Structural insights into Arginine-Serine rich proteins and N-H spin-spin coupling constants

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

"Doctor rerum naturalium"

der Georg-August-Universität Göttingen

vorgelegt von

Shengqi Xiang

aus Wuhu, China

Göttingen, 2012

Mitglied des Betreuungsausschusses (Referent): Prof. Dr. Markus Zweckstetter
NMR-basierte Strukturbiologie, Max-Planck-Institut für biophysikalische Chemie

Mitglied des Betreuungsausschusses: Prof. Dr. Reinhard Lührmann
Abteilung Zelluläre Biochemie, Max-Planck-Institut für biophysikalische Chemie

Mitglied des Betreuungsausschusses: Prof. Dr. Ralf Ficner
Abteilung für Molekulare Strukturbiologie, Georg-August-Universität Göttingen

Tag der mündlichen Prüfung:

Affidavit

I hereby declare that this dissertation has been written independently and with no other sources and aids than quoted.

.....

Shengqi Xiang

Abstract

This thesis is constituted of two parts. In the first part of this thesis, I studied the structural phenomena of SR protein phosphorylation. The serine/arginine-rich (SR) protein family plays multiple functions in the whole process of RNA metabolism such as transcription, RNA splicing, RNA exporting, translation, and nonsense decay of RNA. SR proteins contain at least one essential arginine/serine-rich (RS) domain, from which the protein name is derived. The RS domain mediates both protein-protein and protein-RNA interactions. A number of studies have shown that the serine residues within the RS domains of SR proteins are extensively phosphorylated. The majority of phosphorylation occurs on serine residues in the RS domain. This phosphorylation appears to influence interactions and the subcellular localization of SR proteins, thereby modulating their functions. In spite of the versatile and vital functions of SR proteins, the structure of the whole protein remains unknown so far.

In this study, we use the combined approaches of NMR and MD simulation to study the structures of wild type and phosphorylated RS dipeptide repeat peptides which are derived from the prototypical SR protein ASF/SF2 (200-219). In wild type form, the RS repeat peptide is disordered, which is revealed by its negative heteronuclear NOE values and highly degenerate spectra. Upon phosphorylation, the RS repeat peptide gets structured but is still not as fully rigid as folded proteins. After unbiased MD simulations of both, a sub-ensemble selection procedure was carried out to obtain a representative structural ensemble of the native and phosphorylated ASF/SF2 (200-219) peptides, which constrains the simulated ensembles to reproduce experimental NMR data. The phosphorylated peptide adopts an “arch-like” structure, which get structured at both backbone and side chain levels. The structural changes also show pH dependency, which is related to the charges which require -2 charges on the side chain of serine residues. Hence RE or RD repeats cannot reproduce the same phenomena as phosphorylated RS repeats.

NMR data of the intact full length RS domains in ASF/SF2 or hPrp28 showed that the phosphorylated RS repeats in domain also undergoes a disorder-to-order transition upon phosphorylation, which is similar to the isolated RS peptide. This similarity suggests the

phosphorylated RS peptide is a general model of structure transitions upon phosphorylation for all SR and SR-related proteins.

In the second part of this thesis, I studied the relationship between N-H spin-spin couplings and hydrogen bonds. Hydrogen bonds are essential for the structure of many biochemical compounds. Protein folding, the formation of amyloid aggregates, enzymatic catalysis, drug-receptor interactions, and many other phenomena are intrinsically connected to hydrogen bonding. It has been theoretically predicted that $^1J_{NH}$ becomes more negative upon hydrogen bond formation. In spite of the high accuracy of $^1J_{NH}$ measurement, $^1J_{NH}$ values have not been used for the detection of hydrogen bonds.

In this study, I first measured large numbers of $^1J_{NH}$ values in disordered proteins to serve as a reference. The residue-specific mean values of $^1J_{NH}$ were used as random coil values. The influence of pH and temperature were also checked by systematically measuring the couplings in various conditions. Comparing the random coil values to $^1J_{NH}$ in folded proteins, it was demonstrated that the magnitude of $^1J_{NH}$ was increased by up to 1.6Hz due to hydrogen bond formation. Thus the H-N spin coupling can be a sensitive tool for the detection of hydrogen bonds. The results also provide a basis for further investigating the relation between hydrogen bonds and $^1J_{NH}$ values using theoretical calculations.

Acknowledgements

The research work of this thesis was carried out at the Max-Planck Institute for Biophysical Chemistry, Göttingen, Germany under supervision of **Prof. Dr. Markus Zweckstetter**.

My deepest gratitude goes first and foremost to **Prof. Dr. Markus Zweckstetter**, my supervisor, for his constant guidance, encouragement and the opportunity to work on exciting projects during my PhD.

I want to thank **Prof. Dr. Reinhard Lührmann and Prof. Dr. Ralf Ficner** for agreeing to be members of my thesis committee and giving great help to my projects.

I also want to express my gratitude to the Director of the department, **Prof. Dr. Christian Griesinger,** who has provided an excellent scientific environment with state-of-art equipment for undertaking research.

I owe a special debt of thanks to **Prof. Bert de Groot** and **Vytautas Gapsys** for the successful collaboration on the SR project. Without their splendid work, the SR project could not have been accomplished. I also want to thank **Raghavendran Lakshmi Narayanan** who collaborated with me on the J coupling project and gave me help with the MARS software.

For all the biochemical work in the wet lab, I sincerely thank **Dr. Stefan Becker, Karin Giller**, **Sebastian Wolf**, **Melanie Wegstroth, Ann-Kathrin Brueckner, Laukat Yvonne.** I am very grateful to **Gerhard Wolf** for his manipulation of HPLC and mass spectroscopy. I thank **Volker Klaukien** and **Kerstin Overkamp** for synthesizing peptides in the work. I would like to express my sincere appreciation to **Mrs. Siberer, Mrs Breiner**, **Dr. Dirk Bocklmann** and **Jürgen Arve** for their daily support.

I would like to thank **Dr. Min-Kyu Cho** and **Dr. Hai-Young Kim** not only for their work on the research project, but also for sharing the exciting or frustrating moments with me throughout my PhD and for giving me great help and support since I arrived in Germany. **Dr. Ulirich Dürr** gave me great help on learning NMR spectroscopy. **Dr. Nasrollah Rezaei-Ghaleh** helped me with biophysical analysis, and particularly, gave

the key suggestion for the SR project. I shared lots of fun with **Min-**Kyn, **Ulirich, Han, Stefan, Saskia, Elias and Bejamin** when we were "attempting" to make our NMRs work. I appreciate all the help from my present and former colleagues in NMR2 department: **Mitcheell, Francesca, Stefan, Edward, Nils, Elias, Han, Guowei, Xuejun, Donghan, Hari, Frederik, Michael, David, Davood, Michal, Sebastian, Lukasz, Mariusz, Korvin, Holger, Vinesh.**

I want to thank **Chiritina Bach, Christin Fischer** and other persons in the GGNB team for their kind help and excellent organization.

I am grateful to the joint PhD Program between the Max-Planck Society and the China Academy of Science, which gave me the chance to study at the Max-Planck Institute.

Prof. Yunyu Shi and Prof. Jihui Wu must be credited for their instruction and support of my scientific life.

I want to thank Prof. **Rudolf Kippenhahn**, thought I have never met him. His book, "*100 Milliarden Sonnen*" ignited my interest in science when I was in junior middle school.

Finally, I want to express my heartfelt thanks to my family. I greatly appreciate my parents for their love and support throughout my life. I am deeply grateful to **Lu Zhang**, for her unconditional love and tireless support.

Table of Contents

Abbreviations 1

Part I Structural consequence of RS domain phosphorylation 3

1. Introduction 3

1.1 Intrinsic disorder protein (IDP) 3

1.1.1 Structure and Function of Intrinsic disordered protein (IDP) 3

1.1.2 Structure-Function relation in IDP 5

1.2 Ensemble representations of intrinsically disordered protein 7

1.3 Phosphorylation induced structure changes 10

1.4 SR protein and RS domain 12

1.4.1 The SR protein 12

1.4.2 SR protein functions 18

1.4.3 SR protein in mRNA splicing 19

1.4.4 Function redundancy of SR proteins 23

1.4.5 SR proteins and diseases 23

1.4.6 RS domains in SR proteins 25

1.4.7 Phosphorylation of SR proteins 26

1.4.8 Kinases phosphorylate RS domain 27

1.5 Motivation of this project 31

2. Materials and Methods 33

2.1 Materials 33

2.1.1 Chemicals and Enzymes 33

2.1.2 Bacterial Strains 34

2.1.3 Protein sequence 35

2.1.4 Oligonucleotide primers 35

2.1.5 Equipment 35

2.1.6 Software 37

2.2 Methods 38

2.2.1 Molecular biology methods 38

2.2.1.1 Purification of Plasmid 38

2.2.1.2 PCR amplify 38

2.2.1.3 DNA sequencing 39

2.2.1.4 Double digestion 39

2.2.1.5 Ligation 40

2.2.1.6 Transformation of E.coli 40

2.2.2 Protein Biochemistry Methods 40

2.2.2.1 Expression of RS domain of hPrp28 (DDX23) 40

2.2.2.2 Purification of RS domain of human Prp28 42

2.2.2.3 Expression and purification of SRPK1 42

2.2.2.4 Expression and purification of RS domain of ASF/SF2 43

2.2.2.5 Expression and purification of RS peptide 44

2.2.3 NMR methods 44

2.2.3.1 HSQC (Heteronuclear Single-Quantum Coherence) 44

2.2.3.2 Triple resonance experiments 45

2.2.3.3 Automated projection spectroscopy (APSY) 48

2.2.3.4 Experiments of Coupling measurements 50

2.2.3.5 RDC measurements 51

2.2.3.6 Heteronuclear NOE measurements 52

3. Result 53

3.1 Phosphorylated ASF/SF2 RS domain adopts structure. 53

3.2 Phosphorylation induced structural transition in ASF/SF2 (200-219) peptide 56

3.3 Unphosphorylated ASF/SF2 (200-219) is in random coil status 57

3.4 Structural transition of ASF/SF2 (200-219) upon phosphorylation 64

3.5 Structural ensembles representation of RS repeats peptide 71

3.5.1 Structural ensembles from Molecular Dynamics (MD) simulations 71

3.5.2 Validation of the structural ensembles 76

3.5.3 Backbone conformations 81

3.5.4 Orientations of the side-chains 85

3.6 SR-related proteins undergoing similar structural changes upon phosphorylation as SR proteins. 87

4. Discussion 94

4.1 Structures transition of ASF/SF2 (200-219) peptide upon phosphorylation 94

4.2 Discrepancy with previous works 97

4.3 Orientations of side chains 99

4.4 pH dependency of RS repeats peptide conformation 100

4.5 Universality of RS model peptides 101

4.6 Function redundancy of SR proteins 102

4.7 Binding regulated by phosphorylation 103

4.8 Different kinase specificity 105

4.9 Outlook 109

Part II Probing Hydrogen bonds networks with N-H couplings 111

5. Introduction 111

5.1 Indirect spin-spin coupling 111

5.2 Hydrogen bond 113

5.3 Field dependence of scalar coupling values 115

5.4 The methods to measure scalar couplings 116

5.4.1 Splitting method 116

5.4.2 Intensity modulation method 119

5.5 Motivation of the project 122

6. Methods and Materials 123

6.1 NMR samples 123

6.2 Band-selective-decoupled (BSD) IPAP-HSQC 123

6.3 Intensity modulated HSQC 126

7. Result 129

7.1 The residues-specific $^{1}J_{HN}$ couplings in intrinsically disordered proteins 129

7.2 The $^{1}J_{NH}$ scalar couplings in intrinsically disordered proteins are stable at different temperatures and pH 133

7.3 Field strength dependences of $^{1}J_{NH}$ couplings are negligible for IDP proteins. 136

7.4 Systematic offset between intensity modulated HSQC and BSD-IPAP-HSQC 137

7.5 Negative secondary 1J HN coupling values indicate H-bond 138

7.6 Identify hydrogen bonds by secondary J couplings 143

7.7 Experimental uncertainty 145

8. Discussion 148

8.1 $^{1}J_{NH}$ couplings of intrinsic disordered proteins 148

8.2 The relation between $^{1}J_{NH}$ coupling and hydrogen bonds 150

8.3 Comparison with other Methods to identify hydrogen bond 152

8.4 Outlook 154

9. Bibliography 156

10. Appendix 177

10.1 Pulse program used to measure Hα-Cα. HACANNH 177

10.2 TROSY-HSQC interleaved experiment for measure N-H couplings 184

10.3 HNN pulse program 190

10.4 BSD-IPAP-HSQC pulse program 196

10.5 Intensity modulated HSQC for measuring $^{1}J_{NH}$ 201

10.6 Python scripts for fitting $^{1}J_{NH}$ values 204

10.7 RDC values and Chemical shifts of native ASF/SF2 (200-219) 207

10.8 3bond couplings of native ASF/SF2 (200-219) 208

10.9 RDC values of phosphorylated ASF/SF2 (200-219) 209

10.10 3bond couplings of phosphorylated ASF/SF2 (200-219) 210

10.11 RDC values for H-CO and N-CO from 2 measurements on the same phosphorylated ASF/SF2 (200-219) samples 211

10.12 $^{1}J_{NH}$ values of Tau protein pH 6.0, 278K, at 900MHz 212

10.13 $^{1}J_{NH}$ values of α-synclein at different pH and temperatures, measured at 700MHz. 215

10.14 $^{1}J_{NH}$ values of native ubiquitin at 400MHz, 600MHz and 900MHz field strength, pH 7.0, 298K 218

Publication list 221

Curriculum Vitae 222

Abbreviations

$^{1}D_{C\alpha\text{-}CO}$	Cα-CO residual dipolar coupling
$^{1}D_{C\alpha\text{-}H\alpha}$	Cα-Hα residual dipolar coupling
$^{1}D_{NH}$	N-H residual dipolar coupling
$^{1}J_{NH}$	One bond N-H coupling
$^{3}J_{HNH\alpha}$	Three bond HN-Hα coupling
APSY	Automated projection spectroscopy
ATP	Adenosine-5'-Triphosphate
BSD-IPAP-HSQC	Band selection decoupled IPAP-HSQC
CD	Circular dichroism
CSA	Chemical shift anisotropy
DDX23	DEAD box protein 23
DNA	Deoxyribonucleic acid
DTT	Dithiotreitol
E.coli	Escherichia coli
EDTA	Ethylenediamine tetraacetic acid
HPLC	High performance liquid chromatography
hPrp28	human Prp28
HSQC	Heteronuclear single quantum coherence
im-HSQC	Intensity modulated HSQC
INEPT	Insensitive nuclei enhanced by polarization transfer
IDP	Intrinsic disordered protein
IPAP	In-phase-anti-phase
IPTG	Isopropyl-β-D-thiogalactopyranoside
kD	kilo-Dalton
ml	Milliliter
ms	Milliseconds
MS	Mass spectrometry

NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
ns	Nanoseconds
PCR	Polymerase chain reaction
PDB	Protein data bank
PRE	Paramagnetic relaxation enhancement
RDC	Residual dipolar coupling
RpS	Argine-Phosphoserine dipeptide
RS	Argine-Serine dipeptide
SAXS	Small angle X-ray scattering
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SFRS	Splicing factor, arginine-serine-rich
SR	Serine- Argine dipeptide
SRPK	SR protein kinase
TROSY	Transverse relaxation optimized spectroscopy
μl	Microliter
μs	Microseconds
τ_c	Correlation time for molecular tumbling

Part I Structural consequence of RS domain phosphorylation

1. Introduction

1.1 Intrinsic disorder protein (IDP)

1.1.1 Structure and Function of Intrinsic disordered protein (IDP)

Intrinsic disorder protein is the protein that carries out biological functions but lack a single, well-defined 3-D structure in physiological conditions. Compared to "regular" globular folded proteins, the IDPs exhibit distinct properties, such as resistance to heat and chemical denaturation (1), unusual mobility in SDS pages (2), enhanced proteolytic sensitivity (3), much bigger hydrodynamic radii (4), lack of secondary structure in circular dichroism, and can never be crystallized or missing from X-ray structures.

The fundamental difference between IDP and regular folded proteins is seen on energy landscape topography (5). The topography of globular folded protein has a deep and global minimum, which corresponds to the folded state. Because of the steep energy barriers around, folded proteins reside in this minimum and fluctuate nearby. The energy landscapes of IDP are relatively flat, which do not obtain a global minimum but have many shallow local minima separated by low energy barriers. Thus, the thermal fluctuation enables IDP to sample a relatively large area in conformation space and fluctuate between a diverse set of conformations.

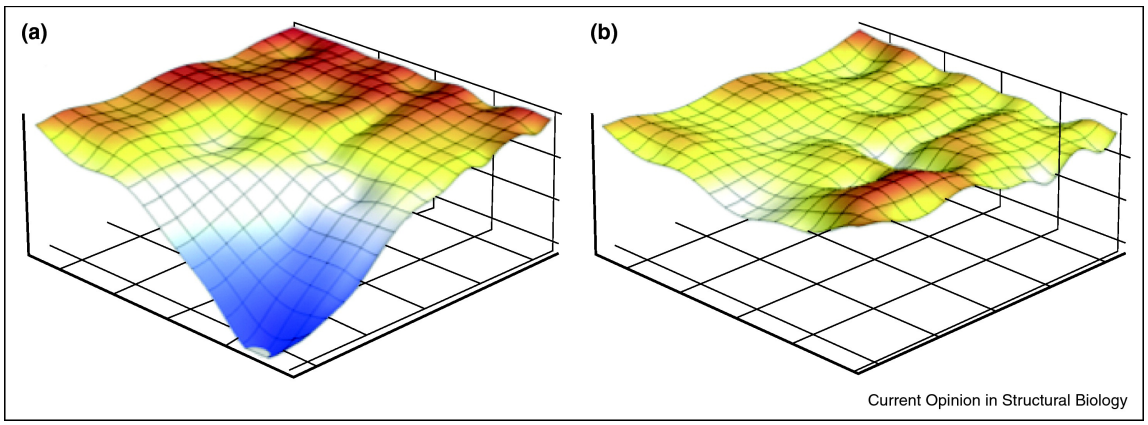


Figure 1.1 The landscape of folded protein (a) and intrinsically disordered protein (b). The folded protein has a defined minimum energy state, while the IDP lacks a deep minimum.(reproduced from (5) with permission)

The primary structure of IDP is different in amino acid composition from folded proteins (6). Certain residues are enriched in IDPs or intrinsic disordered regions (IDR), such as Ala, Arg, Gly, Gln, Ser, Pro, Glu and Lys, whereas some other residues are depleted, such as Trp, Cys, Phe, Ile, Tyr, Val and Leu. The former group has a high flexibility index while the latter ones have low flexibility indices. As a result, the IDPs show increased in net charges and low overall hydrophilia (7).

Although IDP lacks a well-defined 3D structure, it does not mean that they are featureless random coils. Even the highly denatured globular proteins caused by chemical agents are not true random coils. It has been apparent that the IDPs have a transit local order. A significant level of secondary structure elements are identified in IDP by CD, Fourier-transform infrared spectroscopy, and especially NMR methods (8-10). These secondary structure elements in unbound states often adopt similar structure in bound states, which may indicate the recognition segments already partially preformed in free form (11).

On higher structure levels, IDP does not possess a unique tertiary structure. However, there are global parameters that can describe the overall size and shape of the protein, such as the radius of gyration and the hydrodynamic radius. Small angle X-ray scattering and pulsed field gradient NMR studies of IDP both show that the size of IDP molecule is

bigger than the globular folded protein with the same residues number, but is still smaller than the estimated values for a random coil protein of the same size (4, 12). In the same line with this "partially compacted" state of IDP, there are long range contacts identified in IDP by NMR and MD methods (13, 14).

The biological functions of IDP are searched by searching for the key words of IDP functions in databases that show statistically significant correlation (15). The functions most preferential for IDP are in large function areas: 1.transcription and transcription regulation 2.signal transduction and regulation of cell cycle 3.the biogenesis and functioning of nucleic acid containing organelles 4.messenger ribonucleic acid (mRNA) processing 5.the organization and biogenesis of cytoskeleton.

As IDPs are deeply involved in variety of functions, their dysfunctions lead to diverse pathological results (16). Some key proteins in cancers and neurodegenerative diseases, such as p53, α-fetoprotein, α-synclein, Prion and Tau, are IDPs or have long disordered regions. In fact, IDPs are enriched in proteins associated with cancer, cardiovascular disease and diabetes. In neurodegenerative diseases, such as Alzheimer's disease, Parkinson's disease and Huntington Disease, IDPs aggregate to form oligomers and further to form Amyloid fibrils, which is the key process of disease development. This correlation makes IDPs as novel and interesting drug targets.

1.1.2 Structure-Function relation in IDP

The structural disorder property of IDP and IDR enable some unique functions, which include functions as a spacer or linker, entropic brushes, entropic springs, and entropic clocks (17, 18). The space and linker are regions that connect the functional domains or motifs. Their functions are to offer appropriate space separations between motifs, which allow the motifs freely interact with each binding partners. The entropic functions of IDP are due to the entropy penalty when the external forces reduce the conformation freedom (19, 20). To carry out these functions, the disorder to order transition of IDPs is not necessary.

IDPs most often involve molecular recognition. Normally, they undergo induced folding upon binding (21, 22). Unlike folded proteins, the IDP tends to use short and

flexible motifs in binding. These short fragments often have the tendency of the bound form in their free states. Sometimes, two IDPs can bind together and undergo mutual induced folding (23). However, the exact mechanism of folding upon binding is still quite obscure. The folding can happen before, concurrently and after the binding event. The disordered nature of IDP uncouples the specificity and binding strength by the decreasing in conformation entropy upon binding. The entropic penalty reduces the free energy comes from the binding of protein and ligands. By modulating the entropy changes in reaction, it is possible to modulate the proteins interactions and hence regulate the protein functions. In addition, segments of IDP could retain their disordered state in the complex, which is termed as fuzziness(24).

The recognition interfaces of IDPs are larger than folded proteins if normalized by the chain length (25, 26). The interfaces of IDPs also make more hydrophobic-hydrophobic contacts, while the folded proteins use more polar-polar interactions. It is suggested that the folded protein tends to stabilize themselves by the interaction within the same polypeptide chain, whereas the IDPs use more interaction with their partners for stabilization.

From kinetic views of interactions, the flexible, fast moving backbone of IDP speed up the reaction by several factors. First, the binding regions of IDPs keep sampling the binding conformation and expose the primary contact sites for initial contact (27). Second, the increased capture ranges of the segments in disordered proteins make the search for partner more effective (28).

The adaptability of IDPs in binding is another unique feature rooted in the structural property of IDP. In the classical structure-function paradigm, one protein or one part of the protein can only interact with one binding partners. However, some segments of IDPs can interact with different binding partners with different conformations (29). It is particularly important for hub proteins, which sits in the intersection point of many signal transduction pathways such as p53 (30).

1.2 Ensemble representations of intrinsically disordered protein

Under physiological conditions, thermal fluctuation causes the proteins sample certain ranges of different conformations, while the probability of each conformation depends on the terrain of the energy landscape. The energy landscape of folded protein has a well-defined global minimum, and the "funnel" terrain around the minimum is steep. This restricts protein conformations in a folded state corresponding to the global minimum and allows only small fluctuations around it. In this case, one single structure corresponding to the minimum is good approximation of protein conformation.

By contrast, the relatively flat energy landscape of intrinsically disorder protein (IDP) lets the protein sample over a much larger range of conformers. Thus one single structure cannot cover all of the protein states. The comprehensive description of such system requires an ensemble, which consists of a group of structures and covers all accessible states of the protein.

In practice, members of ensemble are selected in such way that the ensemble averages of some "observable" parameters agree with the experimental data. These observable parameters include chemical shifts, residue dipolar couplings (RDC), scalar couplings, paramagnetic relaxation enhancement (PRE) via NMR, average radius of gyration and scattering profile from small angle X-ray scattering (SAXS). The conformations in the ensemble are exchanging with each other in a fairly rapid timescale, so the experimental values correspond to the ensemble average over all structure in the ensemble.

Chemical shifts that give information about local conformation tendency can be estimated from protein structure by a number of programs SHIFTS(31), SPATA(32) or SHIFTX(33). RDCs, which report the relative orientation of the vectors toward the external magnetic field, can be predicted from known protein structures by program PALES(34). There are established Karplus equations to correlate scalar couplings with dihedral angles (35-37). PREs, which provide long-range contact information, can be estimated straightforwardly based on atomic coordinates of protein structure. One has to compute the value for every structure, take the average, and then compare to experimental data.

There are two different approaches to build ensembles. First one is termed as "restrained MD simulation". As shown by its name, it introduces extra terms in energy functions to bias the simulation towards the region in conformation space that agree with experimental data. In such MD simulation, multiple replicas are simulated in parallel and the restrains are not applied to single structures but to the entire group. At each step, the ensemble average over all the replicas are calculated and compared to experimental data. The penalty function is added to enforce the match of back-calculated values and experimental data. This method is applied to α-synclein with PRE data as experiment restrains (38, 39).

Another method to generate ensemble is sub-ensemble construction. First, a library of conformations is generated, and then a subset of this library is selected from this library to reproduce the experimental data. The start library can be obtained from MD simulation for the full length protein (40), or by joining conformations from MD simulation of each small peptide segments consisting the whole sequence altogether (41), or by with statistical coil models such as Flexible-Meccano (42).

Once the ensemble is generated, a small subset of the ensemble is chosen to minimize the difference between predicted and experimental values. The algorithm employs Monte Carlo (43), Sample and Select (44), or evolutionary algorithm (42) to search for a set of conformations that best reproduce the experimental data. This approach was also carried on α-synclein to detect long range contacts (45).

There is a slightly different approach that does not select a subset of conformations, but assigns a probabilistic weight to every structure in library (46). However, the two approaches are equal in principle. The selection approaches can be regarded as a particular case for weight assigning method: for the conformations not selected, their weights are 0. For the conformations in the ensemble, all of them have the same weight 1/N, with N being the number of conformations in the selected ensemble. The Energy-minima Mapping and Weighting (EMW) method employs the both approaches, which chooses the structures to generate ensemble and then assigns weights to these structures. The EMW method was carried out for the C-terminal fragments of p21 protein, which reveals pre-existing bound state conformations in the free peptide.

To assess the performance of the algorithm used in generating ensembles, a method termed as "reference ensemble" was introduced as standard tool for evaluation ensemble generate methods (Figure 1.2). The principle of this method is generating a "reference ensemble", which is made up of preset structures and corresponding weights. The "synthetic" experimental data are calculated from the "reference ensemble". The ensemble building algorithm tested is fed with "synthetic" experimental data, and then the outcome is compared to the "reference ensemble". The sign of success is clear when the algorithm can reproduce the reference ensemble solely based on the synthetic experimental data. This method allows checking the performance of the algorithm in an ideal situation without any experiment error. On the other hand, the same method also could be used to monitor the uncertainty in the final ensemble induced by experiment error.

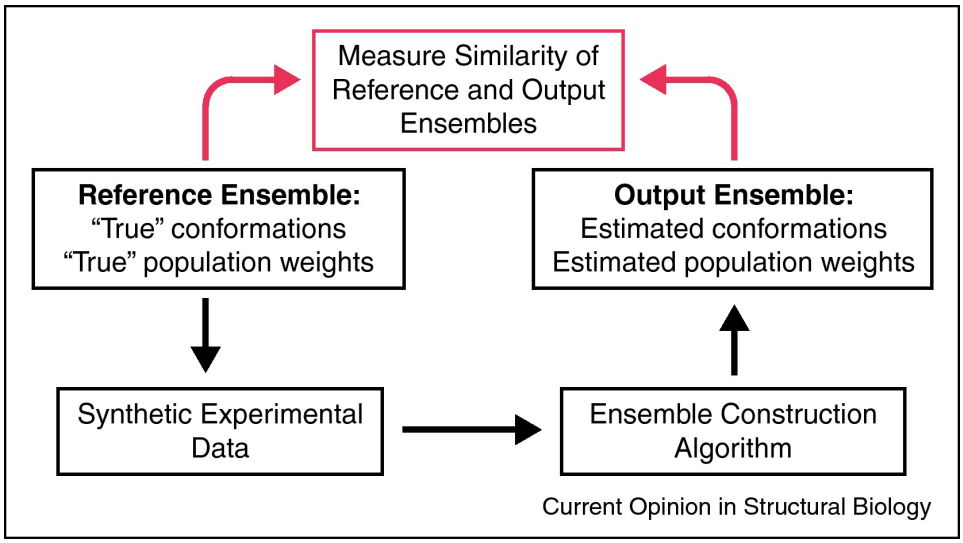


Figure 1.2 A schematic representation of the reference ensemble method for validating ensemble constructing methods (reproduced from (5) with permission)

The two aforementioned approaches of restricted MD simulation and ensemble-selection were tested with this method, which reveal both strengths and disadvantages. Restricted MD simulation of α-synclein using only PRE restrains performs poorly unless large number of restrains are included (39), i.e. more than 4 restrains per residue per replica. However, if the radius of gyration is added as additional information, restricted MD needs much less PRE data; 2 restrain per residue per replica is enough (38). So the

performance of restricted MD depends on both number and type of restrains. With enough experimental data, this method can give accurate model of the protein, although it is difficult to know exactly how many experimental restrains is sufficient.

Sub-ensemble also faces a problem of degeneracy. There are many ensembles can match the experimental data equally well within the experimental error range, but all these ensembles are different from the reference ensemble (41). This means that agreement with experimental data cannot ensure the accuracy of the ensemble. This problem originates from the fact that the experimental data normally is much less than the degrees of freedom, which can uniquely specify the ensemble. Constructing a number of ensembles can alleviate this problem, and then extracting the structural features shared by all these ensembles (47, 48). The characteristics recurring in different ensembles are likely to be accurate. Again, similar to the restrains problem in restricted MD simulation, the number of ensembles needed to identify common structure features is unclear. Moreover, the definition of "common structure features" is ambiguous. An alternative way is to quantitatively estimate the uncertainty of weights of structures in ensemble by Bayesian statistics given the experimental data (41).

In short, all approaches to construct ensemble of IDP have the advantages and limitations of their own. With new developments made to better access and validate the ensembles, the accuracy of IDP structure description will be improved.

1.3 Phosphorylation induced structure changes

Although there are only 20 natural amino acids encoded in genes there are various modifications of amino acid side chains after translation, which are vital for functions of gene products. Phosphorylation is an important post-translation modification in both prokaryotes and eukaryotes that modulates numerous biological functions. One phosphate group was added to side chain hydroxyl group of Ser, Thr or Tyr residues in protein phosphorylation. Phosphorylation is catalyzed by protein kinases, while the reverse reaction is catalyzed by phosphatase.

Gain or loss of the phosphate groups on one protein is able to cause a signal cascade, which results in modifying the enzyme activity or gene expression. Reversible

phosphorylation is one of the most ubiquitous and significant regulatory mechanisms. Aberrant regulation of phosphorylation leads to numerous diseases including several cancers (49).

Genome surveys reveal that up to 30% proteins are phosphorylated *in vivo* and phosphorylation prefers hydrophilic and terminal part along the primary sequence (50). Moreover, sequence analysis shows that disorder-promoting residues are enriched in phosphorylation sites surrounding, which implies that protein phosphorylation predominately occurs within intrinsically disordered regions (51).

Whereas many of thousands of phosphorylation sites are identified, the atomic detail information of structural changes induced by phosphorylation is relatively rare due to experimental difficulties. MD simulation based on known structures plays a valuable role in elucidating structural consequences of phosphorylation (52).

The phosphate group is negatively charged in physiological conditions, which strongly perturbs the electrostatic interaction in proteins. Besides being directly involved in binding contact (53), phosphorylation changes the landscape of a protein, hence alters the conformation of protein (54). Phosphorylation can change the conformation in a variety of ways: from local effects such as stabilizing the structure of the binding groove (55), to inducing allosteric changes distant from the phosphate (56).

Phosphorylation sites can be found on kinase itself, which regulate the activity of kinase. The phosphorylation sites are mainly located in the so-called “activation loop”, which form part of the binding groove for the peptide. In phosphorylated CDK2/cyclin A complex, the pThr160 directly contacts the basic residue 3 residues C- terminal to the phosphorylatable serine on the substrate, which contributes to substrate recognition. In addition, the pThr160 also acts as an organizing center to stabilize the conformation of vicinity by contacting and neutralizing three agrinines with its phosphate group (55). The CDK family is also regulated by phosphorylation on a “glycine rich” loop, which is part of the ATP-binding site. Phosphorylation of two adjacent residues, Thr14 and Tyr15, can inhibit CDK2 activity by misalignment of the ATP (57). In the case of PKA, the phosphorylation of Thr197 in an activation loop not only changes the conformation of

this loop, but also stabilizes the side chain orientation of the serine on the substrate in a position to prime for phosphorylation(58).

The conformation changes are not limited to the region in the vicinity. In one Hck kinase, four distinct conformational motifs are coupled to each other via the interaction limited to the nearest neighbor. Phosphorylation on Tyr416 in the flexible, solvent exposed part of the activation loop causes a shift in the conformation equilibrium of this part. This shift induces some change of contact with the other part of the activation loop, which further changes the rotation of the α-C helix. Specific nearest-neighbor interactions between the α-C helix and N-terminal end hinge correlate the motion of N-terminal hinges with the rotation of the α-C helix. Therefore, effects of Tyr416 phosphorylation are transmitted to regulatory SH2 and SH3 domains (56).

In some cases, more dramatic conformation changes are induced, both order to disorder (59) or disorder to order transitions(60, 61). Phosphorylation of S16 in monomeric phospholamban causes melting of local structure, and further changes the motion of the whole protein (59). Electron paramagnetic resonance (EPR) and Molecular dynamics (MD) studies on the regulatory light chain of smooth muscle myosin reveal a helix structure is induced by phosphorylation in the N-terminal region (60, 61). Phosphorylation at Ser-133 of the kinase inducible domain of CREB (KID) triggers a coil to helix transition and reduces its flexibility, which facilities its binding to the KIX domain of CBP(62). Phosphorylation of Ser14 on glycogen phosphorylase induces a distorted 3_{10} helix at the N-terminal and binds to the surface of the dimer. This structural change strengthens subunit interactions and transforms the enzyme to an active form (63).

1.4 SR protein and RS domain

1.4.1 The SR protein

The SR protein was first identified as splicing factors in Drosophila in the early 1990s (64). The term “SR” refers to high-serine and Argine contents in a protein sequence. The first identified three SR proteins all have one domain rich in Arg-Ser repeats termed RS domain, which becomes the hallmark of SR proteins. Later, in humans, two other SR

proteins, ASF/SF2(65) and SC35(66) are identified, which also obtain RRM domains besides the RS domain.

Member of SR proteins family was identified based on following criteria: the presence of a phosphoepitope recognized by the monoclonal antibody mAb104, complement splicing in HELA S100 extraction, functions in both constitutive and alternative splicing, conservation across species and structure similarity (67, 68).

Only nine proteins in humans are identified as SR proteins: SF2/ASF, SC35, SRp20, SRp40, SRp55, SRp75, SRp30c, 9G8, and SRp54, which are encoded by gene Splicing factor, arginine-serine-rich (*SFRS*) 1-7, 9 and 11 (69). All of them share the same domain architecture: they contain at least one or two RNA recognition motif (RRM) domains at the N terminus and an RS domain at the C-terminus. The RRM domains are in charge of sequence-specific RNA recognition while RS domains mediate protein-protein interaction and RNA-protein interaction (68-70). However, recent cross-linking studies reveal a direct contact between RS domain and critical signals on pre-mRNA sequences (71-73). These are classified as “classical” SR proteins.

A genome-wide study reveals more RS domain containing proteins. Their domain organization is different from “classical” SR proteins. They have one RS domain in the sequence, but the RRM domains are missing. Instead, these proteins have some other domains such as DEAH box domain and PWI domain, which have RNA-binding ability. These proteins are classified as “SR-related” proteins. Some RS domains in “RS related proteins” (U1-70k, Prp28) have many Arg-Asp and Arg-Glu dipeptides repeats and relatively few Arg-Ser repeats(68).

Now it is suggested to divide all SR-contain proteins into three classes based on their domain structures and functions related to splicing (70): Genius “SR proteins” refer to those that have at least one RRM domain and one RS domain, functional in constitutive or alternative splicing. Some SR proteins have a different domain organization compare to “classical” SR proteins, but are still involved in RNA splicing. These proteins are termed as “Additional SR-proteins”. Other proteins that contain one RS domain and other RNA related domains could be labeled as “SR-related factors”. Some the SR-related

factors bind to RNA via domains other than RRM, such as Zinc finger or PWI domains. Other SR-related factors do not directly contact with RNA, but are enzymes or regulators that modulate the splicing. “Classical SR-proteins”, “additional SR proteins”, “RNA binding SR-related factors”, and “RS domain containing factors ” are listed in Tables 1.1 to 1.4. The domain architecture of SR proteins and SR-related proteins are shown in Figure 1.3.

Protein name	**Gene name**	**Key domains**	**Splicing role**	**UniProt**
SF2/ASF	SFRS1	RRM×2, RS	Constitutive and alternative splicing activator	Q07955
SC35	SFRS2	RRM, RS	Constitutive and alternative splicing activator	Q01130
SRp20	SFRS3	RRM, RS	Constitutive and alternative splicing activator	P84103
SRp75	SFRS4	RRM×2, RS	Constitutive and alternative splicing activator	Q08170
SRp40	SFRS5	RRM×2, RS	Constitutive and alternative splicing activator	Q13243
SRp55	SFRS6	RRM×2, RS	Constitutive and alternative splicing activator	Q13247
9G8	SFRS7	RRM, RS, CCHC-type	Constitutive and alternative splicing activator	Q16629

Table 1.1 Classical SR proteins (reproduce from(68))

Protein name	Gene name	Key domains	Splicing role	UniProt
p54	SFRS11	RRM, RS	Alternative splicing repressor	Q05519
SRp30c	SFRS9	RRM×2, RS	Constitutive and alternative splicing regulator	Q13242
SRp38, TASR	FUSIP1	RRM, RS	General splicing repressor	O75494
hTra2α	TRA2A	RRM, RS×2	Splicing activator	Q13595
hTra2β	SFRS10	RRM, RS×2	Splicing activator	P62995
RNPS1	RNPS1	RRM, RS	Constitutive and alternative splicing regulator	Q15287
SRrp35	SRRP35	RRM, RS	Negative regulator of alternative splicing	Q8WXF0
SRrp86, SRrp508	SFRS12	RRM, RS	Positive and negative regulator of alternative splicing	Q8WXA9
U2AF35	U2AF1	RRM, RS, C3H1-type	Constitutive splicing factor	Q01081
U2AF65	U2AF2	RRM×3, RS	Constitutive splicing factor	P26368
U1-70K	SNRP70	RRM, RS	Constitutive splicing factor	P08621
XE7	SFRS17A	RRM, RS	Alternative splicing regulator	Q02040
SRp46	SFRS2B	RRM, RS	Constitutive and alternative splicing regulator	Q9BRL6

Table 1.2 Additional SR proteins (reproduce from(68))

Protein name	Gene name	Key domains	Splicing role	UniProt
Urp	ZRSR2	RRM, RS	Splicing factor	Q15696
HCC1/CAPER	RBM39	RRM, RS	Alternative splicing regulator	Q14498
hSWAP	SFRS16	RS	Alternative splicing regulator	Q8N2M8
Pinin	PNN	RS	Alternative splicing regulator	Q9H307
SRrp129	SFRS2IP	RS	Splicing factor	Q99590
U4/U6·U5 tri-snRNP-	RY-1	RS	Unknown	Q8WVK2
LUC7B1	LUC7L	RS, C2H2-type zinc finger	Unknown	Q9NQ29
Acinus	ACIN1	RRM, RS, SAP	Unknown	Q9UKV3
SR-A1	SFRS19/ SCAF1	RS	Unknown	Q9H7N4
ZNF265	ZRANB2	RS, RANBP2-type zinc	Alternative splicing regulator	O95218
SRm160	SRRM1	RS, PWI	Constitutive and alternative splicing co-activator	Q8IYB3
SRm300	SRRM2	RS	Constitutive and alternative splicing co-activator	Q9UQ35
RBM5	RBM5	RRM×2, RS, RANBP2- and	Unknown	P52756
U2-associated protein	SR140	RRM, RS	Unknown	O15042
RBM23	RBM23	RRM×2, RS	Unknown	Q86U06
SFRS15	SFRS15	RRM, RS	Unknown	O95104

Table 1.3 RNA-binding SR-related factors (reproduce from(68))

Protein name	Gene name	Key domains	Splicing role	UniProt
SRrp53	RSRC1	RS, coiled-coil domain	Unknown	Q96IZ7
hPRP5	DDX46	RS, DEAH box	Spliceosomal rearrangement	Q7L014
hPRP16	DHX38	RS, DEAH box	Splicing factor	Q92620
Prp22/HRH1	DHX8	RS, DEAH box	Spliceosomal rearrangement	Q14562
U5-100k/hPRP28	DDX23	RS, DEAD box	Spliceosomal rearrangement	Q9BUQ8
ClkSty-1	CLK1	RS, kinase domain	SR protein kinase	P49759
ClkSty-2	CLK2	RS, kinase domain	SR protein kinase	P49760
ClkSty-3	CLK3	RS, kinase domain	SR protein kinase	P49761
Prp4k	PRPF4B	RS, kinase domain	SR protein kinase	Q13523
CrkRS	CRKRS	RS, kinase domain	SR protein kinase	Q9NYV4
CDC2L5	CDC2L5	RS, kinase domain	Alternative splicing regulator	Q14004
Cyclin-L1	CCNL1	RS, cyclin-like domain×2	Alternative splicing regulator	Q9UK58
Cyclin-L2	CCNL2	RS, cyclin-like domain×2	Alternative splicing regulator	Q96S94
SR-cyp	PPIG	RS, PPIase cyclophilin-type	Regulates localisation of SR proteins	Q13427
CIR	CIR	RS	Alternative splicing regulator	Q86X95
SRrp130	SFRS18	RS×2	Unknown	Q8TF01

Table 1.4 Other RS domain containing proteins (reproduce from(68))

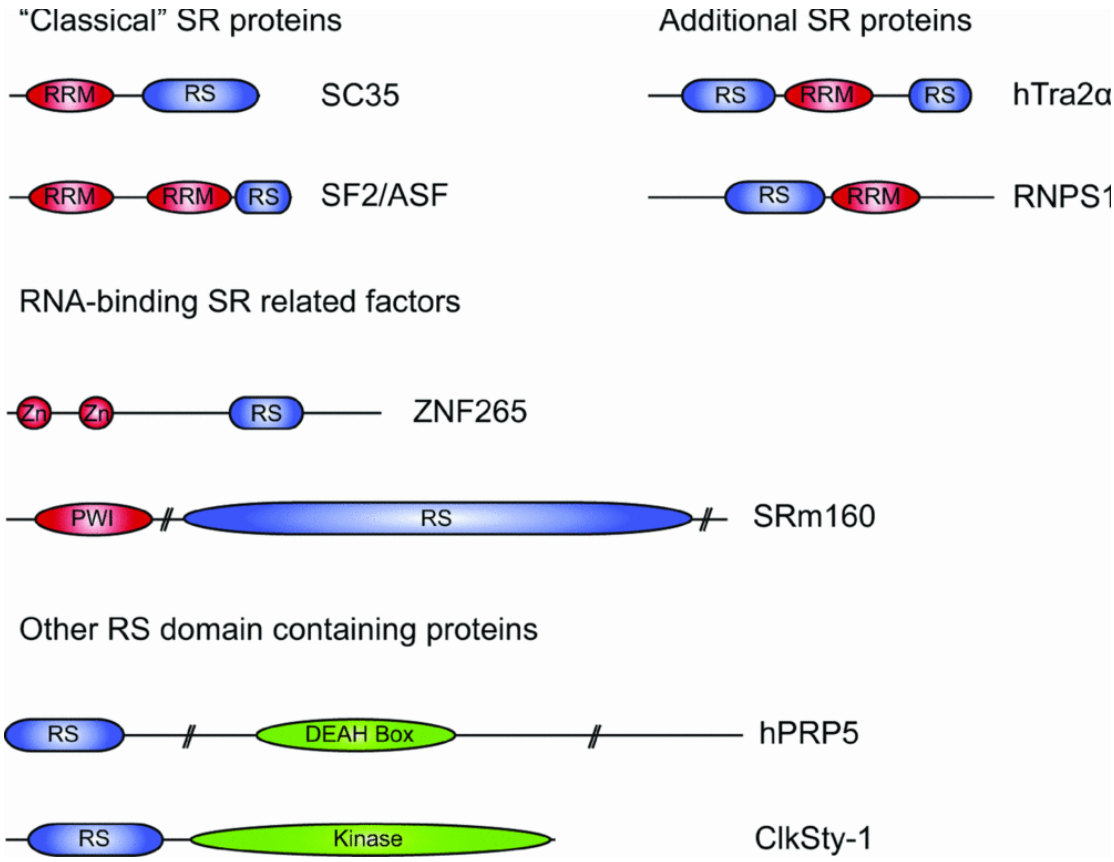


Figure 1.3 Domain structure of SR proteins (reproduce from(68))

1.4.2 SR protein functions

Classical SR proteins are present in all metazoan species as well as some lower eukaryotes, but not all eukaryotes (74). Classical SR proteins are absent in budding yeast, which only has some SR-like proteins (75). Interestingly, 95% of budding yeast genes contains a single intron and alternative splicing is rare. When more SR proteins were being studied, it became evident that the SR protein family is versatile: SR proteins are functional in the whole process of RNA metabolism, from transcription to decay (76).

SR proteins are involved in transcription, splicing, facilitating mRNA exporting, stimulating translation in the cytoplasm, and enhancing nonsense-mediated-decay of mRNA (76). SR proteins bind to the phosphorylated C-terminal domain of RNA polymerase II, which facilitates transcriptional elongation and link transcription with splicing (77). Some SR proteins, such as ASF/SF2, SRp20, 9G8, shuttle between the nucleus and the cytoplasm (78). These SR proteins interact with mRNA nuclear export

receptor TAP/NFX1 (79), by which they affect RNA exporting after splicing (80). SR proteins have also been shown to be related to nonsense-mediated decay (NMD) of mRNA. Overexpressing ASF/SF2, SC35, SRp40 and SRp55 strongly enhance NMD (81). ASF/SF2 stimulates mTOR, which is kinase for various translation factors and inhibits protein phosphatase PP2A, thus enhancing cap recognition and translational initiation (82). SR proteins are required for maintaining genome stability. In vivo depletion of ASF/SF2 in Chicken DT40 cells induced double-stand DNA breaks and gross DNA recombination (83).

1.4.3 SR protein in mRNA splicing

Among various functions of SR proteins, the best known ones play essential roles in both alternative and constitutive splicing (70, 84). In RNA splicing, SR proteins affect recognition and selection of splicing sites, bridging the 3'-end and 5'-end of introns, and play an import role in the incorporation of U4/U5/U6 tri-RNP into spliceosome (Figure 1.4).

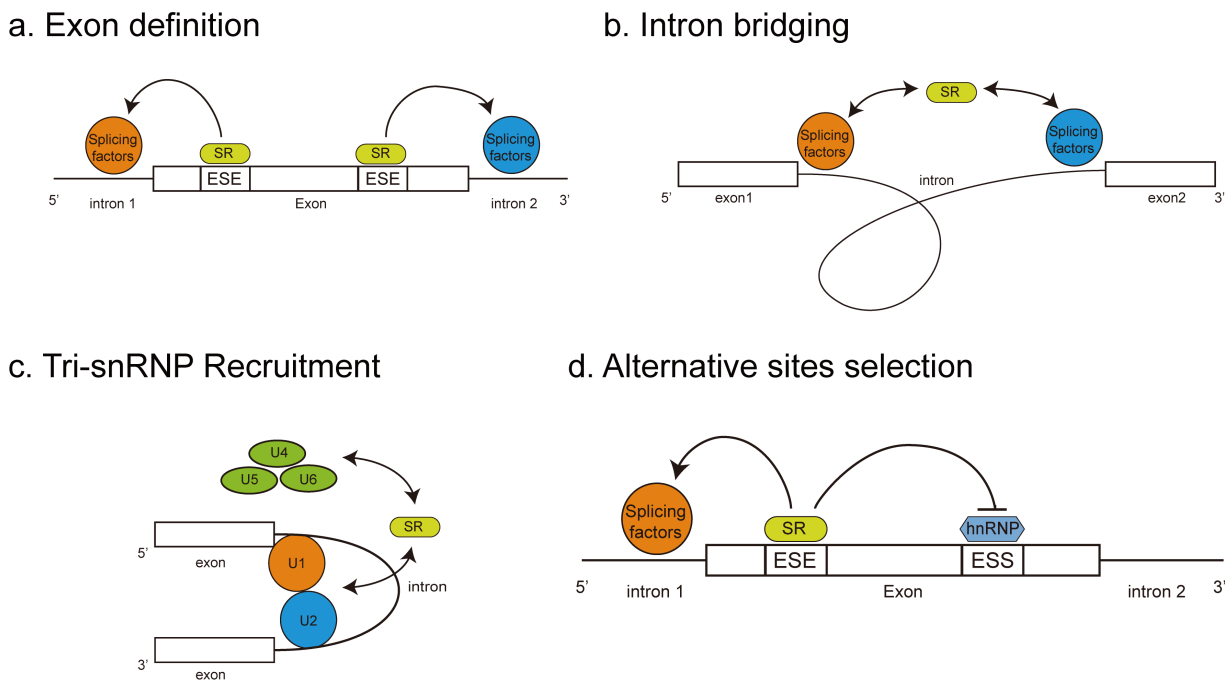


Figure 1.4 Functions of SR proteins in RNA splicing. a) Exon definition. SR proteins recognize the ESE (exonic splicing enhancer) sequences and guide splicing factors to the right position. The khaki sphere represents splicing factors at 3' end of intron while the

blue sphere represents splicing factors at the 5' end of intron. b) Intron bridging. SR proteins interact with splicing factors at both ends of intron to bring them together. c) SR proteins mediate incorporation of U4/U6·U5 tri-snRNP into pre-spliceosome. d) SR proteins antagonist with hnRNP in splicing site selection.

The sequence of splice sites in eukaryotes mRNA is not enough to initiate splicing alone. There are certain sequences in exons or introns, serve as cis-acting elements to facilities splicing site selection, which are termed as exonic splicing enhancers (ESE) and intron splicing enhancers (ISE), respectively. At the first stage of spliceosome, SR proteins bind to certain ESE, and facilitate in two synergic ways spliceosome assembly: first, SR proteins bind to ESEs and then recruit and stabilize other splicing factors on mRNA in the "exon definition" process, such as U1 snRNP at 5'ss (85-87) and U2AF at the 3' splicing site (88-90) (Figure 1.4a). Second, the bindings of SR proteins displace hnRNPs binding to mRNA or prevent them from binding, by which SR proteins antagonize the negative effects of hnRNP (91) (Figure 1.4d). It is also worth noting that two components of the U2AF, U2AF35 and U2AF65 are SR-proteins and SR-related proteins respectively.

The other problem in multi-exon gene splicing is how to ensure the correct order of exons, i.e. how exon skipping is avoided. It is evident that SR proteins binding to ESE also prevent exon skipping (Figure 1.5). Upon binding to ESE, SR proteins favor the splicing on the proximal 3' splicing site rather than on the distal 3'splice site (92). With this preference, SR proteins work as insulators to inhibit pairing of splicing factors at 5' and 3' splicing sites over the ESE sequence, and favor the pairing of closet 5' and 3' splicing sites. Thus, SR protein and ESE sequence together ensure that 5' and 3' splice sites within the same intron are used, therefore suppressing exon skipping.

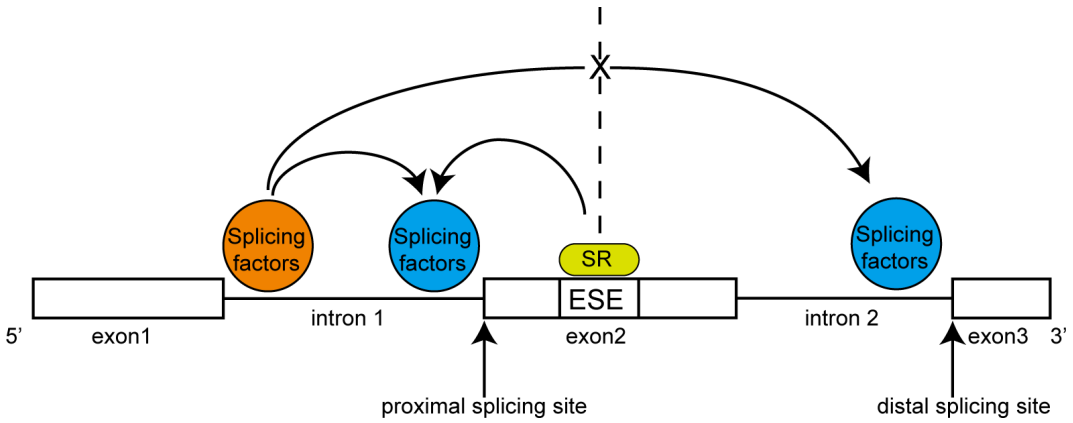


Figure 1.5 SR proteins preventing exon skipping. The color coding is the same as Figure 1.4.

The ESE binding SR proteins also serve as a bridge in correlation of the 3'end and 5'end of the intron through a protein-protein interaction network (Figure 1.4b). It is demonstrated that ASF/SF2 and SC35 can interact either with both U1 small nuclear rib nucleoprotein (snRNP U1-70K) and the 35 kD subunit of the splicing factor U2AF (U2AF35) (93). This is extremely important in higher eukaryotes, in which introns are several kb long, much longer than the exons.

Due to the essential roles of SR proteins in splicing site recognition, it is not surprising to regard SR proteins as master regulators in alternative splicing. Many alternative sites are weak in splicing signals, thus fail to be efficiently recognized by splicing factors, which normally results in exon skipping. SR binding to ESE can compensate the weak signal and recruit splicing factors by the similar mechanism in constitutive splicing.

SR proteins are involved in suppressing or promotion splicing in a substrate-dependent manner: different SR proteins bind to their own ESE respectively and antagonize with each other. The final selection of splicing sites is the result of competition of all SR proteins. Thus, for a particular splicing site, one SR protein can appear as a negative regulator because it promotes splicing on other sites. Moreover, it was shown that the relative positive or negative effects of SR proteins might depend on the relative position of their ESE with respect to the splicing site (94).

There are analogous SR proteins binding sites inside introns that resemble the ESE. Unlike their exonic counterpart, these sequences serve as negative cis-regulators of splicing (95). When SR proteins bind to these repressor elements, they prevent further recruitment of splicing factors, such as demonstrated in adenovirus infection: AFS/SF2 prevents U2 snRNP binding to branch point thus inactivate the 3' splice site. Hence, the SR proteins promote exon inclusion. Thus SR proteins bind to some ESE-like intronic sequence promote splicing at a "cryptic" intron sites instead of "routine" splicing sites (Figure 1.6).

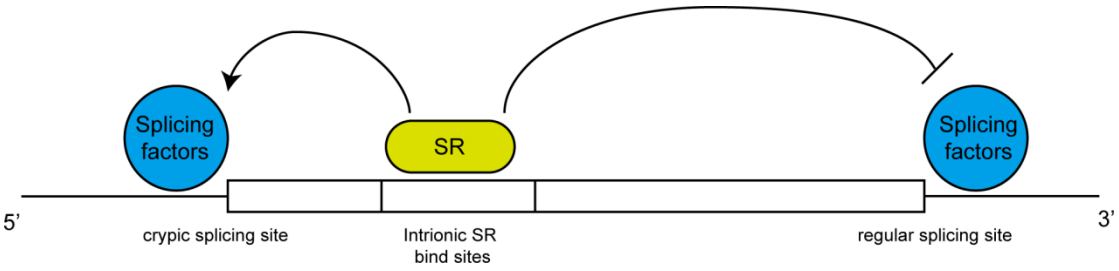


Figure 1.6 SR proteins binding to intronic sequence change splicing sites. The color coding is the same as Figure 1.4.

However, there are several SR-related proteins identified only function as negative regulators without promoting splicing at any other sites on pre-mRNA. For example, SRp35(96), SRp38(97), p54(98) and SRp86 (99, 100) promote exon skipping in alternative splicing and antagonize typical SR protein activities.

U12-dependent introns normally obtain more conservative 5'ss and branch points sequences. However, SR proteins are also shown to participate in U12-dependent splicing by prompting binding of U11 and U12 snRNP to 5'ss and branch point, respectively (101). Therefore, SR proteins appear to have an ancestral origin and were subsequently lost independently in some lineages.

In later stages of spliceosome assembly, SR proteins facilitated the recruitment of U4/U6.U5 tri-snRNP into pre-spliceosome (102)(Figure 1.4c). Interaction between two SR-related proteins, SRrp65 and SRrp110, are required for U4/U6.U5 tri-snRNP incorporation (103). Another SR-related protein, prp28, is important for stability of tri-snRNA, and is also required for tri-snRNP integration into spliceosome (104).

1.4.4 Function redundancy of SR proteins

The redundancy of SR functions is an intriguing puzzle. Because of the assay of SR protein identification, all classic SR proteins have function redundancy: they all can complement splicing in a S100 extract. Both constitute and alternative splicing indeed show little impact in single SR protein knockout cells in most cases, which supports function redundancy of SR proteins in vivo (105, 106). This redundancy can be explained in two ways: first, one ESE sequence may be recognized by several SR proteins (107, 108); second, one exon may obtain several ESE recognized by different SR proteins (109), which act independently or collaboratively (110, 111). These explanations make the implicit assumption that RS domains in SR proteins are exchangeable or at least similar in activity, which is valid to some extent by some domain substitution experiments(112).

However, there are increasing results about deficiency in SR depletion cells and animals. Depletion of ASF/SF2 in chicken cell lines cannot be rescued by the expression of SC35 and SRp40 (113). Depletion of SR protein SRp20(114), SC35(115) and ASF/SF2(116) in mice induces phenotypes in embryonic development indicating that SR proteins are not redundant. In contrast, the cultured cell from knockout mice is viable, indicating that the non-redundant functions of SR proteins are tissue- and developmental stage-dependent. Conditional knockout SR proteins in animals further reveal their non-redundant function in tissue-maintenance and embryogenesis (117).

1.4.5 SR proteins and diseases

Given the essential functions of SR proteins in numerous cell functions, the disruption of SR family protein's functions leads to human disease.

Elevated expression of ASF/SF2, SC35 and SRp20 were observed in some cancer tissues (118, 119); however, the mRNA level of ASF/SF2, SRp40, SRp55 and SRp75 are lower in non-familial colon adenocarcinomas than adjacent non-pathological tissue (120). These different regulation directions indicate that the SR protein levels in cancerous tissues are both tissue- and protein- specific.

ASF/SF2 has specific roles in cancer development. Besides its up-regulation in many human cancers, the gene coding ASF/SF2, SFRS1 is commonly amplified in breast cancer. Increasing expression of ASF/SF2 can transform immortal rodent fibroblasts and induce the formation of sarcomas in nude mice, which can be reversed by the down regulation of ASF/SF2. This transforming activity is unique for ASF/SF2 among other SR proteins. Taken together, these facts support SFRS1 to be a proto-oncogene (121).

In addition, ASF/SF2 regulates the alternative splicing of Ron transcripts, which are well-known proto-oncogenes. ASF/SF2 promotes skipping of exon 11, which results a constitutively active isoform ΔRon. It increases motility on expressing cells, which is required for tumor metastasis. In breast and colon tumors, the accumulation of ΔRon is observed and is correlated with the level of ASF/SF2 (122).

The HIV virus generated more than 40 different mRNAs from its full-length genomic pre-mRNA to generate key viral proteins by a combinational usage of several alternative 5' and 3' splice sites (123). Thus, as the master of alternative splicing regulators, some SR proteins are deeply involved in alternative splicing of HIV viral pre-mRNA and replication of HIV virus accordingly, such as ASF/SF2, SRp75 and SRp40 (124, 125). HIV infection changes the level of splicing factors including SR proteins (126). Current drug developments for treating HIV focus on reverse transcriptase, proteases and gp120 (127). The essential functions of SR in HIV pre-mRNA splicing offer an alternative and novel drug target. Several benzopyridoindole and pyridocarbazole derivatives are identified, which selectively inhibit the ESE-dependent splicing activity of individual SR proteins (128). One of these small chemical compounds was shown to block production of key HIV viral proteins by inhibiting ASF/SF2-dependent splicing (129).

SR proteins are also regulators of pre-mRNAs related to some other human diseases. The absence of the exon 9 of the cystic fibrosis transmembrane conductance regulator (CFTR) is correlated with the occurrence of monosymptomatic and full forms of cystic fibrosis disease (130). ASF/SF2 and SRp40 promote this exclusion by binding to an ISS (131). SC35 has been shown to significantly activate splicing at a cryptic 5'ss splicing site which is introduced by an intronic mutation of E1α pyruvate dehydrogenase. This cryptic splicing site results in a mis-spliced mRNA and a defective protein (132).

Depletion of SC35 by RNAi in primary fibroblasts from patients could restore the normal splicing pattern.

In addition, SR proteins were shown to be the autoantigens in systemic lupus erythematosus patients (133).

1.4.6 RS domains in SR proteins

The RS domain, rich in continual Arg-Ser dipeptide repeats is the marked feature for all SR proteins and SR-related proteins. In RS domain of "Classic SR proteins", such as AF/SF2, RS dipeptide repeats can constitute up to 70% of the sequence. In "SR-related" proteins, however, there are more Arg-Asp and Arg-Glu dipeptide repeats rather than Arg-Ser repeats (134, 135) (Figure 1.7).

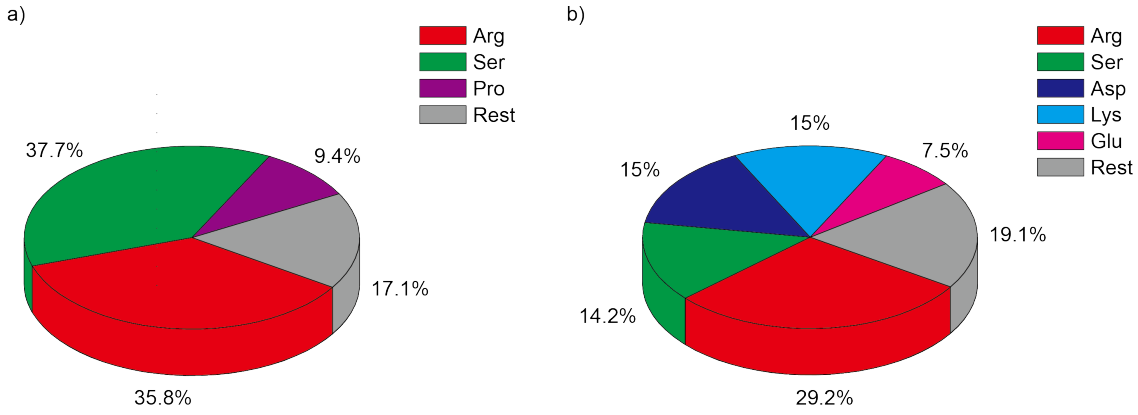


Figure 1.7 Residue compositions of RS domain in classical SR protein ASF/SF2 (a) and SR-related protein hPrp28 (b).

The RS domain of ASF/SF2 can directly contact with RSF1, one splicing regulator in Drosophila (136). The interaction between ASF/SF2 and U1-70k, which is the very initial step in 5'ss recognition, requires RS domains in both proteins. The interactions of ASF/SF2 and SC35 with U1-70k and U2AF35 are RS domain dependent, which is important for bridging 5'ss and 3'ss splicing sites (93). The recruitment of the U4/U6 · U5 tri-snRNP to the pre-spliceosome is similarly mediated by RS domain-mediated interactions with the SR-related proteins SRrp65 and SRrp110 (103). One

peculiarity of these protein-protein interactions is that RS domains binding partners are RS domains from another SR-protein or SR-related proteins.

There is increasing evidence from cross-linked experiments that demonstrate RS domains can directly contact pre-mRNA (71, 72). Initial research showed that RS domains of ASF/SF2 and U2AF65 directly interact with RNA at the branch point in A complex and E complex in spliceosome cycle respectively. Further studies show that during the whole course of spliceosome assembly, different RS domains from different splicing factors sequentially bind to pre-mRNA on different sequences, mainly at 5'ss and at the branchpoint (73). Thus, the interaction with mRNA seems to be a common property of all RS domains.

1.4.7 Phosphorylation of SR proteins

Phosphorylation is one of the most important post-translational modifications. Considering the extraordinary contents of Ser residues in sequence, RS domain is a good substrate for phosphorylation modification. SR proteins isolated from cell extract are highly phosphorylated. The majority of phosphorylation occurs on serines in the RS domain (137). The splicing activity of SR proteins dependent on the phosphorylation contents, both hyper-phosphorylation and hypo-phosphorylation inhibits splicing (138-140). SR proteins undergo several rounds of phosphorylation and de-phosphorylation cycles in the process of splicing.

Some evidence shows that both phosphorylation and dephosphorylation of RS domains regulate the activity of SR proteins, thereby regulating the assembly of spliceosome and splice site selection. For example, phosphorylation of ASF/SF2 enhances the interaction between ASF/SF2 and U1-70K proteins in A complex formation (141). Phosphorylated SR proteins are also required for the U4/U6-U5 tri-snRNP assembly into the spliceosome (102). On the other hand, phosphorylation of ASF/SF2 reduces its interaction with hTra2α, a human homolog of the *Drosophila* splicing regulator Tra2 (142).

Besides splicing activity, phosphorylation of SR proteins affects subcellular localization. Phosphorylation in the RS domain controls SR importing into the nucleus and residing in speckle SR proteins (143, 144). Speckles are dynamic structures that act

as storage sites for pre-RNA splicing factors. Trans-portin-SR, the nuclear import receptor for SR proteins, favors the interaction with phosphorylation RS domain (145). Phosphorylation of RS domain is also required for their recruitment to sites of transcription. Thus the nuclear distributions of SR proteins are also controlled by phosphorylation of the RS domain (146).

1.4.8 Kinases phosphorylate RS domain

Several kinases are responsible for SR protein phosphorylation: Clk/Sty and related kinases (147), SPRK (SR protein kinase) family including SPRK1(147), SPRK2(148), SPRK3 (149) and Atk kinases (150, 151).

SPRK1 and SPRK2 show similar enzyme activity and substrate specificity *in vitro*. But a recently in vivo study showed that SPRK1 is predominately associated with U1 snRNPs, whereas SPRK2 is associated with U4/U6-U5 tri-snRNP (104). The structure of SRPK1 without N-terminal and most insertion domain was determined by X-ray (152).

CLK1 and CLK3 structures are also reported (153). CLK1 contains one kinase domain but without an insertion. Although CLK and SRPK both belong to CMGC group of kinases, CLK1 kinase domain lacks MAPK (mitogen-activated protein kinase), which is between αG and αH in the helix. In addition, CLK kinases contain small insertions between β6 and β9 in the kinase. Interestingly, CLK have an N-terminal extension, which is abundant in isolated Arg-Ser dipeptide, and have autophoshorylation activity on both serine and tyrosine (154).

The molecular details of phosphorylation reactions of SR protein kinases are extensively studied on AFS/SF2 by mutagenesis and mass spectroscopy. The RS domain of ASF/SF2 can be divided into two parts: the N-terminal part (197-228), termed as RS1, has 2 consecutive Arg-Ser stretches which contain 8 and 3 dipeptide respectively; the C-terminal part (229-248), termed as RS2, has less amount of Arg-Ser dipeptide, and its RS repeats are more disperse(Figure 1.8).

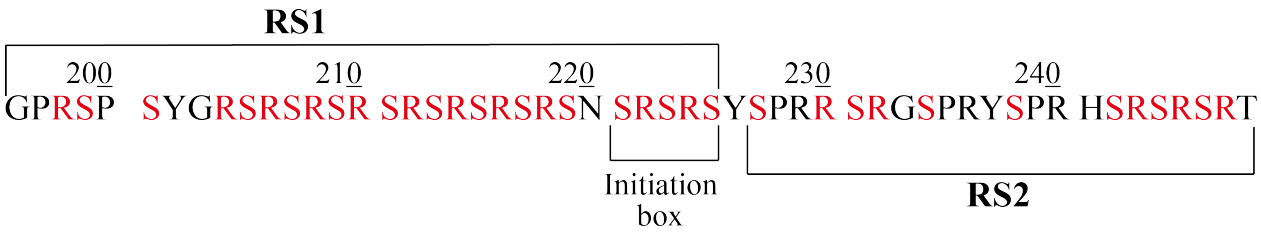


Figure 1.8 Two segments (RS1 and RS2) in RS domain of ASF/SF2. The serine residues are colored in red. The arginine residues in RS dipeptide are also colored in red.

SRPK1 prefers phosphorylating up to 12 serine residues in the N-terminal part (RS1) of the RS domain of ASF/SF2 (155), while CLK does not show such regional specificity and phosphorylates all 20 serine residues in the domain (156). Moreover, CLK can completely phosphorylate RS domain even it is already phosphorylated by SRPK1.

SRPK1 phosphorylate ASF/SF2 in a directional way (157): it binds to one region at the C terminal of RS1 of ASF/SF2 termed as the “initiation box”, and then slides to the N terminal of RS1, phosphorylating 5 to 8 serine residues in the RS1 part in a processive reaction before it dissociates from SR proteins and continues in a distributive manner (155, 156)(Figure 1.9).

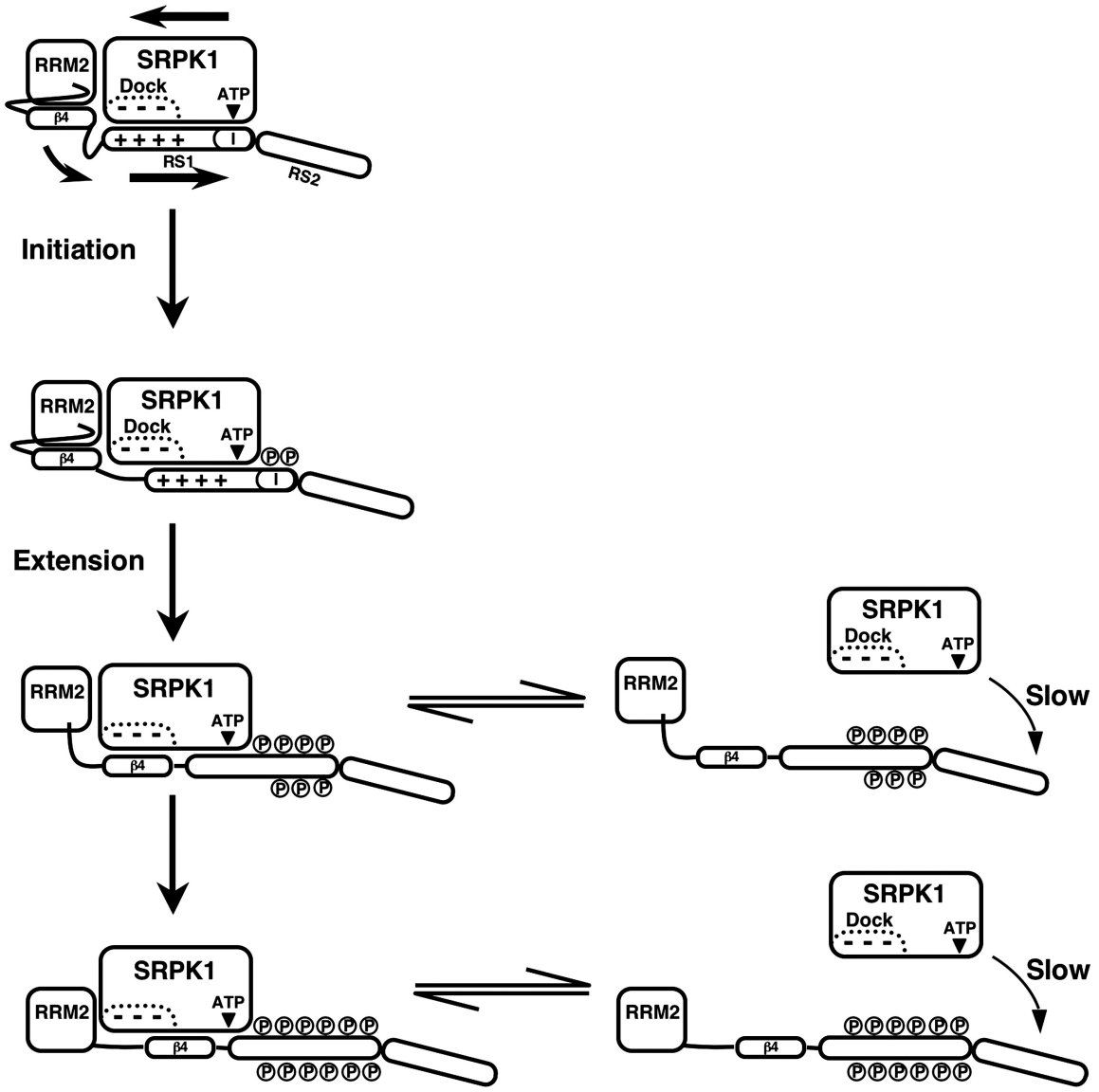


Figure 1.9 Proposed mechanisms for regiospecific phosphorylation of SRPK1 on ASF/SF2. SRPK1 first binds to ASF/SF2 with high-affinity and starts phosphorylation from "initiate box" at the C-terminal of RS1. After phosphorylating 7 to 8 serine residues in RS1 in a sliding manner, SRPK1 disassociates from ASF/SF2 and then rebinds to RS1 or starts to phosphorylate RS2 in a distributive manner.(reproduced from (158))

The mechanism of this processive and directional reaction was revealed by the structure of SRPK1 bound to a small peptide (152) or in the complex with part of ASF/SF2 (RRM2 domain + RS domain) (159). SRPK1 have a docking groove on the big lobe, which can accept some parts of RS domain. The groove is around 6 to 7 residues in

length and constituted by negative charged residues, which facilitate the binding of a basic peptide. Mutations of acid residues in this groove damage the initiation of processive phosphorylation and reduce the directional phosphorylation (155). However, in the two known structures, different sequences are bound to this groove. In the complex with small peptide, this groove was occupied by a small fragment (191-196) precede RS1 (152), while in ASF/SF2 fragment (RRM2+RS) binding structure, N terminal segment (201-210) of RS1 sits in the docking groove (159). The latter observation was unexpected, since this fragment is too far away from activity center as part of substrate, which is phosphorylated during reaction.

The two different structures may account for different stages of directional and processive RS domain phosphorylation, since the RS1 is moving from the C- to N-terminal through the active center during the whole reaction until the last docking element (191-196). Interestingly, the fragment 191-196 is not part of the RS domain but in the last β-sheet of RRM2 domain, which indicates the RRM 2 domain is partially unfolded during processive phosphorylation. Due to the processive mechanism of SRPK1, it prefers long RS repeats such as RS1 part of ASF/SF2.

Compared to the deep and negative charge docking groove in SRPK1, CLK1's corresponding region is shallow and composed of both basic and acid residues (Figure 1.10). This charge distribution makes the binding possible between CLK and RS domains phosphorylated by SRPK1, which possesses alternate positive and negative charges. Thus, CLK and SRPK kinases may effect on the same substrate sequentially: the SR protein phosphorylated by SRPK1 will be the substrate of CLK. It is experimentally proven that CLK1 phosphorylates another 7 serines in RS2 of RS domain of ASF/SF2, after the RS1 of the same domain is phosphorylated by SRPK1. A cell study shows that SR proteins were imported as incorporated to speckle when they are phosphorylated by SRPK kinases, while further phosphorylation by Clk recruited the SR proteins to the transcription active sites.

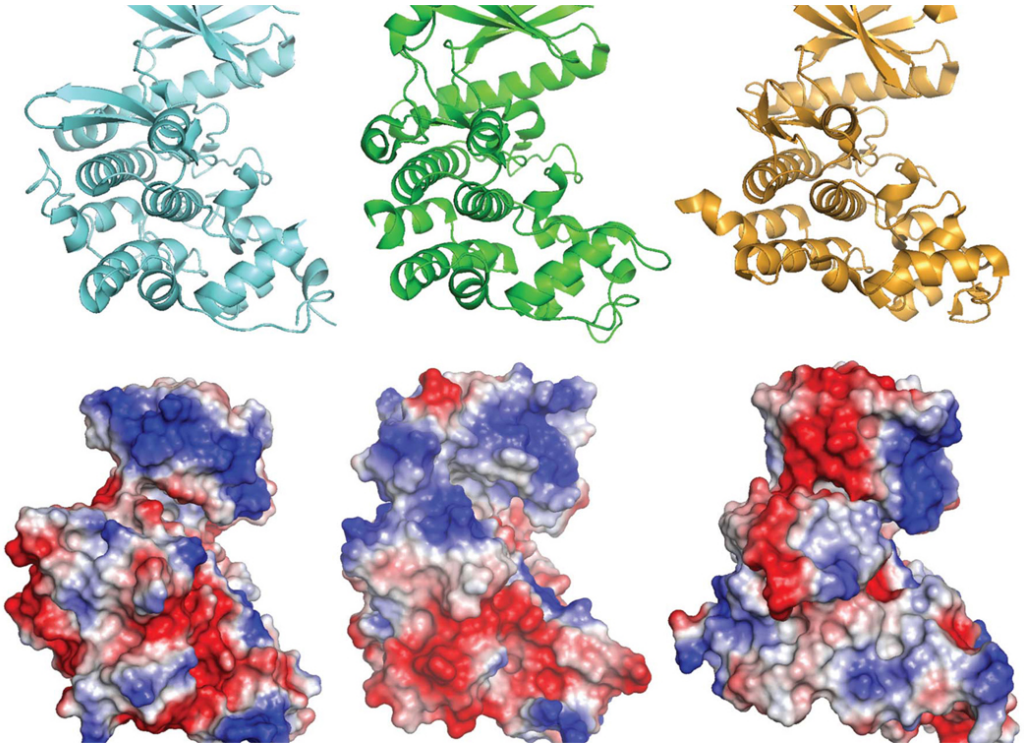


Figure 1.10 The structure (above) and surface electrostatic surface (below) of Kinases for SR proteins. Sky1p is the Yeast SRPK. (Reproduced from (160)with permission)

1.5 Motivation of this project

In spite of the versatile and vital functions of SR proteins, the structure of whole protein is not known so far. This is mainly due to the difficulty to obtain the atomic resolution structure of the RS domain, since RRM domains of SR proteins are solved by X-ray and NMR. Circular dichroism spectra reveal the RS domain is disordered (161). This result also fit with the previous observation that phosphorylation prefer intrinsically disordered regions (51). In an X-ray structure of SRPK1-ASF/SF2 complex, RS repeats docked in a binding groove on SRPK1 and far away from the phosphorylation site is in extended conformation. However, the structure in bound form might not reflect the structure of RS repeats in the free form.

However, there are two molecular dynamic simulation studies that propose different pictures (162, 163). In molecular dynamic (MD) simulations, they did not use the whole

RS domain, but only Arg-Ser dipeptide repeat stretches derived from the N-terminal part of the RS domain. Before phosphorylation, the native (RS)8 peptide (8 times dipeptide repeats) is in a helical conformation. Upon phosphorylation, the backbone gets extended. Particularly, one compact conformation called "Arg-claw" is considered to be the stable and functional form, in which one phosphate from phospho-serine sits in the center and is chelated by guanidinium groups of six or seven arginine residues.

The obvious conflict between experimental data and simulation result was partial eliminated in the second MD study, in which the helix conformation is short and transient. It is worthy to note that both MD study are done in implicit solvent. In both cases, the starting conformations of peptides were set to fully extend or helix in some phosphorylated peptide cases.

2. Materials and Methods

2.1 Materials

2.1.1 Chemicals and Enzymes

During this research work, the following reagents from the indicated suppliers summarized in Table 2.1 were used.

Chemicals/Enzymes	**Supplier Company**
Agar, IPTG, Streptomycine	AppliChem, Darmstadt, Germany
Ammonium chloride (>98 % 15N)	Cambridge Isotope Laboratories, Andover, USA
13C-D-glucose (>98 % 13C)	Spectra Stable Isotopes, Columbia, USA
BamHI, NdeI,NcoI, T4-DNA ligase, Calf-intestinal alkaline phosphatase	Fermentas, St. Leon-Rot, Germany
Coomassie Brilliant Blue R-250	Fluka, Neu-Ulm, Germany
DTT	Gerbu, Gaiberg, Germany
Agarose	GibcoBRL, Karlsruhe, Germany
Bench Mark protein ladder	Invitrogen, Karlsruhe, Germany
Acetic acid, Ammonium sulfate, Ammonium chloride, Ammonium hydrocarbonate, Disodium hydrogen phosphate, Ethanol, Glycerol, Glycine, Hydrochloric acid, Manganese chloride dihydrate, Magnesium sulfate heptahydrate, Sodium chloride, Sodium	Merck, Darmstadt, Germany

dihydrogen phosphate, Sodium hydroxide, TFA, Thiaminechloride hydrochloride, Tris, Urea, Adenosine triphosphate	
Hot Star Taq, Ni-NTA Agarose rasin, QIAGEN Plasmid Midi Kit,	Qiagen, Hilden, Germany
CompleteTM EDTA-free Protease inhitors	Roch Diagnostics, Mannheim,Germany
Acetonitrile, Ampicillin sodium salt, APS, Ethidium bromide, EDTA, HEPES, Imidazole, Magnesium chloride hexahydrate, MES, MOPS, Potassium dihydrogen phosphate, Sodium acetate, TEMED, Tryptone, Yeast extract	Roth, Karlsruhe, Germany
SDS	Serva, Heidelberg, Germany
Turbo Pfu DNA polymerase, dNTP	Stratagene, La Jolla, USA
Glutathione Sepharose 4B	GE Healthcare, Little Chalfont, United Kingdom
Phusion® High-Fidelity PCR Master Mix	Biozym Scientific GmbH, Oldendorf, Germany.
C8E5, n-octanol	Sigma-Aldrich, St.Louis, United States

Table 2.1 Chemicals/enzymes used and a list of suppliers

2.1.2 Bacterial Strains

-*Escherichia coli* BL21 (DE 3). F-, *omp*T, *hsd*Sβ(rβ-mβ-), *dcm*, *gal*, (DE3) tonA.

-*Escherichia coli* Rosessta2 (DE3). F^- *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pRARE2 (Cam^R)

2.1.3 Protein sequences

The ASF/SF2 full length sequence is provided by Dr. Gert Weber from Prof. Markus Wahl's research group, Freie Universität Berlin. The coding sequence of RS domain (1-137) of hPrp28 (DDX23, U5-100K), pCDFDuet-1/ SRPK1 plasmid were provided by Prof. Ficner, Georg-August-University Göttingen.

2.1.4 Oligonucleotide primers

The sequences of Prp28 RS domain (1-137) were amplified by the following primers:

F 5'GCTGAATTCCATATGGCAGGAGAGCTGGCTG 3'

R 5'GTCAAGCTTGGATCCTTACTTAGGCTTCTTATCACCATGTTCATC 3'

RS domain (196-248) was amplified from the plasmid by the following primer:

F 5′GTCCTGCAGGGATC GGGCC AGAAGTCCAAGTTAT G3′

R 5′GGCAAGCTTGTCGACTTATGTACGAGAGCGAGATCTGCTATG3′

The RS8 peptide is the sequence from 200 to 219 of RS domain of ASF/SF2. It was amplified by the following primers:

F 5′GCTGGTACCGCCATGGGCCCAAGTTATGGAAGATCTCGATCTC3′

R 5′GTCGGATCCAAGCTTTTAGCTTCTGCTACGGCTTCTGC3′

2.1.5 Equipments

Laboratory instruments and consumables are summarized in Table 2.2.

Item	**Identifier/company**
Balances	Sartorius B 3100 S, Sartorius, Göttingen, Germany Sartorius AC 210 S, Sartorius, Göttingen, Germany

Centrifuges	Beckmann-Coulter Avanti J-20 and J-301, rotors: JLA 8.100, JLA 9.100, JLA 16.250, JA 25.50 Ti, JA 30.50 Ti, Krefeld, Germany Centrifuges Eppendorf Centrifuge 5415D, Wesseling-Berzdorf, Germany Eppendorf Centrifuge 5804, Wesseling-Berzdorf, Germany
Concentrators	Microcon, YM-3 and YM-10, Amicon, Bedford, USA Concentrators Centricon, YM-3 and YM-10, Amicon, Bedford, USA Centriplus,
Dialysis	Slide-A-Lyzer Dialysis Cassettes, MWCO 10000, 0.1-0.5 ml Capacity, Pierce Biotechnology, Inc., Rockford, IL, USA Spectra Por membranes, MWCO 500,10000, Roth, Karlsruhe, Germany
Electrophoresis	Kodak Electrophoresis documentation and analysis system 120, Eastman Kodak Co., New York, NY, USA Electrophoresis Power Pac 300, BioRad, München, Germany Polyacrylamide gel electrophoresis: Mini- PROTEAN 3 Cell, BioRad, München, Germany Agarose gel electrophoresis: Mini-Sub Cell GT, BioRad, München, Germany
-80°C freezer	MDF-U71V Ultra-low temperature freezer, SANYO Electric Co., Ltd, Osaka, Japan
Filtering	Sterile filter 0,20 µm, Sartorius, Göttingen, Germany
Ni-NTA Agarose	QIAGEN, Hilden
HPLC	System 1: MD-910, PU-980, LG-1580-04, DG-1580-54, AS-950-10, CO-200, JASCO International, Groß-Umstadt, Germany System 2: MD-2010Plus, PU-2080Plus, LG-2080-04, DG-2080-54, AS-2055Plus, CO-200, JASCO International, Groß-Umstadt, Germany Agilent 300SB-C8 4.6X250mm，USA
Incubator	Infors Multitron HT, Einsbach, Germany Certomat R, B. Braun Biotech International, Melsungen, Germany

Lyophylisation	Christ Alpha 2-4, B. Braun Biotech International, Melsungen, Germany
NMR	AVANCE 400 M, Bruker, Karlsruhe, Germany AVANCE 600 MHz, with cryoprobe, z-axis, Bruker, Karlsruhe, Germany AVANCE III 600 MHz, Bruker, Karlsruhe, Germany AVANCE 700 MHz, triple axis gradient, Bruker, Karlsruhe, Germany AVANCE III 800 MHz, triple axis gradient, Bruker, Karlsruhe, Germany AVANCE 900 MHz, with cryoprobe, z-axis gradient, Bruker, Karlsruhe, Germany
NMR tube	Quality NMR Sample Tubes 5 mm and 3mm, Norell, Inc., Landisville, NJ, USA NMR tube Shigemi NMR tube 5 mm, Shigemi Corp., Tokyo, Japan
CD	Chirascan spectrometer, Applied Photophysics Ltd , Leatherhead, UK
UV	Hewlett Packard 8453, HP, Canada

Table 2.2 Equipments

2.1.6 Softwares

Program	Source
Topspin2.1	Bruker Biospin, Karlsruhe, Germany
NMRPipe/NMRDraw	spin.niddk.nih.gov/NMRPipe,(164)
Sparky3	www.cgl.ucsf.edu/home/sparky, T. D. Goddard and D. G. Kneller, University of California, San Francisco
Cara	http://cara.nmr.ch
GAPRO	http://www.aspy.ch
Prosa 6.3	From Prof. Peter Güntert
MARS	http://www3.mpibpc.mpg.de/groups/zweckstetter/_links/software_mars.htm (165)
Pymol	http://www.pymol.org/
Python	http://www.python.org/
Origin 7.5	OriginLab Corporation, Northampton, MA, USA

mddNMR 1.6	http://pc8.nmr.gu.se/~mdd/Downloads/

Table 2.3 Softwares

2.2 Methods

2.2.1 Molecular biology methods

The coding sequence of RS domain(1-137) of human Prp28 (DDX23) was cloned into a modified pET-16 plasmid by NdeI and BamHI.

The RS domain of ASF/SF2 (196-248) was cloned to pETZ2_1a plasmid by NcoI and BamH1 restriction sites. The RS8 peptide was taken from ASF/SF2 (200-219), which was cloned to Petz2_1a in the same way as the RS domain of ASF/SF2.

The pCDFDuet-1 plasmid with SRPK1 sequence from Prof. Ficner was directly used for transformation without any modification.

2.2.1.1 Purification of Plasmid

Purification was done by using the NucleoSpin Plasmid DNA Kit. 3ml E. coli XL2-blue overnight culture can offer up to 15ug plasmid. Plasmid DNA is isolated from bacterial pellets according to manual instructions. For DNA sequencing, all plasmids were produced by QIAprep Midi Kit.

2.2.1.2 PCR amplify

The sequences of RS8 peptide, RS domain of hPrp28 (DDX23) and ASF/SF2 were amplified by PCR reactions.

PCR reaction mixture

2× Phusion Mix	20ul
Primer1(10uM)	2ul
Primer2(10uM)	2ul
Template	10ng
ddH2O	add up to 40ul

PCR cycle: 1. Denaturing at 95°C for 30s

2. Annealing at 55 °C for 1 min

3. Elongation at 72 °C for 1min
4. Repeat to 1 for 32 cycles
5. 10 min at 72°C
6. Keeping in 4°C

2.2.1.3 DNA sequencing

Sequencing service is done by Seqlab (Göttingen, Germany). A 7µl reaction mixture containing 20pmol of primers and 0.6ug plasmid DNA were loaded in a 200µl flat lid PCR tube.

2.2.1.4 Double digestion

BamHI digestion reaction mixture:

DNA (PCR fragment or plasmid)	1µg
BamHI	1µl
BamHI unique Buffer	4µl
H_2O (autoclaved)	add to 40µl

Incubation for 2 hours at 37°C

NcoI digestion reaction is the same as BamHI.

Purification of the fragments by Nucleospin Extract II kit, take up each fragment in 30µl H_2O.

NdeI digestion reaction mixture:

DNA (PCR fragment or plasmid)	30 µl
NdeI	4 µl
10 X buffer orange	5 µl
H_2O (autoclaved)	add to 50 µl

Incubate for 2 hours at 37°C, then add to each reaction 1ul NdeI and go on with digestion overnight at 37°C. Add Calf Intestinal Phosphatase(CIP) 1ul to reaction mixture, digest for 1 hour at 37°C. Purify both fragments with Nucleospin Extract II kit; dissolving each fragment in 30 µl H_2O.

2.2.1.5 Ligation

Reaction mixture

Plasmid (Treated by restriction endonuclease and CIP)	2µl
PCR fragment	9µl
10X Ligation buffer	2µl
T4 DNA ligase	6µl
H_2O (autoclaved)	6µl

Overnight reaction at 14 °C.

2.2.1.6 Transformation of E.coli

1µl plasmid DNA was added to 50µl competent cell stock. After 30 min recovery on ice, cells were subjected to a 45 second heat shock at 42°C. This was then put on ice for 2 min. After that, the cells were supplemented with 200µl 2xYT culture, incubated at 37 °C for 1h. Finally, the cells are plated on LB-agar plates with certain an antibiotic and incubated overnight at 37 °C.

2.2.2 Protein Biochemistry Methods

2.2.2.1 Expression of RS domain of hPrp28 (DDX23)

E.coli Rosessta2 (DE3) cells carrying two plasmids (RS domain and SPRK1) were grown in LB media contained ampicillin and streptomycin. A single colony was used to inoculate a 2 ml LB culture which was grown for 10hrs. Then this culture was transferred to 50ml LB culture for overnight incubation at 37 °C. Finally, the 50ml culture was transferred to 1L LB media. When OD_{600} reached 0.6, expression was induced by 1mM IPTG.

Similar protocols were used for production of unlabeled and ^{13}C-^{15}N labeled hPrp28. In the case of ^{13}C and ^{15}N-labeled sample, LB media is replaced by M9-minimal medium culture which contains the labeled carbon and nitrogen source.

M9 minimal medium: 1L

10* M9 salts	**100 ml**

NH_4Cl	1 g
20 % Glucose	20 ml
1M $MgSO_4$	2 ml
200 mM $CaCl_2$	0.5 ml
5 mg/ml Thiamine'HCl	6 ml
Trace elements	10 ml
ddH_2O	880 ml
Ampicillin	50 mg
Streptomycin	50 mg
Chloromycetin	34 mg

10* M9 salts:	**1L**
Na_2HPO_4	67.8 g
KH_2PO_4	30 g
NaCl	5 g
NH_4Cl	10 g

Trace elements	**100 ml**
$FeSO_4.7H_2O$	0.6 g
$MnCl_2.4H_2O$	0.115 g
$CaCl_2.6H_2O$	0.08 g
$ZnSO_4.7H_2O$	0.07 g
$CuCl_2.2H_2O$	0.03 g
H_3BO_4	0.002 g
$(NH_4)_6MO_7O_{24}*4H_2O$	0.025 g

Stir for 10 min; add 0.5g EDTA, stir overnight (all materials were filtered before use).

Cell pellet was harvested, after 3-4 hours induction at 37 °C, by centrifugation at 6000*g* for 30 min.

2.2.2.2 Purification of RS domain of human Prp28

The bacterial pellet from 1L culture was re-suspended in a 40-60 ml lysis buffer containing 2M LiCl, 50 mM TrisHCl, pH 7.5, 5% glycerol, 2 mM β-mercaptoethanol,0.5mM PMSF. Complete-EDTA free inhibitor mix (1 tablet in 100 ml lysis buffer), and then was lysed by sonication. Spin at 20,000g, 40mins, keep the supernatant.

Hi-Trap Ni-column was first equilibrate with column buffer (500 mM NaCl, 50 mM Tris·HCl ,pH 7.5, 5 % glycerol).The supernatant of sonicated lysate was loaded on the column at 1ml/min speed, and then the column was washed 2 ml/min with a column buffer supplemented with 15 mM imidazole. Elute protein with linear 100 ml gradient from 15 mM to 300 mM imidazole in column buffer. Then His-tag was removed by overnight treatment with TEV enzyme with a 1:1000 ratio.

To obtain the completely phosphorylated RS domain, incubate SPRK1 in 100 mM NaCl, 50 mM Tris·HCl, pH 7.5, 5 % glycerol, 2 mM DTT, 10 mM $MgCl_2$, and 2 mM ATP at 4°C overnight.

The phosphorylated RS domain was also purified by C8 Reversed phase-HPLC. The phosphorylation was checked by HPLC-MS. RP-HPLC and HPLC-MS systems were operated by Gerhard Wolf in the department of NMR-based Structural Biology at MPIBPC.

The peak was collected, frozen, and lyophilized. Then the sample was dissolved in the desired buffer, and dialyzed to the same buffer again, to make sure all the chemical components from HPLC step had been removed. The NMR buffer of hPrp28 (DDX23) was 50mM Na- phosphate buffer, pH6.0, 100mM NaCl.

The wild type and phosphorylated 13C/15N labeled full length hPrp28 was prepared by Prof. Ficner's group at Georg-August-University Göttingen.

2.2.2.3 Expression and purification of SRPK1

The expression protocol of SRPK1 alone was the same as the co-expression of SRPK1 and hPrp28. The cells were harvested and lysed in the same way.

The purification procedure of SRPK1 followed the manual of Glutathione Sepharose 4B from GE Healthcare.

1. The column was washed 5-10 bed volumes of PBS (140mM NaCl, 2.7mM KCl, 10mM Na_2PO_4, and 1.8mM KH_2PO_4) to remove preservative.
2. The column was then equilibrated with 3 bed volumes PBS.
3. The supernatant of lysate was applied to the column, and incubated for 30mins at 4°C.
4. The column was washed with 7 beds volumes of PBS.
5. The bund protein was eluted by 5 bed of elution buffer (10mM reduced Glutathione, 50mM Tris-HCl buffer, pH7.5,100mM NaCl)

The eluted protein was stored at -80°C and used in phosphorylation.

2.2.2.4 Expression and purification of RS domain of ASF/SF2

E.coli BL21 (DE 3) carrying Petz2_1a_ ASF plasmid were grown in LB media containing Kanamycin. The expression protocol, lysis buffer and sonication of RS domain of ASF/SF2 were the same as for the hPrp28 RS domain.

The Qiagen Ni-NTA Superflow Columns were first equilibrated with a column buffer (50 mM Tris·HCl, 1M NaCl, pH 7.5, 5% glycerol) with 20mM imidazole. The supernatant of lysate was adjusted to 20mM imidazole, and then was loaded on the column. The column was washed by column buffer with 20mM imidazole of 5 times volume of resins. The target protein was eluted by 500mM imidazole in column buffer. The eluted protein was concentrated and purified by a C-18 reverse phase HPLC.

For phosphorylation, RS domain of ASF/SF2 was incubated with SRPK1 in 50 mM Tris-HCl buffer, pH7.5, 1M NaCl, 10mM ATP, 10mM $MgCl_2$, at 30°C for 4 hours. The reaction mixture was isolated by C-18 reverse phase HPLC, and the phosphorylation state of RS domain was checked by mass spectroscopy.

The peak was collected, frozen, lyophilized and re-dissolved as RS domain of hPrp28 (hPrp28). The NMR buffer for RS domain of ASF/SF2 was 50mM Na- phosphate buffer, pH7.0, 100mM NaCl.

2.2.2.5 Expression and purification of RS peptide

The RS8 peptide is the sequence 200-219 from RS domain of ASF/SF2. The expression and purification procedures with Hi-Trap Ni-column column are the same as RS domain of hPrp28.

Phosphorylation was done with SRPK1 in 50 mM Tris-HCl buffer, pH7.5, 150mM NaCl, 10mM ATP, 10mM $MgCl_2$, at 30°C for 4 hours. Then the phosphorylated forms were purified as RS domain of ASF/SF2. The NMR buffer for RS domain of ASF/SF2 was a 50mM Na- phosphate buffer, pH7.0, 100mM NaCl.

One 16 residues peptide, (RS)8 were synthesized by Volker Klaukien of the Max Planck Institute for Biophysical Chemistry. The powder of (RS)8 peptide was dissolved in 50mM HEPES buffer,300mM NaCl, and the pH was adjusted to 6.5.

2.2.3 NMR methods

2.2.3.1 HSQC (Heteronuclear Single-Quantum Coherence)

In NMR spectroscopy, nuclear spin is subjected to an external magnetic field B0, which is slightly shielded by the orbiting electron. As a result, the observed resonances differing from each other depend on the chemical environments of each nucleus. This phenomenon is called chemical shift.

HSQC is the most frequently used and also most sensitive heteronuclear spectrum. The pulse programs begin with the excitation of a proton and then transfer the magnetization to the attached heteronucleus (^{15}N, ^{13}C), and subsequently reverse transfer back to proton for detection. Since proton has higher gyromagnetic ratio than carbon and nitrogen, these two steps transfer improve the sensitivity of spectra.

HSQC can offer the chemical shifts of proton and covalently connected carbon or nitrogen atom. ^{1}H-^{15}N HSQC is extremely useful for protein study. Every backbone amide proton and nitrogen pair can give one peak on the spectra, with the proton chemical shift in one dimension and directly connected nitrogen in another dimension. If

the chemical environment of the group changed because of ligand binding, folding, unfolding etc., the corresponding peak will be shifted. Therefore, ^{15}N-HSQC is a powerful tool to monitor the backbone conformation information at atomic resolution.

The ^{1}H-^{15}N HSQC pulse program used in this study was the fast HSQC detection scheme described before (166). The water magnetization was flipped on to the z-axis before detection to avoid saturation of water signal. Flipping of water magnetization allows short inter-scan delay. The pulse program is shown in Figure 2.1. The sweep width for RS domains and isolated peptide were 16ppm (N) and 11ppm (H). The N and H channels were centered at 117 ppm and water position respectively. The spectra were recorded in a data matrix of 256(t1)*512(t2) complex points for each dimension.

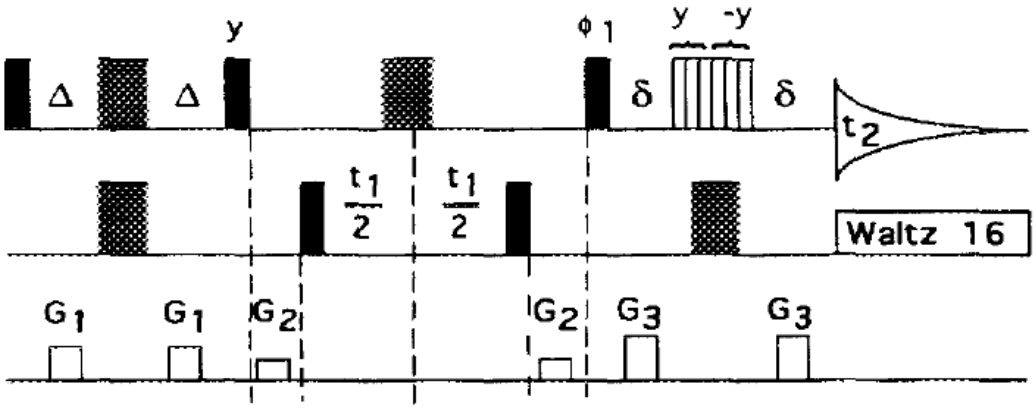


Figure 2.1 Pulse program of fast ^{1}H-^{15}N HSQC. Narrow and wide pulses denote 90° and 180° pulses respectively. All pulse phages are x except where indicated. The comprised watergate is 3-9-19 comprised pulses with delay 166us, 140us, 125us and 110us at 600M, 700M, 800M and 900M respectively. Δ=2.7ms.δ=2.1ms for a total refocus INPEPT time of 2.7ms. Gradient pulses are 1ms half-sine shaped. G1=11.5 G/cm, G2=10 G/cm, G3=25 G/cm. This first nitrogen 90° pulse and receiver phase are cycled (x, -x). (Reproduced from (166) with permission)

2.2.3.2 Triple resonance experiments

Backbone assignment is the first step of NMR study for protein, which means assigning peaks on ^{15}N-HSQC to the residues in the protein sequence. For ^{13}C, ^{15}N-labled-protein normally triple resonance experiments are used. These experiments make

magnetization transfer from protons to C and N nuclei, frequency labeled on those nuclei, and finally transfer back to the proton for detection. All transfer steps rely exclusively on 1J/2J couplings between H, C and N nuclei of protein (Figure 2.2).

In HNCA experiments (167), the 1HN and 15N are correlated with intra-residue Cα and previous Cα. In the spectrum, every 1HN and 15N combinations, which means one peak on the HSQC spectrum, have two peaks result from Cα(i-1), Cα(i). To distinguish Cα(i-1) from Cα(i), HN(CO)CA is also recorded, in which Hi,Ni only correlated to Ca(i-1) by making the magnetization pass though the previous residues' CO.

In principle, only HNCA and HN (CO) CA can yield the backbone assignment. However, the Cα chemical shift alone is not enough to determine the amino acid type and in many cases, Cα chemical shift overlap leads to an ambiguous assignment. Therefore, an improved strategy was developed based on the use of both Cβ and Cα chemical shift. Since the combination of Cβ and Cα chemical shifts for every amino acid type is very characteristic, this information is extremely useful when mapping the sequence of the protein.

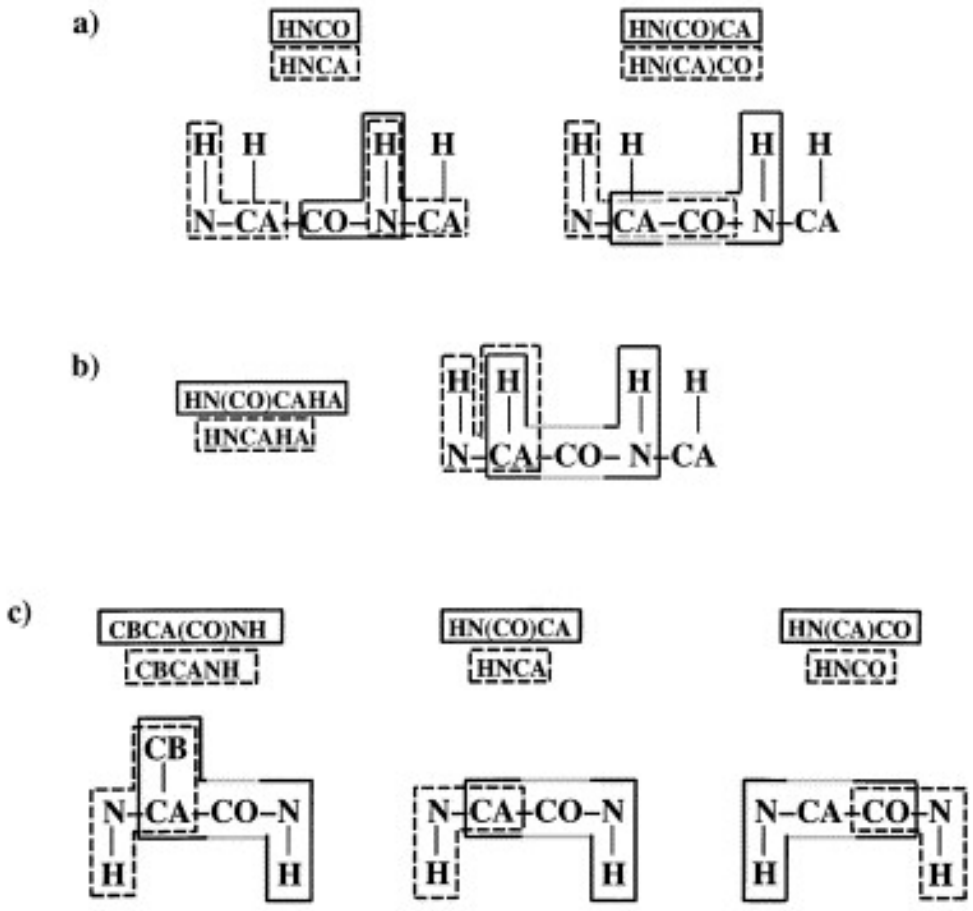


Figure 2.2 Backbone assignment strategy using triple resonance experiments. The boxes drawn around the spin system indicate the connectivity established in the respective experiment. Grey parts of the boxes indicate that a spin is used as a relay spin, without chemical shift evolution. (Reproduced from (167) with permission)

Similar to HNCA and HN (CO) CA, there are also two experiments that offer CB chemical shifts: HNCACB and CBCA (CO) NH. Compared to HNCA, HNCACB has one more transfer step to Cβ after magnetization is on Cα. So in this spectrum, every HN*i* and N*i* combination correlated with 4 peaks result from Cαi-1, Cβ*i-1*, Cαi, Cβi. As in HNCA, another spectrum CBCA (CO) NH is needed to distinguish inter-residue peaks from intra-residue peaks. CBCA (CO) NH uses a different transfer pathway: it starts with aliphatic protons instead of amid protons. By an INEPT step, magnetization is transfer to Cα and Cβ, and frequency evolved on them. In following steps, magnetization goes to Cα, and then transferred to N via CO, so only Cαi-1 and Cβi-1 are observed in this spectrum.

Based on these entire triple resonances spectra, one can group chemical shift values of Cαi-1, Cβi-1, Cαi, Cβi for every amide group. The next step is to find connections, which means looking for another amide group with the Cαi, Cβi chemical shift value matching with the Cαi-1, Cβi-1 chemical shift of this amide group. The fragment of several connected amide group can be mapped to sequence of protein to finally get the complete backbone assignment of the protein (Figure 2).

The triple resonance experiments used for assign RS domain of hPrp28 (DDX23) are listed in Table 2.4.

Experiment	Pulse program	Sweep width (PPM)	Carrier frequency (PPM)	Number of complex points
HNCACB	hncacbgpwg3d from Bruker	10(H)/50(C)/22(N) @900M	4.7(H)/50(C)/117(N)	512(H)×37(C) ×37(N)
CBCA(CO) NH	cbcaconhgpwg3d from Bruker	10(H)/50(C)/22(N) @800M	4.7(H)/46(C)/117(N)	512(H)×40(C) ×30(N)
CT-HNCA(168)	hacannh see Appendix	11(H)/30(C)/22(N) @900M	4.7(H)/55(C)/117(N)	512(H)×40(C) ×45 (N)
HNN(169)	hnngp3d see Appendix	14(H)/22(N_1)/ 22(N_2) @900M	4.7(H)/117(N_1)/ 117(N_2)	512(H)×50(N_1)×50(N)

Table 2.4 Parameters of triple resonance experiments

2.2.3.3 Automated projection spectroscopy (APSY)

Since unfolded proteins cannot be crystallized, NMR is the only source of atomic-resolution information. Unfortunately, the small chemical shift dispersion in "unstructured" protein makes the assignment of unfolded protein a laborious process. As mentioned above, backbone assignment by conventional triple resonance experiments is based on the matching of Cα and Cβ chemical shifts. However, in unfolded proteins, carbon chemical shift deviations from standard values are very small, which lead to serious ambiguity in assignment.

To solve this problem, high-dimensional NMR experiments are employed. High-dimensional NMR experiments reduce resonance overlap by separating peaks in higher dimension space, and eliminate ambiguity during chemical shift matching by direct correlation of a large number of nuclei.

To record high-dimensional NMR experiments in a reasonable time, APSY (Automated projection spectroscopy) experiments were designed (170). The basic idea of APSY is running a series of 2D projection spectra instead of running high dimension experiments. A special algorithm (GAPRO) is used to calculate final peaks list from the peaks in all projections. The peak list can be used as input for automatic assignment software.

There are currently several high-dimensional NMR experiments available. The most powerful APSY pulse program is 7D seq-HNCO(CA)CBCANH, which can offer sequential connectivity of neighboring backbone amide groups and also the Cα, Cβ, CO chemical shifts between them (171). The magnetization of the amide proton of residue i, is transferred with a series of INEPT steps via the scalar couplings $J_{N,NH}$, $J_{N,CO}$, $J_{CO,C\alpha}$, $J_{C\alpha,C\beta}$, $J_{C\alpha,N}$ and $J_{N,HN}$. Along this transfer path, six evolution periods are involved on H^{N}_{i}, Ni, C'_{i-1}, C^{β}_{i-1}, C^{α}_{i-1} and N_{i-1}, respectively. By matching the combination of N and proton chemical shifts, one can connect sequentially neighboring residues. In addition to N and proton chemical shifts, 7D seq-HNCO (CA) CBCANH also provide the Cβ, Cα, CO chemical shifts, which can be used to map the dipeptide fragment to protein sequence.

An alternative approach of 7D APSY is the combination of 5D-CBCACONH and 6D-HNCOCANH. 6D is used to give the sequential connection of neighboring amide group, and 5D-CBCACONH is used to measure the Cα and Cβ chemical shift, which thereby can be used to determine residue type. Since both spectra have less INEPT transfer steps, 6D and 5D have higher sensitivity.

The original APSY pulse programs (5D, 6D and 7D) and GAPRO software are downloaded from www.apsy.ch. The pulse programs are modified to improve water handling. Before and after every comprised pulsed decoupling (CPD) sequence, one 90° pulse proton pulse is added to flip water magnetization. The one before CPD sequence

flipped the water magnetization to the same direction as CPD pulses .The one after CPD sequence flipped the water magnetization back to z-axis. The spectra was then processed by software PROSA(172). The peaks on every projection spectra were picked and the final peak list was calculated by GAPRO. Then the peak lists was fed in MARS to get the assignment (173).

Backbone assignments of native and phosphorylated ASF/SF2 (200-219) were obtained by 7dimensional HNCOCACBNH APSY experiments with 33 projection angles (174). 6D-seq-HNCOCANH APSY and 5D CBCACONH APSY experiments are used to assign hPrp28 (175). 6D and 5D APSY were also applied to RS domain of ASF/SF2.The parameters of APSY experiments are listed in Table 2.5.

Protein	Experiments	Number of projection angles	Number of data points in every projection (real points)
hPrp28	6 D APSY-seq-HNCOCANH	25	1024×180
	5D APSY-CBCACONH	23	1024×160
RS peptide	7D APSY-seq-HNCO(CA)CBCANH	33	1024×180
RS domain of ASF/SF2	6 D APSY-seq-HNCOCANH	25	1024×200
	5D APSY-CBCACONH	23	1024×160

Table 2.5 Parameters of APSY experiments

In 6D and 7D APSY experiments, the carrier frequencies were: Amide proton 8.2 ppm, N 117ppm, CO 172.5ppm, H 4.7 ppm. In 6D APSY Cα were at 53ppm, while in 7D and 5D APSY Cα and Cβ carrier frequency were both at 42ppm.All chemical shifts data are referenced to the DSS signal.

2.2.3.4 Experiments of Coupling measurements

The detail about method of coupling measurements is discussed in the chapter 5.4.

3 bond J couplings, HN-Hα, N-Cγ,CO-Cγ were measured by intensity modulated experiments. $^{3}J_{HN\text{-}H\alpha}$ couplings are measured by intensity modulation in HSQC with

mixing times of 40ms, 45ms and 50ms(176). The Hα magnetization was flipped by Q3 shaped pulse centering at 4 ppm, covering 2ppm.

3J Co-Cγ and 3J N-Cγ couplings of argines' side chains were measured by 2 dimensional spin-echo difference experiments (177, 178) in which mix times are 100ms for N-Cγ and 90ms for Co-Cγ respectively. In N-Cγ couplings measurements, Cγ was flipped by Q3 shaped pulse centering at 25 ppm and covering 24 ppm. In CO-Cγ couplings measurements, Cγ was flipped by 500us sinc pulse which aims at Cγ of arginine.

One-bond J couplings, $J_{Hα-Cα}$, $J_{Cα-Cβ}$, were measured by the splitting method. The $J_{Hα-Cα}$ was measured by constant-time HNCA method without proton decoupling pulse during the Cα evolution. $J_{Cα-Cβ}$ couplings were measured by normal HNCA and HNCOCA methods.

The parameters of couplings measurements are listed in Table 2.6.

Couplings	Sweep width (ppm)	Carrier frequency (ppm)	Number of data points (real points)
HN-Hα	11(H)/16(N)	4.7(H)/117(N)	2048(H)×256(N)
N-Cγ	11(H)/16(N)	4.7(H)/117(N)	1024(H)×256(N)
CO-Cγ	11(H)/16(N)	4.7(H)/117(N)	1024(H)×256(N)
Hα-Cα	10(H)/8 (C)	4.7(H)/53(C)	2048(H) ×80(C)
Cα-Cβ	10(H)/10(C)	4.7/55(C)	1024(H)×256(C)

Table 2.6 Parameters of Couplings measurements

The spectra were processed by NMRpipe software, and analyzed with Sparky. The positions and intensities of peaks are determined by peak picking function of Sparky. The J couplings from intensity modulated experiments are calculated as described previously (176) (177, 178).

2.2.3.5 RDC measurements

The samples were aligned in 5% C8E5/n-octanol liquid crystalline phases (179). The C8E5 was taken out into EP tube with the protection of nitrogen gas. All buffers and

protein samples were cooled by ice. The pure C8E5 was first diluted to 15% by the NMR buffer, and then added to NMR sample to reach the final concentration at 5%. n-octanol was added to the mixture in microliter steps. The transition to Lα phase was reached when the solution just changed from milky to transparent and slightly blue. After placing the sample in the magnet for more than 15mins, the quadrupolar splitting of deuteron was measured.

Three RDCs were measured: D_{N-H}, $D_{H\alpha-C\alpha}$, and $D_{C\alpha-CO}$. The same experiments were carried out twice, once in isotropic samples, once in aligned samples. The RDC values were calculated as the difference of measured couplings in the two sets of experiments.

Trosy-HSCQ interleaved experiments were used to measure HN-N couplings (180). Hα-Cα couplings were measured by HACANNH. To achieve high resolution and intensity, these experiments were recorded with Multi-Dimensional Decomposition (MDD) non-uniform sampling scheme. Cα-Co coupling were measured by spin-state selection HNCO pulse program carried out in a 2D manner (181). The parameters are listed in Table 2.7.

The spectra were processed by NMRpipe software, and analyzed with Sparky. The positions of peaks are determined by peak picking function of Sparky. Non-uniform sampled data were processed by the multiple dimension decomposition method (182).

RDCs	Sweep width (ppm)	Carrier frequency (ppm)	Number of data points (real points)
H-N	11(H)/16(N)	4.7(H)/117(N)	1024(H)×352(N)
Cα-Hα	11(H)/16(N)/10(C)	4.7(H)/117(N)/53(C)	1024(H)×96(N)×96(N) (576×4) by non-uniform sampling
Cα-CO	11(H)/16(N)/8(C)	4.7(H)/117(N)/174(C)	1024(H)×1(N)×256(C)

Table 2.7 Parameters of RDC measurements

2.2.3.6 Heteronuclear NOE measurements

Heteronuclear NOE measurements were carried out by the Bruker pulse program **hsqcnoef3gpsi**. The interscan delay was set to 7.5s to avoid saturation of water.

3. Result

3.1 Phosphorylated ASF/SF2 RS domain adopts structure.

The RS domain (196-248) of ASF/SF2 was cloned to vector Pet-Z2 plasmid and expressed as a fusion protein with the Z domain of staphylococcal protein A. There was some protein eluted at 40mM imidazole, however, most parts of product were eluted by 500mM imidazole in the end. From the 1L M9 culture, more than 20 mg of protein was harvested from the NTA column. The product of important purification steps was sampled and checked by SDS-PAGE, which can be seen in Figure 3.1.

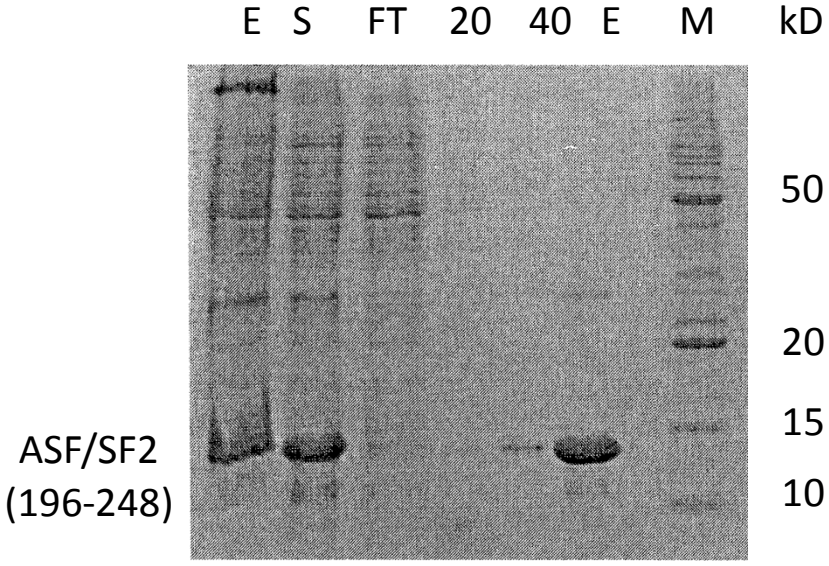


Figure 3.1 SDS gel for ASF/SF2 RS domain (196-248) production. Lane W, the lysis of whole cell; Lane S, supernatant after sonification and centrifuge; Lane FT, flow through of NTA column; Lane 20 and lane 40, elution of 20mM and 40mM imidazole; Lane E, elution by 500mM; Lane M, the protein maker ladder. The molecular weight of every marker line was indicated on right.

The purity of the ASF/SF2 sample was checked by the mass spectroscopy after the raw product from NTA column was further purified by C-18 reverse phase column (Figure 3.2a). Phosphorylated RS domain samples were produced by SRPK1 phosphorylation at 30°C, after 4hours. The mass spectroscopy of the RS domain ASF/SF2 revealed that protein got 12 to 17 phosphorylation groups (Figure 3.2b). Extended reaction times (from

4 hours to 18 hours) were tested; however, no homogeneous phosphorylated sample can be reached. The mixture of species with different phosphorylation numbers can be separated by HPLC.

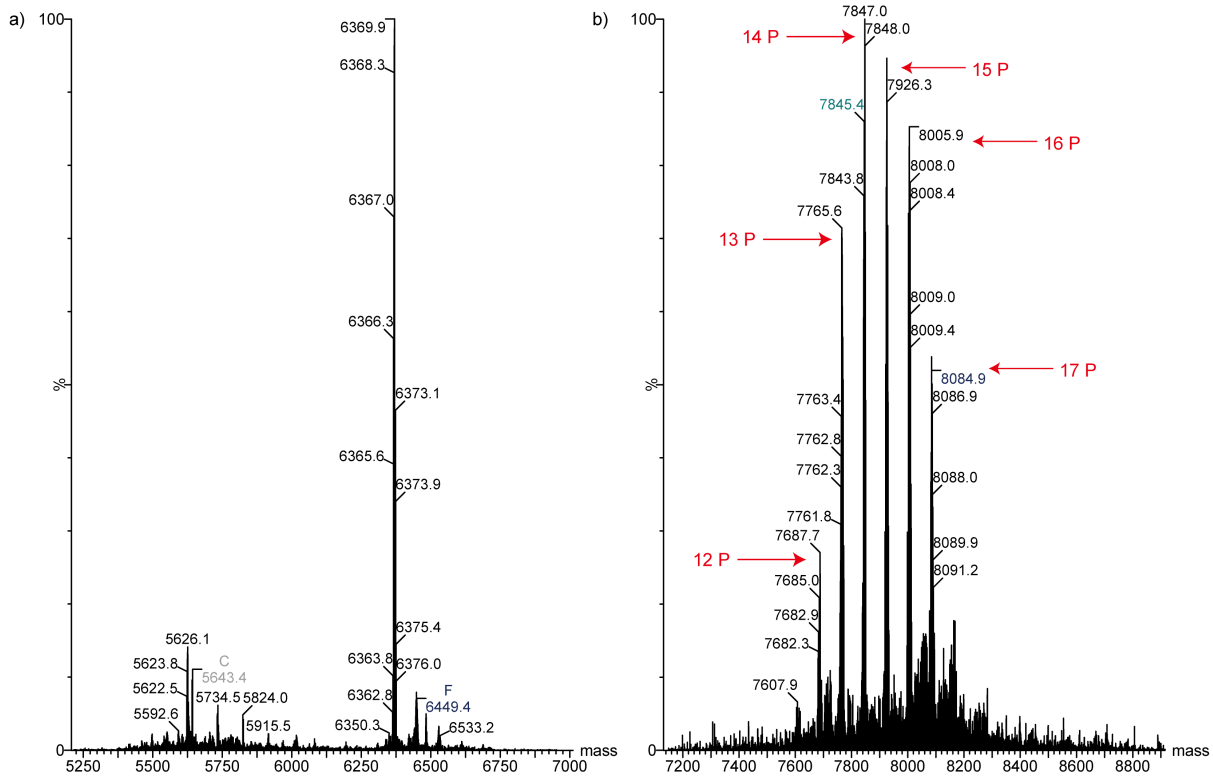


Figure 3.2 Mass spectroscopy of native and phosphorylated RS domain of ASF/SF2. a) Mass spectroscopy of the native RS domain of ASF/SF2 protein. b) Mass spectroscopy of ASF/SF2 protein phosphorylated by SRPK1 at 30°C for 4 hours, $^{15}N/^{13}C$ labeled. The peaks corresponding to different phosphates numbers were labeled in red. This is the same sample for NMR studies.

The ^{15}N-^{1}H HSQC spectrum of ASF/SF2 RS domain was recorded, as seen in Figure 3. Complete backbone assignment was not achieved due to signal overlapping and sample heterogeneity. However, some peaks can be confirmed in phosphorylated serine and arginine (RS) repeat stretches. The phosphorylated serine residues were identified by their characteristic amide proton and Cα chemical shifts. The neighboring arginine residues were identified by the connections in 6D APSY spectra and Cα chemical shifts. These peaks were labeled on the HSQC, and residues in the 8 times RS repeats (204-219)

were circled by dash line (Figure 3.3). NMR parameters of this repeat region showed some distinct properties.

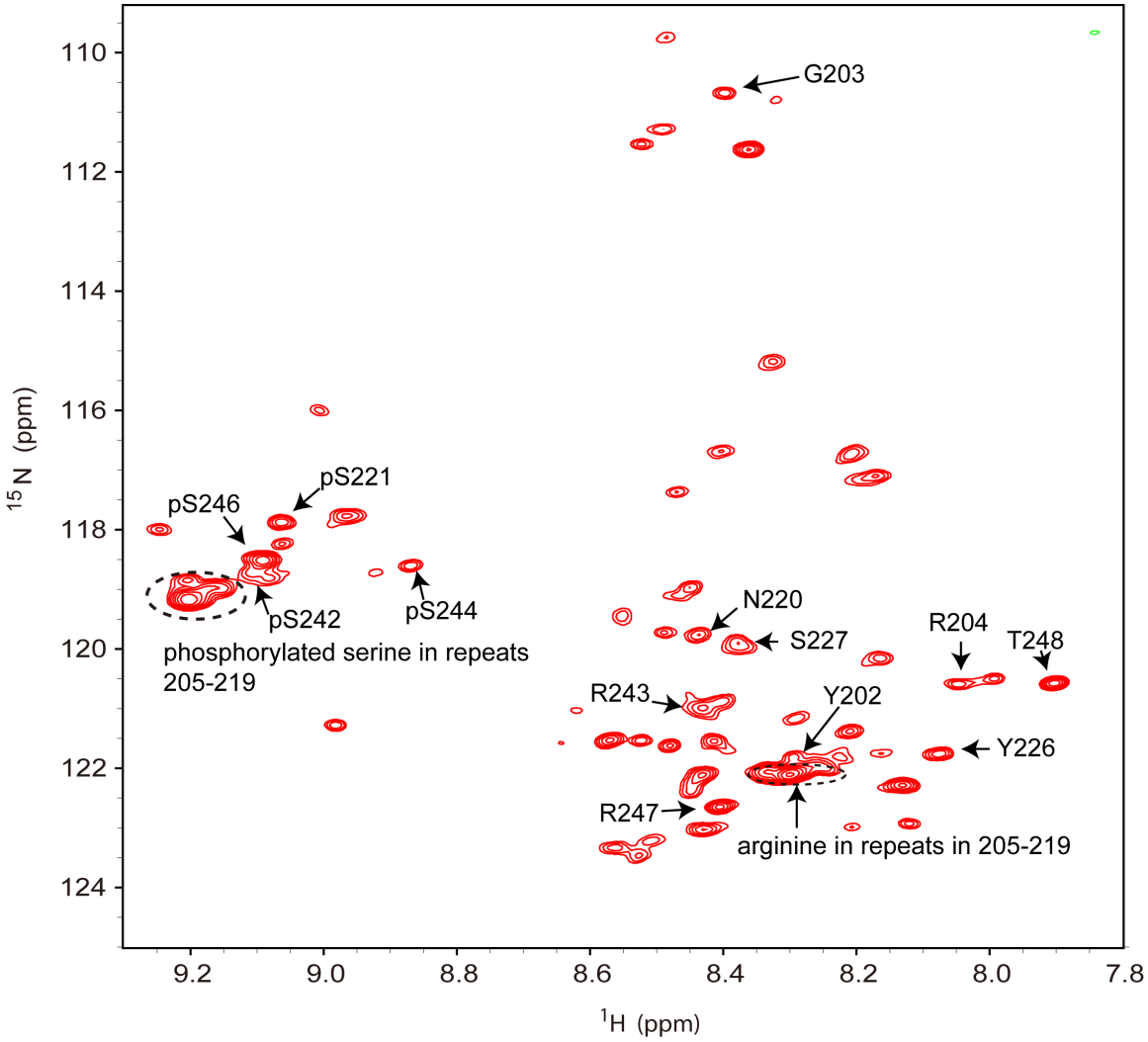


Figure 3.3 ^{15}N-^{1}H HSQC spectrum of phosphorylated RS domain of ASF/SF2, at pH7.0, 288K. The peaks that can be confirmed in the arginine and phosphorylated serine repeat stretch are labeled.

The unphosphorylated RS domain of ASF/SF2 was shown to be disordered by circular dichroism spectra (161). Phosphorylation induced significant structural changes in this domain. The heteronuclear steady state NOE effect gives information about dynamic in pico- to nano- second range. Values of arginine residues in the phosphorylated RS repeat region are 0.50±0.05, while phosphorylated serine peaks gave the value 0.58±0.08, which

is higher than the typical values for intrinsic disordered proteins (183, 184). The repeat region was more rigid than normal disordered proteins.3-bond J coupling between amide proton and Hα, $^{3}J_{HN\text{-}H\alpha}$, dependents on backbone torsion angle ϕ. The $^{3}J_{HN\text{-}H\alpha}$ of phosphorylated serine in stretches serine/arginine repeats was 5.2±0.3Hz, while the value of arginine was 6.3 ±0.3Hz. These values are lower than the predicted random coil values (185).

Taken together, the reducing of flexibility and the preference of backbone torsion angles imply transit structure forming in this phosphorylated region.

3.2 Phosphorylation induced structural transition in ASF/SF2 (200-219) peptide

The structural changes induced by phosphorylation in ASF/SF2 are shown in above experiments. These structural transitions can explain why the phosphorylation of ASF/SF2 can modulate its function in cells. However, the inhomogeneity of phosphorylation on the ASF/SF2 RS domain prevents further study. The NMR sample contains at least 4 species with different phosphorylation numbers. The 4 similar but slightly different proteins increased the signal overlapping severely.

Considering RS repeats in the N terminal part can partly restore functions of the whole domain in mRNA splicing (186), so we separated the arginine and serine repeat part (200-219) from the whole domain as a model peptide (Figure 3.4). The ASF/SF2 (200-219) sequence contains 16 residues of arginine/serine repeats in the N-terminal of the RS domain of ASF/SF2. That is proven to be able to at least partly restore the functions of the whole RS domain in splicing reactions. One (RS) 8 peptide which is constituted by 8 times arginine/serine dipeptide repeats, was chemically synthesized.

Figure 3.4 ASF/SF2 (200-219) peptide in RS domain. The 8 RS dipeptide repeats are labeled in red. The sequence in red was chemically synthesized as (RS)8 peptide.

3.3 Unphosphorylated ASF/SF2 (200-219) is in random coil status

The CD spectra of the (RS)8 peptide at various concentrations showed the same characteristic random coil spectra (Figure 3.5).

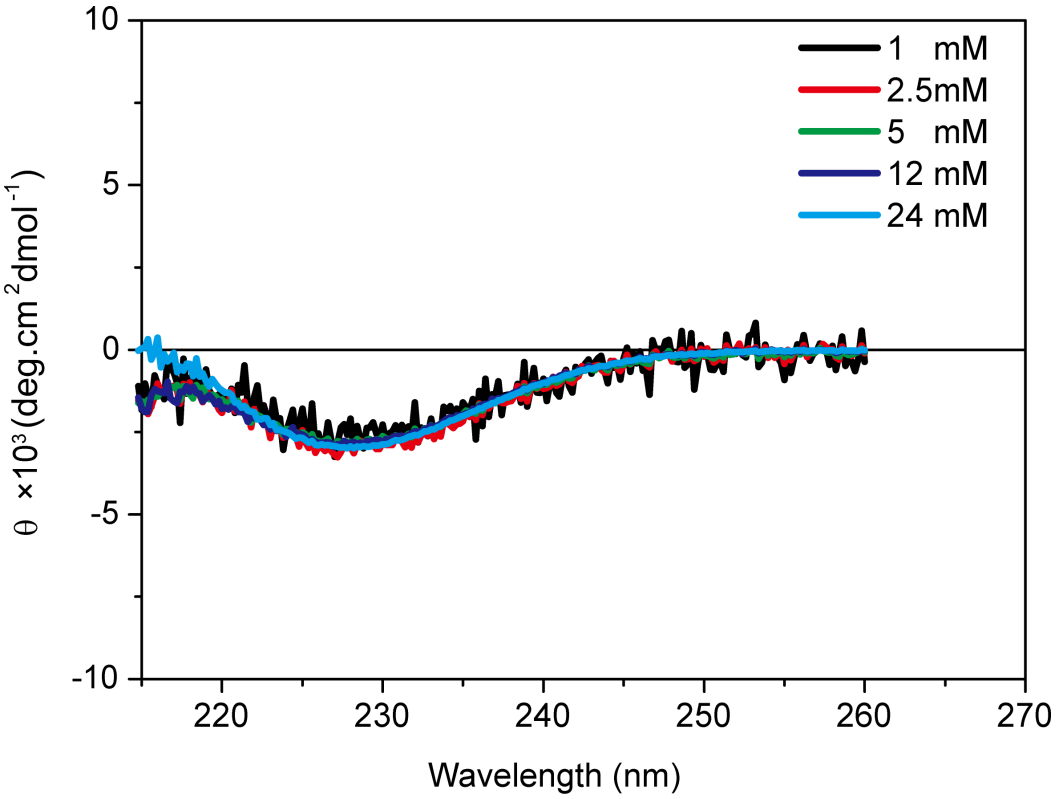


Figure 3.5 Near UV CD spectra of (RS)8 in the concentration range from 1mM to 24mM,50mM HEPES, pH6.5,300mM NaCl.

The chemical shifts of the backbone nucleus are sensitive to secondary structures, because the nuclear shielding dependences on dihedral ϕ and ψ have characteristic values for common secondary structural elements. The Cα and Cβ chemical shifts measured by ^{13}C-^{1}H HSQC are identical to predicted random coil values, which signify the peptide is fully disordered (Figure 3.6a).

To further address the possibility of the formation of any transitory structures, we put the (RS)8 peptide into 6M GuCl, which is a strong denatured condition. The peptide should be completely denatured and in a fully random coil state. On ^{13}C-^{1}H HSQC, the peaks corresponding to Cα and Cβ groups did not shift, which means the peptide is

initially disordered (Figure 3.6b). In other words, the native state has no helix conformation in the native state peptide.

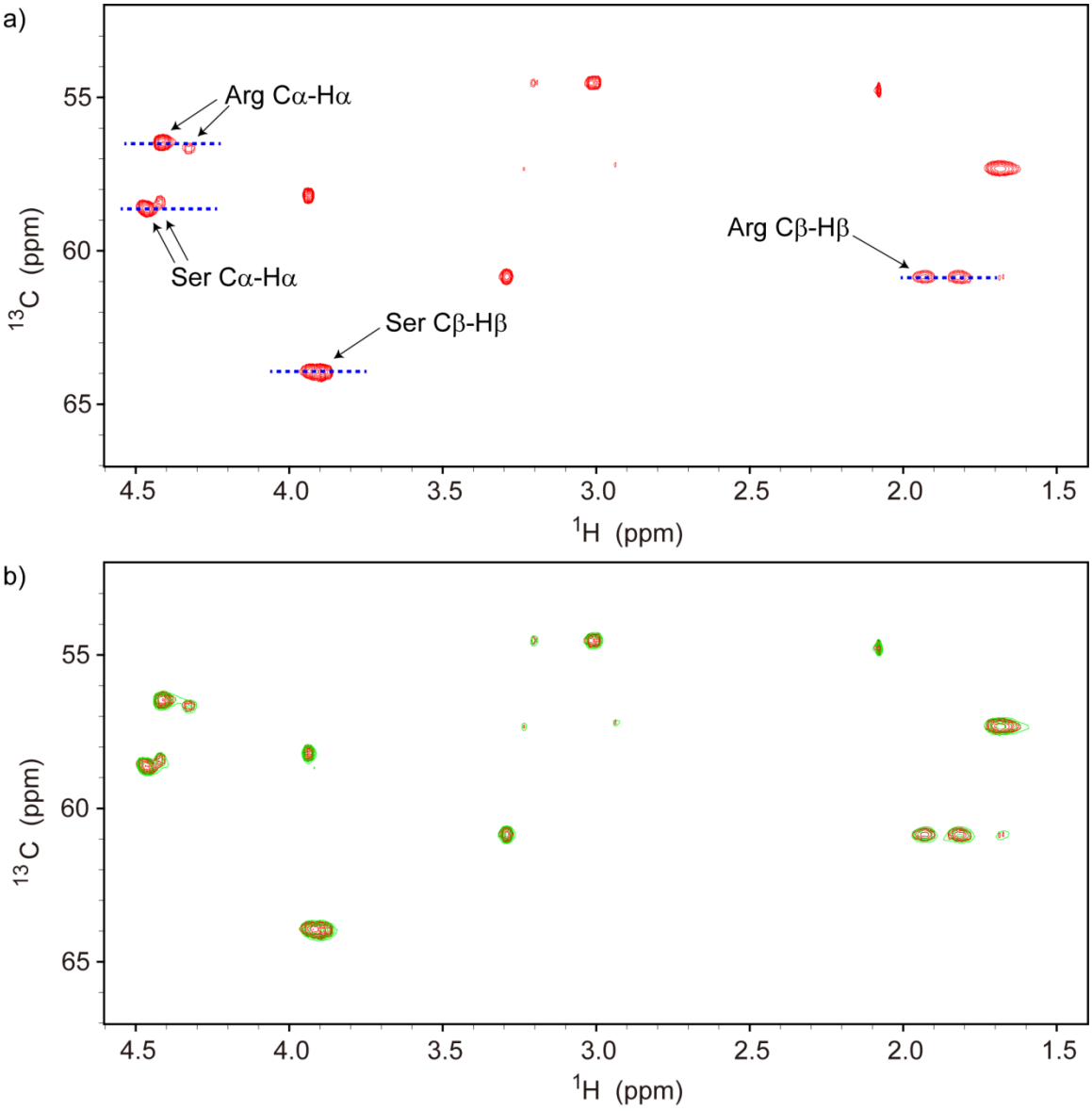
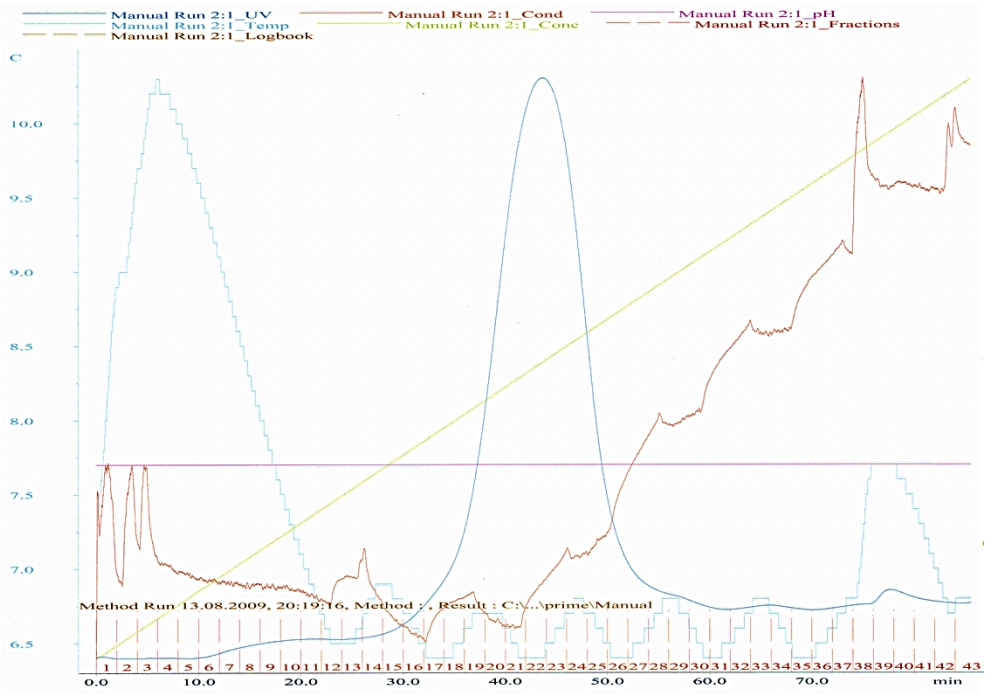


Figure 3.6 Natural abundace ^{1}H-^{13}C HSQC spectrum of (RS)8 peptide in 50mM HEPES, 100mM NaCl, pH 6.5 at 298K . a) ^{13}C-^{1}H HSQC of ASF/SF2 (200-219). Blue dash lines indicate the predicted random coil values of serine and arginine (187, 188). b) Superposition of ^{13}C-^{1}H HSQC OF native ASF/SF2 (200-219) (red) and ASF/SF2 (200-219) treated with 6M GuCl (green).

To further illustrate the structure of ASF/SF2 (200-219) by NMR, the recombinant approach of ASF/SF2 (200-219) was set up. Similar to ASF/SF2, ASF/SF2 (200-219) was expressed in *E.coli* as a fusion protein with the Z domain of staphylococcal protein A. The chromatography curves of NTA column are shown in Figure 3.7a. The fractions from the target protein peak were sampled and analyzed by SDS-PAGE (Figure 3.7b).

a)



b)

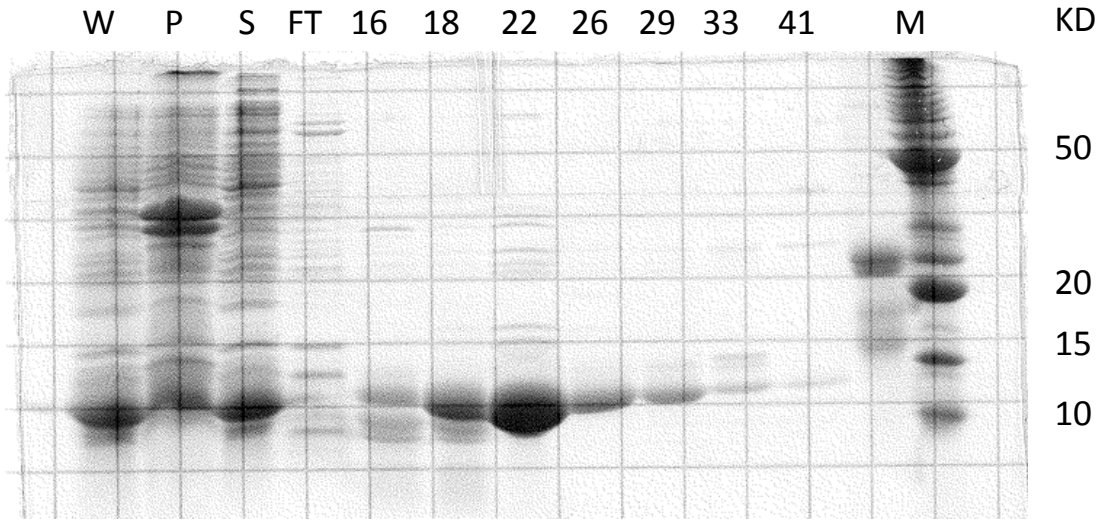


Figure 3.7 Purification of ASF/SF2 (200-219)-Z2 fusion protein. a) Chromatography for NTA column purification of ASF/SF2 (200-219)-Z2 fusion protein. The blue line is absorbance at 280nm; the green line is concentration of imidazole; the red line is conductance; the azure line is temperature. The numbers of fractions are labeled at the bottom of the picture. b) SDS-PAGE of each step and fractions in purification. Lane W, the lysis of the whole cell; Lane P, the precipitation after sonification and centrifuge; Lane S, the supernatant after sonification and centrifuge; Lane FT, flow through of NTA column; Lanes 16-41, the fractions of elusion; Lane M, marker. The molecular weight of each line is labeled on the right side.

The fractions contain target proteins, which were collected and treated with TEV enzyme to cleavage the His-tag and the fused Z domain. The cleavage results were monitored by SDS-PAGE (Figure 3.8). Premature cleavage occurred before TEV enzyme was added. The mass ratios between ASF/SF2(200-219)-Z2 protein and TEV enzyme were scanned to find the best condition, which resulted in the optimal ratio as 1:200.

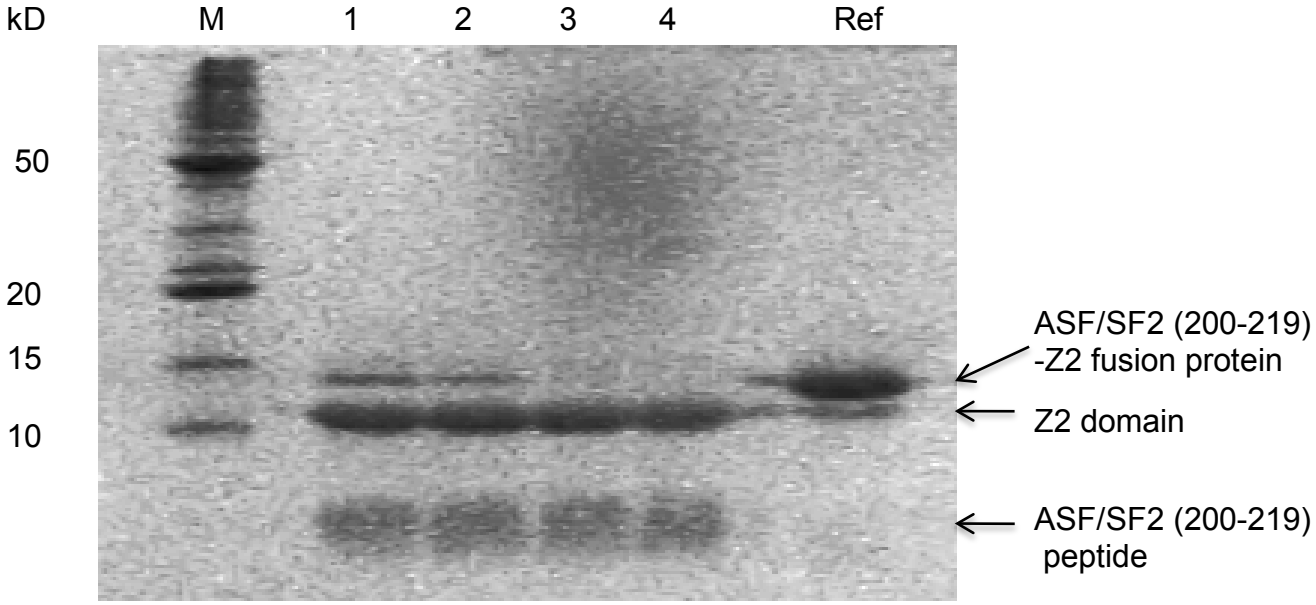


Figure 3.8 TEV cleavage of ASF/SF2 (200-219). Lane M, Marker; Lane 1-4, ASF/SF2 (200-219)-Z2 fusion protein was treated overnight at room temperature by TEV with the mass ratio 1:1000,1:750,1:500,1:200. Lane Ref, the ASF/SF2 (200-219)-Z2 fusion protein sample before cleavage.

After the cleavage, ASF/SF2 (200-219) was isolated by HPLC. Phosphorylation of ASF/SF2 (200-219) was carried out at 30°C, for 4 hours, and then checked by mass spectroscopy. Various reaction conditions were tested, including kinase substrate ratio from 1:10 to 1:200 and reaction time from 4 to 8 hours. The highest phosphorylation number, which can be reached with reasonable amount, was 6. In Figure 3.9, the mass spectroscopy results of native and phosphorylated ASF/SF2 (200-219) peptide are shown in panel a) and b).

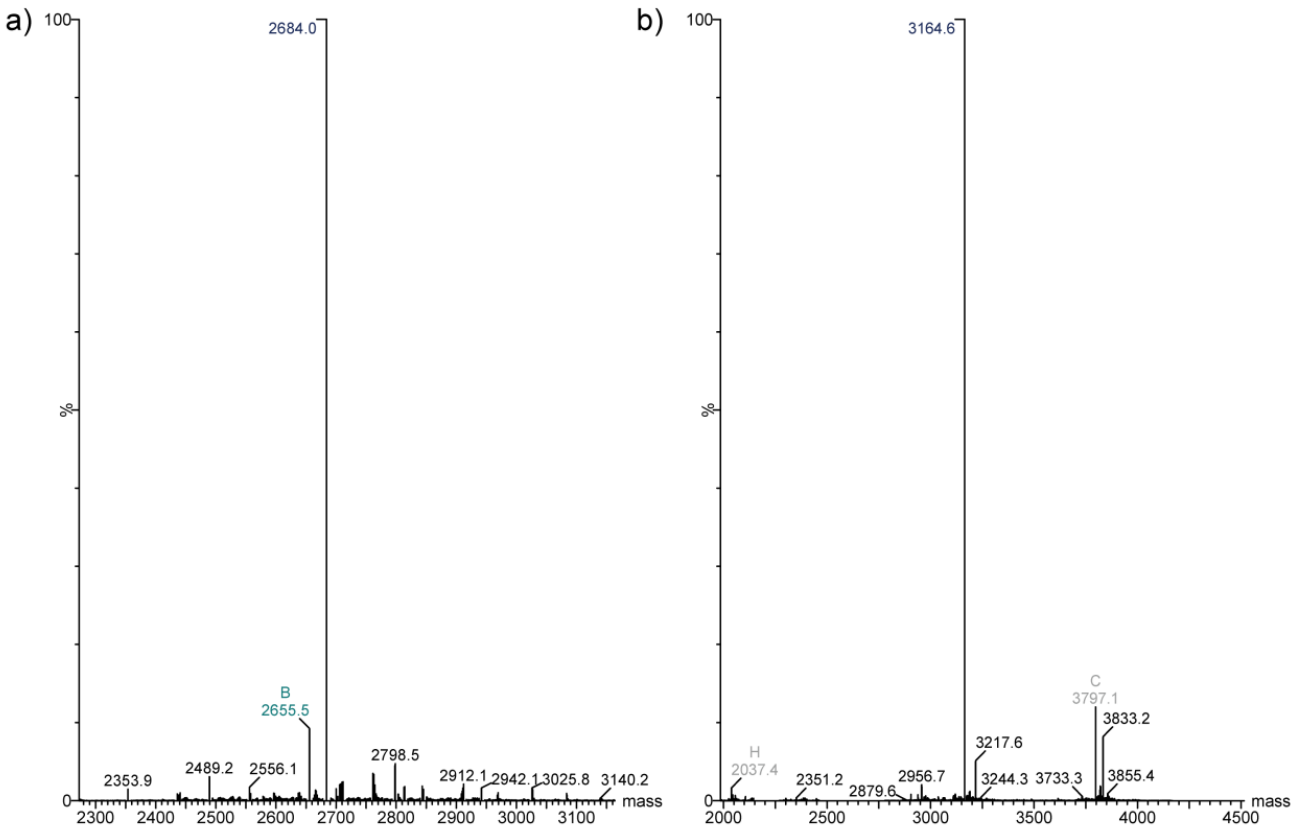


Figure 3.9 Mass spectroscopy of ASF/SF2 (200-219) in a) and major phosphorylation product of ASF/SF2 (200-219) with 6 phosphate groups in b)

On ^{15}N-^{1}H HSQC of ASF/SF2 (200-219), separate peaks belonging to residues precede RS stretch repeats, as labeled in Figure 10a. All arginine residues in repeats collapsed into one blotch except the one on each terminal, as did the serine residues. This indicates the chemical environments of the same type of residues are very similar, which normally is the case in a random coil. The heteronuclear NOE values of ASF/SF2 (200-219)

residues are all negative. It indicates fast motions on the backbone of this peptide (Figure 3.10d), which is also a feature for random coil.

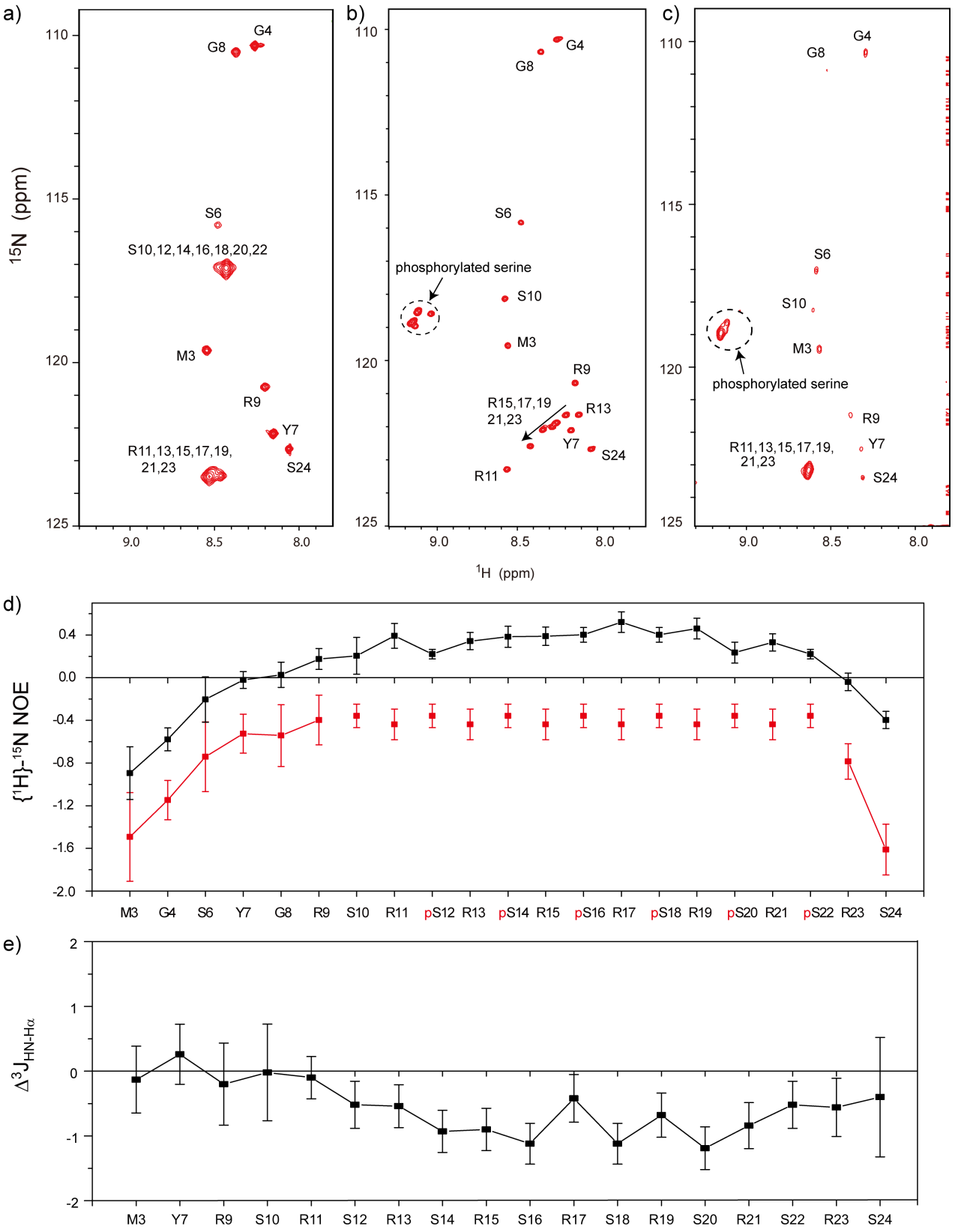

a)
b)
c)
^{15}N (ppm)
^{1}H (ppm)
G4
G8
S6
S10,12,14,16,18,20,22
M3
R9
Y7
R11,13,15,17,19,
21,23
S24
phosphorylated serine
S10
R15,17,19
21,23
R13
R11
d)
{^{1}H}-^{15}N NOE
M3 G4 S6 Y7 G8 R9 S10 R11 pS12 R13 pS14 R15 pS16 R17 pS18 R19 pS20 R21 pS22 R23 S24
e)
$\Delta^3 J_{HN-H\alpha}$
M3 Y7 R9 S10 R11 S12 R13 S14 R15 S16 R17 S18 R19 S20 R21 S22 R23 S24

Figure 3.10 Structural transition of ASF/SF2 (200-219) upon phosphorylation. a-c) ^{15}N-^{1}H HSQC of ASF/SF2(200-219) (a), phosphorylated ASF/SF2(200-219) (b), denatured phosphorylated ASF/SF2(200-219) (c). The assignments are labeled on the spectrum. d) $\{^{1}H\}$ ^{15}N Heteronuclear NOE ratios against sequences of ASF/SF2 (200-219) (red) and phosphorylated ASF/SF2 (200-219) (black). e) The different values of $^{3}J_{HN\text{-}H\alpha}$ couplings between ASF/SF2 (200-219) peptide and phosphorylated ASF/SF2 (200-219) against sequence. The errors of $^{3}J_{HN\text{-}H\alpha}$ couplings were estimated from the distribution of values measured by 3 different mixing times. Due to the signal overlapping, Heteronuclear NOE and 3J coupling values for middle part of RS repeats are averaged values. Values of R11, 13, 15, 17, 19, 21 are average and values of S10, 12, 14,16,18,20 are averaged.

3.4 Structural transition of ASF/SF2 (200-219) upon phosphorylation

^{15}N-^{1}H HSQC of phosphorylated peptide is different from its unphosphorylated form (Figure 3.10b). Backbone assignment of phosphorylated ASF/SF2 (200-219) was achieved by 7D -seq-HNCO (CA) CBCANH APSY experiments. The 6 phosphorylated serine residues are in the middle of RS repeats stretch. The last and first serine residues were not phosphorylated. The assignments are labeled on the spectrum in the Figure 3.10b.

Peaks of the residues in the RS repeat region were better spread. Especially for the arginine residues, completely isolated peaks can be observed for every Arg residue. More importantly, all residues that precede the stretch of RS repeats did not change in phosphorylated form, which means the structural changes are restricted to the RS repeat region.

In contrast to ASF/SF2 (200-219), the heteronuclear NOE values of RpS8 peptide were positive for the middle part of peptide with maximum reach 0.4 (Figure 3.10d). This indicated that upon phosphorylation, the backbone of ASF/SF2 (200-219) became rigid, though not as fully rigid as a folded protein, which normally has even higher values. 3-bond J couplings between amid proton and Hα, $^{3}J_{HN\text{-}H\alpha}$, is dependent on backbone torsion angle ϕ. $^{3}J_{HN\text{-}H\alpha}$ of RS repeats in RpS8 peptide were around 1Hz smaller than the

unphosphorylated form, moreover, this difference was continued and became more significant in the middle of RS repeats (Figure 3.10e). Residues before the RS repeat stretch did not show systematic deviation between the two forms.

In order to find out whether the phosphorylated form deviated from the random coil state, we treated it with 6M GuCl to reach a completely denatured state. In this condition, the phosphorylated ASF/SF2 (200-219) showed similar patterns as its unphosphorylated counterpart. All serine and arginine residues stack to the corresponding blotches by their residue types. Once more, the residues preceding the RS repeats region were affected little by GuCl (Figure 3.10c).

The pK value for the equilibrium between singly and doubly charged phosphate groups is ~6. Hence in the physiology pH range, the charge state of the phosphoserine side chain depends on pH. This changing in charge state made the signals observed in ^{15}N-^{1}H and ^{13}C-^{1}H HSQC of phosphorylated RS peptides also pH dependent. We carried out a pH titration of phosphorylated ASF/SF2 (200-219), from pH 4.0 to pH 8.0. At pH4.0, the peaks from the stretch of RS repeats were highly overlapped on ^{15}N-^{1}H HSQC, which was similar to its unphosphorylated form (Figure 3.11a). The peaks on ^{13}C-^{1}H HSQC, Cα and Cβ chemical shifts of arginine residues were close to random coil values (Figure 3.11b and c). With increasing pH values, the peaks gradually spread on ^{15}N-^{1}H HSQC, and secondary chemical shifts of arginine residues also increased. However, comparing the ^{13}C-^{1}H HSQC at pH 7.4 and pH 8, there was no difference observed (Figure 3.11d).

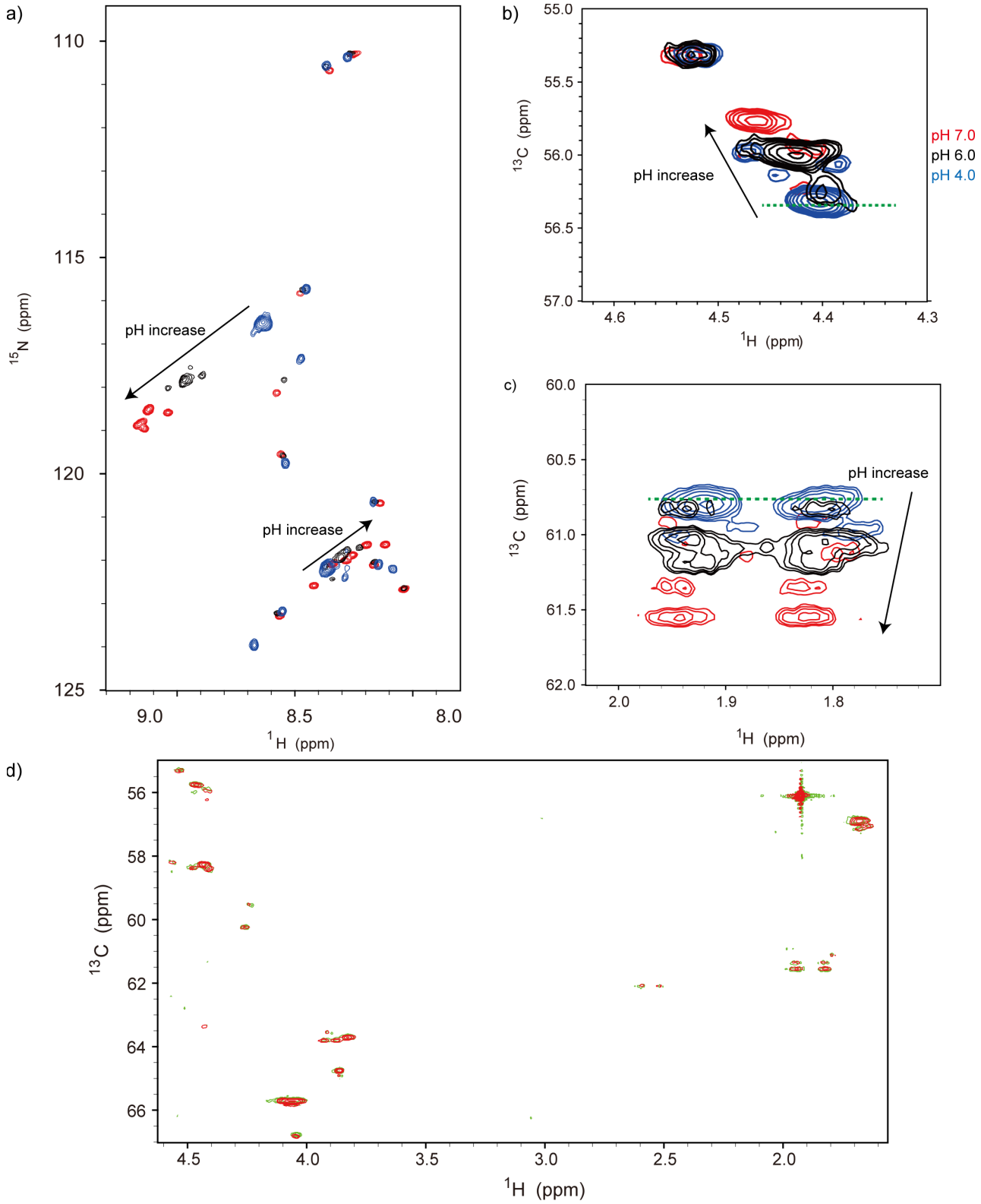


Figure 3.11 pH titration of phosphorylated ASF/SF (200-219) peptide. a) ^{15}N-^{1}H HSQC of peptide at pH 4.0, 6.0 and 7.0 b) Cα-Hα region of arginine residues on ^{13}C-^{1}H

HSQC c) Cβ-Hβ region of arginine residues on ^{13}C-^{1}H HSQC. The green dashed line in panel B and C indicates the predicted random coil values of Cα and Cβ (187, 188) d) Superposition of ^{13}C-^{1}H HSQC of RpS8 peptide at pH 7.5 (red) and pH 8.5 (green)

The secondary chemical shift, which is the deviation of chemical shifts from random coil values, is sensitive to transient secondary structure. The chemical shifts of Cα and carbonyl of phosphorylated ASF/SF2 (200-219) peptide were measured and compared to their unphosphoyrlated counterparts (Figure 3.12). In the arginine and serine repeats region, the phosphorylated peptide showed continuing negative changes, although these changes were smaller, when they were compared to the secondary chemical shifts in secondary elements of the folded protein. The changes were smaller in the center part of the repeats region, while the residues in the N and C-terminals showed fewer differences.

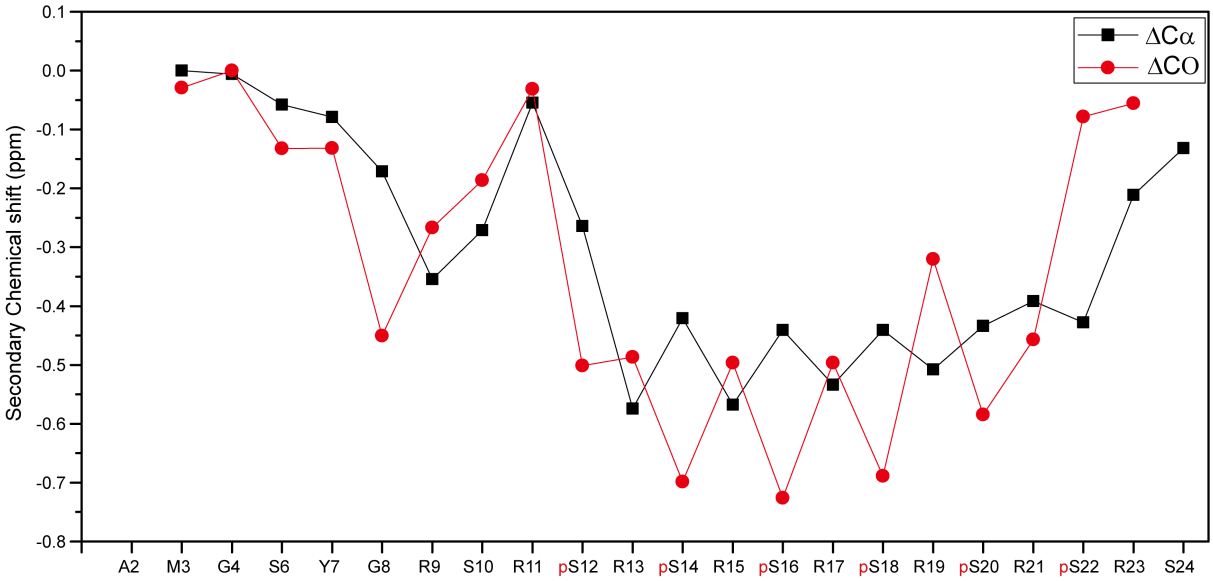


Figure 3.12 The differences of Cα (black) and carbonyl carbon (red) between native and phosphorylated ASF/SF2 (200-219) peptide.

When the protein is weakly aligned in the magnetic field, the residual dipolar couplings (RDC) can be measured, which offers useful information about structure and dynamics. To enable RDC measurements, 5% C8E5/n-octanol liquid crystalline phases were made to align native and phosphorylated ASF/SF2 (200-219) peptides. The splitting of deuterium at 288K was 38Hz and 42Hz for native and phosphorylated ASF/SF2 (200-219) peptides sample respectively. Three backbone RDCs, H-N, Hα-Cα and Cα-Hα, were

measured. The values of native and phosphorylated ASF/SF2 (200-219) peptides were plotted against sequence in the Figure 3.13.

Because the peaks of the residues between S10 and R23 of native ASF/SF2(200-219) peptide collapsed into two peaks, one for all the serine residues and the other for all the arginine residues (Figure 3.10a), the values of this region are actually the averaged values for all serine or all arginine residues in the region. This applies to heteronuclear NOE, 3-bond J couplings, chemical shifts and RDC data. The details are labeled in legends.

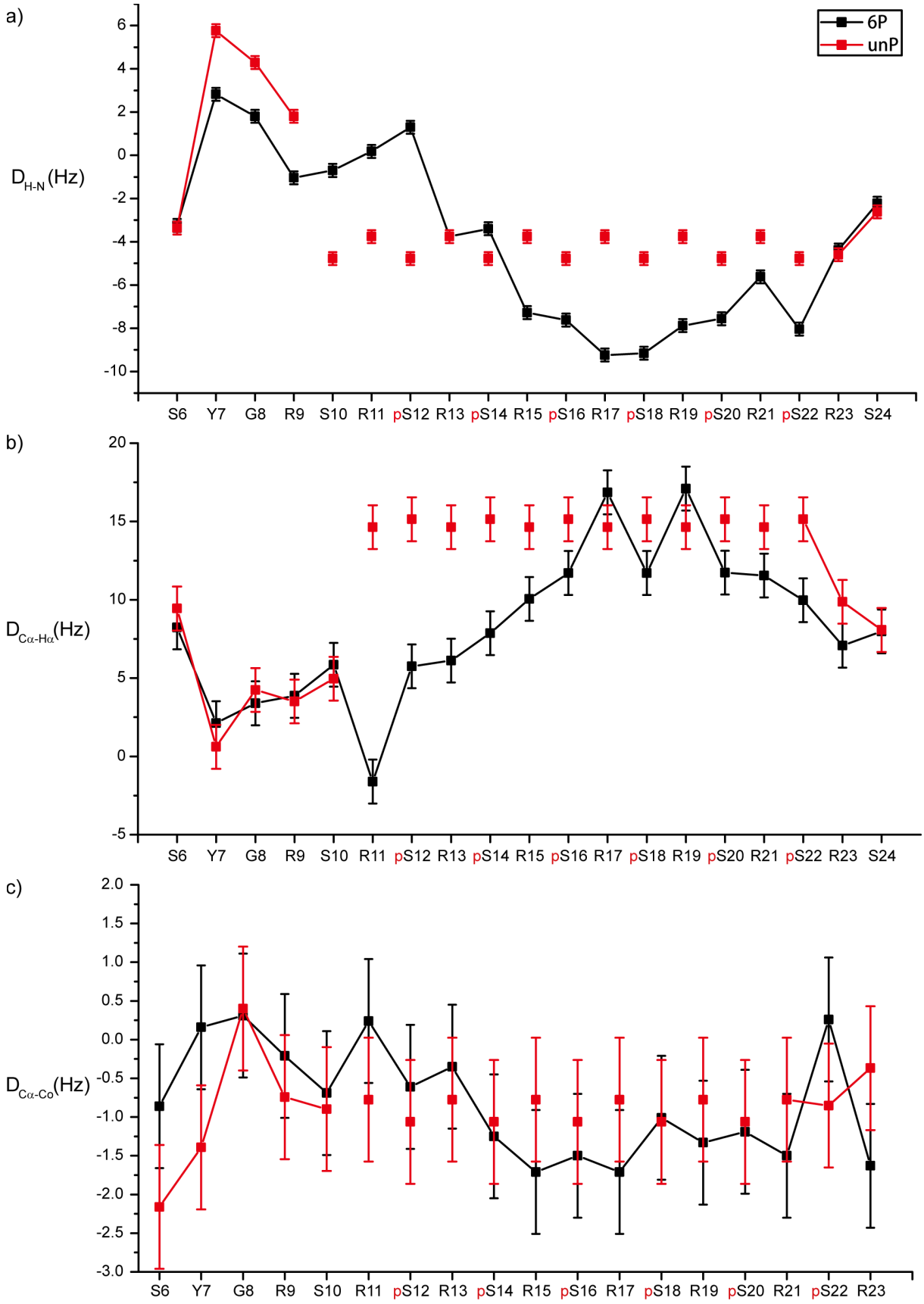

a)
6P
unP
D_{H-N} (Hz)
6
4
2
0
-2
-4
-6
-8
-10
S6 Y7 G8 R9 S10 R11 pS12 R13 pS14 R15 pS16 R17 pS18 R19 pS20 R21 pS22 R23 S24
b)
$D_{C\alpha-H\alpha}$ (Hz)
20
15
10
5
0
-5
S6 Y7 G8 R9 S10 R11 pS12 R13 pS14 R15 pS16 R17 pS18 R19 pS20 R21 pS22 R23 S24
c)
$D_{C\alpha-Co}$ (Hz)
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
-3.0
S6 Y7 G8 R9 S10 R11 pS12 R13 pS14 R15 pS16 R17 pS18 R19 pS20 R21 pS22 R23

Figure 3.13 Backbone RDCs of ASF/SF2(200-219) peptides. H-N (a), Hα-Cα (b) and Cα-CO(c) RDC values are plotted against sequence. The data of native ASF/SF2 (200-219) is in red while phosphorylated peptide data is in black. The coupling values of unphosphorylated form were scaled up base on ^{2}H splitting values. In unphosphorylated form, due to the signal overlapping, values for middle part of RS repeats are averaged values. In all three couplings data, R11, 13, 15, 17, 19, 21 are averaged. Values of S10, 12, 14, 16, 18, 20 are averaged in H-N couplings. Values of S12, 14, 16, 18, 20 are averaged in Cα-Co and Cα-Hα couplings. The errors were estimated by the line widths and signal noise ratios.

The other two RDC, N-CO and HN-CO, were also measured. However, the uncertainty in the data was too high due to the fact that RDC values were small and due to the broad line width of the aligned sample. The NC' and HC'-RDC's were measured twice. The RMSD values between two couplings datasets were 0.6Hz for H-CO and 1.3Hz for N-CO couplings. These RMSD sizes were comparable to the RDC values. These two RDC datasets were not used in further study.

The heteronuclear NOE values clearly indicated the unphosphorylated ASF/SF2 (200-219) was fully disordered, while phosphorylated ASF/SF2 (200-219) is structured but the structure is still partially dynamic. Hence, MD simulation was incorporated to get the structure of the phosphorylated ASF/SF2 (200-219) peptide.

3.5 Structural ensembles representation of RS repeats peptide[1]

3.5.1 Structural ensembles from Molecular Dynamics (MD) simulations

To characterize the structure of phosphorylated RS repeats, a set of unbiased MD simulations was carried out. Peptides of three different topologies were considered for the simulations: wild type RS, RpS (protonation PO_4^-) and RpS (protonation PO_4^{2-}). For every case, 20 independent MD simulations starting from random structure were conducted for 52 ns employing Amber99sb* force field (189, 190). The first 2 ns were used for equilibration and the rest of the 50ns were used for analysis. In total, simulation time of each case reached 1μs.

The resulting MD trajectories were subjected to validation by comparing the experimental data to the values calculated from trajectories. The NMR observables such as RDCs and J couplings were calculated from trajectories by PALES(191) and Karpuls equations. The back-calculated values were then compared to the experimental values.

The structures in the trajectories are dominated by random coil, turn and bend, in agreement with small secondary chemical shifts. However, the back-calculated RDC and J coupling did not match the experimental data. RDCs (NH, HC, CαC') and ^{3}J couplings (HN-Hα) calculated directly from the MD ensembles of the phosphorylated peptide showed high Q(192) and RMSD values when compared to the experimental NMR measurements: QRDC_NH=0.90, Q_{RDC_HC}=0.78, Q_{RDC_caco}=0.92, $RMSD_{3J_HN\text{-}H\alpha}$=0.97 Hz. The high Q values showed that the unbiased MD simulation failed to generate one ensemble representing phosphorylated ASF/SF2 (200-219) structures in solution.

To create a representative structural ensemble which matches the experimental observations a sub-ensemble selection method was therefore employed. RDCs are sensitive to structure and dynamic of proteins, which were therefore used in the study of ensembles of ubiquitin(193) and IDPs (194). In current study, three sets of RDC values (NH, HC, CαC') and a set of $^3J_{HN\text{-}H\alpha}$ couplings were used in the ensemble selection

[1] The MD simulation and ensemble selection part is carried out by Vytautas Gapsys, Research group of computational bimolecular dynamics, Max Planck Institute for biophysical chemistry. For the experimental details, please refer to the thesis of Vytautas Gapsys.

procedure. A custom built sub-ensemble selection procedure that combines a Monte-Carlo search with brute-force scanning was used for the selection of optimal ensembles. The sizes of ensembles varied from 5 to 100 structures, with an increment of 5 peptides. For every fixed ensemble size, 10 independent selections were performed the Q values of which are illustrated in the Figure 3.14.

Increasing the ensemble size enables the structural ensembles to better match NMR constraints, as indicated by the decrease in the Q values for the RDCs and RMSD for the ^{3}J coupling (Figure 3.14). However, selecting a large ensemble increases the risk of overfitting. As the decrease in the Q values and RMSDs saturated at 30 structures for the phosphorylated peptide, this ensemble size was used for further analysis, for both native and phosphorylated peptides. It is important to note that no single structure could accurately reproduce experimentally measured couplings. The single structure from the pool of MD simulation trajectories of phosphorylated ASF/SF2 (200-219) which best matched NMR observables showed high Q and RMSD values: Q_{RDC_NH}=1.34, Q_{RDC_HC}=1.08, Q_{RDC_caco}=1.12, $RMSD_{3J_HN\text{-}H\alpha}$=1.29 Hz. This matched observed heteronuclear NOE values of phosphorylated ASF/SF2 (200-219), which indicated the phosphorylated peptide became structured but was not full rigid.

The wild type and phosphorylated structural ensembles were compared by calculating the Jensen-Shannon divergence between the smoothed 2D projections of the ensembles in a space defined by the principal components of the peptide internal distances between the Cα atoms. For every ensemble, the selection procedure was carried out 10 times, which resulted in 10 ensembles for the native peptide and 10 ensembles for the phosphorylated form. The Jensen-Shannon divergences were calculated between 2 randomly selected ensembles. For divergences of ensembles within the same type, 45 pairwise Jensen-Shannon divergences were calculated. For divergences between native ensembles and ensembles of the phosphorylated form, 100 pairwise Jensen-Shannon divergences were calculated. The calculated divergences were then averaged. The difference between the unphosphorylated and RpS ensembles is significantly larger than the internal differences within each ensemble (Figure 3.15), illustrating a significant difference between the unphosphorylated and phosphorylated ensembles.

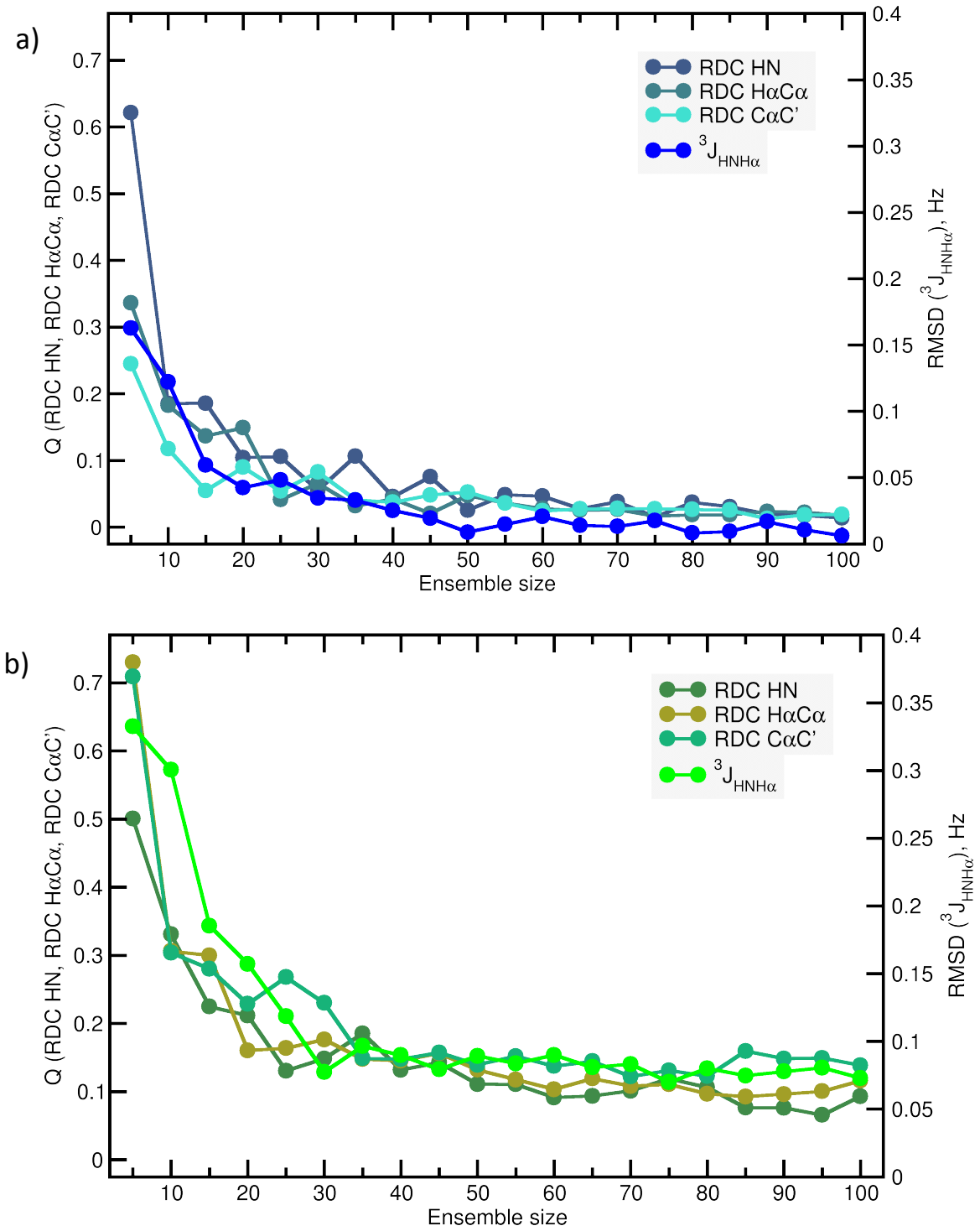


Figure 3.14 Q values change with the sub-ensemble size of phosphorylated (a) and native (b) ASF/SF2 (200-219) peptide.

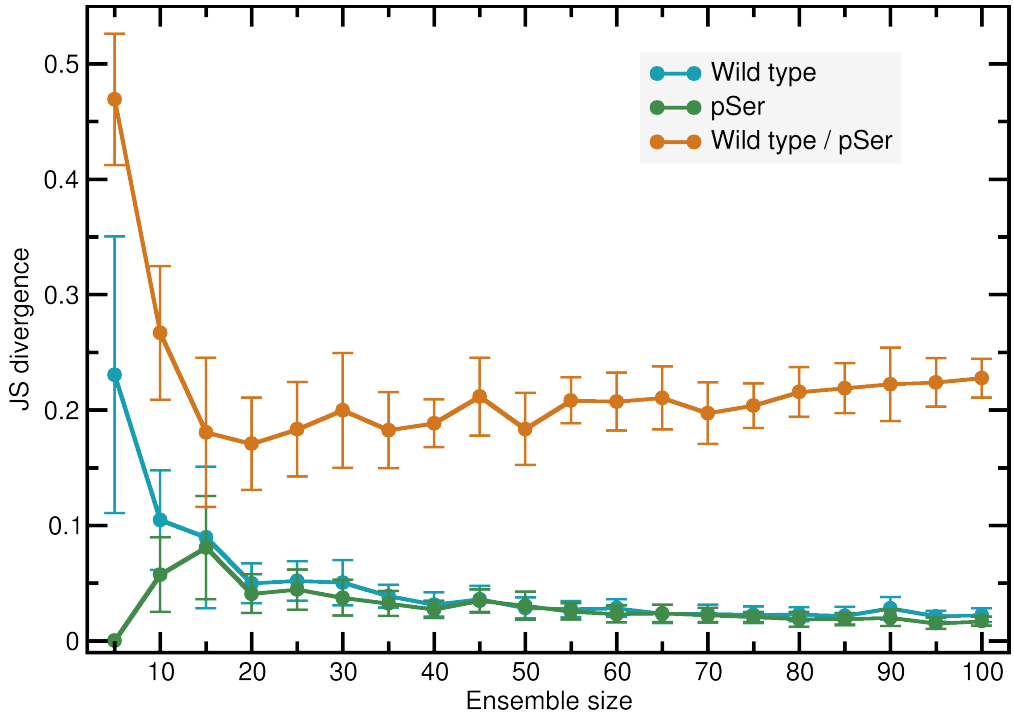
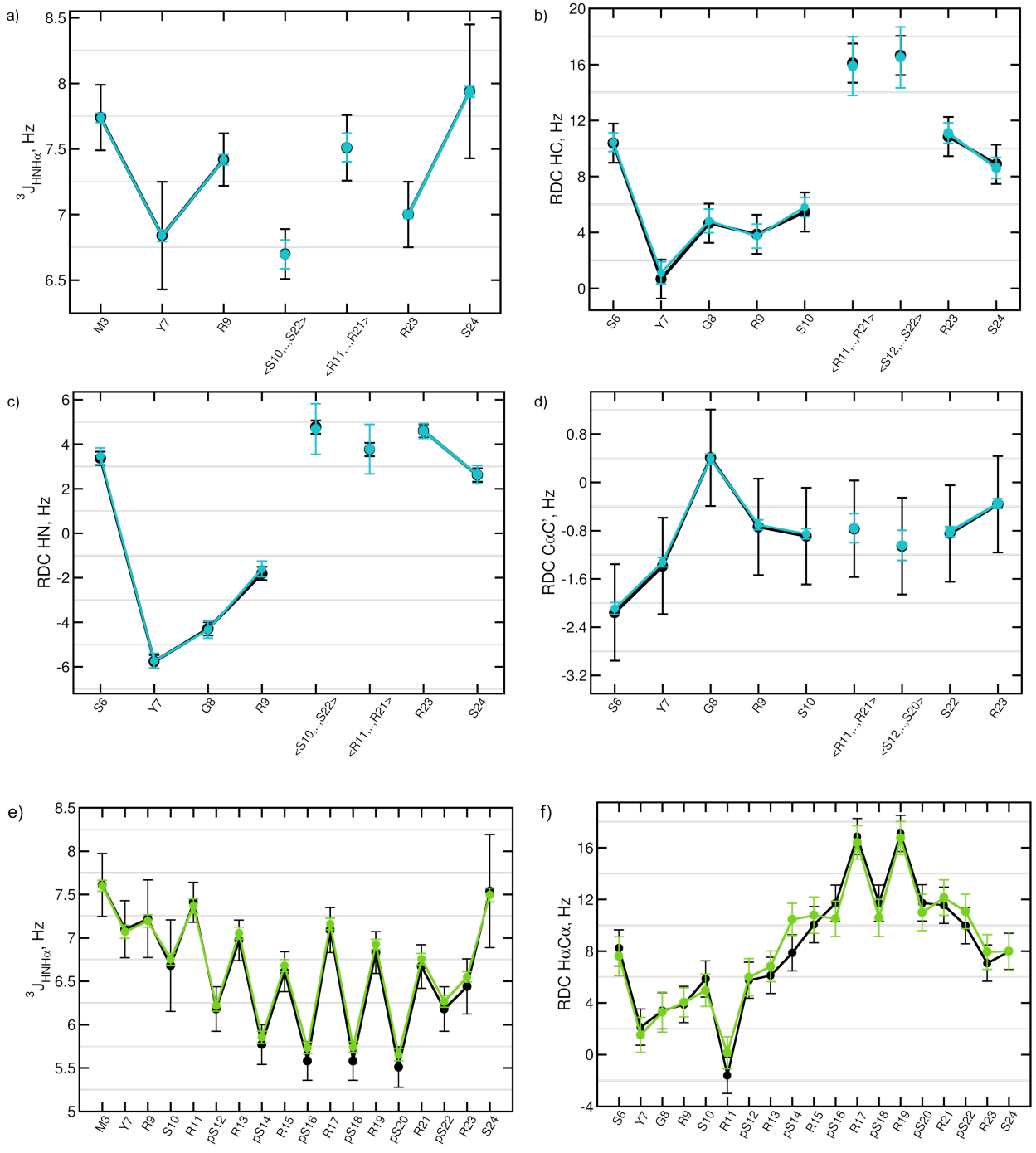


Figure 3.15 The Jensen-Shannon divergences between ensembles within the same type (native or phosphorylated) and between native and phosphorylated ensembles are plotted against ensemble sizes.

With the determined optimal ensemble size of 30 structures, 100 independent selection rounds were separately performed for both the wild type and phosphorylated peptides separately. The obtained 100 ensembles were further used as the structural ensembles for the native and phosphorylated ASF/SF2 (200-219) peptides. The values calculated from ensembles were compared to the experimental values (Figure 3.16).


a)
$^3J_{HNH\alpha}$, Hz
M3
Y7
R9
<S10...S22>
<R11...R21>
R23
S24
b)
RDC HC, Hz
S6
Y7
G8
R9
S10
<R11...R21>
<S12...S22>
R23
S24
c)
RDC HN, Hz
S6
Y7
G8
R9
<S10...S22>
<R11...R21>
R23
S24
d)
RDC CαC', Hz
S6
Y7
G8
R9
S10
<R11...R21>
<S12...S20>
S22
R23
e)
$^3J_{HNH\alpha}$, Hz
M3
Y7
R9
S10
R11
pS12
R13
pS14
R15
pS16
R17
pS18
R19
pS20
R21
pS22
R23
S24
f)
RDC HαCα, Hz
S6
Y7
G8
R9
S10
R11
pS12
R13
pS14
R15
pS16
R17
pS18
R19
pS20
R21
pS22
R23
S24

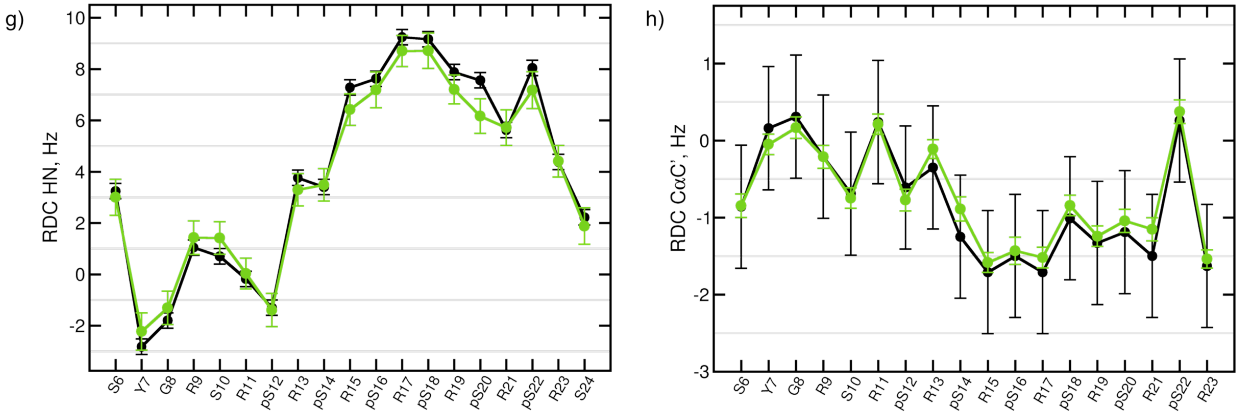


Figure 3.16 Experimental and back-calculated values for the NMR observables used in the selection. Both experimental and back-calculated values of $^{3}J_{HN\text{-}H\alpha}$ (a), Hα-Cα (b), HN-N(c), Cα-C' (d) RDC values from native ASF/SF2 (200-219) peptide are plotted. The same datasets, $^{3}J_{HN\text{-}H\alpha}$ (e), Hα-Cα (f), HN-N (g), Cα-C' (h) RDC values, from phosphorylated ASF/SF2 (200-219) peptide are plotted. Experimental values are all in black while the back-calculated values of native ASF/SF2 (200-219) are in azure whereas the phosphorylated ASF/SF2 (200-219) back-calculated values are in cyan.

3.5.2 Validation of the structural ensembles

To cross-validate the structural ensembles for the native (RS) and phosphorylated (RpS)ASF/SF2 (200-219) peptides, a number of NMR observations that were not included in the ensemble selection, were back-calculated from the selected ensembles and compared to experimental values.

The backbone conformations were validated by $^{1}J_{H\alpha\text{-}C\alpha}$, $^{1}J_{C\alpha\text{-}C\beta}$ and Chemical shifts. The side chain rotamers were validated by $^{3}J_{NC\gamma}$ and $3J_{COC\gamma}$.

For the $^{1}J_{H\alpha\text{-}C\alpha}$ coupling calculations the Karplus equation was re-paremeterized due to an observed discrepancy between values reported in the literature (36, 168) (for details please refer to the thesis of Vytautas Gapsys). Considering the accuracy of the equation, $^{1}J_{H\alpha\text{-}C\alpha}$ couplings back-calculated from the phosphorylated RS repeats ensembles are in close agreement with the experimental measurements of RMSD=0.86 Hz. The wild type ensemble also showed a good match with RMSD=0.99 Hz from the NMR values.

The back-calculated $^{1}J_{C\alpha-C\beta}$ coupling was also in agreement with the experimental measurements: for the native RS repeats ensemble RMSD=1.29 Hz and for the phosphorylated ensemble RMSD=1.21 Hz. The larger RMSD values may be due to an offset in the Karplus equation of $^{1}J_{C\alpha-C\beta}$ which took random coil values from the couplings of free amino acid which is not identical to residues in intrinsically disordered proteins (195).

In the above two calculations, random coil value of serine was used for phosphorylated serine residues, which also introduced systematic error.

The calculated chemical shift values for the Cα and C' atoms showed high Pearson correlation (>0.9) for both, the wild type and phosphorylated ensembles (Table 3.1). It must be noted that chemical shift prediction for phosphoserine is complicated since the empirical predictors are not trained against phosphorylated amino acids. This accounted to the systematic offsets in phosphorylated ensembles.

		Cα		CO	
		Correlation	RMSD (ppm)	Correlation	RMSD (ppm)
SHIFTX2	native ASF/SF2(200-219)	0.99	0.85	0.93	0.68
	phosphorylated ASF/SF2(200-219)	0.99	0.93	0.97	0.70
SPARTA+	native ASF/SF2(200-219)	0.99	0.62	0.97	0.64
	phosphorylated ASF/SF2(200-219)	0.99	0.64	0.94	0.37

Table 3.1 Pearson correlation and RMSD values between experimental data and back-calculated by two software.

Be awarded that the direct replace of phosphorylated serine with serine may be not valid in this case, because phosphorylated serine residue is different from serine in physical and chemical properties. It is hard to distinguish the difference between two states is the direct effects of phosphorylation groups or the effect of structure changes induced by phosphorylation. Due to phosphorylated serine, chemical shifts analysis of Cα and Cβ were not as useful as normal case. The random coil chemical shits of phosphorylated serine are not well established. There is one study use GGXGG peptide to measure random coil values (196). There is not many phosphorylated serine in NMR

database to train empirical predictors. The commonly used chemical shifts software, such as SPATA+, SHIFX2 cannot predict chemical shifts of phosphorylated serine.

Even with this systematic error introduced by phosphorylated serine, the RMS Deviations in our study are smaller than the values reported in the papers introduced the software (197) (198) ,and also comparable to the RMSD in one similar study of IDP ensembles (0.77 ppm for N) (199).

The RDC of H-C' and N-C' were also recorded, but they cannot be used due to the high uncertainty of the experimental values. The size of these 2 RDCs are too small, while the signals of aligned sample are too broad, which makes the accurate measurement unfeasible.

Because the selection of structures was only based on the backbone NMR observables, the side chain rotamers are also compared to validate the ensembls. A considerable agreement between the back-calculated and measured $^{3}J_{NC\gamma}$ (RMSD=0.19 Hz) and $^{3}J_{CoC\gamma}$ (RMSD=0.23 Hz) couplings, which strongly depend on the side-chain χ_1 angle, was observed.

The experimental $^{3}J_{N\text{-}C\gamma}$ values were smaller than the back-calculated values, which are a consequence of systematic errors of the method. There were two sources of systematic errors, passive couplings and T1 relaxation of aliphatic carbon as discussed in the reference (177). The pulse used to flipped Cγ nucleus also flipped the Cβ, so both $^{3}J_{N\text{-}C\gamma}$ and $^{2}J_{N\text{-}C\beta}$ were effective during the mixing time. This small and unknown $^{2}J_{N\text{-}C\beta}$ induced uncertainty in experiments. In this study, the $^{2}J_{N\text{-}C\beta}$ were assumed to be 0.3Hz, the average of two values suggested in paper (177). The second source of error comes from the flipping of aliphatic carbons during the ^{15}N dephasing and rephrasing time. Due to this relaxation, this method underestimated the coupling values 5-15%. In the experiments measuring $^{3}J_{CO\text{-}C\gamma}$, a more selective flip pulse was applied on Cγ without distributing Cβ. And the mixing time was only half of its counterpart in $^{3}J_{N\text{-}C\gamma}$ experiments. Because of these two reasons, the experimental $^{3}J_{CO\text{-}C\gamma}$ values were less affected by errors and matched better with predicted values.

The experimental and back-calculated $^{1}J_{H\alpha\text{-}C\alpha}$, $^{1}J_{C\alpha\text{-}C\beta}$ and chemical shifts of native and phosphorylated ASF/SF2 (200-219) peptides are plotted in Figure 3.17 and 3.18. The experimental and back-calculated values of two J couplings related to χ1 angles, $^{3}J_{N\text{-}C\gamma}$ and $^{3}J_{CO\text{-}C\gamma}$, are plotted in Figure 3.19.

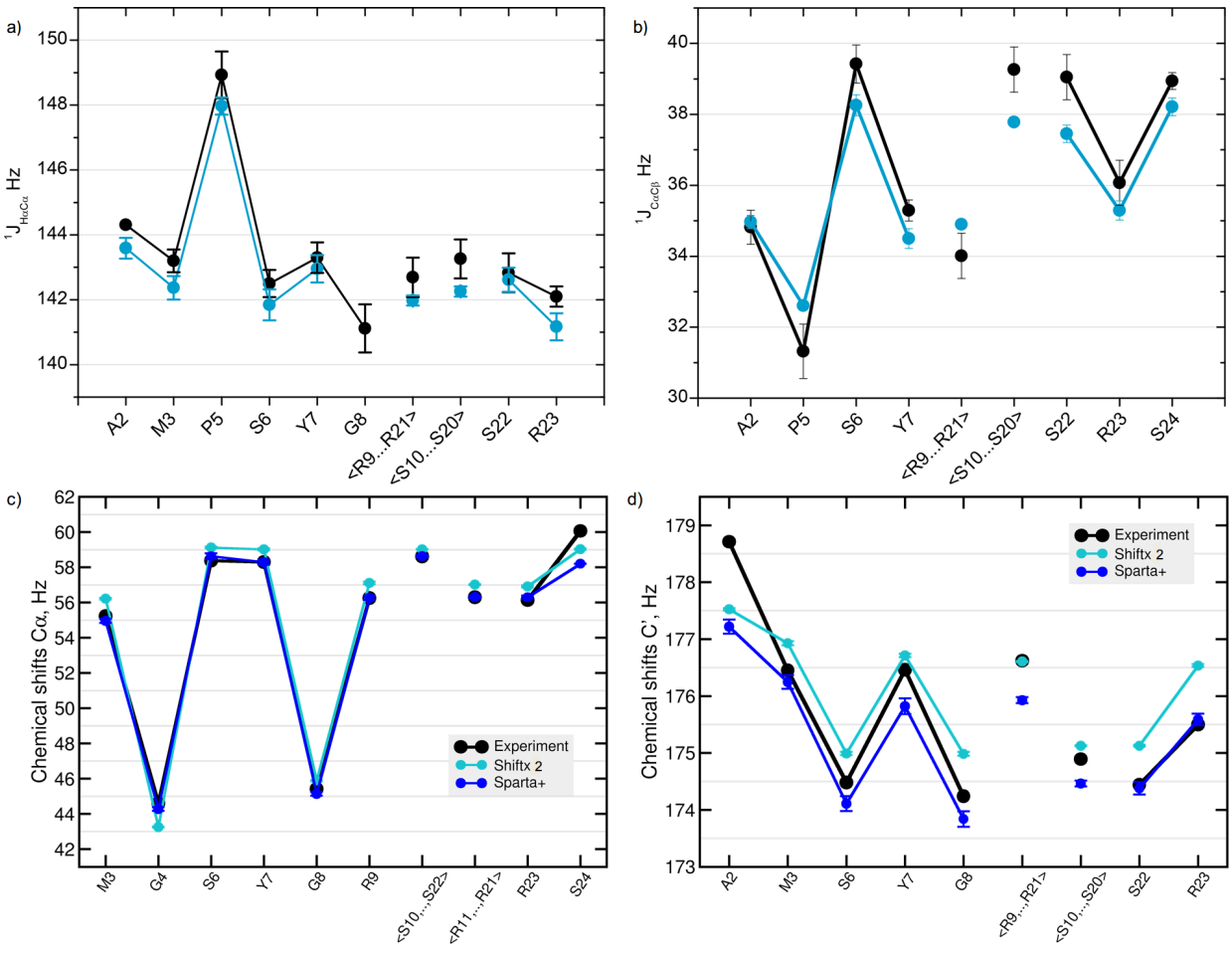


Figure 3.17 Experimental and back-calculated $^{1}J_{H\alpha\text{-}C\alpha}$ (a), $^{1}J_{C\alpha\text{-}C\beta}$ (b) and chemical shifts (c)-(d) of native ASF/SF2 (200-219) peptide are plotted against sequence. Experimental data is in back, while back-calculated $^{1}J_{H\alpha\text{-}C\alpha}$ and $^{1}J_{C\alpha\text{-}C\beta}$ are in azure. Chemical shifts predicted by Shiftx2(198) are plotted in azure and chemical shifts predicted by Sparta+(197) are plotted in blue. The errors of back-calculated couplings are the standard deviations of values from 100 ensembles. The error of experimental data was estimated by intensities and line width of signals.

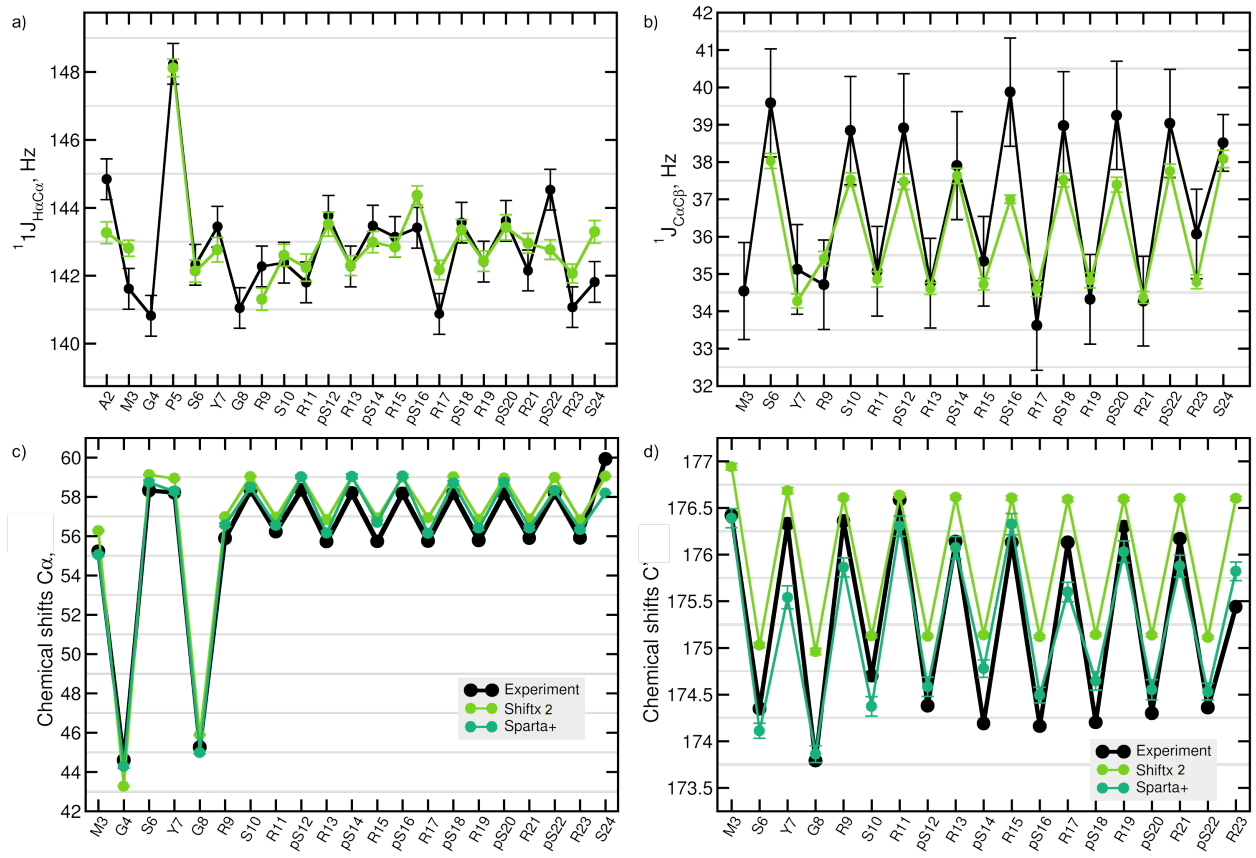


Figure 3.18 Experimental and back-calculated $^{1}J_{Hα-Cα}$ (a), $^{1}J_{Cα-Cβ}$ (b) and chemical shifts (c)-(d) of the phosphorylated ASF/SF2 (200-219) peptide are plotted against sequence. Experimental data is in black, while back-calculated $^{1}J_{Hα-Cα}$ and $^{1}J_{Cα-Cβ}$ are in cyan. Chemical shifts predicted by Shiftx2(198) are plotted in cyan and chemical shifts predicted by Sparta+(197) are plotted in green. The errors of back-calculated couplings are the standard deviations of values from 100 ensembles. The error of experimental data was estimated by intensities and line width of signals.

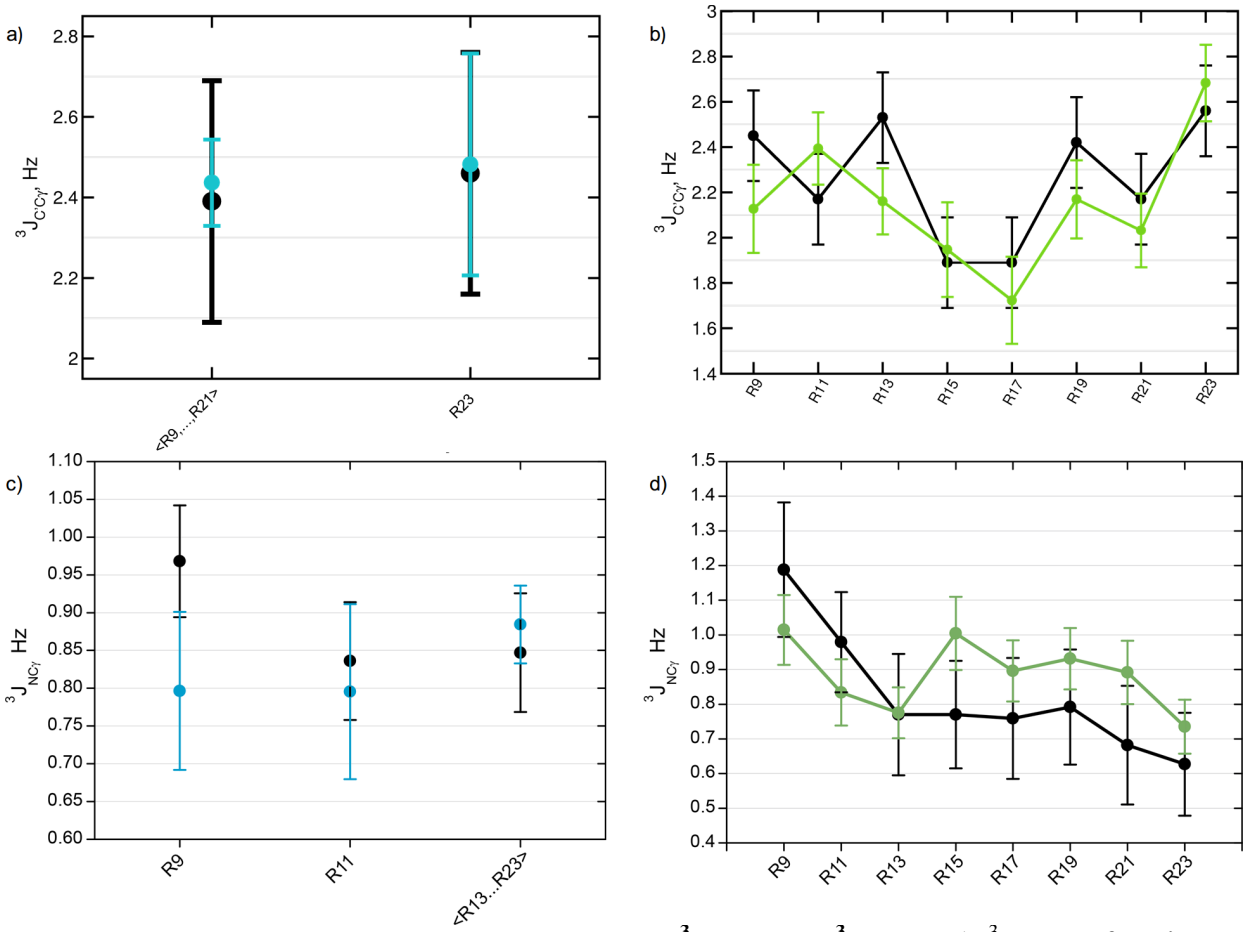


Figure 3.19 Experimental and back-calculated $^3J_{N\text{-}C\gamma}$ and $^3J_{CO\text{-}C\gamma}$. a) $^3J_{CO\text{-}C\gamma}$ of native ASF/SF2 (200-219). c) $^3J_{N\text{-}C\gamma}$ of native ASF/SF2 (200-219). Experimental values are in black, back-calculated values are in azure. b) $^3J_{CO\text{-}C\gamma}$ of phosphorylated ASF/SF2 (200-219) peptide. d) $^3J_{N\text{-}C\gamma}$ of phosphorylated ASF/SF2 (200-219) peptide. Experimental values are in black, back-calculated values are in cyan. The error of experimental data was estimated by intensities and line width of signals.

3.5.3 Backbone conformations

100 ensembles containing 3000 structures were superimposed using the newly developed superpositioning algorithm by minimizing the variance and RMSDs between the molecules in their local neighborhood.[2]

[2] detailed in Vytautas Gapsys' thesis.

Visual inspection of the superpositioned ensembles (Figure 3.20) reveals the highly disordered nature of the unphosphorylated RS peptide, whereas the RpS peptides adopt a specific arch-like structure.

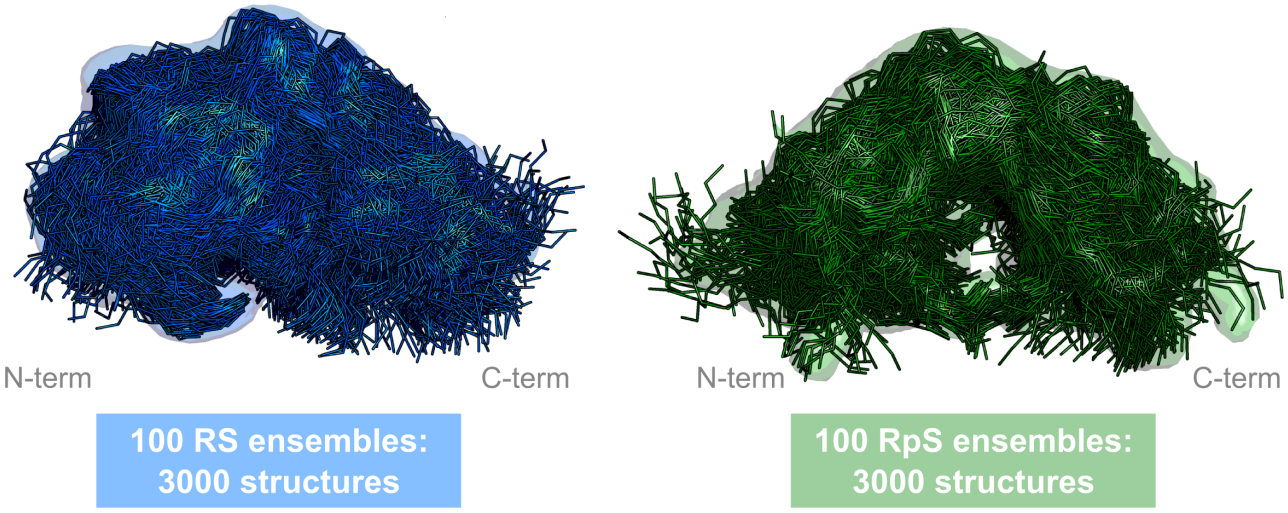
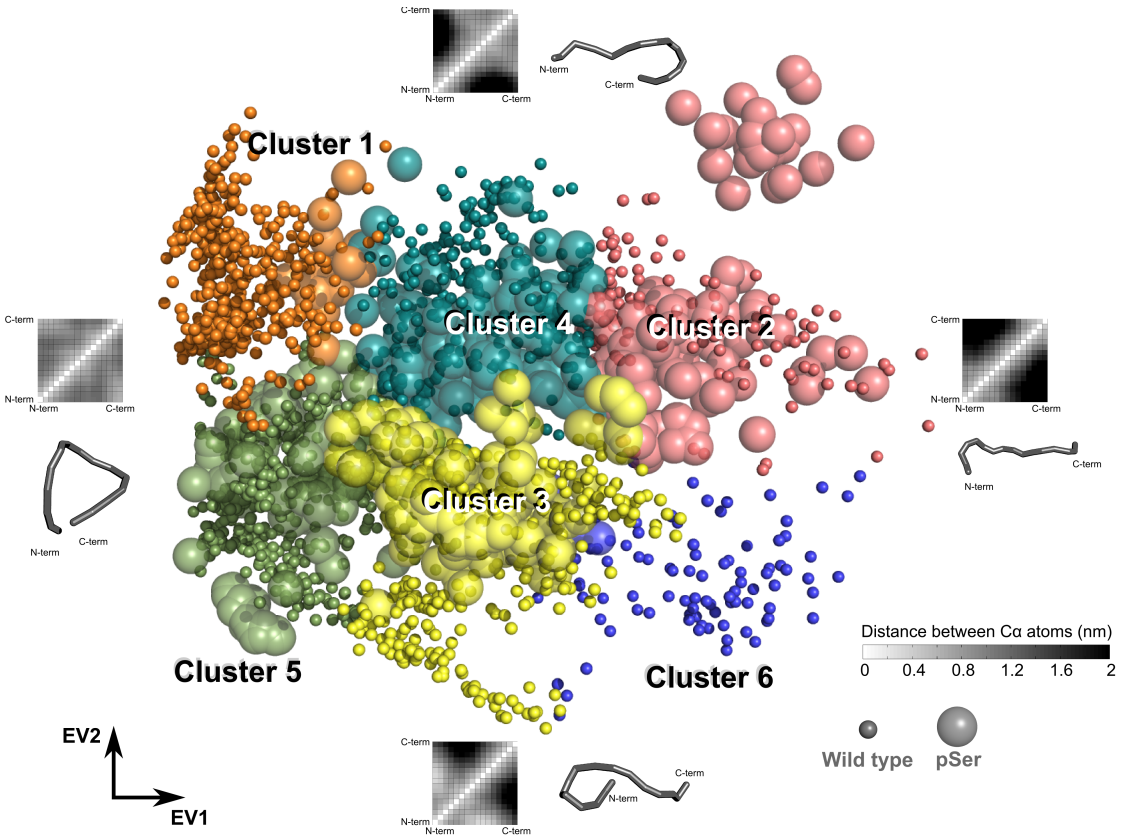


Figure 3.20 Superposition of native (RS) and phosphorylated ASF/SF2 (200-219) peptide (RpS) structures from 100 selected sub-ensembles

To further investigate the conformations adopted in the selected sub-ensembles, a clustering in internal distance space was performed (Figure 3.21) for the 6000 structures (3000 structures, 100 ensembles for each type) of native and phosphorylated peptide together. Six clusters were found to properly depict the populations contained within the ensembles. Phosphorylation appears to induce the depletion of clusters #1 and #6, which corresponds to the collapsed and stretched backbone conformations respectively. The population of the collapsed structures in cluster #5 is also decreased upon phosphorylation of the serine. Reduction of the collapsed and extended structures in the phosphorylated ensembles is compensated by enrichment of the arch-like bent structures corresponding to clusters #2, #3 and #4.


Cluster 1
Cluster 4
Cluster 2
Cluster 3
Cluster 5
Cluster 6
C-term
N-term
EV2
EV1
Distance between Cα atoms (nm)
0 0.4 0.8 1.2 1.6 2
Wild type
pSer

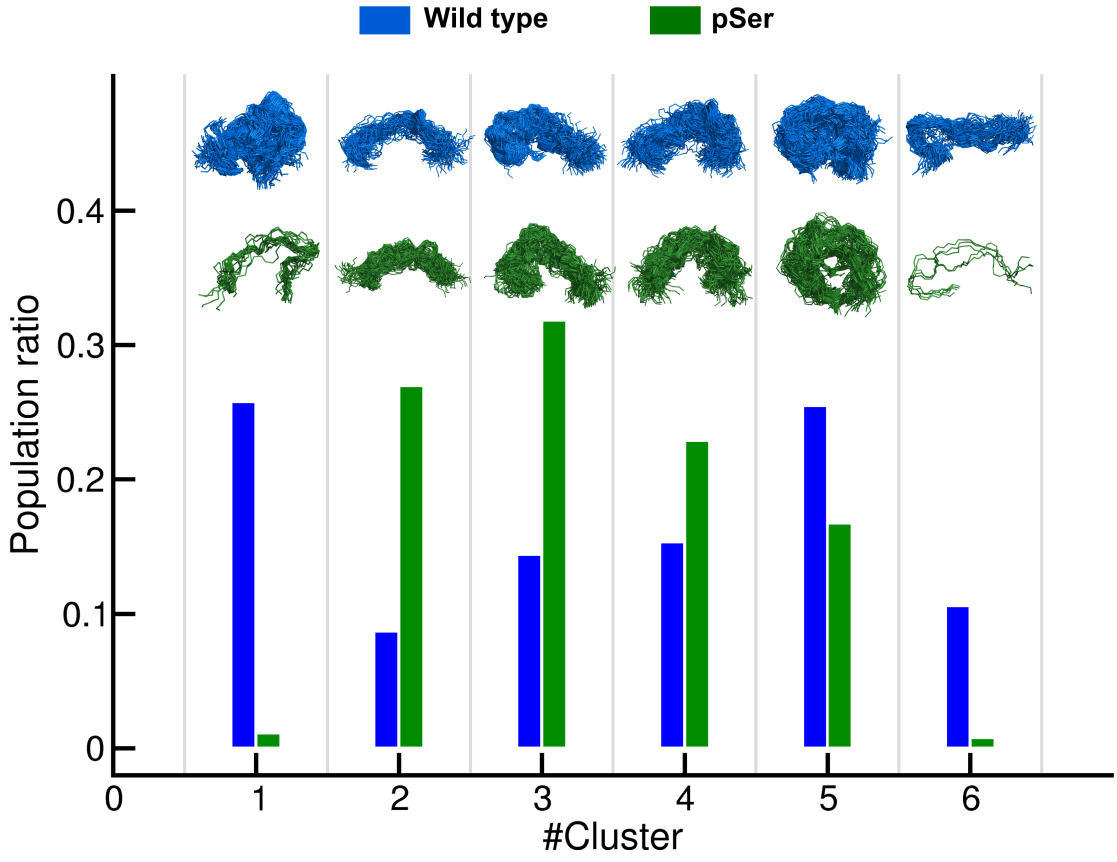

Wild type
pSer
Population ratio
0
0.1
0.2
0.3
0.4
0 1 2 3 4 5 6
#Cluster

Figure 3.21 Clustering analysis of native and phosphorylated ASF/SF2 (200-219) peptide structures. 3000 structures of each type were subjected to analysis. The 6 clusters were labeled on the plane of the first two eigenvectors. Every large sphere represents one phosphorylated ASF/SF2 (200-219) peptide structure and every small dot denotes one native peptide structure. The structures and Cα contact maps of extremes are shown.

A major change in the backbone conformation induced by the phosphorylation was elucidated by training a partial least squares (PLS) model (200) on the 70 RS and 70 RpS ensembles. A binary external parameter was constructed denoting whether a structure was coming from the phosphorylated or wild type molecule pool. Coordinates of the backbone atoms were used as an independent variable for the model. PLS regression using 25 latent vectors was able to discriminate between the wild type and pSer peptides (Figure 3.22). The main contribution to the difference between the peptides captured by the ensemble weighted maximally correlated motion(201) on a single peptide level indicates that the wild type backbone is more collapsed bringing the N- and C-termini closer to each other.

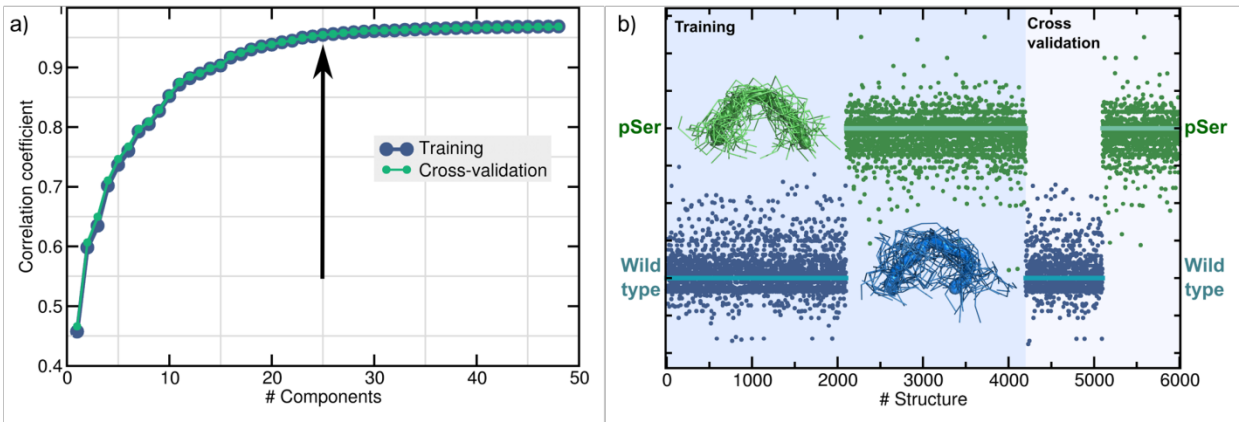


Figure 3.22 PLS analysis of native and phosphorylated ASF/SF2 (200-219) peptide backbone conformation. a) Correlation coefficient plotted against number of latent vectors. The arrow indicated the position of 25, which is the number of latent vectors used in PLS regression. b) The left part with light blue background showed training process with 4200 structures, which contained 2100 native ASF/SF2 (200-219) peptides and 2100 phosphorylated peptides. The right part with white background showed the

cross validation result with 1800 structures, which were half of native peptide and half of phosphorylated peptide. Every point was corresponding to one structure.

3.5.4 Orientations of the side-chains

Values of $^{3}J_{N\text{-}C\gamma}$ and $^{3}J_{CO\text{-}C\gamma}$ indicated changes in χ_1 angle distributions upon phosphorylation. The differences in the arginine χ_1 angle distributions of native and phosphorylated ASF/SF2 (200-219) are shown in Figure 3.23. Arginine and phosphoserine in the RpS ensemble show well defined preferential orientations, depicted by high density regions for the Nε (Arg) and Oγ (Ser, pSer) atoms in Figure 3.24. Side-chain positions of the wild type structures are smeared out; hence, high density areas are scarce, whereas side-chains of the phosphorylated peptides are restrained sampling highly restricted regions (Figure 3.24).

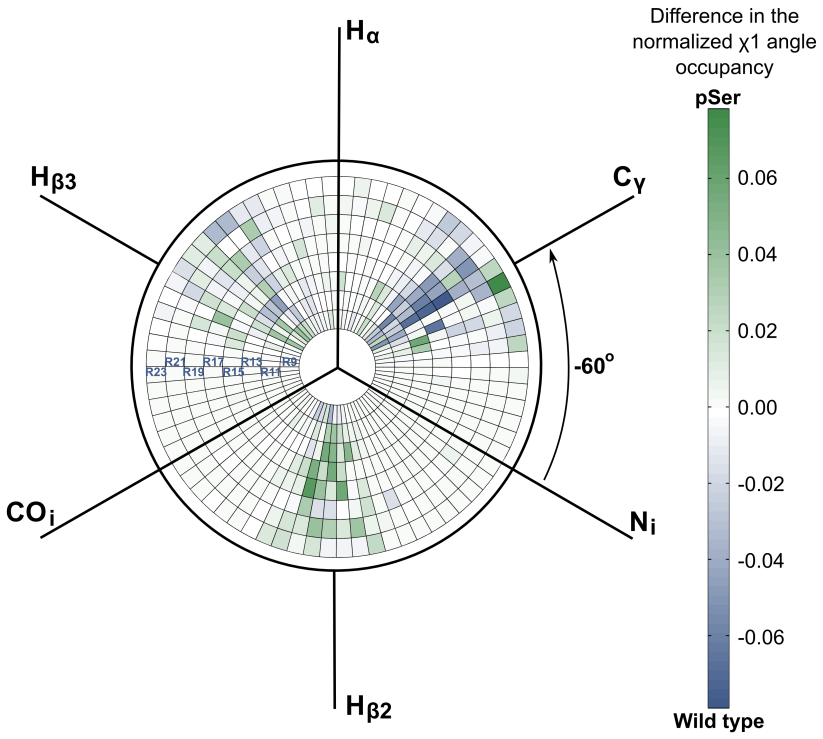


Figure 3.23 The difference of the occupancy of χ1 between native and phosphorylated ASF/SF2 (200-219) peptide. The occupancy of every arginine residue was normalized in such a way that integral over all the angles end up with 1.

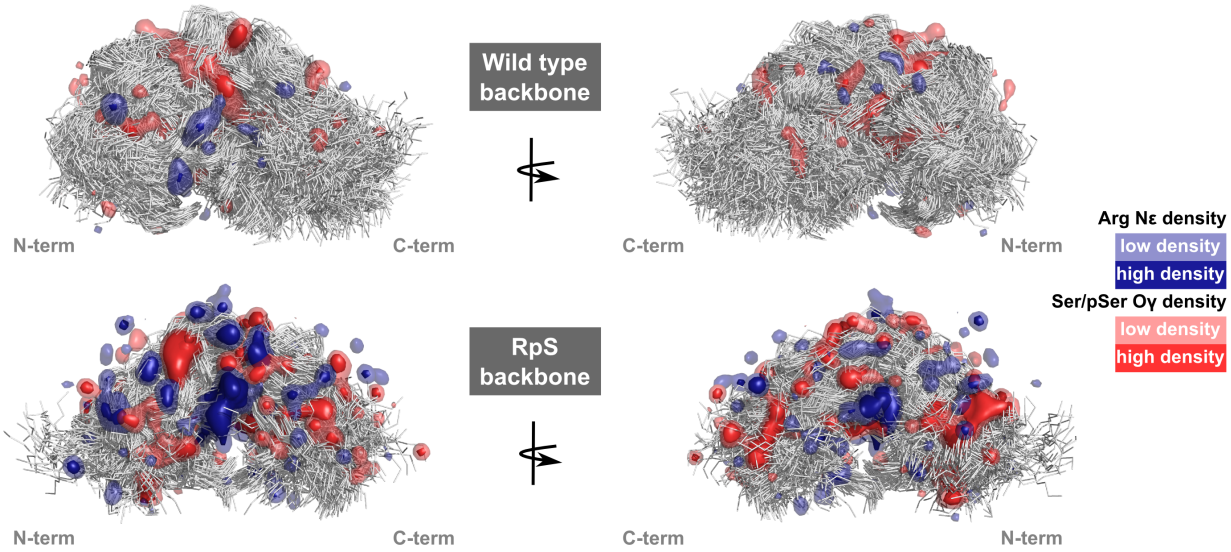


Figure 3.24 The orientation of side chains in native and phosphorylated ASF/SF2 (200-219) peptides. The 3000 structures in 100 sub-ensembles were included in analysis. Backbones are supperpositioned and shown in gray. The densities of Nε in arginine side chains and Oγ in serine side chain are sorted into three grades and plotted on the backbone.

It has been suggested in two previous MD studies (162, 163) that one compact conformation termed as “arginine claw” is formed by the phosphorylated RS repeats peptide. In the search for arginine claw, we counted the number of arginine residues with which phosphoserine side-chains formed hydrogen bonds, in all 3000 structures from the selected 100 NMR based ensembles. In most of the peptides no arginine residue H-bonded to phosphoserine was found. Less than one percent of the phosphoserine made hydrogen bonds with 3 or more arginine side chains, which hints that the arginine claw may be found in the phosphorylated ASF/SF2 (200-219) peptide, but is not a dominating conformation of the side-chains.

Number of arginine residues to which pSer has at least one hydrogen bond	Ratio with respect to all possible Arg-pSer hydrogen bonds (%)
0	60.88
1	28.83
2	9.88
3	0.53
4	0.01
5	0

Table 3.1 Number of arginine residues making hydrogen bond to phosphoserine side chains in the selected ensembles of phosphorylated ASF/SF2 (200-219).

3.6 SR-related proteins undergoing similar structural changes upon phosphorylation as SR proteins.

SR-related proteins are different from classical SR proteins in their domain structures. SR-related proteins normally do not have RRM domains, but do have some other RNA related domains. SR-related proteins also have RS domains. However, their RS domains contain many Arg-Asp and Arg-Glu dipeptide repeats and relatively less Arg-Ser repeats (68). The phosphorylation of the SR-related protein modulates its functions in cell, which is also phosphorylated by the same kinase as SR-proteins (104).

The native and phosphorylated forms of ^{15}N-labeled full length human Prp28 (1-820) and its isolated RS domain (1-257) were provided by Prof. Ficner's group, Georg-August-Universität Göttingen. The ^{15}N-labeled phosphorylated full length human Prp 28 (1-820) and its RS domain (1-257) were produced by SRPK1.

The ^{15}N-^{1}H HSQC spectra of 4 hPrp28 samples (native/phosphorylated, full length/isolated RS domain) were recorded and compared (Figure 3.25). In native and phosphorylated states, the spectra of the full length protein and the RS domain were very similar. Although full length hPrp28 is much longer, it did not show much more signal than the RS domain alone. Full length hPrp28 gave 202 peaks on ^{15}N-^{1}H HSQC while RS domain gave 188, which was picked by the peak picking algorithm in Sparky. The folded

parts of the proteins were mainly not visible on NMR due to slow tumbling. Only the flexible part, such as the disordered RS domain, shows up in spectrum. The full length protein spectrum matches very well with the isolated RS domain, which proves that the isolated RS domain conformations are identical to the domain in intact proteins in both native and phosphorylated states.

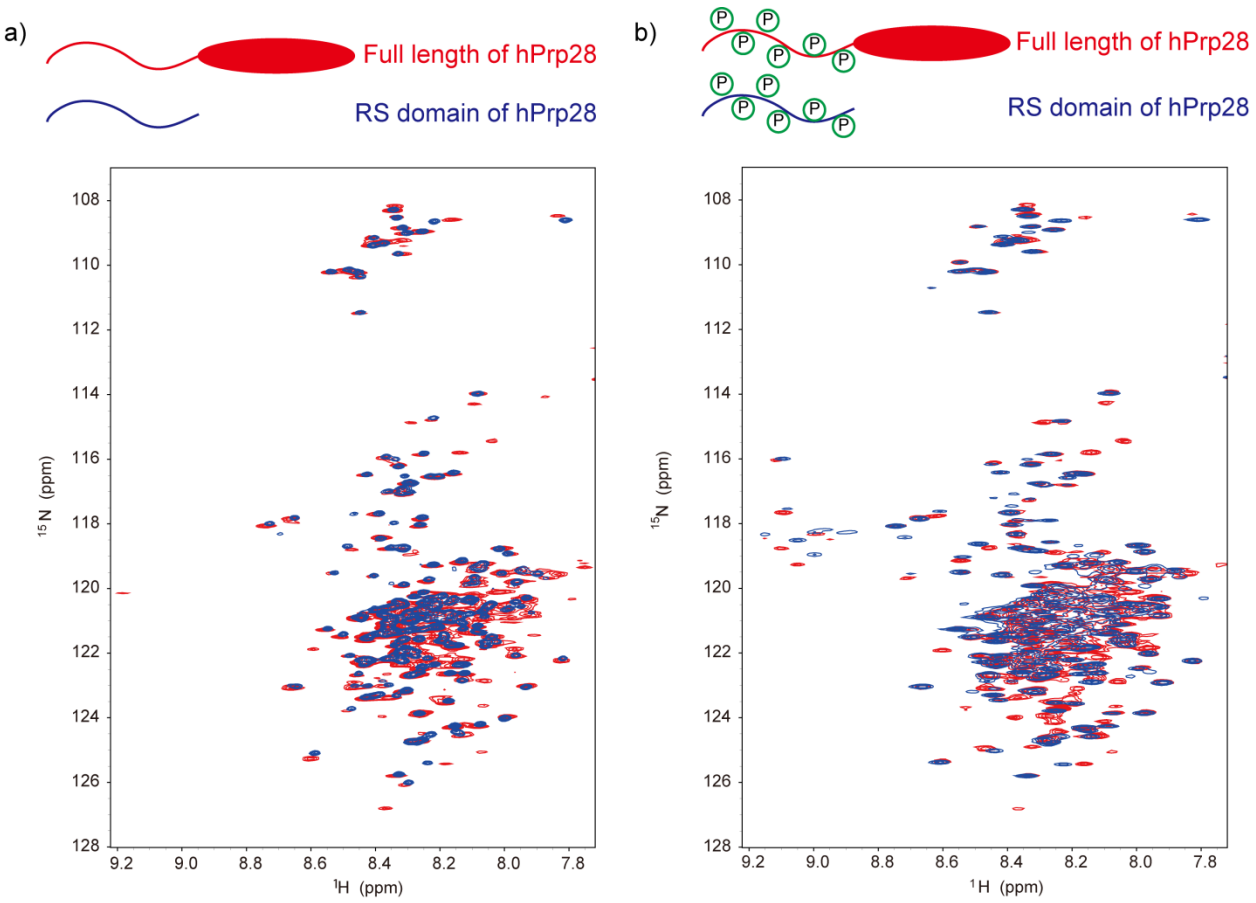


Figure 3.25 The isolated RS domain is similar to the domain in full length hPrp28. a) Superposition of ^{15}N-^{1}H HSQC spectra of RS domain alone (blue) and the full length hPrp28 (red) at pH6.5, 298K.b) Superposition of ^{15}N-^{1}H HSQC spectra of phosphorylated RS domain alone (blue) and phosphorylated full length hPrp28 (red). The spectra were recorded by Dr. Hai-Young Kim, MPI for biophysical chemistry.

The heteronuclear NOE spectra were measured for the isolated RS domain (1-257) in the native form, which is shown in Figure 3.26. Comparing the spectra with and without saturated spectra, the heteronuclear NOE values were nearly zero or negative. Especially

in the region of serine, all signals vanished except one negative peak. The nearly zero or negative heteronuclear NOE values indicated that the RS domain, especially the serine residues including those in RS repeats, is highly flexible.

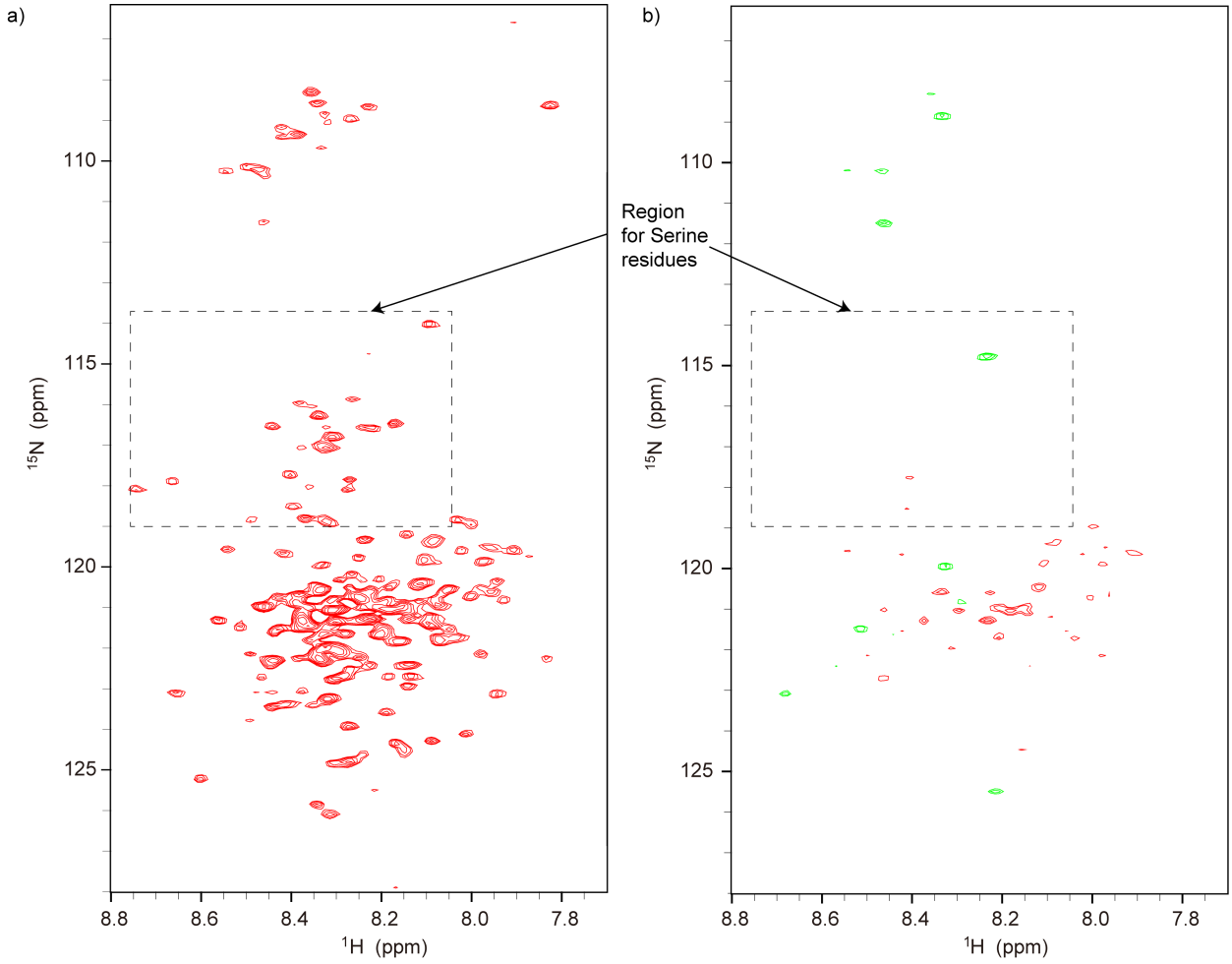


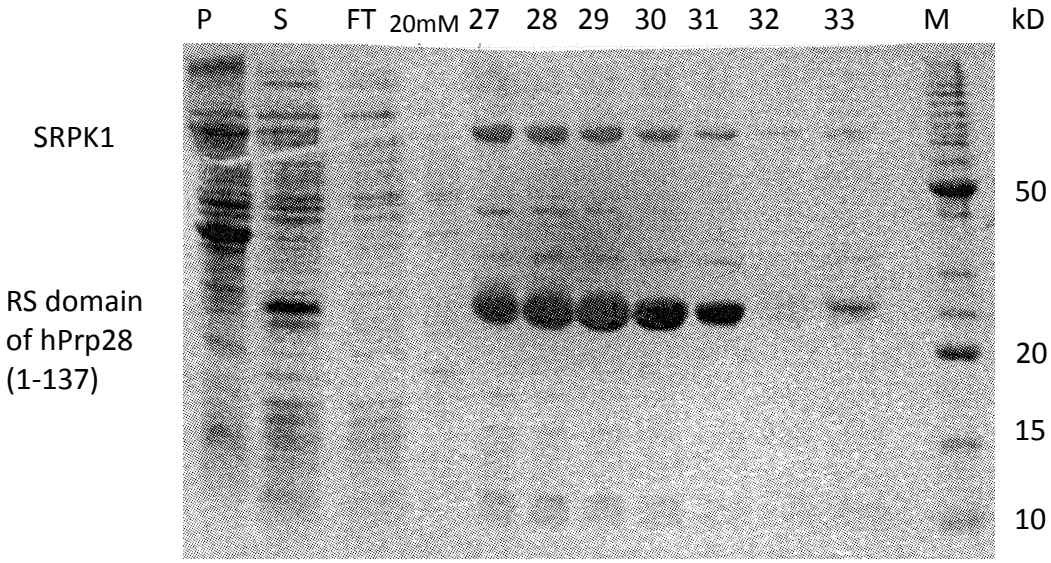
Figure 3.26 {^{1}H} ^{15}N Heteronuclear NOE spectra of hPrp28 RS domain at pH6.5, 298K. The spectra without (a) and with H saturation (b) are shown. The same contour levels were applied to both spectra. The region of serine residues are circled by dashed lines. The spectra were recorded by Dr. Hai-Young Kim, MPI for biophysical chemistry.

One shorter fragment of the RS domain of human Prp28 (1-137) was purified by NTA column, which contains all putative phosphorylation sites. The phosphorylated form of the RS domain was acquired by co-expressing it with SRPK1 kinase, which is described in the "Methods and Materials" section. The NTA chromatography of human Prp28 RS domain purification is shown in Figure 3.27a. The samples of each step in purification

and the fractions of target protein were analyzed by SDS-PAGE (Figure 3.27b). The SRPK1 was also enriched in the fractions of human PRP28. After the phosphorylation reaction, the RS domain was purified by reverse phase HPLC and the phosphorylation number was analyzed by mass spectroscopy (Figure 3.27c).

a)

b)



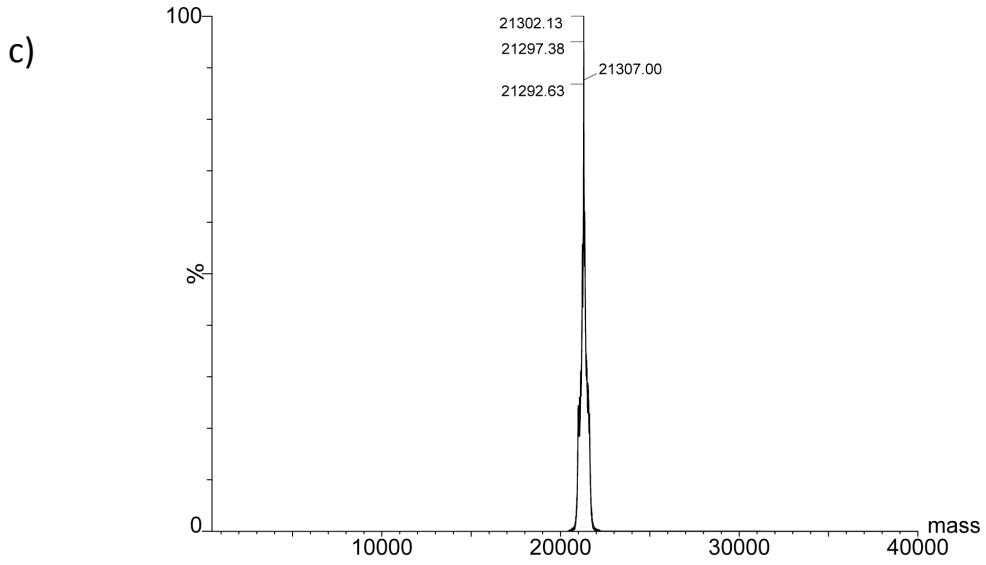


Figure 3.27 Purification of hPrp28 RS domain. a) Chromatography for NTA column purification of DXX23 RS domain. The blue line is absorbance at 280nm; the green line is concentration of imidazole; the red line is conductance; the azure line is temperature; the pink line is pH. The numbers of fractions are labeled at the bottom of picture. b) SDS-PAGE of each step and fraction in purification. Lane P, the precipitation after sonification and centrifuge; Lane S, the supernatant after sonification and centrifuge; Lane FT, flow through of NTA column; Lane 20mM, 20mM imidazole wash of NTA column; Lanes 27-33, the fractions of elusion; Lane M, marker. The molecular weight of each line is labeled on the right side. c) Mass spectroscopy of hPrp28 RS domain phosphorylated by SRPK1.The phosphorylation protocol is presented in "Methods and Materials".

There were 14 extra residues from vector in the N-terminal of protein. The length of whole domain with His tag was 160. Partial backbone assignment of the RS domain was achieved by 6D -seq-HNCOCANH and 5D HNCOCACB APSY experiments. All serine and threonine residues were assigned except Ser63.12 phosphorylation sites, include 11 serine and 1 threonine residues, were identified based on the distinct chemical shifts of phosphorylated amino acid, which is labeled on the ^{15}N-^{1}H HSQC spectrum (Figure 3.28). There is one continuing serine and arginine dipeptide repeat in sequence, which is constituted by 5 serine and 4 arginine residues from 71 to 79. All serines in this repeat are phosphorylated. Heteronuclear NOE shows the phosphorylated serine and arginine repeat

stretch is more rigid than the rest part of the protein (Figure 3.29a). The values were slightly lower than RS domain of ASF/SF2. This may be due to the fact that the hPrp28 sample was recorded at higher temperature and lower pH value. 3 bond J couplings of the HN-Hα of phosphorylated RS repeat stretch were smaller than the random coil values, which suggested the repeat stretch has the same backbone ϕ angle preference as the phosphorylated ASF/SF2 (200-219) peptide (Figure 3.29b). The two phenomena implied the SR-related protein adopted the same structure transition as ASF/SF2 (200-219) upon phosphorylation.

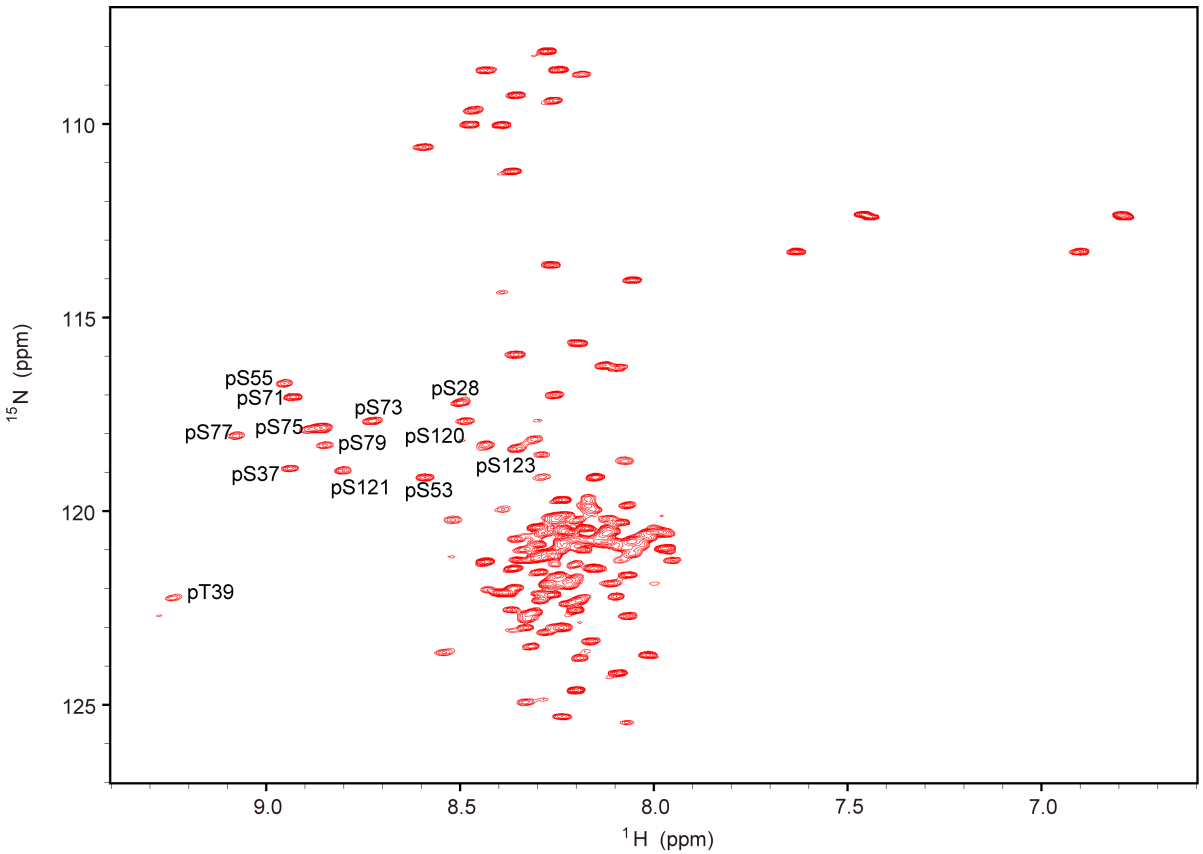


Figure 3.28 ^{15}N-^{1}H HSQC of hPrp28 RS domain at pH 6.0, 298K. The phosphorylated residues are labeled on the spectrum. The spectrum was recorded at 800MHz of H frequency.

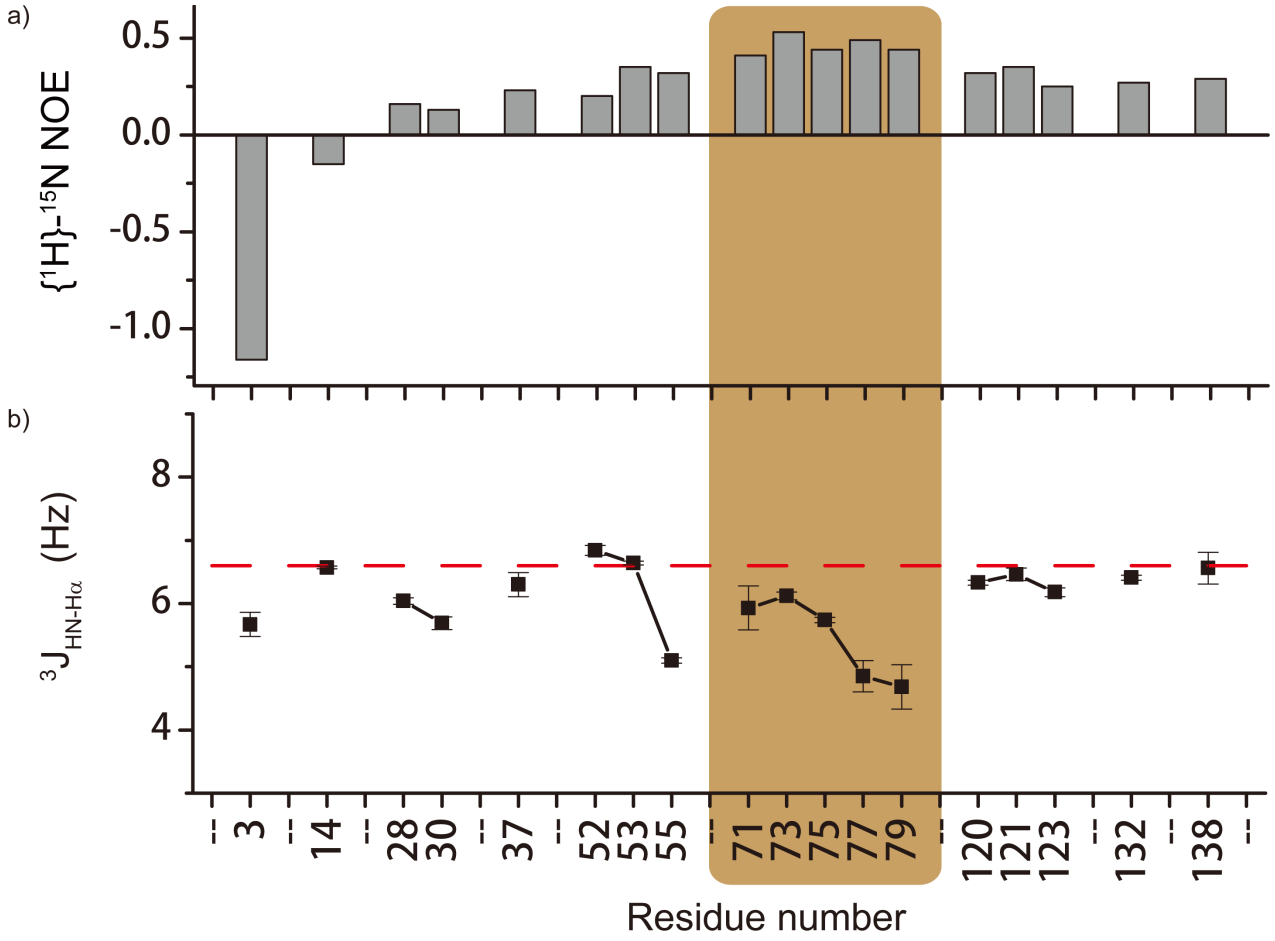


Figure 3.29 a) $\{^1H\}$-^{15}N heteronuclear NOE values of serine residues in hPrp28. b) $^3J_{HN\text{-}H\alpha}$ couplings of serine residues in hPrp28. The red dashed line indicates the random coil value of serine. The arginine-serine repeats stretch is highlighted by beige box.

4. Discussion

4.1 Structures transition of ASF/SF2 (200-219) peptide upon phosphorylation

In this study, we illustrated the atomic resolution structure properties of ASF/SF2 (200-219) peptide in its native and multiple phosphorylated statuses by the combined NMR and MD simulation approach. The result showed that in the native status, the RS repeats stretch was in a random coil state and upon phosphorylation became structured. Moreover, the phosphorylated RS domain of prototypical SR proteins and SR-related protein show similar structure properties as does a phosphorylated ASF/SF2 (200-219) peptide. These similarities suggested that the structure obtained from the RS peptide can be a model for all classical SR and SR-related proteins. It is the first high resolution structure of RS domains.

The random coil nature of unphosphorylated RS peptide was revealed unquestionably by the negative heteronuclear NOE values and its chemical shifts of Cα and Cβ.The severe chemical shifts degeneracy also pointed to a dynamic and fully disordered state, in which the differences of chemical environment were averaged out. In addition, high concentration of strong denaturant did not induce changes of Cα and Cβ chemical shifts, which means the peptide was already in the fully disordered status in its native form. This result matches with a previously reported circular dichroism study of RS domain of ASF/SF2 (161). The random coil status of RS repeats in native form is also in line with the bioinformatics result that phosphorylation predominantly occurs within disordered regions (51). The circular dichroism spectrum of the ASF/SF2 (200-219) peptide at a wide concentration range also support the disordered status of the ASF/SF2 (200-219) peptide.

Upon phosphorylation, the backbone became rigidified, which was proven by the negative to positive sign reversal of heteronuclear NOE values. Considering that the molecular weight of the ASF/SF2 (200-219) peptide is less than 3KD, the herteronuclear NOE around 0.4 is high. The reduced backbone mobility was also supported by the better separation of signals in the ^{1}H-^{15}N HSQC spectrum. The chemical environments of

amid protons could not be averaged. After adding 6M guanidine hydrochloride, the HSQC of RpS8 changed obviously. The spectrum of guanidine treated phosphorylated ASF/SF2 (200-219) peptide looks similar to native ASF/SF2 (200-219) peptide, in which all residues of the same residue type collapsed into one signal. This similarity again indicates that native ASF/SF2 (200-219) peptide is a fully disordered status. The changes of spectrum induced by denaturant treatment also supported that RpS8 peptide adopts a specific structure.

The structural changes induced by phosphorylation are restricted in the phosphorylated RS repeats region. The structural difference between native and phosphorylated ASF/SF2 (200-219) peptide were shown by the difference in $^{3}J_{HN\text{-}H\alpha}$ and different RDC profiles. The difference of $^{3}J_{HN\text{-}H\alpha}$ and RDC values were more significant in the middle of phosphorylated arginine and serine repeats compared to the regions preceding phosphorylated repeats. In $^{1}H\text{-}^{15}N$ HSQC, the peaks of residues before phosphorylated arginine and serine repeats are at the same position as the native form. The phosphorylation did not induce change in that region. The denaturant also did not change signals from the residues before phosphorylation RS repeats.

Although the phosphorylated peptide was more rigid than its native counterpart, it is also not fully rigid as a folded protein. In the sub-ensemble selection, there was no single structure fit to all NMR observables, i.e. there is no single structure, which can represent the conformations of phosphorylated ASF/SF2 (200-219) peptide. Under this situation, the ensemble approach should be used.

The sub-ensemble size of 30 structures was used to reach the reasonable reproduction of experimental data and to avoid the problem of overfitting. The back calculated RDC and $^{3}J_{HN\text{-}H\alpha}$ couplings fit very well with experimental data. However, it is known that one can often find many sets of conformations that will agree with a given dataset(41). Agreement with experimental data alone cannot guarantee the accuracy of ensemble. In prior work, this degeneration problem was circumvented in such a way that the multiple ensembles were generated and the common structure characteristics are extracted (47, 48). In this study, the selection of one ensemble with 30 structures was conducted 100 times for both native and phosphorylated ASF/SF2 (200-219) peptides. Superposition of the

resulting 3000 structures from RpS8 ensemble shows an arch-like structure. This arch-like structure is the common structure characteristic of ensembles. The restricted spread of backbone also fit with the increasing backbone rigid revealed by heteronuclear NOE measurements.

The phosphorylated RS peptide adopts an arch-like structure, which is not a regular secondary structure. The significance of structural differences between native and phosphorylated peptides was demonstrated by the Jesen-Shannon divergence within the same type and between two types.

The helix and sheets contents of native and phosphorylated ASF/SF2 (200-219) peptide did not change significantly. This can partially explain the small (<0.5 ppm) secondary chemical shits observed in the phosphorylated ASF/SF2 (200-219) peptide.

In this study, random coils of serine are used for phosphoserine in the chemical shifts prediction. In such system with a high density of phosphorylated serine residues as RpS8 peptide, the prediction of arginine residues will be also influenced by neighboring effects. It can explain the difference between back-calculated and experimental chemical shifts. The difference is bigger at the phosphoserine residues position.

The structures are validated by back calculated J couplings and chemical shifts. The lack of phosphorylated serine data also affects J coupling calculation. The formula of $^{1}J_{H\alpha-C\alpha}$ and $^{1}J_{C\alpha-C\beta}$ use the deviations from random coils (36, 195). The $^{1}J_{H\alpha-C\alpha}$ and $^{1}J_{C\alpha-C\beta}$ random coil values of phosphorylated serine are also absent. As in chemical shifts calculations, serine random coil values were used in calculation, which introduce systematic error at each serine residue positions.

There is second origin of the difference between back-calculated and experimental J couplings. The formulas for these J couplings predict $^{1}J_{H\alpha-C\alpha}$ with RMSD 1.59Hz and predict $^{1}J_{C\alpha-C\beta}$ with uncertainty of 0.5 Hz (195). The RMSD values between experimental and back-calculated data were less than 1.59Hz, which meant that the selected ensembles can reproduce experimental results. The RMSD was bigger for $^{1}J_{C\alpha-C\beta}$, which was mainly due the improper random coil values for serine residues. The random coil values used to

deduce Karplus equation were from free amino acid, which were not the same as residues in intrinsically disordered protein.

4.2 Discrepancy with previous works

Our results of native and phosphorylated RS repeat peptide were different from the results of two previous works (162, 163). In these two MD papers, (RS)8 peptide containing eight RS repeats, was suggested in, at least partially, α-helix conformation. Phosphorylation induced a melting of helix, and one compact conformation termed as "arginine claw" were found to be formed by the RpS8 peptide.

With the experimental data of ASF/SF2 (200-219), such as negative heternuclear NOE, secondary chemical shifts and signal distribution, there is no doubt of its disordered nature. The deviation of MD results may come from the bias of force field. In one recent paper, (RS)8 peptide conformations from MD with various force fields were compared. The conformations of RS repeats peptide varied with different force fields (163). It was noted that the force field AMBER99SB predicted much fewer helix conformations compared to the result by Amber03 used in a previous study (162). AMBER99SB had been shown to provide a better balance of secondary structure elements, while Amber03 force field preferred helix conformation (190). Therefore, the helix conformation previously observed in (RS)8 peptide is probably an artifact of force field.

Another potential vulnerability of these two studies was that the MD simulation was carried out in an implicit solvent. In one explicit MD simulation, which started with a helical structure, the helix conformation cannot persist within 50ns MD simulation (163).

One compact structure termed "Arg-claw" was supposed to be the conformation induced by phosphorylation. However, such a compact and intensive hydrogen bonded structure was not found in our structures' ensembles. In "Arg-claw", the side chain of one phosphoserine is in the center and is coordinated by side chains of up to 7 arginine residues (162, 163). The guanidinium groups are hydrogen bonded to the phosphate group of the phosphoserine center. In contrast to these results, few arginine residues

(<1%) in our structure ensembles formed more than 3 hydrogen bonds with phosphoserine side chain. The "Arg-claw" is not a dominating conformation. Some structures with high number of hydrogen bonds between arginine and serine side chains are shown in Figure 4.1, which did not resemble the "Arg-claw" structures previously reported (162).

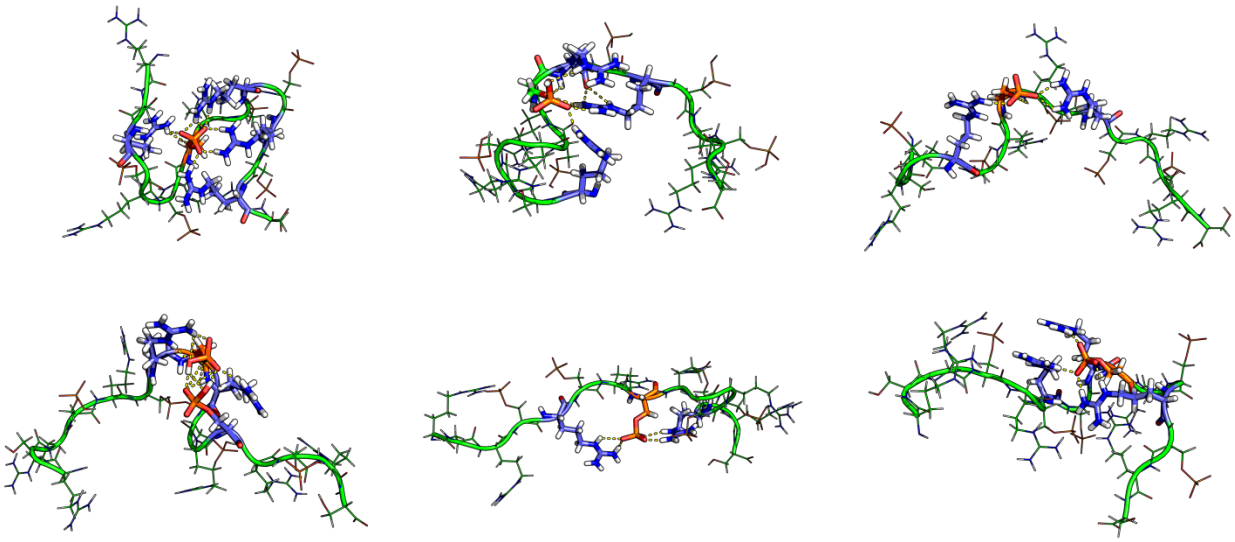


Figure 4.1 Some example structures of the RpS8 peptide with high number of hydrogen bonds between arginine and serine residues. The side chains of arginine residues are in blue and the side chains of serine residues are in orange. The hydrogen bonds are shown as dashed lines.

"Arg-claw" requires high curvature of the backbone, which is easier to achieve in the case of short peptides. For RS repeats stretches in the middle of an RS domain, it is hard to form such a ring-like conformation without collision of flanking sequences. The simulations that resulted "Arg-claw" used either a fully extended or helix structure of RpS8 as initial structure, which is not a good representation for a fully disordered peptide. In addition, the simulations were done in an implicit solvent. When an explicit solvent was used, 50ns MD simulation starting with "Arg-claw" showed a decrease of the hydrogen bond number. This may suggest a breaking down of the whole "Arg-claw" structure (163).

In the complex structure of SRPK1 and ASF/SF2, part of RS repeat stretch, ASF/SF2 (201-220), was in the bound groove of SRPK1 and exhibit an extended conformation (202). However, this was not structure of free form in solution but rather the bound form mainly induced and limited by the groove on the binding partner. The disordered nature of RS domain was supported by the fact that the rest of the RS domain was invisible on the X-ray structure. For the whole RS domain, only the stretch in the dock groove and 2 residues around the active center, all of which form contact and were restricted by SRPK1, showed up.

4.3 Orientations of side chains

The backbone conformation experienced significant changes upon phosphorylation as did the side chains, although very differently.

Unlike the backbone of ASF/SF2 (200-219) peptide, the torsion angle χ1 of arginine residues in the peptide were relatively restricted. The three rotamers were not equally populated. The *gauche*+ rotamer is dominated(Figure 4.2), which matches well with previously reported χ1 angles of arginine residues in denatured lysozyme(203).

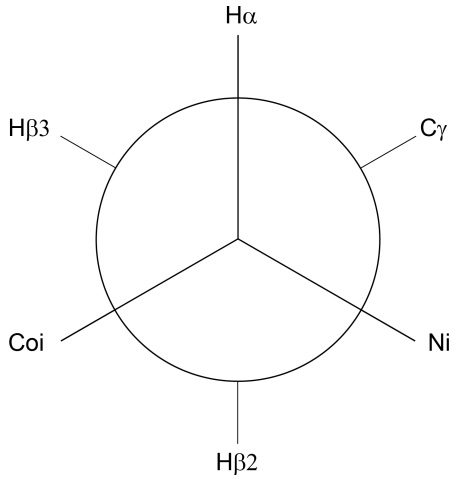


Figure 4.2 Dominant χ1 rotamer of arginine residues in native RS repeats.

Phosphorylation induced a population shift from *gauch*+ to *gauch*- conformation; however, *gauch*+ was still the dominated rotamer. From the point of view of χ1, the distribution of side chain orientations became dispersed upon phosphorylation. But when

the side chains were plotted on the superposing structures from ensembles, the distribution of side chains orientations in phosphorylated ASF/SF2 (200-219) peptide was more concentrated, while the directions of side chains in native peptide were smeared out. Since the χ1 angle distribution in the phosphorylated peptide was more scattered than for the native form, the preferential side chains orientations induced by phosphorylation was the result of disorder to order transition of backbone conformations.

Side chains rather than backbone of residues usually form the contact surface, which mediates the interactions between proteins. In the view of the bulk and charged side chains of phosphoserine and arginine residues, the changes of side chains orientations in the RpS8 peptide will change the shape and electrostatic potential of interaction surfaces significantly. Hence, phosphorylation modulates the interactions of the RS repeats stretch. Speculating from the surface of the RpS8 peptide, the interaction partner of phosphorylated RS repeats should have an interface with the alternating pattern of positive and negative charges.

4.4 pH dependency of RS repeats peptide conformation

The spectrum of phosphorylated ASF/SF2 (200-219) peptide is pH dependent. The ^{1}H-^{15}N HSQC spectrum changed dramatically from pH 4.0 to pH 7.4. The chemical shifts of serine and arginine Cα and Cβ also drifted during pH titration.

It is known that the phosphoserine chemical shift changes with pH because of the charge state of side chains (196). The pKa of protonation of dianionic phosphate group is 6. The pKa values of the other two protonation steps are out of the range of this pH titration. Thus, from pH 4 to pH 8, the charge of phosphate group changed from -1 to -2. The changes of phosphoserine chemical shifts were dominated by the electric charge on side chain. However, the charge of the arginine side chain does not change in this pH titration range. The changes of arginine chemical shifts may indicate the conformation changes of the peptide. At pH 4, when phosphoserine side chain was -1 charged, the arginine signals were overlaid as one big spot on ^{1}H-^{15}N HSQC, which is very similar to native ASF/SF2(200-219). At pH 7.4, the arginine peaks were well separated when the dominated charge state was -2. The different peak distribution implied a transition from

disordered to structured as pH increased. The Cα and Cβ chemical shifts of arginine residues also changed in the same manner. At pH 4, these values were very close to random coil values. As pH increased, the secondary chemical shifts also increased. These observations suggested that RpS repeats requires -2 charges on phosphoserine side chains to become structured. In addition, from pH 7.4 to pH 8, no change of Cα and Cβ were observed because the charge states did not change in this region.

pH 7.4 is usually taken as a physiological pH. *In vitro* splicing experiments are normally carried out in the pH range between 7 and 7.9 (204, 205). Within this pH range, the side chain of phosphoserine is -2 charged. In other words, the -2 charge may be the real biologically relevant state.

Glutamic acid and aspartic acid are usually introduced as a phosphoserine mimic. Both residues are -1 charged in the biologically relevant pH range. Considering different spectra of RpS peptide with -1 and -2 charge on phosphoserine side chains, this approach may be not appropriate when applied to RS repeats. At pH 7, the overall charge of RpS repeats is negative, while RE or RD repeats overall charges are close to neutral. In splicing assay, RD or RE repeats are much less effective than RS repeats in terms of complementing splicing reactions in S100 extract(186). It was also reported before that glutamic acid and aspartic acid cannot reproduce the same phenomena as phosphorylated serine 129 in α-synclein.(206).

In the RS domains of SR-related proteins, there are indeed many RE or RD repeats. The structures and functions of these repeats require further study.

4.5 Universality of RS model peptides

There is experimental evidence that the RS repeats alone can restore the function of whole domains in *in vitro* splicing assays (186). This supported the ideal that the RS repeat stretches can function as “functional elements” independent of flanking sequences. It is known that IDPs interact with their partners *via* short motifs (24). The RS repeat stretches may act in the same manner as performed structural elements (PSE) (207), prestructured motifs (PreSMo) (208) or linear motifs (209). The RS repeat stretches carry

out the crucial interaction, while the remaining residues in the sequence are mainly acting as spacers.

In this study, the phosphorylation induced structure changes in the whole RS domain of ASF/SF2 and SR domain of hPrp28 were similar to the RS peptide case at three points: first, the phosphorylated RS repeats stretches were more rigid than normal IDPs or other parts of the RS domain. Compared to other intrinsic disordered proteins such as Aβ and β-synclein, the values of RpS peptide are higher (183, 184). Since the two other IDPs were measured at lower temperatures (278K), the differences were even more pronounced. Second, the $^{3}J_{HN\text{-}H\alpha}$ values slightly decreased within the RS repeats. In addition, concerning the peaks position on ^{1}H-^{15}N HSQC spectra, the arginine residues in RS repeats drifted to the same region in the RpS8 peptide and two RS domains, which indicates the chemical environments of arginine residues were similar in all cases.

In RS peptide, the phosphorylation induced structure changes were limited within the RS stretch where the phosphates group were attached. There are extra residues in the N-terminal of RS repeats, which do not change upon phosphorylation. This suggested that within the RS domain, the phosphorylation induced structuring in the RS motif whereas the remainder of the domain stays the same, which fit well with the aforementioned "functional elements" concept.

Taken together, the structure transition between native and phosphorylated ASF/SF2 (200-219) peptides is the general model that can be applied to all SR-proteins and SR-related proteins.

4.6 Function redundancy of SR proteins

As mentioned in the introduction, SR proteins have redundant functions to a certain extent. Different individual SR proteins can complement splicing-deficient extracts, which are used as a benchmark to identify SR proteins. Moreover, the RS domain can be exchanged between SR proteins without a loss of biological function. For example, The RS domain can be exchanged between two SR proteins ASF/SF2 and SC35 without any effect on the ability of complement splicing-deficient S100 extracts(210). A further study involved six different human SR proteins showed that the activity of RS domains

differed quantitatively, which is directly related to the number of RS dipeptides in the domains (211).

This redundant function of RS domains can be explained by the universality of RS repeats structure discussed in the previous section. All RS domains contain RS repeat stretches. The same RS stretches in the different RS domains share the same structural feature. All RS stretches adopt the same structure changes upon phosphorylation. The main difference between the RS domains is the amount of RS dipeptides in the domain, which is also the source of "quantitative differences" between RS domains.

It was also shown that individual SR proteins were not functionally equivalent by depletion of single SR protein in cells and animals. Knocking out certain SR proteins results in deficient or lethal phenotypes, which mean the function of that SR protein cannot be accomplished by other SR proteins(113) (114) (115). However, an SR protein does not only contain an RS domain; the function redundancy of RS domain is not equal to functional overlapping of the whole SR protein. The functions carried out by other domains cannot be complemented by another SR protein without the same domain. For example, classical SR proteins have RRM domains, which recognize certain RNA sequences specifically. Different SR proteins recognize distinct RNA sequences, which cannot be simply substituted by another SR protein.

4.7 Binding regulated by phosphorylation

The RS domain is a module in charge of interaction, which mediated protein-protein and protein-RNA interactions. One unusual feature of RS domain is that its binding partners are mainly disordered proteins. Interestingly, the interaction partners often also include one RS domain. The direct interaction between RS domains were confirmed by GST-pull down experiments, which were modulated by phosphorylation in different ways where increased, decreased and unchanged binding affinity were observed(142).

The native RS domain of ASF/SF2 can self-associate, which is surprising considering its highly positive-charged nature (142). Nevertheless it was known that such a highly charged system, such as lysozyme, can form aggregation(212).

The alternatively charged phosphorylated RS repeats may self-associate in the way "polar zipper" does, first proposed by Peruz in a study of hemoglobin (213). Due to the absence of a complex structure, the factors contribute to interaction can only be speculated.

Phosphorylation of RS domain may affect these interactions in two ways: electrostatic potential reversion and entropy contribution.

First, the electrostatic potential was reversed upon phosphorylation. Because every phosphate group is -2 charged in physiological conditions and the number of arginine and serine residues is identical in RS repeats, the phosphorylated RS repeats are negatively charged. In contrast, the native RS repeats are highly positively charged, which make the PI of RS domains usually higher than 12. It was revealed that RS domains contact negatively charged mRNA. The positive charge of native RS domain facilitates this interaction. When the RS domain was phosphorylated, the same charge of RS domain and RNA cause repulsion between them, which may reduce the interaction between the RS domain and RNA.

Another binding event modulated by electrostatic interaction is between Nxf1 (TAP) and ASF/SF2, and 9G8 and SRp20. This interaction can happen without RS domains (214), but the phosphorylated RS domain inhibits the interaction (215, 216). Since the RS domain is not essential for interaction, the indirect effects of the RS domain may be due to long range electrostatic interaction. The arginine residing in adjacent RRM domains are directly involved in this interaction (217). The RS domains are also rich in arginine residues, which contribute to the attraction between molecules with the same charges as a direct interaction fragment. Upon phosphorylation, the reversed charge status of the RS domain causes repulsion in electrostatic interaction between ASF/SF2 and Nxf1, hence changes the affinity between ASF/SF2 and Nxf1.

Second, the conformational entropy contributes to the interaction strengths. It is known that change in conformational entropy can determine whether a proteins-ligand interaction will occur, even with the structures of interfaces unchanged (218). The higher level of disordered states of native ASF/SF2 (200-219) peptide will cause a higher

entropy penalty when folded into a complex. As seen in the c-Myb and KID domain of CREB binds to KIX domain of CBP, the higher disorder level of KID make its affinity to CREB less (219, 220).

On the other hand, the disordered native ASF/SF2 (200-219) peptide will be more favor for interaction under certain situations. The flexible backbone of the native ASF/SF2 (200-219) peptide is easier to be induced into certain bound-form when the interaction occurs in the "induced-fit" model(221). In the scenario of "conformation selection" (222), binding requires that the conformation of bound form already exists in a free status. The disordered native RS repeats samples more conformations than the phosphorylated form. If the bound form structure is not included in the ensemble of phosphorylated RS repeats but it is sampled by native RS repeats, than native form will be more favor for interaction. Hence in both two major models of interaction, the RS repeats may favor the interaction.

One of the changes of backbone conformations upon phosphorylation is that the backbone is more extended, especially at the N-terminal of RS repeats. This conformation may better fit to a binding partner with a long and relatively shadow groove rather than in a deep cavity. This preference of binding sites may also contribute to the different binding properties of RS repeats and its phosphorylation counterparts.

4.8 Different kinase specificity

In a previous study of the phosphorylation of ASF/SF2 by SRPK1, the number of phosphates group added was 15 (223), which is in the range of our results (from 12 to 17, Figure 2 in result part). There are 12 phosphorylation sites in the RS1part, which is known to the preferential substrate for SRPK1. These serine residues were phosphorylated by SRPK1 rapidly, which corresponds to the minus 12 phosphorylation number (223). The remaining 5 phosphorylate groups were added in the RS2 part (223). Since SRPK1 phosphorylates this region inefficiently, instead of a uniform phosphorylation number, species with all different phosphorylation numbers showed up.

The sequence of RS peptide in this study was taken from the N-terminal of the RS domain. It is found that SRPK1 first bound and started phosphorylation from an

"Initiation Box", which is seen in the downstream of the (RS)8 repeat sequence (223). Although this motif was missing, SRPK1 still can phosphorylate ASF/SF2 (200-219) peptides. In all cases, the phosphorylated serine residues were always a consecutive stretch in the middle of RS repeats and the minimum phosphorylation number was more than 5. This phosphorylation pattern matches well with the "sliding" model of SRPK1 phosphorylation mechanism which proposed that RS substrates slide through the active sites of SRPK1 to be phosphorylated without disassociation of the kinase-substrate complex (159). The sliding model of SRPK1 can also give an explanation for its preference of long and undisrupted RS repeats.

SRPK1 phosphorylated 13 serine residues in the RS domain of SR-related protein hPrp28 (see Figure 3.2). It is important to note that all five serine residues in the RS repeats are phosphorylated. An *in vivo* study of hPrp28 revealed that its phosphorylation mediated by SRPK2 is more biologically significant, but knocking out SRPK1 also changes the phosphorylation status of hPrp28 (104). Although SRPK2 and SRPK1 showed different substrate specificities *in vivo*, the specificity are not apparent in an *in vitro* assay (224). The different biological functions of SRPK1 and SRPK2 may not report any difference of their kinase activity but rather their different locations in cell: SRPK1 is predominantly associated with U1snRNA, whereas SRPK2 is associated with U4/U6-U5 tri-sRNP to which hPrp28 also binds. Hence, the kinase (SRPK1 or SRPK2) used for NMR samples did not make a difference. Interestingly, *in vivo* phosphorylation by either SRPK1 or SRPK2 resulted in several phosphorylation species that were revealed by 2D electrophoresis (104). In our study, the phosphorylation of hPrp28 was almost homogeneous (see Figure 3.27). This may be due to the large amounts of kinase and extended incubation time, which makes the phosphorylation reach the maximum.

It was previously reported that CLK1 can phosphorylate all putative sites in ASF/SF2 RS domain (156). However, the full phosphorylation was achieved by a high excess amount of kinase. In the same study, the phosphorylation of CLK1 showed two sequential phases: one fast phase with only around 5 phosphate groups added and a slow phase adding the rest phosphate groups. There is also evidence that CLK1 was less efficient than SRPK1 during the phosphorylation of ASF/SF2 in terms of the number of

incorporated phosphates groups, which was revealed by autoradiography (225). The structure of CLK1 also revealed that the binding groove is absent and the charge distribution is not in favor of a highly positive charged substrate. This suggested that CLK1 prefer a less positive charged substrate.

Taken together, the ASF/SF2 phosphorylation assay of these two kinases CLK1 and SRPK1 so far showed distinct region-specific phosphorylation ability, though they can phosphorylate the remaining part of the protein with low efficiency. SRPK1 prefers the RS1 part of ASF/SF2, and it can inefficiently phosphorylate several serine residues in the RS2 after RS1 phosphorylation. CLK1 on the other hand phosphorylates the RS2 part first, and then phosphorylates the whole domain in favorable conditions. However, the overlap of substrates may not happen in cell. CLK1 are mainly located inside nuclei, and SRPK1 exists in both cytoplasm and nucleus. Phosphorylation by SRPK1 in cytoplasm is required for SR proteins to be imported into the nucleus where SR protein can meet CLK1 (Figure 4.3). It is more likely that SRPK1 and CLK1 exert effects sequentially or synergistically on different regions of the same protein rather than competing against each other.

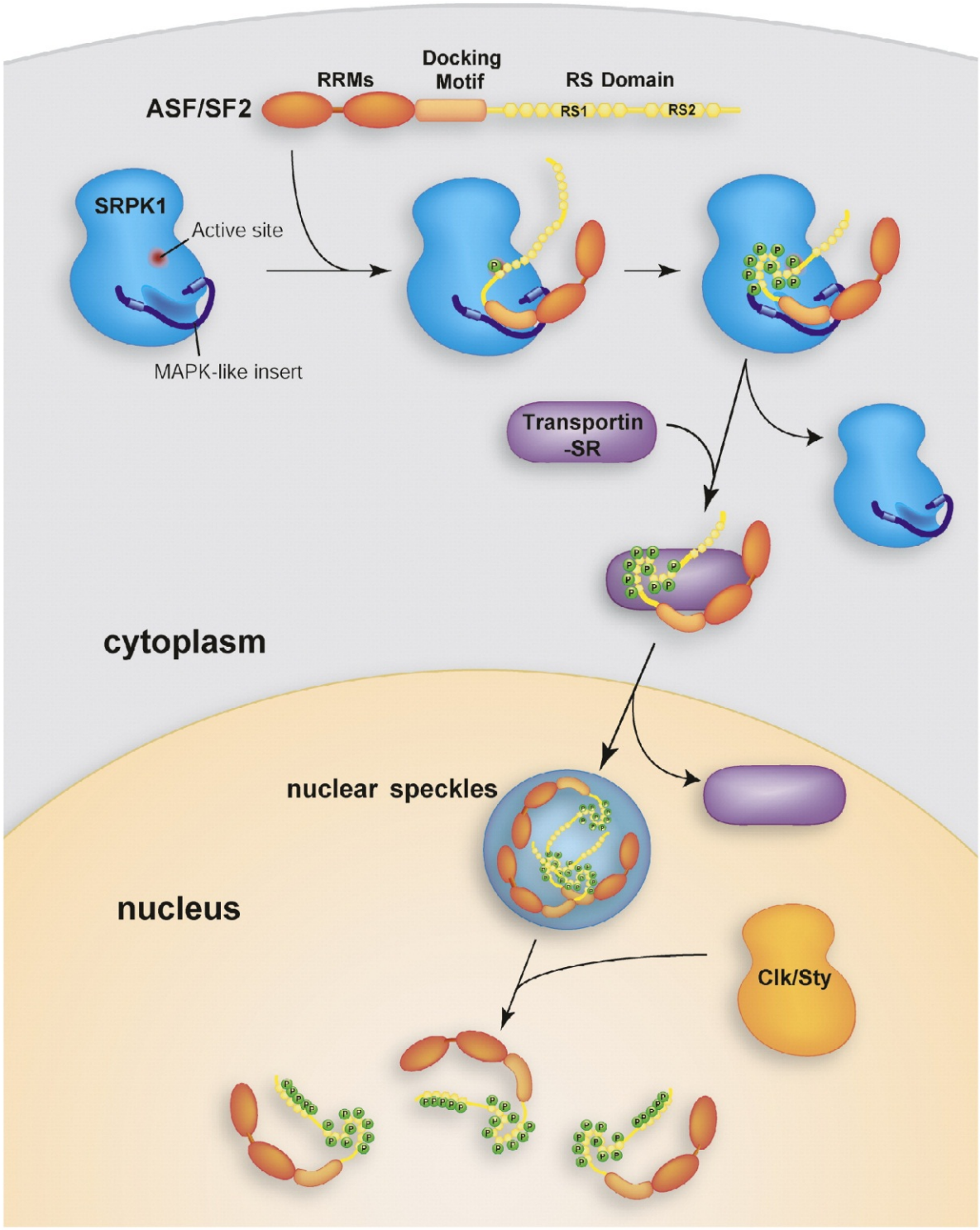


Figure 4.3 Proposed model for sequential phosphorylation of ASF/SF2 by SRPK1 and CLK kinases (reproduced from (152) with permission)

4.9 Outlook

In this study, using the combined NMR and MD approach, we are able to illustrate the structural consequences of phosphorylation on SR proteins. The whole RS domain and the continuous Arg-Ser dipeptide repeats region in it are intrinsically disordered. Upon phosphorylation, the Arg-Ser dipeptide repeats region is still dynamic but adopts an "arch-like" structure. This structural change was observed in the isolated ASF/SF2 (200-219) peptide, RS domain of ASF/SF2 and RS domain of hPrp28. This similarity suggested that the transition upon phosphorylation is universal to all SR and SR-related proteins. This structural change is pH-dependent, requiring -2 charges on the side chain of serine residues. Hence RE or RD repeats cannot reproduce the same phenomena as phosphorylated RS repeats.

The current study of structure is based on the isolated RS peptide and RS domain. It would be of great interest to study the structure of RS repeats in complex with other proteins or RNAs. There are two potential directions for future work.

First, studying the structural changes undergone by RS domain upon binding. There are a few RS domain mediating interactions in which the size of both sides is suitable for NMR study. Two potential target interactions are ASF/SF2 with H3 tail histone and ASF/SF2 with CTD tail of RNA polymerase II.

Second, identifying more binding partners of the RS domain, both in native form and phosphorylated form. Although there are many papers about the functions of SR proteins in cell, the exact binding partners of SR proteins are not clear, particularly in the process of splicesome assembly. Since SR and SR-related proteins contain domains other than the RS domain, it is important to identify the interactions directly mediated by the RS domain, and then further pin down the minimum sequences necessary for binding. The "minimum" complex can be used for later structural studies.

In addition, it is revealed by GST-pull down that the RS domain of ASF/SF2 forms homo-aggregation in its native form, and phosphorylation can inhibit this aggregation(142). It is worth studying this aggregation prosperity using other biochemical and biophysical approaches, such as gel filtration, dynamic light scattering,

NMR and SAXS, etc. Which part, RS1 or RS2, is necessary for aggregation? How does phosphorylation regulate the aggregation? Is this aggregation related to the SR protein location in speckle? These questions need to be answered through further study.

Part II Probing Hydrogen bonds networks with N-H couplings

5. Introduction

5.1 Indirect spin-spin coupling

Indirect spin-spin coupling, also called J coupling, is through-bond interaction, by which the spin of one nucleus perturbs the energy levels of neighboring magnetic nuclei. It is an indirect scalar interaction between two nuclear spins, which originates in the interaction between the nuclei and molecular electrons.

There are 3 main mechanisms for indirect spin-spin coupling, all mediated by electrons of bonds (226, 227). The most important is the Femi contact interaction between the spin of electron and the spin of nucleus, which is mediated by most inner *s* electrons. The electron is perturbed by the interaction between electron spin and nucleus A, which then interacts with nucleus B. The reversed process, the interaction of nucleus B with electron then to the nucleus A, also contributes to the interaction. In this way, the two nuclei A and B are coupled. Femi contact requires an electron with non-zero probability at the nucleus. It is only possible for the most inner *s* electrons.

The other two spin-spin coupling mechanisms are mediated by *p,d*… electrons .The second interaction is between the magnet dipole of the nucleus and the magnetic field caused by orbital motion of the electron. The magnetic field of electron is perturbed by spin A, and then interacts with the magnetic dipole of the other nucleus, spin B. As in the Femi contact, the reverse process also contributes to the mechanism. This interaction requires the non-zero angular momentum of the electron, which is not the case for *s* electrons. In contrast to the Femi contact, the presence of only *s* electrons results in zero orbital interaction. The third mechanism is the interaction between the magnetic dipole of nucleus spin and the electron spin. The magnetic dipole of the electron interacts with the magnetic dipole of spinning nucleus A .The electron perturbed by nucleus A interacts with the other nucleus B and *vice versa*. The term is also zero for *s* electrons.

The total J couplings are expressed as the sum of three terms:

$$J_{AB}^{Total} = J_{AB}^{FC} + J_{AB}^{Obital} + J_{AB}^{Spin-dipole}$$

Although the term "scalar coupling" and "J coupling" are used as synonyms, scalar coupling is the isotropic part of J coupling. In practice, the anisotropic part of J coupling is hard to observe, since it only shows up in aligned samples and is difficult to distinguish from much stronger dipolar couplings.

Since J coupling is mediated by chemical bonds, it provides the information about the connectivity of molecules. The observation of J coupling between two spins is positive proof of existence of a chemical bond. It is not necessary for the bond to be a covalence bond, as J coupling can also be mediated through a weak hydrogen bond (228). The J coupling aids the assignment of resonance enormously. In liquid state NMR, magnetization transfer is usually done by J coupling, such as COSY, TOCSY and numerous triple resonance experiments in bimolecular NMR (229). Moreover, its dependency on conformation makes the J coupling a precious resource for structure information.

J couplings give three parameters: multiplicity, sign and magnitude. Multiplicity is the multiple lines of interesting signals that arise from J coupling. From the number and relative intensities of these lines, the number of the nuclei coupled to the signal can be deduced. The sign of J coupling can be positive and negative, which is dependent both on the gyromagnetic ratios of involved spins and the number of bonds in between. However, the appearance of a high resolution spectrum does not change when the sign of coupling constants are reversed, if the relaxation effects are not taken into account (230).

The strength of J couplings depends on the bond geometry, such as bond length and bond angles. Normally, the more bonds that are involved, the smaller J couplings are. Particularly, 3 bond J couplings are correlated with dihedral angles ϕ by various Karplus equations (231, 232):

$$J(\phi) = A\cos\phi^2 + B\cos\phi + C$$

Some one-bond J couplings, such as $^{1}J_{C\alpha\text{-}H\alpha}$ and $^{1}J_{C\alpha\text{-}C\beta}$ are also related to backbone ϕ and ψ angles by their own equations (36, 195). The torsion angles derived from J couplings provide valuable information for structure determination.

5.2 Hydrogen bond

The hydrogen bond (H-bond) is the attractive force between a hydrogen atom and an electronegative atom such as nitrogen, oxygen or phosphorus, which comes from another molecule or chemical group. The hydrogen is also covalently attached to another electronegative atom. Two electronegative atoms interact with the same hydrogen atom. The electronegative atom covalently attached by hydrogen is termed as the donor. Because of this atom, the hydrogen atom is partially negatively charged, which makes it interact electrostatically with the other electronegative atom, termed the acceptor.

The length of the H-bond is longer than that of a normal covalent bond. The H-bond is generally weaker than a normal covalent bond. The exact energy of the H-bond is dependent on its geometric features.

The H-bonds are essential for the structure of many biochemical compounds. Hydrogen bonds between the complement base pairings are the main force that holds DNA double helixes and secondary structures in RNA molecules. In proteins, the backbone hydrogen bonds maintain the secondary structure elements. Protein folding, the formation of amyloid aggregates, enzymatic catalysis, drug-receptor interactions, and many other phenomena are intrinsically connected to hydrogen bonding.

Hydrogen bonding in proteins involve, in most cases, arrangements of the type C=O•••H–N. The backbone amide proton is usually the donor of the hydrogen bond. The acceptors of the H-bond can be the oxygen atoms of the backbone carbonyl group. Oxygen atoms in hydroxyl groups, which reside in the side chains of some residues, such as threonine, tyrosine and serine, can also be acceptors.

A number of NMR observables are used as indirect evidence as existence of H-bonds, such as exchange rates with solvents (233), temperature coefficients (234), chemical shifts, anisotropies of the chemical shifts (235), and nuclear quadrupole coupling

constants (236), which make NMR spectroscopy a key tool for the detection and characterization of H-bonds.

The amide protons in the H-bonds exchange several magnitudes more slowly than those exposed to solvents (237). The amide protons involved in the H-bonds also shift less than the latter during temperature changes. Temperature coefficients of the amide protons which are more positive than -4.6ppm/K indicate the existence of the H-bonds (234). The H-bond changes the shielding tensors of amide protons in two ways: a decrease of isotropic values, which increases the isotropic chemical shifts, and an increase of the chemical shift anisotropy (CSA). The correlation between the amide proton CSA and the length of the H-bond, which was found to be

$$CSA = B + C/(r_{H\cdots O} - D)^2$$

B= 4.9 ppm, C=1.96 ppm▪ $Å^2$ and D= 1.3Å(235).

The experimental observation of NMR spin-spin coupling constants across the H-bond allowed, for the first time, a direct detection of H-bonding in biomolecules (228, 238). Subsequently, theoretical calculations of trans-hydrogen bond spin-spin couplings provided detailed insight into the nature of H-bonds (239). There are 3 major structural parameters of H-bonds: the distance between H and O r_{HO}, the H··O=C' angle θ_2 and H·· O=C'-N dihedral angle ρ. The experimental J coupling cross H-bond, $^3J_{NC'}$, can be fit to the following formula with these 3 parameters:

$$^{3}J_{NC'}(\theta_2, \rho, r_{HO}) = \{-0.70\cos^2\theta_2 + 0.74\cos\rho\sin^2\theta_2\} * \exp[-3.2(r_{HO} - r_{HO}{}^0)] - 0.01\,Hz$$

Although the formula is based on 34 experimental cross H-bonds couplings in Gb3, it also applies to other proteins. This formula enables the use of cross H-bond couplings as restraints in molecular dynamic simulations, which offer the description of the geometry and the energetics of hydrogen bonds accuracy close to those of high-resolution X-ray structures (240).

5.3 Field dependence of scalar coupling values

In theory, the scalar coupling values should be irrelevant to the external field strength. However, there are two effects, which make the apparent J coupling values vary depending upon external magnetic fields: dynamic frequency shifts (DFS) and dipolar contributions to J coupling. The following chapter is limited to the $^{1}J_{NH}$.

It is well known that two doublet components of ^{15}N relax at different speeds due to cross correlation ^{15}N chemical shift anisotropy and ^{1}H-^{15}N dipolar couplings (241, 242). This interference also causes small resonance frequency shifts of two components. This phenomenon, termed as dynamic frequency shifts (DFS), imposes a change in observed $^{1}J_{HN}$ splitting. The DFS is theoretically described and experimentally measured. The size of DFS is dependent on the static magnetic field and the dynamic properties of amide groups. In a model-free approach, the DFS is the function of order parameter, rotation and internal motion correlation time, and is deduced as follows (243):

$$\delta_{DFS}(B_0) = S^2(20\pi^3)^{-1}h(\sigma_{\parallel} - \sigma_{\perp})(3cos^2\eta - 1)\gamma_N\gamma_H(r_{NH})^{-3} \times [1 + (\gamma_N B_0\tau_c)^{-2}]^{-1}$$

$$= \delta_{DFS}(\infty)[1 + (\gamma_N B_0\tau_c)^{-2}]^{-1}$$

S^2 is the generalized order parameter, which gives the information about the globule and internal motions. h is Planck's constant. $(\sigma_{\parallel} - \sigma_{\perp})$ is the chemical shielding anisotropy of amide N atom. η is the angle between the unique axes of dipolar and CSA tensors. γ_N and γ_H are gyromagnetic ratios of nitrogen and proton. Hence $\gamma_N B_0$ gives the angular Larmor frequency of nitrogen. r_{NH} is the distance between proton and N spin. $\delta_{DFS}(\infty)$ denotes the DFS at slow motion limit where $\gamma_N B_0\tau_c \gg 1$. Plugging in all parameters, $\delta_{DFS}(\infty)$ is -0.54Hz assuming S^2 is 1.

A protein can be weakly aligned by the magnetic field, which is induced by anisotropy of its magnetic susceptibility. This alignment results in a dipolar contribution to the apparent $^{1}J_{NH}$ splitting. The residue dipolar coupling, δ_{dip}, is given as (243)

$$\delta_{dip}(\theta,\phi) = -\left(\frac{{B_0}^2}{15kT}\right)\left[\frac{\gamma_N\gamma_H h}{4\pi^2{r_{NH}}^3}\right] \times [\Delta\chi_a(3cos^2\theta - 1) + \frac{3}{2}\Delta\chi_\gamma(sin^2\theta cos2\phi)]$$

B_0 is the static magnetic field, $\Delta\chi_a$ and $\Delta\chi_\gamma$ are the axial and rhombic components of the magnetic susceptibility tensor. θ and ϕ descript the orientation of the N-H bond vector in the principal axis system of the magnetic susceptibility.

For a diamagnetic protein, the overall magnetic susceptibility anisotropy mainly comes from two sources: backbone peptide bonds and aromatic side-chains. The alignment of protein G is indeed stronger than ubiquitin, because protein G contains a higher proportion of aromatic residues in the primary sequence (243, 244).

These two effects can be extracted by comparing $^1J_{NH}$ obtained at different field strengths. For the folded globule proteins, the DFS of all residues are identical except those in the extremely flexible terminals. So the difference between two magnetic fields can be treated as normal RDC data plus a constant offset. The magnetic susceptibility tensor can be fitted by this process. The studies on ubiquitin and GB3 demonstrate that the actual DFS values fit very well with theoretical values (243, 244). The magnetic susceptibility of ubiquitin obtained by fitting $^1J_{NH}$ also shows good agreement with the one calculated from X-ray structure (243, 244).

5.4 The methods to measure scalar couplings

Since J couplings can provide structural and dynamic information, and the residual dipolar couplings can be measured in the same way as J couplings, various methods were developed to measure J couplings.

Generally there are two ways to measure couplings: the splitting method and the intensity modulation method. These two methods are based on different principles and have their own advantages and shortcomings.

5.4.1 Splitting method

In the splitting method, the coupling values are read out by the displacements between peaks. The principle of such experiments is to allow the couplings to be effective during the chemical shift evolution time, which results in a splitting in that dimension. The displacement between peaks directly gives the coupling values. The design of these experiments is straightforward, usually conducted just by removing decoupling from the

evolution of interesting dimensions. The results are easy to analyze, because they only require basic arithmetic after peak picking. In protein NMR, this method is a common way to measure scalar couplings and residual dipolar couplings, especially for relatively large one-bond couplings, such as $^{1}J_{NH}$ and $^{1}J_{C\alpha\text{-}H\alpha}$ (168, 245).

However, the splitting of signals inevitably increases the signals overlapping, which is more severe in a big system or in an intrinsically disordered protein. Moreover, to get the peaks position accurate enough, high resolution has to be achieved on the dimension, which results in a long measurement time. If the interesting J coupling is very small, much less than the line width of signals, this method becomes problematic.

The problem of signal overlapping can be circumvented by so-called in-phase /anti-phase experiments (245). In this kind of experiment, instead of only recording in-phase terms, the terms of anti-phase are also recorded separately. Adding up or subtracting these two terms will get one of the two split doublets (Figure 5.1). To minimize experiment uncertainty, the two spectra are recorded in an interleaved way.

The IPAP method is widely used in NMR measurements for scalar couplings. IPAP-HSQC is a common pulse program to measure $^{1}J_{HN}$(245). HACANH experiments use the IPAP scheme on the Cα dimension, which can give sequential connection and Hα-Cα couplings at the same time (168). Another backbone coupling Cα-CO can be measured by one HNCO-based pulse program, which makes use of IPAP in its CO dimension (181).

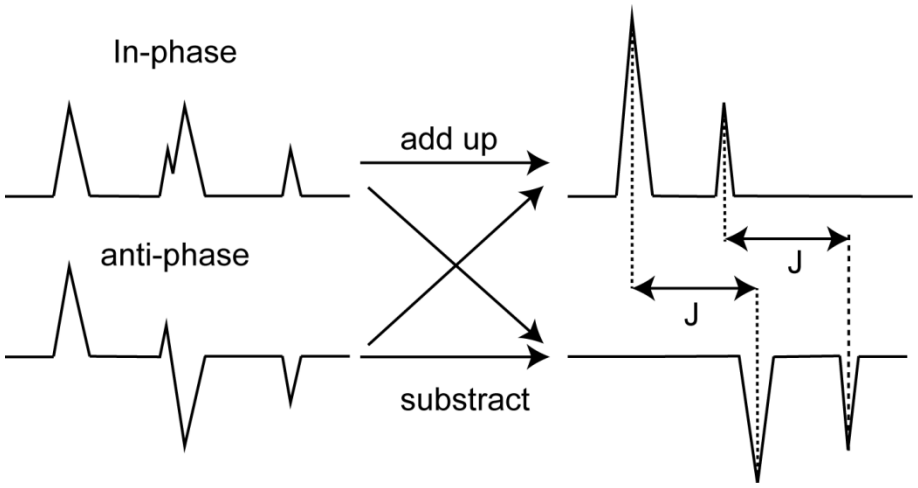


Figure 5.1 IPAP method to solve the overlapping problem

For the couplings which are too small to make peaks separate enough, the E.COSY (exclusive correlation spectroscopy) type experiment can help to solve the problem (246). In heteronuclear experiments, the E.COSY type scheme uses a large coupling to separate two peaks, and lets the small couplings show up in other dimensions. A small unresolved J coupling between I and X, J_{IX}, can be measured with assistance from a large, resolved J coupling between X and a third spin S. Assuming a 3 spin system: I and S both heteronuclear coupled to X. Jsx is large while J_{IX} is too small to direct measure. In t1 evolution, S spin evolves under chemical shift ωs and J couplings Jsx.

$$Sy \rightarrow Sy\cos(\omega_S t_1)\cos(\pi Jsx t_1)-Sx\sin(\omega_S t_1)\cos(\pi Jsx t_1)-$$

$$2SxXz\cos(\omega_S t_1)\sin(\pi Jsx t_1)-2SyXz\sin(\omega_S t_1)\sin(\pi Jsx t_1) \quad (1)$$

Then the magnetizations are transferred from S to I without perturbing the X spin, which results in:

$$Iy\cos(\omega_S t_1)\cos(\pi Jsx t_1)-2SyXz\sin(\omega_S t_1)\sin(\pi Jsx t_1) \quad (2)$$

I spin evolves during t2, in which chemical shifts of I, ω_I, and the scalar couplings J_{IX}:

$$[Iy\cos(\omega_I t_2)-Ix\sin(\omega_I t_2)]\cos(\pi J_{IX} t_2)\cos(\omega_S t_1)\cos(\pi Jsx t_1)$$

$$-2Xz\,[Ix\cos(\omega_I t_2)+Iy\sin(\omega_I t_2)]\sin(\pi J_{IX} t_2)\cos(\omega_S t_1)\cos(\pi Jsx t_1)$$

$$-2Xz\,[Iy\cos(\omega_I t_2)-Ix\sin(\omega_I t_2)]\cos(\pi J_{IX} t_2)\sin(\omega_S t_1)\sin(\pi Jsx t_1)$$

$$+[Ix\cos(\omega_I t_2)+Iy\sin(\omega_I t_2)]\sin(\pi J_{IX} t_2)\sin(\omega_S t_1)\sin(\pi Jsx t_1) \quad (3)$$

Only the first and the last terms are the observable signals. The first term leads to anti-phase signals in both dimensions (Figure 5.2a). The last term generates anti-phase signals split in t1 but in-phase split in t2 (Figure 5.2b). Adding up these two terms will cancel half of the 4 multiples. The diagonally arranged two peaks are separated by J_{SX} in t1

dimension and J_{IX} in t2 dimension (Figure 5.2c). The large displacement of Jsx makes the measurements of J_{IX} possible.

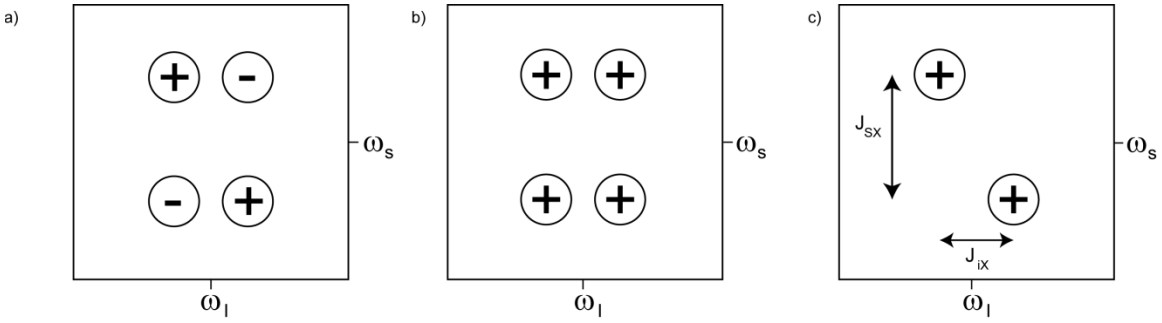


Figure 5.2 Multiple peak pattern of E.COSY.

$^3J_{HNH\alpha}$, $^3J_{HNC\beta}$ and $^3J_{HNCO}$ provide information about the backbone torsion angle ϕ. However, the sizes of these couplings are too small to be measured directly. The E.COSY method is used to take measurements of these smaller couplings. It employs large, one-bond couplings, for instance, $^1J_{C\alpha-H\alpha}$, $^1J_{C\beta-H\beta}$ and $^1J_{C\alpha-Co}$ as Jsx to separate two peaks (247, 248).

5.4.2 Intensity modulation method

In the intensity modulation or so-called "quantitative J correlation" method, the intensities of signals are dephased by J couplings with certain delays. The intensities of resulting signals are compared in order to extract J couplings.

One of the most frequently used intensity modulation methods is a spin-echo difference-based scheme, which is shown in Figure 3a. This method can be used when the S spin is heteronuclear to I spin with scalar couplings J_{IS}. Transverse magnetization of I spin is generated as input for this module. The 180 degree pulse of S can be placed at position a or b. When the pulse is at position b, the J_{IS} coupling is effective for the all 2T period. The resulting signal is proportional to $\cos(2T\pi J_{IS})$. When the 180 degree S pulse is at position a, the J coupling between I and S is refocused. The resulting signal can be used as a reference. The couplings can be extracted by the ratio of these signal intensities.

The spin-echo method is used to measure $^3J_{HN-H\alpha}$, $^3J_{CO-C\gamma}$, $^3J_{N-C\gamma}$ (176-178). One advantage of this method is that the dimensionalities of the experiment can be reduced, because the chemical shift of spin S is not evolved. The ^3J couplings mentioned above can be measured by 2-dimensional N-HSQC based spin-echo experiments. 3D experiments will be needed to measure these couplings if the other methods are used.

One assumption of the spin-echo method is that the relaxation rates of in-phase and anti-phase terms are identical. But this assumption is not valid and systematic errors are introduced. The detailed analysis of this error has been undertaken by Vuister and Bax (35). Generally speaking, the relatively fast relaxations of anti-phase terms result in the underestimating of J couplings.

One variation of this method can be implemented even without a reference spectrum. Instead of using a single T delay, a series of experiments with different T length are carried out. The J couplings can be extracted by fitting the resulting signal intensity and T delays with a certain formula. Since there are always the same in-phase terms involved, the problem of the different relaxation rates can be eliminated. The accuracy of this experiment is high when enough T time points are included in fitting, as illustrated in J-modulated HSQC for measuring $^1J_{HN}$ (243, 249).

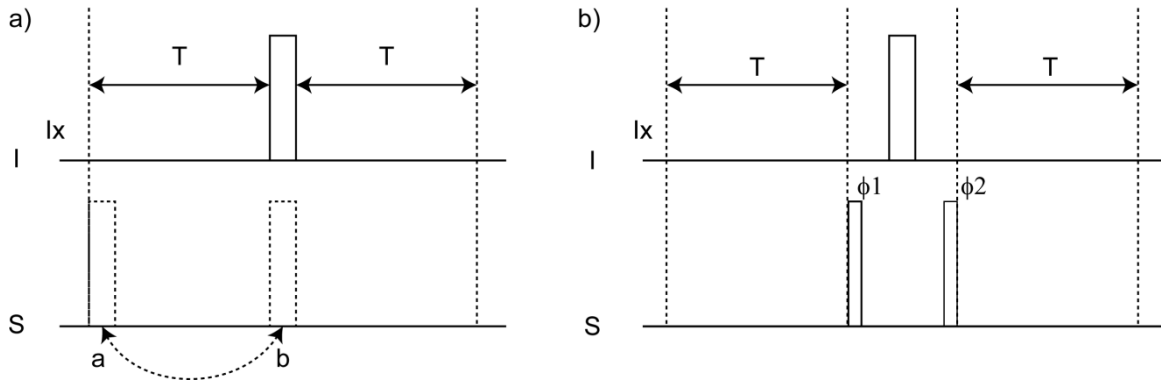


Figure 5.3 Schemes for Intensity modulation experiments.

The second scheme that is based on HMQC for Intensity modulation method is shown in Figure 5.3b. The input magnetization is also Ix. During the first T period, J_{IS} is active:

$$Ix \rightarrow Ix\cos(\pi J_{IS}T)+2IySz\sin(\pi J_{IS}T) \quad (4)$$

In the cross experiment, the second term was selected by phase cycle of $\phi 1$, $\phi 2$ and receiver phase. The $\phi 1$ 90 degree pulse on S turns it to multiple quantum terms.

$$IySzsin(\pi J_{IS}T) \rightarrow -IySy\ sin(\pi J_{IS}T) \quad (5)$$

In t1 evolution time, the chemical shifts of S spin are labeled. The second 90 degree pulse on S spin, $\phi 2$, converts it back to anti-phase terms. (The terms from t1 evolution are omitted for clarity.)

$$-IySy\ sin\ (\pi J_{IS}T) \rightarrow -IySz\ sin\ (\pi J_{IS}T) \quad (6)$$

In the second T period, this anti-phase term is rephrased:

$$-IySz\ sin\ (\pi J_{IS}T) \rightarrow -IzSy\ sin\ (\pi J_{IS}T)cos(\pi J_{IS}T)-Ix\ sin\ (\pi J_{IS}T)\ sin\ (\pi J_{IS}T) \quad (7)$$

So the final signal is proportional to $sin^2(\pi J_{IS}T)$.

In the reference experiment, by changing the phase of $\phi 1$, $\phi 2$ and receiver phase, the first term in equation (4) is selected. Following the same procedure, the result signal of reference experiment is proportional to $cos^2(\pi J_{IS}T)$.

After taking into account a proper scale factor, the ratio of signal intensity in “cross” and “reference” experiment is:

$$Icross/Iref=tan^2(\pi J_{IS}T)$$

J_{IS} couplings can be easily extracted. $^3J_{NH\beta}$ which conveys the information of the side chain torsion angle $\chi 1$ was measured in this way (250).

5.5 Motivation of the project

$^1J_{HN}$ is one of the most frequently measured J couplings in protein NMR. However, there is no equation yet that correlates this coupling with the backbone torsion angles. Moreover, there has not been any reference value of $^1J_{HN}$ in the "non-interacting" condition reported as yet. It is also unknown if the reference is residue type-dependent. In contrast to $^1J_{HN}$, two of other backbone one-bond couplings, $^1J_{C\alpha-H\alpha}$ and $^1J_{C\alpha-C\beta}$, are well known for their relations with backbone ϕ and ψ angles (36, 195). The residue type specific values of $^1J_{C\alpha-H\alpha}$ and $^1J_{C\alpha-C\beta}$ of disordered state have also been reported, and were used as the reference values in the Karplus equation (36, 195).

In theory, six spin-spin coupling constants should be affected by H-bonding. Four of them are directly affected because they are spin-spin coupling constants across the H-bond and two are affected indirectly because they involve atoms participating in the H-bonding. However, only the ^{13}C-^{15}N three-bond scalar couplings couple across the H-bond and the one-bond ^1H-^{15}N scalar coupling, $^1J_{HN}$, can be measured in proteins currently. The magnitude of the $^1J_{HN}$ spin-spin coupling constant is dependent upon the electronic structure of the molecule (251-253). Theoretical studies of small model systems have shown that $^1J_{HN}$ is negative and becomes more negative upon H-bond formation (252, 253). Moreover, conformational changes due to ligand binding can induce changes in $^1J_{HN}$ spin-spin coupling constants (254). $^1J_{HN}$ couplings might therefore be promising as potential hydrogen bond parameters.

In this research, we measured $^1J_{HN}$ in intrinsically disordered proteins to identify the $^1J_{HN}$ values at "non-interacting" states. We checked if there are reference values by looking at the distribution of $^1J_{HN}$ values, and also if the reference values are residue type-dependent. The stability of $^1J_{HN}$ was checked at different temperatures and pH. We then measured the $^1J_{HN}$ in folded proteins to explore whether any possible changes of $^1J_{HN}$ correlated with hydrogen bonds.

6. Methods and Materials

6.1 NMR samples

0.7mM ^{15}N labeled Full length Tau protein (1-441) was dissolved in 57mM Na-phosphate buffer, pH6.0, 42mM NaCl. 0.35mM ^{15}N labeled Tau K18 (245-370) was in 50mM Na-phosphate buffer, pH6.8, 100mM NaCl. The Tau proteins samples were provided by Dr. Eckhard Mandelkow from the German Center for Neurodegenerative Diseases (DZNE) in Bonn.

0.25mM ^{15}N α-synclein was in 50mM Na-phosphate buffer, pH6.0, 100mM NaCl. The $^{1}J_{NH}$ of Tau full length, Tau K18 and α-synclein were measured at 278K. To explore the pH and temperature effects, the α-synclein sample was adjusted to pH5.7, pH6.5 and pH7.4. At all three pH conditions, $^{1}J_{NH}$ couplings were measured at 278K and 288K. At pH5.7, the experiment was also carried out at 298K.

0.5mM ^{15}N labeled ubiquitin was dissolved in 50mM HEPE buffer, pH7.0, 300mM NaCl, which was measured at 298K. One 0.5mM $^{15}N/^{2}H$ labeled ubiquitin sample in the same condition was also used.

The α-synuclein and ubiquitin samples are provided by Dr. Stefan Becker from the Max Planck Institute for Biophysical Chemistry.

GCN (16-31) peptide was synthesized by Kerstin Overkamp of the Max Planck Institute for Biophysical Chemistry. The powder of GCN peptide was dissolved in 50mM Na phosphate buffer, and the pH was adjusted to 6.8.

Chemicals, instruments and software are included in the “methods and materials” section for SR proteins (Chapter 2.1).

6.2 Band-selective-decoupled (BSD) IPAP-HSQC

During the N evolution time in IPAP-HSQC pulsprogram, the couplings between N and aliphatic protons are effective. These long range couplings split the signals further. Resolutions on indirect dimensions normally are not enough to solve the multiple

components. Moreover, the cross-correlated relaxation effect makes the doublets asymmetrical, which distort the line shape of the observed signal (Figure 6.1A). This distortion is a source of systematic error (Figure 6.1B). In gradient-enhanced IPAP-HSQC, E.COSY-type effects, which result from the aliphatic protons coupled to both amide H and N, can further deteriorate the situation (Figure 6.1C).

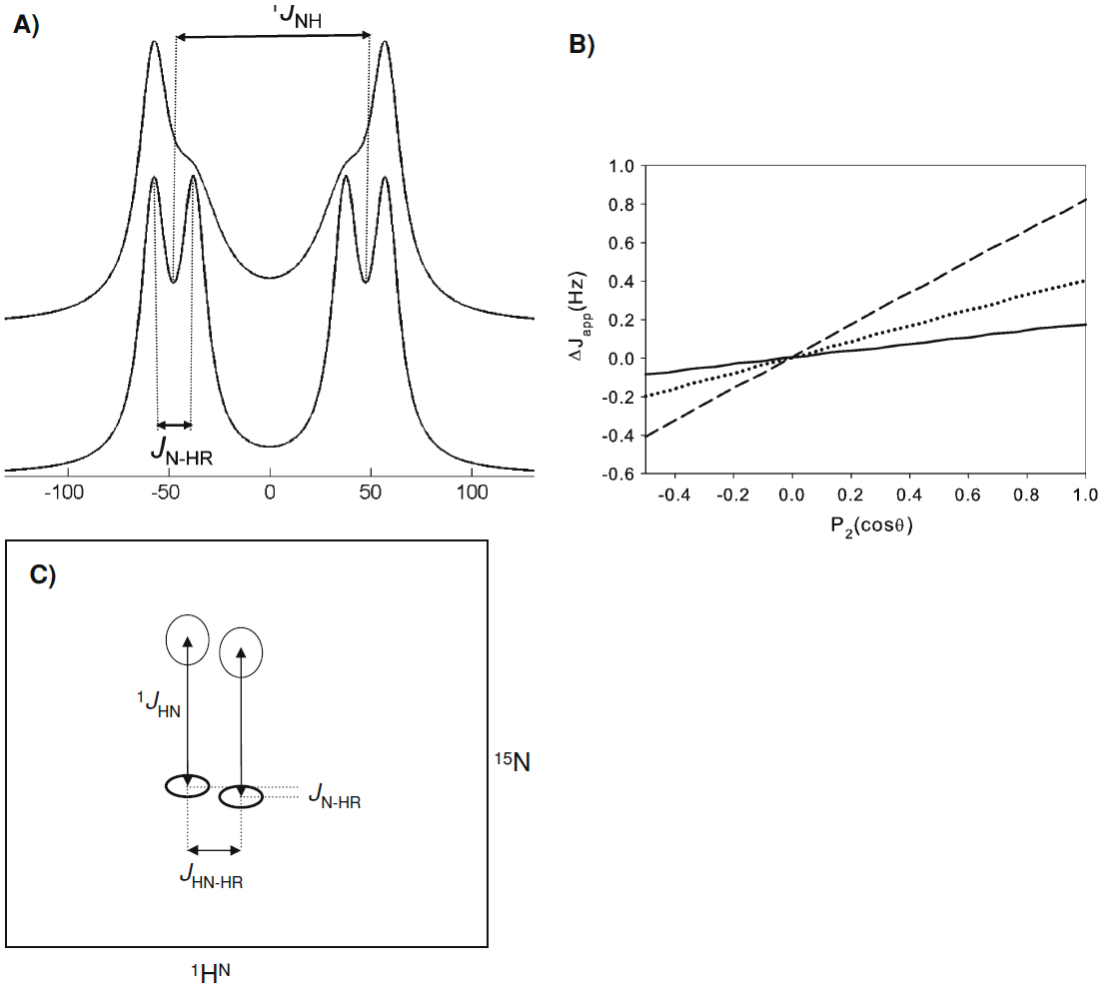


Figure 6.1 The impact of aliphatic proton spin, HR, on the measurement of $^1J_{HN}$. A) cross-correlated relaxation causes different line widths and intensities of the two components. The size of J_{N-HR} is exaggerated for visual purposes. B) Error of $^1J_{HN}$ splitting ΔJapp as a function of $P_2(\cos\theta)$. θ is the H-N-HR angle. Three values of J_{N-HR} are used in the calculation: 1.5Hz (solid line), 3Hz (dotted line) and 4.5Hz (dashed line). C) E.COSY pattern obtained in gradient-enhanced 15N-1H transfer scheme. (Reproduced from (244))

An improved version of IPAP-HSQC was proposed as Band-selective-decoupled (BSD) IPAP-HSQC (244). The pulse program is similar to normal IPAP-HSQC, with one shaped pulse in the middle of the N evolution time to refocus couplings between aliphatic proton and N. This refocusing pulse eliminated the cross-correlation effects, multiple fine structures of peaks, as well as the E.COSY pattern mentioned above. To compensate the off-resonance effect on the amide proton during this pulse, the same pulse was applied at beginning of the N evolution and followed by one 180° pulse on N. The pulse program is shown in Figure 6.2.

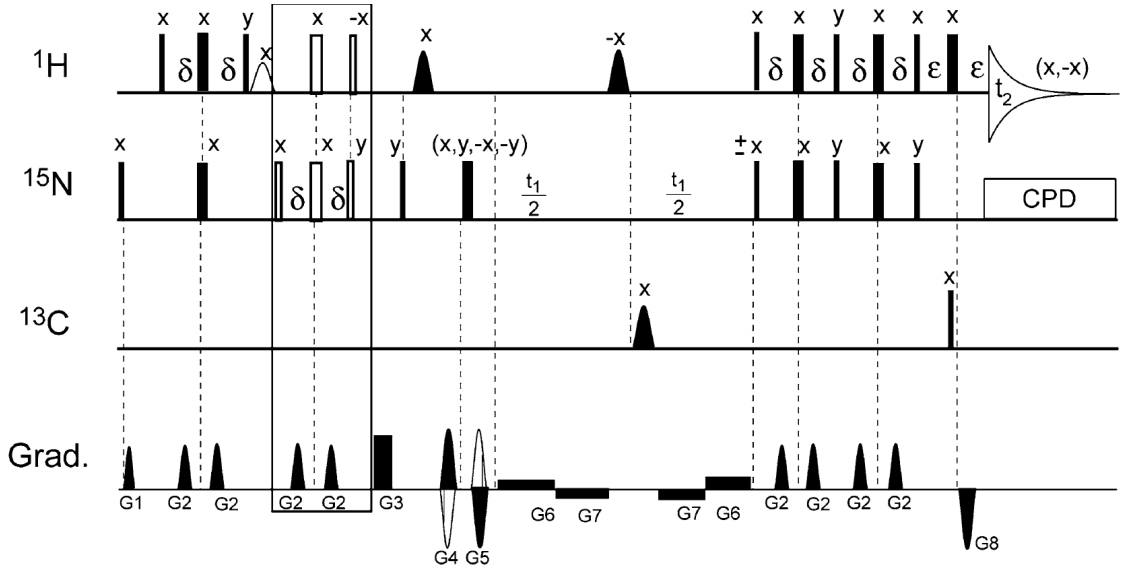


Figure 6.2 Scheme of BSD-IPAP-HSQC. The pulses in the box were only applied in the anti-phase spectrum and were skipped in the in-phase spectrum. The water flip back pulse is a 1.5ms sine-bell shaped pulse, which is only applied in in-phase spectrum. The solid proton shaped pulse were REBURP pulses, centered at 2.4ppm and covering a bandwidth of 2.8ppm. One broadband inversion pulse (BIP)(255), BIP 720.100.10 center at 116 ppm to decouple Cα and CO at the same time. Rance-Kay quadrature detection on t1 was achieved by alternating the phase of the last 90° pulse of the N evolution time between x and –x, and inverting the gradient G4 and G5 at the same time. δ= 2.65ms and ε=0.55ms.

The $^{1}J_{NH}$ couplings of ubiquitin are measured at 400M, 600M, 900M to correct the effects of dynamic frequency shifts (DFS) and residue dipolar couplings which are

caused by anisotropic magnetic susceptibility of protein (243). BSD-IPAP-HQSC were recorded as interleaved in-phase and anti-phase spectra with data matrix 256(t1)*512(t2) complex points for each dimension. Data processing followed the methods outlined in a previous paper (244).

Tau K18 (245-370) was also measured by BSD-IPAP-HSQC on the Bruker Avance 600M and 400M spectrometer with z-axis gradient, triple resonance, and room temperature probes (244). In-phase and anti-phase spectra were recorded interleaved with data matrix 256 (t1)*512(t2) complex points of each dimension.

Protein G $^{1}J_{NH}$ couplings at 500M and 750M were taken from one previous research, which was measured at 25 mM sodium phosphate, 50 mM NaCl, 0.05% w/v NaN3, pH6.5 (244). The DFS are calculated as previously reported (243, 249). After DFS was corrected, the residue couplings due to anisotropic magnetic susceptibility were rectified by taking the difference between $^{1}J_{NH}$ at different fields.

6.3 Intensity modulated HSQC

Intensity modulated HSQC was used to measure $^{1}J_{NH}$. The pulse program is shown in Figure 6.3 (249).

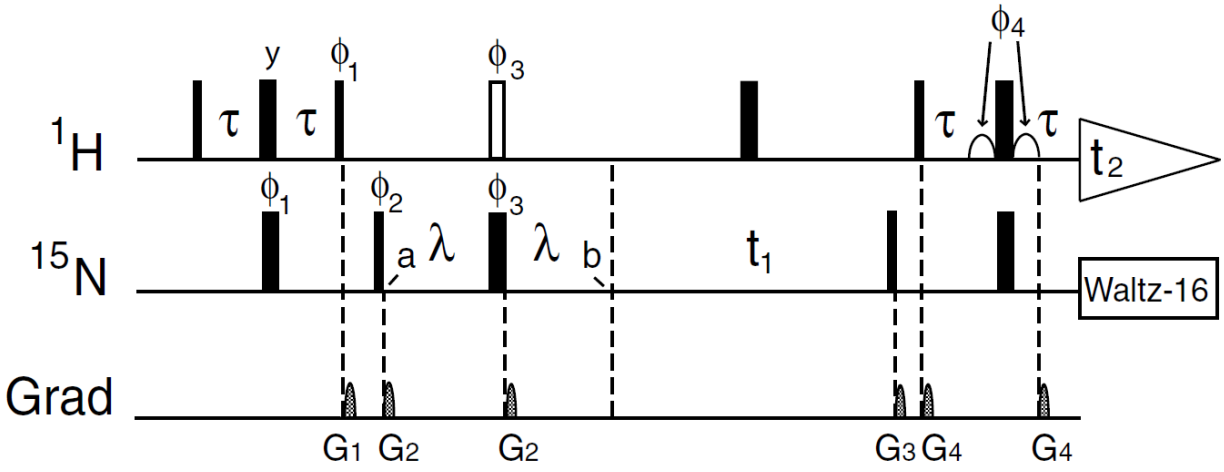


Figure 6.3 Pulse scheme of intensity modulated HSQC. Narrow and wide bars correspond to 90° and 180° pulses. Small white bell shapes are 1ms 90° water flip back pulses in the WATERGATE module for water suppression. The white bar indicates the REBURP pulse which selectively flips amide protons. Except where indicated, all pulses are applied on x. Φ1=8(y),8(-y); Φ2=x,-x; Φ3=2(x),2(y),2(-x),2(-y); Φ4=-x, receiver=x,2(-x),2(x), 2(-x),x,-x, 2(x), 2(-x), 2(x),-x.τ=2.25ms. (Reproduced from (249) with permission)

At point a, anti-phase term 2NyHz is created. During the delay 2λ, $^{1}J_{NH}$ is effective. The terms evolved during 2λ, at point b as follows :

$$2\mathrm{Ny}H_z\cos(\pi \mathrm{J}2\lambda) - N_x\sin(\pi \mathrm{J}2\lambda)$$

After 15N chemical shift evolution t1, the term $2\mathrm{Ny}H_z$ is transferred back to H and results in an observerable signal. So at the end, the single intensity can be described as :

$$I = (A_0\cos(\pi \mathrm{J}2\lambda) + A_1)\exp(-A_2 2\lambda) \qquad (8)$$

A series of experiments are recorded with 15 different λ delays. To minimize the deviation between them, the spectra were recorded in a interleaved way. λ delays were set in such a way that the intensity sampled at 3 zero-crossing. J values are determined by fitting the intensities in each spectrum and λ delays to the equation. A0 is the intensity in the reference spectra, which is usually the first point. The exponential function $\exp(-A_2 2\lambda)$ is the decay of signals due to relaxation. Since the shaped pulse is only the flip amide proton, only one bond of the J couplings are effective during the 2λ delay. The long range coupling discussed in BSD-IPAP-HSQC case are also refouced.

The $^{1}J_{NH}$ of Tau full length protein (1-441), Tau K18 (245-370) and α-synclein at pH6.0 was measured by $^{1}J_{NH}$ modulated HSQC. Data of Tau full length protein was recorded on Bruker Avance 900M spectrometers with z-axis gradient, triple resonance, and cryogenic probes. Tau K18 (245-370) and α-synclein were recorded in the Bruker Avance 700M spectrometer with z-axis gradient, triple resonance, and room temperature probes. The data were recorded interleaved of 15 mixing times: 31.9ms, 31.1ms, 30.7ms, 30.3ms, 29.9ms, 26.5ms, 25.5ms, 25.1ms, 24.6ms, 24.3ms, 21.2ms, 20.3ms, 19.8ms,

19.5ms, and 19.1ms. One 2-dimentional spectrum was recorded at each time point, with data matrix 334(t1)*512(t2) complex points each dimension for Tau, and 256(t1)*512(t2) complex points for Tau K18 (245-370) and α-synclein. The $^{1}J_{NH}$ values were determined by fitting intensities as described previously (249).

The intensities on the same position of each 2D spectrum was read out, and fit to equation (8) mentioned above with Python scripts (found in the Appendix).

7. Result

7.1 The residues-specific $^1J_{HN}$ couplings in intrinsically disordered proteins

The first step to find out the structural properties related to $^1J_{HN}$ couplings is to determine the coupling values of the random coil state as reference. To derive the reference values, we measured the $^1J_{HN}$ scalar couplings in the 441-residue intrinsic disordered protein Tau, which is known to be absent of stable secondary structures, at a static magnetic field strength of 21.1T and a temperature of 278K, pH6.0. A $^1J_{HN}$ modulated HSQC experiment was used to measure 1JHN scalar couplings. The length and native disordered nature of the Tau protein allow measuring large number 1JHN scalar couplings at the same time, and in a condition free of H-bond and chemical denaturants. 367 couplings are obtained which covers all non-proline amino acid types except tryptophan, which is not present in the primary sequence. The values are sorted by amino acid types and mean values are calculated (Table 7.1). As seen in Figure 7.1, the coupling values are dependent on the amino acid type. Glycine has the biggest $^1J_{HN}$ magnitude of 94.23Hz, while valine has the smallest $^1J_{HN}$ magnitude of 93.00Hz. Within the same amino acid type, maximum and minimum values differ by less than 0.6 Hz. Standard deviations within one amino acid type are less than 0.15Hz. The difference within the same amino acid type is smaller than the difference in all residues together as a whole. We propose these mean values as the random coil value for each amino acid type.

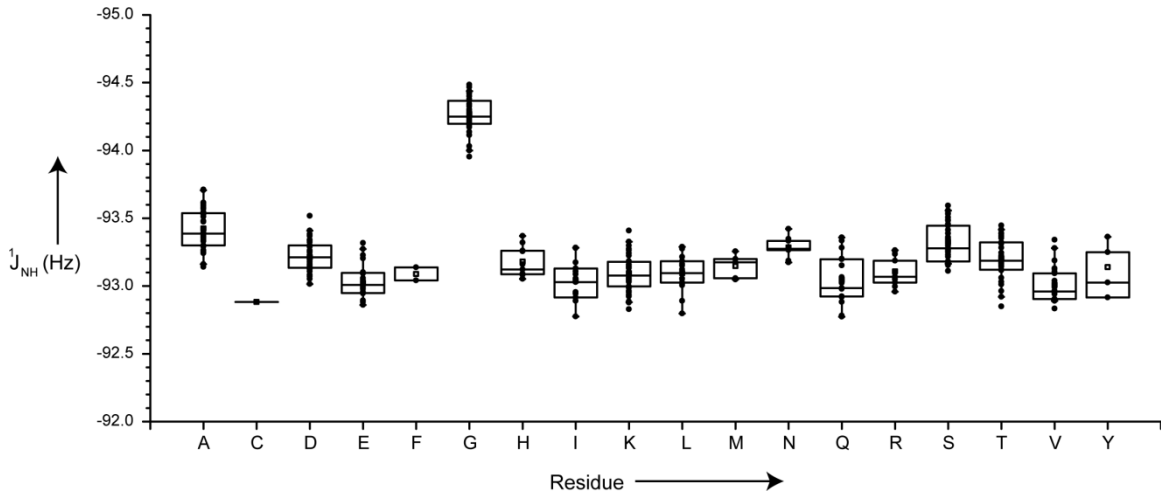


Figure 7.1 $^{1}J_{HN}$ spin-spin coupling constants observed in the intrinsically disordered protein Tau. $^{1}J_{HN}$ were grouped according to the amino acid type and subjected to a box plot analysis. The centre of the box indicates the mean value for each residue type, while the bottom and top correspond to the 25th and 75th percentile, respectively. The end of the whiskers indicates the 95th and 5th percentiles.

Amino acid	# of exp. values in Tau	Random coil value of $^{1}J_{NH}$(Hz)	Standard deviation of $^{1}J_{NH}$(Hz)
Alanine	32	-93.4	0.15
Arginine	12	-93.12	0.10
Asparagine	10	-93.27	0.07
Aspartic acid	29	-93.2	0.13
Cysteine	2	-92.88	0
Glutamine	18	-93.05	0.17
Glutamic acid	27	-93.01	0.13
Glycine	41	-94.23	0.13
Histidine	10	-93.17	0.11
Isoleucine	14	-93.03	0.15
Leucine	19	-93.08	0.12
Lysine	43	-93.08	0.13
Methionine	5	-93.13	0.09
Phenylalanine	3	-93.08	0.07
Serine	43	-93.3	0.14
Threonine	33	-93.2	0.14
Tyrosine	4	-93.13	0.2
Valine	27	-93.00	0.12

Table 7.1 Amino-acid specific random coil values of $^{1}J_{NH}$ spin-spin coupling constants.

The deviations of Tau $^{1}J_{NH}$ couplings from mean values are plotted against the sequence in Figure 7.2. All residues fell into the -0.35Hz~0.35Hz region around mean values, while the majority are restricted to the -0.2Hz~0.2Hz region.

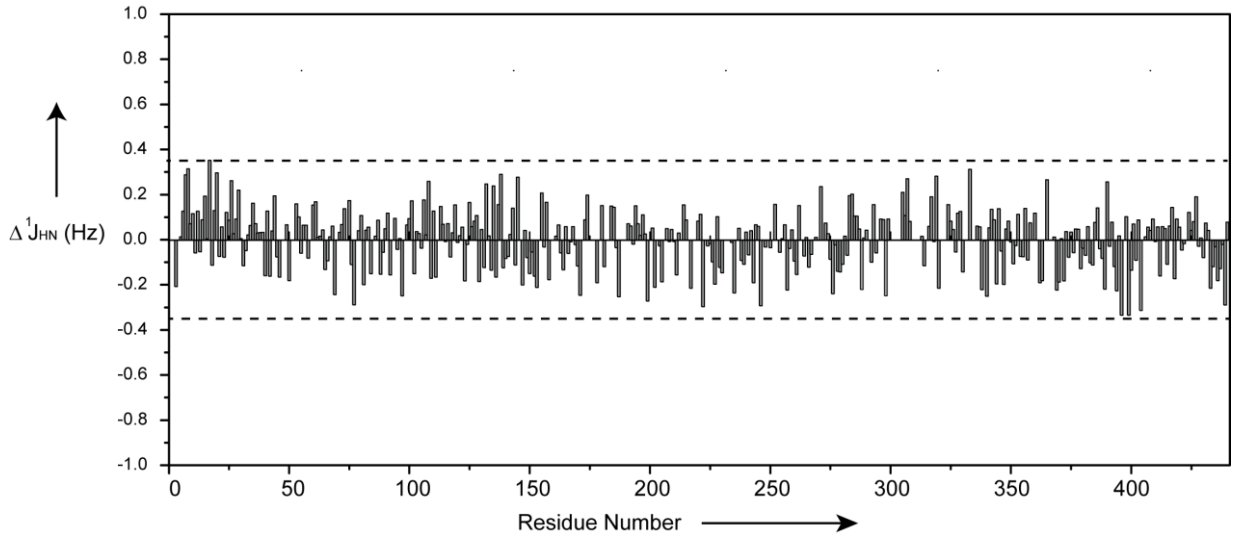


Figure 7.2 Deviation of $^{1}J_{NH}$ spin-spin coupling constants observed in full length Tau from its amino-acid specific random coil values as a function of residue number. The two dash lines indicate the -0.35Hz ~0.35Hz region.

To test the universality of the values derived from Tau, we also measured $^{1}J_{NH}$ scalar couplings of another intrinsically disordered protein, α-synuclein, in the same condition and by the same method. Deviations of α-synuclein $^{1}J_{NH}$ from the mean values are plotted against sequence in Figure 7.3. All residues except K96 are in the range of -0.35Hz ~0.35Hz. Most residues are in the range of -0.2Hz~0.2Hz, with a few exceptions, such as E137 and D119.

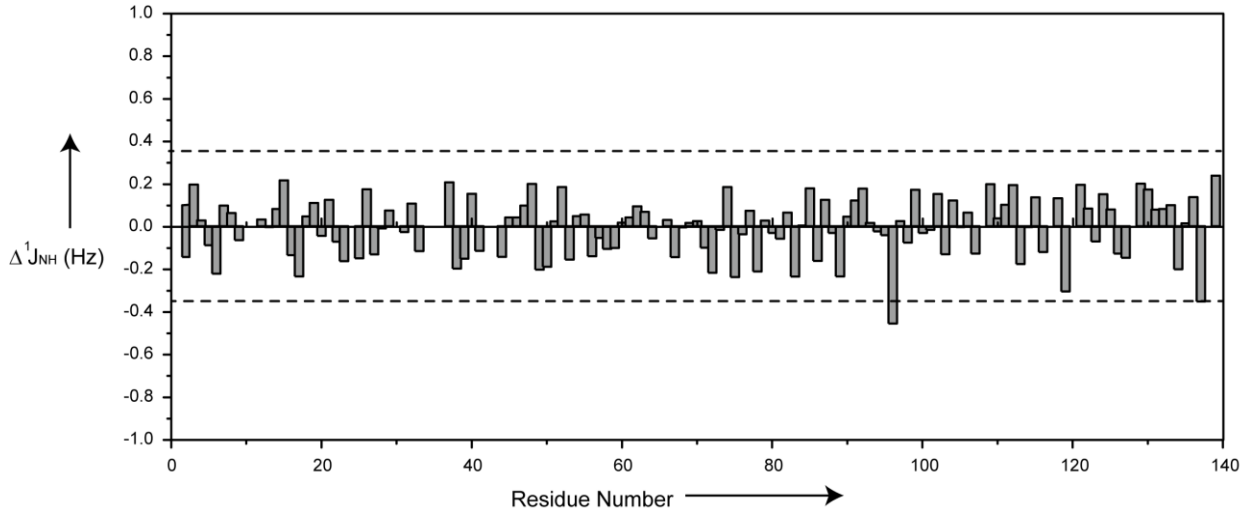


Figure 7.3 Deviation of $^{1}J_{NH}$ spin-spin coupling constants observed in α-synuclein from its amino-acid specific random coil values as a function of residue number. The two dash lines indicate the region between -0.35Hz ~0.35Hz.

To further trace any possible systematic offset between Tau and α-synuclein scalar couplings, the mean values of every residue type from these two proteins are compared (Figure 7.4). The differences between mean values are less than 0.1Hz for all residue types except Ile, which is 0.1Hz. However, there is only 2 Ile in α-synuclein, so the difference may be due to small number of occurrences. Consistency of couplings measured from Tau and α-synuclein shows that the $^{1}J_{HN}$ is not dependent on the protein but only residue types.

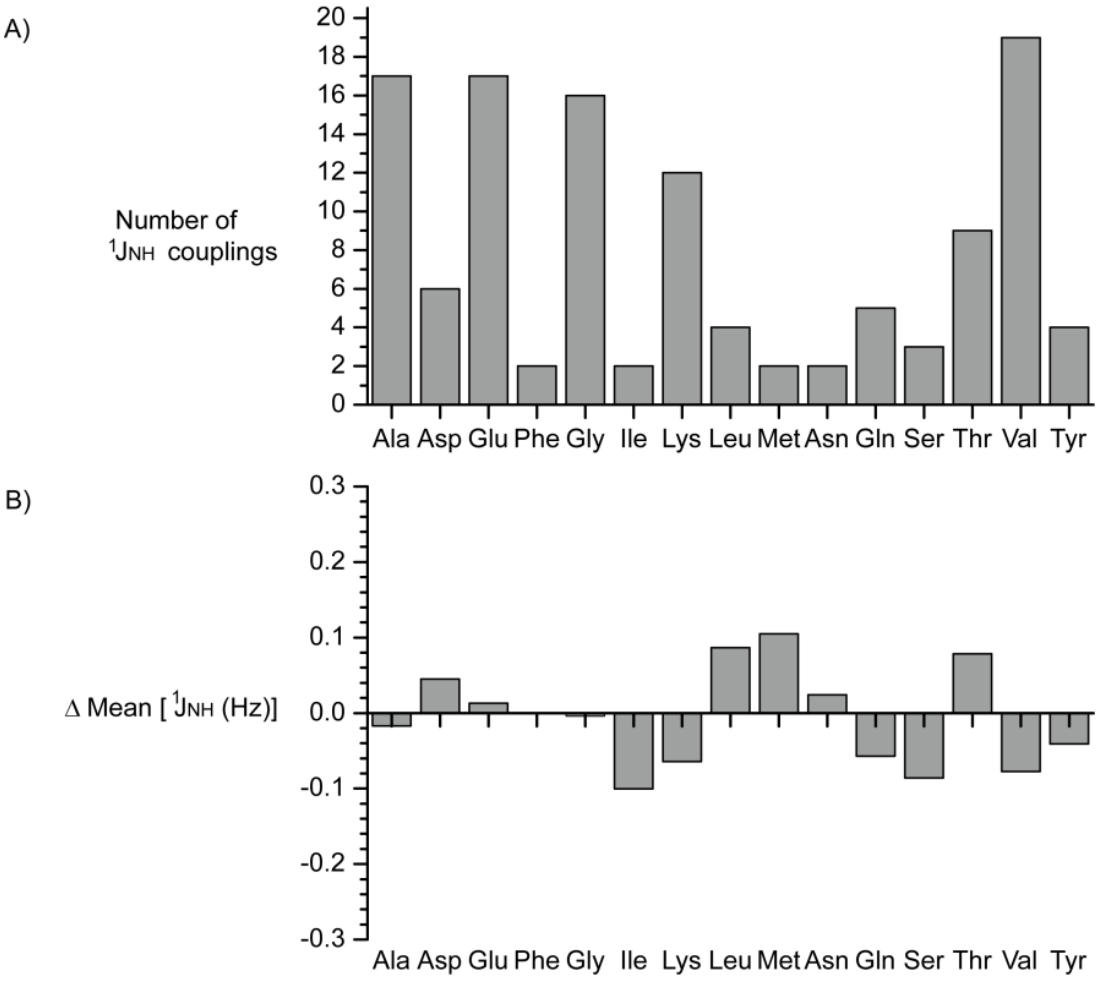


Figure 7.4 Changes in amino-acid specific values of $^{1}J_{NH}$ in the Tau and α-synclein. A) Number of experimental $^{1}J_{NH}$ spin-spin coupling constants for the most common residue types in a-synuclein. B) Differences between mean values of amino-acid specific $^{1}J_{NH}$ constants in Tau and α-synuclein.

7.2 The $^{1}J_{NH}$ scalar couplings in intrinsically disordered proteins are stable at different temperatures and pH

To assess the stability of $^{1}J_{NH}$ scalar coupling values, we measured the $^{1}J_{NH}$ of α-synclein at different temperatures and pH values. The same α-synclein sample was titrated from pH 5.7, pH 6.5 to pH 7.4. At each pH, $^{1}J_{NH}$ values were measured at 278K and 288K. At pH 5.7, the sample was also measured at 298K. The values of different pH or temperatures are compared. The correlation plots are in Figure 7.5, in which the data points are evenly distributed around y=x line. The data are measured by routine IPAP HSQC, in which the long range couplings, such as $^{2}J_{NH\alpha}$, $^{3}J_{NH\beta}$ are not refocused. The RMSD between the groups of $^{1}J_{HN}$ values in comparison are less than 0.15Hz, which is on the range of experimental error of such pulse program (~0.2Hz).

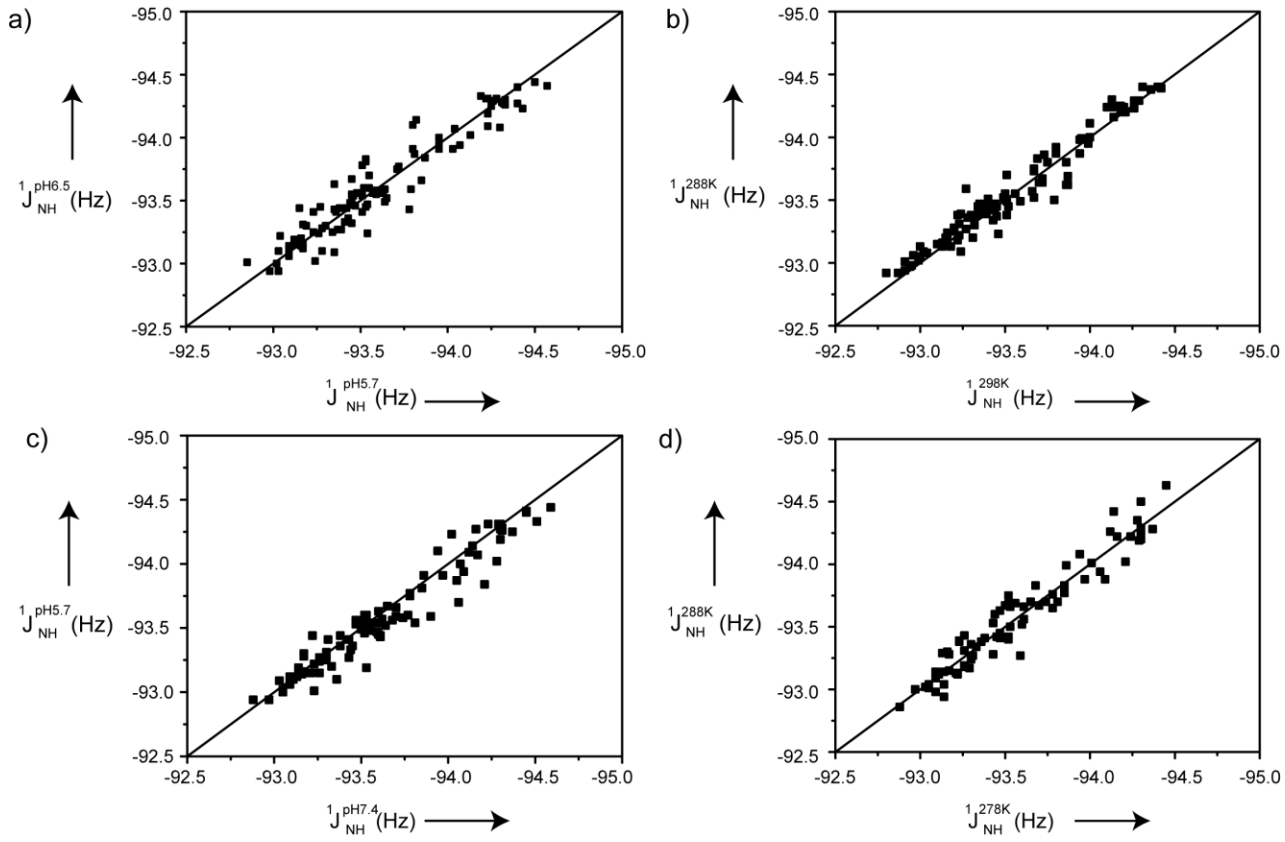


Figure 7.5 a) Correlation between $^1J_{NH}$ of α-synuclein at pH 6.5 and pH5.7, 278K. b) Correlation between $^1J_{NH}$ of α-synuclein at 288K and 298K, pH 5.7. c) Correlation between $^1J_{NH}$ of α-synuclein at pH 5.7 and pH7.4, 278K. d) Correlation between $^1J_{NH}$ of α-synuclein at 278K and 288K, pH 7.4. The y=x line are drawn on all correlation plots.

Correlations	RMSD (Hz)
pH5.7 vs pH6.5,278K	0.15
288K vs 298K, pH5.7	0.09
pH5.7 vs pH7.4, 278K	0.15
278K vs 288K, pH 7.4	0.12

Table 7.2 Residue mean square-root of deviations (RMSD) between $^1J_{NH}$ values of different conditions.

The mean values at different conditions are compared to ensure that there are no systematic drifts with pH or temperature change. As seen from Figure 7.6, the mean

values of different conditions differ around 0.1 Hz or less. It is proved that $^{1}J_{NH}$ is stable within the range of near physiological condition.

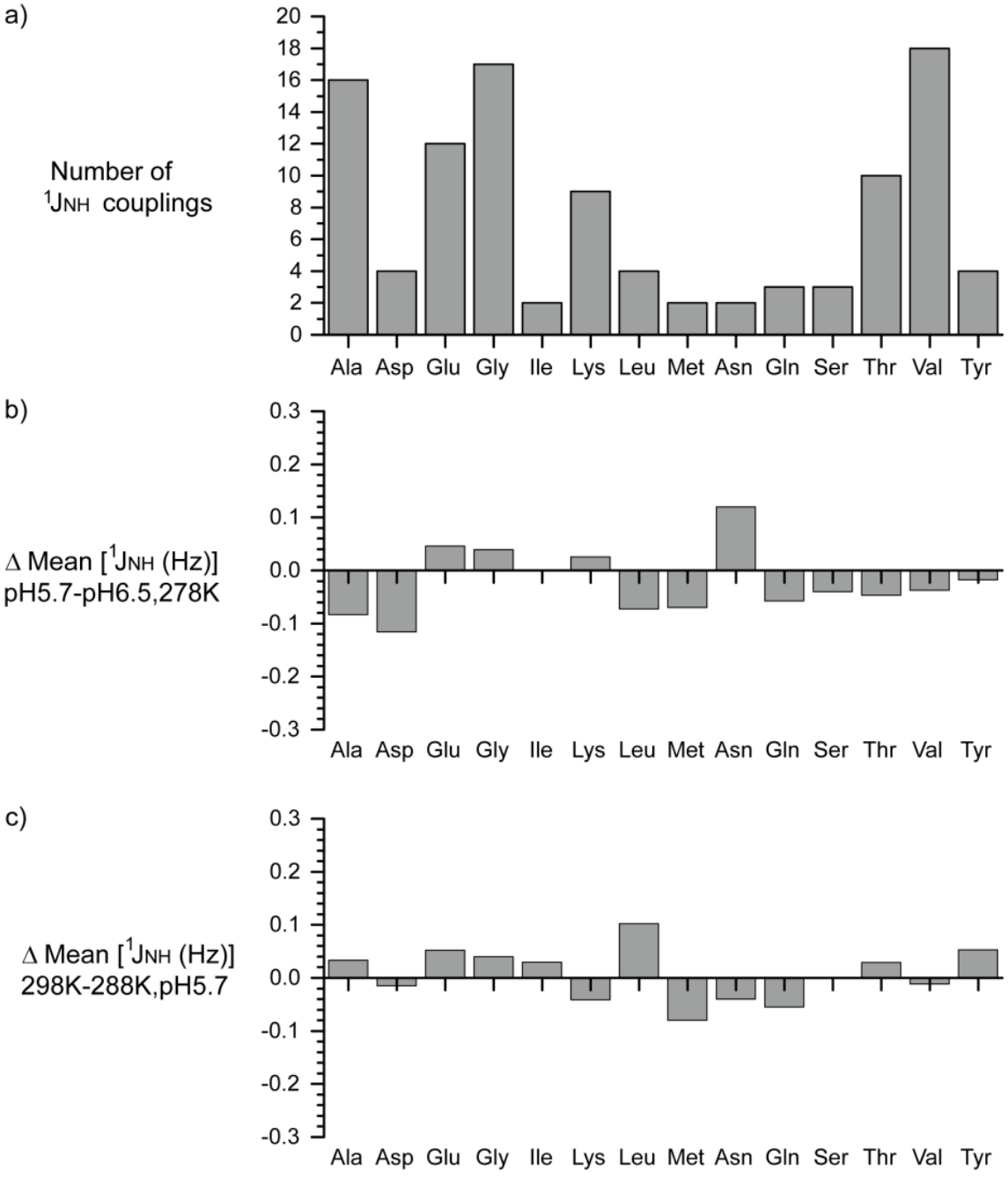


Figure 7.6 Changes in amino-acid specific values of $^{1}J_{NH}$ in the intrinsically disordered proteins a-synuclein with pH and temperature. a) Number of experimental $^{1}J_{NH}$ spin-spin coupling constants for the most common residue types in α-synuclein. b)

Differences between mean values of amino-acid specific $^1J_{NH}$ values in α-synuclein at pH5.7 and pH6.5, 278K. c) Differences between mean values of amino-acid specific $^1J_{NH}$ values in α-synuclein at 298K and 278K, pH6.5.

7.3 Field strength dependences of $^1J_{NH}$ couplings are negligible for IDP proteins.

The $^1J_{NH}$ scalar couplings were measured at different field strengths to check the effects of dynamical frequency shift (DFS) and anisotropic magnetic susceptibility. One fragment of Tau, K18 (245-372), was measured by BSD-IPAP-HSQC at 600MHz and 400MHz.The values measured by the BSD-IPAP-HSQC at different field strengths are compared. The correlation plots are shown in Figure 7.7. Linear fitting between two datasets give an offset as -0.016 Hz.

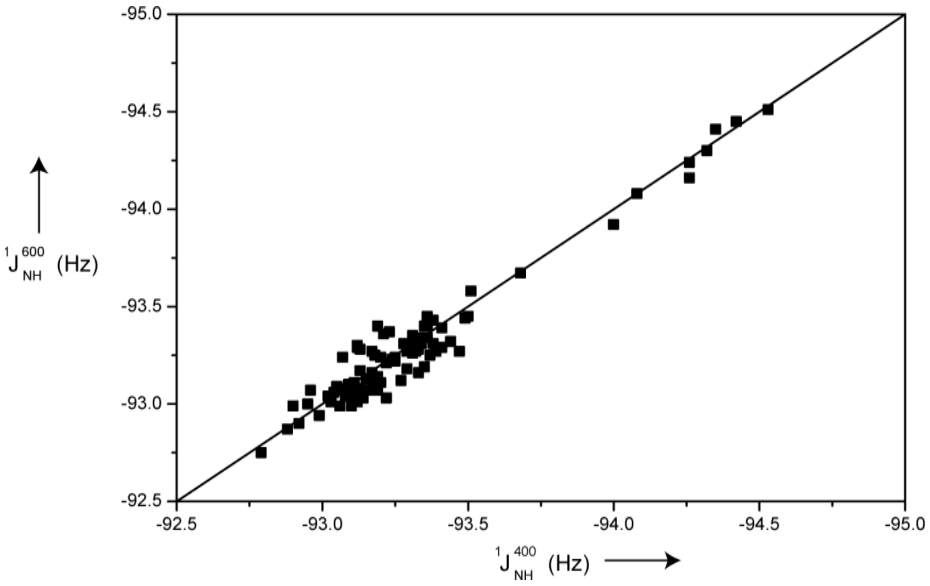


Figure 7.7 Correlation plot of $^1J_{NH}$ of K18 at different magnetic field, pH6.0, 278K. $^1J_{NH}$ values measured by BSD-IPAP-HSQC at 600MHz and 400MHz.The y=x line was drawn on the picture.

The $^1J_{NH}$ of an intrinsic disordered protein does not differ significantly at different magnetic fields, which mean the mean $^1J_{NH}$ values of Tau measured at 900MHz can be applied to data at different field strengths. Due to the fast backbone motion and absence

of 3D structure in intrinsically disordered proteins, these two effects are trivial, so the values at different field strength should be identical.

7.4 Systematic offset between intensity modulated HSQC and BSD-IPAP-HSQC

The $^{1}J_{NH}$ scalar couplings were measured by two different methods: $^{1}J_{NH}$ modulated HSQC and IPAP HSQC. The data from the two experimental methods are compared. The correlation plot of $^{1}J_{NH}$ by intensity modulated HSQC and BSD-IPAP-HSQC are shown in Fig8. There is a slight offset between two datasets, ~0.05Hz. The values from two methods are very close.

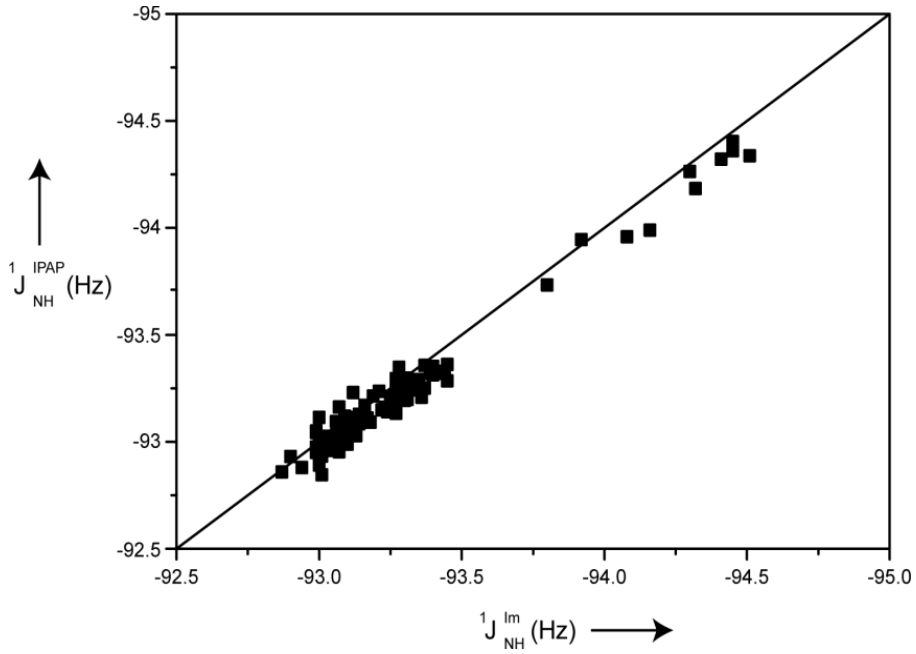


Figure 7.8 Correlation between $^{1}J_{NH}$ of K18 measured by intensity modulation and BSD-IPAP method. The values by intensity modulation HSQC are on average 0.05Hz lower than the value measured by BSD-IPAP-HSQC experiment. The y=x line was drawn on the picture.

The origin of this offset may be due to the different flipping pulses used during the J coupling effect periods in the two experiments. In the intensity modulated experiment, the magnetization of amide protons were flipped by a REBURP pulse and aliphatic

protons were kept on the xy plane, while the aliphatic protons were flipped and amide protons were kept on xy plan in the BSD-IPAP pulse program.

7.5 Negative secondary 1J HN coupling values indicate H-bond

The $^{1}J_{HN}$ values of ubiquitin were measured by BSD-IPAP-HSQC pulse program at 400MHz, 600MHz and 900MHz. The folded proteins have much lower motion, which make the DFS no longer negligible. Dynamic frequency shifts (DFS) at different fields were calculated using the previously reported formula. The relatively rigid folded protein molecules are aligned due to anisotropic magnetic susceptibility, which causes residue dipolar couplings on top of scalar couplings. After DFS was corrected, the residue dipolar couplings were corrected by measuring the difference between J coupling values at various fields.

The DFS values of ubiquitin were calculated by the formula 3 in reference (243). The DFS values of the GB3 protein were directly taken from previous reported values in reference (244). The DFS values of ubiquitin are 0.39Hz, 0.35Hz, 0.32Hz, and 0.23Hz at 900M, 700M, 600M and 400M field strengths respectively. The DFS values of GB3 protein is 0.38Hz and 0.28Hz at 750M and 500M field strengths, given the correlation time 3.3ns and ignore internal dynamic (244, 256).

After two corrections, $^{1}J_{HN}$ couplings of the ubiquitin, which are free of DFS and residue dipolar couplings are obtained. These $^{1}J_{HN}$ couplings are compared to random coil values, and plotted again sequence in Figure 7.9a.

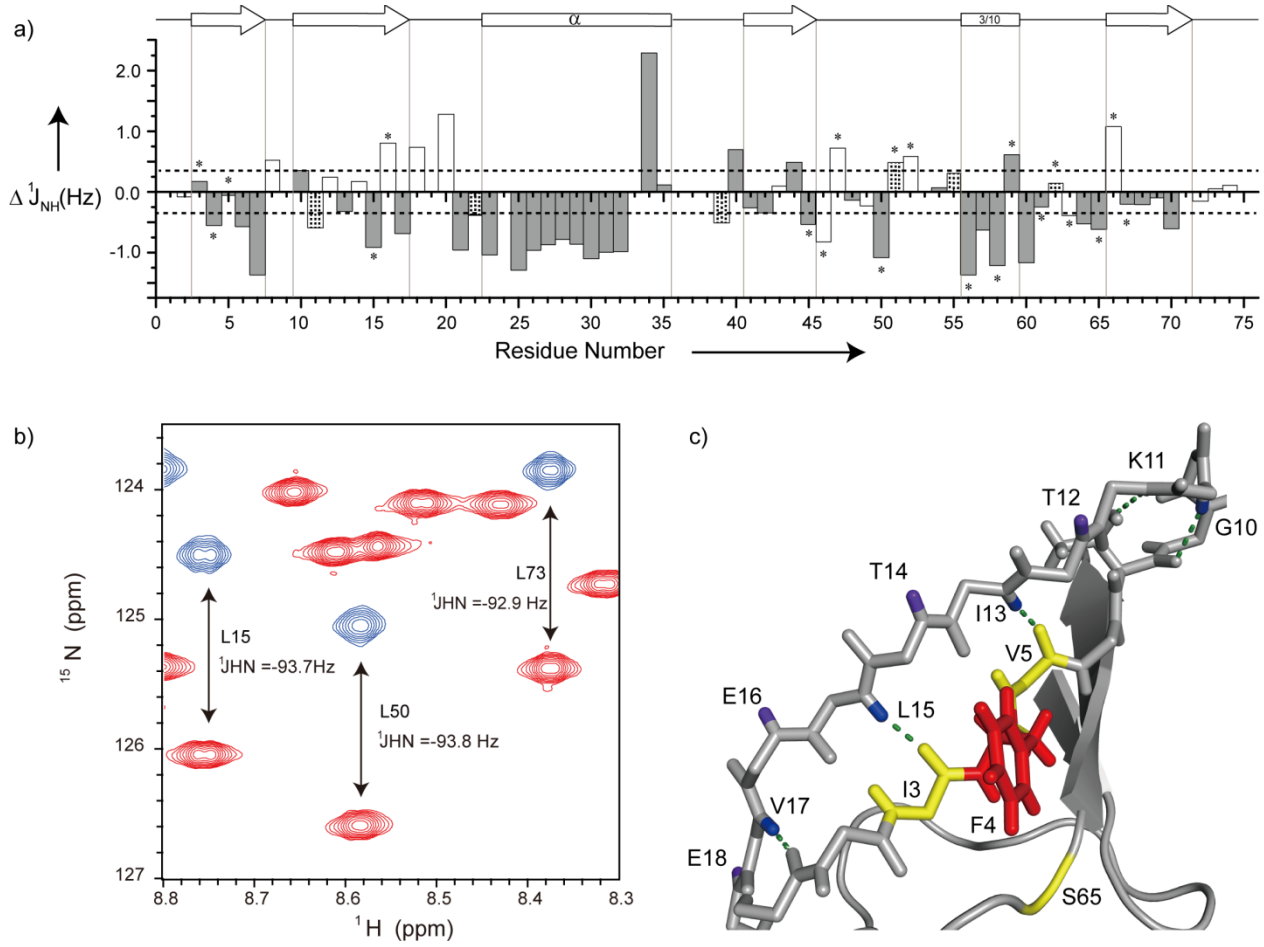


Figure 7.9 $^{1}J_{NH}$ spin-spin coupling constants in ubiquitin. a) Deviation of $^{1}J_{NH}$ spin-spin coupling constants observed in ubiquitin from its amino-acid specific random coil values as a function of residue number. Filled and hatched bars mark amide protons involved in a H-bond to the backbone or a side chain, respectively. Stars indicate spatial proximity (< 6 Å) to aromatic rings. The location of secondary structure elements as found in the crystal structure (PDB code: 1ubq) is shown above. b) Part of the BSD-IPAP hsqc spectra. The upper and lower doublets are shown in red and blue. c) Expanded small region of the crystal structure of ubiquitin shows the H-bonds between β strand 1 and 2. H-bonds are in green. Amide protons, which are involved in H bonds, are colour coded in blue while those not in H-bonds are in purple. The aromatic ring of Phe4 (red) with residues in spatial proximity (< 6 Å) shown in yellow.

Deviations of $^{1}J_{HN}$ couplings from their random coil values in folded proteins are different from the case of intrinsically disordered proteins. The secondary $^{1}J_{HN}$ couplings

which are defined as the deviation from random coil values, $\Delta^1 J_{HN}$, span the range of -2 Hz to 1.5Hz, which is a much larger range compared to the same ranges in intrinsically disordered protein cases (α-synclein, Tau). The residues in secondary elements have bigger $\Delta^1 J_{HN}$ values, while smaller values are present in the loop regions. Strikingly, the majority of the first α-helix, Ile23-Asp32, shows continuous negative values of $\Delta^1 J_{HN}$ with an average of 0.7Hz. High $\Delta^1 J_{HN}$ values were also present in 3/10 helix region. In β-strand, the trends were not clear, and the magnitudes of $\Delta^1 J_{HN}$ are smaller. In β-strand 1, 3, 4, the majority is negative with exceptions at the end. In β-strand 2 an alternating pattern of positive and negative $\Delta^1 J_{HN}$ values is observed.

Considering the position of the H-bond, negative $\Delta^1 J_{HN}$ values are predominantly correlated with H-bonded amid protons. The continuous negative values in α-helix and 3/10 helix were due to H-bond networks within the helix. β-strand 2 is on the edge of β-sheets of ubiquitin, so every second amide proton was involved in an H-bond while the other is solvent-exposure, which results in the alternating positive and negative pattern of β-strand 2 . For the residues in proximity to aromatic rings, the correlation is less prominent due to the coupling constant perturbation from the paramagnetic spin-orbit Ramsey term.

$^1J_{NH}$ scalar couplings of the IgG-binding domain of protein G, GB3, were taken from reference, which was measured by BSD-IPAP –HSQC at 500MHz and 750MHz.The same correction procedure was performed, and the deviation from random coil values are plotted as a function of residue number in Fig10.

GB3 has significantly more aromatic residues in its sequence compare to ubiquitin, 6 out of 56 versus 4 out of 76. The disturbance from aromatic rings is more severe in this protein, with many residues in proximity of 2 or even 3 aromatic rings. However, the correlation between H-bond and negative $\Delta^1 J_{HN}$ is still prominent, particularly in α-helix. The second β-strand, from residue 14 to 20, also shows a similar alternative pattern as in the second β-strand of ubiquitin. This β-strand is also on the edge of the hydrophobic core of GB3, in which the amide proton of every second residue forms an H-bond with β-strand 1 and the others point to the solvent.

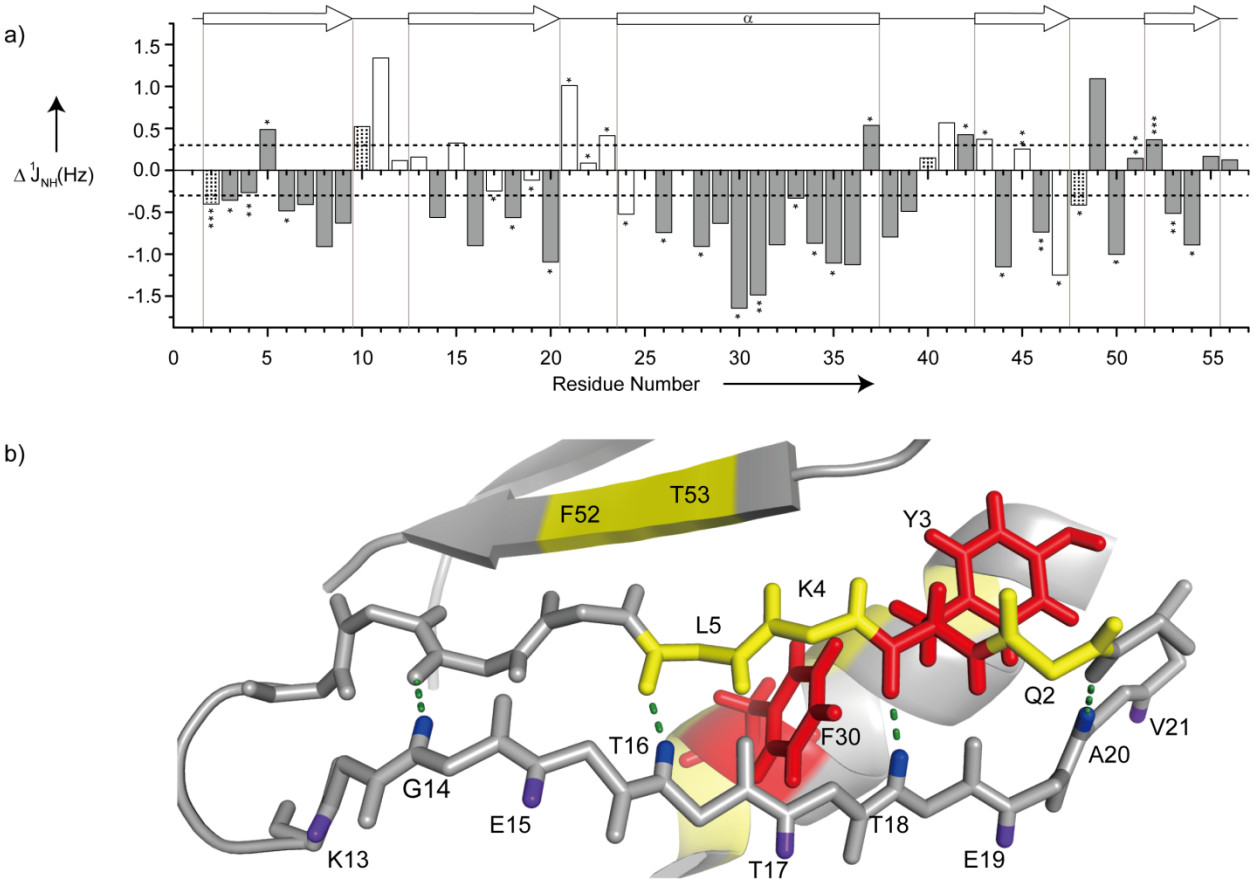


Figure 7.10 $^{1}J_{NH}$ spin-spin coupling constants in protein G. Deviation of $^{1}J_{NH}$ spin-spin coupling constants observed in protein G from their amino-acid specific random coil values were plotted as a function of residue number. Filled bars mark amide protons involved in a backbone H-bond. Stars indicate spatial proximity (< 6 Å) to aromatic rings. Two or three stars indicate that two or three aromatic rings are within 6 Å of either the amide or amide proton. The location of secondary structure elements as found in the 3D structure (PDB code: 2oed) is shown above. (b) Expanded small region of the 3D structure of protein G to show the H-bonds between β strand 1 and 2. . H-bonds are in green. Amide protons, which are involved in H-bonds, are colour coded in blue while those not in H-bonds are in purple. The residues, which are in spatial proximity (< 6 Å) of aromatic rings of Tyr3 and Phe30 (red), are shown in yellow.

$^1J_{NH}$ scalar couplings of another folded protein, DINI, was recorded by routine IPAP-HSQC pulse program, in which the long range scalar couplings $^2J_{NH\alpha}$, $^3J_{NH\beta}$ were effective. This increases the uncertainty of experimental data. Without any correction, the couplings are compared to random coil values (Figure 7.11). The residues in two α-helixes predominantly have negative Δ^1J_{NH} values with few exceptions.

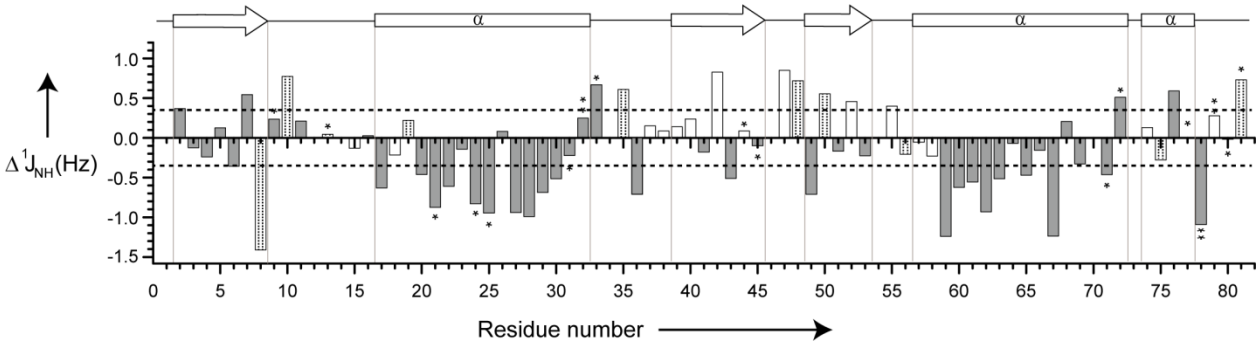


Figure 7.11 $^1J_{NH}$ spin-spin coupling constants in DINI. Deviation of $^1J_{NH}$ spin-spin coupling constants observed in DINI from their amino-acid specific random coil values were plotted as a function of residue number. Filled bars mark amide protons involved in a backbone H-bond. Stars indicate spatial proximity (< 6 Å) to aromatic rings. Two stars indicate that two or three aromatic rings are within 6 Å of either the amide or amide proton. The location of secondary structure elements as found in the 3D structure (PDB code: 1ghh) is shown above.

In all 3 cases, $^1J_{NH}$ scalar couplings of the residues involved in H bonds are higher than random coil coupling values. Without interference from nearby aromatic rings, the residues involved in backbone H-bonds show $^1J_{NH}$ scalar couplings more than 0.3Hz bigger than random coil values, which is bigger than the range of deviation in intrinsic disordered proteins.

The relations between secondary $^1J_{NH}$ couplings with the individual parameter of the H-bond are investigated. The secondary $^1J_{NH}$ couplings values of ubiquitin and the length of corresponding H-bond were plotted in Figure 12a. Although the overall correlation is poor, the short H-bonds in β-sheet tend to be correlated to lower secondary $^1J_{NH}$

couplings, while longer H-bonds in α-helix are more likely with higher second $^1J_{NH}$ couplings. This tendency is clearer if the Δ^1J_{NH} is plotted against dihedral angle ρ = <H•••O=C-N, which to a large extent distinguishes α-helix from β-sheet (Figure 7.12b).

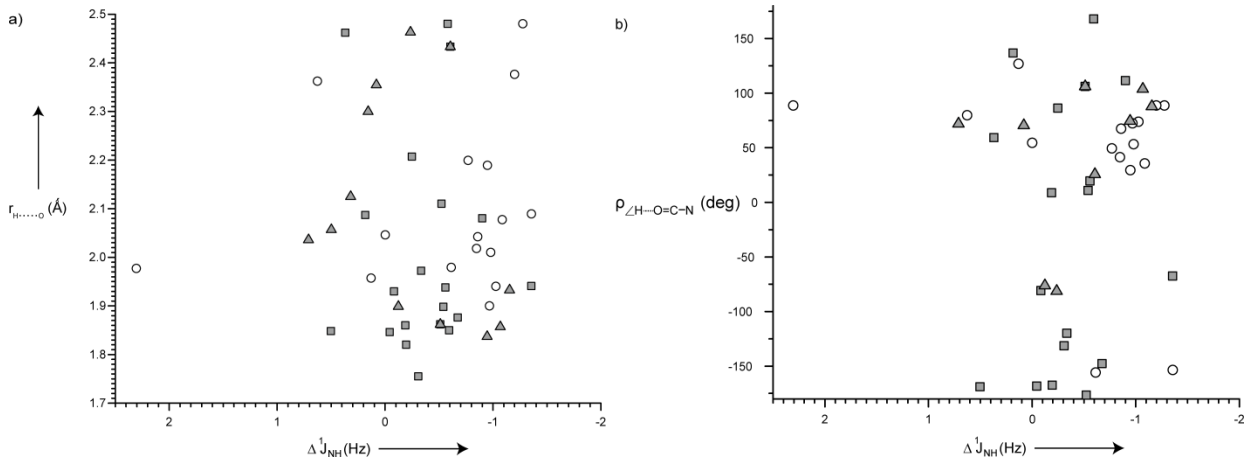


Figure 7.12 a) Comparison of secondary $^1J_{NH}$ values, Δ^1J_{NH}, with H-bond lengths, $r_{H\cdots O}$, in ubiquitin for residues in β-strands (■), loops (▲) and the α-helix (○). b**)** Comparison of secondary $^1J_{NH}$ values,Δ^1J_{NH}, with the dihedral angle ρ= <H•••O=C-N, which to a large extent distinguishes α-helix from β-sheet, in ubiquitin for residues in β-strands (■), loops (▲) and the α-helix (○).

7.6 Identify hydrogen bonds by secondary J couplings

To further validate the relationship between H-bond and $^1J_{NH}$ scalar couplings, GCN peptides were studied. The GCN (16-31) peptide was known to have a helix tendency in its free state. TFE (2, 2, 2-Trifluoroethanol) was used to induce complete helix conformation. The chemical shifts and $^1J_{NH}$ scalar couplings were measured at free state and in the presence of 40% TFE (Figure 7.13). The CD and chemical shifts analysis both revealed that GCN peptide in water only has α-helix tendency, and no complete and stable helixes were formed (Figure 7.14a, black bar and Figure 7.13b). H/D exchange experiments also indicated that there is no stable H-bond formed in free form, in which all amid proton exchanged within 5mins in D2O(Figure 7.14c). The only remaining peak after 10mins in D2O was the proton in hydroxyl group of Tyr17 side chain. In 40% TFE,

however, one α-helix nearly formed based on chemical shifts and CD measurements, in which H bonds should form (Figure 7.13a, light grey bars, Figure 7.14b).

$^1J_{NH}$ scalar couplings of GCN peptide in water were close to random coil values. The values deviated less than 0.35Hz with the exception of R25, which was the upper limit of $^1J_{NH}$ couplings deviation in "H-bond free" residues. When GCN peptide was in 40% TFE solution, the residues in the middle part of the peptide showed a negative shift in $^1J_{NH}$ couplings, while the $^1J_{NH}$ couplings of the residues in the two terminals moved to the positive direction. The decrease of dielectric permittivity in 40% TFE made a positive shift for $^1J_{NH}$ of all residues (257). The negative Δ^1J_{NH} couplings of the residues in the middle of helix were because of the formation of H bonds.

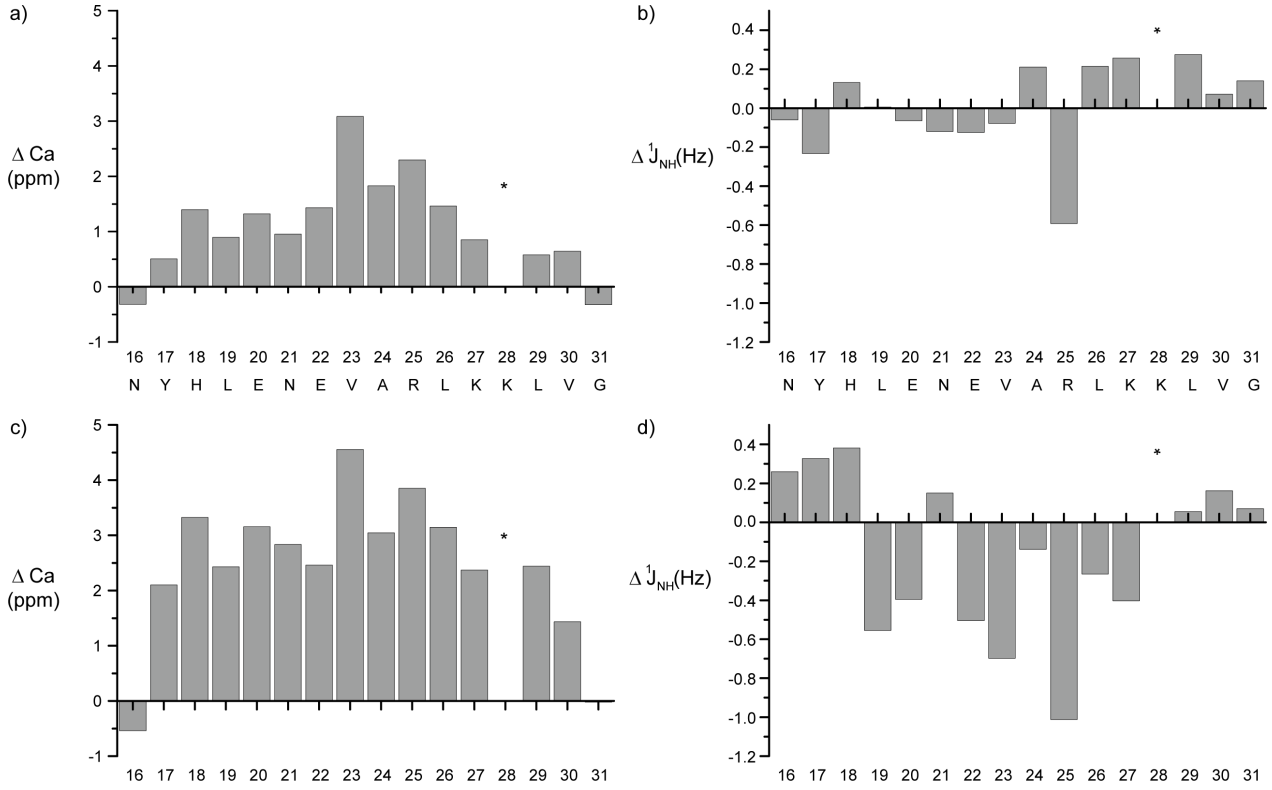


Figure 7.13 a) Secondary Cα chemical shifts of GCN (16-31) peptide in water as a function of residue number. The peptide was dissolved in 50mM Na phosphate buffer, pH was adjusted to 6.8. b) Secondary $^1J_{NH}$ couplings of GCN (16-31) peptide in water as a function of residue number. The random coil values are taken from (187, 188) c) Secondary Cα chemical shifts of GCN (16-31) peptide in 40% TFE as a function of

residue number. d) Secondary $^{1}J_{NH}$ couplings of GCN (16-31) peptide in 40% TFE as a function of residue number. The assignment of L28 is missing, which is marked by star.

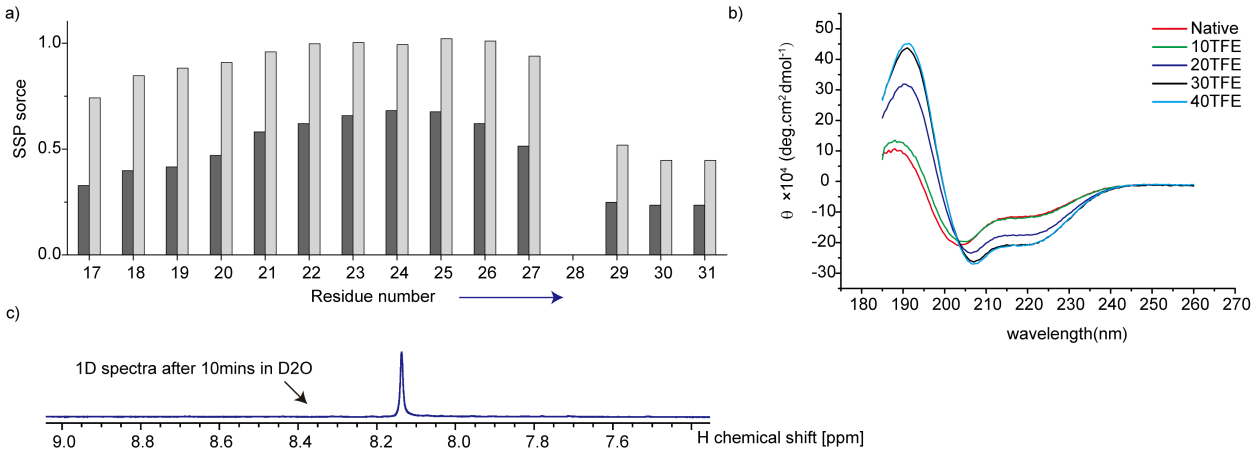


Figure 7.14. a) Secondary structure propensity (SSP) scores of GCN peptide in native state and in 40% TFE. The peptide was dissolved in 50mM Na phosphate buffer, pH was adjusted to 6.8.The SSP scores were predicted by SSP software from Hα,Cα chemical shifts(258). b) CD spectra of GCN16-31 peptide at various TFE concentration, from 0 to 40 %(v/v). c) 1D proton spectrum of GCN16-31 peptide in 100% D_2O after 10 mins. The GCN peptide in pH6.8 buffer was lyophilized and then re-dissolved in 100% D2O. The spectrum was recorded 10 minutes after dissolving.

7.7 Experimental uncertainty

Two similar ubiquitin samples were measured by BSD-IPAP-HSQC at 600MHz. The first one was ^{15}N-single labeled ubiquitin in 50mM HEPES buffer, pH 7.0, 300mM NaCl. The other one was ^{2}D, ^{15}N-labeled in 50mM HEPES buffer, pH 7.4, 300 mM NaCl. The pairwise RMSD between this two data is 0.05Hz, which can be used as the upper limit of experimental uncertainty of BSD-IPAP-HSQC experiments. The correlation plot is shown in Figure 7.15.

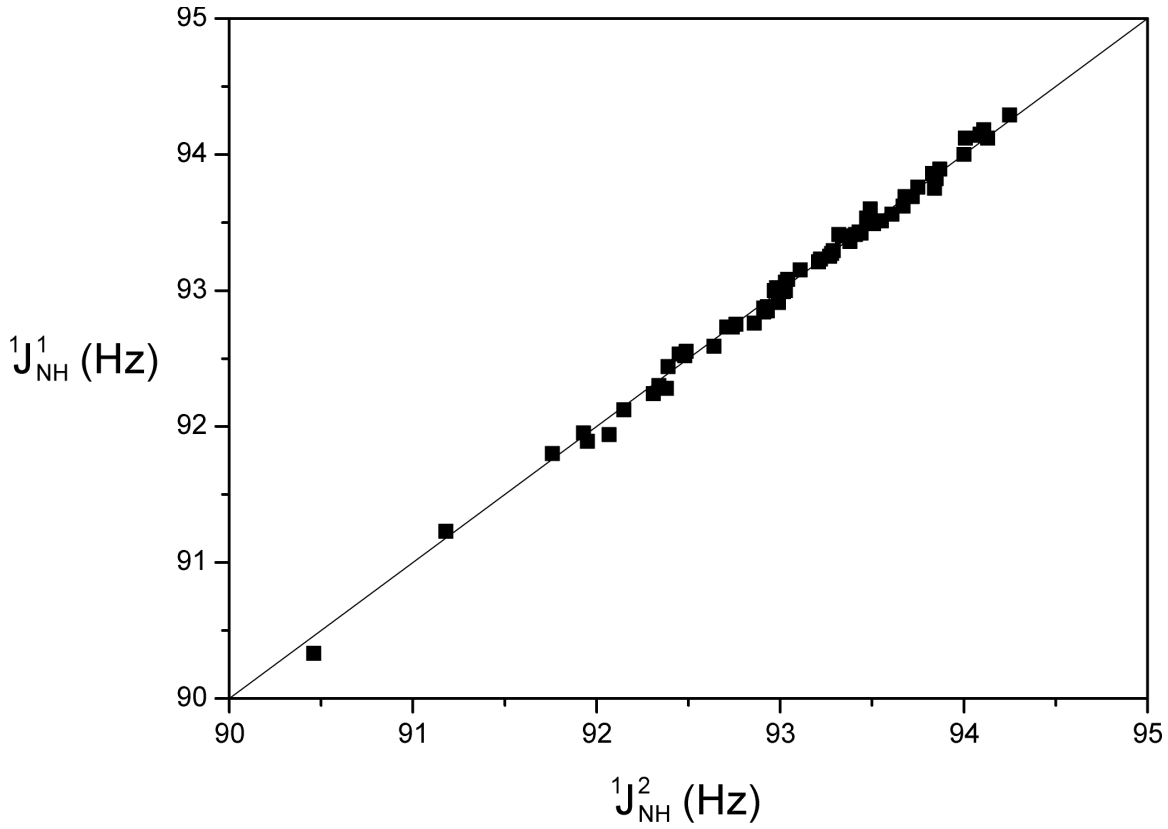


Figure 7.15 Correlation between $^1J_{NH}$ of two ubiquitin samples. The line is y=x line. The x-axis are J couplings of ^2D, ^{15}N-labelled Ubiquitin, y-axis are J couplings of ^{15}N-labelled Ubiquitin. Both couplings were measured by BSD-IPAP HSQC, 298K.

The experimental uncertainty of intensity modulated HSQC was also evaluated by two datasets of Tau K18 in the presence and absence of one small molecule B4A1 which is known to interact with K18. The pairwise RMSD between two datasets is 0.02Hz, which suggests the upper limit of experimental uncertainty. The correlation plot between the two datasets is shown in Figure 7.16.

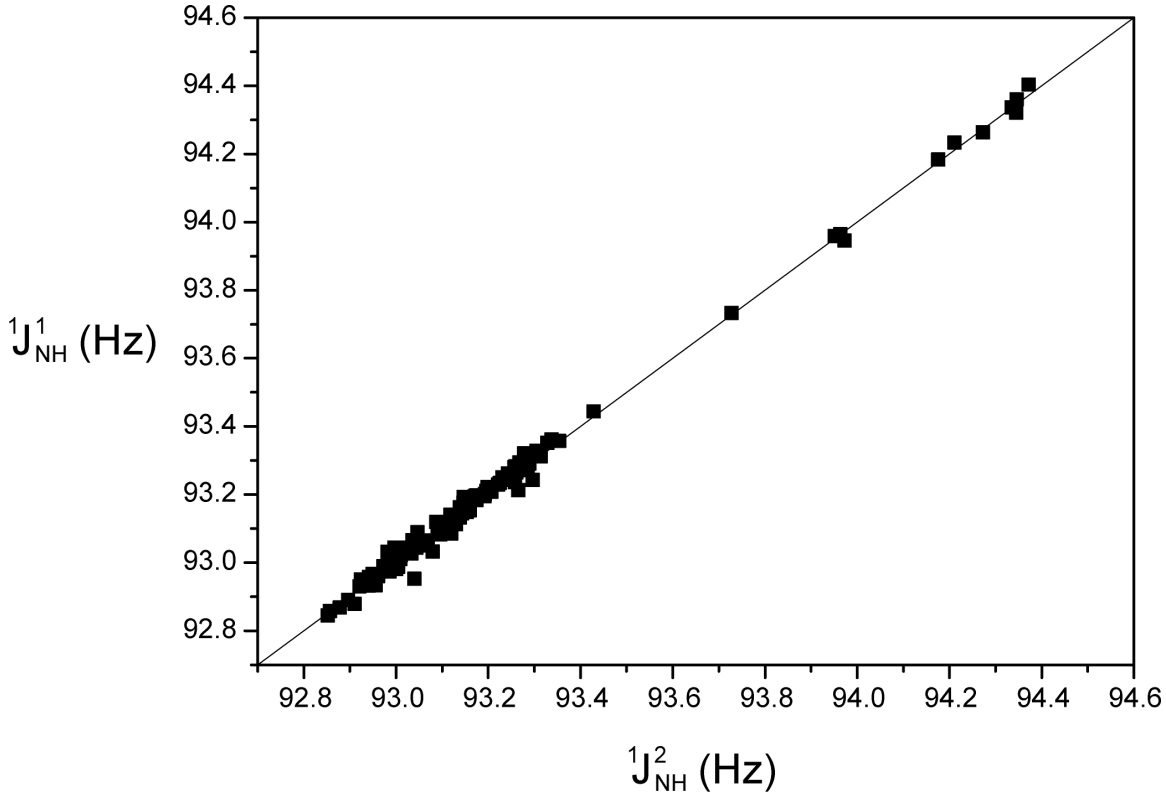


Figure 7.16 Correlation between two datasets of K18 $^1J_{NH}$. Both data set was measured by intensity modulated HSQC. The K18 samples were in 50mM Na-phosphate buffer, pH6.0, 100mM NaCl. The x-axis is K18 in free form while the y-axis is K18 with B4A1 in ratio 1:1.5, 278K.

8. Discussion

8.1 $^1J_{NH}$ couplings of intrinsic disordered proteins

In this study, we propose the mean values of $^1J_{NH}$ couplings as "random coil" values, which are residue type specific. The couplings were measured on the full length Tau protein, which is known to lack stable secondary structures (259, 260). The long sequence (441 residues) of Tau enabled enough reduplications of every residue type within one single sample, which minimized any possible systematic errors between different samples.

The distributions of the 1JNH couplings were all in the range between of -92.77 Hz to -94.47Hz. The values were dependent on the residue type, for instance, the glycine residues showed larger values than other residue types. The statistical parameters, such as standard deviation and range, showed that $^1J_{NH}$ coupling values converged more effectively after sorting by residues types than considering the all J couplings as one whole dataset. The standard deviation of all J couplings together was 0.36Hz and the range was 1.39 Hz. After removing the glycine residues, which were obviously higher than others, the standard deviation was 0.19Hz and the range of data was 0.95Hz. The $^1J_{NH}$ couplings in the same residue type, by contrast, were all within the range, at less than 0.70Hz, and the standard deviations within one residue type were also smaller (Figure 2 and Table 1 in result).

On the other hand, the distributions of residue type specific "random coil" values are very narrow in comparison to other backbone one-bond J couplings. The similar "random coil" values of $^1J_{H\alpha\text{-}C\alpha}$ (without proline and glycine) spans from 141.1 Hz (glutamine) to 143.7Hz (alanine), which covers 2.6Hz (36). The "random coil" values of $^1J_{C\alpha\text{-}C\beta}$ are from 33.6Hz (tyrosine, valine) to 37.4Hz (serine) within the range of 3.8Hz (195). Taking the size of couplings into account, $^1J_{H\alpha\text{-}C\alpha}$ values change ~2% with different residue types while $^1J_{C\alpha\text{-}C\beta}$ changes ~10%. The changes of $^1J_{NH}$ magnitude with residue types are relatively small, at only 0.5% without glycine or 1.4% with glycine. One possible

explanation is that the N-H bond is further away from the side chain, and thus is less affected by different groups on side chains.

The universality of these random coil values were verified by comparing the $^{1}J_{NH}$ couplings of α-synclein to the random coil values which were derived from Tau. Except one residue, the deviation from Tau based random coil values is less than 0.35Hz. The deviations in α-synclein were in a similar range to Tau data. The average values for each residue type were very close to random coil values (maximum difference<0.1Hz, Figure 4 in result), which indicated that there was no systematic difference between the two datasets.

The values do not drift significantly with pH and temperatures as illustrated by the J couplings of α-synclein at various conditions (Figure 6). The pH range in test was from 5.7 to 7.4; the temperature range was 278K to 298K. Although the uncertainties of the experiments were relatively high, the stability of J couplings was revealed by the unchanged mean values, in which the random errors of individual values were averaged out. In the line of this, the biggest difference, which was still less than 0.1Hz, was observed in the residue type with low occurrence.

It is worth noting that apparent J couplings may change with temperature in other cases. The DFS (dynamic frequency shift) is dependent on the dynamic properties of N-H groups, which change with temperature (243, 261). Therefore, the change of DFS results in the different "apparent" J couplings. In the situation of intrinsically disordered α-synclein, the backbone is very flexible. The motions of N-H groups are always fast enough in the temperature range of experiments, which does not change DFS significantly. In the folded protein cases, the motions of the N-H group are mainly dependent on the tumbling time of the whole molecule, which drifts significantly with temperature. The temperature changes of J couplings due to different tumbling time will be not negligible. At 298K, correlation time τ_c of ubiquitin is 4.04ns, while at 278K, the τ_c is ~7.3ns (262). The DFS changes 0.08Hz according to τ_c, assuming the order parameters are the same at the two temperatures.

The residue type-specific random coil values were universal for different proteins and they were stable in the near physiological conditions. Hence, these values were taken as reference for $^1J_{NH}$ couplings.

8.2 The relation between $^1J_{NH}$ coupling and hydrogen bonds

One remarkable characteristic of $^1J_{NH}$ couplings in folded proteins is that the distribution ranges are larger than the ones in intrinsically disordered proteins (IDPs). In intrinsically disorder proteins, such as Tau and α-synuclein, the deviations of $^1J_{NH}$ from random coil values were in the -0.35Hz to +0.35Hz range. In contrast, the deviations in folded proteins were more than 1Hz. The two factors, DFS and residual dipolar couplings, cannot explain this distribution. DFS is one uniform value which is the same for all residues, which can make an offset with maximum to 0.54Hz, but not a larger distribution (243). Residual dipolar coupling due to weak alignment by anisotropic susceptibility is residue-specific, which is dependent on the structure. This effect was measured in ubiquitin (243) and the GB3 protein (244). The sizes of these RDC values were too small to account for the observed range of $^1J_{NH}$ coupling values. Even after the two effects were corrected, the $^1J_{NH}$ coupling values of folded proteins still distributed in a larger range than IDPs'.

The deviations from random coil values were termed as secondary $^1J_{NH}$ coupling, Δ^1J_{NH}. The Δ^1J_{NH} showed significant correlation with the existence of hydrogen bonds. Involving one hydrogen bond increases the magnitude of $^1J_{NH}$ coupling of the amide proton. Apart from the residues near aromatic rings, all residues with secondary $^1J_{NH}$ coupling more negative than -0.35Hz are involved in hydrogen bonds (see Figure 7.9a, Figure 7.10a and Figure 7.11). The connection between negative Δ^1J_{NH} and hydrogen bonds is supported by the data of all three folded proteins under test: ubiquitin, protein G and DINI. This is the first time that the $^1J_{NH}$ couplings have been empirically related to the hydrogen bond, although this relationship has been theoretically predicted before (252, 253, 263).

The correlation between negative Δ^1J_{NH} values and hydrogen bonds was highlighted by the alternative positive and negative secondary J coupling patterns of the β-strand on the

edge of β-sheet, in which every second residue forms a hydrogen bond. In addition, the cross hydrogen bond couplings were not visible in 3_{10}-helix in ubiquitin due to an unfavorable geometry for the hydrogen bonds (264). In contrast, the residues of 3_{10}-helix in ubiquitin showed large negative Δ^1J_{NH} values, which is consistent with the location of hydrogen bonds in NMR and X-ray structures. $^1J_{NH}$ values do not correlate with secondary chemical shifts, demonstrating that $^1J_{NH}$ coupling constants not simply report on the presence of secondary structures, but are intimately linked to the presence of H-bonds.

Here only intramolecular hydrogen bonds were taken into account. The solvent exposure amide proton may also form hydrogen bonds with solvent molecules. However, this kind of hydrogen may also exist in IDPs, in which the amide protons are all accessible to the solvent. Therefore, the difference between the two cannot distinguish the protein-solvent hydrogen bond. In addition, the transitory nature of protein-solvent hydrogen bond makes its effects weaker.

In ubiquitin, Gly10 has a positive secondary $^1J_{NH}$ coupling, but its amide proton was predicted to be in the hydrogen bond based on the X-ray crystal 3D structure of ubiquitin. However, the H/D exchange rate of G10 revealed that it is solvent exposed(237). On the other hand, although Thr22 is not involved in hydrogen bonding based on the crystal structure of ubiquitin (PDB code: 1ubq), it partially hydrogen bonds to its own side chain in the NMR ensemble (PDB code: 1d3z). The H/D exchange rate of Thr22 also proved that it is in hydrogen bonding. The amide group of E34, which is hydrogen bonded to I30, however, had the most positive secondary $^1J_{NH}$ coupling value. The large amide proton chemical shift anisotropy and small deuterium quadrupolar coupling imply the destabilization of this hydrogen bond (235, 236).

There are few amide protons with negative $^1J_{NH}$ couplings that are not involved in the hydrogen bonds, such as Ala46. In addition, $^1J_{NH}$ couplings of the residues in the proximity of aromatic rings are perturbed because of the paramagnetic spin-orbit Ramsey term (265, 266). The residues which are within the 6 Å from the nearby aromatic ring center are labeled. The exceptions of the $^1J_{NH}$ and hydrogen bond correlation were mainly

in these residues. The 6 Å is a rough estimation, since the exact size of this perturbation is still unclear.

Although the correlation between Δ^1J_{NH} values and hydrogen bond is prominent, the values of secondary $^1J_{NH}$ couplings did not show a simple relationship with bond parameters, such as bond length and angles in hydrogen bond. On average, the Δ^1J_{NH} in helix, both α- and 3_{10}- helix, is larger than the Δ^1J_{NH} in β-strands (see Figure 7.9a, Figure 7.10a and Figure 7.12).

Theoretical calculation of $^1J_{NH}$ couplings at the current stage is not accurate enough to enable meaningful comparison with experimental values (252, 263). Our result on the random coil values as well as the relation between Δ^1J_{NH} and hydrogen bonds provide a basis for the theoretical calculation study of $^1J_{NH}$ and hydrogen bonds.

8.3 Comparison with other Methods to identify hydrogen bond

Identifying the hydrogen bond in proteins is a challenging task, since the hydrogen atom is only visible using the best resolution X-ray data. Several NMR observables have been found to relate to hydrogen bonds: H/D exchange, temperature coefficients (234), amide proton chemical shifts anisotropy(CSA) (235), and nuclear quadrupole coupling constants (236). These observables are used to detect hydrogen bonds. In addition, cross hydrogen bond J couplings give direct evidence of hydrogen bond existence. The correlations between Δ^1J_{NH} and hydrogen bonds enable Δ^1J_{NH} to be a powerful tool to identify the hydrogen bond. In comparison to current hydrogen bond detecting methods, Δ^1J_{NH} method has its unique advantages.

Cross hydrogen J couplings are incontrovertible evidence of hydrogen bonds. The trans-hydrogen couplings values give information about the bond parameters (239). The J couplings are measured by long range Trosy-HNCO, which can also provide the chemical shifts of acceptor atoms. In spite of the valuable information offered by cross-hydrogen J couplings, the coupling is difficult to measure due to its small size (<0.5Hz), which require high concentrated sample (>1mM) and $^{15}N/^{13}C$ labeling. Even with help of Trosy and deuteration, this experiment is limited to relatively small proteins (267, 268).

In addition, the cross hydrogen bond couplings were not visible in the 3_{10}-helix in ubiquitin due to an unfavorable geometry for the hydrogen bonds (264).

Temperature coefficient measurements are easy to implement, only needing a series of ^{15}N-^{1}H HSQC. The signal to noise ratio of HSQC is high, and only single labeled samples are needed. However, the change of chemical shifts may be induced by other factors. If the residues are under conformation exchanges, then the populations between different states will be shifted by temperatures (269). This will result in the change of chemical shifts. This contribution can confuse the analysis of temperature coefficients.

The pulse programs and data processing of CSA and quadrupolar coupling constants are complicated. Moreover, the signals levels of these two experiments are low. In the CSA measurements, amide proton magnetization had to be kept on the transverse plane up to 60ms which decrease the signal intensity (235). Measurement of quadrupolar coupling constants need the sample to be 13C/15N double labeled and dissolved in D2O. The pulse program was analogue of CPMG method. However, the magnetization transfers from Hα to Cα, then to Co and finally to N. The long transfer pathway makes the experiment insensitive. The CPMG method employed also makes the results vulnerable to the dynamic properties of the backbone. Without the knowledge of target protein dynamic in detail, as in ubiquitin case, the quadrupolar couplings cannot be reliably calculated (236).

The overwhelming advantage of the $\Delta^{1}J_{NH}$ method is the convenience of implementing experiments. Accurate $^{1}J_{NH}$ values can be measured by intensity-modulated HSQC(243), or BSD-IPAP-HSQC with slightly higher errors (244). BSD-IPAP-HSQC offers an almost identical signal to noise ratio and consumes the same amount of time as routine IPAP-HSQC. Intensity-modulated HSQC normally needs 15 points for fitting, a similar time to R2 experiments. This method only needs a ^{15}N-labled sample and in the H2O, which does not need an extra sample. In fact, the high signals in BSD-IPAP-HSQC enable the $^{1}J_{NH}$ measurement on a natural abundant peptide sample GCN as shown in the result part. Hence this method can apply to many systems in which isotope labeling is not feasible, such as natural products. In this study, the hydrogen bonds in the 3_{10} helix were

identified by Δ^1J_{NH} method. In addition, the hydrogen bond between backbone and side chain, such Thr22, was also found out by this method.

In addition, $^1J_{NH}$ is the NMR parameter, which is a prerequisite for most RDC studies. $^1J_{NH}$ is pre-required to measure backbone N-H RDCs, which are the most commonly used RDC in protein NMR studies.Δ^1J_{NH} values could be a byproduct of RDC studies, which do not require extra time in the schedule of NMR experiments . Higher accurate $^1J_{NH}$ values achieved by BSD-IPAP-HSQC or intensity-modulated HSQC also improve the quality of RDC data.

The main disadvantage of this method is that it is currently unable to extract hydrogen bond parameters such as bond length and angles, from $\Delta^1J_{NH.}$ At this stage, Δ^1J_{NH} is only used to identify the existence of the hydrogen bond. In addition, the relation between Δ^1J_{NH} and hydrogen bonds is affected by aromatic rings. For proteins with large numbers of aromatic residues, this method would be problematic.

8.4 Outlook

The $^1J_{NH}$ values within the same residue type of IDPs are distributed in a range of 0.6Hz. It is possible that other structural factors also contribute to $^1J_{NH}$. Moreover, in folded proteins, some residues show larger positive Δ^1J_{NH} than 0.35Hz, which is the upper limit of Δ^1J_{NH} in IDPs. The reason behind this is unknown. Both experimental and theoretical studies should be conducted to identify the factors causing this.

Unlike CSA and nuclear quadrupolar coupling constants, the size of Δ^1J_{NH} did not show any simple correlation with parameters of hydrogen bonds. The correlation, if it really exists, needs further research to be fully confirmed and developed. Δ^1J_{NH} values of the residues in space proximity of aromatic rings show less correlation with hydrogen bonds, which is due to the interruption from spin-orbit Ramsey term. However, the size and range of this disturbance is not exactly known, which also needs to be addressed by additional research.

In this study, we limited our scope of hydrogen bond types only to N-H···O=C. However, there are other kinds of hydrogen bonds in biomolecules. Their effects on $^1J_{NH}$

also need to be investigated. Cross hydrogen bond couplings are larger in the ones with phosphorus as acceptors (270). It would be advisable to check the Δ^1J_{NH} in these hydrogen bonds. Protons in the C-H group are also likely to be involved in hydrogen bonds (271). It is interesting to see how $^1J_{CH}$ is affected by hydrogen bonds. Hα in proteins were found to form hydrogen bonds. How do the hydrogen bonds affect the $^1J_{C\alpha-H\alpha}$ values, and how does this effect improve the current Karplus equation of $^1J_{C\alpha-H\alpha}$? These questions need to be addressed by further study.

9. Bibliography

1. Hernandez MA, Avila J, & Andreu JM (1986) Physicochemical Characterization of the Heat-Stable Microtubule-Associated Protein Map2. *Eur J Biochem* 154(1):41-48.

2. Cordero OJ, Sarandeses CS, Lopez JL, & Nogueira M (1992) On the Anomalous Behavior on Gel-Filtration and Sds-Electrophoresis of Prothymosin-Alpha. *Biochem Int* 28(6):1117-1124.

3. Tompa P (2002) Intrinsically unstructured proteins. *Trends Biochem Sci* 27(10):527-533.

4. Wilkins DK, *et al.* (1999) Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry-Us* 38(50):16424-16431.

5. Fisher CK & Stultz CM (2011) Constructing ensembles for intrinsically disordered proteins. *Current opinion in structural biology* 21(3):426-431.

6. Dunker AK, *et al.* (2007) The unfoldomics decade: an update on intrinsically disordered proteins. *Bmc Genomics* 9.

7. Uversky VN, Gillespie JR, & Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins-Structure Function and Genetics* 41(3):415-427.

8. Kim TD, Ryu HJ, Cho HI, Yang CH, & Kim J (2000) Thermal behavior of proteins: Heat-resistant proteins and their heat-induced secondary structural changes. *Biochemistry-Us* 39(48):14839-14846.

9. Kriwacki RW, Hengst L, Tennant L, Reed SI, & Wright PE (1996) Structural studies of p21(Waf1/Cip1/Sdi1) in the free and Cdk2-bound state: Conformational disorder mediates binding diversity. *Proceedings of the National Academy of Sciences of the United States of America* 93(21):11504-11509.

10. Schweers O, Schonbrunnhanebeck E, Marx A, & Mandelkow E (1994) Structural Studies of Tau-Protein and Alzheimer Paired Helical Filaments Show No Evidence for Beta-Structure. *Journal of Biological Chemistry* 269(39):24290-24297.

11. Fuxreiter M, Simon I, Friedrich P, & Tompa P (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *Journal of molecular biology* 338(5):1015-1026.

12. Morar AS, Olteanu A, Young GB, & Pielak GJ (2001) Solvent-induced collapse of alpha-synuclein and acid-denatured cytochrome c. *Protein Science* 10(11):2195-2199.

13. Salmon L*, et al.* (2010) NMR Characterization of Long-Range Order in Intrinsically Disordered Proteins. *J Am Chem Soc* 132(24):8407-8418.

14. Mittag T*, et al.* (2010) Structure/Function Implications in a Dynamic Complex of the Intrinsically Disordered Sic1 with the Cdc4 Subunit of an SCF Ubiquitin Ligase. *Structure* 18(4):494-506.

15. Xie HB*, et al.* (2007) Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J Proteome Res* 6(5):1882-1898.

16. Uversky VN, Oldfield CJ, & Dunker AK (2008) Intrinsically disordered proteins in human diseases: Introducing the D(2) concept. *Annu Rev Biophys* 37:215-246.

17. Dunker AK*, et al.* (2001) Intrinsically disordered protein. *J Mol Graph Model* 19(1):26-59.

18. von Ossowski I*, et al.* (2005) Protein disorder: Conformational distribution of the flexible linker in a chimeric double cellulase. *Biophysical journal* 88(4):2823-2832.

19. Mukhopadhyay R & Hoh JH (2001) AFM force measurements on microtubule-associated proteins: the projection domain exerts a long-range repulsive force. *Febs Lett* 505(3):374-378.

20. Trombitas K*, et al.* (1998) Titin extensibility in situ: Entropic elasticity of permanently folded and permanently unfolded molecular segments. *Journal of Cell Biology* 140(4):853-859.

21. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, & Obradovic Z (2002) Intrinsic disorder and protein function. *Biochemistry-Us* 41(21):6573-6582.

22. Dyson HJ & Wright PE (2002) Coupling of folding and binding for unstructured proteins. *Current opinion in structural biology* 12(1):54-60.

23. Demarest SJ*, et al.* (2002) Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* 415(6871):549-553.

24. Fuxreiter M (2012) Fuzziness: linking regulation to protein dynamics. *Mol Biosyst* 8(1):168-177.

25. Gunasekaran K, Tsai CJ, & Nussinov R (2004) Analysis of ordered and disordered protein complexes reveals structural features discriminating between stable and unstable monomers. *Journal of molecular biology* 341(5):1327-1341.

26. Meszaros B, Tompa P, Simon I, & Dosztanyi Z (2007) Molecular principles of the interactions of disordered proteins. *Journal of molecular biology* 372(2):549-561.

27. Csizmok V*, et al.* (2005) Primary contact sites in intrinsically unstructured proteins: the case of calpastatin and microtubule-associated protein 2. *Biochemistry-Us* 44(10):3955-3964.

28. Evans PR & Owen DJ (2002) Endocytosis and vesicle trafficking. *Current opinion in structural biology* 12(6):814-821.

29. Tsai CJ, Ma B, & Nussinov R (2009) Protein-protein interaction networks: how can a hub protein bind so many different partners? *Trends Biochem Sci* 34(12):594-600.

30. Oldfield CJ*, et al.* (2008) Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *Bmc Genomics* 9 Suppl 1:S1.

31. Xu XP & Case DA (2001) Automated prediction of 15N, 13Calpha, 13Cbeta and 13C' chemical shifts in proteins using a density functional database. *Journal of biomolecular NMR* 21(4):321-333.

32. Shen Y & Bax A (2007) Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. *Journal of biomolecular NMR* 38(4):289-302.

33. Neal S, Nip AM, Zhang H, & Wishart DS (2003) Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts. *Journal of biomolecular NMR* 26(3):215-240.

34. Zweckstetter M & 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(15):3791-3792.

35. Vuister GW & Bax A (1993) Quantitative J Correlation - a New Approach for Measuring Homonuclear 3-Bond J(H(N)H(Alpha) Coupling-Constants in N-15-Enriched Proteins. *J Am Chem Soc* 115(17):7772-7777.

36. Vuister GW, Delaglio F, & Bax A (1993) The Use of 1jc-Alpha-H-Alpha Coupling-Constants as a Probe for Protein Backbone Conformation. *Journal of biomolecular NMR* 3(1):67-80.

37. Perez C, Lohr F, Ruterjans H, & Schmidt JM (2001) Self-consistent Karplus parametrization of (3)J couplings depending on the polypeptide side-chain torsion chi(1). *J Am Chem Soc* 123(29):7081-7093.

38. Allison JR, Varnai P, Dobson CM, & Vendruscolo M (2009) Determination of the free energy landscape of alpha-synuclein using spin label nuclear magnetic resonance measurements. *J Am Chem Soc* 131(51):18314-18326.

39. Ganguly D & Chen J (2009) Structural interpretation of paramagnetic relaxation enhancement-derived distances for disordered protein states. *Journal of molecular biology* 390(3):467-477.

40. Fawzi NL, *et al.* (2008) Structure and dynamics of the A ss(21-30) peptide from the interplay of NMR experiments and molecular simulations. *J Am Chem Soc* 130(19):6145-6158.

41. Fisher CK, Huang A, & Stultz CM (2010) Modeling intrinsically disordered proteins with bayesian statistics. *J Am Chem Soc* 132(42):14919-14927.

42. Nodet G, *et al.* (2009) Quantitative Description of Backbone Conformational Sampling of Unfolded Proteins at Amino Acid Resolution from NMR Residual Dipolar Couplings. *J Am Chem Soc* 131(49):17908-17918.

43. Choy WY & Forman-Kay JD (2001) Calculation of ensembles of structures representing the unfolded state of an SH3 domain. *Journal of molecular biology* 308(5):1011-1032.

44. Chen Y, Campbell SL, & Dokholyan NV (2007) Deciphering protein dynamics from NMR data using explicit structure sampling and selection. *Biophysical journal* 93(7):2300-2306.

45. Salmon L, *et al.* (2010) NMR characterization of long-range order in intrinsically disordered proteins. *J Am Chem Soc* 132(24):8407-8418.

46. Yoon MK, *et al.* (2009) Residual structure within the disordered C-terminal segment of p21(Waf1/Cip1/Sdi1) and its implications for molecular recognition. *Protein Science* 18(2):337-347.

47. Huang A & Stultz CM (2008) The effect of a DeltaK280 mutation on the unfolded state of a microtubule-binding repeat in Tau. *PLoS computational biology* 4(8):e1000155.

48. Marsh JA & Forman-Kay JD (2009) Structure and Disorder in an Unfolded State under Nondenaturing Conditions from Ensemble Models Consistent with a Large Number of Experimental Restraints. *Journal of molecular biology* 391(2):359-374.

49. Ostman A, Hellberg C, & Bohmer FD (2006) Protein-tyrosine phosphatases and cancer. *Nature reviews. Cancer* 6(4):307-320.

50. Rao RS & Moller IM (2012) Large-scale analysis of phosphorylation site occupancy in eukaryotic proteins. *Biochimica et biophysica acta* 1824(3):405-412.

51. Iakoucheva LM, *et al.* (2004) The importance of intrinsic disorder for protein phosphorylation. *Nucleic acids research* 32(3):1037-1049.

52. Narayanan A & Jacobson MP (2009) Computational studies of protein regulation by post-translational phosphorylation. *Current opinion in structural biology* 19(2):156-163.

53. Becker S, Groner B, & Muller CW (1998) Three-dimensional structure of the Stat3beta homodimer bound to DNA. *Nature* 394(6689):145-151.

54. Groban ES, Narayanan A, & Jacobson MP (2006) Conformational changes in protein loops and helices induced by post-translational phosphorylation. *PLoS computational biology* 2(4):e32.

55. Russo AA, Jeffrey PD, & Pavletich NP (1996) Structural basis of cyclin-dependent kinase activation by phosphorylation. *Nat Struct Biol* 3(8):696-700.

56. Banavali NK & Roux B (2007) Anatomy of a structural pathway for activation of the catalytic domain of Src kinase Hck. *Proteins* 67(4):1096-1112.

57. Bartova I, Otyepka M, Kriz Z, & Koca J (2004) Activation and inhibition of cyclin-dependent kinase-2 by phosphorylation; a molecular dynamics study reveals the functional importance of the glycine-rich loop. *Protein science : a publication of the Protein Society* 13(6):1449-1457.

58. Cheng Y, Zhang Y, & McCammon JA (2006) How does activation loop phosphorylation modulate catalytic activity in the cAMP-dependent protein kinase: a theoretical study. *Protein science : a publication of the Protein Society* 15(4):672-683.

59. Metcalfe EE, Traaseth NJ, & Veglia G (2005) Serine 16 phosphorylation induces an order-to-disorder transition in monomeric phospholamban. *Biochemistry-Us* 44(11):4386-4396.

60. Espinoza-Fonseca LM, Kast D, & Thomas DD (2007) Molecular dynamics simulations reveal a disorder-to-order transition on phosphorylation of smooth muscle myosin. *Biophysical journal* 93(6):2083-2090.

61. Nelson WD, Blakely SE, Nesmelov YE, & Thomas DD (2005) Site-directed spin labeling reveals a conformational switch in the phosphorylation domain of smooth muscle myosin. *Proceedings of the National Academy of Sciences of the United States of America* 102(11):4000-4005.

62. Solt I, Magyar C, Simon I, Tompa P, & Fuxreiter M (2006) Phosphorylation-induced transient intrinsic structure in the kinase-inducible domain of CREB facilitates its recognition by the KIX domain of CBP. *Proteins* 64(3):749-757.

63. Sprang SR*, et al.* (1988) Structural-Changes in Glycogen-Phosphorylase Induced by Phosphorylation. *Nature* 336(6196):215-221.

64. Roth MB, Zahler AM, & Stolk JA (1991) A conserved family of nuclear phosphoproteins localized to sites of polymerase II transcription. *The Journal of cell biology* 115(3):587-596.

65. Ge H, Zuo P, & Manley JL (1991) Primary structure of the human splicing factor ASF reveals similarities with Drosophila regulators. *Cell* 66(2):373-382.

66. Fu XD & Maniatis T (1990) Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature* 343(6257):437-441.

67. Graveley BR (2000) Sorting out the complexity of SR protein functions. *RNA* 6(9):1197-1211.

68. Long JC & Caceres JF (2009) The SR protein family of splicing factors: master regulators of gene expression. *The Biochemical journal* 417(1):15-27.

69. Shepard PJ & Hertel KJ (2009) The SR protein family. *Genome biology* 10(10):242.

70. Sanford JR, Ellis J, & Caceres JF (2005) Multiple roles of arginine/serine-rich splicing factors in RNA processing. *Biochemical Society transactions* 33(Pt 3):443-446.

71. Shen H, Kan JL, & Green MR (2004) Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. *Molecular cell* 13(3):367-376.

72. Shen H & Green MR (2004) A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly. *Molecular cell* 16(3):363-373.

73. Shen H & Green MR (2007) RS domain-splicing signal interactions in splicing of U12-type and U2-type introns. *Nature structural & molecular biology* 14(7):597-603.

74. Zahler AM, Lane WS, Stolk JA, & Roth MB (1992) SR proteins: a conserved family of pre-mRNA splicing factors. *Genes & development* 6(5):837-847.

75. Kress TL, Krogan NJ, & Guthrie C (2008) A single SR-like protein, Npl3, promotes pre-mRNA splicing in budding yeast. *Molecular cell* 32(5):727-734.

76. Zhong XY, Wang P, Han J, Rosenfeld MG, & Fu XD (2009) SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. *Molecular cell* 35(1):1-10.

77. Fededa JP & Kornblihtt AR (2008) A splicing regulator promotes transcriptional elongation. *Nature structural & molecular biology* 15(8):779-781.

78. Caceres JF, Screaton GR, & Krainer AR (1998) A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes & development* 12(1):55-66.

79. Huang Y, Gattoni R, Stevenin J, & Steitz JA (2003) SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Molecular cell* 11(3):837-843.

80. Huang Y & Steitz JA (2001) Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA. *Molecular cell* 7(4):899-905.

81. Zhang Z & Krainer AR (2004) Involvement of SR proteins in mRNA surveillance. *Molecular cell* 16(4):597-607.

82. Michlewski G, Sanford JR, & Caceres JF (2008) The splicing factor SF2/ASF regulates translation initiation by enhancing phosphorylation of 4E-BP1. *Molecular cell* 30(2):179-189.

83. Li X & Manley JL (2005) Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* 122(3):365-378.

84. Lin S Fau - Fu X-D & Fu XD (SR proteins and related factors in alternative splicing. (0065-2598 (Print)).

85. Kohtz JD*, et al.* (1994) Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature* 368(6467):119-124.

86. Zahler AM & Roth MB (1995) Distinct functions of SR proteins in recruitment of U1 small nuclear ribonucleoprotein to alternative 5' splice sites. *Proceedings of the National Academy of Sciences of the United States of America* 92(7):2642-2646.

87. Zuo P & Manley JL (1994) The human splicing factor ASF/SF2 can specifically recognize pre-mRNA 5' splice sites. *Proceedings of the National Academy of Sciences of the United States of America* 91(8):3363-3367.

88. Wang Z, Hoffmann HM, & Grabowski PJ (1995) Intrinsic U2AF binding is modulated by exon enhancer signals in parallel with changes in splicing activity. *RNA* 1(1):21-35.

89. Li Y & Blencowe BJ (1999) Distinct factor requirements for exonic splicing enhancer function and binding of U2AF to the polypyrimidine tract. *The Journal of biological chemistry* 274(49):35074-35079.

90. Zuo P & Maniatis T (1996) The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. *Genes & development* 10(11):1356-1368.

91. Zhu J, Mayeda A, & Krainer AR (2001) Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Molecular cell* 8(6):1351-1361.

92. Ibrahim EC, Schaal TD, Hertel KJ, Reed R, & Maniatis T (2005) Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers. *Proceedings of the National Academy of Sciences of the United States of America* 102(14):5002-5007.

93. Wu JY & Maniatis T (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* 75(6):1061-1070.

94. Goren A*, et al.* (2006) Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers. *Molecular cell* 22(6):769-781.

95. Kanopka A, Muhlemann O, & Akusjarvi G (1996) Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature* 381(6582):535-538.

96. Cowper AE, Caceres JF, Mayeda A, & Screaton GR (2001) Serine-arginine (SR) protein-like factors that antagonize authentic SR proteins and regulate alternative splicing. *The Journal of biological chemistry* 276(52):48908-48914.

97. Shin C & Manley JL (2002) The SR protein SRp38 represses splicing in M phase cells. *Cell* 111(3):407-417.

98. Wu JY, Kar A, Kuo D, Yu B, & Havlioglu N (2006) SRp54 (SFRS11), a regulator for tau exon 10 alternative splicing identified by an expression cloning strategy. *Molecular and cellular biology* 26(18):6739-6747.

99. Barnard DC, Li J, Peng R, & Patton JG (2002) Regulation of alternative splicing by SRrp86 through coactivation and repression of specific SR proteins. *RNA* 8(4):526-533.

100. Li J*, et al.* (2003) Regulation of alternative splicing by SRrp86 and its interacting proteins. *Molecular and cellular biology* 23(21):7437-7447.

101. Hastings ML & Krainer AR (2001) Functions of SR proteins in the U12-dependent AT-AC pre-mRNA splicing pathway. *RNA* 7(3):471-482.

102. Roscigno RF & Garcia-Blanco MA (1995) SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. *RNA* 1(7):692-706.

103. Makarova OV, Makarov EM, & Luhrmann R (2001) The 65 and 110 kDa SR-related proteins of the U4/U6.U5 tri-snRNP are essential for the assembly of mature spliceosomes. *The EMBO journal* 20(10):2553-2563.

104. Mathew R*, et al.* (2008) Phosphorylation of human PRP28 by SRPK2 is required for integration of the U4/U6-U5 tri-snRNP into the spliceosome. *Nature structural & molecular biology* 15(5):435-443.

105. Longman D, Johnstone IL, & Caceres JF (2000) Functional characterization of SR and SR-related genes in Caenorhabditis elegans. *The EMBO journal* 19(7):1625-1637.

106. Kawano T, Fujita M, & Sakamoto H (2000) Unique and redundant functions of SR proteins, a conserved family of splicing factors, in Caenorhabditis elegans development. *Mechanisms of development* 95(1-2):67-76.

107. Tacke R & Manley JL (1995) The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *The EMBO journal* 14(14):3540-3551.

108. Tacke R, Tohyama M, Ogawa S, & Manley JL (1998) Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell* 93(1):139-148.

109. Schaal TD & Maniatis T (1999) Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. *Molecular and cellular biology* 19(1):261-273.

110. Lynch KW & Maniatis T (1995) Synergistic interactions between two distinct elements of a regulated splicing enhancer. *Genes & development* 9(3):284-293.

111. Hertel KJ & Maniatis T (1998) The function of multisite splicing enhancers. *Molecular cell* 1(3):449-455.

112. Shin C, Kleiman FE, & Manley JL (2005) Multiple properties of the splicing repressor SRp38 distinguish it from typical SR proteins. *Molecular and cellular biology* 25(18):8334-8343.

113. Wang J, Takagaki Y, & Manley JL (1996) Targeted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability. *Genes & development* 10(20):2588-2599.

114. Jumaa H, Wei G, & Nielsen PJ (1999) Blastocyst formation is blocked in mouse embryos lacking the splicing factor SRp20. *Current biology : CB* 9(16):899-902.

115. Wang HY, Xu X, Ding JH, Bermingham JR, Jr., & Fu XD (2001) SC35 plays a role in T cell development and alternative splicing of CD45. *Molecular cell* 7(2):331-342.

116. Xu X*, et al.* (2005) ASF/SF2-regulated CaMKIIdelta alternative splicing temporally reprograms excitation-contraction coupling in cardiac muscle. *Cell* 120(1):59-72.

117. Ding JH*, et al.* (2004) Dilated cardiomyopathy caused by tissue-specific ablation of SC35 in the heart. *The EMBO journal* 23(4):885-896.

118. Fischer DC*, et al.* (2004) Expression of splicing factors in human ovarian cancer. *Oncology reports* 11(5):1085-1090.

119. Stickeler E, Kittrell F, Medina D, & Berget SM (1999) Stage-specific changes in SR splicing factors and alternative splicing in mammary tumorigenesis. *Oncogene* 18(24):3574-3582.

120. Ghigna C, Moroni M, Porta C, Riva S, & Biamonti G (1998) Altered expression of heterogenous nuclear ribonucleoproteins and SR factors in human colon adenocarcinomas. *Cancer research* 58(24):5818-5824.

121. Karni R*, et al.* (2007) The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nature structural & molecular biology* 14(3):185-193.

122. Ghigna C*, et al.* (2005) Cell motility is controlled by SF2/ASF through alternative splicing of the Ron protooncogene. *Molecular cell* 20(6):881-890.

123. Stoltzfus CM & Madsen JM (2006) Role of viral splicing elements and cellular RNA binding proteins in regulation of HIV-1 alternative RNA splicing. *Current HIV research* 4(1):43-55.

124. Asang C, Hauber I, & Schaal H (2008) Insights into the selective activation of alternatively used splice acceptors by the human immunodeficiency virus type-1 bidirectional splicing enhancer. *Nucleic acids research* 36(5):1450-1463.

125. Exline CM, Feng Z, & Stoltzfus CM (2008) Negative and positive mRNA splicing elements act competitively to regulate human immunodeficiency virus type 1 vif gene expression. *Journal of virology* 82(8):3921-3931.

126. Dowling D*, et al.* (2008) HIV-1 infection induces changes in expression of cellular splicing factors that regulate alternative viral splicing and virus production in macrophages. *Retrovirology* 5:18.

127. Barbaro G, Scozzafava A, Mastrolorenzo A, & Supuran CT (2005) Highly active antiretroviral therapy: current state of the art, new agents and their pharmacological interactions useful for improving therapeutic outcome. *Current pharmaceutical design* 11(14):1805-1843.

128. Soret J*, et al.* (2005) Selective modification of alternative splicing by indole derivatives that target serine-arginine-rich protein splicing factors. *Proceedings of the National Academy of Sciences of the United States of America* 102(24):8764-8769.

129. Bakkour N*, et al.* (2007) Small-molecule inhibition of HIV pre-mRNA splicing as a novel antiretroviral therapy to overcome drug resistance. *PLoS pathogens* 3(10):1530-1539.

130. Pagani F*, et al.* (2000) Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a nonevolutionary conserved intronic element. *The Journal of biological chemistry* 275(28):21041-21047.

131. Buratti E, Stuani C, De Prato G, & Baralle FE (2007) SR protein-mediated inhibition of CFTR exon 9 inclusion: molecular characterization of the intronic splicing silencer. *Nucleic acids research* 35(13):4359-4368.

132. Mine M*, et al.* (2003) Splicing error in E1alpha pyruvate dehydrogenase mRNA caused by novel intronic mutation responsible for lactic acidosis and mental retardation. *The Journal of biological chemistry* 278(14):11768-11772.

133. Neugebauer KM, Merrill JT, Wener MH, Lahita RG, & Roth MB (2000) SR proteins are autoantigens in patients with systemic lupus erythematosus. Importance of phosphoepitopes. *Arthritis and rheumatism* 43(8):1768-1778.

134. Teigelkamp S, Mundt C, Achsel T, Will CL, & Luhrmann R (1997) The human U5 snRNP-specific 100-kD protein is an RS domain-containing, putative RNA

helicase with significant homology to the yeast splicing factor Prp28p. *RNA* 3(11):1313-1326.

135. Theissen H*, et al.* (1986) Cloning of the human cDNA for the U1 RNA-associated 70K protein. *Embo J* 5(12):3209-3217.

136. Labourier E*, et al.* (1999) Antagonism between RSF1 and SR proteins for both splice-site recognition in vitro and Drosophila development. *Genes & development* 13(6):740-753.

137. Gui JF, Lane WS, & Fu XD (1994) A serine kinase regulates intracellular localization of splicing factors in the cell cycle. *Nature* 369(6482):678-682.

138. Kanopka A*, et al.* (1998) Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins. *Nature* 393(6681):185-187.

139. Prasad J, Colwill K, Pawson T, & Manley JL (1999) The protein kinase Clk/Sty directly modulates SR protein activity: both hyper- and hypophosphorylation inhibit splicing. *Molecular and cellular biology* 19(10):6991-7000.

140. Sanford JR & Bruzik JP (1999) Developmental regulation of SR protein phosphorylation and activity. *Genes & development* 13(12):1513-1518.

141. Xiao SH & Manley JL (1997) Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev* 11(3):334-344.

142. Xiao S-H & Manley JL (1998) Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2. *The EMBO journal* 17(21):6359-6367.

143. Caceres JF, Misteli T, Screaton GR, Spector DL, & Krainer AR (1997) Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *The Journal of cell biology* 138(2):225-238.

144. Lai MC, Lin RI, Huang SY, Tsai CW, & Tarn WY (2000) A human importin-beta family protein, transportin-SR2, interacts with the phosphorylated RS domain of SR proteins. *The Journal of biological chemistry* 275(11):7950-7957.

145. Kataoka N, Bachorik JL, & Dreyfuss G (1999) Transportin-SR, a nuclear import receptor for SR proteins. *The Journal of cell biology* 145(6):1145-1152.

146. Misteli T*, et al.* (1998) Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. *The Journal of cell biology* 143(2):297-307.

147. Manley JL & Tacke R (1996) SR proteins and splicing control. *Genes Dev* 10(13):1569-1579.

148. Kuroyanagi N, Onogi H, Wakabayashi T, & Hagiwara M (1998) Novel SR-protein-specific kinase, SRPK2, disassembles nuclear speckles. *Biochem Biophys Res Commun* 242(2):357-364.

149. Nakagawa O*, et al.* (2005) Centronuclear myopathy in mice lacking a novel muscle-specific protein kinase transcriptionally regulated by MEF2. *Genes Dev* 19(17):2066-2077.

150. Blaustein M*, et al.* (2005) Concerted regulation of nuclear and cytoplasmic activities of SR proteins by AKT. *Nature structural & molecular biology* 12(12):1037-1044.

151. Patel NA*, et al.* (2005) Molecular and genetic studies imply Akt-mediated signaling promotes protein kinase CbetaII alternative splicing via phosphorylation of serine/arginine-rich splicing factor SRp40. *The Journal of biological chemistry* 280(14):14302-14309.

152. Ngo JC*, et al.* (2005) Interplay between SRPK and Clk/Sty kinases in phosphorylation of the splicing factor ASF/SF2 is regulated by a docking motif in ASF/SF2. *Molecular cell* 20(1):77-89.

153. Bullock AN*, et al.* (2009) Kinase domain insertions define distinct roles of CLK kinases in SR protein phosphorylation. *Structure* 17(3):352-362.

154. Prasad J & Manley JL (2003) Regulation and substrate specificity of the SR protein kinase Clk/Sty. *Molecular and cellular biology* 23(12):4139-4149.

155. Hagopian JC*, et al.* (2008) Adaptable molecular interactions guide phosphorylation of the SR protein ASF/SF2 by SRPK1. *Journal of molecular biology* 382(4):894-909.

156. Velazquez-Dones A*, et al.* (2005) Mass spectrometric and kinetic analysis of ASF/SF2 phosphorylation by SRPK1 and Clk/Sty. *The Journal of biological chemistry* 280(50):41761-41768.

157. Ma CT*, et al.* (2008) Ordered multi-site phosphorylation of the splicing factor ASF/SF2 by SRPK1. *Journal of molecular biology* 376(1):55-68.

158. Ma C-T, Hagopian JC, Ghosh G, Fu X-D, & Adams JA (2009) Regiospecific Phosphorylation Control of the SR Protein ASF/SF2 by SRPK1. *Journal of molecular biology* 390(4):618-634.

159. Ngo JC*, et al.* (2008) A sliding docking interaction is essential for sequential and processive phosphorylation of an SR protein by SRPK1. *Molecular cell* 29(5):563-576.

160. Ghosh G & Adams JA (2011) Phosphorylation mechanism and structure of serine-arginine protein kinases. *FEBS Journal* 278(4):587-597.

161. Labourier E*, et al.* (1998) Interaction between the N-terminal domain of human DNA topoisomerase I and the arginine-serine domain of its substrate determines phosphorylation of SF2/ASF splicing factor. *Nucleic acids research* 26(12):2955-2962.

162. Hamelberg D, Shen T, & McCammon JA (2007) A proposed signaling motif for nuclear import in mRNA processing via the formation of arginine claw. *Proceedings of the National Academy of Sciences of the United States of America* 104(38):14947-14951.

163. Sellis D*, et al.* (2012) Phosphorylation of the arginine/serine repeats of lamin B receptor by SRPK1-insights from molecular dynamics simulations. *Biochimica et biophysica acta* 1820(1):44-55.

164. Delaglio F*, et al.* (1995) Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. *Journal of biomolecular NMR* 6(3):277-293.

165. Jung YS & Zweckstetter M (2004) Mars - robust automatic backbone assignment of proteins. *Journal of biomolecular NMR* 30(1):11-23.

166. Mori S, Abeygunawardana C, Johnson MO, & Vanzijl PCM (1995) Improved Sensitivity of HSQC Spectra of Exchanging Protons at Short Interscan Delays Using a New Fast HSQC (FHSQC) Detection Scheme That Avoids Water Saturation. *Journal of Magnetic Resonance, Series B* 108(1):94-98.

167. Sattler M, Schleucher J, & Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Mag Res Sp* 34(2):93-158.

168. Zweckstetter M & Bax A (2001) Single-Step Determination of Protein Substructures Using Dipolar Couplings: Aid to Structural Genomics. *J Am Chem Soc* 123(38):9490-9491.

169. Panchal SC, Bhavesh NS, & Hosur RV (2001) Improved 3D triple resonance experiments, HNN and HN(C)N, for H-N and N-15 sequential correlations in (C-13, N-15) labeled proteins: Application to unfolded proteins. *Journal of biomolecular NMR* 20(2):135-147.

170. Hiller S, Fiorito F, Wuthrich K, & Wider G (2005) Automated projection spectroscopy (APSY). *Proceedings of the National Academy of Sciences of the United States of America* 102(31):10876-10881.

171. Hiller S, Wasmer C, Wider G, & Wuthrich K (2007) Sequence-specific resonance assignment of soluble nonglobular proteins by 7D APSY-NMR spectroscopy. *J Am Chem Soc* 129(35):10823-10828.

172. Güntert P, Dötsch V, Wider G, & Wüthrich K (1992) Processing of multi-dimensional NMR data with the new software PROSA. *Journal of biomolecular NMR* 2(6):619-629.

173. Narayanan RL*, et al.* (2010) Automatic Assignment of the Intrinsically Disordered Protein Tau with 441-Residues. *J Am Chem Soc* 132(34):11906-11907.

174. Hiller S, Wasmer C, Wider G, & Muthrich K (2007) Sequence-specific resonance assignment of soluble nonglobular proteins by 7D APSY-NMR Spectroscopy. *J Am Chem Soc* 129(35):10823-10828.

175. Fiorito F, Hiller S, Wider G, & Wuthrich K (2006) Automated resonance assignment of proteins: 6D APSY-NMR. *Journal of biomolecular NMR* 35(1):27-37.

176. Permi P, Kilpelainen I, Annila A, & Heikkinen S (2000) Intensity modulated HSQC and HMQC: Two simple methods to measure (3)J(HNH alpha) in proteins. *Journal of biomolecular NMR* 16(1):29-37.

177. Hu JS & Bax A (1997) Chi 1 angle information from a simple two-dimensional NMR experiment that identifies trans 3JNC gamma couplings in isotopically enriched proteins. *Journal of biomolecular NMR* 9(3):323-328.

178. Hu JS, Grzesiek S, & Bax A (1997) Two-dimensional NMR methods for determining (chi 1) angles of aromatic residues in proteins from three-bond J(C'C gamma) and J(NC gamma) couplings. *J Am Chem Soc* 119(7):1803-1804.

179. Ruckert M & Otting G (2000) Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J Am Chem Soc* 122(32):7793-7797.

180. Kontaxis G, Clore GM, & Bax A (2000) Evaluation of cross-correlation effects and measurement of one-bond couplings in proteins with short transverse relaxation times. *J Magn Reson* 143(1):184-196.

181. Permi P, Rosevear PR, & Annila A (2000) A set of HNCO-based experiments for measurement of residual dipolar couplings in N-15, C-13, (H-2)-labeled proteins. *Journal of biomolecular NMR* 17(1):43-54.

182. Jaravine V, Ibraghimov I, & Orekhov VY (2006) Removal of a time barrier for high-resolution multidimensional NMR spectroscopy. *Nature methods* 3(8):605-607.

183. Bertoncini CW*, et al.* (2007) Structural Characterization of the Intrinsically Unfolded Protein β-Synuclein, a Natural Negative Regulator of α-Synuclein Aggregation. *Journal of molecular biology* 372(3):708-722.

184. Rezaei-Ghaleh N, Giller K, Becker S, & Zweckstetter M (2011) Effect of Zinc Binding on β-Amyloid Structure and Dynamics: Implications for Aβ Aggregation. *Biophysical journal* 101(5):1202-1211.

185. Smith LJ*, et al.* (1996) Analysis of main chain torsion angles in proteins: Prediction of NMR coupling constants for native and random coil conformations. *Journal of molecular biology* 255(3):494-506.

186. Cazalla D*, et al.* (2002) Nuclear export and retention signals in the RS domain of SR proteins. *Molecular and cellular biology* 22(19):6871-6882.

187. Kjaergaard M, Brander S, & Poulsen FM (2011) Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. *Journal of biomolecular NMR* 49(2):139-149.

188. Kjaergaard M & Poulsen FM (2011) Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. *Journal of biomolecular NMR* 50(2):157-165.

189. Best RB & Hummer G (2009) Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J Phys Chem B* 113(26):9004-9015.

190. Hornak V*, et al.* (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65(3):712-725.

191. Zweckstetter M & 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(15):3791-3792.

192. Showalter SA & Bruschweiler R (2007) Quantitative molecular ensemble interpretation of NMR dipolar couplings without restraints. *J Am Chem Soc* 129(14):4158-+.

193. Lange OF*, et al.* (2008) Recognition Dynamics Up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution. *Science* 320(5882):1471-1475.

194. Bernadó P*, et al.* (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proceedings of the National Academy of Sciences of the United States of America* 102(47):17002-17007.

195. Cornilescu G, Bax A, & Case DA (2000) Large variations in one-bond (C alpha-13C beta)-C-13 J couplings in polypeptides correlate with backbone conformation. *J Am Chem Soc* 122(10):2168-2171.

196. Bienkiewicz EA & Lumb KJ (1999) Random-coil chemical shifts of phosphorylated amino acids. *Journal of biomolecular NMR* 15(3):203-206.

197. Shen Y & Bax A (2010) SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *Journal of biomolecular NMR* 48(1):13-22.

198. Han B, Liu Y, Ginzinger SW, & Wishart DS (2011) SHIFTX2: significantly improved protein chemical shift prediction. *Journal of biomolecular NMR* 50(1):43-57.

199. Jensen MR, Salmon L, Nodet G, & Blackledge M (2010) Defining Conformational Ensembles of Intrinsically Disordered and Partially Folded Proteins Directly from Chemical Shifts. *J Am Chem Soc* 132(4):1270-1272.

200. Krivobokova T, Briones R, Hub JS, Munk A, & de Groot BL (2012) Partial Least-Squares Functional Mode Analysis: Application to the Membrane Proteins AQP1, Aqy1, and CLC-ec1. *Biophysical journal* 103(4):786-796.

201. Hub JS & de Groot BL (2009) Detection of functional modes in protein dynamics. *PLoS computational biology* 5(8):e1000480.

202. Ngo JCK*, et al.* (2008) A sliding docking interaction is essential for sequential and processive phosphorylation of an SR protein by SRPK1. *Molecular cell* 29(5):563-576.

203. Hennig M*, et al.* (1999) Side-chain conformations in an unfolded protein: chi1 distributions in denatured hen lysozyme determined by heteronuclear 13C, 15N NMR spectroscopy. *Journal of molecular biology* 288(4):705-723.

204. Hernandez N & Keller W (1983) Splicing of in vitro synthesized messenger RNA precursors in HeLa cell extracts. *Cell* 35(1):89-99.

205. Bessonov S, Anokhina M, Will CL, Urlaub H, & Luhrmann R (2008) Isolation of an active step I spliceosome and composition of its RNP core. *Nature* 452(7189):846-850.

206. Paleologou KE*, et al.* (2008) Phosphorylation at Ser-129 but Not the Phosphomimics S129E/D Inhibits the Fibrillation of α-Synuclein. *Journal of Biological Chemistry* 283(24):16895-16905.

207. Fuxreiter M, Simon I, Friedrich P, & Tompa P (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *Journal of molecular biology* 338(5):1015-1026.

208. Lee S-H*, et al.* (2012) Understanding Pre-Structured Motifs (PreSMos) in Intrinsically Unfolded Proteins. *Current Protein and Peptide Science* 13(1):34-54.

209. Puntervoll P*, et al.* (2003) ELM server: A new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic acids research* 31(13):3625-3630.

210. Chandler SD, Mayeda A, Yeakley JM, Krainer AR, & Fu XD (1997) RNA splicing specificity determined by the coordinated action of RNA recognition motifs in SR proteins. *Proceedings of the National Academy of Sciences of the United States of America* 94(8):3596-3601.

211. Graveley BR, Hertel KJ, & Maniatis T (1998) A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. *Embo Journal* 17(22):6747-6756.

212. Deonier RC & Williams JW (1970) Self-Association of Muramidase (Lysozyme) in Solution at 25 Degrees, Ph 7.0, and 1=0.20. *Biochemistry-Us* 9(22):4260-&.

213. Debaere I*, et al.* (1992) Polar Zipper Sequence in the High-Affinity Hemoglobin of Ascaris-Suum - Amino-Acid-Sequence and Structural Interpretation. *Proceedings of the National Academy of Sciences of the United States of America* 89(10):4638-4642.

214. Huang Y, Gattoni R, Stévenin J, & Steitz JA (2003) SR Splicing Factors Serve as Adapter Proteins for TAP-Dependent mRNA Export. *Molecular cell* 11(3):837-843.

215. Huang Y, Yario TA, & Steitz JA (2004) A molecular link between SR protein dephosphorylation and mRNA export. *Proceedings of the National Academy of Sciences of the United States of America* 101(26):9666-9670.

216. Lai M-C & Tarn W-Y (2004) Hypophosphorylated ASF/SF2 Binds TAP and Is Present in Messenger Ribonucleoproteins. *Journal of Biological Chemistry* 279(30):31745-31749.

217. Tintaru AM*, et al.* (2007) Structural and functional analysis of RNA and TAP binding to SF2/ASF. *Embo Rep* 8(8):756-762.

218. Tzeng SR & Kalodimos CG (2012) Protein activity regulation by conformational entropy. *Nature* 488(7410):236-240.

219. Radhakrishnan I*, et al.* (1997) Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. *Cell* 91(6):741-752.

220. Radhakrishnan I, Perez-Alvarado GC, Dyson HJ, & Wright PE (1998) Conformational preferences in the Ser133-phosphorylated and non-phosphorylated forms of the kinase inducible transactivation domain of CREB. *Febs Lett* 430(3):317-322.

221. Koshland DE (1958) Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences of the United States of America* 44(2):98-104.

222. Burgen AS (1981) Conformational changes and drug action. *Fed Proc* 40(13):2723-2728.

223. Ma CT, Hagopian JC, Ghosh G, Fu XD, & Adams JA (2009) Regiospecific Phosphorylation Control of the SR Protein ASF/SF2 by SRPK1. *Journal of molecular biology* 390(4):618-634.

224. Wang HY*, et al.* (1998) SRPK2: A differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells. *Journal of Cell Biology* 140(4):737-750.

225. Colwill K, *et al.* (1996) SRPK1 and Clk/Sty protein kinases show distinct substrate specificities for serine/arginine-rich splicing factors. *Journal of Biological Chemistry* 271(40):24569-24575.

226. Ramsey NF (1953) Electron Coupled Interactions between Nuclear Spins in Molecules. *Phys Rev* 91(2):303-307.

227. Venanzi TJ (1982) Nuclear magnetic resonance coupling constants and electronic structure in molecules. *Journal of Chemical Education* 59(2):144.

228. Pervushin K, *et al.* (1998) NMR scalar couplings across Watson–Crick base pair hydrogen bonds in DNA observed by transverse relaxation-optimized spectroscopy. *Proceedings of the National Academy of Sciences* 95(24):14147-14151.

229. Cavanagh J (2007) *Protein NMR spectroscopy : principles and practice* (Academic Press, Amsterdam ; Boston) 2nd Ed pp xxv, 885 p.

230. Mason J (1987) *Multinuclear NMR* (Plenum Press, New York) pp xx, 639 p.

231. Karplus M (1963) Vicinal Proton Coupling in Nuclear Magnetic Resonance. *J Am Chem Soc* 85(18):2870-2871.

232. Karplus M (1959) Contact Electron-Spin Coupling of Nuclear Magnetic Moments. *The Journal of chemical physics* 30(1):11-15.

233. Wagner G (1983) Characterization of the distribution of internal motions in the basic pancreatic trypsin inhibitor using a large number of internal NMR probes. *Quarterly Reviews of Biophysics* 16(01):1-57.

234. Cierpicki T, Zhukov I, Byrd RA, & Otlewski J (2002) Hydrogen Bonds in Human Ubiquitin Reflected in Temperature Coefficients of Amide Protons. *J Magn Reson* 157(2):178-180.

235. Tjandra N & Bax A (1997) Solution NMR measurement of amide proton chemical shift anisotropy in N-15-enriched proteins. Correlation with hydrogen bond length. *J Am Chem Soc* 119(34):8076-8082.

236. Liwang AC & Bax A (1997) Solution NMR characterization of hydrogen bonds in a protein by indirect measurement of deuterium quadrupole couplings. *J Magn Reson* 127(1):54-64.

237. Bougault C, Feng L, Glushka J, Kupce E, & Prestegard JH (2004) Quantitation of rapid proton-deuteron amide exchange using hadamard spectroscopy. *Journal of biomolecular NMR* 28(4):385-390.

238. Cordier F & Grzesiek S (1999) Direct observation of hydrogen bonds in proteins by interresidue (3h)J(NC ') scalar couplings. *J Am Chem Soc* 121(7):1601-1602.

239. Barfield M (2002) Structural Dependencies of Interresidue Scalar Coupling h3JNC‘ and Donor 1H Chemical Shifts in the Hydrogen Bonding Regions of Proteins. *J Am Chem Soc* 124(15):4158-4168.

240. Gsponer J, Hopearuoho H, Cavalli A, Dobson CM, & Vendruscolo M (2006) Geometry, energetics, and dynamics of hydrogen bonds in proteins: structural information derived from NMR scalar couplings. *J Am Chem Soc* 128(47):15127-15135.

241. Gueron M, Leroy JL, & Griffey RH (1983) Proton Nuclear Magnetic-Relaxation of N-15-Labeled Nucleic-Acids Via Dipolar Coupling and Chemical-Shift Anisotropy. *J Am Chem Soc* 105(25):7262-7266.

242. Goldman M (1984) Interference Effects in the Relaxation of a Pair of Unlike Spin-1/2 Nuclei. *J Magn Reson* 60(3):437-452.

243. Tjandra N, Grzesiek S, & Bax A (1996) Magnetic field dependence of nitrogen-proton J splittings in N-15-enriched human ubiquitin resulting from relaxation interference and residual dipolar coupling. *J Am Chem Soc* 118(26):6264-6272.

244. Yao LS, Ying JF, & Bax A (2009) Improved accuracy of N-15-H-1 scalar and residual dipolar couplings from gradient-enhanced IPAP-HSQC experiments on protonated proteins. *Journal of biomolecular NMR* 43(3):161-170.

245. Ottiger M, Delaglio F, & Bax A (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J Magn Reson* 131(2):373-378.

246. Griesinger C, Sorensen OW, & Ernst RR (1986) Correlation of Connected Transitions by Two-Dimensional Nmr-Spectroscopy. *Journal of Chemical Physics* 85(12):6837-6852.

247. Weisemann R, *et al.* (1994) Determination of HN,Hα and HN,C′ coupling constants in 13C, 15N-labeled proteins. *Journal of biomolecular NMR* 4(2):231-240.

248. Wang AC & Bax A (1996) Determination of the backbone dihedral angles phi in human ubiquitin from reparametrized empirical Karplus equations. *J Am Chem Soc* 118(10):2483-2494.

249. de Alba E & Tjandra N (2006) On the accurate measurement of amide one-bond N-15-H-1 couplings in proteins: Effects of cross-correlated relaxation, selective pulses and dynamic frequency shifts. *J Magn Reson* 183(1):160-165.

250. Madsen JC, Sorensen OW, Sorensen P, & Poulsen FM (1993) Improved Pulse Sequences for Measuring Coupling-Constants in C-13,N-15-Labeled Proteins. *Journal of biomolecular NMR* 3(2):239-244.

251. Cremer D & Grafenstein J (2007) Calculation and analysis of NMR spin-spin coupling constants. *Phys Chem Chem Phys* 9(22):2791-2816.

252. Pecul M, Leszczynski J, & Sadlej J (2000) The shielding constants and scalar couplings in N-H center dot center dot center dot O=C and N-H center dot center dot center dot N=C hydrogen bonded systems: An ab initio MO study. *J Phys Chem A* 104(34):8105-8113.

253. Sahakyan AB, Shahkhatuni AG, Shahkhatuni AA, & Panosyan HA (2008) Electric field effects on one-bond indirect spin-spin coupling constants and possible Biomolecular perspectives. *J Phys Chem A* 112(16):3576-3586.

254. Ma JH, Gruschus JM, & Tjandra N (2009) N-15-H-1 Scalar Coupling Perturbation: An Additional Probe for Measuring Structural Changes Due to Ligand Binding. *J Am Chem Soc* 131(29):9884-9885.

255. Smith MA, Hu H, & Shaka AJ (2001) Improved broadband inversion performance for NMR in liquids. *J Magn Reson* 151(2):269-283.

256. Hall JB & Fushman D (2003) Characterization of the overall and local dynamics of a protein with intermediate rotational anisotropy: Differentiating between conformational exchange and anisotropic diffusion in the B3 domain of protein G. *Journal of biomolecular NMR* 27(3):261-275.

257. Sahakyan AB, Shahkhatuni AA, Shahkhatuni AG, & Panosyan HA (2008) Dielectric permittivity and temperature effects on spin-spin couplings studied on acetonitrile. *Magn Reson Chem* 46(1):63-68.

258. Marsh JA, Singh VK, Jia Z, & Forman-Kay JD (2006) Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. *Protein Science* 15(12):2795-2804.

259. Cleveland DW, Hwo SY, & Kirschner MW (1977) Physical and chemical properties of purified tau factor and the role of tau in microtubule assembly. *Journal of molecular biology* 116(2):227-247.

260. Mukrasch MD*, et al.* (2009) Structural polymorphism of 441-residue tau at single residue resolution. *PLoS biology* 7(2):e34.

261. Bruschweiler R (1996) Cross-correlation-induced J coupling. *Chem Phys Lett* 257(1-2):119-122.

262. Leitz D, Vogeli B, Greenwald J, & Riek R (2011) Temperature dependence of 1HN-1HN distances in ubiquitin as studied by exact measurements of NOEs. *J Phys Chem B* 115(23):7648-7660.

263. Tuttle T, Kraka E, Wu AA, & Cremer D (2004) Investigation of the NMR spin-spin coupling constants across the hydrogen bonds in ubiquitin: The nature of the hydrogen bond as reflected by the coupling mechanism. *J Am Chem Soc* 126(16):5093-5107.

264. Juranic N, Moncrieffe MC, Likic VA, Prendergast FG, & Macura S (2002) Structural dependencies of (h3)J(NC ') scalar coupling in protein H-bond chains. *J Am Chem Soc* 124(47):14221-14226.

265. Ma J, Gruschus JM, & Tjandra N (2009) 15N-1H scalar coupling perturbation: an additional probe for measuring structural changes due to ligand binding. *J Am Chem Soc* 131(29):9884-9885.

266. Grafenstein J & Cremer D (2004) Analysis of the paramagnetic spin-orbit transmission mechanism for NMR spin-spin coupling constants using the paramagnetic spin-orbit density distribution. *Chem Phys Lett* 383(3-4):332-342.

267. Wang YX*, et al.* (1999) Measurement of (3h)J(NC ') connectivities across hydrogen bonds in a 30 kDa protein. *Journal of biomolecular NMR* 14(2):181-184.

268. Cordier F, Nisius L, Dingley AJ, & Grzesiek S (2008) Direct detection of N-H center dot center dot center dot O=C hydrogen bonds in biomolecules by NMR spectroscopy. *Nat Protoc* 3(2):235-241.

269. Baxter NJ, Hosszu LLP, Waltho JP, & Williamson MP (1998) Characterisation of low free-energy excited states of folded proteins. *Journal of molecular biology* 284(5):1625-1639.

270. Grzesiek S, Cordier F, & Dingley AJ (2002) Scalar couplings across hydrogen bonds. *Methods in Enzymology*, ed Thomas L. James VDUS (Academic Press), Vol Volume 338, pp 111-133.

271. Cordier F, Barfield M, & Grzesiek S (2003) Direct observation of C-alpha-H-alpha center dot center dot center dot O=C hydrogen bonds in proteins by interresidue (h3)J(C alpha C ') scalar couplings. *J Am Chem Soc* 125(51):15750-15751.

10. Appendix

10.1 Pulse program used to measure Hα-Cα. HACANNH

```
#include "bits.mz"
#include <Avance.incl>
#include <Grad.incl>

;started from hnca-ge.abx
;hcannh-3D with grad enhancement 1/7/00

;p1 = 90 deg (10us) 1H pulse @pl1
;p30 = 90 deg (60us) 1H pulse @pl30
;p7 = 90 deg (50us) 15N pulse @pl7
;p31 = 90 deg (180us) 15N pulse @pl31
;p5 = selective 90 deg (53.0us) 13CA pulse @pl5
;p4 = selective 180 deg (23.7*2us) 13CA pulse @pl4
;p6 = 180 deg (191.7us=180deg) 13C pulse @sp1 using sinc1.0
;p6 = 90 deg (191.7us=90deg) 13C pulse @sp2 using sinc1.0

#define NITRO          ; selection by hand
#define CA
;#define FIDCHECK2D
;#define BS


define loopcounter NLOOP
define loopcounter CLOOP

"NLOOP=38"          ;maximum "NLOOP=76"
"CLOOP=49"          ;maximum "CLOOP=126"

"d11=50m"
"d12=100u"


"d20=p7-p4"
"d21=p7-p1"
;"d28=p4-p1"

;Gradient pulses
"p17=0.75m"
"p20=1.5m"
"p21=1.1m"
"p22=1.0m"
```

```
"p23=1.0m"
"p24=0.200m"
"p25=0.074m"            ; optimization of gradient selection
"p26=2.705m"
"p27=2.705m"
"p28=2.0m"
;"p29=1.5m"

"d2=1.5m-p21"
"d3=2.65m-p22"
"d4=2.65m-p23"
"d6=0.25m-p24"
"d7=0.25m-p25"
"d15=1.5m-p17"


;-----Carbon Evolution----------
"d10 = 8u"
;"d16 = 14m-0.63661977*p5-4u-d10-p21-p6-p7"
;"d17 = 14m-0.63661977*p5-11u-p21-p7"
;--- if BS compensating pulse is used
"d16 = 14m-0.63661977*p5-4u-p6-p21-p7-d10"
"d17 = d16+d10-8u-4u"

"in10=inf1/2"
"in17=in10"   ; for C dimension

;----Nitrogen Evolution---------
; !!!! all increment MUST satisfy the following relations
;   ******* in22 = in25  &&  in18 = in24 + in25 ********
"d22=7m-p6*0.5"             ;d22 =6.25m for in22 = 125u and 50 incs
"d18=d22-5.5m"              ;in18 = 225u
"d25=d22-p4*2-d24-5u"       ;in25 = 125u
"d24=5u"                    ;in24 = 100u   for 15N incr of 700u
"d19=5.4m-p26-300u-10u-p1"
;--------Nitrogen increasement-----
"in25=10u"
"in24=inf2/2-2*in25"
"in22=in25"
"in18=in24+in25"


1       ze
        d12
        10u
        10u ru1
```

```
          10u ru2
;------- selection by hand
#ifdef FIDCHECK2D
          10u iu2                ; for recording 2D IP-(H)CA(N)NH spectrum
          d12 dp14               ; if 2D (HCA)NNH, as Sx is required instead of Sy
#endif
;------- END selection by hand
          10u
2         10u
          1m
          d11 do:N
          d12
3         d12*4.0
4         d12*5.0
5         d12*5.0
6         d12*4.0 do:N
7         5u do:C1
          10u pl4:C1
          d1 BLKGRAD
          1m UNBLKGRAD
          10u pl1:H
          10u pl7:N
;------- removal of boltzmann magnetization ---
          (p7 ph7):N (p4 ph0):C1
          2u
       p20:gp20
          1.0m
;------- INEPT HA -> CA ----------------------
          (p1 ph0):H
       10u
       p21:gp21
          d2
          (center (p1*2 ph1):H (p4*2 ph8):C1)
          ;(d28 p1*2 ph1):H (p4*2 ph8):C1
       10u
       p21:gp21
          d2
          (p1 ph2):H
       10u

;------- flip back ----------------------
; problems with water flip back !!
;          if "l2==1" goto 13
;       (p2 ph0):H
;          if "l2==2" goto 14
;13      (p2 ph20):H
```

```
;14     50u pl1:H
       300u pl5:C1

;------- IPAP selection -----------------------
         if "l2==1" goto 15
         (p5 ph0):C1
         2u
         p17:gp20
         d15 pl4:C1
         (center (p4*2 ph5):C1 (p1*2 ph0):H)
         ;(p4*2 ph5):C1 (d28 p1*2 ph0):H
         2u
         p17:gp20
         d15 pl5:C1
         (p5 ph3):C1
         4u
         p20:gp20
         1m
         (p5 ph14):C1
         2u
         if "l2==2" goto 16
15       20u                    ; symmetry gradients
         p17:gp20
         d15
         p17:gp20
         d15
         p20:gp20
         1m

         (p5 ph0):C1
         2u
;------- end IPAP selection -----------------------

;---------------- CT evolution on CA -------------------
16       d10
         (p6:sp1 ph0):C2              ;carbonyl decoupling with sinc
         2u
       p21:gp19
         d16 pl4:C1
         (d20 p4*2 ph13):C1 (p7*2 ph0):N
         2u
       p21:gp19
         d17 pl5:C1
       4u
         (p6:sp1 ph0):C2              ;BS compensating pulse
       2u
```

```
        8u pl5:C1

        (p5 ph6):C1
        5u
      p28:gp28
        300u pl5:C1
      (p1 ph1):H              ;trim H decoup on
      5u pl30:H
      5u cpds1:H
;--------------- CT evolution on N -------------------
        (p7 ph10):N
        d22
        (p6:sp1 ph0):C2             ;carbonyl decoupling with sinc
        d22
        (p7*2 ph7):N
        d24
      5u pl4:C1
        (p4*2 ph0):C1
        d25
        (p6:sp1 ph0):C2             ;carbonyl decoupling with sinc
        d18
        0.1m do:H           ;H decoupling off
      10u pl1:H
      (p1 ph3):H              ;trim pulse
      d19 pl1:H
;alternating Rance-Kay coherence encoding
      if "l1==1" goto 10
      if "l1==2" goto 20
10      p26:gp26              ;coherence encoding gradient
       100u
      goto 30
20      p27:gp27              ;coherence encoding gradient
       100u
30      200u pl1:H
;end coherence encoding
        (p7 ph9):N (d21 p1 ph0):H
;--------- Rance-Kay transfer N -> HN --------------------
        2u
      p22:gp22
        d3
        (d21 p1*2 ph0):H (p7*2 ph7):N
        2u
      p22:gp22
        d3
        (p1 ph1):H (p7 ph8):N
        2u
```

```
      p23:gp23
         d4
         (d21 p1*2 ph0):H (p7*2 ph7):N
         2u
      p23:gp23
         d4
         (d21 p1 ph0):H (p7 ph0):N
         4u
      p24:gp24
         d6 pl31:N
         (p1*2 ph0):H
         2u
      p25:gp25
         d7 BLKGRAD
         (2u ph0)
         go=2 ph31 cpd2:N
      10u do:N
         1m
         d11 wr #0 if #0 zd
; ---------- 3D IPAP selection
#ifndef FIDCHECK2D
 d12 iu2
 lo to 3 times 2
 d12 ru2
#endif



#ifdef NITRO
d12 ip9
d12 ip9
d12 iu1
lo to 4 times 2
d12 ru1
d12 dd22
d12 id18
d12 id24
d12 id25
lo to 5 times NLOOP
d12 rd22
d12 rd24
d12 rd25
d12 rd18
#endif
#ifdef CA
d12 ip6
```

```
lo to 6 times 2
d12 id10
d12 dd17
d12 ip31
d12 ip31
lo to 7 times CLOOP
#endif
#ifdef BS
d12 ip13
lo to 7 times td1
#endif
d12
d12 do:C1
d12 do:N
exit


ph0=0
ph1=1
ph2=1 3
;ph2=3
ph3=3
ph5=0
;ph6=0 2
ph6=0
ph7=0
ph8=1
ph9=2
ph10=0 0 2 2
ph11=0
ph13=0
;ph13=(360) 11                ;BS compensation
ph14=2
ph20=2
ph31=0 2 2 0
```

10.2 TROSY-HSQC interleaved experiment for measure N-H couplings

```
 ;hsqc15N.new
;D. Lee, Nov. 2002
;15N-1H HSQC correlations without water saturation
;The delay for 3-9-19 watergate (d5) should be matched
;with 1/d;d=distance of next null point (in Hz).
;S. Mori et al, JMR B108, 94-98 (1995)
;pl1   : power for 1H
;pl2   : power for 13C
;pl3   : power for 15N
;pl13  : power for 15N waltz16 decoupling
;p1    : 90 degree hard pulse 1H
;p3    : 90 degree hard pulse 13C
;p4    : 180 degree hard 13C pulse (225d for 5/600)
;p5    : 90 degree hard pulse 15N
;pcpd3 : 90 deg cpd-pulse15N(waltz16,160u)
;d1    : relaxation delay
;d2    : INEPT delay (~2.7m)
;d5    : delay for 3-9-19=1/(Hz between nulls)
;in0   : 1/(2 SW) (Hz)
;p21   : 500u (Gradient in first INEPT)
;p22   : 500u (Gradient for z-filter)
;p23   : 1m (Gradient for second INEPT)
;gpz1  : 19%
;gpz2  : 30%
;gpz3  : 65%

;trosy.new
;D. Lee, Nov. 2002

;optimization of water flip back:
;- optimize watergate (o1,sp2,ph26)
;- optimize water flip back (sp1,ph16,ph18,ph17,ph19)

;K. Pervushin et al, PNAS, 94, 12366 (1997)

;pl1   : power for 1H
;pl2   : power for 13C
;pl3   : power for 15N

;sp1   : water flipback power
;sp2   : water flipback power in watergate
;spnam1: gauss128_5
;spnam2: gauss128_5
```

```
;p1    : 90 degree hard pulse 1H
;p3    : 90 degree hard pulse 13C
;p4    : 180 degree hard pulse 13C (225d for 5/600)
;p5    : 90 degree hard pulse 15N
;p11   : water flipback pulse (1.5m)


;p21   : 500u (Gradient in first INEPT)
;p22   : 500u (Gradient in second INEPT)
;p23   : 900u (Gradient in watergate)
;gpz1  : 19%
;gpz2  : 15%
;gpz3  : 32%


;d1    : relaxation delay
;d2    : INEPT delay (~2.7ms)
;in0   : 1/(2 SW) (Hz)

#include <Avance_dl.incl>

define delay INEPT_W
define delay INEPT_D

define delay INEPT1
define delay INEPT2
define delay INEPT3
define delay INEPT4

#define GRADIENT1  10u p21:gp1 200u
#define GRADIENT2  10u p22:gp2 200u
#define GRADIENT3  10u p23:gp3 200u

"p2=2*p1"
"p6=2*p5"

;"in0=inf1/2"
"in10=inf1/2"
;"d0=in0/2-p5*2/3.14159-p1"
"d10=in10/2-p5*2/3.14159-p1"
"d3=d5/2-p5"
"INEPT_D=d2-p21-210u"
"INEPT_W=d2-(p23+210u+p1*2.3846+d5*2.5)"


"in0=inf1/2"
```

```
"d0=in0/2-(p3*2 + 1.5u)"
"INEPT1=d2-(p21+p11+210u)-10u"
"INEPT2=d2-(p22+p11+210u)-10u"
"INEPT3=d2-(p23+p11+210u)-10u"
"INEPT4=d2-(p23+p11+210u)-10u-p5"

"l2 = 1"

;=========================trosy=================================
=
1  10u ze
2  1m
   10u*5
3  10u*2
4  10u
   d1
   20u pl1:f1
   20u pl2:f2
   20u pl3:f3
   20u LOCKH_ON
;----------------------------------------first INEPT
5  10u
   (p1 ph20):f1
   10u
   (p11:sp1 ph18:r):f1
   GRADIENT1
   INEPT1 pl1:f1
   (center(p2 ph21):f1 (p6 ph20):f3)
   GRADIENT1
   INEPT1
   (p11:sp1 ph19:r):f1
   10u
   (p1 ph23):f1
;----------------------------------------15N evolution
if "l2 %2 == 1" goto 31
   (p5 ph2):f3
goto 32
31 (p5 ph1):f3
32 d0
   (p3 ph23 1.5u p4 ph20 1.5u p3 ph23):f2
   d0
;----------------------------------------second INEPT
   (p1 ph10):f1
   10u
   (p11:sp1 ph17:r):f1
   GRADIENT2
```

```
  INEPT2 pl1:f1
  (center(p2 ph20):f1 (p6 ph20):f3)
  GRADIENT2
  INEPT2
  (p11:sp1 ph16:r):f1
  10u
  (center(p1 ph20):f1 (p5 ph12):f3)
;---------------------------------------WATERGATE
  GRADIENT3
  INEPT3
  (p11:sp2 ph26:r):f1
  10u
  (center(p2 ph20):f1 (p6 ph20):f3)
  10u
  (p11:sp2 ph26:r):f1
  GRADIENT3
  INEPT4 LOCKH_OFF
  (p5 ph11):f3
;---------------------------------------acquisition
  go=2 ph31
  ;1m mc #0 to 2 F1EA(ip10*2 & ip12*2 & ip17*2 & iu2,id0)
   1m wr #0 if #0 zd

;10u do:f1
;10u do:f2
;10u do:f3
;10u do:f4
;10u LOCKH_OFF



;=========================hsqc============================
;1  10u ze
6  1m do:f3
  d1 pl1:f1
  20u pl3:f3
  20u LOCKH_ON
;---------------------------------------first INEPT
  (p1 ph20):f1
  GRADIENT1
  INEPT_D
  ;(center(p2 ph21):f1 (p6 ph1):f3)
  (center(p2 ph21):f1 (p6 ph3):f3)
  GRADIENT1
  INEPT_D
  (p1 ph21):f1
```

```
  GRADIENT2
;---------------------------------------15N evolution
  ;(p5 ph1):f3
   (p5 ph3):f3
  ;(refalign (d0 p2 ph23 d0):f1 center (p3 ph23 1.5u p4 ph20 1.5u p3 ph23):f2)
  (refalign (d10 p2 ph23 d10):f1 center (p3 ph23 1.5u p4 ph20 1.5u p3 ph23):f2)
  (p5 ph20):f3
  GRADIENT2
;----------------------------------------second INEPT
  (p1 ph22):f1
  GRADIENT3
  INEPT_W
  (p1*0.2308 ph21 d5 p1*0.6923 ph21 d5 p1*1.4615 ph21):f1
  (d3 p6 ph1 d3):f3
  (p1*1.4615 ph23 d5 p1*0.6923 ph23 d5 p1*0.2308 ph23):f1
  GRADIENT3
  INEPT_W pl13:f3 LOCKH_OFF
;---------------------------------------acquisition
  ;go=2 ph31 cpd3:f3
 go=6 ph30 cpd3:f3
 ; 1m do:f3 mc #0 to 2 F1PH(ip1,id0)
  1m do:f3  wr #1 if #1 zd


;=====phase =======
;======trosy========
10u ip10*2
10u ip12*2
10u ip17*2
10u iu2
;======hsqc=====
10u  ip3

lo to  3 times 2

;======increament=======
;=====trosy======
10u  id0
;=====hsqc=======
10u  id10

lo to  4 times l1

10u do:f1
;10u do:f2
10u do:f3
```

```
;10u do:f4
10u LOCKH_OFF
exit

;=========================trosy=================================
=======
ph1 =1 3 2 0
ph2 =1 3 0 2
ph31=1 3 2 0
ph10=3 3 3 3
ph11=0 0 0 0
ph12=3 3 3 3


ph16=0
ph17=1
ph18=2
ph19=3

ph20=0
ph21=1
ph22=2
ph23=3

ph26=2

;================hsqc================
;ph1 =0 2
ph3=0 2
;ph31=2 0
ph30=2 0

;ph20=0
;ph21=1
;ph22=2
;ph23=3
```

10.3 HNN pulse program

```
;hnngp3d
;avance-version
;3D HNN (800 MHz)
;o2p = 56.0 ppm
;Panchal et el J Biomol NMR 20, 135, 2001
;1) exclude Rance-Kay trick
;2) use maximum resolution for t2 including also period k
;3) water flip back and final watergate for water suppression

;sequence checkd and corrected by DMO

;typed by jeet at 800 MHz

prosol relations =<triple>

#include <Avance.incl>
#include <Grad.incl>

;################ constant values ###############
"d2=2m"
;"d4=2.7m-p2-3u-4u-p16-d16"
"d5=2.7m-p11-3u-20u-3u-p16-d16"        ;added DMO
"d6=5.4m"                        ;k
"d7=13.5m"                         ;Tn
"d11=30m"
"d14=p14"
"d15=12m"                        ;Tcn=12-17m
"d16=200u"   ;shengqi

;################ center pulses ################
"d3=p21-p1"
"d8=(p14-p21*2)*0.5"

;################ incremented time t1 ##########
"d0=3u"                          ;t1/2
"d9=d0+d14+d7+d8"                    ;t1/2 NOTE that d9<=d8+3u

;################ incremented time t2 ##########
"d20=d9-d8"                        ;t2/2 decreased d20
;d7=d9= 15ms = Tn
"d10=3u"                         ;t2/2 increased
"d21=d10+d9+d8+d7-d6-p1-p21*2-24u"

"in0=inf1/2"
```

```
"in10=inf2/2"

"in9=in0"
"in20=in10"

"spoff1=0"
"spoff2=0"
"spoff3=0"
"spoff4=0"
;"spoff6=0"
"spoff5=bf2*(cnst21/1000000)-o2"

aqseq 312

1 ze
  d11 pl13:f3
2 d11 do:f3
  d1 pl1:f1
  20u pl2:f2
  50u UNBLKGRAD
;------- start 90-degree on N to get rid of boltzmann ---------
          (p21*2 ph7):f3 (p5 ph0):f2
          2u
       p20:gp20           ;(0.5ms @ 8G/cm)
          1.0m  pl11:f2
;------- start 90-degree on hn -----------------------
  (p1 ph0)
   10u
  d2 pl3:f3
  p15:gp5    ;(0.5ms @ 8G/cm)
  d16
  (d3 p1*2 ph0) (p21*2 ph0):f3
  10u
  p15:gp5
  d16
  d2
  (p1 ph9):f1
  3u
  20u pl11:f1
  (p11:sp1 ph7:r):f1
  3u
  p16:gp1    ;(1ms @11G/cm)
  d16 pl1:f1
  (p21 ph1):f3

;#### Start t1 d0=A ; d7=15m=B ; d9=C=15m-3u ; d6=5.4m=k #########
```

```
  (d6 p1 ph9 20u pl19 4u cpds1 ph7):f1 (d0 p14:sp5 ph0 d7 p14:sp3 ph0):f2 (d0 d14 d7
d8 p21*2 ph0 d9):f3

;############ end t1 #############

  (p21 ph2):f3
  4u do:f1
  20u pl1:f1
  (p1 ph10):f1

  4u
  p16:gp2
  d16
  (p1 ph9):f1
  3u
  10u pl19:f1
  10u
  4u cpds1 :f1
  (p13:sp2 ph2):f2
  d15  pl11:f2                          ;Tcn = 12-16 ms
  (d8 p21*2 ph0):f3 (p14:sp3 ph0):f2
  d15                    ;Tcn = 12-16 ms
  (p13:sp4 ph0):f2
  4u do:f1
  20u pl1:f1
  (p1 ph10):f1
  4u
  p16:gp3
  d16
  (p1 ph9):f1
  3u
  20u pl19:f1
  4u cpds1 :f1
  (p21 ph8):f3
  5u

;#### Start t2 d10=F ; d20=D=15m-3u ; d7=E=15m ; d6=5.4m=k #########

  (d21 4u do 20u pl1 p1 ph10):f1 (d20 p14:sp3 ph0 d7 p14:sp5 ph0):f2 (d20 d8 p21*2 ph0
d7 d14 d8 d10):f3

;######## end t2 ###############

  (p21 ph0):f3
  10u
```

```
  (p11:sp1 ph3:r):f1
  3u
  20u pl1:f1
  (p1 ph0):f1
  3u
  p16:gp4
  d16
  4u
  d5 pl11:f1                    ;d4 substituted by d5 DMO
  (p11:sp1 ph4:r):f1                    ;p2 substituted by p28 DMO
  3u
  20u pl1:f1
  (d3 p1*2 ph5) (p21*2 ph0):f3
  3u
  20u pl11:f1
  (p11:sp1 ph4:r):f1
  3u
  d5                            ;d4 substituted by d5 DMO
  p16:gp4
  d16 pl16:f3
  4u BLKGRAD
  go=2 ph31 cpd3:f3
  d11 do:f3 mc #0 to 2
  F1PH(ip1, id0 & dd9)
  F2PH(rd0 & rd9 & ip8, id10 & dd20)
exit

ph0=0
ph1=0 0 2 2
ph2=0 2
ph3=2
ph4=0
ph5=2
ph7=0
ph8=0 0 0 0 2 2 2 2
ph9=1
ph10=3
ph31=0 0 2 2 2 2 0 0

;d1: relaxation delay; 1-5*T1
;d11: delay for disk I/O [30 msec]
;pl1: f1 power H high power pulse
;p1: f1 90 degree high power pl1 pulse
;p11: f1 90 degree shaped pulse using sp1
;p28: f1 90 degree H2O shaped pulse  sp6
;d2: 2.25-2.75m (Jhn=92Hz)
```

;d3: p21-p1 (center pulse)
;pl2: f2 power C13 hard pulse
;pl3: f3 power N15 hard pulse
;p21: f3 90 degree high power N15 high power pl3 pulse
;p5: f2 90 degree high power C13 high power pl2 pulse
;pl11: 120db
;p13: 90 degree 13C shape
;p14: 180 degree 13C shape
;gpz1: 37%
;gpnam1: sine.100 shape of gradient pulse
;sp5: f2 power 180 (CA) soft shaped
;p14: f2 soft shaped pulse 180 null at 18200 Hz
;spnam1: use shape sinc.1000
;spoff1: 0
;spnam5: use shape G3.256
;spoff5: use offset 18200.0 Hz (600 MHz) on CO, null CA
;spnam6: use shape sinc.1000
;spoff6: 0
;spnam3: use shape G3.256
;spoff3: use offset 0.0 Hz (600 MHz) on CA, null CO
;d6: 5.4m (k in sequence)
;pl19: f1 power for cpds1 decoupling
;cpd1: dipsi2 (along x)
;pcpd3: f1 cpds1 pulse length (60-70u) use 65u
;d0: t1/2
;d7: Tn [12-16m] use 15m
;d9: d0+d14+d7+d8 (Tn-t1/2) SET ALWAYS d9<=d8+3u
;gpz1: 43 % power level
;gpnam1: sine.100 shape of gradient pulse
;gpz2: 27 % power level
;gpnam2: sine.100 shape of gradient pulse
;sp2: f2 power 90 (CA) soft shaped
;spnam2: use shape G4.256
;spoff2: use offset 0.0 on CA, null CO
;sp4: f2 power 90 (CA) soft shaped
;spnam4: use shape G4.256tr (time reverse shape)
;spoff4: use offset 0.0 on CA, null CO
;d15: [12-18m] Tcn, use 16m
;gpz3: 37%
;gpnam3: sine.100 shape of gradient pulse
;d20: t2/2 (Tn-t2/2) [d9-d8]
;d10: t2/2
;sp1: f1 power H2O soft shaped pulse
;p11: (2m) half-gaussian soft pulse with power sp1
;spnam1: hg.1000 half gaussian shape
;p16: gradient pulse 1 ms

```
;gpz4: 57%
;gpnam4: sine.100 shape of gradient pulse
;gpz5: 50%
;gpnam5: sine.100 shape of gradient pulse
;cnst21: CO chemical shift
;d4: 2.7m-p2-3u-4u-p16-d16
;pl16: f3 power level N15 cpd decoupling
;cpd3: cpdprg3 decoupling sequence [garp]
;pcpd3: f3 decoupling 90 degree pulse [>170 u]
;l1: td1/2 STATES/TPPI
;l2: td2/2 STATES/TPPI
;in0: 1/(2* SW(N15)) = DW(N15)
;in9: SET ALWAYS in9=in0
;nd0: 2
;in10: 1/(2* SW(N15)) = DW(N15)
;in20: SET ALWAYS in20=in10
;nd10: 2
;NS: 8*n
;DS: 16
;td1: number of experiments N15
;td2: number of experiments N15
;MC2: STATES-TPPI
```

10.4 BSD-IPAP-HSQC pulse program

```
#include "bits.mz"
#include <Avance.incl>
#include <Grad.incl>

"d6=12.5m"
"d10=3u"
;"d14=3u"
;"d0=d6+d10+p4*4+p9*2+d14*2+3u"
"d11=50m"
"d12=200u"
;"d18=12.5m-3u-p26"
;"d19=50ms"
"d20=10u"
;"d7=d20*4+p9*4+37u"
"d7=d20*4+p14-63u"

;"d21=2.4m-p21-203u-p7"
"d21=2.4m-p21-203u-p7"
;"d23=2.4m-p23"
"d23=2.4m-p23"
;"d24=2.4m-p24"
"d24=2.4m-p24"
;"d25=p25+110u"
;"d25=p25+p9+110u+p9+10u"
"d25=p25+p9+110u+p19+10u"

"in20=inf1/4"
"spoffs13=bf1*(cnst21/1000000)-o1"

1       ze
        1m
        1m
2       d11 do:f3 do:f2
        d12
4       d12*2
5       d12*3
6       10u
        1m
        10u
        1m
        1m do:f2
        1m do:f3
```

```
        d1
        1m
        10u pl1:f1
        10u pl7:f3
        10u UNBLKGRAD ;setnmr3|0 setnmr0|34|32|33 ctrlgrad 0
        (p7 ph1):f3
        p21:gp28
        100u
;------- start 90-degree on hn ----------------------
        (p1 ph0):f1
        3u
        p21:gp21
        d21
        200u
        (center (p1*2 ph0):f1 (p7*2 ph0):f3)
        d21
        p21:gp21
        203u
        (p1 ph8):f1             ;INEPT to 15N
        3u pl2:f1
        6u
        3u pl1:f1
if "l1==1" goto 110
       (p7 ph1):f3
        3u gron22
        2.7m
        3u groff
       (center (p1 ph0 p1*2.2 ph1 p1 ph0):f1 (p7*2 ph0):f3)
        3u gron22
        2.7m
        3u groff
        (center(p1 ph0):f1 (p7 ph0):f3)
        goto 111
110     3u
        (p10:sp5 ph4:r):f1
        6u
        (p7 ph0):f3
        3u gron22
        1.35m
        3u groff
        (p1 ph0 p1*2.2 ph1 p1 ph0):f1
        3u gron22
        1.35m
        3u groff
        (p7*2 ph0):f3
        3u gron22
```

```
        1.35m
        3u groff
       (p1 ph0 p1*2.2 ph1 p1 ph0):f1
        3u gron22
        1.35m
        3u groff
        (p7 ph0):f3
111     5u
        p20:gp20
        200u pl16:f2
        5u
        3u pl2:f1
;---------start f3 evolution --------------------------------
       (p7 ph7):f3
        d7
        (p13:sp13 ph11):f1
       if "l2==2" goto 27
        p26:gp26
        100u
        (p7*2 ph9):f3
        p26:gp27
        goto 26
27      p26:gp27
        100u
        (p7*2 ph9):f3
        p26:gp26
26      3u gron5
        d20
        3u groff
        5u
        3u gron3
        d20
        3u groff
       (p13:sp13 ph12):f1
       (p14:sp14 ph0):f2     ;hsec 13C dec
        3u gron3
        d20
        3u groff
        5u
        3u gron5
        d20 pl1:f1
        3u groff
       (center (p1 ph0):f1 (p7 ph5):f3)
        p23:gp23
        d23
       (center (p1*2 ph0):f1 (p7*2 ph0):f3)
```

```
      p23:gp23
      d23
     (center (p1 ph1):f1 (p7 ph1):f3)
      p24:gp24
      d24
     (center (p1*2 ph0):f1 (p7*2 ph6):f3)
      p24:gp24
      d24
     (p1 ph0):f1
      d25
     (p1*2 ph0):f1
      5u pl17:f2
     (p9 ph0):f2
      5u
     (p19:sp11 ph0):f2
      p25:gp25

999    100u pl30:f3
      5u pl19:f2
      5u BLKGRAD;setnmr3^0 setnmr0^34^32^33 ctrlgrad 7
      go=2 ph31 cpd3:f3
      10u do:f3 do:f2
      1m
      d11 wr #0 if #0 zd



      d12  iu1
      lo to 4 times 2
      d12  ru1
      d12*0.5 ip5*2
      d12*0.5  iu2
      lo to 5 times 2
      d12*0.5 ru2
      d12*0.5 id20
      d12 ip7*2
      d12 ip31*2
      lo to 6 times l4



1m
1m do:f2
1m do:f3
1m ;setnmr3^0
1m ;RESET
```

```
1m
exit

ph0=0
ph1=1
ph2=2
ph3=3
ph4=0
ph5=0
ph6=0
ph7=0
ph8=1
ph9=0 1 2 3
ph10=2
ph11=0 0
ph12=2 2
ph31=0 2 0 2
```

10.5 Intensity modulated HSQC for measuring $^1J_{NH}$

```
;hsqc15N.new
;D. Lee, Nov. 2002

;15N-1H HSQC correlations without water saturation
;The delay for 3-9-19 watergate (d5) should be matched
;with 1/d;d=distance of next null point (in Hz).


;S. Mori et al, JMR B108, 94-98 (1995)

;pl1   : power for 1H
;pl2   : power for 13C
;pl3   : power for 15N
;pl13  : power for 15N waltz16 decoupling

;p1    : 90 degree hard pulse 1H
;p3    : 90 degree hard pulse 13C
;p4    : 180 degree hard 13C pulse (225d for 5/600)
;p5    : 90 degree hard pulse 15N
;pcpd3 : 90 deg cpd-pulse15N(waltz16,160u)

;d1    : relaxation delay
;d2    : INEPT delay (~2.7m)
;d5    : delay for 3-9-19=1/(Hz between nulls)
;in0   : 1/(2 SW) (Hz)

;p21   : 500u (Gradient in first INEPT)
;p22   : 500u (Gradient for z-filter)
;p23   : 1m (Gradient for second INEPT)
;gpz1  : 19%
;gpz2  : 30%
;gpz3  : 65%


#include <Avance_dl.incl>

define delay INEPT_W
define delay INEPT_D
define delay Tau

#define GRADIENT1  10u p21:gp1 200u
#define GRADIENT2  10u p22:gp2 200u
```

```
#define GRADIENT3  10u p23:gp3 200u
#define GRADIENT4  10u p22:gp4 200u
#define GRADIENT5  10u p22:gp5 200u

"p2=2*p1"
"p6=2*p5"

"in0=inf1/2"


"d0=in0/2-p1"
"d3=d5/2-p5"
"INEPT_D=d2-p21-210u"
"INEPT_W=d2-(p23+210u+p1*2.3846+d5*2.5)"

"spoffs1=bf1*cnst1/1000000-o1"

1  10u ze
2  1m do:f3
   d1 pl1:f1
   20u pl3:f3
   20u LOCKH_ON
   20u
     "Tau=vd-(0.5*p12-p5)-210u-p22"
;----------------------------------------first INEPT
   (p1 ph20):f1
   GRADIENT1
   INEPT_D
   (center(p2 ph21):f1 (p6 ph1):f3)
   GRADIENT1
   INEPT_D
   ;(p1 ph21):f1
  (p1 ph11):f1
   GRADIENT2
;----------------------------------------15N evolution

(p5 ph12):f3

 GRADIENT4
 Tau   pl0:f1
 (center (p12:sp1 ph13):f1 (p6 ph13):f3)
  GRADIENT4
  Tau pl1:f1
  (d0 p2 ph20 d0):f1
   (p5 ph20):f3
   GRADIENT5
```

```
;----------------------------------------second INEPT
  (p1 ph22):f1
  GRADIENT3
  INEPT_W
  (p1*0.2308 ph21 d5 p1*0.6923 ph21 d5 p1*1.4615 ph21):f1
  (d3 p6 ph1 d3):f3
  (p1*1.4615 ph23 d5 p1*0.6923 ph23 d5 p1*0.2308 ph23):f1
  GRADIENT3
  INEPT_W pl13:f3 LOCKH_OFF
;---------------------------------------acquisition
  go=2 ph31 cpd3:f3
  1m do:f3 mc #0 to 2
 F1I(ivd, 15)
 F1PH(ip12&ip13,id0)
10u do:f1

10u do:f3
10u LOCKH_OFF
exit

ph1 =2 0

ph31=0 2 2 0 0 2 2 0 2 0 0 2 2 0 0 2

ph11=1 1 1 1 1 1 1 1 3 3 3 3 3 3 3 3
ph12=0 2
ph13=0 0 1 1 2 2 3 3

ph20=0
ph21=1
ph22=2
ph23=3
```

10.6 Python scripts for fitting $^{1}J_{NH}$ values

```
import numpy as np
import os
import pylab as plt
import scipy.optimize as opt

def shiftlist(proton,nitrogen):

A=np.genfromtxt('DeUbi400bax1.list',skip_header=2,dtype=[('assignment','|S10'),('w1','<
f8'),('w2','<f8'),('DataHeight','<f8'),('SN','<f8')])
    B=np.zeros(len(A),dtype=[('assignment', '|S10'), ('w1', '<f8'), ('w2', '<f8')])
    B['assignment']=A['assignment']
    B['w1']=A['w1']+nitrogen
    B['w2']=A['w2']+proton
    np.savetxt('DeUbi400bax2.list',B, fmt='    %10s    %1.3f    %1.3f    ', newline='\n')
    return B


def changeto3D():

A=np.genfromtxt('Tau2D.list',skip_header=1,dtype=[('assignment','|S10'),('w1','<f8'),('w2
','<f8')])
    B=np.zeros(15*len(A),dtype=[('assignment', '|S10'), ('w1', '<f8'), ('w2', '<f8'),('w3',
'<f8')])
    for i in range(15):
        s='-'+str(i+1)+'-'

B['assignment'][i*len(A):(i+1)*len(A)]=np.core.defchararray.replace(A['assignment'][0:l
en(A)] ,'-',s)
        B['w2'][i*len(A):(i+1)*len(A)]=A['w1'][0:len(A)]
        B['w3'][i*len(A):(i+1)*len(A)]=A['w2'][0:len(A)]
        B['w1'][i*len(A):(i+1)*len(A)]=i+1
    np.savetxt('3D.list',B, fmt='    %10s    %1.3f    %1.3f    %1.3f', newline='\n')
    return B,A

def changetoMatrix():

#C=np.genfromtxt('SynEGCG10Im2.list',skip_header=1,dtype=[('assignment','|S10'),('w1
','<f8'),('w2','<f8'),('w3','<f8'),('Height','<f8'),('SN','<f8')])

C=np.genfromtxt('Im.list',skip_header=2,dtype=[('assignment','|S10'),('w1','<f8'),('w2','<f
8'),('w3','<f8'),('Height','<f8')])
    matrix=np.zeros((15,int(len(C)/15)))
```

```
    for i in range(len(C)/15):
      buff=np.zeros(15)
      for j in range(15):
        buff[int(C['w1'][i*15+j])-1]=C['Height'][i*15+j]# change between w1 and w2 based on table structure.
      matrix[:,i]=buff[:]
    np.savetxt('matrix.out', matrix, delimiter='      ')
    return matrix

f=lambda x,A0,A1,A2,j:(A0*np.cos(np.pi*j*2*x/1000)+A1)*np.exp(-A2*2*x/1000)

def  fitting():
 matrix=np.genfromtxt('matrix.out')

A=np.genfromtxt('Tau2D.list',skip_header=1,dtype=[('assignment','|S10'),('w1','<f8'),('w2','<f8')])

#A=np.genfromtxt('synEGCG.list',skip_header=1,dtype=[('assignment','|S10'),('w1','<f8'),('w2','<f8'),('w1Hz','<f8'),('w2Hz','<f8')])

mixtime=np.array([31.9,31.1,30.7,30.3,29.9,26.5,25.8,25.1,24.6,24.3,21.2,20.5,19.8,19.5,19.1])
 x=int(matrix.shape[1])
 result=np.zeros(x,dtype=[('assignment','|S10'),('J','<f8'),('error','<f8')])

resultfull=np.zeros(x,dtype=[('assignment','|S10'),('A0','<f8'),('A1','<f8'),('A2','<f8'),('j','<f8')])
 for i in range(x):
      y=matrix[:,i]/matrix[:,i].max()
      #y=matrix[:,i]/np.abs(matrix[:,i]).min()
      popt,pcov=opt.curve_fit(f,mixtime,y,p0=[1,0.1,10,100])
      result['assignment'][i]=A['assignment'][i]
      result['J'][i]=popt[-1]
      result['error'][i]=pcov[-1,-1]
      resultfull['assignment'][i]=result['assignment'][i]
      resultfull['A0'][i]=popt[0]
      resultfull['A1'][i]=popt[1]
      resultfull['A2'][i]=popt[2]
      resultfull['j'][i]=popt[3]
 np.savetxt('result.out',result, fmt='   %10s   %1.3f   %1.6f', newline='\n')
 np.savetxt('resultfull.out',resultfull,fmt='     %10s     %1.6f     %1.6f    %1.6f     %1.6f',newline='\n')
 return result,resultfull

def  singleplot(resname):
  matrix=np.genfromtxt('matrix.out')
```

```
mixtime=np.array([31.9,31.1,30.7,30.3,29.9,26.5,25.5,25.1,24.6,24.3,21.2,20.3,19.8,19.5,
19.1])
  dataseries1=np.zeros((15,2))

A=np.genfromtxt('resultfull.out',skip_header=0,dtype=[('assignment','|S10'),('A0','<f8'),('
A1','<f8'),('A2','<f8'),('j','<f8')])
  idx=A['assignment']==resname
  dataseries=matrix[:,idx]/matrix[:,idx].max()
  dataseries1[:,0]=mixtime
  dataseries1[:,1]=np.transpose(matrix[:,idx])
  np.savetxt('datatrace.out',dataseries1)
  parameter=A[idx]
  x=np.linspace(mixtime[-1],mixtime[0],num=500)
  y=f(x,parameter['A0'],parameter['A1'],parameter['A2'],parameter['j'])
  plt.scatter(mixtime,dataseries)
  p1=plt.plot(x,y)
  p2=plt.scatter(mixtime,dataseries)
  plt.suptitle(resname,fontsize=12)
  plt.show()
  return int(matrix.shape[1])==int(A.shape[0])
```

10.7 RDC values and Chemical shifts of native ASF/SF2 (200-219)

res number	res name	RDC HN(Hz)	error RDC HN(Hz)	RDC HαCα(Hz)	error RDC HαCα (Hz)	RDC CαCo(Hz)	error RDC CαCo(Hz)	CS Cα (ppm)	CS CO (ppm)
1	*G*	-		-		-		-	-
2	A	-		-		-		-	178.71
3	M	-		-		-		55.23	176.45
4	G	-		-		-		44.61	-
5	P	-		-		-		-	-
6	S	3.366	0.3	10.384	1.4	-2.156	0.8	58.38	174.48
7	Y	-5.764	0.3	0.671	1.4	-1.386	0.8	58.3	176.46
8	G	-4.29	0.3	4.664	1.4	0.407	0.8	45.42	174.24
9	R	-1.804	0.3	3.861	1.4	-0.737	0.8	56.25	176.62
10	S	4.774	0.3	5.456	1.4	-0.891	0.8	58.62	174.89
11,13,15,17,19,21	R	3.762	0.3	16.104	1.4	-0.77	0.8	56.3	176.62
12,14,16,18,20	S	4.774	0.3	16.654	1.4	-1.056	0.8	58.62	174.89
21	R	3.762	0.3	16.104	1.4	-0.77	0.8	56.3	176.62
22	S	4.774	0.3	16.654	1.4	-0.847	0.8	58.62	174.44
23	R	4.598	0.3	10.857	1.4	-0.363	0.8	56.13	175.5
24	S	2.618	0.3	8.8803	1.4	-		60.06	-

10.8 3bond couplings of native ASF/SF2 (200-219)

res number	res name	$^{3}J_{NH-H\alpha}$ (Hz)	error $^{3}J_{NH-H\alpha}$ (Hz)	$^{3}J_{NC\gamma}$ (Hz)	error $^{3}J_{NC\gamma}$ (Hz)	$^{3}J_{COC\gamma}$ (Hz)	error $^{3}J_{COC\gamma}$ (Hz)	$^{1}J_{H\alpha C\alpha}$ (Hz)	error $^{1}J_{H\alpha C\alpha}$ (Hz)	$^{1}J_{C\alpha C\beta}$ (Hz)	error $^{1}J_{C\alpha C\beta}$ (Hz)
1	G	-				-		-		-	
2	A	-				-		144.31		34.82	0.48
3	M	7.74	0.25			-		143.20	0.35	-	
4	G	-				-		-		-	
5	P	-				-		148.93	0.72	31.32	0.76
6	S	-				-		142.50	0.42	39.42	0.54
7	Y	6.84	0.41			-		143.30	0.47	35.29	0.30
8	G	-				-		141.12	0.74	-	
9	R	7.42	0.20	0.97	0.07	-		142.69	0.61	34.01	0.64
10	S	6.70	0.19			-		143.26	0.60	39.26	0.64
11	R	7.51	0.40	0.84	0.08	-		142.69	0.61	34.01	0.64
12	S	6.70	0.19			-		143.26	0.60	39.26	0.64
13,15,17,19,21	R	7.51	0.25	0.85	0.08	2.39	0.30	142.69	0.61	34.01	0.64
14,16,18,20,22	S	6.70	0.19			-		143.26	0.60	39.26	0.64
22	S	6.70	0.19			-		142.83	0.60	39.05	0.64
23	R	7.00	0.25	0.85	0.08	2.46	0.30	142.10	0.31	36.07	0.64
24	S	7.94	0.51			-		-		38.94	0.24

10.9 RDC values of phosphorylated ASF/SF2 (200-219)

res number	res name	RDC HN(Hz)	error RDC HN(Hz)	RDC HC(Hz)	error RDC HC(Hz)	RDC CαCo(Hz)	error RDC CαCo(Hz)	CS Cα(ppm)	CS Co(ppm)
1	G	-		-		-		-	-
2	A	-		-		-		-	-
3	M	-		-		-		55.23	176.42
4	G	-		-		-		44.6	-
5	P	-		-		-		-	-
6	S	3.24	0.3	8.24	1.4	-0.86	0.8	58.33	174.35
7	Y	-2.82	0.3	2.12	1.4	0.16	0.8	58.22	176.33
8	G	-1.8	0.3	3.39	1.4	0.31	0.8	45.25	173.79
9	R	1.04	0.3	3.87	1.4	-0.21	0.8	55.9	176.36
10	S	0.7	0.3	5.85	1.4	-0.69	0.8	58.35	174.7
11	R	-0.18	0.3	-1.61	1.4	0.24	0.8	56.24	176.59
12	S	-1.3	0.3	5.75	1.4	-0.61	0.8	58.35	174.38
13	R	3.76	0.3	6.12	1.4	-0.35	0.8	55.72	176.14
14	S	3.4	0.3	7.87	1.4	-1.25	0.8	58.2	174.19
15	R	7.28	0.3	10.05	1.4	-1.71	0.8	55.73	176.13
16	S	7.62	0.3	11.71	1.4	-1.5	0.8	58.18	174.16
17	R	9.24	0.3	16.86	1.4	-1.71	0.8	55.76	176.13
18	S	9.16	0.3	11.71	1.4	-1.01	0.8	58.18	174.2
19	R	7.88	0.3	17.1	1.4	-1.33	0.8	55.79	176.3
20	S	7.56	0.3	11.73	1.4	-1.19	0.8	58.18	174.3
21	R	5.62	0.3	11.55	1.4	-1.5	0.8	55.91	176.17
22	S	8.04	0.3	9.97	1.4	0.26	0.8	58.19	174.36
23	R	4.38	0.3	7.07	1.4	-1.63	0.8	55.92	175.44
24	S	2.22	0.3	7.98	1.4	-		59.93	-

10.10 3bond couplings of phosphorylated ASF/SF2 (200-219)

res number	res name	$^{3}J_{NH-H\alpha}$ (Hz)	error $^{3}J_{NH-H\alpha}$ (Hz)	$^{3}J_{NC\gamma}$ (Hz)	error $^{3}J_{NC\gamma}$ (Hz)	$^{3}J_{CoC\gamma}$ (Hz)	error $3J_{CoC\gamma}$ (Hz)	$^{1}J_{H\alpha C\alpha}$ (Hz)	error $^{1}J_{H\alpha C\alpha}$ (Hz)	$^{1}J_{CaCb}$ (Hz)	error $^{1}J_{CaCb}$ (Hz)
1	G	-				-		-		-	
2	A	-				-		144.84	0.60	-	
3	M	7.61	0.36			-		141.61	0.60	34.54	1.30
4	G	-				-		140.82	0.60	-	
5	P	-				-		148.24	0.60	-	
6	S	-				-		142.32	0.60	39.58	1.45
7	Y	7.10	0.33			-		143.44	0.60	35.12	1.20
8	G	-				-		141.05	0.60	-	
9	R	7.22	0.45	1.19	0.19	2.45	0.20	142.27	0.60	34.71	1.20
10	S	6.68	0.53			-		142.38	0.60	38.84	1.45
11	R	7.41	0.23	0.98	0.14	2.17	0.20	141.80	0.60	35.07	1.20
12	S	6.18	0.26			-		143.76	0.60	38.91	1.45
13	R	6.97	0.23	0.77	0.18	2.53	0.20	142.27	0.60	34.75	1.20
14	S	5.77	0.23			-		143.47	0.60	37.90	1.45
15	R	6.61	0.23	0.77	0.16	1.89	0.20	143.14	0.60	35.34	1.20
16	S	5.58	0.22			-		143.41	0.60	39.87	1.45
17	R	7.09	0.26	0.76	0.17	1.89	0.20	140.87	0.60	33.62	1.20
18	S	5.58	0.22			-		143.56	0.60	38.97	1.45
19	R	6.83	0.24	0.79	0.17	2.42	0.20	142.41	0.60	34.32	1.20
20	S	5.51	0.23			-		143.61	0.60	39.25	1.45
21	R	6.67	0.25	0.68	0.17	2.17	0.20	142.15	0.60	34.27	1.20
22	S	6.18	0.26			-		144.53	0.60	39.03	1.45
23	R	6.44	0.32	0.63	0.15	2.56	0.20	141.07	0.60	36.07	1.20
24	S	7.54	0.65			-		141.81	0.60	38.51	0.76

10.11 RDC values for H-CO and N-CO from 2 measurements on the same phosphorylated ASF/SF2 (200-219) samples.

Res number	D_{N-CO}(Hz)		D_{HN-CO}(Hz)	
	1st	2nd	1st	2nd
3	-0.56	0.43	0.84	1.66
4				
6	-0.19	0.6	4.34	0.94
7	-0.72	0.51	1.13	1.4
8	-0.61	-1.65	0.26	0.07
9	-0.35	-0.28	-2.24	-0.72
10	0.52	0.12	0.28	0.83
11	-0.18	-0.39	-0.05	0.55
12	0.25	-0.41	2.28	1.9
13	-0.4	-0.35	0.99	0.8
14	-0.14	0.11	0.6	2.25
15	-0.71	-0.57	2.57	2.33
16	0.4	-0.21	2.75	2.83
17	-0.57	-0.37	1.4	2.65
18	0.48	0.42	3.53	5.13
19	-0.28	-0.19	2.72	2.51
20	0.66	-0.41	4.04	1.9
21	-0.21	0.33	2.17	3.17
22	0.27	-0.22	5.12	2.53
23	-0.18	0.39	0.98	1.71
24	-0.09	0.03	1.59	2.42

10.12 $^{1}J_{NH}$ values of Tau protein pH 6.0, 278K, at 900MHz

Residue	$^{1}J_{NH}$(Hz)	Residue	$^{1}J_{NH}$(Hz)	Residue	$^{1}J_{NH}$(Hz)
E3	-93.23	E53	-92.86	A103	-93.38
R5	-93.10	D54	-93.11	E104	-92.99
Q6	-92.92	G55	-94.33	E105	-93.06
E7	-92.73	S56	-93.26	A106	-93.24
E9	-92.95	E57	-92.96	G107	-94.25
V10	-92.89	E58	-93.10	I108	-92.78
M11	-93.20	G60	-94.11	G109	-94.44
E12	-92.89	S61	-93.15	D110	-93.08
D13	-93.26	E62	-93.01	T111	-93.37
H14	-93.09	T63	-93.19	S113	-93.17
G16	-94.26	S64	-93.28	L114	-93.03
T17	-92.85	D65	-93.34	E115	-93.03
Y18	-93.25	A66	-93.51	D116	-93.14
G19	-94.14	K67	-93.06	E117	-93.10
L20	-92.80	S68	-93.26	A118	-93.39
G21	-94.34	T69	-93.45	A119	-93.26
D22	-93.15	T71	-93.17	G120	-94.28
R23	-93.19	A72	-93.35	H121	-93.16
K24	-92.95	E73	-92.88	V122	-92.95
D25	-93.12	D74	-93.21	T123	-93.39
Q26	-92.78	V75	-92.83	Q124	-93.07
G27	-94.24	T76	-93.31	A125	-93.25
G28	-94.18	A77	-93.71	R126	-93.05
Y29	-92.92	L79	-93.05	M127	-93.06
T30	-93.20	V80	-92.90	V128	-92.90
M31	-93.26	D81	-93.41	S129	-93.51
Q33	-93.02	E82	-92.98	K130	-93.03
D34	-93.15	G83	-94.21	S131	-93.45
Q35	-92.88	A84	-93.57	K132	-92.83
E36	-92.95	G86	-94.25	D133	-93.19
G37	-94.24	K87	-92.99	G134	-94.40
D38	-93.18	Q88	-93.20	T135	-92.96
T39	-93.17	A89	-93.47	G136	-94.43
D40	-93.37	A90	-93.37	S137	-93.17
A41	-93.29	A91	-93.30	D139	-93.33
G42	-94.43	Q92	-93.20	K140	-93.16
L43	-93.05	H94	-93.08	K141	-93.15
K44	-92.88	T95	-93.25	A142	-93.40
E45	-93.10	E96	-93.02	K143	-92.94
S46	-93.49	I97	-93.28	G144	-94.38
L48	-93.10	E99	-92.96	A145	-93.14
Q49	-92.98	G100	-94.17	G147	-94.47
T50	-93.38	T101	-93.03	K148	-93.03
T52	-93.20	T102	-93.35	T149	-93.28

Residue	$^1J_{NH}$(Hz)	Residue	$^1J_{NH}$(Hz)	Residue	$^1J_{NH}$(Hz)
K150	-93.23	T212	-93.17	Q269	-93.03
I151	-93.09	S214	-93.17	G271	-94.03
A152	-93.58	L215	-93.01	G273	-94.19
T153	-93.41	T217	-93.42	K274	-93.05
G156	-94.30	T220	-93.12	V275	-93.09
A157	-93.25	R221	-93.00	Q276	-93.28
A158	-93.60	E222	-93.32	I277	-93.06
G161	-94.25	K224	-93.10	I278	-93.17
Q162	-92.98	K225	-93.09	N279	-93.42
K163	-93.13	V226	-93.11	K280	-93.18
G164	-94.40	A227	-93.62	K281	-93.06
E165	-92.99	V228	-92.90	L282	-93.16
A166	-93.48	V229	-93.13	D283	-93.01
N167	-93.29	R230	-93.26	L284	-92.89
A168	-93.36	K234	-93.09	S285	-93.22
T169	-93.22	S235	-93.56	N286	-93.17
I171	-93.28	S237	-93.27	V287	-92.96
A173	-93.33	S238	-93.41	Q288	-93.36
K174	-92.88	A239	-93.53	S289	-93.31
A178	-93.61	K240	-93.04	K290	-93.03
K180	-92.92	S241	-93.39	C291	-92.88
T181	-93.32	R242	-93.07	G292	-94.37
S184	-93.17	L243	-93.28	S293	-93.17
S185	-93.18	Q244	-92.98	K294	-93.13
G186	-94.30	T245	-93.14	D295	-93.21
E187	-93.27	A246	-93.71	N296	-93.19
S191	-93.25	V248	-93.04	I297	-92.94
G192	-94.21	M250	-93.18	K298	-93.28
D193	-93.23	D252	-93.05	H299	-93.09
R194	-92.96	L253	-93.10	V300	-93.28
S195	-93.25	K254	-93.13	S305	-93.11
G196	-94.25	N255	-93.27	V306	-92.90
Y197	-93.03	V256	-92.94	Q307	-92.77
S198	-93.30	K257	-93.30	V313	-92.99
S199	-93.59	S258	-93.36	D314	-93.32
G201	-94.21	K259	-93.03	L315	-93.39
S202	-93.53	I260	-93.13	K317	-92.89
G204	-94.29	G261	-94.42	V318	-93.01
T205	-93.39	S262	-93.17	T319	-92.92
G207	-94.22	T263	-93.20	S320	-93.54
S208	-93.33	E264	-93.09	K321	-93.08
R209	-93.06	N265	-93.27	C322	-92.88
S210	-93.33	L266	-93.22	S324	-93.17
R211	-93.26	K267	-93.14	L325	-93.01

Residue	$^1J_{NH}$(Hz)
G326	-94.22
N327	-93.33
I328	-92.92
H329	-93.05
H330	-93.32
K331	-93.41
G333	-93.95
Q336	-92.98
V337	-92.95
K338	-93.24
V339	-93.01
K340	-93.33
S341	-93.27
E342	-92.89
K343	-92.99
L344	-93.29
D345	-93.07
F346	-93.14
K347	-93.27
D348	-93.16
R349	-93.03
V350	-93.02
Q351	-93.15
S352	-93.35
K353	-92.96
I354	-93.11
G355	-94.34
S356	-93.18
L357	-93.18
D358	-93.13
I360	-92.92
H362	-93.37
V363	-93.19
G365	-94.00
N368	-93.27
K369	-93.30
K370	-93.26
I371	-93.03
E372	-93.20
T373	-93.17
H374	-93.26
K375	-93.04
L376	-93.15
T377	-93.16

Residue	$^1J_{NH}$(Hz)
F378	-93.04
R379	-93.24
E380	-93.06
N381	-93.35
A382	-93.36
K383	-93.18
A384	-93.53
K385	-93.00
T386	-93.06
D387	-93.25
H388	-93.26
G389	-94.49
A390	-93.16
E391	-93.05
I392	-92.96
V393	-93.13
Y394	-93.36
K395	-93.06
V398	-92.90
V399	-93.34
S400	-93.46
G401	-94.20
D402	-93.30
T403	-93.12
S404	-93.71
R406	-93.10
H407	-93.12
L408	-93.05
S409	-93.23
N410	-93.29
V411	-92.95
S412	-93.48
S413	-93.26
T414	-93.15
G415	-94.38
S416	-93.26
I417	-92.89
D418	-93.38
M419	-93.05
V420	-92.95
D421	-93.26
S422	-93.34
Q424	-92.92
L425	-93.05

Residue	$^1J_{NH}$(Hz)
A426	-93.34
T427	-93.01
L428	-93.12
A429	-93.41
D430	-93.29
E431	-92.95
V432	-92.97
S433	-93.54
A434	-93.54
S435	-93.35
L436	-93.28
A437	-93.55
K438	-93.10
Q439	-93.33
G440	-94.19

10.13 $^{1}J_{NH}$ values of α-synclein at different pH and temperatures, measured at 700MHz. The missing data due to signal overlapping and broadening are labeled as "N.A."

Residues	$^{1}J_{NH}$(Hz)					
	278K			288K		298K
	pH 5.7	pH 6.5	pH 7.4	pH 5.7	pH 7.4	pH 5.7
V3	-92.9	-93.0	-93.0	-92.9	-93.0	-92.9
F4	-93.2	-93.5	-93.9	-93.5	-93.8	-93.4
M5	-93.5	-93.7	-93.6	-93.5	N.A.	-93.8
G7	-94.1	-94.3	N.A.	-94.3	N.A.	-94.1
L8	-93.3	-93.3	N.A.	-93.4	N.A.	-93.2
S9	-93.6	-93.6	N.A.	-93.6	N.A.	-93.6
K12	-93.3	-93.2	-93.2	-93.4	N.A.	-93.3
E13	-93.1	-93.4	N.A.	-93.3	N.A.	-93.2
G14	-94.3	-94.4	-94.2	-94.2	-94.2	-94.2
V15	-93.0	-92.9	-93.2	-93.0	-93.4	-93.0
A17	-94.1	-93.6	-94.1	-94.0	-94.3	-94.0
A18	-93.7	-93.6	-94.1	-93.6	-93.9	-93.7
A19	-93.8	-93.9	-94.2	-93.8	-94.0	-93.7
E20	-93.6	-93.6	-93.6	-93.4	N.A.	-93.3
K21	-93.1	-93.2	-93.1	-93.5	-93.3	-93.4
T22	-93.4	-93.4	-93.3	-93.4	-93.3	-93.4
G25	-94.3	-94.3	-94.3	-94.2	N.A.	-94.2
V26	-93.4	-93.2	-93.2	-93.2	-93.1	-93.2
A27	-94.1	N.A.	N.A.	-94.0	-94.2	-94.0
E28	-93.6	-93.6	-93.7	-93.2	-93.8	-93.5
A29	-94.1	-93.8	-93.9	N.A.	-94.1	-93.3
G31	-94.3	-94.3	-94.3	-94.3	-94.3	-94.3
K32	-93.4	-93.5	N.A.	-93.3	N.A.	-93.4
T33	-93.6	-93.5	-93.5	-93.4	N.A.	-93.4
V37	-93.0	-93.2	-93.1	-93.0	-93.0	-92.9
L38	-93.8	-93.5	-93.9	-93.7	-93.8	-93.7
Y39	-93.5	-93.3	N.A.	-93.4	-93.4	-93.3
V40	-93.1	-93.1	-93.1	-93.1	-93.1	-93.0
G41	-94.4	-94.5	-94.6	-94.4	N.A.	-94.4
T44	-93.5	-93.5	-93.5	-93.5	-93.4	-93.5
V48	-93.0	-93.0	-93.1	-92.9	-93.0	-92.9
V49	N.A.	-93.1	-93.1	-93.1	-93.1	-93.2
H50	-93.6	N.A.	-93.8	-93.6	N.A.	-93.9
G51	-94.3	-94.2	-94.5	-94.4	N.A.	-94.3
V52	-93.1	N.A.	-93.1	-93.0	-93.0	N.A.
A53	-94.0	-94.1	-94.3	-94.0	N.A.	-94.0

V55	-93.3	-93.3	-93.3	-93.1	-93.2	N.A.
A56	N.A.	-94.1	N.A.	-93.9	N.A.	-93.8
E57	-93.4	-93.4	-93.6	-93.4	-93.6	-93.2
K58	-93.5	-93.5	-93.8	-93.4	-93.7	-93.4
T59	-93.5	-93.5	-93.6	-93.5	-93.3	-93.3
K60	-93.5	-93.4	-93.6	-93.7	N.A.	N.A.
E61	-93.4	-93.8	-93.6	-93.5	-93.7	N.A.
Q62	-93.6	-93.5	-93.5	N.A.	-93.8	-93.4
V63	-93.2	N.A.	N.A.	-93.1	-93.2	-93.0
T64	-93.5	-93.5	N.A.	-93.4	-93.3	-93.4
V66	-93.1	-93.2	-93.2	-93.1	-93.3	-93.0
G67	-94.4	-94.4	-94.5	-94.4	-94.6	-94.4
G68	-94.3	-94.3	-94.3	-94.3	-94.5	-94.3
A69	-93.9	-93.8	-94.1	-93.9	N.A.	-93.8
V70	-93.1	-93.1	-93.1	-93.1	-93.0	-93.0
V71	-93.3	-93.2	-93.3	-93.2	-93.2	-93.1
T72	-93.6	-93.4	-93.6	-93.5	-93.5	-93.5
G73	-94.3	-94.2	-94.2	-94.2	N.A.	-94.1
V74	-93.1	-93.0	-93.1	-92.9	-93.1	-92.8
T75	-93.6	-93.6	-93.5	-93.5	-93.7	-93.5
A76	-94.0	-94.0	-94.1	-93.8	N.A.	-93.9
V77	-93.2	-93.2	-93.2	-93.0	-93.1	-93.0
A78	-94.3	-94.3	-94.4	-94.1	-94.3	-94.0
Q79	-93.3	-93.4	-93.4	-93.4	N.A.	-93.5
K80	-93.6	-93.6	N.A.	-93.4	N.A.	-93.5
T81	-93.5	-93.5	-93.5	-93.4	-93.4	-93.3
V82	-93.2	-93.2	-93.2	-93.1	-93.2	-93.2
E83	-93.6	N.A.	-93.7	-93.6	-93.7	-93.5
G84	-94.3	-94.3	-94.3	-94.3	N.A.	-94.3
A85	-93.7	-93.9	-93.7	-93.6	N.A.	-93.9
G86	-94.4	-94.6	-94.5	-94.4	N.A.	-94.4
S87	-93.5	-93.5	-93.6	-93.5	N.A.	-93.4
I88	-93.3	-93.3	-93.2	-93.2	-93.3	-93.1
A89	-94.1	-94.0	N.A.	-94.0	N.A.	-93.9
A90	-94.0	-94.0	-94.1	-93.8	-93.9	-93.7
A91	-93.9	-94.0	N.A.	-93.7	N.A.	-93.9
T92	-93.2	-93.0	-93.2	-93.1	-93.4	-93.1
G93	-94.2	-94.4	-94.0	-94.2	N.A.	-94.3
F94	-93.4	N.A.	-93.4	-93.4	-93.4	-93.4
V95	-93.3	-93.2	-93.3	-93.2	-93.2	-93.1
K96	-93.9	-94.0	-93.9	-93.9	-94.0	-93.7
K97	N.A.	-93.1	-93.1	N.A.	-93.1	N.A.
D98	-93.7	-93.5	-93.7	-93.5	-93.7	-93.6
Q99	-93.2	-93.1	-93.5	-93.2	N.A.	N.A.

L100	-93.6	-93.6	-93.7	-93.5	-93.7	-93.4
G101	-94.3	-94.3	-94.3	-94.3	-94.2	-94.3
K102	-93.2	-93.5	N.A.	-93.2	-93.4	-93.3
N103	-93.6	-93.8	-93.9	-93.6	N.A.	-93.7
E104	-93.3	-93.4	-93.3	-93.3	-93.4	-93.2
E105	-93.1	-93.6	-93.4	-93.4	-93.6	N.A.
G106	-94.3	-94.2	-94.3	-94.3	-94.3	-94.2
A107	-94.1	-93.8	-94.1	-93.9	-94.4	-93.9
Q109	-93.2	-93.2	-93.3	-93.2	-93.3	-93.2
E110	-93.3	N.A.	-93.4	-93.2	-93.5	-93.2
G111	-94.2	-94.2	-94.3	-94.2	-94.2	-94.2
I112	-93.1	-93.1	-93.0	-93.0	-93.0	-93.0
E114	-93.1	N.A.	N.A.	-93.6	-93.3	-93.3
D115	-93.5	-93.3	-93.6	-93.4	-93.7	-93.4
M116	-93.8	-93.5	N.A.	-93.5	-93.7	-93.7
V118	-93.2	-93.2	-93.1	-93.0	-92.9	-92.9
D119	-93.9	-93.8	-94.0	-93.8	-93.9	-93.8
D121	-93.5	N.A.	-93.5	-93.5	-93.5	-93.5
N122	-93.6	-93.6	-93.5	-93.5	-93.7	-93.5
E123	-93.5	-93.6	-93.5	-93.4	-93.7	-93.4
A124	-93.8	-93.7	-93.8	N.A.	-93.7	-93.6
Y125	-93.1	-93.3	-93.4	-93.0	-93.4	-92.9
E126	-93.6	-93.5	-93.5	-93.4	-93.6	-93.4
S129	-93.4	-93.2	-93.4	-93.3	-93.3	-93.3
E130	-93.3	-93.4	-93.3	-93.3	-93.4	-93.3
G132	-94.3	-94.3	-94.3	-94.2	N.A.	-94.1
Y133	-93.4	N.A.	-93.4	-93.3	-93.4	-93.2
Q134	-93.9	N.A.	N.A.	-93.5	-93.7	N.A.
D135	-93.6	-93.6	-93.5	-93.6	-93.5	N.A.
Y136	-93.2	-93.3	-93.3	-93.2	-93.2	-93.2
E137	-93.8	-93.7	-93.8	-93.7	-93.8	-93.7
E139	-93.2	-93.1	-93.3	-93.1	-93.3	-93.0

10.14 $^{1}J_{NH}$ values of native ubiquitin at 400MHz, 600MHz and 900MHz field strength, pH 7.0, 298K. I36 was not included because its amid proton chemical shift is close to the edge of the selective pulse excitation profile. T9 and G75 were excluded because of low sensitivity.

Residue	$^{1}J_{NH}$(Hz) 400M	$^{1}J_{NH}$(Hz) 600M	$^{1}J_{NH}$(Hz) 900M
Q2	-92.97	-92.99	-92.90
I3	-92.65	-92.64	-92.48
F4	-93.37	-93.22	-92.95
V5	-92.95	-92.71	-92.67
K6	-93.37	-93.43	-93.01
T7	-94.2	-94.09	-93.76
L8	-92.38	-92.48	-92.61
T9	N.A.	N.A.	N.A.
G10	-93.64	-93.49	-93.40
K11	-93.52	-93.43	-93.41
T12	-92.68	-92.49	-92.26
I13	-93.2	-93.03	-92.88
T14	-92.84	-92.76	-92.65
L15	-93.83	-93.72	-93.58
E16	-92.02	-91.93	-91.86
V17	-93.47	-93.41	-93.28
E18	-92.15	-92.07	-91.85
S20	91.95	91.95	-92.00
D21	-93.96	-93.87	-93.78
T22	-93.43	-93.38	-93.36
I23	-93.96	-93.87	-93.95
N25	-94.22	-94.11	-93.84
V26	-93.76	-93.67	-93.45
K27	-93.66	-93.55	-93.19
A28	-94.03	-93.85	-93.68
K29	-93.67	-93.61	-93.31
I30	-93.79	-93.68	-93.31
Q31	-93.9	-93.83	-93.90
D32	-93.9	-93.98	-93.55
K33	-92.87	-92.74	-92.54
E34	-90.54	-90.46	-90.14
G35	-93.98	-94.00	-94.03
I36	N.A.	N.A.	N.A.
D39	-93.4	-93.32	-93.16

Q40	-92.15	-92.15	-92.06
Q41	-93.22	-93.27	-93.44
R42	-93.35	-93.34	-93.57
L43	-92.92	-92.93	-93.03
I44	-92.42	-92.34	-92.30
F45	-93.59	-93.51	-93.67
A46	-93.91	94.01	-94.01
G47	-93.28	-93.21	-93.00
K48	-93.03	-92.99	-92.97
Q49	-93.02	-93.04	-92.99
L50	-93.97	-93.85	-93.65
E51	-92.47	-92.38	-92.33
D52	-92.45	-92.39	-92.44
R54	-92.86	-92.91	-92.82
T55	-92.57	-92.45	-92.23
L56	-94.18	-94.13	-93.88
S57	-93.73	-93.75	-93.71
D58	-94.26	-94.09	-94.07
Y59	-92.32	-92.31	-92.13
N60	-94.15	-94.25	-94.19
I61	-93.07	-93.03	-92.99
Q62	-92.91	-92.93	-93.14
K63	-93.26	-93.28	-93.23
E64	-93.35	-93.29	-93.21
S65	-93.84	-93.84	-93.82
T66	91.85	91.76	91.58
L67	-93.08	-92.97	-92.87
H68	-93.08	-92.98	-92.74
L69	-93.10	-93.02	-93.06
V70	-93.45	-93.47	-93.62
L71	-92.97	-93.00	-93.14
R72	-93.12	-93.11	-93.24
L73	-92.86	-92.91	-92.87
R74	-92.99	-92.86	-92.88
G75	N.A.	N.A.	N.A.

Publication list

Min-Kyu Cho, **Shengqi Xiang**, Hai-Young Kim, Stenfan Becker, & Markus Zweckstetter: Cold-induced changes in the protein ubiquitin. (2012) PloS one 7(6): e37270.

Shengqi Xiang, Raghavendran L. Narayanan, Stefan Becker and Markus Zweckstetter: N-H Spin-Spin Couplings: Probing Hydrogen-Bonds in Proteins. (Submitted)

Francesca Munari, Nasrollah Rezaei-Ghaleh, **Shengqi Xiang**, Wolfgang Fischle, Markus Zweckstetter: Structural Plasticity in Human Heterochromatin protein 1β. (Submitted)

Curriculum Vitae

Personal information

Name:	Shengqi Xiang
Birthday:	September 26, 1982
Birthplace:	Wuhu, China
Citizenship:	China

Education

2000-2004	Undergraduate studies of biology School of Life Sciences, University of Science & Technology of China, Hefei, China.
2004	Bachelor of Science, Life Sciences.
2005-2007	Graduate student, Structural Biology., University of Science & Technology of China, Hefei, China.
2007-2008	Exchange student of MPS-CAS Doctoral Training Program Department of NMR-based Structural Biology, Max Planck Institute for biophysical chemistry, Göttingen, Germany.
2009-2012	Doctoral thesis at Department of NMR-based Structural Biology, Max Planck Institute for biophysical chemistry, Göttingen, Germany. (Supervisor: Prof.Dr. M. Zweckstetter) Subject: Structural insights into Argenine-Serine rich proteins and N-H spin-spin coupling constants

Göttingen, December, 2012