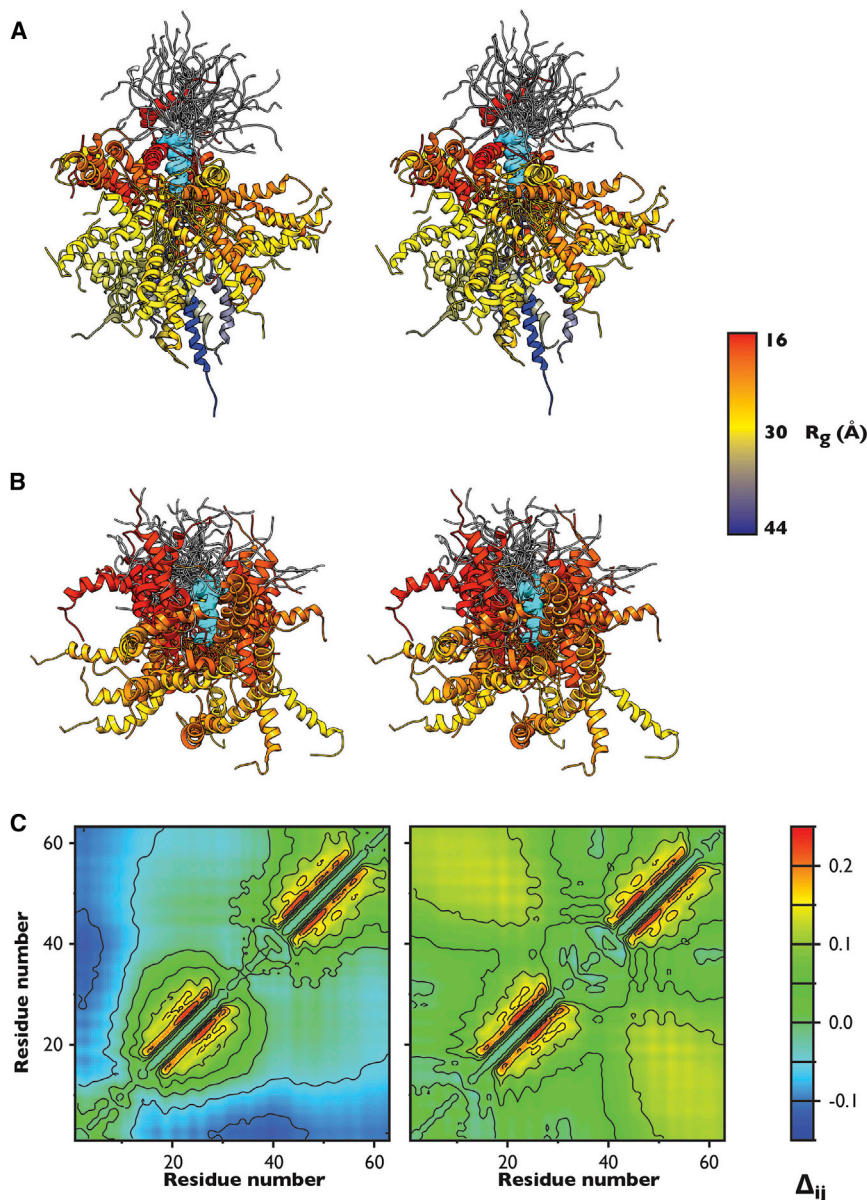


Figure 5. NMR- and SAXS-Derived Model Ensemble of PaaA2

(A and B) Stereo figures of a model ensemble of 50 conformers selected randomly from the initial NMR pool (A) and the final NMR- and SAXS-derived 50-member model ensemble of PaaA2 (B). All structures are superposed on the first helix (colored in cyan). The second helix is colored according to the R_g of the conformer (the coloring scheme is indicated by the bar).

(C) Contact maps showing chain proximity in the random model ensemble (left) and the final model ensemble derived by combining SAXS and NMR data (right). The definition of the metric Δ_{ij} is described in the Experimental Procedures.

See also Figure S5.

interactions were not observed for PaaA2. This can be explained either by a transient formation of the long-range contacts or a low fractional population of the conformers displaying such interactions. Although PaaA2 may assume a certain level of compactness, it is clear that the protein neither adopts a globular conformation nor contains a hydrophobic core. Although the former is supported by the analytical SEC and DLS experiments, the latter is evidenced by the thermal unfolding data. The moderate decrease in α helicity with increasing temperature is probably caused by a gradual melting of the two α helices. The compactness of the PaaA2 model ensemble is also interesting from a biological point of view. It has been shown that PaaA2 is extremely sensitive to proteolytic degradation *in vivo* (Hallez et al., 2010). However, in comparison with a random-coil IDP, this compactness could serve to reduce the antitoxin's

own τ_c value, these possible sources of artifact were considered acceptable. Thus, there is a gray area between the well-established traditional structure calculation approaches for globular proteins and the well-evolved approaches for determining conformational ensembles for IDPs lacking preformed elements. The results for PaaA2 presented here suggest that the application of NOEs during the calculation may be justified for at least IDPs containing preformed α helices.

The final model ensemble of the PaaA2 antitoxin displays two features that are of interest for its biological function. First, the model ensemble displays a higher level of compactness than one would expect from an IDP-like protein. This feature, which is perceived in the SAXS data because PaaA2 lies just under the Flory region (Figure 3B), seems to be typical for IDPs containing long-range contacts (Bernadó and Svergun, 2012). However, the NOE cross-peaks corresponding to these long-range

sensitivity to proteolysis in order to preserve the intricate balance necessary for the regulation and function of the *paaR2-paaA2-parE2* module. Indeed, it seems that the proteolysis of the antitoxin seems to be an important regulation mechanism in type II TA systems (Brzozowska and Zielenkiewicz, 2013). For instance, for the *hipBA* TA module, it has been shown that the activation of the HipA toxin depends on the amount of HipB antitoxin in the cell and that the degradation of HipB plays an important role in the regulation of the *hipBA* system (Rotem et al., 2010; Hansen et al., 2012).

Second, the structural study of PaaA2 reveals that the antitoxin is a monomer composed of two α helices connected by a highly flexible linker. Interestingly, these pre-existing α helices align well with the toxin-binding elements of the archetypal RelB antitoxins. Therefore, we hypothesize that these helices act as molecular recognition elements that allow PaaA2 to efficiently

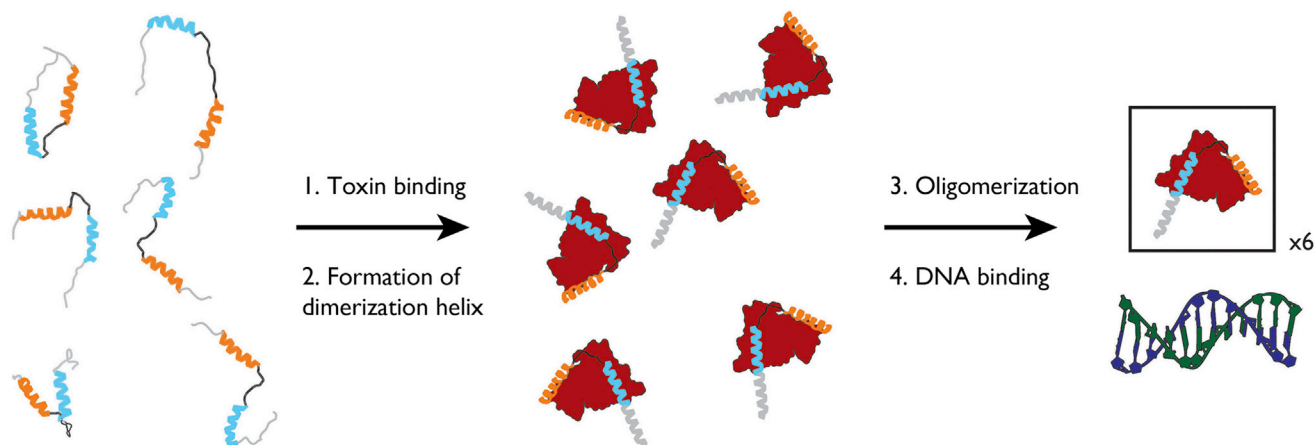


Figure 6. Schematic Model for the Roles of PaaA2 in the Context of ParE2 Neutralization, Oligomerization of the ParE2-PaaA2 Complex, and DNA Recognition

PaaA2 exists in solution as two preformed α helices (helix 1 and 2 colored in cyan and orange, respectively) connected by a highly flexible linker (depicted in black). The N and C termini (shown in gray) of PaaA2 are disordered in the unbound state. The two preformed α helices serve as initial contact points for efficient neutralization of the toxin (1). Upon binding of ParE2 (shown in red), the α helices extend thereby locking down the toxin-antitoxin interaction, which in turn causes the N terminus to adopt an α -helical structure (2). Through this newly formed “dimerization” helix, the oligomerization process toward attaining the final heterododecameric state may be initiated (3). Finally, the heterododecamer may recognize and bind the operator DNA (4).

bind the ParE2 toxin. Given that the helices are preformed, PaaA2 probably interacts with its cognate toxin through the mechanism of conformational selection: i.e., many IDPs are not fully disordered, but rather adopt a number of conformations that favor their bound state (Wright and Dyson, 2009; Tompa, 2011). This strategy is recurrent for IDPs because, compared with the fully disordered state, the occurrence of preformed elements would kinetically and energetically be more favorable for the recognition and binding of their target partner (Fuxreiter et al., 2004). Moreover, the role of PaaA2 possibly extends beyond just toxin neutralization. Recently, it has been shown that the ParE2-PaaA2 complex consists of a stable heterododecamer in solution (Sterckx et al., 2012) and that this toxin-antitoxin complex binds the promoter region of the *paaR2-paaA2-parE2* module in vivo (Hallez et al., 2010). For many examples, the importance of intrinsically disordered segments in DNA-protein interactions has been recognized (Fuxreiter et al., 2011). Based on the binding of PaaA2 to ParE2 through conformational selection, a model encompassing the role of PaaA2 in toxin neutralization, oligomerization of the toxin-antitoxin complex, and subsequent DNA recognition (and thus autoregulation of the TA operon) may be proposed (Figure 6). Although it is conceivable that PaaA2 has an important part to play in these processes following toxin recognition and binding, this remains speculative without further experimental evidence.

EXPERIMENTAL PROCEDURES

Production and Purification of *E. coli* O157 PaaA2

The gene coding for PaaA2 was cloned and expressed in *E. coli* as described before (Sterckx et al., 2012). The used construct equips PaaA2 with an N-terminal FLAG tag and ParE2 with a C-terminal His tag. The purification of the ParE2-PaaA2 complex was performed as described previously (Sterckx et al., 2012). PaaA2 was obtained from the ParE2-PaaA2 complex via an unfolding/refolding protocol (see Supplemental Information).

Analytical SEC and DLS

Analytical SEC experiments were performed using a Superdex 75 HR 10/30 column (GE Healthcare) pre-equilibrated with running buffer (20 mM phosphate, pH 6.6, 500 mM NaCl) for at least one column volume. A 500 μ l PaaA2 sample at a concentration of 2 mg/ml was injected and eluted with a flow rate of 0.3 ml/min. The column was calibrated with molecular-mass standards from Bio-Rad under the same conditions, and the apparent molecular mass of PaaA2 was determined according to Uversky (1993). DLS experiments were performed on a DynaPro™ plate reader (Wyatt Technology) in 20 mM phosphate, pH 6.6, 500 mM NaCl at room temperature. Data were acquired for different concentrations (typically 1–10 mg/ml).

CD Spectroscopy

CD spectra were recorded on a J-715 spectropolarimeter (Jasco). Continuous scans were taken using a 1 mm cuvette, a scan rate of 50 nm/min, a band width of 1.0 nm, and a resolution of 0.5 nm. Six accumulations were taken at room temperature in 20 mM phosphate, pH 6.6, 150 mM NaCl, and a PaaA2 concentration of 0.2 mg/ml. The raw CD data (ellipticity θ in mdeg) were normalized for protein concentration and the number of residues according to Equation 1, yielding the mean residue ellipticity ($[\theta]$ in deg \cdot cm 2 \cdot mol $^{-1}$)

$$[\theta] = \frac{\theta \cdot M_r}{n \cdot C \cdot l}, \quad (\text{Equation 1})$$

where M_r , n , C , and l denote the molecular mass (Da), the number of amino acids, the concentration (mg/ml), and the cuvette path length (cm), respectively. Thermal unfolding experiments were performed by gradually increasing the temperature from 278 to 368 K at a constant rate of 1.5 K/min. To follow the change in α helicity, the mean residue ellipticity was plotted as a function of temperature.

NMR Assignment and Relaxation Experiments

For the NMR experiments, a uniformly labeled ^{13}C - ^{15}N PaaA2 sample was produced in Spectra9 medium (Cambridge Isotope Laboratories) and subsequently purified as described. The final sample contained 1 mM protein (20 mM phosphate, pH 6.6, 500 mM NaCl) and 10% D_2O for the lock. All NMR spectra were acquired at 25°C on a Varian NMR Direct-Drive Systems 599.78 MHz spectrometer equipped with room temperature triple-resonance PFG-Z probe. The details concerning recorded spectra, data processing,

relaxation analysis, and model-free analysis are described in the [Supplemental Information](#).

RDC Experiments

Experimental ^1H - ^{15}N RDCs were measured for PaaA2 at 0.33 mM in 20 mM phosphate, pH 6.6, 500 mM NaCl. The sample was measured on a 600 MHz Varian spectrometer at 25°C. The protein was aligned in a suspension of filamentous phages (14 mg/ml) giving rise to a D_2O splitting of 15.3 Hz. The RDCs were measured in an HNCO-type experiment modified to allow for spin coupling measurements in the ^{13}C dimension (Rasia et al., 2011). Theoretical ^1H - ^{15}N RDCs were calculated from each of the structures in the derived model ensemble using PALES (Zweckstetter and Bax, 2000) and subsequently averaged over all conformers.

SAXS

SAXS studies were conducted at the X33 Beamline (DESY, Hamburg, Germany). Prior to analysis, the sample was dialyzed into a buffer matching the NMR conditions (50 mM MES, pH 6.6, 500 mM NaCl) and concentrated (Amicon Ultra UltraCel 3K). Aliquots of the sample were prepared at different concentrations (1, 3, 6, and 10 mg/ml). The dialysis buffer was filtered (0.22 μm) and kept as a buffer blank for the experiments. Scattering curves were collected for the different protein concentrations, and buffer data were collected before and after each sample measurement. At low salt conditions, aspecific oligomerization interfered with the SAXS analysis (data not shown), which was alleviated by increasing the conductivity of the experimental conditions. Although the use of high salt concentrations substantially improved the quality of the data, a slight concentration dependency was still noticeable. Therefore, the scattering curves collected at different concentrations were used to obtain a final zero concentration scattering curve through extrapolation (Mangenot et al., 2002; Miyamoto et al., 2010). The information on data collection and derived structural parameters is given in [Table S2](#) according to the guidelines provided by Jacques et al. (2012).

Model Ensemble Calculation

The results obtained from the NMR and SAXS experiments were combined in search of the model conformational ensemble that describes well the solution behavior of PaaA2. In the first step, a pool of 5,000 structures was generated based solely on the NMR data (CS and NOEs; [Table 1](#)) through a modified NMR structure calculation protocol. Model ensembles of varying size (N) were selected from the 5,000-member starting pool using the recorded SAXS data through EOM. Because the χ^2 values reported by EOM are very similar (1.43 ± 0.08), they could not be used on their own to select a model ensemble that best fits the experimental. This is why the search protocol was refined by applying a block jackknife procedure on the EOM-derived SAXS ensembles. The details of the calculation method are reported in the [Supplemental Information](#). The scripts used for the SAXS block jackknife calculations are available upon request. The statistics of the final model conformational ensemble of PaaA2 are reported in [Table 1](#).

ACCESSION NUMBERS

The data of the final model conformational ensemble of PaaA2 have been deposited in the Protein Data Bank under ID code 3ZBE and in the Protein Ensemble Database under ID code 5AAA.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and four tables and can be found with this article online at <http://dx.doi.org/10.1016/j.str.2014.03.012>.

AUTHOR CONTRIBUTIONS

Y.G.J.S. designed and carried out the experimental work, performed the data collection and analysis, performed research, and wrote the manuscript. A.N.V. performed the determination of the ensemble calculation, assisted with data processing and analysis, and contributed to the writing of

the manuscript. N.A.J.v.N. performed NMR data collection and assisted with data processing and analysis. A.G.-P. assisted with the SAXS data collection. W.F.V. and L.B. performed the NMR structure calculations and assisted with the $\delta 2\text{D}$ analysis. T.J. and L.V.M. assisted with the bioinformatics analysis. J.K., M.R.J., and M.B. performed the RDC experiments and analyzed the data. N.A.J.v.N. and R.L. conceived and directed the study and contributed to experimental design, data analysis, and writing of the manuscript.

ACKNOWLEDGMENTS

This research was supported by research grants from Onderzoeksraad – Vrije Universiteit Brussel, VIB, Fonds voor Wetenschappelijk Onderzoek – Vlaanderen, and the Hercules Foundation. Y.G.J.S. acknowledges the receipt of an individual predoctoral fellowship from Fonds voor Wetenschappelijk Onderzoek – Vlaanderen, whereas A.G.-P., L.B., and A.N.V. are Fonds voor Wetenschappelijk Onderzoek postdoctoral researchers. W.F.V. was funded by Brussels Institute for Research and Innovation (Innoviris) grant BB2B 2010-1-12. T.J. is a postdoctoral fellow at the FNRS. L.V.M. thanks the Fonds de la Recherche Scientifique (FRSM-3.4530.04) and the Fonds Brachet for financial support. J.K. is a fellow of the IDP-byNMR Marie Curie action of the EC (contract 264257). M.R.J. and M.B. acknowledge financial support from the French Agence Nationale de la Recherche Programme Jeunes Chercheuses et Jeunes Chercheurs Protein-Disorder (to M.R.J.) and l'Agence Nationale de la Recherche Maladie d'Alzheimer et Maladies Apparentées Project TAU-TAUSTRICT (to M.B.). For the SAXS experiments, we acknowledge use of the X33 Beamline (DESY, Hamburg, Germany).

We are grateful to Sarah Haesaerts for technical support during protein purification, the SWITCH laboratory at VIB-Belgium for access to DLS facilities, and Dr. Rodrigo Gallardo for help during data collection and analysis. Many thanks and appreciation are addressed to Cesyen Cedeno, Mickael Krzeminski, and the SAXIER forum of the BioSAXS group in Hamburg for highly valuable discussions.

Received: November 9, 2012

Revised: March 14, 2014

Accepted: March 15, 2014

Published: April 24, 2014

REFERENCES

- Bernadó, P., and Svergun, D.I. (2012). Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol. Biosyst.* 8, 151–167.
- Bernadó, P., Mylonas, E., Petoukhov, M.V., Blackledge, M., and Svergun, D.I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* 129, 5656–5664.
- Bernadó, P., Modig, K., Grela, P., Svergun, D.I., Tchorzewski, M., Pons, M., and Akke, M. (2010). Structure and dynamics of ribosomal protein L12: an ensemble model based on SAXS and NMR relaxation. *Biophys. J.* 98, 2374–2382.
- Bertini, I., Fragai, M., Luchinat, C., Melikian, M., Mylonas, E., Sarti, N., and Svergun, D.I. (2009). Interdomain flexibility in full-length matrix metalloproteinase-1 (MMP-1). *J. Biol. Chem.* 284, 12821–12828.
- Bertini, I., Giachetti, A., Luchinat, C., Parigi, G., Petoukhov, M.V., Pierattelli, R., Ravera, E., and Svergun, D.I. (2010). Conformational space of flexible biological macromolecules from average data. *J. Am. Chem. Soc.* 132, 13553–13558.
- Blower, T.R., Salmond, G.P.C., and Luisi, B.F. (2011). Balancing at survival's edge: the structure and adaptive benefits of prokaryotic toxin-antitoxin partners. *Curr. Opin. Struct. Biol.* 21, 109–118.
- Boehr, D.D., Nussinov, R., and Wright, P.E. (2009). The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* 5, 789–796.
- Brzozowska, I., and Zielenkiewicz, U. (2013). Regulation of toxin-antitoxin systems by proteolysis. *Plasmid* 70, 33–41.

- Buts, L., Lah, J., Dao-Thi, M.-H., Wyns, L., and Loris, R. (2005). Toxin-antitoxin modules as bacterial metabolic stress managers. *Trends Biochem. Sci.* **30**, 672–679.
- Camilloni, C., De Simone, A., Vranken, W.F., and Vendruscolo, M. (2012). Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry* **51**, 2224–2231.
- Christensen, S.K., Pedersen, K., Hansen, F.G., and Gerdes, K. (2003). Toxin-antitoxin loci as stress-response-elements: ChpAK/MazF and ChpBK cleave translated RNAs and are counteracted by tmRNA. *J. Mol. Biol.* **332**, 809–819.
- Christensen-Dalsgaard, M., Jørgensen, M.G., and Gerdes, K. (2010). Three new RelE-homologous mRNA interferases of *Escherichia coli* differentially induced by environmental stresses. *Mol. Microbiol.* **75**, 333–348.
- Cooper, C.R., Daugherty, A.J., Tachdjian, S., Blum, P.H., and Kelly, R.M. (2009). Role of *vapBC* toxin-antitoxin loci in the thermal stress response of *Sulfolobus solfataricus*. *Biochem. Soc. Trans.* **37**, 123–126.
- Csizmók, V., Szollosi, E., Friedrich, P., and Tompa, P. (2006). A novel two-dimensional electrophoresis technique for the identification of intrinsically unstructured proteins. *Mol. Cell. Proteomics* **5**, 265–273.
- De Jonge, N., Garcia-Pino, A., Buts, L., Haesaerts, S., Charlier, D., Zangger, K., Wyns, L., De Greve, H., and Loris, R. (2009). Rejuvenation of CcdB-poisoned gyrase by an intrinsically disordered protein domain. *Mol. Cell* **35**, 154–163.
- Doreleijers, J.F., Sousa da Silva, A.W., Krieger, E., Nabuurs, S.B., Spronk, C.A.E.M., Stevens, T.J., Vranken, W.F., Vriend, G., and Vuister, G.W. (2012). CING: an integrated residue-based structure validation program suite. *J. Biomol. NMR* **54**, 267–283.
- Durand, D., Vivès, C., Cannella, D., Pérez, J., Pebay-Peyroula, E., Vachette, P., and Fieschi, F. (2010). NADPH oxidase activator p67^(phox) behaves in solution as a multidomain protein with semi-flexible linkers. *J. Struct. Biol.* **169**, 45–53.
- Dyson, H.J., and Wright, P.E. (2005). Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197–208.
- Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans*. (Montpelier, VT: Capital City Press, The Society for Industrial and Applied Mathematics).
- Feldman, H.J., and Hogue, C.W.V. (2000). A fast method to sample real protein conformational space. *Proteins* **39**, 112–131.
- Fink, A.L. (2005). Natively unfolded proteins. *Curr. Opin. Struct. Biol.* **15**, 35–41.
- Fuxreiter, M., Simon, I., Friedrich, P., and Tompa, P. (2004). Prefolded structural elements feature in partner recognition by intrinsically unstructured proteins. *J. Mol. Biol.* **338**, 1015–1026.
- Fuxreiter, M., Simon, I., and Bondos, S. (2011). Dynamic protein-DNA recognition: beyond what can be seen. *Trends Biochem. Sci.* **36**, 415–423.
- García-Pino, A., Balasubramanian, S., Wyns, L., Gazit, E., De Greve, H., Magnuson, R.D., Charlier, D., van Nuland, N.A.J., and Loris, R. (2010). Allosteric and intrinsic disorder mediate transcription regulation by conditional cooperativity. *Cell* **142**, 101–111.
- Gerdes, K., Christensen, S.K., and Løbner-Olesen, A. (2005). Prokaryotic toxin-antitoxin stress response loci. *Nat. Rev. Microbiol.* **3**, 371–382.
- Hallez, R., Geeraerts, D., Sterckx, Y., Mine, N., Loris, R., and Van Melderen, L. (2010). New toxins homologous to ParE belonging to three-component toxin-antitoxin systems in *Escherichia coli* O157:H7. *Mol. Microbiol.* **76**, 719–732.
- Hansen, S., Vulić, M., Min, J., Yen, T.-J., Schumacher, M.A., Brennan, R.G., and Lewis, K. (2012). Regulation of the *Escherichia coli* HipBA toxin-antitoxin system by proteolysis. *PLoS ONE* **7**, e39185.
- Hayes, F., and Van Melderen, L. (2011). Toxins-antitoxins: diversity, evolution and function. *Crit. Rev. Biochem. Mol. Biol.* **46**, 386–408.
- Huang, F., Rajagopalan, S., Settanni, G., Marsh, R.J., Armoogum, D.A., Nicolaou, N., Bain, A.J., Lerner, E., Haas, E., Ying, L., and Fersht, A.R. (2009). Multiple conformations of full-length p53 detected with single-molecule fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. USA* **106**, 20758–20763.
- Jacques, D.A., Guss, J.M., Svergun, D.I., and Trehwella, J. (2012). Publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 620–626.
- Jensen, M.R., and Blackledge, M. (2008). On the origin of NMR dipolar waves in transient helical elements of partially folded proteins. *J. Am. Chem. Soc.* **130**, 11266–11267.
- Jensen, M.R., Markwick, P.R.L., Meier, S., Griesinger, C., Zweckstetter, M., Grzesiek, S., Bernadó, P., and Blackledge, M. (2009). Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. *Structure* **17**, 1169–1185.
- Jensen, M.R., Salmon, L., Nodet, G., and Blackledge, M. (2010). Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts. *J. Am. Chem. Soc.* **132**, 1270–1272.
- Lah, J., Simic, M., Vesnaver, G., Marianovsky, I., Glaser, G., Engelberg-Kulka, H., and Loris, R. (2005). Energetics of structural transitions of the addition antitoxin MazE: is a programmed bacterial cell death dependent on the intrinsically flexible nature of the antitoxins? *J. Biol. Chem.* **280**, 17397–17407.
- Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486.
- Leplae, R., Geeraerts, D., Hallez, R., Guglielmini, J., Drèze, P., and Van Melderen, L. (2011). Diversity of bacterial type II toxin-antitoxin systems: a comprehensive search and functional analysis of novel families. *Nucleic Acids Res.* **39**, 5513–5525.
- Leyrat, C., Schneider, R., Ribeiro, E.A., Jr., Yabukarski, F., Yao, M., Gérard, F.C.A., Jensen, M.R., Ruigrok, R.W.H., Blackledge, M., and Jamin, M. (2012). Ensemble structure of the modular and flexible full-length vesicular stomatitis virus phosphoprotein. *J. Mol. Biol.* **423**, 182–197.
- Madl, T., Gabel, F., and Sattler, M. (2011). NMR and small-angle scattering-based structural analysis of protein complexes in solution. *J. Struct. Biol.* **173**, 472–482.
- Maisonneuve, E., Shakespeare, L.J., Jørgensen, M.G., and Gerdes, K. (2011). Bacterial persistence by RNA endonucleases. *Proc. Natl. Acad. Sci. USA* **108**, 13206–13211.
- Mangenot, S., Leforestier, A., Vachette, P., Durand, D., and Livolant, F. (2002). Salt-induced conformation and interaction changes of nucleosome core particles. *Biophys. J.* **82**, 345–356.
- Marsh, J.A., and Forman-Kay, J.D. (2011). Ensemble modeling of protein disordered states: experimental restraint contributions and validation. *Proteins* **80**, 556–572.
- Marsh, J.A., Dancheck, B., Ragusa, M.J., Allaire, M., Forman-Kay, J.D., and Peti, W. (2010). Structural diversity in free and bound states of intrinsically disordered protein phosphatase 1 regulators. *Structure* **18**, 1094–1103.
- Metcalfe, E.E., Zmoon, J., Thomas, D.D., and Veglia, G. (2004). ¹H/¹⁵N heteronuclear NMR spectroscopy shows four dynamic domains for phospholamban reconstituted in dodecylphosphocholine micelles. *Biophys. J.* **87**, 1205–1214.
- Mittag, T., and Forman-Kay, J.D. (2007). Atomic-level characterization of disordered protein ensembles. *Curr. Opin. Struct. Biol.* **17**, 3–14.
- Mittag, T., Marsh, J., Grishaev, A., Orlicky, S., Lin, H., Sichi, F., Tyers, M., and Forman-Kay, J.D. (2010). Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase. *Structure* **18**, 494–506.
- Miyamoto, Y., Nishimura, S., Inoue, K., Shimamoto, S., Yoshida, T., Fukuhara, A., Yamada, M., Urade, Y., Yagi, N., Ohkubo, T., and Inui, T. (2010). Structural analysis of lipocalin-type prostaglandin D synthase complexed with biliverdin by small-angle X-ray scattering and multi-dimensional NMR. *J. Struct. Biol.* **169**, 209–218.
- Nodet, G., Salmon, L., Ozenne, V., Meier, S., Jensen, M.R., and Blackledge, M. (2009). Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. *J. Am. Chem. Soc.* **131**, 17908–17918.
- Ozenne, V., Bauer, F., Salmon, L., Huang, J.-R., Jensen, M.R., Segard, S., Bernadó, P., Charavay, C., and Blackledge, M. (2012). *Flexible-meccano*: a

- tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics* 28, 1463–1470.
- Rasia, R.M., Lescop, E., Palatnik, J.F., Boisbouvier, J., and Brutscher, B. (2011). Rapid measurement of residual dipolar couplings for fast fold elucidation of proteins. *J. Biomol. NMR* 51, 369–378.
- Rotem, E., Loinger, A., Ronin, I., Levin-Reisman, I., Gabay, C., Shoresh, N., Biham, O., and Balaban, N.Q. (2010). Regulation of phenotypic variability by a threshold-based mechanism underlies bacterial persistence. *Proc. Natl. Acad. Sci. USA* 107, 12541–12546.
- Salmon, L., Nodet, G., Ozenne, V., Yin, G., Jensen, M.R., Zweckstetter, M., and Blackledge, M. (2010). NMR characterization of long-range order in intrinsically disordered proteins. *J. Am. Chem. Soc.* 132, 8407–8418.
- Shah, D., Zhang, Z., Khodursky, A.B., Kaldalu, N., Kurg, K., and Lewis, K. (2006). Persisters: a distinct physiological state of *E. coli*. *BMC Microbiol.* 6, 53–61.
- Sterckx, Y.G.-J., Garcia-Pino, A., Haesaerts, S., Jové, T., Geerts, L., Sakellaris, V., Van Melderen, L., and Loris, R. (2012). The ParE2-PaaA2 toxin-antitoxin complex from *Escherichia coli* O157 forms a heterododecamer in solution and in the crystal. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 68, 724–729.
- Tompa, P. (2002). Intrinsically unstructured proteins. *Trends Biochem. Sci.* 27, 527–533.
- Tompa, P. (2011). Unstructural biology coming of age. *Curr. Opin. Struct. Biol.* 21, 419–425.
- Tompa, P. (2012). On the supertertiary structure of proteins. *Nat. Chem. Biol.* 8, 597–600.
- Uversky, V.N. (1993). Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry* 32, 13288–13298.
- Wells, M., Tidow, H., Rutherford, T.J., Markwick, P., Jensen, M.R., Mylonas, E., Svergun, D.I., Blackledge, M., and Fersht, A.R. (2008). Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl. Acad. Sci. USA* 105, 5762–5767.
- Wishart, D.S., Sykes, B.D., and Richards, F.M. (1991). Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol.* 222, 311–333.
- Wright, P.E., and Dyson, H.J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* 293, 321–331.
- Wright, P.E., and Dyson, H.J. (2009). Linking folding and binding. *Curr. Opin. Struct. Biol.* 19, 31–38.
- Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*. (New York: John Wiley & Sons, Inc.).
- Zweckstetter, M., and Bax, A. (2000). Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. *J. Am. Chem. Soc.* 122, 3791–3792.

Targeting the disordered C terminus of PTP1B with an allosteric inhibitor

Navasona Krishnan¹, Dorothy Koveal², Daniel H Miller^{3,4}, Bin Xue^{1,7}, Sai Dipikaa Akshinthala¹, Jaka Kragelj⁵, Malene Ringkjøbing Jensen⁵, Carla-Maria Gauss¹, Rebecca Page², Martin Blackledge⁵, Senthil K Muthuswamy^{1,6}, Wolfgang Peti^{3,4} & Nicholas K Tonks^{1*}

PTP1B, a validated therapeutic target for diabetes and obesity, has a critical positive role in HER2 signaling in breast tumorigenesis. Efforts to develop therapeutic inhibitors of PTP1B have been frustrated by the chemical properties of the active site. We define a new mechanism of allosteric inhibition that targets the C-terminal, noncatalytic segment of PTP1B. We present what is to our knowledge the first ensemble structure of PTP1B containing this intrinsically disordered segment, within which we identified a binding site for the small-molecule inhibitor MSI-1436. We demonstrate binding to a second site close to the catalytic domain, with cooperative effects between the two sites locking PTP1B in an inactive state. MSI-1436 antagonized HER2 signaling, inhibited tumorigenesis in xenografts and abrogated metastasis in the NDL2 mouse model of breast cancer, validating inhibition of PTP1B as a therapeutic strategy in breast cancer. This new approach to inhibition of PTP1B emphasizes the potential of disordered segments of proteins as specific binding sites for therapeutic small molecules.

Disruption of the normal patterns of protein phosphorylation results in aberrant regulation of signal transduction and has been implicated in the etiology of a variety of major human diseases. The ability to modulate signaling pathways selectively holds enormous therapeutic potential. The first drugs directed against protein tyrosine kinases (PTKs) represent breakthroughs in cancer therapy. For example, the humanized antibody trastuzumab (Herceptin) targets the PTK HER2 (ERBB2), which is amplified and/or overexpressed in ~25% of breast tumors, where it is associated with poor prognosis^{1,2}. Although trastuzumab is a treatment of choice, the overall success rate is low, and patients develop resistance to the drug. Similar problems have limited the success of other PTK-directed inhibitors^{3,4}. Therefore, it is anticipated that alternative therapies, which target simultaneously different signaling enzymes and processes, may be more effective than targeting individual PTKs alone. Consequently, a major problem remains the identification of such alternative therapies.

The focus on PTKs for drug development ignores the other major component of phosphorylation-dependent regulation of signaling. Protein phosphorylation is a reversible process in which the coordinated and competing activities of kinases and phosphatases are important for determining signaling outcome, but the protein tyrosine phosphatases (PTPs) remain a largely untapped resource for drug development. Since its discovery 25 years ago, PTP1B has become a highly validated therapeutic target for diabetes and obesity⁵. Consequently, there have been major programs in industry focused on developing small-molecule inhibitors of PTP1B that followed standard procedures of targeting the active site; however, these efforts have been frustrated by technical challenges arising from the chemistry of PTP catalysis. Although it was possible to generate potent, specific and reversible inhibitors of

PTP1B, such molecules were highly charged and thus of limited drug development potential⁶.

PTP1B function is not restricted to metabolic regulation; it is overexpressed in breast tumors together with HER2 (refs. 7,8). Mice expressing activated alleles of HER2 in mammary glands develop multiple mammary tumors and frequent metastases to the lung; however, when such mice were crossed with PTP1B-null mice, tumor development was delayed, and the incidence of lung metastases was decreased. Conversely, targeted overexpression of PTP1B alone was sufficient to drive mammary tumorigenesis⁷. These observations suggest that PTP1B may have a positive role in promoting signaling events associated with breast tumorigenesis. Therefore, inhibition of PTP1B function may represent a novel therapeutic strategy to address not only diabetes and obesity but also mammary tumorigenesis and malignancy. Consequently, new approaches to inhibition of PTP1B, which circumvent the problems with active site-directed small-molecule inhibitors, are required to reinvigorate drug development efforts against this highly validated target.

Our approach has been to develop allosteric inhibitors of PTP1B that bind to unique sites on the enzyme distinct from the active site. PTP1B was purified originally from human placenta as a 37-kDa catalytic domain comprising residues 1–321 (ref. 9) and has been the focus of attention to date for mechanistic analysis as well as for drug screening. Nevertheless, PTP1B exists *in vivo* as a longer protein of ~50 kDa, in which the C-terminal segment, which is deleted from the 37-kDa protein, serves a regulatory function¹⁰. We have demonstrated that an aminosterol natural product, MSI-1436 (also known as trodusquemine)¹¹, inhibited the full-length form of PTP1B preferentially in a reversible, selective manner. We have identified the binding sites for MSI-1436 in PTP1B and defined the mechanism of inhibition. Furthermore, we have demonstrated that by targeting PTP1B, MSI-1436 attenuated HER2 signaling, resulting

¹Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA. ²Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, Rhode Island, USA. ³Department of Molecular Pharmacology, Physiology and Biotechnology, Brown University, Providence, Rhode Island, USA. ⁴Department of Chemistry, Brown University, Providence, Rhode Island, USA. ⁵Protein Dynamics and Flexibility, Institut de Biologie Structurale Jean-Pierre Ebel, CEA, CNRS, Grenoble, France. ⁶Department of Medical Biophysics, Ontario Cancer Institute, Campbell Family Institute for Breast Cancer Research, University of Toronto, Toronto, Canada. ⁷Present address: Department of Molecular and Cell Biology, University of California–Berkeley, Berkeley, California, USA. *e-mail: tonks@cshl.edu

in extensive inhibition of tumor growth and abrogation of metastasis to the lung in HER2-positive animal models of breast cancer. Overall, these data establish that PTP1B is a bona fide target for therapeutic intervention in HER2-positive cancer and illustrate a novel mechanism for specific inhibition of PTP1B through which such intervention may be achieved.

RESULTS

MSI-1436 was a noncompetitive inhibitor of PTP1B

Using a small-molecule fluorescent substrate 6,8-difluoro-4-methylumbelliferyl phosphate (DiFMUP), we observed no change in K_m but a decrease in V_{max} (Supplementary Results, Supplementary Fig. 1). MSI-1436 was a more potent inhibitor of PTP1B₁₋₄₀₅, which contains the noncatalytic C-terminal segment of the protein, than of the PTP1B₁₋₃₂₁ catalytic domain (Fig. 1a,b). MSI-1436 was a reversible noncompetitive inhibitor of both forms of the enzyme, although it had a sevenfold higher affinity for PTP1B₁₋₄₀₅ (0.6 μ M) compared to PTP1B₁₋₃₂₁ (4 μ M). Assays performed with tyrosine-phosphorylated reduced carboxamidomethylated, maleylated lysozyme yielded a similar K_i of 600 nM for PTP1B₁₋₄₀₅. Inhibition was independent of incubation time and was reversible (Supplementary Fig. 2). Although MSI-1436 inhibited PTP1B₁₋₄₀₅, it was less effective against a panel of seven PTPs and two dual-specificity phosphatases (Supplementary Table 1).

We investigated the stoichiometry of the enzyme-inhibitor complex using isothermal titration calorimetry and demonstrated that binding of MSI-1436 to the 'short form' of PTP1B₁₋₃₂₁ was exothermic and fitted a one-site reaction with a K_d of 2 μ M. In contrast, the isotherm generated by binding to PTP1B₁₋₄₀₅ was distinct and best fitted a two-site model, with a K_d of 0.3 μ M for one site and 3.0 μ M for the second site (Fig. 1c). To validate the two-site model for

PTP1B₁₋₄₀₅, we used [³H]MSI-1436 and measured inhibitor binding directly by scintillation counting, revealing a stoichiometry of 2 mol/mol for PTP1B₁₋₄₀₅ and 1 mol/mol for PTP1B₁₋₃₂₁ (Fig. 1d). The calculated binding constants correlated well with the enzyme kinetics data. Scatchard analysis revealed a pronounced concave curvature in the plot, indicative of positive cooperativity between the two sites in PTP1B₁₋₄₀₅. Furthermore, a Hill plot indicated a Hill co-efficient (n_H) of 1.3, also consistent with positive cooperativity (Fig. 1e). In addition, we performed a displacement titration with a previously characterized allosteric inhibitor, a derivative of benzbromarone¹². We observed that this benzbromarone derivative displaced MSI-1436 completely from PTP1B₁₋₃₂₁, whereas only 50% displacement was achieved with PTP1B₁₋₄₀₅ (Fig. 1f). These data illustrate that MSI-1436 is a selective, noncompetitive, reversible inhibitor of PTP1B, using a novel mechanism of inhibition with a unique binding site in the C terminus of the protein.

MSI-1436 induced a conformational change in PTP1B

In previous attempts to crystallize forms of PTP1B containing the C-terminal, noncatalytic segment, we noted that this segment remained largely unstructured. Consistent with this, dynamic light scattering revealed that the radius of hydration (R_h) of PTP1B₁₋₃₉₃ (32.8 $\text{\AA} \pm 0.2 \text{\AA}$ s.e.m.) was $>8 \text{\AA}$ larger than that of PTP1B₁₋₃₀₁ (24.1 $\text{\AA} \pm 0.4 \text{\AA}$; error represents s.e.m.), demonstrating that PTP1B residues 300–393 must adopt an extended structure in solution (Supplementary Fig. 3). Although such intrinsically disordered segments in proteins have been implicated in oligomerization, such as dimer formation^{13,14}, we demonstrated that MSI-1436 did not cause the phosphatase to oligomerize. In fact, PTP1B₁₋₄₀₅ was retained on the gel filtration column for longer in the presence compared to the absence of MSI-1436; the calculated molecular

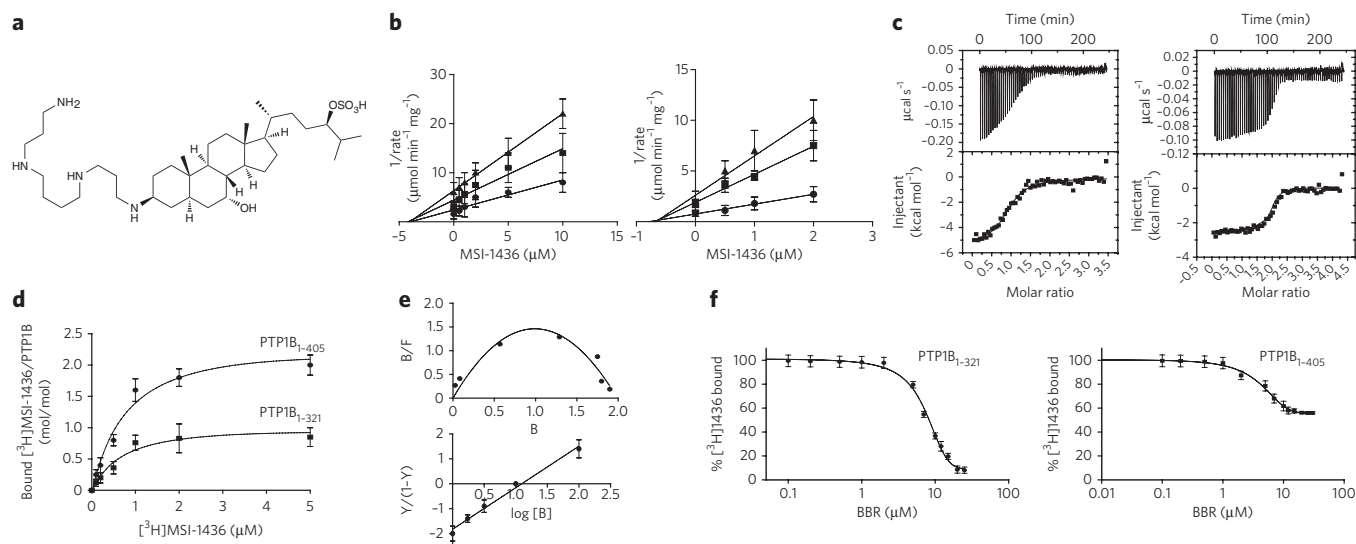
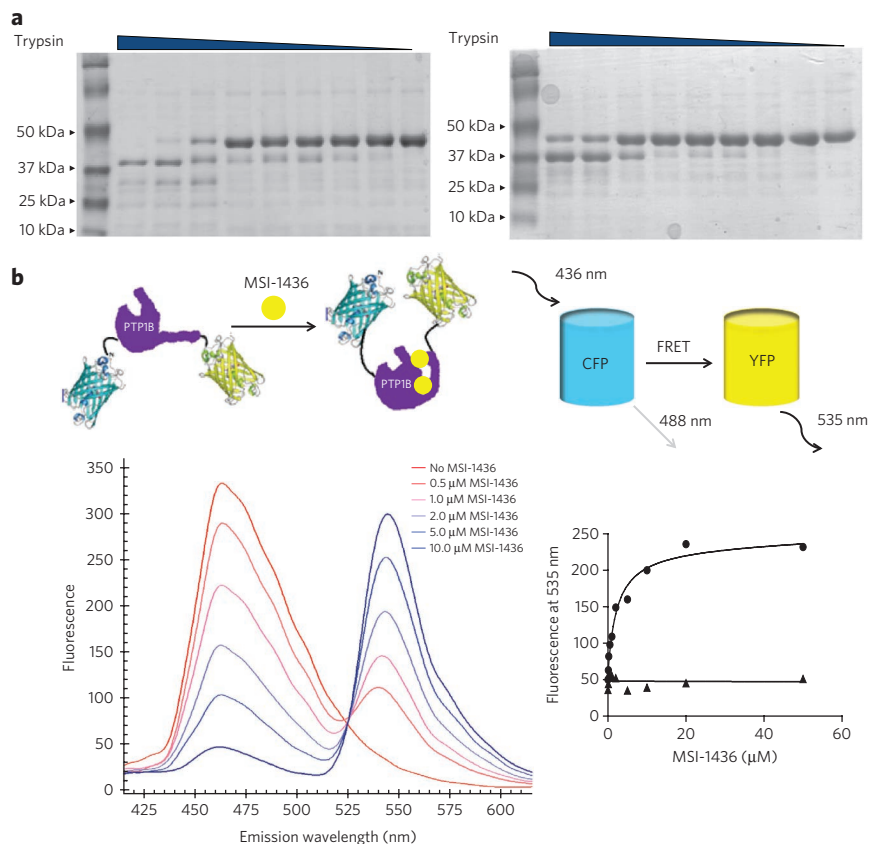


Figure 1 | MSI-1436 was a noncompetitive inhibitor of PTP1B. (a) Structure of MSI-1436. (b) Dixon plot for PTP1B₁₋₃₂₁ (left) and PTP1B₁₋₄₀₅ (right) showing $1/\text{rate}$ versus varying concentrations of MSI-1436 at substrate concentrations of 2 μ M (filled circles), 5 μ M (filled squares) and 10 μ M (filled triangles). The lines intersect on the x axis to give a K_i of 4 μ M for PTP1B₁₋₃₂₁ and 0.6 μ M for PTP1B₁₋₄₀₅. Data are representative of three independent experiments. (c) Isothermal titration calorimetry analysis. Top, binding thermograms for PTP1B₁₋₃₂₁ (left) and PTP1B₁₋₄₀₅ (right). Bottom, nonlinear least-square curves were fitted to enthalpies integrated from the individual titrations. The binding constants are 300 nM and 1 μ M for PTP1B₁₋₄₀₅ and 2 μ M for PTP1B₁₋₃₂₁, respectively. Data are representative of three independent experiments. (d) Binding of radiolabeled MSI-1436. PTP1B₁₋₄₀₅ (filled circles) or PTP1B₁₋₃₂₁ (filled squares) (100 nM) was incubated with varying concentrations of [³H]1436, and then the PTP1B-[³H]1436 complex was separated from unbound compound, and bound radioactivity (c.p.m.) was measured. The stoichiometry was found to be 1:1 for PTP1B₁₋₃₂₁-MSI-1436 and 1:2 for the PTP1B₁₋₄₀₅-MSI-1436 complex. Data represent mean \pm s.e.m. from three independent experiments. (e) Scatchard analysis of MSI-1436 binding to PTP1B₁₋₄₀₅ (top). B and F represent bound and free MSI-1436 (μ M). Hill plot (bottom), where Y represents fraction of binding sites that are occupied by MSI-1436 on PTP1B₁₋₄₀₅. The slope yields a Hill coefficient of 1.3 and a K_d of 0.7 μ M. Data are representative of three independent experiments. (f) Displacement titrations. PTP1B₁₋₃₂₁ (left) or PTP1B₁₋₄₀₅ (right) (100 nM) was incubated with 5 μ M tritiated MSI-1436 for 5 min in assay buffer. Following the incubation, varying concentrations of a benzbromarone derivative (BBR) (0–10 μ M) were added for 30 min. The extent of MSI-1436 displaced by BBR was determined by measuring the residual [³H]MSI-1436 bound to the enzyme. The experiment was repeated three times, and error bars represent s.e.m.

Figure 2 | MSI-1436 induced a conformational change in PTP1B.

(a) SDS-PAGE following limited proteolysis of PTP1B₁₋₄₀₅ with trypsin in the absence (left) and presence (right) of MSI-1436. Purified PTP1B₁₋₄₀₅ (2 μM) was incubated without (left) or with (right) MSI-1436 (5 μM) for 30 min with trypsin/PTP1B (mol/mol) ratios of (from left to right) 1:10, 0.5:10, 0.25:10, 0.125:10, 0.06:10, 0.03:10, 0.01:10, 0.005:10 and 0:10. Reactions were terminated with SDS sample buffer, and the digested products were analyzed by 20% SDS-PAGE and stained with Coomassie blue. **(b)** Top, schematic model of the PTP1B fusion protein tagged with CFP and YFP on the N and C terminus, respectively, showing a conformational change induced by binding of the inhibitor (MSI-1436) that results in a change in FRET. Bottom left, the effect of MSI-1436 binding to PTP1B₁₋₄₀₅ was followed by FRET. Representative spectra from the experiment in which CFP-PTP1B₁₋₄₀₅-YFP was titrated with MSI-1436, covering a concentration range of 0–10 μM and revealing the inhibitor-induced decrease in CFP emission at 475 nm and concomitant increase in the YFP emission at 535 nm. Bottom right, the increase in YFP with addition of MSI-1436 was plotted against the concentration of MSI-1436 to derive a K_d of 0.8 μM.



weight of PTP1B₁₋₄₀₅ was ~47,000 in the absence and ~43,000 in the presence of the inhibitor (**Supplementary Fig. 4**). In contrast, there was no change in the elution of PTP1B₁₋₃₂₁, which displayed an apparent molecular weight of 37,000 (**Supplementary Fig. 5**). Intrinsically disordered segments of proteins are also known to be sensitive to proteases. Therefore, we tested the effect of saturating concentrations of MSI-1436 on sensitivity to trypsin and observed that PTP1B₁₋₄₀₅ adopted a more trypsin-resistant state (**Fig. 2a** and **Supplementary Fig. 6**). These data suggest that inhibitor binding may impose a more compact structure on the flexible C terminus in PTP1B₁₋₄₀₅, thereby decreasing the effective hydrodynamic radius of the protein and protecting the enzyme from proteolysis.

We generated a construct in which the two fluorophores CFP and YFP were fused to the N and C terminus of PTP1B, respectively, to investigate MSI-1436-induced conformational changes by intramolecular fluorescence resonance energy transfer (FRET). The CFP-PTP1B-YFP fusion protein was excited at 440 nm. The emission peak of CFP is expected to be at 475 nm; however, if CFP and YFP were brought into close proximity by inhibitor binding, the energy from CFP would be transferred to YFP, causing the YFP emission peak at 535 nm and a concomitant decrease in the emission peak for CFP at 475 nm (**Fig. 2b**). At low concentrations of MSI-1436 (tenfold below the K_i), we observed a strong emission peak at 475 nm (CFP emission) and a relatively smaller peak at 535 nm (YFP emission). As the concentration of MSI-1436 was increased, a strong FRET signal between CFP and YFP was observed. A plot of the emission at 535 nm against the concentration of MSI-1436 yielded a K_d of 0.8 μM for MSI-1436 binding, in close agreement with the K_i (**Fig. 2b**). In contrast, excitation of CFP-PTP1B₁₋₃₂₁-YFP at 440 nm did not result in FRET at any concentration of MSI-1436 tested (**Supplementary Fig. 7**), highlighting the importance of the C-terminal segment for the recognition of the inhibitor by PTP1B. The data suggest that binding of MSI-1436 induced a conformational change that brought the C terminus of CFP-PTP1B₁₋₄₀₅-YFP closer to the N terminus,

which agrees with our gel filtration data, indicating that inhibitor binding induced a more compact structure.

The structure of PTP1B₁₋₃₉₃ defined by NMR spectroscopy

To understand the interaction of MSI-1436 with the long form of PTP1B, structural data were essential. Biomolecular NMR spectroscopy is a technique that permits structural analysis of extended, flexible segments of biomacromolecules. To overcome the technical complexities that are characteristic of NMR analysis of proteins ≥ 35 kDa, we expressed all of the proteins in D₂O-based medium, and transverse relaxation optimized spectroscopy (TROSY) versions of all of the two- and three-dimensional experiments were recorded at a ¹H Larmor frequency of 800 MHz or 850 MHz¹⁵. To overcome spectral overlap, two-dimensional [¹H,¹⁵N] TROSY spectra of singly amino-acid-labeled (with ¹⁵N-Leu, ¹⁵N-Tyr, ¹⁵N-Phe or ¹⁵N-Val) PTP1B were used both to verify existing assignments and to facilitate the assignment of additional residues. The chemical shift index calculated from the deviations of C α and C β chemical shifts from random coil values (RefDB¹⁶) correlated very well with secondary structures observed in the crystal structure of PTP1B (Protein Data Bank (PDB) code 1SUG¹⁷).

Peaks corresponding to residues 300–393 of PTP1B in the two-dimensional [¹H,¹⁵N] TROSY spectrum were on average five times more intense than those from PTP1B₁₋₃₀₁ and clustered in the center of the spectrum (7.5–8.5 p.p.m., ¹H dimension; **Fig. 3a**). This is typical for flexible and mostly unstructured amino acids. Cross-peaks in the two-dimensional [¹H,¹⁵N] TROSY spectrum of PTP1B₁₋₃₀₁ overlaid almost perfectly with the corresponding peaks in the two-dimensional [¹H,¹⁵N] TROSY spectrum of PTP1B₁₋₃₉₃ (**Fig. 3a** and **Supplementary Fig. 8**). This indicates that residues 300–393 were flexible and intrinsically disordered, did not interact with and extended away from PTP1B₁₋₃₀₁.

Secondary structure propensity scores derived from the chemical shift index revealed that PTP1B residues 300–393 were not entirely unstructured; instead, they contained two α -helices, composed of

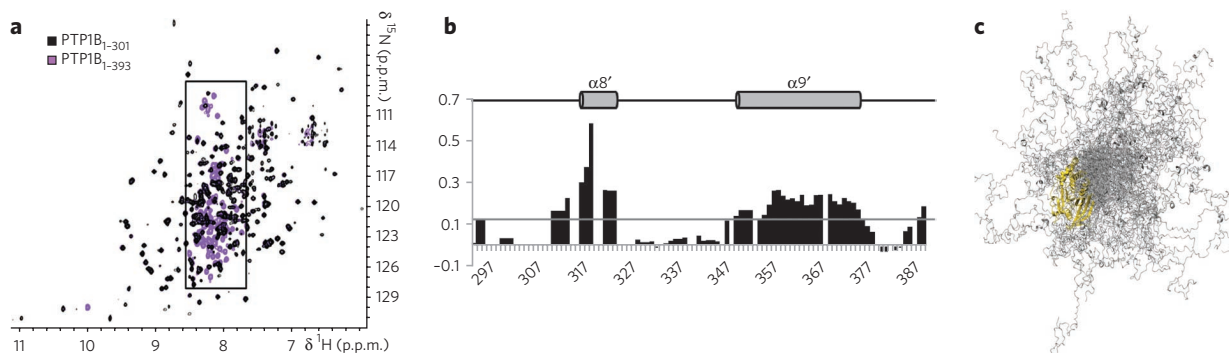


Figure 3 | PTP1B residues 300–393 were flexible and predominantly disordered. (a) Overlay of the two-dimensional $[\text{H},^{15}\text{N}]$ TROSY spectra of PTP1B₁₋₃₀₁ (black) and PTP1B₁₋₃₉₃ (purple). Boxed region (7.5–8.5 p.p.m., $^{\text{H}}$ dimension) marks PTP1B residues 300–393. (b) Secondary structure propensity scores for PTP1B residues 300–393. A score of 1 indicates a fully populated α -helix, whereas a score of 0.5 indicates an α -helix that is $\sim 50\%$ populated in solution. Helices $\alpha 8'$ and $\alpha 9'$, which are 20% populated, are denoted with a gray cylinder and labeled. (c) Representative ensemble of 100 conformers that are superimposed over PTP1B residues 1–284 (gold). PTP1B residues 285–398 (gray) sample a wide range of conformational space.

residues 320–327 ($\alpha 8'$) and 360–377 ($\alpha 9'$), which were, on average, $\sim 33\%$ and $\sim 22\%$ populated in solution, respectively (Fig. 3b).

To gain further insights into the structure of PTP1B₁₋₃₉₃, we modeled an ensemble structure of PTP1B₁₋₃₉₃, rather than a single structure, using the crystal structure of PTP1B (PDB code 1SUG), NMR chemical shift and SAXS data (Supplementary Fig. 9a–e). Experimental chemical shifts were matched with predicted $\text{C}\alpha$, $\text{C}\beta$, N and H^{N} chemical shifts for PTP1B residues 285–393. We built 15,000 starting conformers, and the ensemble exhibited, as expected, a Gaussian distribution of R_g values. However, this average-calculated SAXS curve did not fit the experimental curve, which, again as expected, did not follow a Gaussian distribution; rather, it was composed of more-than-average compact and extended conformations (Supplementary Fig. 9f,g). Therefore, from this initial pool, sub-ensembles containing different numbers of conformers were selected for which the average theoretical curve reproduced the experimental SAXS curve (Supplementary Fig. 9h). Selected ensembles of 100 conformers reproduced the SAXS curve and the NMR chemical shifts for the 108 C-terminal residues remarkably well. Thus, these ensembles provide an atomic representation of PTP1B₁₋₃₉₃ that is in excellent agreement with all of the experimental data (Fig. 3c).

MSI-1436 bound to the disordered C terminus of PTP1B

To define the primary binding site within the C-terminal segment of PTP1B, we tested the interaction between PTP1B₁₋₃₉₃ and MSI-1436 at atomic resolution using NMR spectroscopy. We recorded two-dimensional $[\text{H},^{15}\text{N}]$ TROSY spectra of PTP1B₁₋₃₀₁ and PTP1B₁₋₃₉₃ with MSI-1436 (Fig. 4a,b). No chemical shift perturbations (CSPs) were detected upon addition of MSI-1436 to PTP1B₁₋₃₀₁; however, NMR CSP mapping showed that Arg371, Arg373 and Val375 in $\alpha 9'$ and residues Leu299, His310, Ile311, Val334 and Ser393 were most affected by MSI-1436 binding to PTP1B₁₋₃₉₃ (Fig. 4c,d), consistent with the importance of the C-terminal segment for the interaction with MSI-1436.

We generated truncation mutants of PTP1B and observed that the inhibitory potency of MSI-1436 was lower for the two truncation mutants PTP1B₁₋₃₆₇ and PTP1B₁₋₃₄₆ than it was for PTP1B₁₋₄₀₅ and PTP1B₁₋₃₉₄; the potency values for the truncation mutants were similar to that of the catalytic domain construct PTP1B₁₋₃₂₁ (Supplementary Fig. 10). These data indicate that the primary binding site must be located between residues 367–394, which contain helix $\alpha 9'$ (Fig. 4c). Three point mutations to convert residues Val370, Ser372 and Arg373 to alanine did not affect enzymatic

activity or inhibition by MSI-1436 (data not shown). Nevertheless, helix-destabilizing mutations, replacing the residues with proline, increased the K_i for MSI-1436 10–20 fold. The S372P and R373P mutations, which were centered in helix $\alpha 9'$, had a greater effect than V370P (Fig. 5a). Taken together, these data show that the primary PTP1B binding site of MSI-1436 is helix $\alpha 9'$.

Identification of a second binding site for MSI-1436

The ability of MSI-1436 to inhibit PTP1B₁₋₃₂₁, albeit with reduced activity compared to PTP1B₁₋₄₀₅, suggested the presence of a binding site within this truncated form of the phosphatase. Previously, an allosteric site (exosite) within the PTP1B catalytic domain was identified that binds a series of benzofuran compounds (benzbromarones) that bear no structural similarity to MSI-1436. These compounds were shown to stabilize an inactive state of the enzyme in which the active site cleft is locked in the open conformation¹². Mutation of residues known to be critical for benzobromarone binding also impaired inhibition by MSI-1436 (Supplementary Fig. 11 and Supplementary Tables 2 and 3). The most prominent effect on K_i was observed following mutation of Leu192 to alanine; the mutant displayed a K_i of 8 μM , which was more than tenfold greater than that of wild-type PTP1B₁₋₄₀₅. This and the results of the competition studies between benzobromarone and MSI-1436 (Fig. 1f) indicate that there is some overlap between this exosite and the second MSI-1436 binding site.

Nevertheless, there are unique features to the second MSI-1436 binding site. No interaction was detected between PTP1B₁₋₃₀₁ and MSI-1436 by NMR spectroscopy, indicating that residues within the 20 amino acids at the C terminus of the catalytic domain were important for this second binding site. In fact, NMR CSP mapping revealed that Leu299, His310 and Ile311 were affected by MSI-1436 binding, consistent with a contribution to the modest inhibition of PTP1B₁₋₃₂₁ by MSI-1436. Residues 302–321 in PTP1B flank helix $\alpha 7$, which, as shown in our ensemble PTP1B structure, can adopt numerous conformations. Mutation of residues Val287, Lys292 and Leu294 to alanine impaired inhibition by MSI-1436 (Supplementary Tables 2 and 3), indicating the importance of $\alpha 7$ helix in mediating MSI-1436 binding. This indicates that MSI-1436 targets a site that is distinct from but incorporates features of the already characterized allosteric site, stabilizing an inactive conformation of PTP1B by a novel mechanism that involves the C-terminal segment and a second binding site close to the catalytic segment.

The importance of the combined effects of both binding sites for MSI-1436 in the 'long form' of PTP1B was revealed by the fact

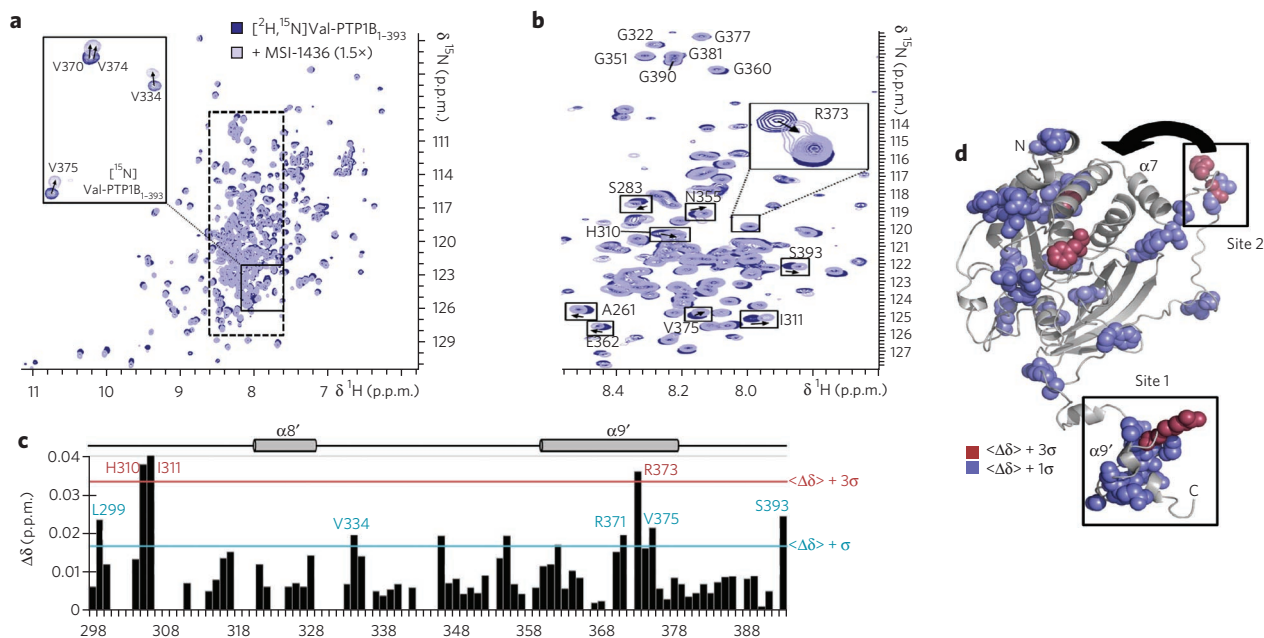


Figure 4 | PTP1B residues were perturbed upon MSI-1436 binding. (a) Overlay of the two-dimensional [$^2\text{H},^{15}\text{N}$] TROSY spectra of [$^2\text{H},^{15}\text{N}$] PTP1B₁₋₃₉₃ with (light blue) and without (dark blue) MSI-1436 (1.5 \times molar excess). Inset: two-dimensional [$^2\text{H},^{15}\text{N}$] TROSY spectra of [^{15}N]Val-PTP1B₁₋₃₉₃ (same coloring scheme). Dashed box marks the region of the spectrum shown in b. (b) Segment of the two-dimensional [$^2\text{H},^{15}\text{N}$] TROSY spectrum corresponding to PTP1B residues 300-393. (c) CSP mapping of MSI-1436 binding to PTP1B residues 300-393. Colored lines indicate one (blue) and three (red) s.d. from the mean $<\Delta\delta>$. (d) CSPs mapped onto the structure of a single conformer of PTP1B₁₋₃₉₃, using the same color scheme as in c. MSI-1436 binding sites 1 and 2 are highlighted, and the close proximity of binding site 2 to the exosite is highlighted by an arrow.

that a double point mutant in which one residue from each site was mutated, Ser372 to proline and Leu192 to alanine, preserved catalytic activity but was insensitive to MSI-1436 (Fig. 5b and Supplementary Fig. 11b).

MSI-1436 attenuated HER2-mediated tumorigenesis

To determine whether this novel mechanism for inhibition of PTP1B has therapeutic potential, we examined the effects of MSI-1436 in models of HER2-dependent breast cancer. We used MCF10A mammary epithelial cells that express a well-characterized chimeric form of HER2/ErbB2 (10A.B2), the activity of which can be induced with a small-molecule dimerizer, AP1510 (ref. 18). Activation of HER2 by AP1510 induced characteristic changes in acinar morphogenesis, proliferation and luminal apoptosis in three-dimensional culture, all of which were blocked by MSI-1436 (Supplementary Fig. 12a). Furthermore, MSI-1436 inhibited migration of 10A.B2 cells in two-dimensional culture to a similar extent to PTP1B-directed shRNA (Supplementary Fig. 12b). Similarly, MSI-1436 attenuated migration of BT474 and SKBR3 cells, which are HER2 positive, but not MCF-7 and MDA-MB-231 cells, which are HER2 negative (Supplementary Fig. 12c). Finally, we observed that MSI-1436 inhibited proliferation (Supplementary Fig. 12d) but had a limited effect on apoptosis (Supplementary Fig. 12e) in a panel of breast cancer cells that are sensitive to the HER2/EGF receptor dual kinase inhibitor lapatinib¹⁹. These data illustrate that MSI-1436 is an inhibitor of HER-induced signaling in cells; however, it is possible that MSI-1436 may also affect other pathways by targeting PTP1B.

The effect of MSI-1436 was also assessed on breast tumorigenesis in animals. Using a xenograft model, in which HER2-positive BT474 cells expressing luciferase were injected orthotopically into the mammary fat pad, we tested the effect of administering MSI-1436 both as soon as the cells were injected and following the appearance of palpable tumors. In both cases, treatment with the PTP1B inhibitor intraperitoneally reduced tumor burden (Fig. 6a). In addition, we tested the effects of MSI-1436 in the ND2L2 transgenic

mouse model of human breast cancer that expresses an activated mutant form of Neu, the rat homolog of human HER2, and develops multiple mammary tumors with metastasis to the lung^{7,20}. ND2L2 animals were allowed to develop palpable tumors and then were randomly assigned into two groups to receive vehicle control (saline) or MSI-1436 (5 mg kg⁻¹) intraperitoneally every third day. MSI-1436-treated animals displayed a marked decrease in tumor size and tumor number relative to the control group (Fig. 6b). Even more remarkable was the observation that all of the control animals developed lung metastases, although the extent of metastasis varied within the group. In contrast, metastasis was essentially abrogated in the MSI-1436-treated animals. In two separate studies, conducted with six animals in each group, we detected only one drug-treated animal that developed a small metastasis in the lung, whereas all of

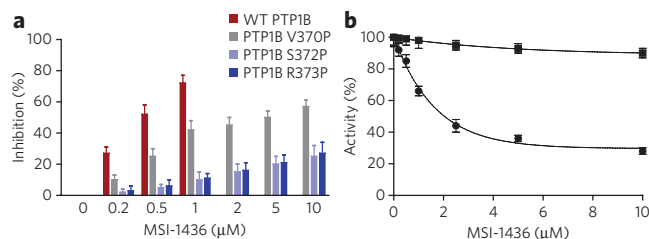


Figure 5 | Effect of mutations in the C terminus of PTP1B on inhibition by MSI-1436. (a) The bar graph compares the extent of inhibition of PTP1B₁₋₄₀₅ and three point mutants located in the NMR-detected helix $\alpha 9'$ (residues 360-379 in the C terminus). The inhibition constants (k_i) were 0.8 μM for PTP1B₁₋₄₀₅ and 10 μM , 15 μM and 18 μM for the V370P, S372P and R373P mutants, respectively. Data represent mean \pm s.e.m. from three independent experiments. WT, wild type. (b) Introduction of two point mutations in PTP1B₁₋₄₀₅, L192A and S372P, resulted in complete loss of inhibition by MSI-1436. Data represent mean \pm s.e.m. from three independent experiments.

the animals in the untreated control group presented with a high occurrence of lung metastasis (Fig. 6c). These data highlight the potential of MSI-1436 as an effective approach to counter HER2-induced mammary tumorigenesis and malignancy.

MSI-1436 specifically targeted PTP1B in breast cancer models

We used a variety of approaches to demonstrate that inhibition of PTP1B was the foundation for the mechanism by which MSI-1436 antagonized HER2 function. First, we tested its effects on tyrosine

phosphorylation-dependent signaling. Using the NDL2 mice, we observed a pronounced increase in tyrosine phosphorylation of p62^{DOK} in tumor samples from animals treated with MSI-1436. Concomitantly, we also observed a decrease in phosphorylation of p44/42 MAP kinase upon MSI-1436 treatment (Supplementary Fig. 13). These signaling changes are consistent with PTP1B being the target of MSI-1436 (ref. 21).

To demonstrate a direct interaction with PTP1B, we developed a binding assay in which MSI-1436 was linked to NHS-activated

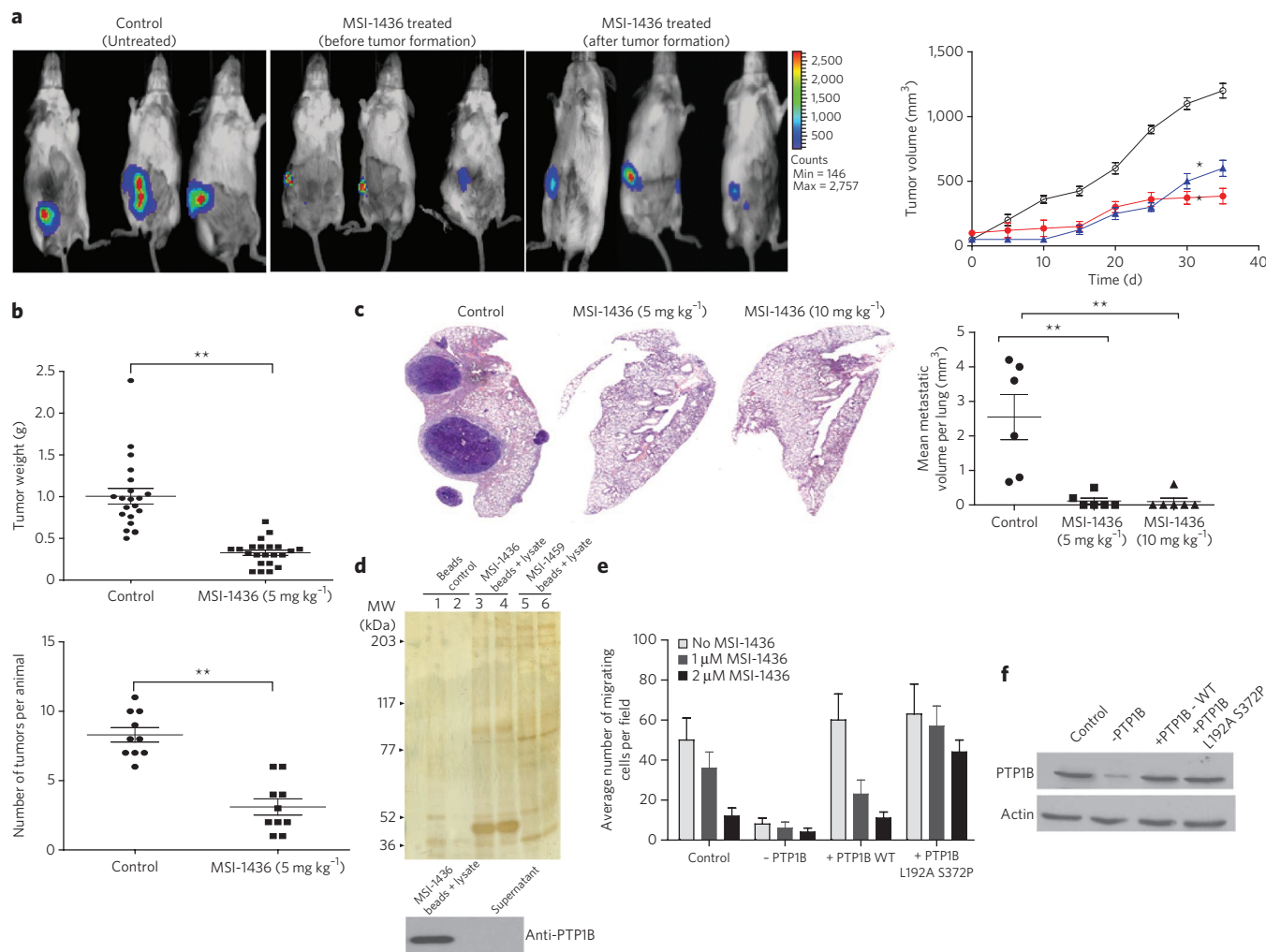


Figure 6 | Effect of the allosteric inhibitor of PTP1B, MSI-1436, in cell and animal models of HER2-positive breast cancer. (a) Effect of MSI-1436 on tumor xenografts. BT474 cells (5×10^6) were injected orthotopically into the mammary fat pad of SCID-beige mice. MSI-1436 (5 mg kg^{-1}) or saline (control) was injected intraperitoneally as soon as the surgery was performed or once the tumors reached $\sim 100 \text{ mm}^3$ in volume. Tumor volume was assessed with a digital caliper for control (open circles) and treated with MSI-1436. Quantification of tumor volume, measured using a digital caliper, over time (right). Student's *t*-test, $^{**}P < 0.01$. Representative images of three independent experiments are shown. **(b)** Distribution of tumor weight (top) and tumor number (bottom) in NDL2 animals treated with MSI-1436 (5 mg kg^{-1} ; filled squares) or vehicle control (filled circles). Student's *t*-test, $^{**}P < 0.001$. Data are representative of three independent experiments with more than six animals in each experiment. **(c)** Effect of MSI-1436 on lung metastases, monitored by histopathological analysis of hematoxylin and eosin-stained lung sections taken from the animals. Mice were treated with saline or MSI-1436 at 5 mg kg^{-1} or 10 mg kg^{-1} . Each treatment group contained six animals. Left, representative images of lung sections from control and MSI-1436-treated animals subjected to hematoxylin and eosin staining. Right, quantification of the whole lung obtained from all of the animals in each group (control and MSI-1436-treated). Student's *t*-test, $^{**}P < 0.001$. Data are representative of three independent experiments with more than six animals in each experiment. **(d)** Tumor lysate was incubated with MSI-1436 immobilized to NHS-activated Sepharose beads for 30 min at 4°C . The beads were washed, and the samples were subjected to SDS-PAGE and silver staining (top). Both beads and supernatant were subjected to immunoblotting using anti-PTP1B (FG6) (bottom) to demonstrate that PTP1B was cleared from the lysate by immobilized MSI-1436. Data are representative of three independent experiments. MW, molecular weight. **(e)** Migration of control and PTP1B-depleted MCF10-NeuNT cells, in the absence and presence of doxycycline, was compared to migration of MCF10-NeuNT cells overexpressing wild-type (WT) PTP1B or the double mutant PTP1B^{L192A S372P}, in the presence of doxycycline and in the absence (pale gray bars) and presence of MSI-1436 (gray bars, $1 \mu\text{M}$; dark gray bars, $2 \mu\text{M}$). Data are representative of three independent experiments done in duplicates. **(f)** PTP1B knockdown and re-expression of wild-type (WT) and L192A S372P mutant PTP1B in MCF10A-Neu-NT cells were analyzed by immunoblotting with anti-PTP1B (FG6). Data are representative of three independent experiments. The full blot image is in **Supplementary Figure 20**. In **a–c** and **e**, error bars represent s.e.m., $n = 3$.

magnetic beads via its free amino group on the spermine tail (**Supplementary Fig. 14**). The structurally related inactive analog MSI-1459 (β -spermine, 5α , cholenic acid methyl ester), which has a free amino group like MSI-1436 but does not inhibit PTP1B (**Supplementary Fig. 15**), was included as a negative control. Bead-immobilized MSI-1436 bound recombinant PTP1B_{1–405} and PTP1B_{1–321} *in vitro*, whereas MSI-1459 did not (**Supplementary Fig. 16a**). Furthermore, upon mixing bead-immobilized MSI-1436 with tumor lysates from NDL2 mice, a major band at 50 kDa, confirmed as PTP1B by immunoblotting, was retained by the MSI-1436 beads but was not bound by MSI-1459 (**Fig. 6d**). Similar results were obtained with lysates of MCF10A cells (**Supplementary Fig. 16b**).

To establish that inhibition of PTP1B was the critical step underlying the effects of MSI-1436, we turned to the PTP1B^{L192A S372P} double mutant, which preserved catalytic activity but was insensitive to MSI-1436 (**Fig. 5b**). We used a transformed breast epithelial cell line, MCF10A-NeuNT, which expresses a PTP1B-directed shRNA from a doxycycline-inducible lentiviral vector. In the absence of doxycycline, migration of these cells was attenuated by MSI-1436. In contrast, doxycycline treatment essentially ablated PTP1B expression, which inhibited migration (**Fig. 6e**). We reconstituted expression of either the wild type or the L192A S372P double mutant forms of PTP1B to a similar level as that of endogenous PTP1B in the control cells (**Fig. 6f**). Reconstitution with wild-type PTP1B rescued cell migration, which remained sensitive to inhibition by MSI-1436; however, although reconstitution with the PTP1B^{L192A S372P} double mutant rescued cell migration, this process was no longer sensitive to MSI-1436 (**Fig. 6e**). Similarly, we tested the effects of MSI-146 on the growth of orthotopic BT474 xenografts in which PTP1B had been suppressed by RNAi then rescued by reexpressing wild-type or MSI-1436-resistant forms of PTP1B. In two independent experiments, we observed that tumor xenografts from BT474 cells reexpressing wild-type PTP1B were more sensitive to MSI-1436 than xenografts from BT474 cells reexpressing the mutant protein (**Supplementary Fig. 17**). These data illustrate that PTP1B is the primary target of MSI-1436 in these effects on HER2-induced cell migration.

DISCUSSION

Through its ability to antagonize insulin and leptin signaling, PTP1B is a highly validated metabolic regulator and therapeutic target for diabetes and obesity. Although there is also evidence to support a role in cancer, this function is less well defined, with reports suggesting both tumor-suppressive and oncogenic functions²². Although there are reports linking PTP1B to the downregulation of several growth-promoting signaling pathways, such as inhibitory effects on EGFR signaling²³, PTP1B-deficient mice do not display increased incidence of tumors⁷. Furthermore, although PTP1B has been implicated in the regulation of cadherin-dependent cell adhesion and apoptosis, the literature contains contradictions that prevent the precise definition of a tumor suppressor role for the phosphatase. In fact, the *PTPN1* gene encoding PTP1B is located on chromosome 20q13.1–13.2 (ref. 24), which is amplified in several cancers, and PTP1B is highly expressed in several tumor types²⁵. For example, HER2 amplification has been associated with overexpression of PTP1B^{7,26}. Furthermore, PTP1B has a positive role in promoting the tumor phenotype in gastric cancer, where its expression is correlated with metastasis^{27,28} as well as in prostate²⁹ and colon³⁰ cancer cells. However, this oncogenic function of PTP1B also remains to be fully defined. Although PTP1B has been reported to have a positive role in breast cancer cell invasion^{7,31}, it has also been suggested that the phosphatase is not important for breast tumor maintenance³². Nevertheless, through using a PTP1B-directed allosteric inhibitor, we now provide compelling evidence for the validation of PTP1B as a therapeutic target in HER2-positive breast cancer.

Given the importance of the unmet medical needs that could be addressed by targeting PTP1B, many pharmaceutical companies and academic laboratories took up the challenge of generating inhibitors of the phosphatase. The current focus has been on producing active site-directed inhibitors, but this has encountered challenges owing to the charged nature of the active site and the susceptibility of PTPs to irreversible inactivation due to covalent modification of an essential, catalytic cysteine. The most potent, selective inhibitors are based upon nonhydrolyzable analogs of a phosphorylated tyrosine substrate, such as fluorophosphonates³³; however, these inhibitors are highly charged, with limited membrane permeability. As industry has set a hurdle of oral bioavailability for the next generation of treatments of diabetes and obesity, these substrate analogs have limited drug development potential. A second problem, and one that also impeded the early stages of development of kinase inhibitors, is the issue of inhibitor selectivity. In the case of kinase inhibitors, most now not only target the active site but also the ATP binding pocket within the active site, a common feature of all kinases. Furthermore, the ability of an inhibitor to act on more than one kinase has often proven to be advantageous³⁴. This concern about inhibitor selectivity is fueled by the structural similarities among the catalytic domains of the members of the PTP family. In particular, TCPTP, which is the closest relative of PTP1B, sharing 75% sequence identity in the catalytic domain, regulates inflammation, and ablation of the gene encoding TCPTP in mice is lethal³⁵. Although there are data to suggest that some inhibition of TCPTP may be of benefit to promoting insulin signaling^{36,37}, drug discovery programs have always sought inhibitors that distinguish between these two enzymes. Our approach addresses these concerns. We devised methods to overcome technical problems associated with using recombinant forms of PTP1B that contain the C-terminal segment, including susceptibility to proteolysis. By targeting the unique, noncatalytic segment that is unrelated to TCPTP, or any other member of the PTP family, such allosteric inhibitors would have the potential to be highly specific for PTP1B. Also, the inhibitors we identify in this way will have been missed in screens conducted to date in industry, which have used primarily the truncated form of the enzyme comprising only the catalytic domain.

In this study, we identified two binding sites for MSI-1436 in the long form of PTP1B, a novel site in the C-terminal segment (helix $\alpha 9'$) and an extended site incorporating helix $\alpha 7$ and residues 299, 310 and 311. Our data suggest a model in which one molecule of the cholestane MSI-1436 could be bound between $\alpha 7$ and the $\alpha 9'$ helices, in a manner analogous to the cholesterol-binding motif that has been identified in the crystal structure of the β -adrenergic receptor³⁸. We noted a similar spatial arrangement and identity of residues in the $\alpha 7$ and $\alpha 9'$ helices in the C-terminal segment of PTP1B to those engaged by cholesterol in two critical α -helices in the β -adrenergic receptor (**Supplementary Fig. 18**). PTP1B helix $\alpha 7$, together with helix $\alpha 3$ and $\alpha 6$, also forms a previously identified allosteric inhibition site in PTP1B. Benzofuran compounds have been shown to bind to this site and change the ability of helix $\alpha 3$ and the WPD loop to adopt the catalytically competent conformation of PTP1B¹². This suggests a mechanism by which the two binding sites for MSI-1436 may communicate to yield the positive cooperativity we observed. Thus, a first molecule of MSI-1436 sandwiched between $\alpha 7$ and the C-terminal $\alpha 9'$ helix may bring the $\alpha 7$ helix close to $\alpha 3$ and $\alpha 6$, forming the second site where another molecule of MSI-1436 would bind. Molecular dynamic simulations revealed that the best conformation with least binding energy between MSI-1436 and PTP1B was consistent with this proposed mechanism of inhibition (**Supplementary Table 4 and Supplementary Fig. 19**).

The approach of generating allosteric inhibitors has also been applied to several protein kinases with the goal of producing cancer therapeutics. For example, GNF-5 engages a myristate-binding pocket in the ABL PTK, which alters the ATP binding site and may

represent an approach to overcome resistance to ATP-competitive inhibitors of BCR-ABL³⁹. However, another unique feature of our approach is that we targeted a segment of PTP1B that is intrinsically disordered rather than addressing defined binding pockets. Intrinsically disordered segments of proteins do not have stable secondary or tertiary structure; instead, as we have noted for the C-terminal segment of PTP1B, there is an ensemble of rapidly interconverting structures. It is interesting to note that disordered proteins are particularly prevalent in signaling and disease, where the potential for specificity in the interactions is stimulating efforts to exploit them for therapeutic development¹³. Attempts have been made to target Myc-Max interactions by this approach⁴⁰, but our definition of the effects of MSI-1436 on PTP1B is to our knowledge one of the first examples of a therapeutic candidate that works by this mechanism.

In conclusion, inhibition of PTP1B may offer a new approach to attenuating HER2-dependent signaling in a clinical setting. As has been reported for PTK inhibitors, such as lapatinib^{41,42} and gefitinib⁴³, the effects of MSI-1436 appear to be primarily cytostatic rather than cytotoxic. This suggests that inhibition of PTP1B may perhaps be most effective in combinatorial therapies, together with other signal transduction-based drugs and cytotoxic agents. Trastuzumab is the treatment of choice for HER2-dependent cancer, although there are problems with *de novo* and acquired resistance. It will be of interest to test whether inclusion of MSI-1436 together with trastuzumab, to target simultaneously the PTK and its downstream signaling, will prolong the time needed for resistance to develop or even overcome resistant states. In addition, unlike PTK inhibitors, which have major side effects, MSI-1436 is well tolerated in patients¹¹. Thus, we are poised for a clinical trial to test this new therapeutic strategy in HER2-positive cancer.

Received 11 December 2013; accepted 16 April 2014;
published online 20 May 2014

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Accession codes. BioMagResBank: All of the chemical shift data for PTP1B_{1–393} were deposited under accession number [19224](#).

References

- Slamon, D.J. *et al.* Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* **244**, 707–712 (1989).
- Tiwari, R.K., Borgen, P.I., Wong, G.Y., Cordon-Cardo, C. & Osborne, M.P. HER-2/neu amplification and overexpression in primary human breast cancer is associated with early metastasis. *Anticancer Res.* **12**, 419–425 (1992).
- Engelman, J.A. & Settleman, J. Acquired resistance to tyrosine kinase inhibitors during cancer therapy. *Curr. Opin. Genet. Dev.* **18**, 73–79 (2008).
- Rexer, B.N. & Arteaga, C.L. Intrinsic and acquired resistance to HER2-targeted therapies in HER2 gene-amplified breast cancer: mechanisms and clinical implications. *Crit. Rev. Oncog.* **17**, 1–16 (2012).
- Tonks, N.K. Protein tyrosine phosphatases—from housekeeping enzymes to master regulators of signal transduction. *FEBS J.* **280**, 346–378 (2013).
- Andersen, J.N. & Tonks, N.K. Protein tyrosine phosphatase-based therapeutics: lessons from PTP1B. *Top. Curr. Genet.* **5**, 201–230 (2004).
- Julien, S.G. *et al.* Protein tyrosine phosphatase 1B deficiency or inhibition delays ErbB2-induced mammary tumorigenesis and protects from lung metastasis. *Nat. Genet.* **39**, 338–346 (2007).
- Bentires-Alj, M. & Neel, B.G. Protein-tyrosine phosphatase 1B is required for HER2/Neu-induced breast cancer. *Cancer Res.* **67**, 2420–2424 (2007).
- Tonks, N.K., Diltz, C.D. & Fischer, E.H. Purification of the major protein-tyrosine-phosphatases of human placenta. *J. Biol. Chem.* **263**, 6722–6730 (1988).
- Hao, L., Tiganis, T., Tonks, N.K. & Charbonneau, H. The noncatalytic C-terminal segment of the T cell protein tyrosine phosphatase regulates activity via an intramolecular mechanism. *J. Biol. Chem.* **272**, 29322–29329 (1997).
- Lantz, K.A. *et al.* Inhibition of PTP1B by trodusquemine (MSI-1436) causes fat-specific weight loss in diet-induced obese mice. *Obesity (Silver Spring)* **18**, 1516–1523 (2010).
- Wiesmann, C. *et al.* Allosteric inhibition of protein tyrosine phosphatase 1B. *Nat. Struct. Mol. Biol.* **11**, 730–737 (2004).
- Metallo, S.J. Intrinsically disordered proteins are potential drug targets. *Curr. Opin. Chem. Biol.* **14**, 481–488 (2010).
- Cuchillo, R. & Michel, J. Mechanisms of small-molecule binding to intrinsically disordered proteins. *Biochem. Soc. Trans.* **40**, 1004–1008 (2012).
- Pervushin, K., Riek, R., Wider, G. & Wuthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. USA* **94**, 12366–12371 (1997).
- Zhang, H., Neal, S. & Wishart, D.S. RefDB: a database of uniformly referenced protein chemical shifts. *J. Biomol. NMR* **25**, 173–195 (2003).
- Barford, D., Flint, A.J. & Tonks, N.K. Crystal structure of human protein tyrosine phosphatase 1B. *Science* **263**, 1397–1404 (1994).
- Muthuswamy, S.K., Li, D., Lelievre, S., Bissell, M.J. & Brugge, J.S. ErbB2, but not ErbB1, reinitiates proliferation and induces luminal repopulation in epithelial acini. *Nat. Cell Biol.* **3**, 785–792 (2001).
- Konecny, G.E. *et al.* Activity of the dual kinase inhibitor lapatinib (GW572016) against HER-2-overexpressing and trastuzumab-treated breast cancer cells. *Cancer Res.* **66**, 1630–1639 (2006).
- Siegel, P.M., Dankort, D.L., Hardy, W.R. & Muller, W.J. Novel activating mutations in the neu proto-oncogene involved in induction of mammary tumors. *Mol. Cell. Biol.* **14**, 7068–7077 (1994).
- Dubé, N., Cheng, A. & Tremblay, M.L. The role of protein tyrosine phosphatase 1B in Ras signaling. *Proc. Natl. Acad. Sci. USA* **101**, 1834–1839 (2004).
- Julien, S.G., Dube, N., Hardy, S. & Tremblay, M.L. Inside the human cancer tyrosine phosphatase. *Nat. Rev. Cancer* **11**, 35–49 (2011).
- Sangwan, V. *et al.* Protein-tyrosine phosphatase 1B modulates early endosome fusion and trafficking of Met and epidermal growth factor receptors. *J. Biol. Chem.* **286**, 45000–45013 (2011).
- Brown-Shimer, S. *et al.* Molecular cloning and chromosome mapping of the human gene encoding protein phosphotyrosyl phosphatase 1B. *Proc. Natl. Acad. Sci. USA* **87**, 5148–5152 (1990).
- Tonks, N.K. & Muthuswamy, S.K. A brake becomes an accelerator: PTP1B—a new therapeutic target for breast cancer. *Cancer Cell* **11**, 214–216 (2007).
- Wiener, J.R. *et al.* Overexpression of the protein tyrosine phosphatase PTP1B in human breast cancer: association with p185c-erbB-2 protein expression. *J. Natl. Cancer Inst.* **86**, 372–378 (1994).
- Wang, J., Chen, X., Liu, B. & Zhu, Z. Suppression of PTP1B in gastric cancer cells *in vitro* induces a change in the genome-wide expression profile and inhibits gastric cancer cell growth. *Cell Biol. Int.* **34**, 747–753 (2010).
- Wang, J. *et al.* PTP1B expression contributes to gastric cancer progression. *Med. Oncol.* **29**, 948–956 (2012).
- Lessard, L. *et al.* PTP1B is an androgen receptor-regulated phosphatase that promotes the progression of prostate cancer. *Cancer Res.* **72**, 1529–1537 (2012).
- Zhu, S., Bjorge, J.D. & Fujita, D.J. PTP1B contributes to the oncogenic properties of colon cancer cells through Src activation. *Cancer Res.* **67**, 10129–10137 (2007).
- Johnson, K.J. *et al.* PTP1B suppresses prolactin activation of Stat5 in breast cancer cells. *Am. J. Pathol.* **177**, 2971–2983 (2010).
- Balavenkatraman, K.K. *et al.* Epithelial protein-tyrosine phosphatase 1B contributes to the induction of mammary tumors by HER2/Neu but is not essential for tumor maintenance. *Mol. Cancer Res.* **9**, 1377–1384 (2011).
- He, R., Zeng, L.F., He, Y., Zhang, S. & Zhang, Z.Y. Small molecule tools for functional interrogation of protein tyrosine phosphatases. *FEBS J.* **280**, 731–750 (2013).
- Kane, R.C. *et al.* Sorafenib for the treatment of advanced renal cell carcinoma. *Clin. Cancer Res.* **12**, 7271–7278 (2006).
- Wiede, F. *et al.* T cell protein tyrosine phosphatase attenuates T cell signaling to maintain tolerance in mice. *J. Clin. Invest.* **121**, 4758–4774 (2011).
- Tiganis, T. PTP1B and TCPTP—nonredundant phosphatases in insulin signaling and glucose homeostasis. *FEBS J.* **280**, 445–458 (2013).
- Galic, S. *et al.* Coordinated regulation of insulin signaling by the protein tyrosine phosphatases PTP1B and TCPTP. *Mol. Cell. Biol.* **25**, 819–829 (2005).
- Hanson, M.A. *et al.* A specific cholesterol binding site is established by the 2.8 Å structure of the human β₂-adrenergic receptor. *Structure* **16**, 897–905 (2008).
- Zhang, J. *et al.* Targeting Bcr-Abl by combining allosteric with ATP-binding-site inhibitors. *Nature* **463**, 501–506 (2010).
- Hammoudeh, D.I., Follis, A.V., Prochownik, E.V. & Metallo, S.J. Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc. *J. Am. Chem. Soc.* **131**, 7390–7401 (2009).

41. Agulnik, M. *et al.* Phase II study of lapatinib in recurrent or metastatic epidermal growth factor receptor and/or erbB2 expressing adenoid cystic carcinoma and non adenoid cystic carcinoma malignant tumors of the salivary glands. *J. Clin. Oncol.* **25**, 3978–3984 (2007).
42. Hinow, P., Wang, S.E., Arteaga, C.L. & Webb, G.F. Relocating job wise? A mathematical model separates quantitatively the cytostatic and cytotoxic effects of a HER2 tyrosine kinase inhibitor. *Theor. Biol. Med. Model.* **4**, 14 (2007).
43. Jänne, P.A., Taffaro, M.L., Salgia, R. & Johnson, B.E. Inhibition of epidermal growth factor receptor signaling in malignant pleural mesothelioma. *Cancer Res.* **62**, 5242–5247 (2002).

Acknowledgments

This research was supported by US National Institutes of Health (NIH) grants CA53840 and GM55989 and the Cold Spring Harbor Laboratory Cancer Centre Support Grant CA45508 to N.K.T.; by NIH grants GM100910 and GM098482, American Diabetes Association grant 1-14-ACN-31 and a Brown University Research Seed Fund grant by the Vice President for Research to W.P.; and by financial support from the French Agence Nationale de la Recherche (ANR) through ANR JCJC ProteinDisorder to M.R.J. and ANR MALZ TAUSTRICT to M.B. J.K. is a fellow of the IDPbyNMR Marie Curie action of the European Commission (contract no 264257).

The 800-MHz NMR data were recorded at Brandeis University; the instrument was purchased with support from NIH S10-RR017269. NMR data (500 MHz and 850 MHz) were recorded in the Brown University Structural Biology Facility, which is

generously supported by Brown University. N.K.T. is also grateful for support from the following foundations; The Gladowsky Breast Cancer Foundation, The Don Monti Memorial Research Foundation, Hansen Memorial Foundation, West Islip Breast Cancer Coalition for Long Island, Glen Cove CARES, Find a Cure Today (FACT), Constance Silveri, Robertson Research Fund and the Masthead Cove Yacht Club Carol Marcincuk Fund. PTP1B inhibitor MSI-1436 was provided by Ohr Pharmaceuticals and Genaera Corporation.

Author contributions

N.K. performed the biochemical characterization of PTP1B inhibition by MSI-1436 and tested the effects of the inhibitor in cell and animal models; D.K., D.H.M., J.K., M.R.J., R.P., M.B. and W.P. designed, performed and analyzed NMR-based structural work. C.M.G. performed the homology modeling, molecular docking and dynamics simulation. B.X., S.D.A. and S.K.M. helped with the breast cancer studies. N.K., W.P. and N.K.T. analyzed the data and wrote the manuscript, which was reviewed by all authors. N.K.T. directed the study.

Competing financial interests

The authors declare competing financial interests: details accompany the [online version of the paper](#).

Additional information

Supplementary information and chemical probe information is available in the [online version of the paper](#). Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Correspondence and requests for materials should be addressed to N.K.T.

ONLINE METHODS

Reagents. All of the chemicals and reagents were purchased from Sigma-Aldrich unless otherwise specified. Global tyrosyl phosphorylation was detected by anti-phosphotyrosine 4G10 (no. 05-321, Millipore). Anti-p62DOK (no. 3912), anti-phospho-Erk (no. 9101) and total ERK (no. 9102) antibodies were obtained from Cell Signaling Technology. For loading controls for corresponding phosphotyrosine proteins, the membrane was stripped by Restore Stripping Buffer (no. 21059, Thermo Scientific) and reprobed with antibodies against the individual protein. For the FRET reporter, ECFP and EYFP cDNA was PCR amplified and cloned on either side of PTP1B in the pET28b vector. A spacer containing residues GSGSG was used between the protein and the fluorophores on each side to provide a flexible connection and also to allow rotational mobility as the individual domains of the protein fold separately. CFP-PTP1B-YFP was expressed under IPTG induction in BL21 cells (*Escherichia coli*) and purified with Ni-NTA matrix exploiting the C-terminal His tag.

Enzyme kinetics. PTP assays were performed in black polystyrol 96-well plates using DiFMUP as substrate. DiFMUP (10 μ M) was added to assay buffer (50 mM Hepes, 100 mM NaCl, 0.1% BSA, 2 mM DTT, 2 mM EDTA, pH 6.5) containing 10 nM purified PTP1B in a final volume of 100 μ l. The fluorescence emitted at 450 nm was monitored continuously for 20 min using a Gemini XPS fluorescence plate reader. For assays using radiolabeled substrate, reduced carboxamidomethylated and maleylated lysozyme (RCML) was phosphorylated on tyrosine to a stoichiometry up to 0.8 mol 32 P per mol of protein by recombinant GST-FER kinase and [γ - 32 P]ATP, and activity was measured as described in ref. 44.

Binding assay with [3 H]MSI-1436. Direct binding assays were performed using tritiated MSI-1436. Histidine-tagged PTP1B (100 nM) was incubated with varying concentrations of [3 H]MSI-1436 for 30 min in assay buffer (50 mM Hepes, 100 mM NaCl, 0.1% BSA, 2 mM DTT, 2 mM EDTA, pH 6.5) at 25 $^{\circ}$ C. Protein-bound and free [3 H]MSI-1436 were separated by incubating the protein-inhibitor mixture with 50 μ l of 50% Ni-NTA beads for 10 min at 25 $^{\circ}$ C. The beads were washed with assay buffer containing 150 mM NaCl three times, and the inhibitor bound to protein was determined by scintillation counting. The amount of bead-bound protein was estimated by BCA protein quantification. For Scatchard analysis, the number of moles of [3 H]MSI-1436 bound to PTP1B was obtained directly from scintillation counting, and the number of moles of free (F) [3 H]MSI-1436 was calculated by subtracting the bound from the total number of moles of [3 H]MSI-1436 used in each assay. The ratio of B/F was plotted against B. The extent of positive cooperativity was estimated using a Hill plot, where Y represents the fraction of binding sites that are occupied on the protein and 1 - Y represents the fraction of binding sites that are not occupied. Using the equation $(Y/(1 - Y)) = [\text{ligand}]^n/K_d$, the K_d and the Hill coefficient (n) were obtained. For displacement titrations, PTP1B (100 nM) was saturated with [3 H]MSI-1436, and the unbound ligand was separated as mentioned above. The enzyme saturated with [3 H]MSI-1436 was titrated against sulfanamido-benzbromarone (0–10 μ M). The residual MSI-1436 bound to the enzyme was measured by scintillation counting.

Isothermal titration calorimetry. The studies were performed using a VP-ITC isothermal titration calorimetry system (GE Healthcare). The titrations were performed at 25 $^{\circ}$ C in degassed buffer (50 mM HEPES pH 7.0, 100 mM NaCl, 1 mM DTT and 2% DMSO). Protein concentration in the calorimeter cell was 50–100 μ M. All of the protein samples used in the titration were dialyzed completely against the buffer. Inhibitor (10 mM) in DMSO was diluted to make a stock solution of 500 μ M in degassed buffer. The instrument was calibrated by using the heat of dilution of NaCl in water. By titrating buffer into protein solution, the heat generated due to protein dilution was estimated and was found to be negligible. The heat of ligand dilution was corrected by subtracting the average heat of injection after saturation. Origin software was used to analyze and fit the data.

Gel filtration. PTP1B (5 μ M) was incubated with 10 μ M MSI-1436 in a final volume of 200 μ l buffer (50 mM HEPES, pH 7.0, 100 mM NaCl, 0.2 mM EDTA, 2 mM DTT) at 4 $^{\circ}$ C. The reaction mixture was then subjected to gel filtration on a Superdex 200 column (HR30/10; Pharmacia Biotech Inc.). Prior to loading the MSI-1436-saturated protein sample, the column was equilibrated with 50 mM HEPES, pH 7.0, 100 mM NaCl, 0.2 mM EDTA and 2 mM DTT containing 100 μ M MSI-1436 to ensure that the protein was bound to the inhibitor

throughout the run. The void volume (V_o) and total volume (V_t) were measured using blue dextran and Coomassie blue dye, respectively. Fractions of 0.5 ml were collected, and the protein concentration in each sample was determined by Bradford assay or by recording the UV absorption spectrum at 280 nm. Alternatively, the fractions were precipitated with 10% trichloroacetic acid and subjected to SDS-PAGE followed by staining of the gel with Coomassie blue. Relative retention (R_t) was determined from the absolute retention volumes (V_{abs}) according to the equation $R_t = (V_{\text{abs}} - V_o)/(V_t - V_o)$. The column was calibrated with ferritin, catalase, cytochrome C and BSA as molecular weight standards.

Limited proteolysis. PTP1B (2 μ M) was incubated with MSI-1436 (5 μ M) for 10 min at 4 $^{\circ}$ C before adding varying amounts of trypsin for 30 min at room temperature. The trypsin/PTP1B (mol/mol) ratios were 1:10, 0.5:10, 0.25:10, 0.125:10, 0.06:10, 0.03:10, 0.01:10, 0.005:10 and 0:10. Reactions were terminated after 5 min by adding phenylmethylsulfonyl fluoride (PMSF; 2 mM final concentration) and hot SDS sample buffer at a 3:2 (v/v) ratio of SDS sample buffer/sample and analyzed by SDS-PAGE using 10% polyacrylamide separating gels.

FRET. Fluorescence spectra were recorded from 400 nm to 600 nm with excitation at 435 nm using a Gemini XPS fluorescence microplate reader. Excitation and emission bandwidths were set to be 0.5 nm. The emission of 10 nM CFP-PTP1B-YFP in 50 mM HEPES, pH 7.0, 100 mM NaCl, 2 mM DTT and 0.1% BSA was measured in the absence of MSI-1436, for which a very low FRET signal was observed. MSI-1436 was added to CFP-PTP1B-YFP, and FRET was monitored as the concentration of MSI-1436 was increased from 1 nM to 10 μ M. The Forster radius (R_o), FRET efficiency (E) and distance (R) were calculated using the software PTI Felix32 Analysis, version 1.2. Errors for E and R were estimated to be ~5%.

Protein expression for NMR studies. A human PTP1B catalytic domain construct (residues 1–301; PTP1B_{1–301}) and PTP1B_{1–393} (residues 1–393) were subcloned into pRP1B. The plasmid was transformed into *E. coli* BL21 (DE3) RIL cells (Agilent), and protein expression was induced using 1 mM IPTG at 18 $^{\circ}$ C for 18h. To prepare uniformly 2 H, 13 C, 15 N-labeled and 2 H, 15 N-labeled protein, cultures were grown in M9 minimal medium containing selective antibiotics and [13 C]D-d₇-glucose (4 g/l) and/or 15 NH₄Cl (1 g/l) in 99% D₂O (Cambridge Isotope Laboratories or Isotec). A single round of 50% D₂O adaptation was necessary to increase the yield of PTP1B. To incorporate single 15 N-isotopically labeled amino acids into PTP1B, cultures were grown in minimal medium, replacing an unlabeled amino acid with the corresponding 15 N-labeled amino acid where appropriate. PTP1B_{1–301} was labeled with [15 N]l-valine, [15 N]l-tyrosine, [15 N]l-phenylalanine or [15 N]l-leucine, whereas PTP1B_{1–393} was labeled with [15 N]l-valine or [15 N]l-phenylalanine. Cells were harvested by centrifugation and stored at -80 $^{\circ}$ C.

Protein purification. Cell pellets were resuspended in ice-cold lysis buffer (50 mM Tris, pH 8.0, 500 mM NaCl, 5 mM imidazole, 0.1% Triton X-100 and EDTA-free protease inhibitor tablets (Roche)) and lysed by high-pressure cell homogenization (Avestin C3 Emulsiflex). The bacterial lysate was clarified by centrifugation at 45,000g for 60 min at 4 $^{\circ}$ C. After filtration, the supernatant was loaded onto a HisTrap HP column (GE Healthcare) equilibrated with 50 mM Tris, pH 7.5, 5 mM imidazole and 500 mM NaCl, and the His₆-tagged protein was eluted using an imidazole gradient of 5–500 mM. Fractions containing PTP1B were pooled and cleaved with tobacco etch virus (TEV) protease overnight at 4 $^{\circ}$ C while being dialyzed against 50 mM Tris, pH 8.0, 500 mM NaCl. Cleaved protein was further purified using Ni²⁺-NTA immobilized metal affinity chromatography followed by size exclusion chromatography (SEC; Superdex 75 26/60; GE Healthcare), equilibrated in NMR buffer (50 mM HEPES, pH 6.8, 150 mM NaCl, 0.5 mM TCEP) to a purity of >98% and had a final yield of 40 mg PTP1B_{1–301} or 20 mg PTP1B_{1–393} per liter of LB cell culture. Purified PTP1B_{1–301} and PTP1B_{1–393} were concentrated to 1 mM and 0.3 mM, respectively. For long-term storage, the protein was flash frozen in liquid nitrogen and stored at -80 $^{\circ}$ C.

NMR spectroscopy. NMR data of PTP1B_{cat} and PTP1B_{1–393} were collected at 298 K on Bruker Avance 500 MHz and 800 MHz spectrometers and a Bruker Avance III HD 850 MHz spectrometer, each equipped with a TCI HCN z-gradient cryoprobe. All of the NMR spectra were processed with Topspin

2.1/3.0/3.1 (Bruker) and analyzed using CARA (<http://cara.nmr.ch>). TROSY versions of a three-dimensional HNCA, three-dimensional HN(CO)CA, three-dimensional HNCACB and three-dimensional HN(CO)CACB were recorded on 1 mM [²H,¹⁵N,¹³C]PTP1B_{cat} and 0.3 mM [²H,¹⁵N,¹³C]PTP1B₁₋₃₉₃ samples. A TROSY version of a three-dimensional ¹⁵N-resolved [¹H,¹H] NOESY ($T_m = 120$ ms) was also recorded on 1 mM [²H,¹⁵N]PTP1B_{cat} to verify assignments using sequential backbone ¹H^N-¹H^N NOE values. Additionally, two-dimensional [¹H,¹⁵N] TROSY spectra of single ¹⁵N-isotopically labeled amino acid samples (¹⁵N-Leu, ¹⁵N-Tyr, ¹⁵N-Phe or ¹⁵N-Val) were recorded.

Small-angle X-ray scattering. All data were collected on a Rigaku BioSAXS-1000 (Anode: FRE⁺-SuperBright) equipped with a Dectris Pilatus 100K detector at 10 °C using 1-h experiments. Immediately before data collection, proteins were verified to be monomeric and were exchanged into NMR buffer using SEC. Buffer scattering curves were subtracted from sample scattering curves using the Rigaku SAXSLab Software. PTP1B₁₋₃₀₁ was recorded at 2.0 mg/ml and 3.5 mg/ml concentrations, whereas PTP1B₁₋₃₉₃ was recorded at 3.6 mg/ml and 5.0 mg/ml. Guinier approximation, $I(q) = I(0)\exp(-q^2R_g^2/3)$, was performed on each independent scattering experiment and averaged to determine the radius of gyration. The linearity of the Guinier region and the forward scattering intensity, $I(0)$, which is proportional to the molecular weight of the sample, were used to verify that samples were monodisperse in solution.

Dynamic light scattering. Dynamic light scattering (DLS) data for PTP1B₁₋₃₀₁ and PTP1B₁₋₃₉₃ were acquired using a Viscotek 802 DLS instrument. All of the proteins were measured in triplicate at a concentration of 20 μM in 50 mM HEPES, pH 6.8, 150 mM NaCl, 0.5 mM TCEP. All of the data were analyzed using the OmniSIZE software, and each experiment reflects the average of ten measurements, each of which was acquired for 10 s at 25 °C.

Cell migration and invasion assays. 10A.B2 cells expressing an inducible HER2 chimera⁴⁵ were used to quantify cell invasion with BD BioCoat Matrigel invasion chambers with 8.0-μm pore size. 10A.B2 cells (1×10^6) were grown in the insert. After 24–48 h, the cells retained inside the insert were removed, and those that migrated to the other side of the insert were fixed and stained with KARYOMAX Giemsa Stain (GIBCO, Invitrogen) and counted. Suppression of PTP1B by RNAi in MCF-10AN cells was achieved using a retroviral expression vector (pMLP) that uses a murine stem cell virus (MSCV) backbone⁴⁵. The pMLP vector containing the validated PTP1B shRNA (5'-CTTTGACCATAGTCGGATT-3') and nontarget control shRNA targeting firefly luciferase were used to create stable cell lines. MCF10AN cells, ~30% confluent, were infected with retrovirus encoding shRNA for ~15 h using polybrene (8 μg/ml). The infection medium was then replaced with fresh growth medium and cultured for 24 h, following which cells were selected with puromycin (2 μg/ml) for 7 d. MCF10ANeuNT cells expressing doxycycline inducible control and PTP1B shRNA miRs were obtained from Mohamed Bentires-Alj (Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland)³². For experiments with inducible miRs, 500 ng/ml of doxycycline was added to the medium 12 h after seeding the cells and refreshed every 12 h. For experiments overexpressing wild-type and mutant forms of PTP1B, cells were transfected with PMT2-PTP1B WT or PMT2-PTP1BL192A/S372P for 24 h using X-tremeGene HP DNA transfection reagent (Roche), following which cells were untreated or treated with MSI-1436 (1 μM or 2 μM), and cell invasion was studied as described above.

Three-dimensional-morphogenesis assays. For three-dimensional cultures, ~5,000 cells were plated on chamber slides coated with growth factor-reduced Matrigel (BD Biosciences) as described. To activate chimeric ErbB2 protein, 1 μM AP1510 was added to the growth medium. For studies with MSI-1436, 2.5 μM of the inhibitor was used. At least 100 acini were imaged every 4 d using a Zeiss Axiovert 200M microscope and AxioVision 4.5 software (Zeiss).

Cell proliferation assays. Cells (1×10^4) were plated in 96-well plates for 12–96 h in growth medium in the absence and presence of MSI-1436 (2 μM and 5 μM). Cell proliferation was measured using the CellTiter-Glo (CTG; Promega) assay according to the manufacturer's instructions using a luminescence plate reader.

Annexin/PI staining. Cells (1×10^5) were plated in six-well plates in growth medium. The cells were untreated or treated with MSI-1436 (2 μM and 5 μM), stained with annexin and PI according to the manufacturer's instructions

(Invitrogen) and analyzed by flow cytometry (Becton Dickinson LSRII Cell Analyzer).

Animal use. The animal protocol application was reviewed and approved by the Cold Spring Harbor Laboratory Animal Care and Use committee and was designated protocol no. 10-4. The Institution has an Animal Welfare Assurance on file with the Office of Laboratory Animal Welfare (number A3280-01).

PTP1B inhibitor treatment in NDL2 animals. NDL2 animals with palpable tumors were randomly grouped to receive saline (control) or MSI-1436. MSI-1436 was administered (5 mg/kg or 10 mg/kg body weight, with an injection volume of 100 μl) every third day intraperitoneally for 45 d. Control animals received an injection at the same time of 100 μl of saline. Solutions were freshly prepared before each administration. Mice were sacrificed 45 d after the treatment, and individual tumors were removed to measure the weight.

Xenograft studies. Luciferase-expressing BT474 cells (5×10^6) were injected orthotopically into mammary fat pads of SCID-beige mice (Tectonic lab) in 15 μl of a 1:1 mixture with DMEM and growth factor-reduced Matrigel (BD Biosciences). The animals were randomly assigned to two groups to receive saline (control) or MSI-1436 (5 mg/kg) (treated) intraperitoneally every third day. In another study, animals injected with BT474 cells (5×10^6) were allowed to develop tumors. Once the tumors reached ~100 mm³ in volume, treatment with MSI-1436 (5 mg/kg) was initiated. Tumor growth was monitored periodically by injecting 100 μl of *l*-luciferin (15 mg/kg) and imaging the animals using a Xenogen Imager (Xenogen IVIS-200 Optical *in vivo* Imaging System). Tumor volume in mm³ was measured three times weekly by palpation using the formula: volume = width² × length/2.

Mammary tissue harvesting for immunoblotting. After harvesting, connective tissue was carefully dissected from the tumor to exclude stromal cells from whole-tumor cell lysate to be used for immunoblotting. Tissue samples were ground into a powder under liquid nitrogen and lysed for 30 min on ice in lysis buffer (50 mM HEPES, pH 7.4; 10% glycerol; 1% Triton X-100, 150 mM NaCl; 1 mM EGTA, pH 8.0; 1 mM sodium orthovanadate; 20 mM NaF; complete EDTA-free protease cocktail inhibitors). The cell lysates were cleared by centrifugation at 12,000g for 15 min at 4 °C. Lysates (20 μg) were used for immunoblotting.

MSI-1436 binding assay in cell and tumor lysates. MSI-1436 was immobilized on magnetic beads through covalent linkage using its primary amine group. NHS-activated magnetic beads equilibrated in DMSO were incubated with 1 mM MSI-1436 and 100 mM triethylamine for 8 h at 4 °C. The reaction was terminated by pelleting the magnetic beads and gently aspirating the supernatant. Free NHS groups were blocked with 0.8 M aminoethanol. The beads were washed twice sequentially with 100 mM Tris-HCl, pH 8.0, and 100 mM acetate buffer, pH 4.5, containing 150 mM NaCl.

The coupled magnetic beads were incubated with purified recombinant PTP1B (0.5 μg) for 6–8 h at 4 °C in 100 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM DTT and 0.1% (w/v) BSA. Subsequently, the beads were washed in 100 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM DTT, 0.01% Nonidet P-40, 1% glycerol and 0.1% BSA and eluted by heat denaturation in SDS-PAGE sample buffer. The bound protein was separated by 10% SDS-PAGE and subjected to Coomassie blue or silver staining.

For experiments with cell or tumor lysates, 0–1 mg lysate protein was used. Preliminary analysis of derivatives of MSI-1436 suggested that the cholestane moiety of the compound is critical for its binding to PTP1B. Considering the hydrophobic nature of the cholestane core, we anticipated that the interaction would be stabilized predominantly by hydrophobic forces. Hence, we eliminated detergent in the wash buffer. We also observed that the interaction was resistant to salt concentration up to 200 mM (NaCl). Consequently, we chose a final buffer composition of 50 mM HEPES, pH 7.0, 150 mM NaCl, 2 mM TCEP, 5% (v/v) glycerol. MSI-1459 (3β-spermine, 5α, cholenic acid methyl ester), a structurally related inactive analog of MSI-1436, was included as a negative control for binding specificity. Beads coupled to another related compound (MSI-1459) with a primary amine group but poor inhibitory potency against PTP1B was used as a negative control.

Histology. To evaluate the presence and absence of tumors in a secondary site, the entire lung from untreated and treated animals was sectioned and stained with H&E. Whole-slide, digitized images of H&E-stained tissue were

captured using the Aperio ScanScope XT automated scanning system (Vista, CA). Images can be viewed from a public database maintained by Cold Spring Harbor Laboratory. For volumetric measurement of total lung metastasis the slides stained with H&E were analyzed by Aperio software.

Statistical analysis. Statistical analysis was performed using GraphPad Prism software, and two-tailed Student's *t*-tests were applied.

44. Meng, T.C., Hsu, J.S.F. & Tonks, N.K. Development of a modified in-gel assay to identify protein tyrosine phosphatases that are oxidized and inactivated *in vivo*. *Methods* **35**, 28–36 (2005).
45. Lin, G., Aranda, V., Muthuswamy, S.K. & Tonks, N.K. Identification of PTPN23 as a novel regulator of cell invasion in mammary epithelial cells from a loss-of-function screen of the 'PTP-ome'. *Genes Dev.* **25**, 1412–1425 (2011).

10 Bibliography

Alexander, J., Lim, D., Joughin, B.A., Hegemann, B., Hutchins, J.R.A., Ehrenberger, T., Ivins, F., Sessa, F., Hudecz, O., Nigg, E.A., et al. (2011). Spatial exclusivity combined with positive and negative selection of phosphorylation motifs is the basis for context-dependent mitotic signaling. *Sci. Signal.* 4, ra42.

Alvarez, E., Northwood, I.C., Gonzalez, F.A., Latour, D.A., Seth, A., Abate, C., Curran, T., and Davis, R.J. (1991). Pro-Leu-Ser/Thr-Pro is a consensus primary sequence for substrate protein phosphorylation. Characterization of the phosphorylation of c-myc and c-jun proteins by an epidermal growth factor receptor threonine 669 protein kinase. *J. Biol. Chem.* 266, 15277–15285.

Apic, G., Gough, J., and Teichmann, S.A. (2001). Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J. Mol. Biol.* 310, 311–325.

Avruch, J. (2007). MAP kinase pathways: The first twenty years. *Biochim. Biophys. Acta* 1773, 1150–1160.

Bardwell, A.J., Frankson, E., and Bardwell, L. (2009). Selectivity of Docking Sites in MAPK Kinases. *J. Biol. Chem.* 284, 13165–13173.

Bermel, W., Bertini, I., Felli, I.C., Gonnelli, L., Koźmiński, W., Piai, A., Pierattelli, R., and Stanek, J. (2012). Speeding up sequence specific assignment of IDPs. *J. Biomol. NMR* 53, 293–301.

Bermel, W., Felli, I.C., Gonnelli, L., Koźmiński, W., Piai, A., Pierattelli, R., and Zawadzka-Kazimierzczuk, A. (2013). High-dimensionality ¹³C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. *J. Biomol. NMR* 57, 353–361.

Bernadó, P., and Blackledge, M. (2010). Structural biology: Proteins in dynamic equilibrium. *Nature* 468, 1046–1048.

Bernadó, P., Bertoncini, C.W., Griesinger, C., Zweckstetter, M., and Blackledge, M. (2005a). Defining long-range order and local disorder in native alpha-synuclein using residual dipolar couplings. *J. Am. Chem. Soc.* 127, 17968–17969.

Bernadó, P., Blanchard, L., Timmins, P., Marion, D., Ruigrok, R.W.H., and Blackledge, M. (2005b). A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc. Natl. Acad. Sci. U. S. A.* 102, 17002–17007.

Bertini, I., Gupta, Y.K., Luchinat, C., Parigi, G., Peana, M., Sgheri, L., and Yuan, J. (2007). Paramagnetism-based NMR restraints provide maximum allowed probabilities for the different conformations of partially independent protein domains. *J. Am. Chem. Soc.* 129, 12786–12794.

Bhattacharyya, R.P., Reményi, A., Good, M.C., Bashor, C.J., Falick, A.M., and Lim, W.A. (2006). The Ste5 Scaffold Allosterically Modulates Signaling Output of the Yeast Mating Pathway. *Science* 311, 822–826.

Bhunia, A., Mohanram, H., and Bhattacharjya, S. (2012). Structural determinants of the specificity of a membrane binding domain of the scaffold protein Ste5 of budding yeast: Implications in signaling by the scaffold protein in MAPK pathway. *Biochim. Biophys. Acta BBA - Biomembr.*

1818, 1250–1260.

Blackledge, M. (2005). Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings. *Prog. Nucl. Magn. Reson. Spectrosc.* *46*, 23–61.

Boivin, S., Kozak, S., and Meijers, R. (2013). Optimization of protein purification and characterization using Thermofluor screens. *Protein Expr. Purif.* *91*, 192–206.

Borsello, T., Centeno, C., Riederer, I.M., Haefliger, J.-A., and Riederer, B.M. (2007). Phosphorylation-dependent dimerization and subcellular localization of islet-brain 1/c-Jun N-terminal kinase-interacting protein 1. *J. Neurosci. Res.* *85*, 3632–3641.

Bouvignies, G., Bernadó, P., Meier, S., Cho, K., Grzesiek, S., Brüschweiler, R., and Blackledge, M. (2005). Identification of slow correlated motions in proteins using residual dipolar and hydrogen-bond scalar couplings. *Proc. Natl. Acad. Sci. U. S. A.* *102*, 13885–13890.

Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Rovesti, P., Sharff, A., Smart, O.S., Vonrhein, C., et al. (2011). BUSTER version 1.10.0. *Camb. U. K. Glob. Phasing Ltd.*

Buchan, D.W.A., Minneci, F., Nugent, T.C.O., Bryson, K., and Jones, D.T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.* *41*, W349–W357.

Buday, L., and Tompa, P. (2010). Functional classification of scaffold proteins and related molecules. *FEBS J.* *277*, 4348–4355.

Camilloni, C., De Simone, A., Vranken, W.F., and Vendruscolo, M. (2012). Determination of Secondary Structure Populations in Disordered States of Proteins Using Nuclear Magnetic Resonance Chemical Shifts. *Biochemistry (Mosc.)* *51*, 2224–2231.

Campen, A., Williams, R.M., Brown, C.J., Meng, J., Uversky, V.N., and Dunker, A.K. (2008). TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept. Lett.* *15*, 956–963.

Carver, J.P., and Richards, R.E. (1972). A general two-site solution for the chemical exchange produced dependence of T₂ upon the carr-Purcell pulse separation. *J. Magn. Reson.* *1969* *6*, 89–105.

Chang, C.I., Xu, B., Akella, R., Cobb, M.H., and Goldsmith, E.J. (2002). Crystal structures of MAP kinase p38 complexed to the docking sites on its nuclear substrate MEF2A and activator MKK3b. *Mol. Cell* *9*, 1241–1249.

Clark-Lewis, I., Sanghera, J.S., and Pelech, S.L. (1991). Definition of a consensus sequence for peptide substrate recognition by p44mpk, the meiosis-activated myelin basic protein kinase. *J. Biol. Chem.* *266*, 15180–15184.

Cohen, G.B., Ren, R., and Baltimore, D. (1995). Modular binding domains in signal transduction proteins. *Cell* *80*, 237–248.

Cohen-Katsenelson, K., Wasserman, T., Khateb, S., Whitmarsh, A.J., and Aronheim, A. (2011). Docking interactions of the JNK scaffold protein WDR62. *Biochem. J.* *439*, 381–390.

- Collins, M.O., Yu, L., Campuzano, I., Grant, S.G.N., and Choudhary, J.S. (2008). Phosphoproteomic analysis of the mouse brain cytosol reveals a predominance of protein phosphorylation in regions of intrinsic sequence disorder. *Mol. Cell. Proteomics MCP* 7, 1331–1348.
- Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* 13, 289–302.
- Davey, N.E., Van Roey, K., Weatheritt, R.J., Toedt, G., Uyar, B., Altenberg, B., Budd, A., Diella, F., Dinkel, H., and Gibson, T.J. (2012). Attributes of short linear motifs. *Mol. Biosyst.* 8, 268–281.
- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6, 277–293.
- Deshmukh, L., Schwieters, C.D., Grishaev, A., Ghirlando, R., Baber, J.L., and Clore, G.M. (2013). Structure and Dynamics of Full-Length HIV-1 Capsid Protein in Solution. *J. Am. Chem. Soc.* 135, 16133–16147.
- Diella, F., Haslam, N., Chica, C., Budd, A., Michael, S., Brown, N.P., Trave, G., and Gibson, T.J. (2008). Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front. Biosci. J. Virtual Libr.* 13, 6580–6603.
- Dinkel, H., Van Roey, K., Michael, S., Davey, N.E., Weatheritt, R.J., Born, D., Speck, T., Krüger, D., Grebnev, G., Kuban, M., et al. (2014). The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic Acids Res.* 42, D259–D266.
- Doig, A.J. (2002). Recent advances in helix–coil theory. *Biophys. Chem.* 101–102, 281–293.
- Dunker, A.K., Obradovic, Z., Romero, P., Garner, E.C., and Brown, C.J. (2000). Intrinsic protein disorder in complete genomes. *Genome Inform. Workshop Genome Inform.* 11, 161–171.
- Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., et al. (2001). Intrinsically disordered protein. *J. Mol. Graph. Model.* 19, 26–59.
- Dunker, A.K., Silman, I., Uversky, V.N., and Sussman, J.L. (2008). Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.* 18, 756–764.
- Dutta, K., Shi, H., Cruz-Chu, E.R., Kami, K., and Ghose, R. (2004). Dynamic influences on a high-affinity, high-specificity interaction involving the C-terminal SH3 domain of p67phox. *Biochemistry (Mosc.)* 43, 8094–8106.
- Dyson, H.J., and Wright, P.E. (2002). Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* 12, 54–60.
- Eghbalian, H.R., Wang, L., Bahrami, A., Assadi, A., and Markley, J.L. (2005). Protein energetic conformational analysis from NMR chemical shifts (PECAN) and its use in determining secondary structural elements. *J. Biomol. NMR* 32, 71–81.
- Ember, B., and LoGrasso, P. (2008). Mechanistic characterization for c-jun-N-Terminal Kinase 1 α 1. *Arch. Biochem. Biophys.* 477, 324–329.

10. Bibliography

- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132.
- Erikson, E., and Maller, J.L. (1985). A protein kinase from *Xenopus* eggs specific for ribosomal protein S6. *Proc. Natl. Acad. Sci. U. S. A.* 82, 742–746.
- Erikson, E., and Maller, J.L. (1986). Purification and characterization of a protein kinase from *Xenopus* eggs highly specific for ribosomal protein S6. *J. Biol. Chem.* 261, 350–355.
- Feldman, H.J., and Hogue, C.W.V. (2000). A fast method to sample real protein conformational space. *Proteins Struct. Funct. Bioinforma.* 39, 112–131.
- Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230.
- Fisher, C.K., and Stultz, C.M. (2011). Constructing ensembles for intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 21, 426–431.
- Fisher, C.K., Huang, A., and Stultz, C.M. (2010). Modeling Intrinsically Disordered Proteins with Bayesian Statistics. *J. Am. Chem. Soc.* 132, 14919–14927.
- Francis, D.M., Rózycki, B., Koveal, D., Hummer, G., Page, R., and Peti, W. (2011). Structural basis of p38 α regulation by hematopoietic tyrosine phosphatase. *Nat. Chem. Biol.* 7, 916–924.
- Freund, C., Kühne, R., Yang, H., Park, S., Reinherz, E.L., and Wagner, G. (2002). Dynamic interaction of CD2 with the GYF and the SH3 domain of compartmentalized effector molecules. *EMBO J.* 21, 5985–5995.
- Frishman, D., and Argos, P. (1995). Knowledge-based protein secondary structure assignment. *Proteins* 23, 566–579.
- Frith, M.C., Saunders, N.F.W., Kobe, B., and Bailey, T.L. (2008). Discovering Sequence Motifs with Arbitrary Insertions and Deletions. *PLoS Comput Biol* 4, e1000071.
- Fu, M., and Holzbaur, E.L.F. (2013). JIP1 regulates the directionality of APP axonal transport by coordinating kinesin and dynein motors. *J. Cell Biol.* 202, 495–508.
- Fuxreiter, M., Simon, I., Friedrich, P., and Tompa, P. (2004). Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J. Mol. Biol.* 338, 1015–1026.
- Fuxreiter, M., Tompa, P., and Simon, I. (2007). Local structural disorder imparts plasticity on linear motifs. *Bioinforma. Oxf. Engl.* 23, 950–956.
- Garai, A., Zeke, A., Gogl, G., Toro, I., Ferenc, F., Blankenburg, H., Barkai, T., Varga, J., Alexa, A., Emig, D., et al. (2012). Specificity of Linear Motifs That Bind to a Common Mitogen-Activated Protein Kinase Docking Groove. *Sci. Signal.* 5, ra74.
- Gibson, T.J. (2009). Cell regulation: determined to signal discrete cooperation. *Trends Biochem. Sci.* 34, 471–482.

- Glatz, G., Gógl, G., Alexa, A., and Reményi, A. (2013). Structural mechanism for the specific assembly and activation of the extracellular signal regulated kinase 5 (ERK5) module. *J. Biol. Chem.* *288*, 8596–8609.
- Göbl, C., Madl, T., Simon, B., and Sattler, M. (2014). NMR approaches for structural analysis of multidomain proteins and complexes in solution. *Prog. Nucl. Magn. Reson. Spectrosc.* *80*, 26–63.
- Goldsmith, E.J. (2011). Three-Dimensional Docking in the MAPK p38 α . *Sci. Signal.* *4*, pe47–pe47.
- Good, M., Tang, G., Singleton, J., Remenyi, A., and Lim, W.A. (2009). Scaffold-Assisted Catalysis: A Novel Domain In the Ste5 Scaffold Protein is Required to Unlock the MAPK Fus3 for Phosphorylation by the MAPKK Ste7. *Cell* *136*, 1085–1097.
- Good, M.C., Zalatan, J.G., and Lim, W.A. (2011). Scaffold Proteins: Hubs for Controlling the Flow of Cellular Information. *Science* *332*, 680–686.
- Graf, J., Nguyen, P.H., Stock, G., and Schwalbe, H. (2007). Structure and dynamics of the homologous series of alanine peptides: a joint molecular dynamics/NMR study. *J. Am. Chem. Soc.* *129*, 1179–1189.
- Guerry, P., Salmon, L., Mollica, L., Ortega Roldan, J.-L., Markwick, P., van Nuland, N.A.J., McCammon, J.A., and Blackledge, M. (2013). Mapping the Population of Protein Conformational Energy Sub-States from NMR Dipolar Couplings. *Angew. Chem. Int. Ed.* *52*, 3181–3185.
- Guo, C., and Whitmarsh, A.J. (2008). The beta-arrestin-2 scaffold protein promotes c-Jun N-terminal kinase-3 activation by binding to its nonconserved N terminus. *J. Biol. Chem.* *283*, 15903–15911.
- Hagarman, A., Measey, T.J., Mathieu, D., Schwalbe, H., and Schweitzer-Stenner, R. (2010). Intrinsic Propensities of Amino Acid Residues in GxG Peptides Inferred from Amide I' Band Profiles and NMR Scalar Coupling Constants. *J. Am. Chem. Soc.* *132*, 540–551.
- Hamelberg, D., Mongan, J., and McCammon, J.A. (2004). Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* *120*, 11919–11929.
- Hammes, G.G., Chang, Y.-C., and Oas, T.G. (2009). Conformational selection or induced fit: a flux description of reaction mechanism. *Proc. Natl. Acad. Sci. U. S. A.* *106*, 13737–13741.
- Hansen, M.R., Mueller, L., and Pardi, A. (1998). Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat. Struct. Biol.* *5*, 1065–1074.
- Hansmann, U.H.E. (1997). Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* *281*, 140–150.
- Hao, B., Oehlmann, S., Sowa, M.E., Harper, J.W., and Pavletich, N.P. (2007). Structure of a Fbw7-Skp1-cyclin E complex: multisite-phosphorylated substrate recognition by SCF ubiquitin ligases. *Mol. Cell* *26*, 131–143.
- Henzler-Wildman, K., and Kern, D. (2007). Dynamic personalities of proteins. *Nature* *450*, 964–972.

10. Bibliography

- Heo, Y.-S., Kim, S.-K., Seo, C.I., Kim, Y.K., Sung, B.-J., Lee, H.S., Lee, J.I., Park, S.-Y., Kim, J.H., Hwang, K.Y., et al. (2004). Structural basis for the selective inhibition of JNK1 by the scaffolding protein JIP1 and SP600125. *EMBO J.* *23*, 2185–2195.
- Ho, D.T., Bardwell, A.J., Abdollahi, M., and Bardwell, L. (2003). A docking site in MKK4 mediates high affinity binding to JNK MAPKs and competes with similar docking sites in JNK substrates. *J. Biol. Chem.* *278*, 32662–32672.
- Ho, D.T., Bardwell, A.J., Grewal, S., Iverson, C., and Bardwell, L. (2006). Interacting JNK-docking sites in MKK7 promote binding and activation of JNK mitogen-activated protein kinases. *J. Biol. Chem.* *281*, 13169–13179.
- Hollingsworth, S.A., Berkholz, D.S., and Karplus, P.A. (2009). On the occurrence of linear groups in proteins. *Protein Sci. Publ. Protein Soc.* *18*, 1321–1325.
- Huang, J., Ozenne, V., Jensen, M.R., and Blackledge, M. (2013). Direct Prediction of NMR Residual Dipolar Couplings from the Primary Sequence of Unfolded Proteins. *Angew. Chem. Int. Ed.* *52*, 687–690.
- Huang, J., Warner, L.R., Sanchez, C., Gabel, F., Madl, T., Mackereth, C.D., Sattler, M., and Blackledge, M. (2014). Transient Electrostatic Interactions Dominate the Conformational Equilibrium Sampled by Multidomain Splicing Factor U2AF65: A Combined NMR and SAXS Study. *J. Am. Chem. Soc.* *136*, 7068–7076.
- Hung, L.-H., and Samudrala, R. (2003). Accurate and automated classification of protein secondary structure with PsiCSI. *Protein Sci. Publ. Protein Soc.* *12*, 288–295.
- Hutti, J.E., Jarrell, E.T., Chang, J.D., Abbott, D.W., Storz, P., Toker, A., Cantley, L.C., and Turk, B.E. (2004). A rapid method for determining protein kinase phosphorylation specificity. *Nat. Methods* *1*, 27–29.
- Iakoucheva, L.M. (2004). The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* *32*, 1037–1049.
- Iakoucheva, L.M., Brown, C.J., Lawson, J.D., Obradović, Z., and Dunker, A.K. (2002). Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.* *323*, 573–584.
- Iešmantavičius, V., Dogan, J., Jemth, P., Teilum, K., and Kjaergaard, M. (2014). Helical Propensity in an Intrinsically Disordered Protein Accelerates Ligand Binding. *Angew. Chem. Int. Ed.* *53*, 1548–1551.
- Ittner, L.M., Ke, Y.D., and Götz, J. (2009). Phosphorylated Tau interacts with c-Jun N-terminal kinase-interacting protein 1 (JIP1) in Alzheimer disease. *J. Biol. Chem.* *284*, 20909–20916.
- Jensen, M.R., and Blackledge, M. (2008). On the origin of NMR dipolar waves in transient helical elements of partially folded proteins. *J. Am. Chem. Soc.* *130*, 11266–11267.
- Jensen, M.R., Houben, K., Lescop, E., Blanchard, L., Ruigrok, R.W.H., and Blackledge, M. (2008). Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein. *J. Am. Chem. Soc.* *130*, 8055–8061.

- Jensen, M.R., Markwick, P.R.L., Meier, S., Griesinger, C., Zweckstetter, M., Grzesiek, S., Bernadó, P., and Blackledge, M. (2009). Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. *Struct. Lond. Engl.* 1993 17, 1169–1185.
- Jensen, M.R., Salmon, L., Nodet, G., and Blackledge, M. (2010). Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts. *J. Am. Chem. Soc.* 132, 1270–1272.
- Jensen, M.R., Communie, G., Ribeiro, E.A., Martinez, N., Desfosses, A., Salmon, L., Mollica, L., Gabel, F., Jamin, M., Longhi, S., et al. (2011). Intrinsic disorder in measles virus nucleocapsids. *Proc. Natl. Acad. Sci. U. S. A.* 108, 9839–9844.
- Jensen, M.R., Zweckstetter, M., Huang, J., and Blackledge, M. (2014). Exploring Free-Energy Landscapes of Intrinsically Disordered Proteins at Atomic Resolution Using NMR Spectroscopy. *Chem. Rev.* 114, 6632–6660.
- Jha, A.K., Colubri, A., Freed, K.F., and Sosnick, T.R. (2005a). Statistical coil model of the unfolded state: resolving the reconciliation problem. *Proc. Natl. Acad. Sci. U. S. A.* 102, 13099–13104.
- Jha, A.K., Colubri, A., Zaman, M.H., Koide, S., Sosnick, T.R., and Freed, K.F. (2005b). Helix, sheet, and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library. *Biochemistry (Mosc.)* 44, 9691–9702.
- Jung, Y.-S., and Zweckstetter, M. (2004). Mars -- robust automatic backbone assignment of proteins. *J. Biomol. NMR* 30, 11–23.
- Kabsch, W. (2010). XDS. *Acta Crystallogr. D Biol. Crystallogr.* 66, 125–132.
- Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637.
- Kami, K., Takeya, R., Sumimoto, H., and Kohda, D. (2002). Diverse recognition of non PxxP peptide ligands by the SH3 domains from p67phox, Grb2 and Pex13p. *EMBO J.* 21, 4268–4276.
- Karplus, M. (1959). Contact Electron Spin Coupling of Nuclear Magnetic Moments. *J. Chem. Phys.* 30, 11–15.
- Kathiriya, J.J., Pathak, R.R., Clayman, E., Xue, B., Uversky, V.N., and Davé, V. (2014). Presence and utility of intrinsically disordered regions in kinases. *Mol. Biosyst.*
- Kato, Y., Ito, M., Kawai, K., Nagata, K., and Tanokura, M. (2002). Determinants of Ligand Specificity in Groups I and IV WW Domains as Studied by Surface Plasmon Resonance and Model Building. *J. Biol. Chem.* 277, 10173–10177.
- Kato, Y., Nagata, K., Takahashi, M., Lian, L., Herrero, J.J., Sudol, M., and Tanokura, M. (2004). Common Mechanism of Ligand Recognition by Group II/III WW Domains. *J. Biol. Chem.* 279, 31833–31841.
- Kazimierczuk, K., Zawadzka, A., and Koźmiński, W. (2009). Narrow peaks and high dimensionalities: exploiting the advantages of random sampling. *J. Magn. Reson. San Diego Calif*

1997 197, 219–228.

Keshet, Y., and Seger, R. (2010). The MAP Kinase Signaling Cascades: A System of Hundreds of Components Regulates a Diverse Array of Physiological Functions. In *MAP Kinase Signaling Protocols*, R. Seger, ed. (Humana Press), pp. 3–38.

Keyse, S.M. (2000). Protein phosphatases and the regulation of mitogen-activated protein kinase signalling. *Curr. Opin. Cell Biol.* 12, 186–192.

Kiefhaber, T., Bachmann, A., and Jensen, K.S. (2012). Dynamics and mechanisms of coupled protein folding and binding reactions. *Curr. Opin. Struct. Biol.* 22, 21–29.

Kjaergaard, M., and Poulsen, F.M. (2011). Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. *J. Biomol. NMR* 50, 157–165.

Kjaergaard, M., Brander, S., and Poulsen, F.M. (2011). Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. *J. Biomol. NMR* 49, 139–149.

Kõivomägi, M., Valk, E., Venta, R., Iofik, A., Lepiku, M., Balog, E.R.M., Rubin, S.M., Morgan, D.O., and Loog, M. (2011). Cascades of multisite phosphorylation control Sic1 destruction at the onset of S phase. *Nature* 480, 128–131.

Kõivomägi, M., Örd, M., Iofik, A., Valk, E., Venta, R., Faustova, I., Kivi, R., Balog, E.R.M., Rubin, S.M., and Loog, M. (2013). Multisite phosphorylation networks as signal processors for Cdk1. *Nat. Struct. Mol. Biol.* 20, 1415–1424.

Kook, S., Zhan, X., Kaoud, T.S., Dalby, K.N., Gurevich, V.V., and Gurevich, E.V. (2013). Arrestin-3 Binds c-Jun N-terminal Kinase 1 (JNK1) and JNK2 and Facilitates the Activation of These Ubiquitous JNK Isoforms in Cells via Scaffolding. *J. Biol. Chem.* 288, 37332–37342.

Kornev, A.P., Haste, N.M., Taylor, S.S., and Eyck, L.F.T. (2006). Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc. Natl. Acad. Sci. U. S. A.* 103, 17783–17788.

Kornev, A.P., Taylor, S.S., and Ten Eyck, L.F. (2008). A helix scaffold for the assembly of active protein kinases. *Proc. Natl. Acad. Sci. U. S. A.* 105, 14377–14382.

Kragelj, J., Ozenne, V., Blackledge, M., and Jensen, M.R. (2013). Conformational propensities of intrinsically disordered proteins from NMR chemical shifts. *Chemphyschem Eur. J. Chem. Phys. Phys. Chem.* 14, 3034–3045.

Krzeminski, M., Marsh, J.A., Neale, C., Choy, W.-Y., and Forman-Kay, J.D. (2013). Characterization of disordered proteins with ENSEMBLE. *Bioinforma. Oxf. Engl.* 29, 398–399.

Kukekov, N.V., Xu, Z., and Greene, L.A. (2006). Direct interaction of the molecular scaffolds POSH and JIP is required for apoptotic activation of JNKs. *J. Biol. Chem.* 281, 15517–15524.

Kuriyan, J., Petsko, G.A., Levy, R.M., and Karplus, M. (1986). Effect of anisotropy and anharmonicity on protein crystallographic refinement. An evaluation by molecular dynamics. *J. Mol. Biol.* 190, 227–254.

- Kyriakis, J.M., and Avruch, J. (2012). Mammalian MAPK signal transduction pathways activated by stress and inflammation: a 10-year update. *Physiol. Rev.* 92, 689–737.
- Labudde, D., Leitner, D., Krüger, M., and Oschkinat, H. (2003). Prediction algorithm for amino acid types with their secondary structure in proteins (PLATON) using chemical shifts. *J. Biomol. NMR* 25, 41–53.
- Lacroix, E., Viguera, A.R., and Serrano, L. (1998). Elucidating the folding problem of alpha-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *J. Mol. Biol.* 284, 173–191.
- Lange, O.F., Lakomek, N.-A., Farès, C., Schröder, G.F., Walter, K.F.A., Becker, S., Meiler, J., Grubmüller, H., Griesinger, C., and de Groot, B.L. (2008). Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* 320, 1471–1475.
- Laughlin, J.D., Nwachukwu, J.C., Figuera-Losada, M., Cherry, L., Nettles, K.W., and LoGrasso, P.V. (2012). Structural mechanisms of allostery and autoinhibition in JNK family kinases. *Struct. Lond. Engl.* 1993 20, 2174–2184.
- Leader, D.P., and Milner-White, E. (2012). Structure Motivator: A tool for exploring small three-dimensional elements in proteins. *BMC Struct. Biol.* 12, 26.
- Leader, D.P., and Milner-White, E.J. (2011). The structure of the ends of α -helices in globular proteins: effect of additional hydrogen bonds and implications for helix formation. *Proteins* 79, 1010–1019.
- Lee, C.M., Onésime, D., Reddy, C.D., Dhanasekaran, N., and Reddy, E.P. (2002). JLP: A scaffolding protein that tethers JNK/p38MAPK signaling modules and transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* 99, 14189–14194.
- Lee, T., Hoofnagle, A.N., Kabuyama, Y., Stroud, J., Min, X., Goldsmith, E.J., Chen, L., Resing, K.A., and Ahn, N.G. (2004). Docking motif interactions in MAP kinases revealed by hydrogen exchange mass spectrometry. *Mol. Cell* 14, 43–55.
- Lei, X.-F., Kim Kaneyama, J., Arita Okubo, S., Offermanns, S., Itabe, H., Miyazaki, T., and Miyazaki, A. (2014). Identification of Hic-5 as a Novel Scaffold for the MKK4/p54 JNK Pathway in the Development of Abdominal Aortic Aneurysms. *J. Am. Heart Assoc.* 3, e000747.
- Leone, V., Marinelli, F., Carloni, P., and Parrinello, M. (2010). Targeting biomolecular flexibility with metadynamics. *Curr. Opin. Struct. Biol.* 20, 148–154.
- Lescop, E., Schanda, P., and Brutscher, B. (2007). A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J. Magn. Reson. San Diego Calif 1997* 187, 163–169.
- Lim, W.A. (2002). The modular logic of signaling proteins: building allosteric switches from simple binding domains. *Curr. Opin. Struct. Biol.* 12, 61–68.
- Lindorff-Larsen, K., Best, R.B., Depristo, M.A., Dobson, C.M., and Vendruscolo, M. (2005). Simultaneous determination of protein structure and dynamics. *Nature* 433, 128–132.

10. Bibliography

- Lindorff-Larsen, K., Trbovic, N., Maragakis, P., Piana, S., and Shaw, D.E. (2012). Structure and dynamics of an unfolded protein examined by molecular dynamics simulation. *J. Am. Chem. Soc.* *134*, 3787–3791.
- Liu, B.A., Jablonowski, K., Shah, E.E., Engelmann, B.W., Jones, R.B., and Nash, P.D. (2010). SH2 Domains Recognize Contextual Peptide Sequence Information to Determine Selectivity. *Mol. Cell. Proteomics* *9*, 2391–2404.
- Lupas, A., Van Dyke, M., and Stock, J. (1991). Predicting coiled coils from protein sequences. *Science* *252*, 1162–1164.
- MacArthur, M.W., and Thornton, J.M. (1991). Influence of proline residues on protein conformation. *J. Mol. Biol.* *218*, 397–412.
- Marsh, J.A., and Forman-Kay, J.D. (2009). Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints. *J. Mol. Biol.* *391*, 359–374.
- Marsh, J.A., Singh, V.K., Jia, Z., and Forman-Kay, J.D. (2006). Sensitivity of secondary structure propensities to sequence differences between β - and β -synuclein: Implications for fibrillation. *Protein Sci. Publ. Protein Soc.* *15*, 2795–2804.
- Marsh, J.A., Baker, J.M.R., Tollinger, M., and Forman-Kay, J.D. (2008). Calculation of Residual Dipolar Couplings from Disordered State Ensembles Using Local Alignment. *J. Am. Chem. Soc.* *130*, 7804–7805.
- Matsuda, S., Yasukawa, T., Homma, Y., Ito, Y., Niikura, T., Hiraki, T., Hirai, S., Ohno, S., Kita, Y., Kawasumi, M., et al. (2001). c-Jun N-terminal kinase (JNK)-interacting protein-1b/islet-brain-1 scaffolds Alzheimer's amyloid precursor protein with JNK. *J. Neurosci. Off. J. Soc. Neurosci.* *21*, 6597–6607.
- Matsuda, S., Matsuda, Y., and D'Adamio, L. (2003). Amyloid beta protein precursor (A β PP), but not A β PP-like protein 2, is bridged to the kinesin light chain by the scaffold protein JNK-interacting protein 1. *J. Biol. Chem.* *278*, 38601–38606.
- Mayer, B.J. (2001). SH3 domains: complexity in moderation. *J. Cell Sci.* *114*, 1253–1263.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* *40*, 658–674.
- McGrath, D.A., Balog, E.R.M., Kõivomägi, M., Lucena, R., Mai, M.V., Hirschi, A., Kellogg, D.R., Loog, M., and Rubin, S.M. (2013). Cks confers specificity to phosphorylation-dependent CDK signaling pathways. *Nat. Struct. Mol. Biol.* *20*, 1407–1414.
- Meharena, H.S., Chang, P., Keshwani, M.M., Oruganty, K., Nene, A.K., Kannan, N., Taylor, S.S., and Kornev, A.P. (2013). Deciphering the Structural Basis of Eukaryotic Protein Kinase Regulation. *PLoS Biol* *11*, e1001680.
- Meier, S., Grzesiek, S., and Blackledge, M. (2007). Mapping the Conformational Landscape of Urea-Denatured Ubiquitin Using Residual Dipolar Couplings. *J. Am. Chem. Soc.* *129*, 9799–9807.

- Meng, W., Swenson, L.L., Fitzgibbon, M.J., Hayakawa, K., Ter Haar, E., Behrens, A.E., Fulghum, J.R., and Lippke, J.A. (2002). Structure of mitogen-activated protein kinase-activated protein (MAPKAP) kinase 2 suggests a bifunctional switch that couples kinase activation with nuclear export. *J. Biol. Chem.* 277, 37401–37405.
- Meyer, D., Liu, A., and Margolis, B. (1999). Interaction of c-Jun amino-terminal kinase interacting protein-1 with p190 rhoGEF and its localization in differentiated neurons. *J. Biol. Chem.* 274, 35113–35118.
- Mielke, S.P., and Krishnan, V.V. (2009). Characterization of protein secondary structure from NMR chemical shifts. *Prog. Nucl. Magn. Reson. Spectrosc.* 54, 141–165.
- Mistry, J., Coghill, P., Eberhardt, R.Y., Deiana, A., Giansanti, A., Finn, R.D., Bateman, A., and Punta, M. (2013). The challenge of increasing Pfam coverage of the human proteome. *Database* 2013, bat023–bat023.
- Mittag, T., Orlicky, S., Choy, W.-Y., Tang, X., Lin, H., Sicheri, F., Kay, L.E., Tyers, M., and Forman-Kay, J.D. (2008). Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proc. Natl. Acad. Sci.* 105, 17772–17777.
- Mittermaier, A., and Kay, L.E. (2006). New Tools Provide New Insights in NMR Studies of Protein Dynamics. *Science* 312, 224–228.
- Mooney, L.M., and Whitmarsh, A.J. (2004). Docking interactions in the c-Jun N-terminal kinase pathway. *J. Biol. Chem.* 279, 11843–11852.
- Mukhopadhyay, N.K., Price, D.J., Kyriakis, J.M., Pelech, S., Sanghera, J., and Avruch, J. (1992). An array of insulin-activated, proline-directed serine/threonine protein kinases phosphorylate the p70 S6 kinase. *J. Biol. Chem.* 267, 3325–3335.
- Mukrasch, M.D., Markwick, P., Biernat, J., Bergen, M. von, Bernadó, P., Griesinger, C., Mandelkow, E., Zweckstetter, M., and Blackledge, M. (2007). Highly populated turn conformations in natively unfolded tau protein identified from residual dipolar couplings and molecular simulation. *J. Am. Chem. Soc.* 129, 5235–5243.
- Muñoz, V., and Serrano, L. (1994). Elucidating the folding problem of helical peptides using empirical parameters. *Nat. Struct. Mol. Biol.* 1, 399–409.
- Muñoz, V., and Serrano, L. (1995). Helix design, prediction and stability. *Curr. Opin. Biotechnol.* 6, 382–386.
- Nakamura, K., and Johnson, G.L. (2007). Noncanonical Function of MEKK2 and MEK5 PB1 Domains for Coordinated Extracellular Signal-Regulated Kinase 5 and c-Jun N-Terminal Kinase Signaling. *Mol. Cell. Biol.* 27, 4566–4577.
- Narayanan, C., Weinstock, D.S., Wu, K.-P., Baum, J., and Levy, R.M. (2012). Investigation of the Polymeric Properties of α -Synuclein and Comparison with NMR Experiments: A Replica Exchange Molecular Dynamics Study. *J. Chem. Theory Comput.* 8, 3929–3942.
- Nihalani, D., Meyer, D., Pajni, S., and Holzman, L.B. (2001). Mixed lineage kinase-dependent JNK activation is governed by interactions of scaffold protein JIP with MAPK module components.

10. Bibliography

EMBO J. 20, 3447–3458.

Nihalani, D., Wong, H.N., and Holzman, L.B. (2003). Recruitment of JNK to JIP1 and JNK-dependent JIP1 phosphorylation regulates JNK module dynamics and activation. *J. Biol. Chem.* 278, 28694–28702.

Nodet, G., Salmon, L., Ozenne, V., Meier, S., Jensen, M.R., and Blackledge, M. (2009). Quantitative Description of Backbone Conformational Sampling of Unfolded Proteins at Amino Acid Resolution from NMR Residual Dipolar Couplings. *J. Am. Chem. Soc.* 131, 17908–17918.

Orlicky, S., Tang, X., Willems, A., Tyers, M., and Sicheri, F. (2003). Structural basis for phosphodependent substrate selection and orientation by the SCFCdc4 ubiquitin ligase. *Cell* 112, 243–256.

Ozenne, V., Schneider, R., Yao, M., Huang, J., Salmon, L., Zweckstetter, M., Jensen, M.R., and Blackledge, M. (2012a). Mapping the potential energy landscape of intrinsically disordered proteins at amino acid resolution. *J. Am. Chem. Soc.* 134, 15138–15148.

Ozenne, V., Bauer, F., Salmon, L., Huang, J.-R., Jensen, M.R., Segard, S., Bernadó, P., Charavay, C., and Blackledge, M. (2012b). Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinforma. Oxf. Engl.* 28, 1463–1470.

Pawson, T. (1995). Protein modules and signalling networks. *Nature* 373, 573–580.

Perkins, J.R., Diboun, I., Dessailly, B.H., Lees, J.G., and Orengo, C. (2010). Transient Protein-Protein Interactions: Structural, Functional, and Network Properties. *Structure* 18, 1233–1243.

Piana, S., and Laio, A. (2007). A bias-exchange approach to protein folding. *J. Phys. Chem. B* 111, 4553–4559.

Pierce, L.C.T., Salomon-Ferrer, R., Augusto F de Oliveira, C., McCammon, J.A., and Walker, R.C. (2012). Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J. Chem. Theory Comput.* 8, 2997–3002.

Prestegard, J.H., Bougault, C.M., and Kishore, A.I. (2004). Residual dipolar couplings in structure determination of biomolecules. *Chem. Rev.* 104, 3519–3540.

Prestegard, J.H., Sahu, S.C., Nkari, W.K., Morris, L.C., Live, D., and Gruta, C. (2013). Chemical Shift Prediction for Denatured Proteins. *J. Biomol. NMR* 55, 201–209.

Raman, M., Chen, W., and Cobb, M.H. (2007). Differential regulation and properties of MAPKs. *Oncogene* 26, 3100–3112.

Rasia, R.M., Lescop, E., Palatnik, J.F., Boisbouvier, J., and Brutscher, B. (2011). Rapid measurement of residual dipolar couplings for fast fold elucidation of proteins. *J. Biomol. NMR* 51, 369–378.

Ren, S., Uversky, V.N., Chen, Z., Dunker, A.K., and Obradovic, Z. (2008). Short Linear Motifs recognized by SH2, SH3 and Ser/Thr Kinase domains are conserved in disordered protein regions. *BMC Genomics* 9 Suppl 2, S26.

- Rincón, M., and Davis, R.J. (2009). Regulation of the immune response by stress-activated protein kinases. *Immunol. Rev.* 228, 212–224.
- Van Roey, K., Gibson, T.J., and Davey, N.E. (2012). Motif switches: decision-making in cell regulation. *Curr. Opin. Struct. Biol.* 22, 378–385.
- Van Roey, K., Dinkel, H., Weatheritt, R.J., Gibson, T.J., and Davey, N.E. (2013). The switches.ELM resource: a compendium of conditional regulatory interaction interfaces. *Sci. Signal.* 6, rs7.
- Rogers, J.M., Wong, C.T., and Clarke, J. (2014). Coupled Folding and Binding of the Disordered Protein PUMA Does Not Require Particular Residual Structure. *J. Am. Chem. Soc.* 136, 5197–5200.
- Romero, P., Obradovic, Z., Kissinger, C.R., Villafranca, J.E., Garner, E., Guillot, S., and Dunker, A.K. (1998). Thousands of proteins likely to have long disordered regions. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.* 437–448.
- Rose, B.A., Force, T., and Wang, Y. (2010). Mitogen-activated protein kinase signaling in the heart: angels versus demons in a heart-breaking tale. *Physiol. Rev.* 90, 1507–1546.
- Rossomando, A.J., Payne, D.M., Weber, M.J., and Sturgill, T.W. (1989). Evidence that pp42, a major tyrosine kinase target protein, is a mitogen-activated serine/threonine protein kinase. *Proc. Natl. Acad. Sci. U. S. A.* 86, 6940–6943.
- Rózycki, B., Kim, Y.C., and Hummer, G. (2011). SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. *Struct. Lond. Engl.* 19, 109–116.
- Rückert, M., and Otting, G. (2000). Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. *J. Am. Chem. Soc.* 122, 7793–7797.
- Salmon, L., Bouvignies, G., Markwick, P., Lakomek, N., Showalter, S., Li, D.-W., Walter, K., Griesinger, C., Brüschweiler, R., and Blackledge, M. (2009). Protein conformational flexibility from structure-free analysis of NMR dipolar couplings: quantitative and absolute determination of backbone motion in ubiquitin. *Angew. Chem. Int. Ed Engl.* 48, 4154–4157.
- Salmon, L., Nodet, G., Ozenne, V., Yin, G., Jensen, M.R., Zweckstetter, M., and Blackledge, M. (2010). NMR characterization of long-range order in intrinsically disordered proteins. *J. Am. Chem. Soc.* 132, 8407–8418.
- Salmon, L., Bouvignies, G., Markwick, P., and Blackledge, M. (2011). Nuclear magnetic resonance provides a quantitative description of protein conformational flexibility on physiologically important time scales. *Biochemistry (Mosc.)* 50, 2735–2747.
- De Sanctis, D., Beteva, A., Caserotto, H., Dobias, F., Gabadinho, J., Giraud, T., Gobbo, A., Guijarro, M., Lentini, M., Lavault, B., et al. (2012). ID29: a high-intensity highly automated ESRF beamline for macromolecular crystallography experiments exploiting anomalous scattering. *J. Synchrotron Radiat.* 19, 455–461.
- Sass, H.J., Musco, G., Stahl, S.J., Wingfield, P.T., and Grzesiek, S. (2000). Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *J. Biomol. NMR* 18, 303–309.

10. Bibliography

Satake, T., Otsuki, K., Banba, Y., Suenaga, J., Hirano, H., Yamanaka, Y., Ohno, S., and Hirai, S. (2013). The interaction of Kinesin-1 with its adaptor protein JIP1 can be regulated via proteins binding to the JIP1-PTB domain. *BMC Cell Biol.* 14, 12.

Schwalbe, M., Ozenne, V., Bibow, S., Jaremko, M., Jaremko, L., Gajda, M., Jensen, M.R., Biernat, J., Becker, S., Mandelkow, E., et al. (2014). Predictive atomic resolution descriptions of intrinsically disordered hTau40 and α -synuclein in solution from NMR and small angle scattering. *Struct. Lond. Engl.* 22, 238–249.

Schwarzinger, S., Kroon, G.J., Foss, T.R., Wright, P.E., and Dyson, H.J. (2000). Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. *J. Biomol. NMR* 18, 43–48.

Schwarzinger, S., Kroon, G.J., Foss, T.R., Chung, J., Wright, P.E., and Dyson, H.J. (2001). Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc.* 123, 2970–2978.

Schweitzer-Stenner, R. (2012). Conformational propensities and residual structures in unfolded peptides and proteins. *Mol. Biosyst.* 8, 122–133.

Serrano, L. (1995). Comparison between the ϕ Distribution of the Amino Acids in the Protein Database and NMR Data Indicates that Amino Acids have Various ϕ Propensities in the Random Coil Conformation. *J. Mol. Biol.* 254, 322–333.

Sharrocks, A.D., Yang, S.H., and Galanis, A. (2000). Docking domains and substrate-specificity determination for MAP kinases. *Trends Biochem. Sci.* 25, 448–453.

Shen, Y., and Bax, A. (2007). Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. *J. Biomol. NMR* 38, 289–302.

Shen, Y., and Bax, A. (2012). Identification of helix capping and b-turn motifs from NMR chemical shifts. *J. Biomol. NMR* 52, 211–232.

Shi, Z., Chen, K., Liu, Z., and Kallenbach, N.R. (2006). Conformation of the backbone in unfolded proteins. *Chem. Rev.* 106, 1877–1897.

Shortle, D., and Ackerman, M.S. (2001). Persistence of Native-Like Topology in a Denatured Protein in 8 M Urea. *Science* 293, 487–489.

De Simone, A., Cavalli, A., Hsu, S.-T.D., Vranken, W., and Vendruscolo, M. (2009). Accurate random coil chemical shifts from an analysis of loop regions in native states of proteins. *J. Am. Chem. Soc.* 131, 16332–16333.

Smith, L.J., Bolin, K.A., Schwalbe, H., MacArthur, M.W., Thornton, J.M., and Dobson, C.M. (1996). Analysis of Main Chain Torsion Angles in Proteins: Prediction of NMR Coupling Constants for Native and Random Coil Conformations. *J. Mol. Biol.* 255, 494–506.

Spivack, J.G., Erikson, R.L., and Maller, J.L. (1984). Microinjection of pp60v-src into *Xenopus* oocytes increases phosphorylation of ribosomal protein S6 and accelerates the rate of progesterone-induced meiotic maturation. *Mol. Cell. Biol.* 4, 1631–1634.

- Stockinger, W., Brandes, C., Fasching, D., Hermann, M., Gotthardt, M., Herz, J., Schneider, W.J., and Nimpf, J. (2000). The reelin receptor ApoER2 recruits JNK-interacting proteins-1 and -2. *J. Biol. Chem.* *275*, 25625–25632.
- Sugita, Y., and Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* *314*, 141–151.
- Takekawa, M., Tatebayashi, K., and Saito, H. (2005). Conserved Docking Site Is Essential for Activation of Mammalian MAP Kinase Kinases by Specific MAP Kinase Kinase Kinases. *Mol. Cell* *18*, 295–306.
- Tamiola, K., and Mulder, F.A.A. (2012). Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins. *Biochem. Soc. Trans.* *40*, 1014–1020.
- Tamiola, K., Acar, B., and Mulder, F.A.A. (2010). Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J. Am. Chem. Soc.* *132*, 18000–18003.
- Tang, X., Orlicky, S., Mittag, T., Csizmok, V., Pawson, T., Forman-Kay, J.D., Sicheri, F., and Tyers, M. (2012). Composite low affinity interactions dictate recognition of the cyclin-dependent kinase inhibitor Sic1 by the SCFCdc4 ubiquitin ligase. *Proc. Natl. Acad. Sci. U. S. A.* *109*, 3287–3292.
- Tanoue, T., Adachi, M., Moriguchi, T., and Nishida, E. (2000). A conserved docking motif in MAP kinases common to substrates, activators and regulators. *Nat. Cell Biol.* *2*, 110–116.
- Taylor, S.S., and Kornev, A.P. (2011). Protein Kinases: Evolution of Dynamic Regulatory Proteins. *Trends Biochem. Sci.* *36*, 65–77.
- Tjandra, N., and Bax, A. (1997). Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* *278*, 1111–1114.
- Tokunaga, Y., Takeuchi, K., Takahashi, H., and Shimada, I. (2014). Allosteric enhancement of MAP kinase p38 α 's activity and substrate selectivity by docking interactions. *Nat. Struct. Mol. Biol.* *21*, 704–711.
- Tolman, J.R., Flanagan, J.M., Kennedy, M.A., and Prestegard, J.H. (1995). Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proc. Natl. Acad. Sci.* *92*, 9279–9283.
- Tompa, P. (2002). Intrinsically unstructured proteins. *Trends Biochem. Sci.* *27*, 527–533.
- Tompa, P. (2012). Intrinsically disordered proteins: a 10-year recap. *Trends Biochem. Sci.* *37*, 509–516.
- Tompa, P., and Fuxreiter, M. (2008). Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. *Trends Biochem. Sci.* *33*, 2–8.
- Vacic, V., Oldfield, C.J., Mohan, A., Radivojac, P., Cortese, M.S., Uversky, V.N., and Dunker, A.K. (2007). Characterization of molecular recognition features, MoRFs, and their binding partners. *J. Proteome Res.* *6*, 2351–2366.
- Vaishnav, M., MacFarlane, M., and Dickens, M. (2011). Disassembly of the JIP1/JNK molecular

10. Bibliography

scaffold by caspase-3-mediated cleavage of JIP1 during apoptosis. *Exp. Cell Res.* *317*, 1028–1039.

Vallurupalli, P., Bouvignies, G., and Kay, L.E. (2012). Studying “Invisible” Excited Protein States in Slow Exchange with a Major State Conformation. *J. Am. Chem. Soc.* *134*, 8148–8161.

Voter, A.F. (1997). Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events. *Phys. Rev. Lett.* *78*, 3908–3911.

Vuister, G.W., and Bax, A. (1993). Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNH.alpha.) coupling constants in ¹⁵N-enriched proteins. *J. Am. Chem. Soc.* *115*, 7772–7777.

Wang, Y., and Jardetzky, O. (2002a). Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci. Publ. Protein Soc.* *11*, 852–861.

Wang, Y., and Jardetzky, O. (2002b). Investigation of the neighboring residue effects on protein chemical shifts. *J. Am. Chem. Soc.* *124*, 14075–14084.

Wang, C.-C., Chen, J.-H., Lai, W.-C., and Chuang, W.-J. (2007). 2DCSi: identification of protein secondary structure and redox state using 2D cluster analysis of NMR chemical shifts. *J. Biomol. NMR* *38*, 57–63.

Wang, Z., Harkins, P.C., Ulevitch, R.J., Han, J., Cobb, M.H., and Goldsmith, E.J. (1997). The structure of mitogen-activated protein kinase p38 at 2.1-Å resolution. *Proc. Natl. Acad. Sci. U. S. A.* *94*, 2327–2332.

Wasserman, T., Katsenelson, K., Daniliuc, S., Hasin, T., Choder, M., and Aronheim, A. (2010). A Novel c-Jun N-terminal Kinase (JNK)-binding Protein WDR62 Is Recruited to Stress Granules and Mediates a Nonclassical JNK Activation. *Mol. Biol. Cell* *21*, 117–130.

Weatheritt, R.J., and Gibson, T.J. (2012). Linear motifs: lost in (pre)translation. *Trends Biochem. Sci.* *37*, 333–341.

Wells, M., Tidow, H., Rutherford, T.J., Markwick, P., Jensen, M.R., Mylonas, E., Svergun, D.I., Blackledge, M., and Fersht, A.R. (2008). Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 5762–5767.

Whisenant, T.C., Ho, D.T., Benz, R.W., Rogers, J.S., Kaake, R.M., Gordon, E.A., Huang, L., Baldi, P., and Bardwell, L. (2010). Computational prediction and experimental verification of new MAP kinase docking sites and substrates including Gli transcription factors. *PLoS Comput. Biol.* *6*.

Whitmarsh, A.J. (2006). The JIP family of MAPK scaffold proteins. *Biochem. Soc. Trans.* *34*, 828–832.

Whitmarsh, A.J., Cavanagh, J., Tournier, C., Yasuda, J., and Davis, R.J. (1998). A Mammalian Scaffold Complex That Selectively Mediates MAP Kinase Activation. *Science* *281*, 1671–1674.

Whitty, A. (2008). Cooperativity and biological complexity. *Nat. Chem. Biol.* *4*, 435–439.

Wishart, D.S. (2011). Interpreting protein chemical shift data. *Prog. Nucl. Magn. Reson. Spectrosc.* *58*, 62–87.

- Wishart, D.S., and Sykes, B.D. (1994). The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR* 4, 171–180.
- Wishart, D.S., Sykes, B.D., and Richards, F.M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry (Mosc.)* 31, 1647–1651.
- Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L., and Sykes, B.D. (1995a). ^1H , ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* 6, 135–140.
- Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S., and Sykes, B.D. (1995b). ^1H , ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J. Biomol. NMR* 5, 67–81.
- Won, A.P., Garbarino, J.E., and Lim, W.A. (2011). Recruitment interactions can override catalytic interactions in determining the functional identity of a protein kinase. *Proc. Natl. Acad. Sci.* 108, 9809–9814.
- Wright, P.E., and Dyson, H.J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* 293, 321–331.
- Xie, H., Vucetic, S., Iakoucheva, L.M., Oldfield, C.J., Dunker, A.K., Obradovic, Z., and Uversky, V.N. (2007). Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. *J. Proteome Res.* 6, 1917–1932.
- Xie, J., Onnockx, S., Vandenbroere, I., Degraef, C., Erneux, C., and Pirson, I. (2008). The docking properties of SHIP2 influence both JIP1 tyrosine phosphorylation and JNK activity. *Cell. Signal.* 20, 1432–1441.
- Xie, X., Gu, Y., Fox, T., Coll, J.T., Fleming, M.A., Markland, W., Caron, P.R., Wilson, K.P., and Su, M.S. (1998). Crystal structure of JNK3: a kinase implicated in neuronal apoptosis. *Struct. Lond. Engl.* 1993 6, 983–991.
- Xu, Z., Kukekov, N.V., and Greene, L.A. (2003). POSH acts as a scaffold for a multiprotein complex that mediates JNK activation in apoptosis. *EMBO J.* 22, 252–261.
- Xue, B., Dunbrack, R.L., Williams, R.W., Dunker, A.K., and Uversky, V.N. (2010). PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta* 1804, 996–1010.
- Yang, S., Blachowicz, L., Makowski, L., and Roux, B. (2010). Multidomain assembled states of Hck tyrosine kinase in solution. *Proc. Natl. Acad. Sci. U. S. A.* 107, 15757–15762.
- Zarrinpar, A., Park, S.-H., and Lim, W.A. (2003). Optimization of specificity in a cellular protein interaction network by negative selection. *Nature* 426, 676–680.
- Zhang, F., Strand, A., Robbins, D., Cobb, M.H., and Goldsmith, E.J. (1994). Atomic structure of the MAP kinase ERK2 at 2.3 Å resolution. *Nature* 367, 704–711.
- Zhang, H., Neal, S., and Wishart, D.S. (2003). RefDB: a database of uniformly referenced protein

10. Bibliography

chemical shifts. *J. Biomol. NMR* 25, 173–195.

Zhou, H., Zheng, M., Chen, J., Xie, C., Kolatkar, A.R., Zarubin, T., Ye, Z., Akella, R., Lin, S., Goldsmith, E.J., et al. (2006a). Determinants That Control the Specific Interactions between TAB1 and p38? *Mol. Cell. Biol.* 26, 3824–3834.

Zhou, T., Sun, L., Humphreys, J., and Goldsmith, E.J. (2006b). Docking interactions induce exposure of activation loop in the MAP kinase ERK2. *Struct. Lond. Engl.* 1993 14, 1011–1019.

Zweckstetter, M. (2008). NMR: prediction of molecular alignment from structure using the PALES software. *Nat. Protoc.* 3, 679–690.

Zweckstetter, M., and Bax, A. (2000). Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. *J. Am. Chem. Soc.* 122, 3791–3792.

Zweckstetter, M., Hummer, G., and Bax, A. (2004). Prediction of charge-induced molecular alignment of biomolecules dissolved in dilute liquid-crystalline phases. *Biophys. J.* 86, 3444–3460.