



**Structural characterization of
the leukemia drug target ABL kinase
and
unfolded polypeptides
by novel solution NMR techniques**

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Navratna Vajpai

aus Kanpur, Indien

Basel, 2010

Original document stored on the publication server of the University of Basel
edoc.unibas.ch



This work is licenced under the agreement „Attribution Non-Commercial No Derivatives
– 2.5 Switzerland“. The complete text may be viewed here:
creativecommons.org/licenses/by-nc-nd/2.5/ch/deed.en

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Stephan Grzesiek (Faculty responsible)

Dr. Wolfgang Jahnke (Co-referee)

Basel, den 14.10.2008

Prof. Dr. Eberhard Parlow, Dekan



Attribution-Noncommercial-No Derivative Works 2.5 Switzerland

You are free:



to Share — to copy, distribute and transmit the work

Under the following conditions:



Attribution. You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).



Noncommercial. You may not use this work for commercial purposes.



No Derivative Works. You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work. The best way to do this is with a link to this web page.
- Any of the above conditions can be waived if you get permission from the copyright holder.
- Nothing in this license impairs or restricts the author's moral rights.

Your fair dealing and other rights are in no way affected by the above.

This is a human-readable summary of the Legal Code (the full license) available in German:
<http://creativecommons.org/licenses/by-nc-nd/2.5/ch/legalcode.de>

Disclaimer:

The Commons Deed is not a license. It is simply a handy reference for understanding the Legal Code (the full license) — it is a human-readable expression of some of its key terms. Think of it as the user-friendly interface to the Legal Code beneath. This Deed itself has no legal value, and its contents do not appear in the actual license. Creative Commons is not a law firm and does not provide legal services. Distributing of, displaying of, or linking to this Commons Deed does not create an attorney-client relationship.

To my beloved family and all my respected teachers

Acknowledgements

From experience these pages of a PhD thesis are the most widely read pages of the entire publication, primarily, because it can catch even readers with no or little scientific knowledge and secondly but more importantly, because this page is written with all soul and heart to appreciate all those people who have helped in one way or another during the time of writing this dissertation. Indeed, at times while writing these pages, lack of words in comparison to the contribution made by people has debarred me of expressing how much it meant to me.

This is **high time** to mention that I am deeply indebted and express my gratitude to all the lovely people who have always been supportive and encouraging to accomplish this thesis. To start with, I would like to thank my thesis supervisor Prof. Stephan Grzesiek for his continuous support and encouragement since the start of this thesis. I am grateful to him for giving me an opportunity to pursue my thesis and introducing me to the fascinating world of protein NMR spectroscopy. During the course of my thesis, he has educated me on many aspects related to experiments as well as theory. His approach to the problems, suggestions and stimulating discussions has always endured me to learn more deeply about the subject. Apart from scientific discussions, his caring nature, ever availability and perseverance for perfection has taught me another lesson to build my overall personality. I feel highly privileged to have him as my guide.

I could not have wished for a better project than ABL kinase, which involves the famous blood cancer drug Glivec. I pay my heartfelt gratitude to Dr. Wolfgang Jahnke and other colleagues, Dr. Andre Strauss, Dr. Gabriele Fendrich, Dr. Paul W. Manley and Dr. Sandra W. Cowan-Jacob for an excellent collaboration. Your contributions, detailed comments and fruitful discussions has made possible to identify several key issues related to kinase that may lead to development of more potent drugs. I would also like to thank Dr. Jahnke for kindly accepting to be my referee. His inspiring words and positive vision has definitely helped in the success of the project. Adding to this, I would like to pass on my thanks to other colleagues at Novartis, Dr. Marcel JJ Blommers, Dr. Cesar Fernandez and Ms. Chrystelle Henry for their help and hospitality.

I thank Dr. Martin Blackledge and Dr. Alessandro Pintar for fruitful collaborations and their interest in our work.

I would like to thank Prof. Tilman Schirmer for accepting to be chairman of my PhD Thesis examination.

My earnest thanks goes to the very lovely and friendly colleagues in the group of Prof Grzesiek whose intellectual knowledge, both in biology as well as spectroscopy, has always boosted me to achieve my goal.

The one person who has always been there to help overcome my difficulties is Martin Allan. Many fruitful discussions with him, which include both scientific as well as social issues, have always been highly pleasurable. His hospitality and kindness especially during my initial days in Switzerland is simply outstanding. His gentle but very elegant critics have improved the quality of my work especially during the writing part of my thesis. In spite of his very busy schedule, he has always found some time for reading my thesis and given me plenty of valuable suggestions. On a lighter part, I am also thankful to him for inviting me to attend his marriage with Carmen Chan (now Ms. Allan), allowing me to attend first such kind of function in the past six years. I sincerely wish for their great future and hope that they will reach to the best scalable heights. I thank Martin and Carmen for every contribution they made.

I would like to thank Dr. Hans Juergen Sass for many fruitful discussions and his help in setting up MD simulations of polypeptides. His vast experience and knowledge in both biology and NMR has always benefited me whenever approached. I also would like to acknowledge him for reading my thesis and providing me his valuable suggestions.

My sincere thanks also go to Martin Gentner for nice collaboration, stimulating discussions and his help in preparation of few figures for my thesis. I wish to thank Lydia Nisius and Jie-rong Huang especially for voluntarily reading some chapters of my thesis and giving me their valuable suggestions. I thank Romel Bobby for nice collaboration and friendly behavior. I also wish to thank Sebastian and Sonja ji for their help especially during the initial days of my thesis; Sebastian for introducing me to the world of residual dipolar couplings that has been applied in most of my projects and Sonja ji for

introducing me to a NMR analysis package to say the least. I would also like to thank Regula Aregger for a nice collaboration and for her help in preparation of few NMR samples.

Prof. K. Pervushin, Dr. Daniel, Ms. Klara, Marco, Wei, Yaroslav, Luminita, Pernille, Judith, Maciej have all been great colleagues for the past few years. Their friendly and helping behavior had been one of the mainstays of my happy-go-living stay in Switzerland. I thank all of them from the bottom of my heart and wish them a great future ahead. Here, I would also like to thank Sara Paulilo and Dr. Brian Cutting for their hospitality and also Brian for his nice suggestions on the text of ABL kinase.

All the administrative and/or technical help very generously came from Debbie, Mr. Buerki, Ms. Margret, Mr. Wyss, Ms. Barbara, Mr. Beat Schumacher and people in the printing section, Ms. I. Singh and Ms. V. Grieder. Being unknown to the German language, their help at all levels have made my process very easy for me. I sincerely thank them for all that they have done to me.

The excellent scientific environment in the department is mainly set by the presence of many lovely colleagues and their cheerful faces. I really feel privileged to have worked in such a motivating environment.

Here, I also wish to acknowledge Prof. N. Chandrakumar who taught me my first lesson of NMR during my Master's thesis at IIT Madras. His words of encouragement actually inspired me to move into the fascinating world of NMR.

My thesis would not have been a success without the help and support of many wonderful Indian people around me. I feel honored and proud to have such a nice-Indian-family in Basel, whose love and affection never allowed me feel alone or miss my family. The kind of affection they have given me always make/made me feel that I have several homes in Basel. A "thanks" is certainly a very small word for everything they have bestowed on me. Social get-togethers, picnics, celebration of festivals have all made my stay highly pleasant and will remain as cherished memories all through my life.

Anurag bhaisahab and Divya bhabhi-1 have been the most adorable. I have always been cared and loved like a younger member of their family. I am deeply indebted to them but certainly feel that mere my words can't express for the love and affection they have showered on me. Another person who has been most supportive especially during the course of my thesis is Joshi ji. For the last two months, I was a permanent invitee for the dinner at his place. His kind motive was to avail some more of my time for writing my thesis. I pay my earnest gratitude to Joshi ji for his incomparable help and all great moments he has shared with me. I wish him and his wife Ashwini a great future ahead. Ratnesh bhaisahab-Richa (Parjai) ji, Vivek bhaisahab-Nidhi ji, Sudeep bhai-Rejina ji, Jenish-Jhanvi, Prasad, Abiraj, Shivani ji-Naveen bhaisahab, Harish-Manu, Murali bhaisahab-Reshmi ji and Sachin bhai-Abhi have all been excellent people around for past few years. This list will not be complete without including the names of Gudda bhai and Divya bhabhi-2, Srijit bhai-Brinda ji, Charuji-Mayank bhai, Akshata-Shantanu and Ago-Srinjoy. Their emotional support, backing and care had been incredible during all my good and bad times. I express my heartfelt gratitude to all of you for everything you have done to me. I would like to thank Senthil and Siva for their hospitality on the very first eve in Basel.

I wish to thank Anna, Karim, and Sonali for their wonderful friendship and support during my stay at Basel. I would also like to thank relatively new members in the Indian battalion Sabyasachi, Swarna, Sandeep, Ranjini, Ramya ji, Satrajit, Vijay-Shankar, Somedutta and Vimal for their encouraging words. I also wish to thanks many more friends Arundhati ji, Varadha bhai-Rashi ji, Hemant bhaiyya, Atul, Manoj, Saman, Helen, Kaustubh, Anil, Nikhil, Dushyant, Kaushal, Jitu-1, Vipu, Tapas, Ashish, Amit, Jitu-2, Sanju, Shiv and Bibha who despite of not living in the same city have always been highly encouraging and motivating. Thanks for being there with me. I wish you all a great life ahead.

This thesis wouldn't have been written without the blessings, constant emotional support and encouragement of my parents and siblings Ashuda, Puneet and Geetu. I feel proud to have such a great family. I am blessed to have great cousins G2 bhaiyya, Beetu, Divya, Shruti, Aayush, Pratyush, Richa, Gunjan, Tanu and Komal, uncles and aunts who

have always made my day whenever I spoke to them. Their encouragement and motivation is always a boon for me. I am extremely delighted to have Amita as a very special friend. Many heartfelt thanks goes to her as she always stood by me as a strong support. I am sure this success will definitely make all of them happy and proud. I wish all of you success in all your endeavors.

Special thanks to Rejina ji, Nidhi ji, Vivek bhaisahab, Jenish, Thapliyal ji, Parjai ji, Anurag bhaisahab and bhabhi ji for correcting some chapters of my thesis at different stages of writing.

Last but not the least, I wish to thank the almighty for giving me all wisdom and opportunities to reach at this level. My stay in Switzerland has been excellent since the day I have landed here. Wonderful people around me at work and in my social life have all contributed for this success. The most quality facilities, the lovely scenic beauty and excellent places nearby are simply the bonus to every other detail. Million thanks to everyone who have been a part of my life and contributed me to grow as an individual.

-Navratna Vajpai

Summary

In this thesis, novel weak alignment techniques and new biochemical strategies like selective isotope labeling in combination with other high-resolution solution NMR methods have been applied to characterize folded and unfolded polypeptides. This includes the characterization of the solution conformations of the leukemia drug target Abelson (ABL) kinase in complex with three clinical drugs (imatinib, nilotinib and dasatinib), unstructured/urea-denatured polypeptides, and the transcriptional repressor in the highly conserved Notch pathway, HES1.

Solution NMR studies of ABL kinase in complex with three clinical inhibitors

Aberrant forms of ABL kinase are important drug targets for the treatment of chronic myelogenous leukaemia (CML). The results of this thesis provide the first detailed characterization of solution conformations of ABL tyrosine kinase in complex with three effective clinical inhibitors imatinib, nilotinib and dasatinib. In solution, a centrally located regulatory segment termed the activation loop adopts the non-ATP binding inactive conformation in complex with imatinib and nilotinib, and preserves the ATP-binding active conformation in complex with dasatinib. However, relaxation studies and/or line broadening of some resonances in the activation loop and the phosphate-binding loop (P-loop) indicate presence of microsecond to millisecond dynamics for all the investigated ABL-inhibitor complexes. These results contribute to our understanding of drug resistance and support the rational design of improved kinase inhibitors (Manley *et al.*, 2006, Vajpai *et al.*, 2008a, Vajpai *et al.*, 2008b).

Conformational studies of unstructured polypeptides by residual dipolar couplings

The characterization of unfolded states of polypeptide chains is of high significance with regard to their role in biological processes and to understanding protein folding. Here, we have investigated the influence of amino acid substitutions X on the conformation of unfolded model peptides EGAAXAASS as monitored by backbone RDCs. The RDCs show a specific dependence on the substitutions X that correlates to steric or hydrophobic interactions with the adjacent amino acids. RDC profiles along the nonapeptide sequence show large variations for a few amino acid substitutions. In particular, RDCs for glycine

and proline indicate less or more order than the other amino acids, respectively. The RDCs for aromatic substitutions tryptophan/ tyrosine or isoleucine give evidence of kink or stiffness in the polypeptide backbone (Dames *et al.*, 2006).

For a quantitative description, these experimental results were compared to the predictions from the statistical coil model, which derives amino acid specific local conformations from the torsion angle distribution of non- α , non- β structures in folded proteins, or all-atom molecular dynamics (MD) simulations. While the coil model reproduced, to some extent, the observed RDC pattern for most substitutions, MD simulations showed stronger deviations from the experimental data. This indicates specific deficiencies in both the statistical coil model and the MD simulations. For the coil model, the discrepancy may be related to imperfect modeling of the side chains, while for MD simulations, inadequate sampling of the conformational space in the time used for the simulations may be the most plausible reason.

Side-chains conformations in urea-denatured proteins: a study by 3J scalar couplings and residual dipolar couplings

In order to probe the conformational behavior of the side-chains in unfolded states, we have measured an extensive set of six three-bond scalar couplings ($^3J_{\text{NH}\beta}$, $^3J_{\text{C}'\text{H}\beta}$ and $^3J_{\text{H}\alpha\text{H}\beta}$) and two $^1D_{\text{C}\beta\text{H}\beta}$ residual dipolar couplings (RDCs) on urea-denatured proteins, ubiquitin and protein G. Interpretation of the 3J couplings by a model of mixed staggered χ_1 rotamers yields excellent agreement and also provides stereoassignments for $^1\text{H}^\beta$ methylene protons. Independent analysis of $^1D_{\text{C}\beta\text{H}\beta}$ RDCs obtained in polyacrylamide gels show good correlation with the RDCs predicted from the χ_1 populations obtained from the 3J data and a coil model ensemble of 50000 conformers according to the coil library backbone angle distribution. The study validates coil model as a good first approximation of the unfolded state. However, individual variations from the coil averages of up to 40% are highly significant and must originate from sequence- and residue-specific interactions. The deviations between the measured and predicted values also indicate that the local backbone geometries may be improved by incorporation of the additional RDC information (Vajpai *et al.*, 2010).

Backbone resonance assignment of the 31 kDa of homodimer of apo-HES1

HES1 acts as an effector of highly conserved intercellular Notch signaling pathway by repressing the expression of target genes. The backbone resonance assignment and homology modeling of the 31 kDa homodimer of apo-HES1 are reported. The obtained results are being used for further structural studies on HES1.

Results from this thesis have been published in the following peer-reviewed articles:

1. Dames S.A., Aregger R., Vajpai N., Bernado P., Blackledge M., and Grzesiek S.,
Residual dipolar couplings in short peptides reveal systematic conformational preferences of individual amino acids
J Am Chem Soc 2006 **128**; 13508-13514
2. Vajpai N., Strauss A., Fenderich G., Manley P.W., Jacob S., Jahnke W., and Grzesiek S.
Backbone NMR resonance assignment of the Abelson kinase domain in complex with imatinib.
Biomol NMR Assgn 2008 **2**: 41-42
3. Vajpai N., Strauss A., Fenderich G., Manley P.W., Jacob S., Grzesiek S., and Jahnke W.
Solution conformations and dynamics of ABL kinase inhibitor complexes determined by NMR substantiate the different binding modes of imatinib/nilotinib and dasatinib. * §
J Biol Chem 2008 **283**; 18292-18302
4. Vajpai N., Gentner M., Huang J.R., Blackledge M., and Grzesiek S.
Side-chain χ_1 conformations in urea-denatured ubiquitin and protein G from 3J coupling constants and residual dipolar couplings
J Am Chem Soc 2010 **132**; 3196-3203

* This paper has been selected as Paper of the Week by the Editorial Board of the *Journal of Biological Chemistry (JBC)*.

§ This paper has been selected for the Faculty of 1000 Biology.

Contents

<i>Acknowledgements</i>	v
<i>Summary</i>	x
<i>Contents</i>	xiv
<i>Abbreviations and symbols</i>	xvi
Chapter 1: Introduction	1
Structure determination of biomacromolecules	1
Theory of residual dipolar couplings	7
Chapter 2: Solution NMR studies of ABL kinase in complex with three clinical inhibitors	17
Abstract	17
Background	18
Serine/Threonine kinases	18
Tyrosine kinases	19
Protein tyrosine kinases as targets for inhibitor design	20
Role of structural biology in drug design	21
Abelson tyrosine kinase	23
Structural characterization of ABL kinase complexes by high-resolution solution NMR techniques	31
Original Publications	34
Chapter 3: Conformational studies of unstructured oligopeptides by residual dipolar couplings	55
Abstract	55
Background	56
Quantitative characterization of unfolded states by NMR spectroscopy	58
Section 3.1: Conformational preferences of individual amino acids in short peptides revealed by residual dipolar couplings	63
Original Publication	64
Section 3.2: Residual dipolar couplings of nonapeptides as predicted from molecular dynamic simulations	73
Introduction	73
Materials and Methods	74
Results and Discussion	74
Conclusions	76
Acknowledgements	76

Chapter 4: Side-chain χ_1 conformations in urea-denatured proteins: a study by 3J coupling constants and residual dipolar couplings	78
Original Publication	80
Chapter 5: Backbone resonance assignment and homology modeling of the 31 kDa protein dimer of HES1: a transcriptional repressor protein in the Notch signaling pathway	106
Abstract	106
Background	107
Overview of Notch signaling	107
Overview of HES/E(spl) family:	110
Structural studies of HES1	112
Materials and Methods	113
NMR samples and experiments:	113
Results	114
Backbone resonance assignment of HES1	114
Homology modeling of the Orange domain of HES1	117
Chapter 6: Conclusions and perspectives	119
Bibliography	122
<u>CURRICULUM VITAE</u>	137
<u>Work Experience</u>	137

Abbreviations and symbols

{ ¹ H}- ¹⁵ N NOE	¹⁵ N steady state NOE upon ¹ H saturation
A	diagonal alignment tensor with components A _{xx} , A _{yy} , A _{zz}
Å	0.1 nm
A _{zz}	z component of the diagonalized alignment tensor, $D = A_{zz} \left[\frac{(3\cos^2\theta - 1)}{2} + \frac{\eta}{2} \sin^2\theta \cos 2\varphi \right]$
ABL	abelson
AHBP	alternating hydrogen bond potentials
ATP	adenosine tri-phosphate
bHLH	basic helix-loop-helix
BCR	breakpoint cluster region
BMRB	BioMagResBank, www.bmrwisc.edu
CIDNP	chemically induced dynamic nuclear polarization
CML	chronic myelogenous leukemia
CSA	chemical shift anisotropy
δ	chemical shift
δ _{ij}	Kronecker symbol
D	dipolar coupling
D _{max} ^{IS}	$-\frac{\mu_0}{4\pi} \frac{h}{2\pi} \frac{\gamma_I \gamma_S}{r_{IS}^3}$, maximal solid state dipolar coupling
DHPC	diheptanoyl-phosphatidylcholine
DMPC	dimyristoyl-phosphatidylcholine
EGF	epidermal growth factor
γ _x	gyromagnetic ratio of nucleus X
η	(A _{xx} -A _{yy})/A _{zz} , asymmetry parameter
h	6.6·10 ⁻³⁴ Js, Planck's constant
ΔH	enthalpy of transition
HMQC	heteronuclear multiple quantum coherence
HSQC	heteronuclear single quantum coherence

INEPT	insensitive nucleus enhancement by polarization transfer
I_z	z component of the spin operator
3J	three-bond scalar coupling
kDa	kilo dalton
μ_0	$4\pi \cdot 10^{-7}$ V·s/(A·m), vacuum permeability
MD	molecular dynamics
NOE	nuclear Overhauser effect
NICD	Notch intracellular domain
$P_2(x)$	$1/2 \cdot (3x^2 - 1)$, 2nd order Legendre polynomial
PDB	RCSB Protein Data Bank, www.rcsb.org/pdb
Pf1	filamentous phage
P-loop	phosphate-binding loop
ppm	parts per million
PRE	paramagnetic relaxation enhancement
PTK	protein tyrosine kinase
R_1, R_2	longitudinal and transverse relaxation rate
r_{IJ}	internuclear distance between I and J
RDC	residual dipolar coupling
rmsd	root mean square deviation
ROE	rotating frame Overhauser effect
S	order parameter
S_{ij}	Saupe order matrix
SAXS	small-angle scattering experiments
θ, φ	polar angles
T_1, T_2	longitudinal and transverse relaxation time
TCEP	tris(carboxyethyl)phosphine hydrochloride
TEMED	N,N,N',N'-tetramethylethylenediamine
TOCSY	total correlation spectroscopy
TROSY	transverse relaxation optimized spectroscopy
U	uniformly isotope labeled

Atoms and angles are referred to according to IUPAC nomenclature.

Chapter 1:

Introduction

Structure determination of biomacromolecules

For the past few decades, progress in all areas of structural biology has shown that there are no real limitations to determining the three-dimensional structures of considerable size and complexity. Today biomolecular structures are solved at ever-increasing rates, sizes and qualities, as the applied methods are continuously improving. The methods that structural biologists use to determine structures include X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, electron microscopy and atomic force microscopy. These structural determination techniques are not competing techniques, but rather complement each other. For example, the information from cryo-electron microscopy and atomic force microscopy yield structures of extremely large systems or whole cells (Baumeister, 2002) at relatively low resolution, that can, however, often be combined with atomic resolution structures.

The oldest and most precise method to obtain high-resolution structural information is the diffraction of X-rays from a crystalline material. The important developments in crystallography like seleno-methionine derivatives, cryo-freezing, robotic crystallization, synchrotron radiations and improvement in refinement techniques have made possible to solve crystal structures at high precision. Structures with very high molecular weight up to MDa units (Ban *et al.*, 2000) and with resolutions as good as 0.54 Å have already been obtained (Jelsch *et al.*, 2000). Since its first success in solving a biomolecular structure of sperm whale myoglobin in the late 1950's (Kendrew *et al.*, 1958), X-ray crystallography has reported over 45000 structures in Protein Data Bank (PDB). However, production of good crystals is a limiting factor.

In the past two decades, NMR has established itself as a powerful method for high-resolution structure determination of biological macromolecules in solution. Both X-ray and NMR complement each other as the two techniques provide different type of information in different environments. Together, they can provide an atomic detail

picture of macromolecular structure and dynamics that help in understanding of life processes at molecular level.

While crystallography has deposited more structures in the Protein Data Bank (PDB), NMR is unique in extracting dynamical information on biological macromolecules over a large range of timescales. This makes it a more efficient technique for determining ligand binding and mapping interaction surfaces of the protein/ligand complexes. The most severe drawback of solution NMR is the molecular size limitation, which results from the increasing resonance linewidths at increasing molecular weight and extensive signal overlap. The first *de novo* NMR structure of a globular protein, the bull seminal protease inhibitor (BUSI), was solved by the Wüthrich group in 1985 (Williamson *et al.*, 1985), but since then advances in hardware, such as stronger magnets and cryoprobe, NMR methodology combined with molecular biology and recently developed isotopic labeling methods have expanded the range of proteins amenable to structural determination. With transverse relaxation-optimized spectroscopy (TROSY) (Pervushin *et al.*, 1997) molecular limits has increased by an order of magnitude. The Kay group has achieved resonance assignment and characterized interdomain dynamics of enzyme malate synthase G, an 81 kDa protein (Tugarinov *et al.*, 2002, Tugarinov & Kay, 2003b, Tugarinov & Kay, 2003c, Tugarinov & Kay, 2003a, Tugarinov *et al.*, 2003, Korzhnev *et al.*, 2004). Even spectra of multimeric proteins up to 900 kDa have been obtained (lysine decarboxylase (810 kDa), (Tugarinov *et al.*, 2004); GroEL-GroES complex (900 kDa) (Fiaux *et al.*, 2002). Recently, the Kay group has characterized dynamics of 20S proteasome, a multimeric protein of size 670 kDa (Sprangers & Kay, 2007).

Rapid data acquisition techniques have been developed to significantly reduce the spectrometer time, as large numbers of spectra are required in the course of single investigation of structure, dynamics and interactions of a protein. These techniques include GFT (Kim & Szyperski, 2003, Atreya & Szyperski, 2005), nonuniform sampling (Rovnyak *et al.*, 2004, Marion, 2005), Hadamard spectroscopy (Kupce & Freeman, 2003a, Kupce & Freeman, 2003b), single scan NMR (Frydman *et al.*, 2002, Frydman *et al.*, 2003), projection–reconstruction (Kupce & Freeman, 2003c, Kupce & Freeman, 2004, Kupce & Freeman, 2005) and filter diagonalization (Chen *et al.*, 2000,

Mandelstam, 2000, Hu *et al.*, 2000). These methods avoid the strictures imposed by the conventional sampling strategy in which the multidimensional NMR experiments are recorded by systematically and independently incrementing each of the indirect evolution delays. A particularly powerful illustration is provided by a 4D ^{13}C , ^{13}C -edited NOESY spectrum recorded using methyl-TROSY techniques and nonuniform sampling on malate synthase G, a protein of size 81.4 kDa (Tugarinov *et al.*, 2005).

NMR has a wide range of applications, for example NMR of solids (crystalline or powders) is applied in inorganic chemistry and material sciences to characterize polymers. High-resolution solid-state NMR has allowed characterization of non-crystalline membrane proteins (de Groot, 2000) and amyloid fibrils (Petkova *et al.*, 2002, Balbach *et al.*, 2002) that are not amenable to other structural determination techniques. New methods are being developed to determine 3D structures using similar methods as for structure determination in the liquid state. These developments are very important, since it is still difficult to crystallize membrane proteins, while no crystals are required for NMR.

With the recent improvements in sensitivity and techniques, NMR has now been used in drug discovery and structural genomics, and has shown its potential for playing a greater role in the pharmaceutical and biotech industries. A major application of NMR is in structure-based drug design. Using structure-activity relationship (SAR) by NMR (Shuker *et al.*, 1996), novel lead compounds are constructed in a rational way that cannot be found using conventional methods. Recently, NMR has characterized dynamics in the key regions of pharmaceutically important proteins targets (Vogtherr *et al.*, 2006, Honndorf *et al.*, 2008). These studies hopefully will lead us to rationale design of more potent drugs.

Another important application is Magnetic Resonance Imaging (MRI) (Lauterbur, 1973, Stehling *et al.*, 1991), which allows *in vivo* imaging of human tissue. MRI based computer tomography is used to obtain slice images of a human body, for example in cancer diagnostics. An extension of this is localized spectroscopy, which allows studying metabolism in different tissues or organs *in vivo*.

A major drawback with NMR is its inherent insensitivity, which leads to requirement of large amounts of sample. In a recent study, using an in-house built microslot waveguide probe, Maguire *et al.* resolved signal of ~ 1.6 nmol of RNase A (Maguire *et al.*, 2007). This microslot is a dual-layer, metallic, planar structure with its largest dimension much smaller than the wavelength of the electromagnetic wave and its width much smaller than the height of the dielectric. This structure concentrates flux density and has properties ideal for the detection of magnetic flux density. The ability to generate spectra from such a small amount of protein is of extreme importance, as many biologically important proteins cannot be expressed in standard NMR required quantities.

Most NMR studies of protein structure are based on analysis of the nuclear Overhauser effect, NOE; (Overhauser, 1953), between different protons in the protein. Because the NOE depends on the inverse sixth power of the distance between the nuclei, they can be converted into distance restraints that can be used in molecular dynamics- or distance geometry-type structure calculations. Other quantitative restraints that are traditionally used in structure calculations are dihedral angle restraints, which are restricted to groups of atoms separated by three bonds. Over large distances, uncertainties in these short-range restraints will add up, which means that NMR structures of large, elongated systems (*e.g* for B-form DNA) will be poor overall even though individual regions of the structure will be well-defined. Also, as a consequence of the local short-range nature of these restraints, it is difficult to accurately define the relative orientation of regions, which are apart from each other in the three-dimensional structure of the molecule.

Over the last ten years, advances in the NMR methodology that provide long-range information have tremendously improved the accuracy of the structures. Partial alignment of samples is prerequisite to the observation of these long-range information, such as residual dipolar couplings (RDCs) and chemical shift anisotropy (CSA), that come from anisotropies in spin interactions (Prestegard *et al.*, 2004). These anisotropic data contain information about internuclear vector orientations in a global common reference frame and therefore yield the orientation of internuclear vectors relative to each other, irrespective of their distance. Early applications used the field-induced order of liquid crystalline media to indirectly induce nonisotropic distributions through collisional

interactions of the molecules of interest (Saupe & Englert, 1963, Bothner-By *et al.*, 1981, Tjandra & Bax, 1997) or they used high magnetic fields and inherent anisotropies in magnetic susceptibilities of the molecules of interest to directly induce nonisotropic distributions (Tolman *et al.*, 1995). Later it turned out that one can use various other kinds of media such as compressed and charged gels, phages, purple membrane etc., to *partially* orient samples, so that these interactions no longer averages to zero but has some small *residual* value (~1000 times less compared to the dipolar couplings observed in solid-state NMR). The most prominent amongst these long-range restraints are residual dipolar couplings. As RDCs can be easily measured with high accuracy and provide complementary information to short-range NOEs, they became integral part of many structural refinement protocols, especially, for the multi-domain biomolecules. Since the recognition of the potential of RDCs in protein structure determination, applications have spread to nucleic acid structure, carbohydrate structure, protein-ligand interactions, protein domain relationships, high-throughput strategies for structural genomics, and studies of motional amplitudes in flexible assemblies.

Due to high sensitivity and global distribution throughout the protein, directly bonded $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$ and $^1\text{H}^{\text{C}\alpha}\text{-}^{13}\text{C}^{\alpha}$ are most commonly measured dipolar couplings, and can be observed with minimal modifications of pulse sequences. A variety of experimental schemes have been designed for extraction of other backbone RDCs in $^{15}\text{N}/^{13}\text{C}$ or $^{15}\text{N}/^{13}\text{C}^2\text{H}$ -labeled proteins. RDCs have shown substantial improvements in the quality of NMR structures (Drohat *et al.*, 1999, Pan *et al.*, 2002, Wiesner *et al.*, 2002). Refinement with RDCs leads to novel insights into NMR structures and in turn into biological functions. Crystal structures of the same protein can result in structural discrepancies, making it difficult to determine the physiologically relevant structure (Lipsitz & Tjandra, 2004). *De novo* structures have been determined solely based on RDCs measured in different media (Delaglio *et al.*, 2000, Hus *et al.*, 2001, Delaglio, 2000, Simon & Sattler, 2002, Rohl & Baker, 2002, Beraud *et al.*, 2002, Meier *et al.*, 2007a, Bouvignies *et al.*, 2006). RDCs are sensitive to motion from picoseconds to milliseconds, and thus complement relaxation studies in characterizing slow dynamics ($> 10^{-9}$ sec) in biomolecules (Tolman *et al.*, 1997, Fischer *et al.*, 1999, Tolman, 2001, Honndorf *et al.*, 2008, Peti *et al.*, 2002). Recently, RDCs have shown large applications

in characterization of partially unfolded or unstructured states (Shortle & Ackerman, 2001, Ding *et al.*, 2004). The detailed analysis of experimental data and its interpretation on the basis of statistical coil model (Bernado *et al.*, 2005, Jha *et al.*, 2005) has been reviewed recently (Meier *et al.*, 2008). Many eukaryotic proteins are multi-domain rather than single-domain with each domain have an average size of about 153 residues (Orengo *et al.*, 1999). Because of their size, most of these domains are amenable to heteronuclear NMR methods. However, determining relative domain orientations has been challenging using the traditional NMR short-range restraints. As demonstrated in many reports, domain orientations can be established using long range orientational constrains derived from measurements of RDCs, reviewed by (Tolman, 2001, Kay, 2001, Prestegard *et al.*, 2004). Studies of multi-domain proteins systems is an important area of collaboration between RDC-NMR and structural genomics (Al-Hashimi & Patel, 2002). The expanding repertoire of multi-domain protein structures determined by RDCs is highlighting important differences between the domain orientations determined in solution and their solid-state counterparts, possibly due to the crystal contacts.

Theory of residual dipolar couplings

Dipolar couplings arise from the interaction of two magnetically active nuclei (Figure 1.1). In the heteronuclear case, the transverse spin operator components oscillate rapidly and average to zero. Thus, the energy of the interaction for a heteronuclear spin-coupled pair of S and I, in the presence of an external magnetic field, can be given as:

$$H_d = -\frac{\mu_0}{4\pi} \frac{h}{2\pi} \gamma_I \gamma_S I_z S_z \frac{3\cos^2 \xi - 1}{2r_{IS}^3} \quad (1)$$

where μ_0 , magnetic permittivity of vacuum; h , Planck's constant; ξ is the angle of the internuclear vector with the magnetic field and r_{IS} is the fixed internuclear distance (assumption); and γ_I , γ_S are the gyromagnetic ratio and I_z , S_z are the longitudinal component of the nuclei I and S, respectively.

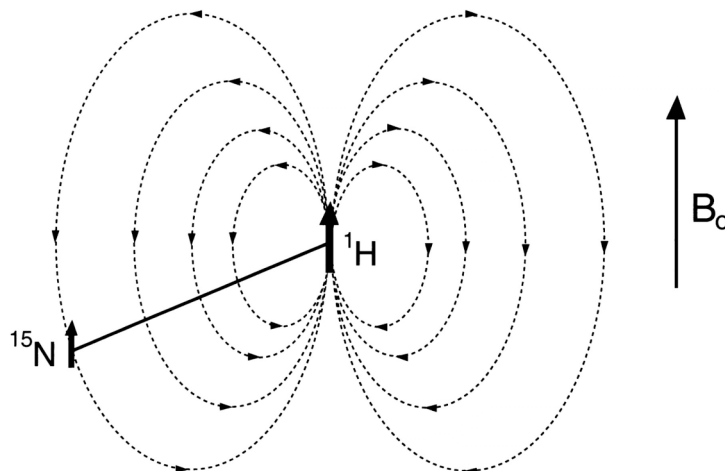


Figure 1.1: Dipolar coupling illustrated for a ^{15}N - ^1H spin pair. ^{15}N and ^1H magnetic moments are aligned parallel (or antiparallel) to the static magnetic field, B_0 . The total magnetic field in the B_0 direction at the ^{15}N position can increase or decrease relative to B_0 , depending on the orientation of the ^{15}N - ^1H vector and the spin state of the proton (parallel or antiparallel to B_0). Taken from (Bax, 2003).

In solution, due to motional averaging, equation (1) transforms to

$$\langle H_d \rangle = -\frac{\mu_0}{4\pi} \frac{h}{2\pi r_{IS}^3} \gamma_I \gamma_S I_z S_z \left\langle \frac{3\cos^2 \xi - 1}{2} \right\rangle \quad (2)$$

where $\langle \rangle$ denotes the time and ensemble averaging.

The dipolar field of spin I adds to the local magnetic field and thus changes the resonance frequency of spin S by a value depending on the orientation of the internuclear vector, the magnetic moment and their distance. For spin-1/2 nuclei, half of the spins are parallel to the magnetic field and half are antiparallel. Thus, spin I will either increase or decrease the magnetic field at spin S, by an equal amount, resulting in a doublet of resonances. The frequency separation for the doublet can be given by equation (3).

$$D^{IS} = D_{\max}^{IS} \left\langle \frac{3\cos^2 \xi - 1}{2} \right\rangle \text{ or } D_{\max}^{IS} \langle P_2(\cos \xi) \rangle \quad (3)$$

where $D_{\max}^{IS} = -\frac{\mu_0}{4\pi} \frac{h}{2\pi} \frac{\gamma_I \gamma_S}{r_{IS}^3}$ is the maximal splitting obtained.

In an isotropic solution, rotational Brownian diffusion rapidly averages the internuclear dipolar interaction of equation (3) to exactly zero, and therefore, the valuable orientational information is lost. However, when the protein is dissolved in a slightly anisotropic aqueous medium, complete averaging of the dipolar interaction does not occur. Practically useful alignment for biomolecules leaves residual (to a large extent averaged) dipolar couplings of tens of Hz from the several-kHz-couplings observed in solids where no averaging occurs.

The spin parts of heteronuclear J -coupling ($H_J = hJ \cdot I_z S_z$) and dipolar coupling Hamiltonians, equation (1), are identical and simply add up to the observed splitting. Thus, the value of the RDC is determined by comparison of the splitting in an aligned state with a reference spectrum in isotropic phase where only the J-splitting is detected.

RDC description based on the Saupe order matrix

For a rigid molecule, a unit vector \hat{r} in the internuclear direction can be expressed by coordinates (c_x, c_y, c_z) that are fixed relative to an arbitrary, time dependent molecular frame, given by the unit vectors $\hat{e}_x, \hat{e}_y, \hat{e}_z$:

$$\hat{r} = c_x \cdot \hat{e}_x + c_y \cdot \hat{e}_y + c_z \cdot \hat{e}_z = \begin{pmatrix} \hat{e}_x & \hat{e}_y & \hat{e}_z \end{pmatrix} \cdot \begin{pmatrix} c_x \\ c_y \\ c_z \end{pmatrix} \quad (4)$$

It is often convenient to express the dipolar coupling in a molecular fixed frame. The residual dipolar coupling is proportional to $\langle P_2(\cos\xi) \rangle$ where $\cos\xi$ is the scalar product between \hat{r} and a unit vector \hat{b} parallel to the magnetic field axis:

$$\cos\xi = \frac{\sqrt{b} \cdot \sqrt{r}}{\left| \sqrt{r} \right|} = \begin{pmatrix} \sqrt{b} \cdot \sqrt{e}_x & \sqrt{b} \cdot \sqrt{e}_y & \sqrt{b} \cdot \sqrt{e}_z \end{pmatrix} \cdot \begin{pmatrix} c_x \\ c_y \\ c_z \end{pmatrix} = (C_x \quad C_y \quad C_z) \cdot \begin{pmatrix} c_x \\ c_y \\ c_z \end{pmatrix} = \sum_{i=x,y,z} C_i c_i$$

(5)

where $C_{x,y,z}$ describe the direction cosines of \hat{b} relative to the coordinate system $\hat{e}_x, \hat{e}_y, \hat{e}_z$.

Equation (3), therefore, can be described by applying equation (5)

$$\langle P_2(\cos\xi) \rangle = \frac{3}{2} \left\langle \left(\sum_{i=x,y,z} C_i c_i \right)^2 \right\rangle - \frac{1}{2} = \frac{3}{2} \sum_{\substack{i=x,y,z \\ j=x,y,z}} \langle C_i c_j \rangle c_i c_j - \frac{1}{2} \quad (6)$$

or,

$$\langle P_2(\cos\xi) \rangle = \sum_{\substack{i=x,y,z \\ j=x,y,z}} c_i S_{ij} c_j = \begin{pmatrix} c_x & c_y & c_z \end{pmatrix} \cdot \begin{pmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{pmatrix} \cdot \begin{pmatrix} c_x \\ c_y \\ c_z \end{pmatrix} \quad (7)$$

using the identity, $c_x^2 + c_y^2 + c_z^2 = 1$.

Equation (6) describes the molecular average orientation in the magnetic field and represents the symmetric 3×3 Saupe order matrix S_{ij} given by (Saupe & Englert, 1963)

$$S_{ij} = \frac{1}{2} \langle 3 \cos\xi_i \cos\xi_j - \delta_{ij} \rangle \quad (8)$$

where $i, j = x, y, z$ are the axes of the molecular Cartesian coordinate system, ξ_k where $k = i, j$ is the angle of the axis with the magnetic field and δ_{ij} is the Kronecker delta symbol.

As the Saupe order matrix is real, symmetric and traceless, it can be diagonalized by the rotation to the principle axis coordinate system. Thus, equation (8) can be written as

$$D = \langle P_2(\cos \xi) \rangle = D_{\max} [S'_{xx} c'^2_{xx} + S'_{yy} c'^2_{yy} + S'_{zz} c'^2_{zz}] \quad (9)$$

where, the primes denote the quantities within the principle axis system.

The internuclear vector orientation (Equation 9) in the principle axis frame is conveniently described by polar coordinates (θ, φ) as

$$D = D_{\max} S'_{zz} \left[\frac{(3\cos^2 \theta - 1)}{2} + \frac{\eta}{2} \sin^2 \theta \cos 2\varphi \right] \quad (10a)$$

and is frequently written as

$$D = A_{zz} \left[\frac{(3\cos^2 \theta - 1)}{2} + \frac{\eta}{2} \sin^2 \theta \cos 2\varphi \right] \quad (10b)$$

where, $A_{zz} = D_{\max} S'_{zz}$ is referred to as the magnitude of the dipolar coupling tensor, $\eta = (S'_{xx} - S'_{yy})/S'_{zz}$ or $\eta = (2/3)(A_{xx} - A_{yy})/A_{zz}$ is the rhombicity and the alignment matrix used in its traceless form, \mathbf{A} with elements $|A_{zz}| > |A_{yy}| > |A_{xx}|$. The geometric dependence of the RDC is exclusively orientational, if the distance r_{IS} is known. If r_{IS} is not known, both distance and orientation influence the RDC.

These equations also indicate that the relationship between D and (θ, φ) is many-to-one, since there exists manifolds of (θ, φ) points that give rise to the same dipolar coupling. Thus, the dipolar coupling does not uniquely define the orientation but restricts it to be on the surface of a distorted cone (Figure 1.2). Because the direction of a second rank tensor interaction cannot be distinguished from its inverse, the dipolar coupling actually defines two cones of possible bond vector orientations, in opposing directions. However, this ambiguity of the bond vector orientation can be removed by measuring the RDC data in a second medium under different alignment conditions. Thus, measuring the residual dipolar coupling in two alignment media, the orientation can be defined to the intersections of these lines (Ramirez & Bax, 1998).

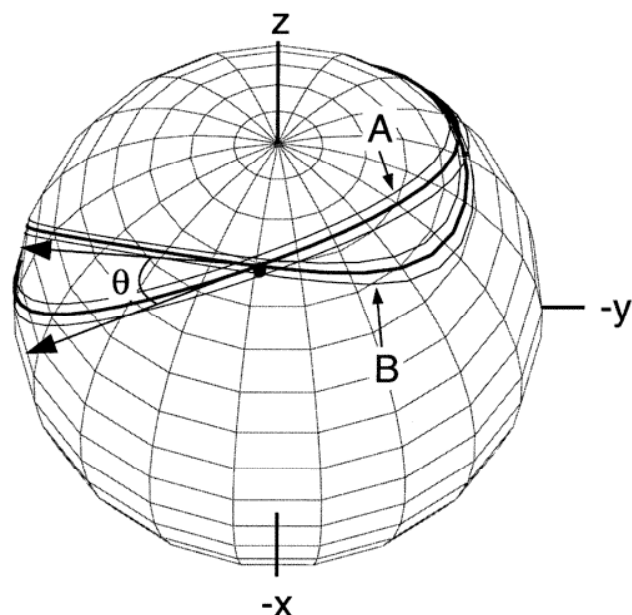


Figure 1.2: RDCs determine the orientation of internuclear vectors to a line with a relatively small error. Measuring the residual dipolar couplings in two alignment media can lift the ambiguity of the vector direction. The orientation can be defined to the intersections of these lines. Taken from (Ramirez & Bax, 1998).

Tunable weak alignment has first been achieved by placing proteins into dilute aqueous liquid crystalline media, composed of dihexanoyl- (or diheptanoyl-) phosphatidylcholine and dimyristoyl-phosphatidylcholine (DHPC/DMPC) (Tjandra & Bax, 1997). DMPC and DHPC are uncharged and interactions with the solute occur by steric collision. In this situation, simple obstruction models could predict the steric alignment tensor to high accuracy (Zweckstetter & Bax, 2000). Electrostatic contributions can be introduced by doping bicelles with charged amphiphiles, thus leading to a different alignment tensor (Ramirez & Bax, 1998).

Other alignment media that are commonly used: filamentous phage Pf1 (Hansen *et al.*, 1998) or other rodlike viruses (fd, TMV) (Clare *et al.*, 1998), lamellar phases consisting of ether/alcohol mixtures (“Otting media”), mechanically stressed polyacrylamide gels (Sass *et al.*, 2000, Tycko *et al.*, 2000) or charged copolymer gels (Meier *et al.*, 2002), liquid crystalline “Helfrich phases” (Prosser *et al.*, 1998) and purple membrane of *Halobacterium salinarum* with bacteriorhodopsin (Sass *et al.*, 1999) in two-dimensional crystalline arrangement. In addition, artificially coupled paramagnetic groups can also achieve alignment.

The identification of suitable medium for a particular application is not necessarily trivial. It is not simply sufficient that medium does not perturb molecular structure; it must also induce a proper level of alignment. Alignment must be sufficient to give measurable RDCs but not so large as to introduce spectral complexity. Several factors beyond simple concentration of the orienting medium must also be taken into account when attempting to predict the level of order. For example, the overall charge and charge distribution of a protein must be considered when attempting to orient it in an electrically charged medium; *e.g.* a positively charged protein will interact strongly with negatively charged filamentous phage, leading to broad lines and poor resolution. A highly asymmetric charge distribution (large quadrupole moment) will also lead to greatly enhanced RDCs. In some cases, raising ionic strength can alleviate problems with strong charge-induced association or orientation. However, this solution can be problematic with less salt-tolerant high-sensitivity cryogenic probes.

In this thesis, filamentous phage Pf1 phages and mechanically stressed polyacrylamide gels were used for partial alignment. A brief discussion of these two media is as follows:

Filamentous phage Pf1

Pf1 phage is a 7,349-nucleotide DNA-phage where the circular DNA is packaged with coat protein at a 1:1 nucleotide: coat protein-ratio. The Pf1 phages forms rods of ca 20,000 Å length and 60 Å diameter and spontaneously align by their intrinsic diamagnetic susceptibility in the magnetic field (Figure 1.3). Pf1-phages can be grown in *Pseudomonas aeruginosa* and are commercially available (ASLA biotech). Phages have a net negative surface charge and biomolecules are therefore mainly aligned via electrostatic interactions. Positively charged biomolecules at a pH below their pI thus might interact too strongly with the phages. Other rod-shaped viruses like fd and tobacco mosaic virus have been reported to have a similar orienting effect (Clare *et al.*, 1998).

Magnetic alignment of the Pf1 phage can be monitored by 1D ²H NMR spectra. The splitting of the HOD signal arises from the large deuterium quadrupole moment that is not isotropically averaged for water bound to the aligned phage particles. The observed quadrupole splitting increases with phage concentration indicating that the degree of

ordering of the water can be tuned by adjusting the phage concentration. Below a certain concentration threshold (~10-20 mg/ml), the dependence is non-linear (Zweckstetter & Bax, 2001). The alignment is tunable by addition of salt. At NaCl concentrations of up to 600 mM and above 16 mg/ml phage concentration, pH 7.2 (Zweckstetter & Bax, 2001), the dependence is linear. pH-values recommended originally are 6.5-8.0 and NaCl-concentrations below 100 mM (Hansen *et al.*, 1998). Phages have a tendency to aggregate at pH values below 6.

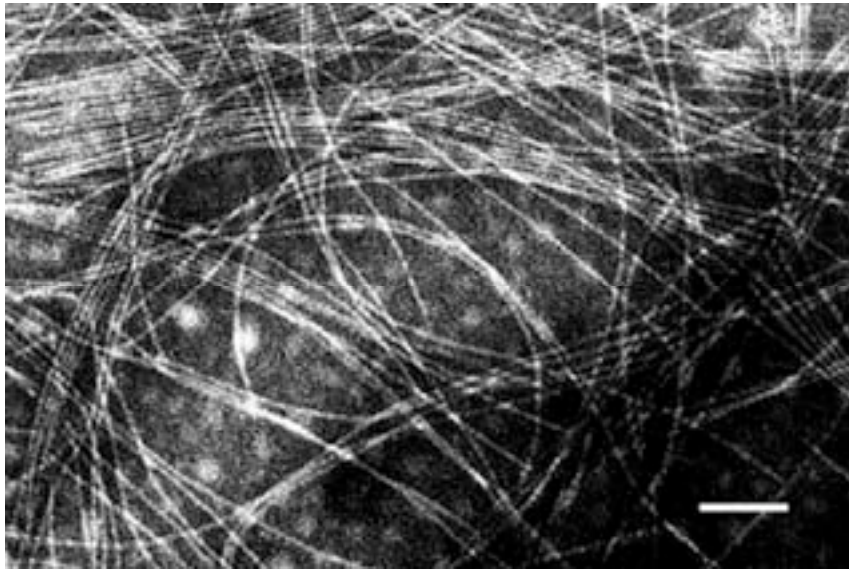


Figure 1.3: The electron microscopy picture of the Pf1 filamentous phages. (Taken from <http://www.asla-biotech.com/asla-phage.htm>)

Advantages of filamentous phages in residual dipolar coupling experiments

- Alignment is extremely stable for a long time under the physiological conditions and over a wide temperature range
- Alignment can be tuned by changing phage concentration and/or salt concentration.
- Macromolecule of interest can be easily separated by ultracentrifugation
- No effect on the rotational correlation time of nucleic acids

Disadvantage of filamentous phages

- As the pH range for Pfl phages is small (6.5-8.0), they cannot be used under harsh solvent conditions which are prerequisite for denatured proteins, such as 8 M urea and low pH, to study protein folding
- Expensive medium
- Too strong interactions for high-pI proteins

Sample preparation

Phages are rebuffed by washing with the desired buffer and centrifuging at 95,000 rpm (320,000 g) in a table ultracentrifuge for one hour. Supernatant is discarded and phage is resuspended (preferably with a Teflon tube). Washing is repeated twice. The sample volume is adjusted to the desired phage concentration.

Mechanically stressed polyacrylamide gels

Mechanical stress introduces anisotropy into the pores of a gel. Thus solute molecules align by steric clashes with the anisotropic pores in uncharged gels (Sass *et al.*, 2000, Tycko *et al.*, 2000) or by additional electrostatic interactions for charged gels. The pore size and diffusion properties of polyacrylamide gels can be tuned by adjusting the acrylamide and N,N'-methylenebisacrylamide concentration from stocks of 29.2% w/v and 0.78% w/v respectively. A certain mechanical stability of the gels is required for the orientation experiments. Good results are obtained at concentrations of $\geq 4\%$ (w/v) acrylamide. Radial compression can be obtained via a commercially available device (www.newera-nmr.com) (Figure 1.4) where a gel, originally polymerized with a 6 mm diameter, is pressed into the NMR tube of 4.2 mm inner diameter through a Teflon funnel via air pressure from a piston (Chou *et al.*, 2001). Radially compressed gels yield larger alignment than vertically compressed gels. The residual alignment in stressed polyacrylamide gels is steric. However, electrostatic alignment can be obtained if up to 50 % of the acrylamide monomers are replaced by acrylic acid in the polymerization reaction (Meier *et al.*, 2002).

Sample preparation

Polymerization is started in the gel-cylinder (of the device shown in Figure 1.4) sealed with parafilm on one side by the addition of 0.1% w/v ammonium persulfate and 0.5 % w/v TEMED. The gels are pushed out from the gel-cylinder and washed for 5-6 hours at 37 °C with water and dried in a drying oven at 37 °C for several hours (over night). After this process, the gels are dehydrated. These gels are then reswollen in the gel-cylinder with the desired protein solution in buffer for several hours (over night), and then pushed into NMR sample tube. Mechanical stress, in this case, is applied radially as the gels are originally polymerized in a tube of larger diameter than the NMR tube.

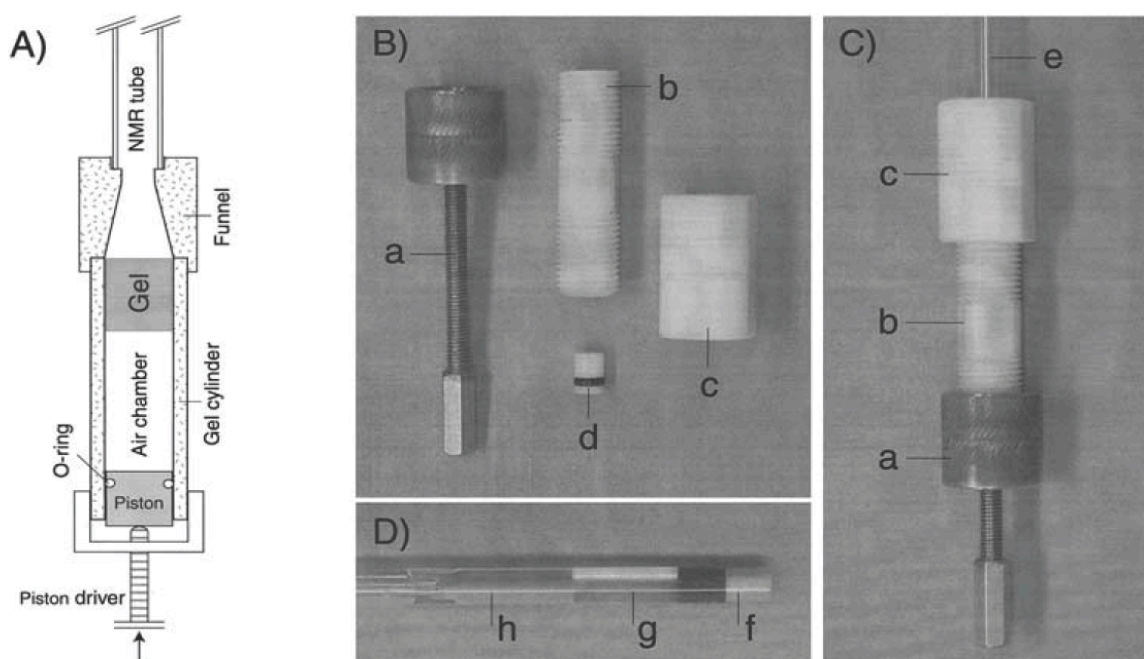


Figure 1.4: Apparatus for stretching the gel and inserting it into the open-ended NMR tube. (A) Schematic drawing. (B, C) Photograph of the disassembled and assembled gel-stretcher. (D) Open-ended NMR tube with the shigemi plunger above the gel. The various components are: (a) Piston driver, (b) gel cylinder, (c) funnel, (d) piston with o-ring, (e) open-ended NMR tube, (f) vespel bottom plug of assembled NMR cell with Teflon sleeve, (g) stretched gel, (h) Shigemi plunger. Detailed dimensions of the gel-stretcher can be downloaded from <http://www.newera-nmr.com>. Taken from (Chou *et al.*, 2001).

Advantages of polyacrylamide gels

- Acrylamide gels are chemically inert and can be used under harsh solvent conditions like 8 M urea to study protein unfolding (Shortle & Ackerman, 2001).
- Protein can be recovered from gels by mincing the gel and placing it in buffer followed by centrifugation and concentration of the supernatant.
- Cheap material used in the preparation.
- Membrane proteins can be studied under charged conditions (Jones & Opella, 2004, Cierpicki & Bushweller, 2004)

Disadvantages of acrylamide gels

- Sometime too strong alignment is observed especially for the larger proteins, leading to line broadening
- Sometimes weak alignment is obtained
- Unwanted signal near amide protons region

Chapter 2:

Solution NMR studies of ABL kinase in complex with three clinical inhibitors

Abstract

The abelson (ABL) tyrosine kinase is an important drug target in the treatment of chronic myelogenous leukemia (CML). Highly effective drugs imatinib (Glivec®; Novartis), nilotinib (Tasigna®; Novartis) and dasatinib (Sprycel®; BMS) have been developed against the oncogenic BCR-ABL fusion protein. Much structural knowledge on ABL-inhibitor complexes has been generated using X-ray crystallography (Schindler *et al.*, 2000, Nagar *et al.*, 2002, Cowan-Jacob *et al.*, 2004), but very little is known about the solution structure and dynamics of ABL kinase.

Preceding this thesis, isotope labeling of ABL kinase catalytic domain in Baculovirus-infected insect cells was performed by our collaborators at Novartis Pharma (Basel), either uniformly or specifically for certain amino acid types.

My work describes the detailed characterization of ABL kinase in complex with all the three aforementioned clinical drugs by high-resolution NMR spectroscopy. To enable solution studies, almost complete backbone resonance assignment for ABL-imatinib complex and partial backbone assignments for ABL kinase in complex with nilotinib and dasatinib was achieved. This allowed structural characterization by chemical shift mapping and residual dipolar couplings (RDCs). The results of this study show that for imatinib and nilotinib complexes the activation loop adopts the inactive conformation, although ^{15}N relaxation studies indicated the dynamics associated with the activation loop. The RDC data clearly show that the dasatinib complex preserves the active conformation, contrary to the predictions based upon molecular modeling. However, line broadening of residues in the activation loop and P-loop indicate presence of microsecond to millisecond dynamics. This study proposes more extensive dynamics in case of complexes in the active state than in the inactive state.

Background

Protein kinases are one of the largest protein families, comprising about 2% of all eukaryotic genes. The chemical activity of a kinase involves removing a γ -phosphate group from ATP and covalently attaching it to one of the three amino acids (serine, threonine or tyrosine) that have a free hydroxyl group. By adding phosphate groups to substrate proteins, they direct the activity, localization and overall function of many proteins, and serve to coordinate the activity of almost all cellular processes. The diversity of essential functions mediated by kinases is shown by the conservation of some 50 distinct kinase families between yeast, invertebrate and mammalian kinomes. Kinases are particularly prominent in signal transduction and co-ordination of complex functions such as the cell cycle. Since kinases have profound effects on a cell, their activity is highly regulated. They are turned on or off by phosphorylation (sometimes by the kinase itself - cis-phosphorylation/- autophosphorylation), by binding of activator or inhibitor proteins, small molecules or by controlling their location in the cell relative to their substrates.

Most kinases act on both serine and threonine, while some act on tyrosine, and a few (dual specificity kinases) act on all three (Coffin *et al.*, 1999). A brief description of the two major classes of enzymes is as follows:

Serine/Threonine kinases

Serine/Threonine protein kinases phosphorylate the OH group of serine or threonine, which have similar sidechains. Activity of these protein kinases can be regulated by specific events (e.g. DNA damage), as well as numerous chemical signals, including cAMP/cGMP, diacylglycerol, and Ca^{2+} . A class of serine/threonine kinase is also regulated by calmodulin. Calmodulin-dependent kinases have a calmodulin binding domain, characterized by a high proportion of basic amino acid residues and having a propensity for formation of an amphiphilic α -helix, residing outside the catalytic domain (Hanks *et al.*, 1988). Most of the members of ser/thr kinase are cytoplasmic proteins. However, serine/threonine kinase domains are also present in the family of mammalian transmembrane receptors (Mathews & Vale, 1991, Lin *et al.*, 1992, Matsuzaki *et al.*,

1993, ten Dijke *et al.*, 1993); members of this class have been shown to bind transforming growth factor- β s (TGF- β s) and activins (Miyazono *et al.*, 1994, ten Dijke *et al.*, 1994).

Some of the subfamilies of ser/thr kinases and their respective function are as follows: Members of cyclic nucleotide-dependent subfamily such as cAMP-dependent-protein kinase regulate metabolism of glycogen, sugar and lipid inside the cell; members of Ca^{2+} /calmodulin-dependent subfamily play role in neurotransmitter secretion, transcription factor regulation, and glycogen metabolism. In contrast to these two, effects of calcium-phospholipid-dependent subfamily are cell-type specific such as for the iris dilator muscle wherein it leads to contraction on phosphorylation and in neurons of the central nervous system where it undergoes neuronal excitement, *etc.* The members of casein kinase subfamily have been involved in cell cycle control, DNA repair, regulation of the circadian rhythm and other cellular processes. The mitogen-activated protein kinases (MAPKs) respond to extracellular stimuli (mitogens) and regulate various cellular activities, such as gene expression, mitosis, differentiation, and cell survival/apoptosis. Members of Raf-Mos family belong to the proto-oncogene subfamily and function to stimulate growth of cells. Raf inhibition has become the target for new anti-metastatic cancer drugs as they inhibit the MAPK cascade and reduce cell proliferation. The AKT subfamily regulates cell proliferation and is important for insulin actions in cells; the members of this subfamily are also responsible for activation of phosphoinositide 3-kinase (PI3-kinase).

Tyrosine kinases

Protein tyrosine kinases (PTKs) can catalyze the transfer of the γ -phosphate from ATP to the hydroxyl group of tyrosine residues both within the kinase itself (autophosphorylation) or other proteins in downstream signaling pathways (Collett *et al.*, 1980, Hunter & Sefton, 1980, Pawson, 1994b). These enzymes are involved in key cell functions such as proliferation, differentiation and anti-apoptotic signaling. They are broadly classified as receptor PTKs and cellular or non-receptor PTKs.

Receptor PTKs possess an extracellular ligand binding domain, a transmembrane domain and an intracellular catalytic domain (van der Geer *et al.*, 1994, Tracy *et al.*, 1995, Hanks

& Hunter, 1995, Taylor *et al.*, 1995). The transmembrane domain anchors the receptor in the plasma membrane, while the extracellular domains bind growth factors. Characteristically, the extracellular domains are comprised of one or more identifiable structural motifs e.g. immunoglobulin-like domains, EGF-like domains, cadherin-like domains, etc. The intracellular kinase domains of receptor PTKs are divided into two classes: those containing a stretch of amino acids separating the kinase domain and those in which the kinase domain is continuous. Signal transduction takes place via phosphorylation that involves a kinase cascade. This cascade leads to amplification of the signal. Proteins that bind to the intracellular domain of receptor tyrosine kinases in a phosphotyrosine-dependent manner and transduce signal include RasGAP, PI3-kinase, phospholipase C gamma, phosphotyrosine phosphatase SHP and adaptor proteins such as Shc, Grb2 and Crk.

In contrast to receptor PTKs, cellular PTKs are located in the cytoplasm, nucleus or anchored to the inner leaflet of the plasma membrane (Pawson, 1994a, Pawson, 1995, Cohen *et al.*, 1995). They are grouped into eight families: SRC, JAK, ABL, FAK, FPS, CSK, SYK and BTK. Each family consists of several members. With the exception of homologous kinase domains (SRC Homology 1, or SH1 domains), and some protein-protein interaction domains (SH2 and SH3 domains), they have few common structural details. Of those cellular PTKs whose functions are known, many, such as SRC, are involved in cell growth. In contrast, FPS PTKs are involved in differentiation, ABL PTKs are involved in growth inhibition, and FAK activity is associated with cell adhesion. Some members of the cytokine receptor pathway interact with JAKs, which phosphorylate the transcription factors, STATs. While there are still other PTKs that activate pathways, the components and functions of which remain undetermined.

Protein tyrosine kinases as targets for inhibitor design

Unregulated activation of these enzymes through mechanisms such as gene amplification, mutation or viral factors can lead to various forms of cancer as well as benign proliferative conditions. Indeed, more than 70% of the known oncogenes and proto-oncogenes involved in cancer, code for PTKs. This identification directs one to concentrate extensively on targeted therapies as a more specific and effective way for

blockade of cancer progression. Currently, there are drugs in clinical trials that target all stages of signal transduction: from the receptor tyrosine kinases that initiate intracellular signaling, through second-messenger generators and kinases involved in signaling cascades, to the kinases that regulate the cell cycle that governs cellular fate. A general mechanism of a tyrosine kinase is shown in Figure 2.1.

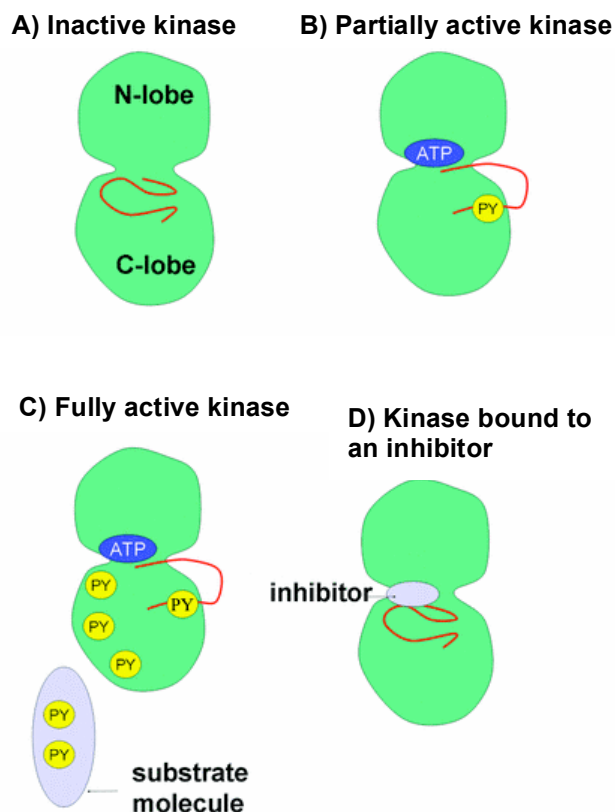


Figure 2.1: Activation mechanism of a tyrosine kinase. A) The N- and C-lobes of the kinase domain of an inactive kinase are indicated with activation loop in red. B) Binding of ATP leads to phosphorylation of tyrosine residues (PY) that results in partial kinase activation and a conformational change of the activation loop. C) Full kinase activation follows with further autophosphorylation and phosphorylation of substrate molecules. D) Binding of an ATP-competitive inhibitor, such as imatinib, prevents kinase phosphorylation and activation. Taken from (Chase & Cross, 2006).

Role of structural biology in drug design

Over the past few years, drug design has profited from structures, which provided clear insight into the mechanism of inhibition. In particular, X-ray crystallography has revealed various active and inactive conformational states of kinases, which are implicated in their regulation and modulation by inhibitors (Huse & Kuriyan, 2002). The active states are

characterized by certain conformations of the activation loop (a centrally located regulatory element in protein kinases), phosphate-binding loop (P-loop) and helix C, which orientate the catalytic machinery to phosphorylate substrates; in the inactive states one or more of these elements are in different conformations, such that substrate binding and/or catalysis cannot occur (Figure 2.2).

Studies on kinases have shown that the catalytic domains of eukaryotic serine/threonine and tyrosine kinases are highly conserved in sequence and structure (Nagar *et al.*, 2002). The catalytic domain has a bilobal structure. The N-lobe contains β -sheet and one conserved α -helix (helix C). The C-lobe is largely helical. At the interface between the two lobes, a number of highly conserved residues form the ATP-binding pocket and the catalytic machinery (Figure 2.2). This ATP-binding site, together with less conserved surrounding pockets, has been the focus of inhibitor design that has exploited differences in kinase structure and flexibility in order to achieve selectivity (Nagar *et al.*, 2002).

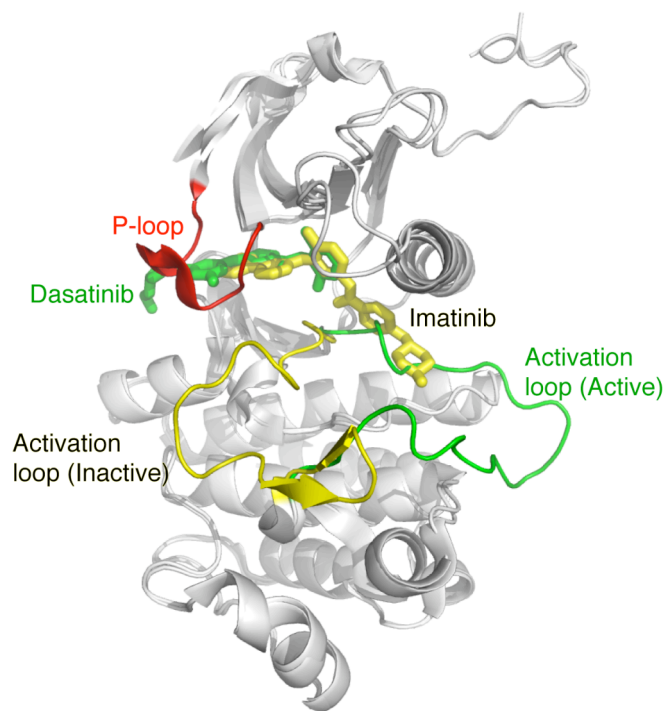


Figure 2.2: ABL kinase architecture in complex with two inhibitors. Complexes of imatinib (inactive; yellow) and dasatinib (active; green) are aligned with respect to the main chain C α atoms.

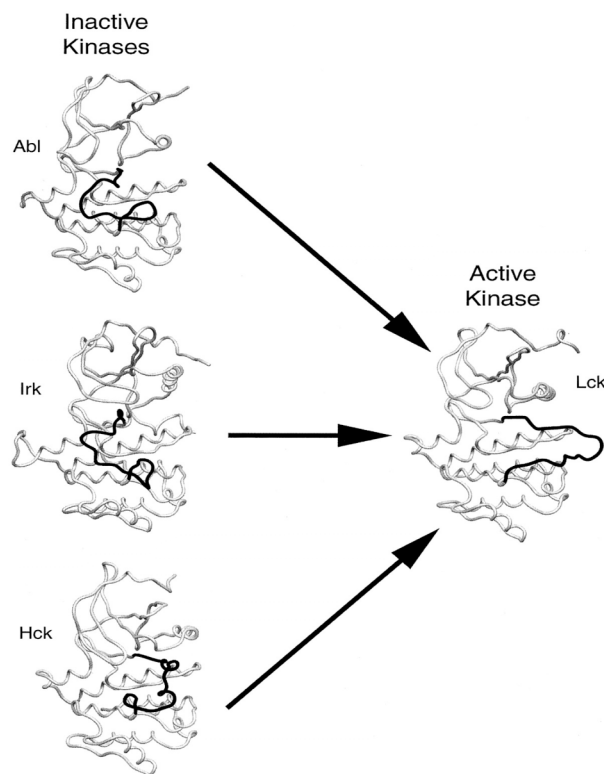


Figure 2.3: Different inactive conformation of kinases. Taken from (Nagar *et al.*, 2002).

Inhibitors can bind to kinase in either its active, ATP-binding state or in its inactive, non-ATP binding state. Inhibitors that bind to the inactive state are often thought to be more specific, as protein kinases display a considerably higher conformational variability in their inactive state (Nagar *et al.*, 2002). Their conformational variability can be the basis for more selective inhibitors, which are difficult to achieve given that all active kinases have an ATP-binding pocket with some high degree of homology (Figure 2.3). Many clinical inhibitors take advantage of this highly specific inactive state of their target.

Abelson tyrosine kinase

Abelson tyrosine kinase is one such important drug target because the expression of constitutively activated BCR-ABL fused oncogene (caused by the reciprocal translocation of genetic material of chromosome 9 and 22) leads to lethal condition of chronic myelogenous leukemia (CML). In BCR-ABL, breakpoint cluster region (BCR) protein replaces the N-terminal autoregulatory domain of the abelson (ABL) protein (Figure 2.4), which deregulates many signal transduction pathways. This altered activation causes resistance to apoptosis, enhanced proliferation and altered adhesion

properties, promoting the premature release of the precursor cells from the bone marrow into circulation.

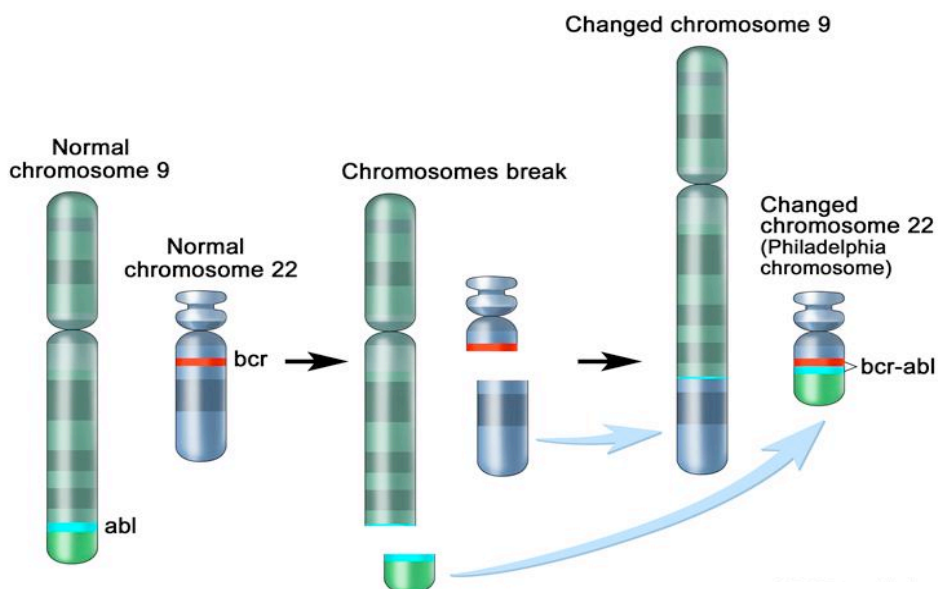
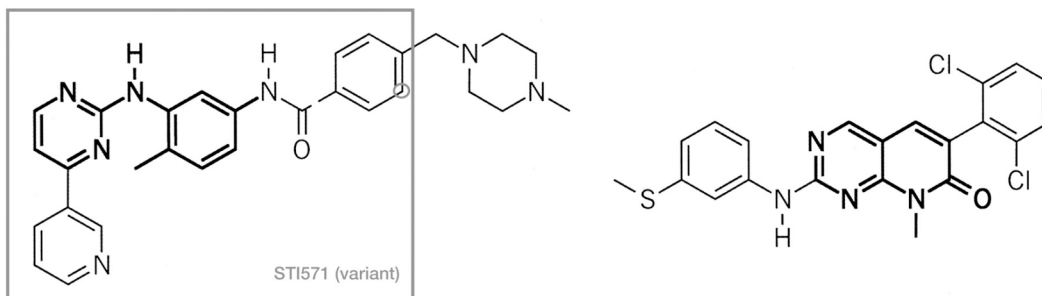


Figure 2.4: Reciprocal translocation of ABL gene with BCR gene. This translocation replaces the autoregulatory N-terminal cap of ABL gene by an oligomerization domain of BCR gene. This results in a constitutively activated ABL kinase, which leads to chronic myelogenous leukemia (CML).

Many small molecule inhibitors have been developed against the aberrantly activated ABL. A series of inhibitors based on the 2-phenylaminopyrimidine class of pharmacophores have been identified that have exceptionally high affinity and specificity for ABL. The presence of a highly polar side-chain (N-methylpiperazine) markedly improved the potency of a member of this class named imatinib (Glivec® or STI-571, identified by Novartis; Figure 2.5), which is currently a frontline therapy against chronic phase CML, gastrointestinal stromal tumours (GIST) and a number of other malignancies. Imatinib also inhibits the ABL protein of non-cancer cells, but cells normally have additional redundant tyrosine kinases that allow them to continue to function even if ABL tyrosine kinase is inhibited. Inhibition of the BCR-ABL tyrosine kinase also stimulates its entry into the nucleus, where it is unable to perform any of its normal anti-apoptotic functions (Vigneri & Wang, 2001). The clinical success of imatinib has enhanced our understanding of the pharmacology of kinase inhibition.



Imatinib (STI-571)

PD173955

Figure 2.5: Chemical structures of imatinib and PD173955. The core compounds from which these two inhibitors were developed are shown in bold lines. In a, a gray box outlines the imatinib variant, and a gray circle denotes the position where a carbon atom is replaced by a nitrogen atom in the variant. Lead compound of both inhibitors are shown in bold lines. Taken from (Nagar *et al.*, 2002)

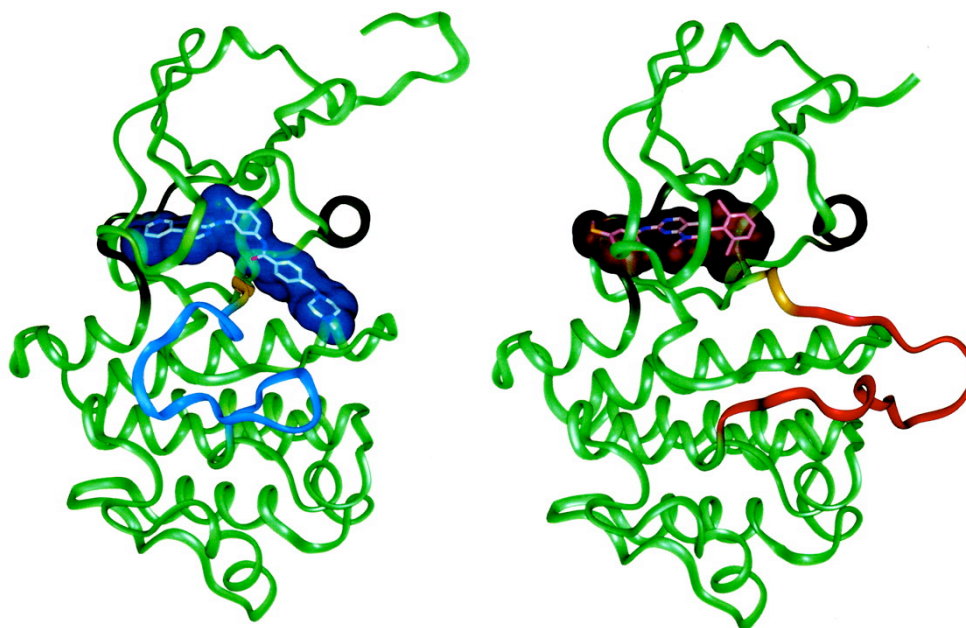


Figure 2.6: Ribbon representation of the structure of the ABL kinase domain (green) in complex with imatinib (left) and PD173955 (right). The activation loops and the van der Waals surfaces corresponding to the inhibitors are colored blue and black for imatinib and PD173955, respectively. The DFG motif situated at the NH2 terminus of the activation loop is colored gold. Helix C and the interlobe connector are colored dark green. Taken from (Nagar *et al.*, 2002)

X-ray crystallographic studies of ABL-inhibitor complexes

Until recently, small molecule inhibitors of ABL kinase that have been discovered almost invariably bind to the kinase domain at the interfacial cleft between the two lobes displacing ATP. These compounds have been shown to bind the kinase in different conformations. For example, crystal studies of ABL bound to PD173955 (Parke-Davis; Figure 2.5), an inhibitor based on the pyrido-[2,3-d]pyrimidine core compounds, show that the activation loop of ABL resembles that of an active kinase (Figure 2.6). In contrast, imatinib (Nagar *et al.*, 2002) and its variant that lacks the piperazinyl group (Schindler *et al.*, 2000) exhibit an inactive bound conformation of the inhibitor (Figure 2.5).

The detailed analysis of the ABL-imatinib complex (Figure 2.6) reveals that the pyrimidine and the pyridine rings of the drug overlap with the ATP-binding site and are surrounded by a hydrophobic cage. The rest of the molecule is wedged between the activation loop and the helix C, which locks the kinase in an inactive conformation. In addition, the normally smooth contour of the phosphate-binding loop of ABL is distorted by imatinib binding, adding further to the unique conformational requirements for optimal kinase inhibition. These conformation-specific binding requirements contribute to imatinib's selectivity, particularly with regard to the closely related kinase SRC, which imatinib does not inhibit.

Structural studies on ABL kinase domain have allowed us to recognize the mechanisms wherein mutant forms of kinase were resistant to imatinib (Gorre *et al.*, 2001, Cowan-Jacob *et al.*, 2004). These studies opened the possibility for the design of second-generation BCR-ABL inhibitors to inhibit wild-type oncoprotein and maintain activity against imatinib-resistant mutants.

Nilotinib (Tasigna® or AMN107; Novartis) and Dasatinib (Sprycel® or BMS-354825; Bristol-Meyer-Squib) (Figure 2.7) are two such targeted drugs developed to treat imatinib-resistant CML. Both drugs inhibit wild type BCR-ABL and all clinically relevant imatinib-resistant mutant forms with the exception of the T315I (gatekeeper) mutation (Shah *et al.*, 2004, Weisberg *et al.*, 2005). Nilotinib is a close analogue of

imatinib with approximately 20-fold higher potency regarding BCR-ABL kinase inhibition. Nilotinib has good clinical efficacy in imatinib-resistant patients and is a well-tolerated drug. Dasatinib has been developed as a dual SRC/ABL inhibitor, but was subsequently shown to affect a wider array of kinases (Figure 2.8). Like nilotinib, it has excellent clinical efficacy and is generally well tolerated. However, dasatinib displays some unique side effects, most notably pleural effusions and cytopenias (Guilhot *et al.*, 2007).

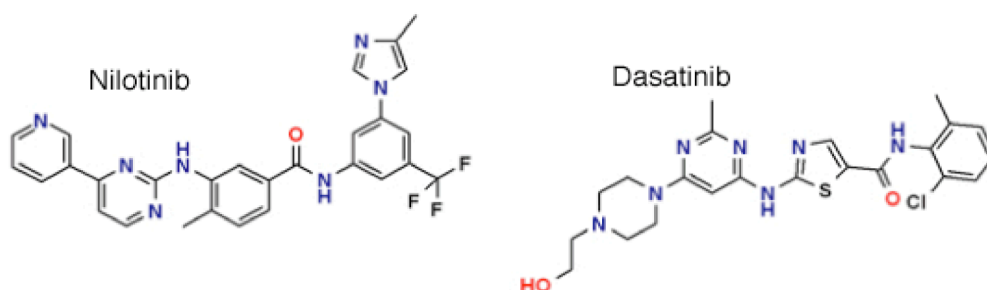


Figure 2.7: Chemical structures of nilotinib and dasatinib. Nilotinib is a member of 2-phenylaminopyrimidine class of compounds whereas dasatinib is a thiazole- and pyrimidine-based kinase inhibitor.

Crystallographic studies of ABL kinase in complex with nilotinib and dasatinib show differences in the binding mode of these inhibitors. These studies demonstrated that nilotinib specifically recognizes an inactive and unphosphorylated conformation of ABL. As expected, this conformation is very similar to that with imatinib. In contrast, dasatinib binds to the active conformation of ABL (Tokarski *et al.*, 2006).

Despite the logic behind the approach of targeting an inactive conformation of a protein kinase, there are also potential advantages of targeting an active conformation. The active conformation requires conservation of the 3D structure and, therefore, is likely to be more potent and less tolerant of resistant mutation. Thus, dasatinib requires less stringent conformational requirements due to its binding to the active conformation. Studies have shown that dasatinib is 300 times more potent than imatinib and offers no resistance to most of the imatinib resistant forms. But, as many kinases have similar active conformations, the specificity of this drug is more limited. One could hypothesize that the

simultaneous inhibition of several targets by dasatinib causes some of the unique side effects.

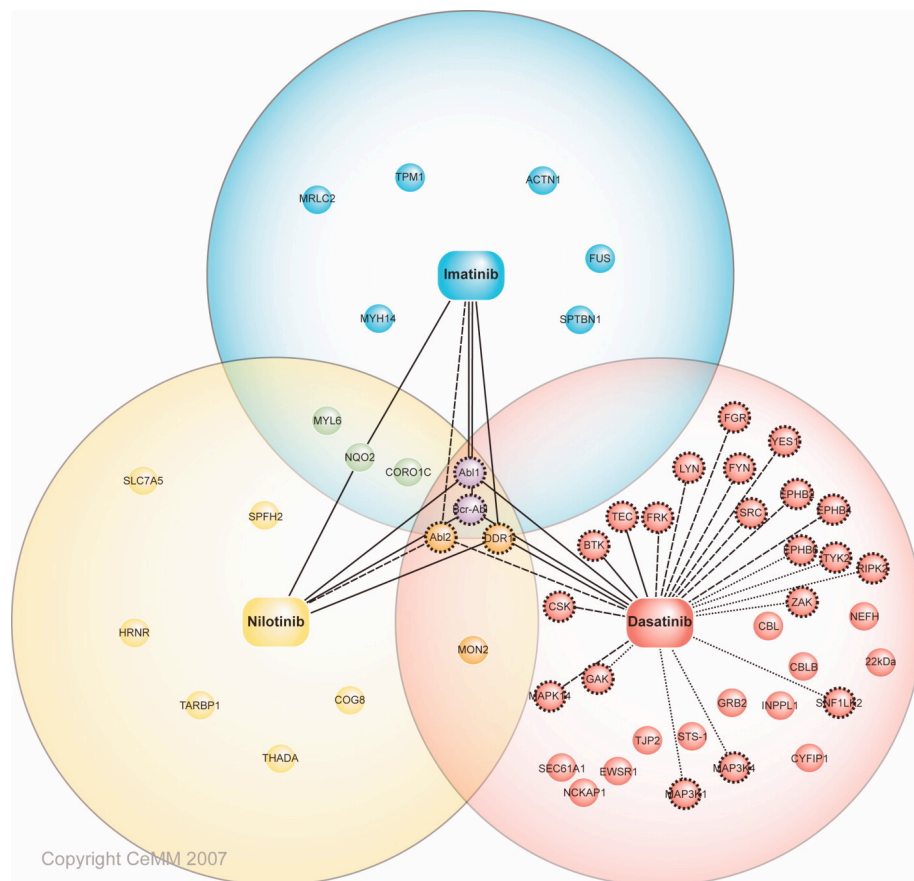


Figure 2.8: Drug interaction network of imatinib, nilotinib and dasatinib. Schematic representation of the observed drug protein interactions in K562 cells for the studied BCR-ABL inhibitors. Proteins are depicted as circles and kinases as dotted rim. Taken from (Eggert & Superti-Furga, 2008).

Recently, a new class of compounds have been identified that potently inhibits the BCR-ABL activity but through a novel allosteric, non-ATP-competitive mechanism. These compounds maintain potency against some clinically relevant imatinib-resistant BCR-ABL mutants. A detailed investigation of these inhibitors is described in (Adrian *et al.*, 2006).

The analysis of the structures of various ABL-inhibitor complexes highlights the important role that the conformational plasticity of protein kinases can play in providing

routes for the development of specific small molecule inhibitors of kinases. Although the hundreds of protein kinases encoded by the human genome are structurally very similar, their conformational dynamics differ. Understanding these dynamics could enable us to design more potent inhibitors. While crystallography does give precise information about the kinase conformation in which the ligand binds, it describes only those molecular states that could be crystallized. The only way to obtain definitive insight into the conformational transitions of ABL kinase domain is likely to be through the use of Nuclear Magnetic Resonance spectroscopy and this was the primary objective of my thesis.

Solution NMR studies on kinases

Nuclear Magnetic Resonance (NMR) spectroscopy is complementary to crystal studies, in that it permits a description of these complexes in physiologically more relevant conditions. In particular, it allows monitoring of a wide range of fast motional processes by relaxation analysis and of local large-scale motions by evaluation of residual dipolar couplings. To date, only a few NMR studies provided some insight on the dynamics of tyrosine kinases. Specifically, the mobility of the Eph receptor (Wiesner *et al.*, 2006) and p38 MAPK kinases (Vogtherr *et al.*, 2006) have been evaluated from chemical shift changes and line broadening effects. Very recently, the dynamics in the p38 MAP kinase-SB203580 complex have been characterized on the basis of residual dipolar couplings (Honndorf *et al.*, 2008). Coincidentally, all the above NMR studies were reported on the kinases that have been expressed in *E. coli*, where efficient methods for the ^{15}N -, $^{13}\text{C}/^{15}\text{N}$ - and $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ -labeling have been described (Bax, 1994, Venters *et al.*, 1995). NMR work on the kinases was mainly limited because of their relatively large size, poor solubility, and the fact that they can often be produced only in expression systems that do not allow cost-effective labeling with ^{13}C , ^{15}N , and ^2H isotopes.

Previous NMR studies on ABL kinase

Prior to this thesis, a number of technical advances by our collaborators at Novartis Pharma (Basel) have made it possible to study ABL kinase complexes in solution by high-resolution NMR techniques. New biochemical techniques, such as amino-acid-type selective isotope labeling with the baculovirus-infected Sf9 insect cell expression system

(Strauss *et al.*, 2003) have been successfully developed at Novartis. While amino-acid type selective isotopic labeling of ABL kinase greatly simplified protein NMR spectra and reduced signal overlap, the few remaining signals could not be assigned since the sequential assignment strategy was not applicable. A strategy was introduced for the resonance assignment of selectively labeled ABL kinase, whereby resonances close to the active site were assigned based on the paramagnetic relaxation enhancement (Cutting *et al.*, 2004). This strategy involved a spin-labeled paramagnetic ligand (an analog of imatinib) for which the three-dimensional structure of the complex is known. It provided some information about the kinase dynamics on the pico- to nanosecond time scale; however, the information about the more relevant micro- to millisecond could not be derived. More importantly, this study was limited by possible inaccurate resonance assignment due to its dependence on the crystal structure especially for the flexible active site, which may not represent the conformational ensemble in solution (Lin, 1999). Thus, to achieve more information for sequential resonance assignment, uniformly ^{15}N - and $^{13}\text{C}/^{15}\text{N}$ -labeled ABL kinase was produced (Strauss *et al.*, 2005). High (>90%) isotope incorporation rates were obtained for all the ABL kinase inhibitor complexes making them amenable to NMR studies.

Experimental strategy and limitations:

In order to interpret structural and dynamic details at atomic level, resonances have to be assigned to their respective position in the protein sequence. A high degree of incorporation of isotope labeling in ABL kinase complexes permitted structural studies by NMR; however, low solubility, slow tumbling and signal overlap limit this study to only the most sensitive state-of-the-art NMR experiments. Line narrowing techniques, such as uniform or partial ^2H labeling, were not attempted because the insect cell expression system does not allow cost-effective ^2H -labeling. For the ABL kinase complex, ambiguity in the resonance assignment of backbone CA could not be resolved by the combination of HNCA, HN(CO)CA and ^{15}N -edited NOESY. Supplemental information for this system was therefore taken from amino-acid type selectively isotope labeled samples.

Structural characterization of ABL kinase complexes by high-resolution solution NMR techniques

Based on previous NMR studies on ABL, my work demonstrates detailed investigation of ABL-inhibitor complexes by NMR spectroscopy. To enable solution studies, almost complete backbone resonance assignment was achieved for ABL-imatinib complex. This backbone resonance assignment was difficult, though a prerequisite for structural studies by NMR. Missing assignments are due to exchange broadening beyond detection in the NMR spectra and cluster in the flexible activation loop with an exception of the N-terminal Gly225 and His361 lining the ABL-imatinib binding surface (Figure 2.9). All other key residues, which are relevant for characterizing the activation loop conformation and dynamics could thus be assigned. Secondary chemical shifts agree with the secondary structure and thus support the assignment (Figure 2.10). The assignments have been deposited in the BMRB under accession number 15488 and have been published in the Biomolecular NMR Assignments (Vajpai *et al.*, 2008a).

The above results formed the basis of detailed structural analysis. Partial backbone resonance assignments were achieved for nilotinib, dasatinib and XCD710 complexes, which allowed residual dipolar couplings and ^{15}N relaxation studies to characterize the solution conformation and dynamics of ABL-inhibitor complexes. These studies clearly demonstrate that in solution, the conformational ensemble of imatinib and nilotinib complex closely resembles the inactive state of the crystal structure, although relaxation studies and line broadening of some resonances in the activation loop indicate presence of residual dynamics. For the dasatinib complex, RDC data show that the ensemble of solution conformations is close to the active conformation as determined in the crystal. However, line broadening of residues in the activation loop and P-loop indicate presence of microsecond to millisecond dynamics.

The new data provide insights into the structural dynamics of ABL protein kinase and help clarify its physiologically relevant binding modes. The detailed results on the characterization of ABL kinase inhibitor complexes have been published in the Journal of Biological Chemistry (Vajpai *et al.*, 2008b).

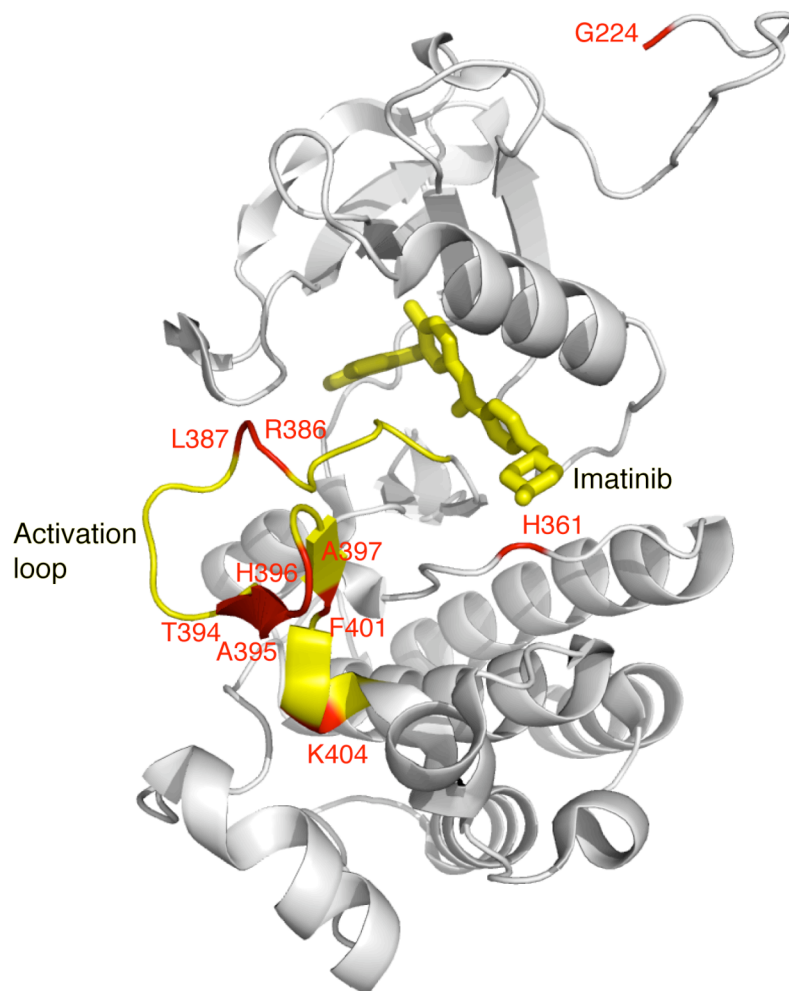


Figure 2.9: Crystal structure of ABL kinase in complex with imatinib (PDB entry: 1IEP). Unassigned amino acids (labeled in red), excluding proline, are mapped on the crystal structure. Activation loop and imatinib are colored in yellow.

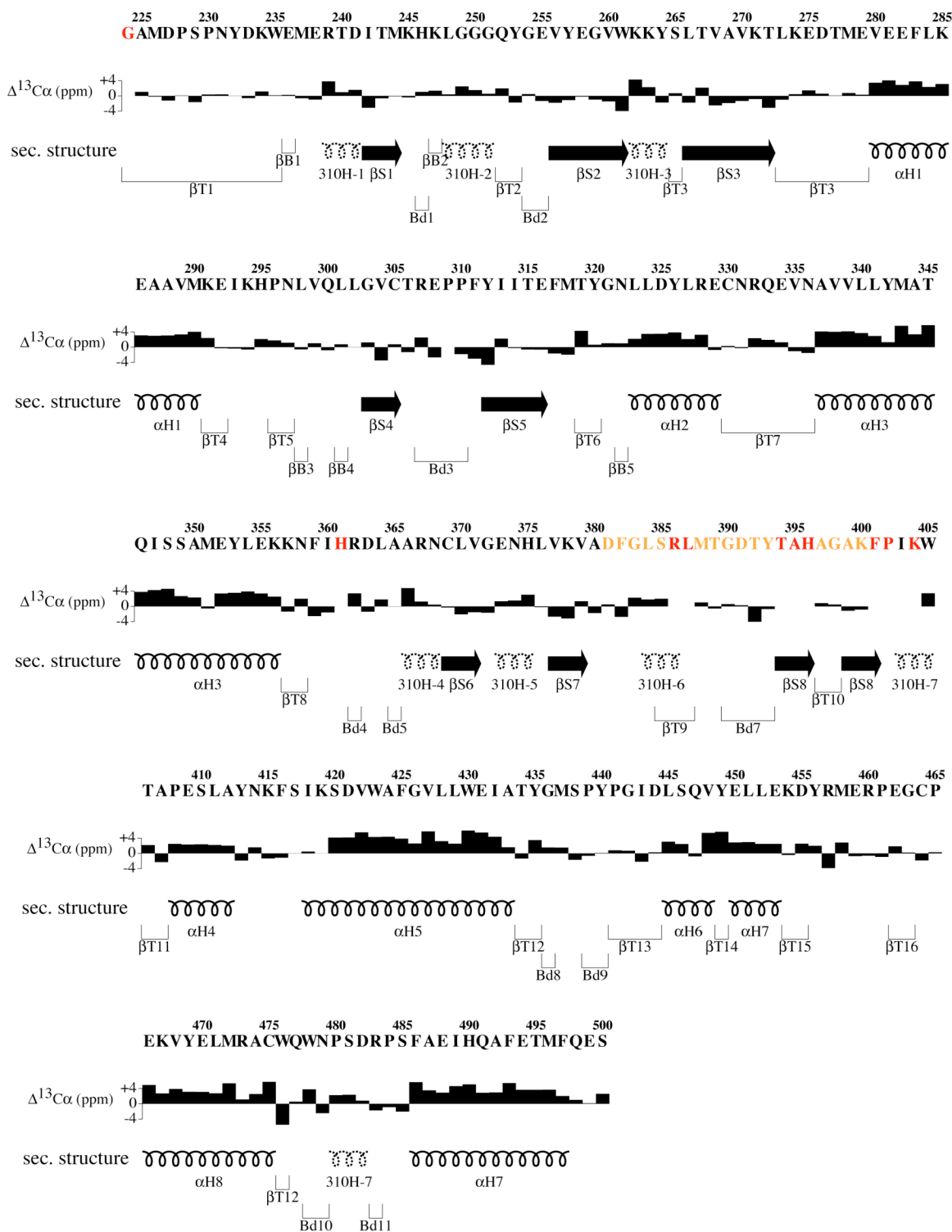


Figure 2.10: Secondary structure and secondary chemical shifts match in the ABL-imatinib complex backbone resonance assignment. Amino acids lacking backbone resonance assignments are shown in red. Activation loop is shown in orange (D381-P402). Random coil shifts are taken from (Spera & Bax, 1991).

Original Publications

P.W. Manley, S.W. Jacob, G. Fendrich, A. Strauss, **N. Vajpai**, S. Grzesiek, W. Jahnke

Bcr-Abl binding modes of dasatinib, imatinib and nilotinib: An NMR study

Blood 2006 **108**: 747 (ASH Annual Meeting Abstracts)

Vajpai N., Strauss A., Fenderich G., Manley P.W., Jacob S., Jahnke W., and Grzesiek S.

Backbone NMR resonance assignment of the Abelson kinase domain in complex with imatinib.

Biomol NMR Assgn 2008 **2**: 41-42

Vajpai N., Strauss A., Fenderich G., Manley P.W., Jacob S., Grzesiek S., and Jahnke W.

Solution conformations and dynamics of ABL kinase inhibitor complexes determined by NMR substantiate the different binding modes of imatinib/nilotinib and dasatinib.

J Biol Chem 2008 **283**: 18292-18302

blood

Leading the world in reporting basic and applied
hematology research

JOURNAL OF
THE AMERICAN
SOCIETY OF
HEMATOLOGY

[Home](#) | [About 'Blood'](#) | [Authors](#) | [Subscriptions](#) | [Permissions](#) | [Advertising](#) | [Public Access](#) | [Contact Us](#)

Institution: KANTONSSPITAL | [Sign In via User Name/Password](#)

SEARCH: [Blood \(ASH Annual Meeting Abstracts\) 2006 108: Abstract 747](#)
© 2006 [American Society of Hematology](#)

GO

Current Issue

First Edition

Archives

Submit to Blood

Search Blood

ASH™

Meeting Abstracts

E-Mail Alerts

Oral Sessions

Bcr-Abl Binding Modes of Dasatinib, Imatinib and Nilotinib: An NMR Study.

Paul W. Manley¹, Sandra W. Cowan-Jacob^{1,*}, Gabriele Fendrich^{1,*},
André Strauss^{1,*}, Navratna Vapai^{2,*}, Stephan Grzesiek^{2,*} and
Wolfgang Jahnke^{1,*}

¹ Novartis Institutes for BioMedical Research, Novartis Pharma AG, Basel, Switzerland and ² Structural Biology Department, University of Basel, Switzerland.

Abstract

Following the discovery that point mutations in the kinase domain of Bcr-Abl reduce the binding affinity of imatinib and lead to drug resistance in CML patients, efforts have been directed towards the discovery of new drugs which inhibit these resistant enzymes. Two such agents are dasatinib and nilotinib. Whereas, like imatinib, x-ray analysis of crystal structures of nilotinib in complex with the Abl kinase domain reveal that this agent binds to an inactive, DFG-out conformation of the enzyme, similar studies have shown that dasatinib binds to the catalytically active state of the enzyme (Tokarski *et al*, Cancer Res. 2006). However, based upon *in silico* methods using homology models of the imatinib-binding inactive conformation of Abl some reports claim that dasatinib is capable of binding to both the active and inactive forms of the enzyme. To help address this conundrum we have employed nuclear magnetic resonance (NMR) spectroscopy to study the different conformational characteristics and dynamic changes of the Abl protein obtained upon adding ligands.

Selectively isotope labelled (¹⁵N and ¹³C) Abl kinase in the unphosphorylated state was produced according to published methods (Strauss *et al*, J. Biomolecular NMR 2005). Chemical shift data of the protein in solution were recorded by NMR spectroscopy, both with and without ligand. By measuring residual dipolar couplings (RDC) between the back-bone amide nitrogen and hydrogen atoms of amino-acid residues in the vicinity of the conserved DFG-motif (residues 370 – 410), the conformational states and equilibria of the activation loop of the kinase were established. Upon adding imatinib to the unliganded Abl, both chemical shift and RDC data show characteristic signals for residues M388, Y393 and G398, which are entirely consistent with the conformational equilibrium moving to the inactive state, in which the activation loop adopts the DFG-out conformation. In the case of nilotinib, NMR spectroscopy revealed chemical shift patterns and couplings involving the same three residues confirming that the drug binds to the same inactive conformation as imatinib, both in solution as well as in the crystalline state. In contrast, upon adding dasatinib to Abl, NMR data of the complex show distinctly different chemical shifts and RDC values, confirming that the protein assumes a different conformational state, with the activation loop adopting the active conformation, in accordance with the crystallographic evidence. Even upon adding dasatinib to a complex of imatinib with Abl in the inactive conformation, the kinase conformation changed to a state indistinguishable to that observed upon adding dasatinib to unliganded protein.

In conclusion, these studies show that the three Abl kinase inhibitors all interact with the protein in solution with the same binding modes as observed in x-ray crystallographic studies, but show no evidence for dasatinib binding to the inactive "DFG-out" conformation. This case study also demonstrates the power of NMR spectroscopy in evaluating solution structures of ligand-protein complexes. Further experiments are in progress evaluating other structurally different Abl kinase inhibitors.

Footnotes

* Corresponding author

Disclosures: Paul W. Manley, Sandra W. Cowan-Jacob, Gabriele Fendrich, André Strauss and Wolfgang Jahnke are full-time employees of Novartis Pharma Ltd.; Stephan Grzesiek has received Research Funding from Novartis Pharma Ltd.

[CiteULike](#) [Connotea](#) [Del.icio.us](#) [Digg](#) [Reddit](#) [Technorati](#) [What's this?](#)

Click for information regarding free online access to various full-text *Blood* articles

[Home](#) | [About 'Blood'](#) | [Authors](#) | [Subscriptions](#) | [Permissions](#) | [Advertising](#) | [Public Access](#) | [Contact Us](#)

Genentech
BIO@NCOLOGY™

biogen idec

BloodOnline is supported in part by
[Genentech BioOncology](#) and [Biogen Idec](#)

Copyright © 2006 by American Society of Hematology Online ISSN: 1528-0020

Backbone NMR resonance assignment of the Abelson kinase domain in complex with imatinib

Navratna Vajpai · André Strauss · Gabriele Fendrich ·
Sandra W. Cowan-Jacob · Paul W. Manley ·
Wolfgang Jahnke · Stephan Grzesiek

Received: 26 November 2007 / Accepted: 9 January 2008
© Springer Science+Business Media B.V. 2008

Abstract Imatinib (Glivec or Gleevec) potently inhibits the tyrosine kinase activity of BCR-ABL, a constitutively activated kinase, which causes chronic myelogenous leukemia (CML). Here we report the first almost complete backbone assignment of c-ABL kinase domain in complex with imatinib.

Keywords Tyrosine kinases · Glivec · BCR-ABL · Chronic myelogenous leukemia · Selective labeling

Biological context

Tyrosine kinases are important mediators in signal transduction pathways and are tightly regulated by several mechanisms. Aberrant activation of these enzymes may lead to diseases, and mutations in kinase genes are implicated in many forms of cancer (Greenman et al. 2007). A number of tyrosine kinases are therefore attractive drug targets for the discovery of inhibitors, which can modulate the activity of these enzymes. The BCR-ABL fusion protein, having a constitutively activated Abelson (ABL)

kinase domain, is one such important drug target, which has been clinically validated by imatinib (Glivec; Novartis Pharma AG), an efficacious and well-tolerated treatment for chronic myelogenous leukemia (CML) (Ren 2005). Much structural knowledge on ABL-inhibitor complexes has been generated using X-ray crystallography (Nagar et al. 2002; Cowan-Jacob et al. 2007), but very little is known about the solution structure and dynamics of ABL kinase.

To enable NMR studies on the solution behavior of ABL kinase, we have obtained backbone resonance assignment of ABL kinase domain in complex with imatinib. The selected construct of the ABL kinase domain (denoted here as ABL, residues GAMDP-S229-S500) contains 277 residues and has a typical kinase bilobal structure, with the N-terminal lobe containing β -sheets and the conserved helix C, and the C-terminal lobe being mainly helical. At the interface of the two lobes is the ATP-binding pocket, which is also the binding site of most ABL inhibitors.

Methods and materials

Since ABL is not readily expressed in *E. coli*, we have developed a protocol for the $^{13}\text{C}/^{15}\text{N}$ isotope labeling of the ABL kinase catalytic domain in Baculovirus-infected insect cells, either uniformly or selectively for certain amino acid types (Strauss et al. 2003; Strauss et al. 2005).

Uniformly $^{13}\text{C},^{15}\text{N}$ or selectively labeled samples (Strauss et al. 2003; Strauss et al. 2005) of the ABL-imatinib complex were prepared as 0.4 mM solutions in 250 μl (Shigemi microtubes) of buffer containing 95% H_2O , 5% D_2O , 20 mM BisTris, 100 mM NaCl, 2 mM EDTA, 3 mM DTT at pH 6.5 with an ABL:imatinib ratio of 1:1. NMR spectra were recorded at 293 K on Bruker

Electronic supplementary material The online version of this article (doi:10.1007/s12104-008-9079-7) contains supplementary material, which is available to authorized users.

N. Vajpai · S. Grzesiek (✉)
Biozentrum, University of Basel, Klingelbergstrasse 70,
Basel, Switzerland
e-mail: stephan.grzesiek@unibas.ch

A. Strauss · G. Fendrich · S. W. Cowan-Jacob ·
P. W. Manley · W. Jahnke (✉)
Novartis Institutes for Biomedical Research, Basel, Switzerland
e-mail: wolfgang.jahnke@novartis.com

DRX 600 (cryo or room temperature probe) or DRX 800 MHz (cryoprobe) spectrometers. Due to the lack of deuteration, the sensitivity of NMR experiments involving $^{13}\text{C}^\beta$ or other side chain nuclei was very low. Therefore, backbone assignments had to be performed with non-TROSY versions of the HNCO, HNCA, HN(CO)CA (Grzesiek and Bax 1992), and an ^{15}N -edited ^1H - ^1H NOESY. The assignment was aided and verified by spectra of a total of 15 selectively labeled samples that had ^{15}N , $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ nuclei introduced specifically for certain amino acid types (see supplementary material Table 1). All NMR data were processed and analysed using the NMRPipe (Delaglio et al. 1995) and NMRView (Johnson and Blevins 1994) software suites. ^1H , ^{15}N and ^{13}C are referenced relative to the frequency of the ^2H lock resonance of water.

Extent of assignment and data deposition

The achieved assignments comprise 96% of all backbone $^1\text{H}^N$, ^{15}N , $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ resonances, covering 254 of the

264 non-proline residues (Fig. 1). Unassigned residues consist of the *N*-terminal glycine, seven residues within the activation loop (consisting of residues D381-P402), and H361. Line broadening of adjacent residues indicates that most of the missing residues are broadened beyond detection due to intermediate conformational exchange. Despite the difficulty in observing the entire activation loop, the assignments include at least 14 key residues involved in ligand binding. The assignments have been deposited in the BMRB (accession number 15488).

Acknowledgements We would like to thank Drs. Sonja Alexandra Dames and Martin Allan for their help during the initial phase of the project. This work was supported by SNF grant 31-109712 (S.G.) and Novartis Pharma AG.

References

- Cowan-Jacob SW, Fendrich G, Floersheimer A, Furet P, Liebetanz J, Rummel G, Rheinberger P, Centeleghe M, Fabbro D, Manley PW (2007) Structural biology contributions to the discovery of drugs to treat chronic myelogenous leukaemia. *Acta Crystallogr D Biol Crystallogr* 63:80–93
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe – a multidimensional spectral processing system based on unix pipes. *J Biomol NMR* 6:277–293
- Greenman C, Stephens P, Smith R et al (2007) Patterns of somatic mutation in human cancer genomes. *Nature* 446:153–158
- Grzesiek S, Bax A (1992) Improved 3D triple-resonance NMR techniques applied to a 31-kDa protein. *J Magn Reson* 96:432–440
- Johnson BA, Blevins RA (1994) NMR View – a computer-program for the visualization and analysis of NMR data. *J Biomol NMR* 4:603–614
- Nagar B, Bornmann WG, Pellicena P, Schindler T, Veach D R, Miller WT, Clarkon B, Kuriyan J (2002) Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). *Cancer Res* 62:4236–4243
- Pervushin K, Riek R, Wider G, Wuthrich K (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci USA* 94:12366–12371
- Ren R (2005) Mechanisms of BCR-ABL in the pathogenesis of chronic myelogenous leukaemia. *Nat Rev Cancer* 5:172–183
- Strauss A, Bitsch F, Cutting B, Fendrich G, Graff P, Liebetanz J, Zurini M, Jahnke W (2003) Amino-acid-type selective isotope labeling of proteins expressed in Baculovirus-infected insect cells useful for NMR studies. *J Biomol NMR* 26:367–372
- Strauss A, Bitsch F, Fendrich G, Graff P, Knecht R, Meyhack B, Jahnke W (2005) Efficient uniform isotope labeling of Abl kinase expressed in Baculovirus-infected insect cells. *J Biomol NMR* 31:343–349

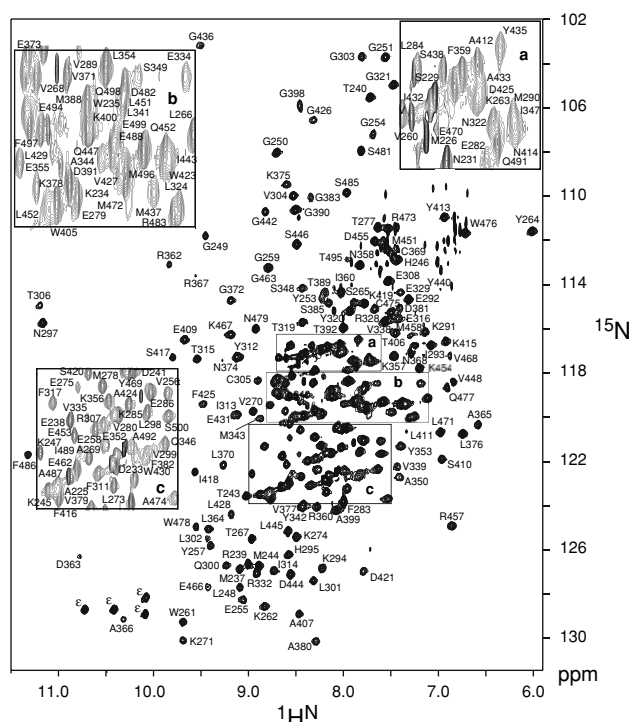


Fig. 1 $^1\text{H}^N$ - ^{15}N HSQC-TROSY (Pervushin et al. 1997) spectrum of uniformly ^{15}N -labeled ABL kinase domain-imatinib complex with assignment information. An 'ε' denotes unassigned W^ϵ NH sidechain resonances, Subpanels a, b and c show enlarged regions of the respective rectangular boxes in the main spectrum

Backbone NMR resonance assignment of the Abelson kinase domain in complex with imatinib

Navratna Vajpai, André Strauss, Gabriele Fendrich, Sandra W. Cowan-Jacob, Paul W. Manley, Wolfgang Jahnke*, Stephan Grzesiek*

Novartis Institutes for Biomedical Research, Basel, Switzerland

Biozentrum, University of Basel, Klingelbergstrasse 70, Basel, Switzerland

**To whom correspondence should be addressed at stephan.grzesiek@unibas.ch or wolfgang.jahnke@novartis.com*

Supplementary Material

Table S1: Isotope labeled amino acids of Abl samples and labeling culture conditions used for the assignments

Sample Name	U- ¹⁵ N	U- ¹³ C	¹³ C ^α	¹³ CO	Medium ¹	Insect cells ²	Imatinib added ³
15N-Tyr	Y				SF-4/C-x	Sf21	-
15N-Val	V				SF-4/C-x	Sf21	-
15N-Gly	G				SF-4/C-x	Sf9	+
15N-Met	M				SF-4/C-x	Sf9	+
15N-Phe	F				SF-4/C-x	Sf9	+
15N-Ile	I				SF-4/C-x	Sf9	+
15N-Leu	L				SF-4/C-x	Sf9	+
15N-Thr	T				SF-4/C-x	Sf9	+
15N-Ala	A				SF-4/C-x	Sf9	+
15N-Trp	W				SF-4/C-x	Sf9	+
15N/13C-Gly	G	G			SF-4/C-x	Sf9	+
DFG1	GMLY		G	F	SF-4/C-x	Sf9	+
DFG2	LVI	G			SF-4/C-x	Sf9	+
LIMY	LIMY	LIMY			BE2000-CN-x	Sf9	+
FGMY	FGMY	FGMY		LT	BE2000-CN- x	Sf9	-
13C/15N-uniform	All	All			BE2000 -CN	Sf9	+

¹Labeling medium for expression as given in (Strauss et al., 2003; Strauss et al., 2005): SF-4/C-x, BE2000-CN-x, BE2000-CN; x denotes the labeled amino acids

² Insect cells for expression.

³ Addition of 10-20 μ M imatinib to the culture medium to stabilize the expressed ABL protein and to suppress phosphorylation

Solution Conformations and Dynamics of ABL Kinase-Inhibitor Complexes Determined by NMR Substantiate the Different Binding Modes of Imatinib/Nilotinib and Dasatinib^{*†‡}

Received for publication, February 20, 2008, and in revised form, April 18, 2008. Published, JBC Papers in Press, April 22, 2008, DOI 10.1074/jbc.M801337200

Navratna Vajpai[‡], André Strauss[§], Gabriele Fendrich[§], Sandra W. Cowan-Jacob[§], Paul W. Manley[§], Stephan Grzesiek^{‡1}, and Wolfgang Jahnke^{§2}

From the [§]Novartis Institutes for BioMedical Research, 4002 Basel, Switzerland and the [‡]Biozentrum, University of Basel, 4056 Basel, Switzerland

Current structural understanding of kinases is largely based on x-ray crystallographic studies, whereas very little data exist on the conformations and dynamics that kinases adopt in the solution state. ABL kinase is an important drug target in the treatment of chronic myelogenous leukemia. Here, we present the first characterization of ABL kinase in complex with three clinical inhibitors (imatinib, nilotinib, and dasatinib) by modern solution NMR techniques. Structural and dynamical results were derived from complete backbone resonance assignments, experimental residual dipolar couplings, and ¹⁵N relaxation data. Residual dipolar coupling data on the imatinib and nilotinib complexes show that the activation loop adopts the inactive conformation, whereas the dasatinib complex preserves the active conformation, which does not support contrary predictions based upon molecular modeling. Nanosecond as well as microsecond dynamics can be detected for certain residues in the activation loop in the inactive and active conformation complexes.

Protein kinases play critical roles in intracellular signal transduction pathways, deregulation of which can lead to a variety of pathological states and diseases such as cancer. These enzymes are therefore tightly regulated with multiple layers of control, including phosphorylation, myristoylation, and interaction with SH2³ and SH3 or other regulatory domains. Modulation of kinase activity by therapeutic agents is a clinically validated concept, with many kinases considered to be attractive drug

targets. ABL kinase is such a target because the expression of the BCR-ABL fusion protein (caused by unfaithful repair of DNA strand breaks in bone marrow hematopoietic stem cells and subsequent t(9,22) chromosome translocation) leads to life-threatening chronic myelogenous leukemia (1, 2). In BCR-ABL, the breakpoint cluster region BCR protein replaces the N-terminal autoregulatory domain of the Abelson ABL protein to give a constitutively activated tyrosine kinase, which deregulates signal transduction pathways, causing uncontrolled proliferation and impaired differentiation of progenitor cells.

X-ray crystallography has revealed various active and inactive conformational states of kinases, which are implicated in their regulation and modulation by inhibitors (3). The active states are characterized by certain conformations of the activation loop, phosphate-binding loop (P-loop), and helix C, which orient the catalytic machinery to phosphorylate substrates; in the inactive states, one or more of these elements are in different conformations, such that substrate binding and/or catalysis cannot occur. An important determinant is the orientation of the conserved Asp-Phe-Gly motif within the activation loop. For efficient catalysis, this motif adopts a “DFG-in” conformation. In contrast, the “DFG-out” conformation has this motif displaced from the orientation needed for binding the substrate ATP to phosphorylate and activate downstream signaling proteins. Such a DFG-out conformation has been observed in many inactive kinases, including ABL, IRK, KIT, and FLT3 tyrosine kinases (4–7) as well as the serine/threonine kinases p38 MAPK and BRAF (8, 9).

Different kinase inhibitors can bind to and stabilize different kinase conformations, as exemplified in Fig. 1 for different ABL inhibitors. Crystallographic studies have shown that the tyrosine kinase inhibitor imatinib (Glivec®/Gleevec®), a highly effective treatment for chronic phase chronic myelogenous leukemia (10), binds within the catalytic site of the inactive form of ABL with the activation loop in a DFG-out conformation (6, 11–15). This conformation is very similar to that with nilotinib (16) (Tasigna®), a more potent and selective ABL inhibitor developed to inhibit imatinib-resistant mutant forms of BCR-ABL, which frequently emerge in advanced stages of chronic myelogenous leukemia and lead to relapse and disease progression (17). In contrast, crystallographic studies have shown that the multi-targeted ABL and SRC family kinase inhibitor dasat-

* This work was supported by Swiss National Science Foundation Grant 31-109712 (to S. G.) and Novartis Pharma AG. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† This article was selected as a Paper of the Week.

‡ The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Fig. 1.

¹ To whom correspondence may be addressed. E-mail: stephan.grzesiek@unibas.ch.

² To whom correspondence may be addressed. E-mail: wolfgang.jahnke@novartis.com.

³ The abbreviations used are: SH, Src homology; MAPK, mitogen-activated protein kinase; BisTris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; TCEP, tris(2-carboxyethyl)phosphine; RDC, residual dipolar coupling; NOE, nuclear Overhauser effect; HSQC, heteronuclear single quantum coherence.

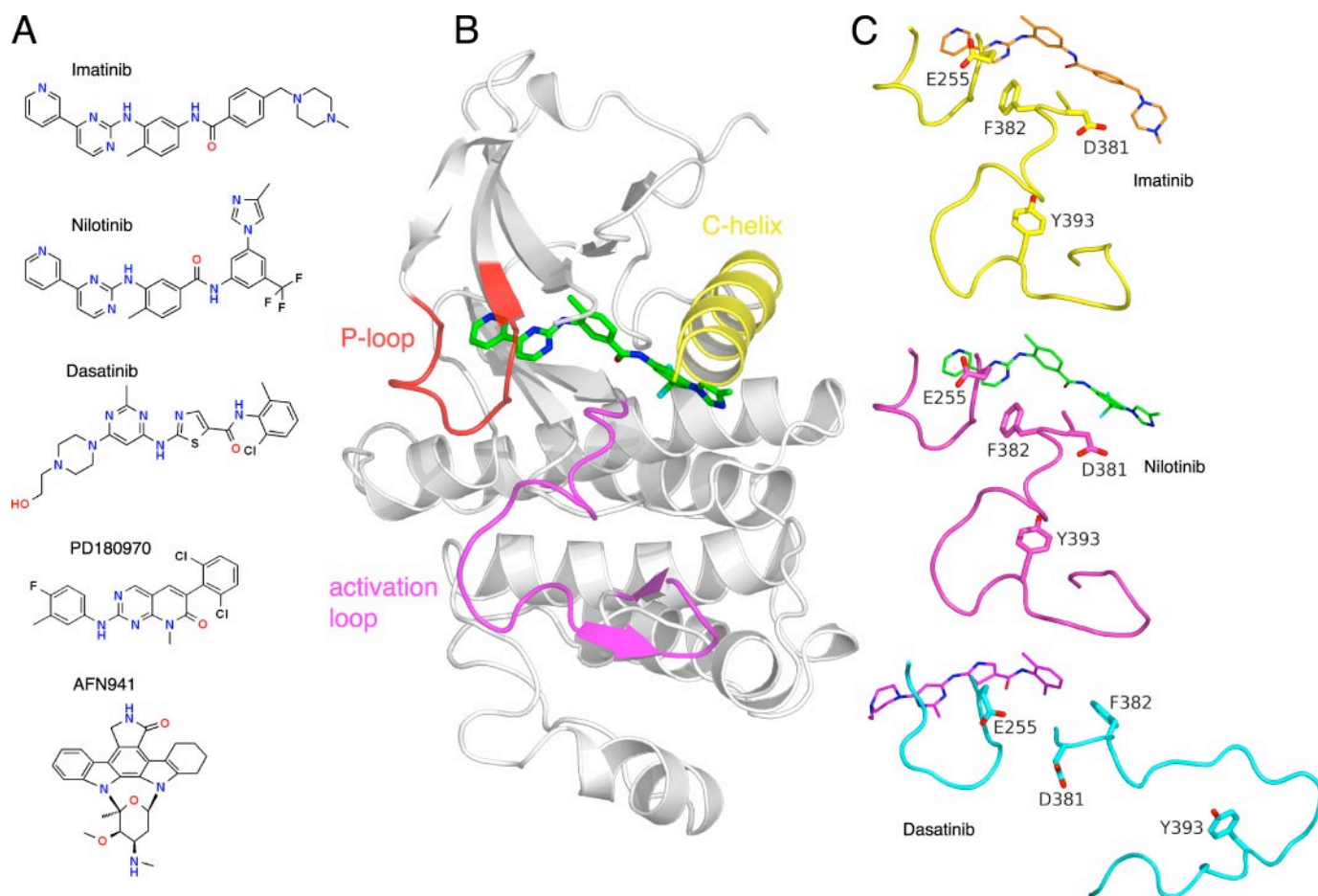


FIGURE 1. Chemical structure of ABL kinase and its inhibitors. *A*, chemical structures of the compounds discussed in this work: imatinib, nilotinib, dasatinib, PD180970, and AFN941 (tetrahydrostaurosporine). *B*, ribbon diagram showing ABL kinase (gray) with nilotinib (green carbons) bound, highlighting the P-loop (red), the activation loop (magenta), and helix C (yellow). *C*, details of the conformations of the P-loop and activation loop with imatinib bound (yellow, ABL; gold, imatinib) (Protein Data Bank code 1IEP), with nilotinib bound (magenta, ABL; green, nilotinib) (Protein Data Bank code 3CS9), and with dasatinib bound (cyan, ABL; magenta, dasatinib) (Protein Data Bank code 2GQG). All three structures are shown in the same orientation based on superposition of the protein coordinates. Residues highlighted in the activation loops are Asp³⁸¹, Phe³⁸², and Tyr³⁹³. In the P-loop, Glu²⁵⁵ is shown as an indication of the difference in conformation between the active and inactive ABL conformations.

inib (SPRYCEL[®]) (18) binds to the active DFG-in conformation of ABL (19). However, based on molecular modeling (19, 20), it has been hypothesized that dasatinib can also bind to the inactive DFG-out conformation, and even without experimental support, this notion is becoming established. Crystallographic studies have shown that other promiscuous kinase inhibitors also bind to active conformations of ABL, including compounds of the pyrido[2,3-*d*]pyrimidine class (21–23) such as PD180970 and the staurosporine derivative AFN941 (11). A clear understanding of the physiologically relevant binding modes of BCR-ABL inhibitor complexes is of utmost importance for the rational design of potent and selective inhibitors that can also counteract the emergence of drug-resistant mutant forms of BCR-ABL (12).

Crystallographic analysis is limited to those biomolecular states that crystallize, which may not capture the full ensemble of conformations that are available in solution under physiologically more relevant conditions. These crystallographic states may be artificially stabilized by crystal contacts while highly dynamical parts of structures remain invisible. In principle, NMR spectroscopy can provide the missing characterization of conformational ensembles and the dynamics of their intercon-

version. However, NMR work on kinases has been severely limited by their relatively large size, poor solubility, and the fact that they can often be produced only in expression systems that do not allow cost-effective labeling by ¹³C, ¹⁵N, and ²H isotopes. Only recently have a few NMR studies provided some limited insight on the dynamics of the Eph receptor (24) and p38 MAPK (25) kinases from chemical shift changes and line broadening effects.

In this study, we have applied new techniques such as expression of isotopically labeled ABL kinase in baculovirus-infected insect cells and residual dipolar couplings, which provide precise geometrical information, to characterize the solution conformations and dynamics of the ABL kinase domain in complex with the three clinically used inhibitors: imatinib, nilotinib, and dasatinib.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—Expression and purification of uniform and amino acid-selective ¹³C/¹⁵N isotope-labeled ABL kinase in baculovirus-infected insect cells were carried out as described previously using the construct His₆-TEVsite-GAMDP-hABL(Ser²²⁹–Ser⁵⁰⁰) (26, 27).

Solution Conformations of ABL in Complex with Inhibitors

Selectively [$U\text{-}^{13}\text{C}/^{15}\text{N}$]Phe-Gly-Met-Tyr (FGMY)- and [^{13}CO]Leu-Thr-labeled ABL kinase was expressed in custom-made BioExpress 2000 medium containing [$U\text{-}^{13}\text{C}/^{15}\text{N}$]Phe-Gly-Met-Tyr and [^{13}CO]Leu-Thr (Cambridge Isotope Laboratories, Inc.) as described (28), but without addition of imatinib to the culture medium. FGMY-labeled His-ABL kinase was isolated by nickel affinity chromatography (nickel-nitrilotriacetic acid, Qiagen) with imidazole elution, yielding 20 mg of heterogeneously phosphorylated kinase from two 0.5-liter cultures. Incubation with AcTEVTM protease (100 units/mg of His-ABL; Invitrogen) and YOP protein-tyrosine phosphatase (1000 units/ml of reaction; New England Biolabs) for 15 h at 6 °C removed the His tag and dephosphorylated the protein. Inhibitor (imatinib, nilotinib, dasatinib, PD180970, or AFN941) was added to aliquots of the reaction from 20 mM stock solutions in Me₂SO. The ABL complexes were then purified by size exclusion chromatography (Superdex 75 HR10/30 column, GE Healthcare) in 20 mM BisTris, 150 mM NaCl, 1 mM EDTA, and 3 mM TCEP (pH 6.5), except for the ABL-PD180970 complex, for which 20 mM Tris, 100 mM NaCl, 1 mM EDTA, and 2.5 mM TCEP (pH 7.6) was used. Purified complexes were concentrated (Ultrafree-0.5, 10 kDa, Millipore) to 230–330 μM. Protein concentration, purity, and stoichiometry were determined by high pressure liquid chromatography for each complex. Liquid chromatography/mass spectrometry analysis showed an incorporation of $^{13}\text{C}/^{15}\text{N}$ label of 95% and 12% residual monophosphorylation for the purified FGMY-labeled ABL kinase.

NMR Samples—Uniformly $^{13}\text{C}/^{15}\text{N}$ - and ^{15}N -labeled samples of ABL-imatinib complexes (1:1) were prepared as 0.4 mM solutions in 250 μl of 95% H₂O and 5% D₂O, 20 mM BisTris, 100 mM NaCl, 2 mM EDTA, and 3 mM dithiothreitol or TCEP (pH 6.5). Selectively labeled samples of imatinib, nilotinib, dasatinib, and PD180970 complexes (1:1) were prepared as solutions (0.32, 0.32, 0.22, and 0.22 mM respectively) in either 95% H₂O and 5% D₂O, 20 mM BisTris, 150 mM NaCl, 2 mM EDTA, and 3 mM dithiothreitol or TCEP (pH 6.5) (imatinib, nilotinib, and dasatinib) or 95% H₂O and 5% D₂O, 20 mM Tris, 100 mM NaCl, 1 mM EDTA, and 2.5 mM TCEP (pH 7.6) (PD180970). Similar preparations were tested for an ABL-AFN941 complex. However, this complex precipitated even at the low concentration of 0.1 mM, and no NMR data could be acquired. Non-isotropic samples of selectively labeled imatinib and nilotinib (dasatinib) complexes were prepared by adding 30 mg/ml (20 mg/ml) filamentous phage Pf1 (Asla Biotech).

NMR Resonance Assignments and Measurement of Residual Dipolar Coupling (RDC) Values—NMR spectra were recorded at 293 K on Bruker DRX 600 MHz (with and without a CryoProbe) and 800 MHz (equipped with a TCI CryoProbe) spectrometers. All spectrometers were equipped with triple-resonance, triple-axis pulsed-field gradient probes. Backbone assignments followed standard triple-resonance strategies with two- and three-dimensional experiments, including HNC0, HNCA, HN(CO)CA, and ^{15}N -edited $^1\text{H}\text{-}^1\text{H}$ nuclear Overhauser effect (NOE) spectroscopy. All NMR data were processed using the NMRPipe suite of programs (29) and analyzed with NMRView (30) to obtain assignments. RDCs were obtained as differences in the splitting observed in the two-

dimensional $^1\text{H}\text{-}^{15}\text{N}$ in-phase anti-phase experiments (31) under anisotropic and isotropic conditions.

NMR Relaxation Experiments and Analysis—Standard ^{15}N relaxation measurements (T_1/T_2 , $\{^1\text{H}\}\text{-}^{15}\text{N}$ NOE) were recorded on the ABL-imatinib complex (uniformly and selectively labeled samples) at 800 MHz. T_1/T_2 decay curves were fitted by an in-house written routine implemented in MATLAB (MathWorks, Inc.) using a simplex search minimization and Monte Carlo estimation of errors (see Fig. 7, A and B). Lipari-Szabo model-free analysis of ^{15}N relaxation data was achieved using the TENSOR2 suite of programs (supplemental Fig. 1) (32).

RESULTS

Resonance Assignment of the ABL-Imatinib Complex—Assignment of backbone NMR resonances was initially performed for the ABL kinase domain (GAMDP-Ser²²⁹-Ser⁵⁰⁰, human ABL1, isoform 1A, 32 kDa) in its non-phosphorylated form and in complex with imatinib. We have shown previously that efficient production of well folded ABL with uniform $^{13}\text{C}/^{15}\text{N}$ isotope labeling is possible by the baculovirus Sf9 insect cell expression system (26). NMR analysis of the ABL complex was difficult for two reasons. 1) Because of solubility problems, ABL concentrations in the NMR samples had to be less than ~0.4 mM. 2) The assignment had to be carried out using protonated protein because cost-effective deuterium labeling is currently not possible in the insect cell system. Consequently, the short transverse relaxation times of the 32-kDa complex allowed only HNC0, HNCA, HNC0CA, and ^{15}N -edited NOE spectroscopy backbone assignment experiments and prevented the use of CBCA-type experiments, which would have yielded distinctive amino acid-type information (33). Supplemental information about amino acid types was therefore obtained from a total of 15 additional, selectively labeled ABL samples. Further details on these samples and the obtained chemical shifts are described elsewhere (34). The available assignments comprise 96% of all backbone $^1\text{H}^{\text{N}}$, ^{15}N , $^{13}\text{C}^{\alpha}$, and ^{13}CO resonances, covering 254 of the 264 non-proline residues (Fig. 2). Unassigned residues consist of the N-terminal glycine, several residues within the activation loop, and His³⁶¹ lining the imatinib-binding surface. Line broadening of adjacent residues indicates that most of the missing residues are broadened beyond detection because of intermediate conformational exchange.

Design of Isotope Labeling Scheme for the Study of Various Inhibitor Complexes—After fully assigning the resonances of the ABL-imatinib complex, a strategy employing selective amino acid labeling was devised for the rapid and unambiguous resonance assignment of key residues in the other inhibitor complexes. To select the best suited labeling scheme, residual $^1\text{H}\text{-}^{15}\text{N}$ dipolar couplings were predicted based upon the crystal structures of the various inhibitor complexes and the orientation tensor of the ABL-imatinib complex measured in Pf1 phages (35). Residues for selective labeling were then chosen to maximize the differences of the predicted dipolar couplings in the “inactive” DFG-out and “active” DFG-in conformations. Thus, maximal experimental differentiation by RDC data between these two conformations should be achieved.

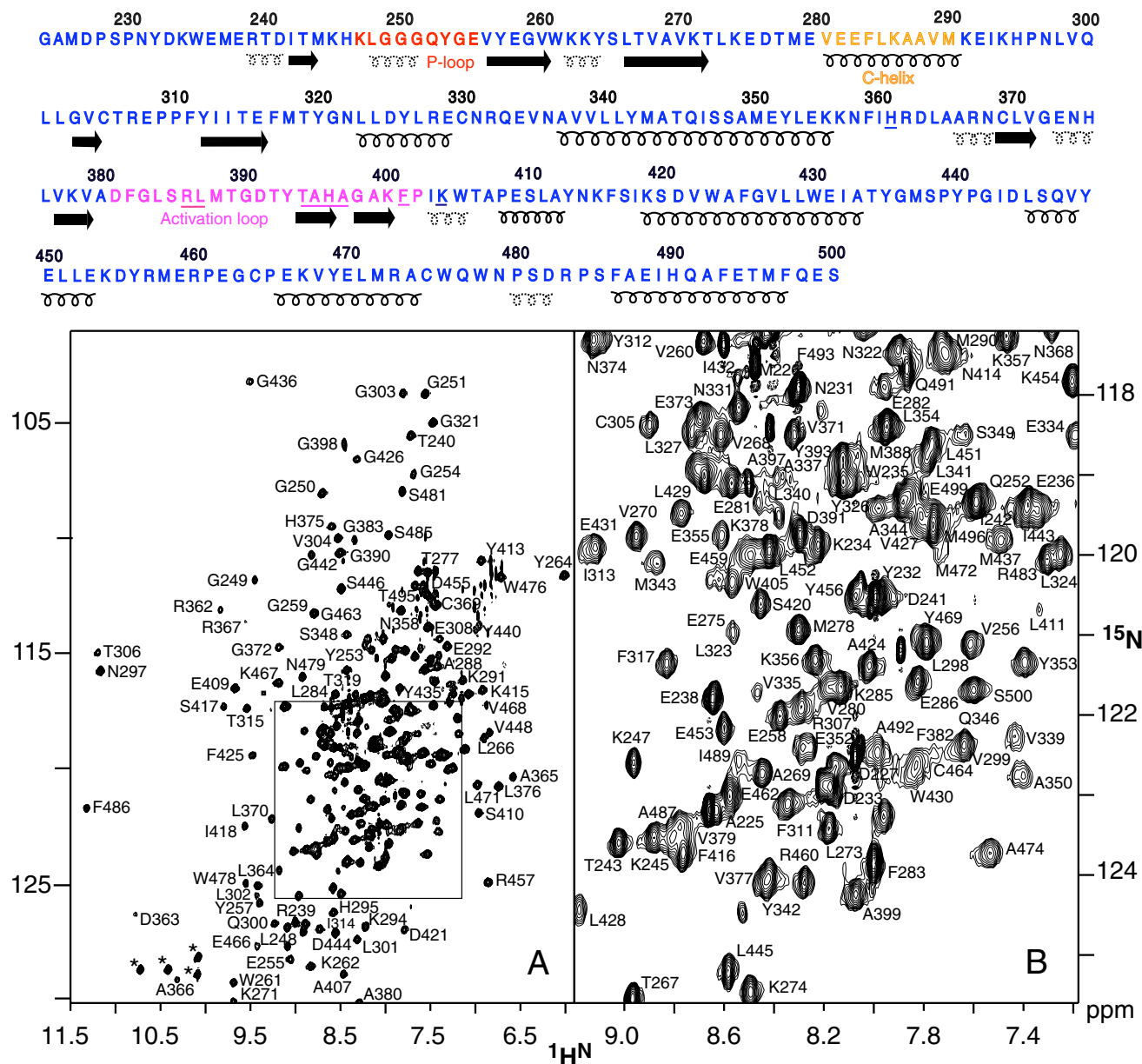


FIGURE 2. A, ^1H - ^{15}N HSQC-transverse relaxation optimized spectrum of uniformly ^{15}N -labeled ABL-imatinib complex with assignment of resonances. Asterisks indicate unassigned Trp^e side chain resonances. B, enlarged region of the box shown in A. The amino acid sequence of the ABL kinase domain (GAMDP-Ser²²⁹-Ser⁵⁰⁰) with its secondary structure and the mentioned activation loop (magenta), P-loop (red), and helix C (gold) is shown at the top. Boldface helices indicate α -helices, whereas dotted helices indicate 3_{10} -helices. Unassigned residues are underlined. The spectrum was acquired for 1 h at a sample concentration of ~ 0.4 mM.

The chosen labeling scheme (FGMY) consists of uniform $^{13}\text{C}/^{15}\text{N}$ labeling for Phe, Gly, Met, and Tyr and specific ^{13}C labeling for Thr and Leu residues. FGMY labeling covers five residues (Gly²⁴⁹, Gly²⁵⁰, Gly²⁵¹, Tyr²⁵³, and Gly²⁵⁴) in the P-loop and seven key residues (Phe³⁸², Gly³⁸³, Ser³⁸⁵, Met³⁸⁸, Gly³⁹⁰, Tyr³⁹³, and Gly³⁹⁸) in the activation loop, with the Ser labeling being the consequence of metabolic scrambling of isotope-labeled glycine. The most important residues (Gly²⁴⁹, Met³⁸⁸, and Tyr³⁹³) are preceded by ^{13}C -labeled Leu or Thr such that they are distinctively detectable in an HNCO experiment. Besides the amino acid-type information, this selective labeling scheme had the advantage that all resonances were completely free of overlap. The ultimate assignment of resonances was achieved by a combination of three- and two-di-

mensional versions of HNCA, HNCO, and HN(CO)CA experiments.

Chemical Shift Analysis—The chemical shift of a nucleus is a sensitive probe of its local environment and can therefore serve as a fingerprint for different molecular conformations. Comparison of the HSQC fingerprint spectra of the ABL-inhibitor complexes demonstrates that the two DFG-out complexes (imatinib and nilotinib) possess a high degree of similarity (Fig. 3A), but are very distinct from the dasatinib DFG-in complex (Fig. 3B). The HSQC spectrum of the dasatinib complex has a strong resemblance to that of the PD180970 (DFG-in) complex (not shown), but the differences are larger than those between the two DFG-out complexes as quantified by the average ^1H - ^{15}N chemical shift differences ($\Delta\delta_{\text{ave}}$) shown in Fig. 3C.

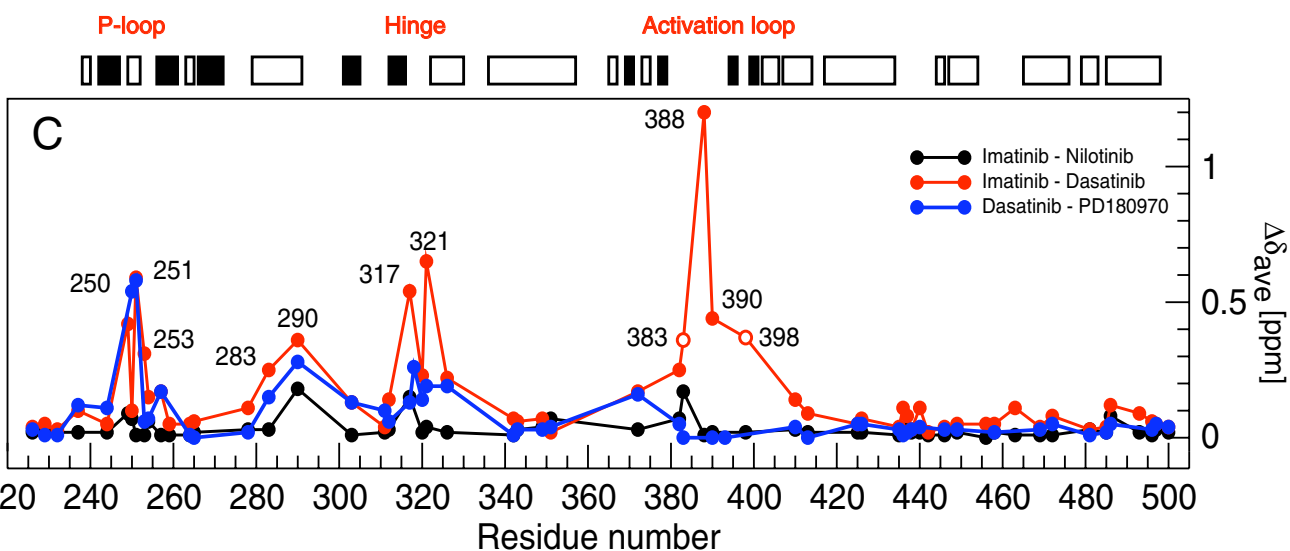
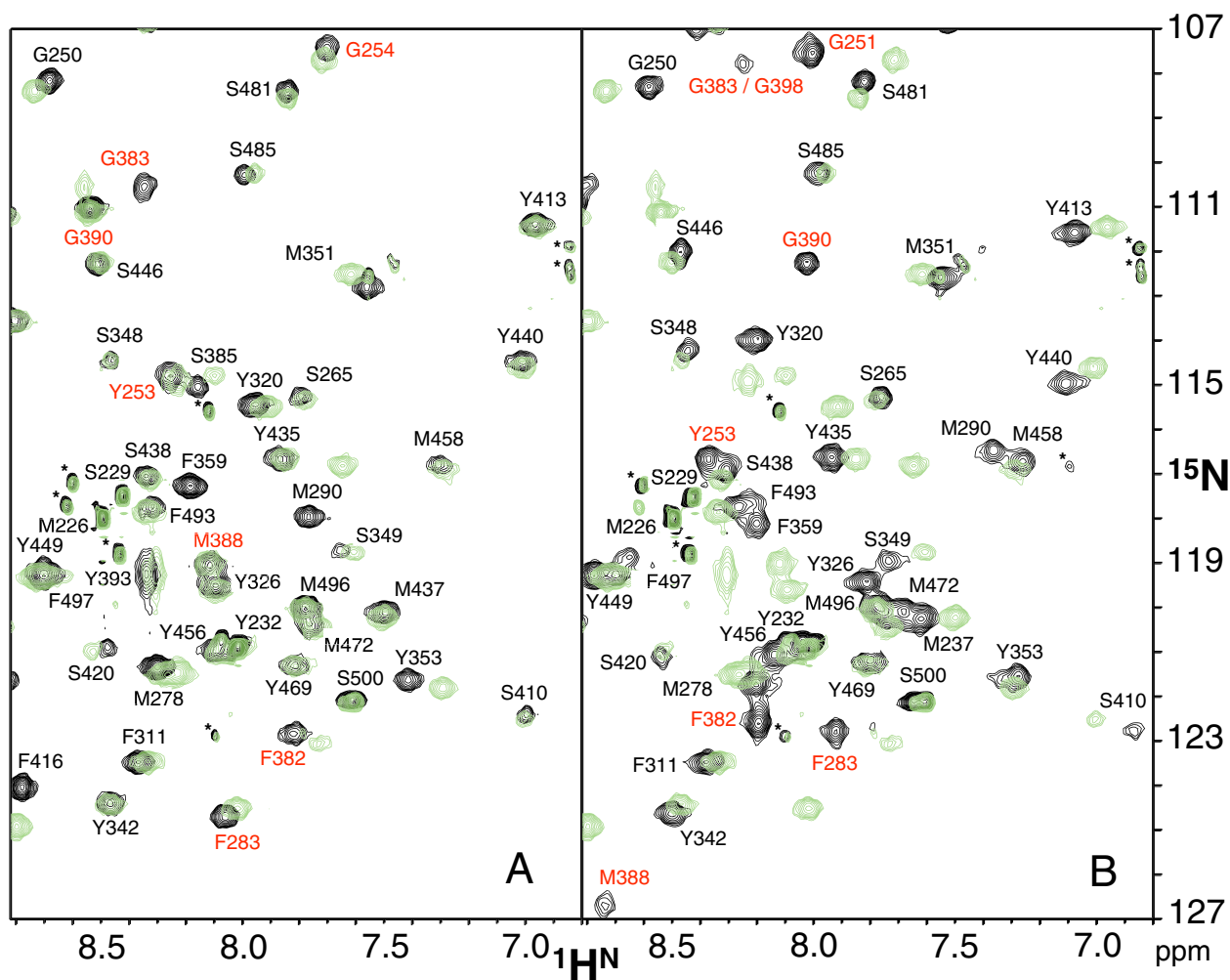
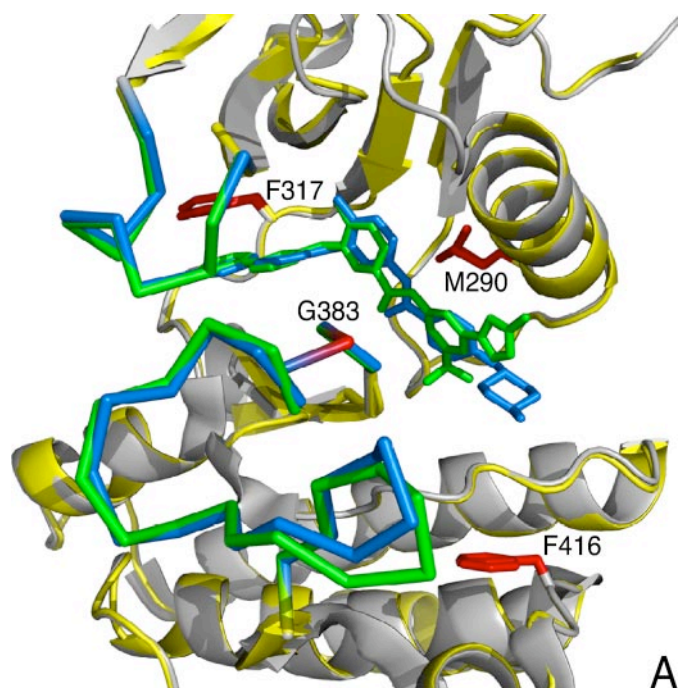
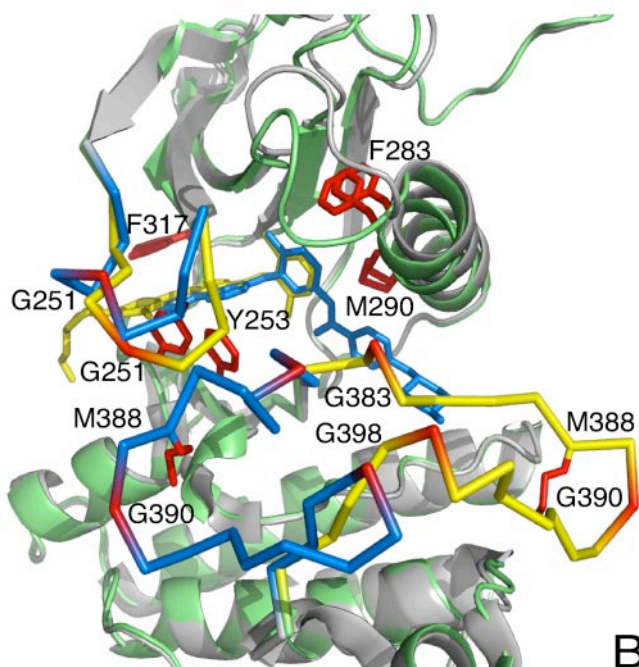


FIGURE 3. Chemical shift analysis of selectively labeled (FGMY) ABL-inhibitor complexes. *A* and *B*, extracted region from the $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$ HSQC spectra of selectively FGMY-labeled ABL kinase in complex with imatinib (*A*; black) and dasatinib (*B*; black). Resonances of the ABL-nilotinib complex (green) are shown for comparison. Residues labeled in red show the largest chemical shift changes. Asterisks indicate unassigned resonances of a low molecular mass impurity. *C*, weighted chemical shift differences $\Delta\delta_{\text{ave}} = (\Delta\delta^2(\text{N})/50 + \Delta\delta^2(\text{H})/2)^{1/2}$ between imatinib and nilotinib (inactive-inactive; black), imatinib and dasatinib (inactive-active; red), and dasatinib and PD180970 (active-active; blue) complexes. Open circles for Gly³⁸³/Gly³⁹⁸ indicate ambiguous assignments for dasatinib. Boxes at the top indicate secondary structure elements showing β -sheets (black) and helices (white). The spectra were acquired for 6 h at a sample concentration of ~ 0.3 mM (~ 0.2 mM) for imatinib and nilotinib (dasatinib) complexes.



A



B

FIGURE 4. Mapping of largely shifted residues on three-dimensional structures of ABL-inhibitor complexes. Overlaid structures of ABL-inhibitor complexes are shown. *A*, ABL-imatinib (gray; Protein Data Bank code 1IEP) and ABL-nilotinib (yellow; code 3CS9). The P-loop, activation loop, and inhibitors are colored in blue and green for imatinib and nilotinib complexes, respectively. *B*, ABL-imatinib (gray; code 1IEP) and ABL-dasatinib (green; code 2GQG, molecule B). The P-loop, activation loop, and inhibitors are colored in blue and yellow for imatinib and dasatinib complexes, respectively. Residues indicated in red show relatively larger changes in the chemical shifts (>0.1 ppm for *A* and >0.3 for *B*).

Between imatinib and nilotinib, chemical shift differences larger than 0.1 ppm are detected only for Met²⁹⁰, Phe³¹⁷, and Gly³⁸³, which are in direct contact with the inhibitors. Much stronger differences are observed between the dasatinib and imatinib complexes in the region of the P-loop around Gly²⁵⁰,

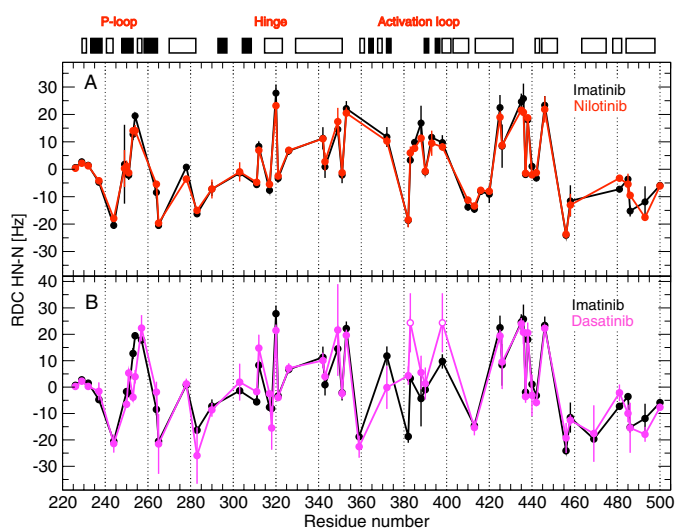


FIGURE 5. Analysis of RDCs (measured in Pf1 phages) of selectively labeled (FGMY) ABL-inhibitor complexes. Shown is a comparison of the experimental $^1\text{H}^{\text{N}}$, ^{15}N RDCs obtained for imatinib and nilotinib complexes (*A*) and imatinib and dasatinib complexes (*B*) shown along the primary sequence. Open circles for Gly³⁸³/Gly³⁹⁸ indicate ambiguous assignments for dasatinib. Error bars indicate variances of two independently measured data sets. Boxes at the top indicate secondary structure elements: β -sheets (black) and helices (white).

the hinge region residues Phe³¹⁷ and Gly³²¹, and the activation loop around Met³⁸⁸. Whereas the residues in the P-loop and hinge region participate in direct interactions with the inhibitor, the affected residues in the activation loop (Met³⁸⁸ and Gly³⁹⁰) are not in direct contact, and hence, their chemical shift changes are in agreement with an allosteric reorientation of the activation loop (Fig. 4).

Residual Dipolar Couplings and Solution Structures—A more quantitative description of the conformations in the different inhibitor complexes was obtained by RDCs. These can be induced in solution by the weak alignment of biomacromolecules (36) and provide a measure of the orientation of internuclear vectors with respect to a fixed coordinate system. Thus, the RDC of the amide N-H bond ($^1D_{\text{NH}}$) is given as follows: $^1D_{\text{NH}} = ^1D_{\text{NH,max}}(P_2(\cos\theta) + \eta/2\sin^2\theta\cos 2\phi)$, where $^1D_{\text{NH,max}}$ is a constant depending on the degree of orientation, P_2 is the second Legendre polynomial, η is the rhombicity of the alignment tensor, and θ and ϕ are polar coordinates of the N-H vector in the principal axis system of the alignment tensor (36). Because RDCs can be measured with high precision, their geometric dependence makes them a powerful tool to study solution conformations and compare them with other structural models such as solid-state x-ray crystal structures.

Weak alignment of the selectively labeled ABL-inhibitor complexes was achieved by the addition of filamentous bacteriophage Pf1 (35). Large $^1D_{\text{NH}}$ RDCs (~ 30 Hz) were obtained for the imatinib, nilotinib, and dasatinib complexes, which indicated substantial alignment and allowed for high sensitivity detection. For the PD180970 complex, the spectral quality was insufficient, and several key resonances were unobservable because of intermediate conformational exchange.

For the imatinib and nilotinib complexes (DFG-out), the RDC values are strikingly similar throughout the protein (Fig. 5A), implying very similar solution structures and dynamics for

Solution Conformations of ABL in Complex with Inhibitors

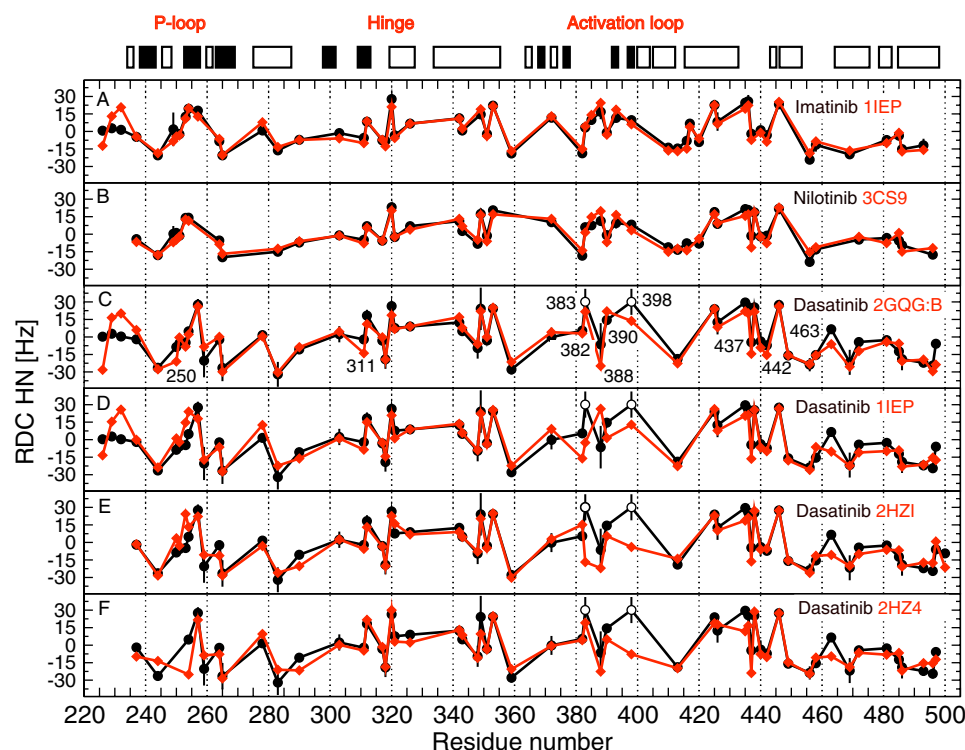


FIGURE 6. Predicted versus experimental RDCs. The RDC values predicted from the crystal structures are shown in red, and the experimental RDC values are shown in black. A, ABL-imatinib (Protein Data Bank code 1IEP); B, ABL-nilotinib (code 3CS9); C, ABL-dasatinib (code 2GQG, molecule B); D, ABL-dasatinib (code 11EP); E, ABL-dasatinib (PD180970; code 2HZ1); F, ABL-dasatinib (AFN941; code 2HZ4). Open circles for Gly³⁸³/Gly³⁹⁸ indicate ambiguous assignments for dasatinib. The alignment tensor was calculated using the experimental RDCs of the complexes mentioned above and the coordinates taken from the crystal structures indicated in parentheses.

both complexes. In marked contrast, RDCs for the dasatinib complex (Fig. 5B) differ substantially from those for the imatinib and nilotinib complexes in both the activation loop and the P-loop. Thus, the solution conformation of the dasatinib complex clearly differs from that of the imatinib and nilotinib complexes, which corroborates the results of the chemical shift analysis.

To characterize the conformations of the activation loop (Asp³⁸¹–Pro⁴⁰²), P-loop (Lys²⁴⁷–Glu²⁵⁵), and hinge region (Phe³¹⁷–Leu³²³) in detail, theoretical RDC values were calculated for each complex. For this, alignment tensors were determined employing a linear fit procedure (37) using the respective crystal structures and the measured RDCs, but excluding the activation loop, P-loop, hinge region, and the flexible residues at the N terminus (<Met²³⁷) and C terminus (>Phe⁴⁹³). Using these alignment tensors together with the crystal coordinates, RDC values were predicted for the entire protein, with the previously excluded regions included. These theoretical values were then compared with the experimental RDC values for the imatinib, nilotinib, and dasatinib complexes (Fig. 6).

For the imatinib and nilotinib complexes (Fig. 6, A and B), it is evident that, besides the flexible N-terminal region, all RDC values throughout the entire protein including the loop regions are in perfect agreement with the crystal structures. In particular, this is the case for Phe³⁸², Gly³⁸³, Ser³⁸⁵, Met³⁸⁸, Gly³⁹⁰, Tyr³⁹³, and Gly³⁹⁸ in the activation loop. This demonstrates that the ABL-imatinib and ABL-nilotinib complexes adopt the

inactive DFG-out conformation in solution and that any dynamic variations from the crystal structure coordinates must be small.

Taking into account the slightly larger experimental errors for the dasatinib complex, the agreement between measured RDC values and those predicted from the crystal structure (Protein Data Bank code 2GQG) is also very good for this complex. The asymmetric unit of the crystal structure 2GQG contains two ABL molecules, one in the phosphorylated form (molecule A; phospho-Tyr³⁹³) and one in the non-phosphorylated form (molecule B). Both structures have almost identical backbone conformations. Predictions are shown in Fig. 6C for the non-phosphorylated form because the ABL protein used for solution NMR was also non-phosphorylated. Besides the flexible N and C termini, moderate deviations between measured and predicted RDCs outside of the experimental error are observed only for turn residues Gly²⁵⁰, Phe³¹¹, Met⁴³⁷, Gly⁴⁴², and Gly⁴⁶³, which are all part of loop regions. For all other unambiguous

assignments, very close agreement is observed. In particular, Phe³⁸², Met³⁸⁸, and Gly³⁹⁰ within the activation loop could be detected and assigned unambiguously. Because of exchange broadening (see below), a further glycine resonance could only be assigned in an ambiguous way either to Gly³⁸³ or Gly³⁹⁸ (shown as open circles in Fig. 6, C–F). However, for both the unambiguous and the two possible ambiguous assignments in the activation loop, the experimental RDCs correspond very closely to the prediction of the 2GQG structure. Very similar agreement is found when the experimental data are compared with the crystal conformation of the dasatinib complex with phosphorylated ABL (2GQG, molecule A) (data not shown). Thus, we conclude that, in solution, the activation loop of the dasatinib complex predominates in the active DFG-in conformation corresponding to that of the 2GQG crystal structure.

To estimate the discriminative power of RDC values for different conformations of the activation loop, we compared the experimental RDCs of the dasatinib complex with those predicted from both the “inactive state” protein in complex with imatinib (Fig. 6D) and the “active state” protein in complex with the structurally unrelated ABL kinase inhibitors PD180970 (Fig. 6E) and AFN941 (Fig. 6F). For the inactive state imatinib complex, all of the predictions for the activation loop region are outside of the error limits of the measured RDCs, showing that the dasatinib complex does not sample the inactive imatinib conformation to a significant extent.

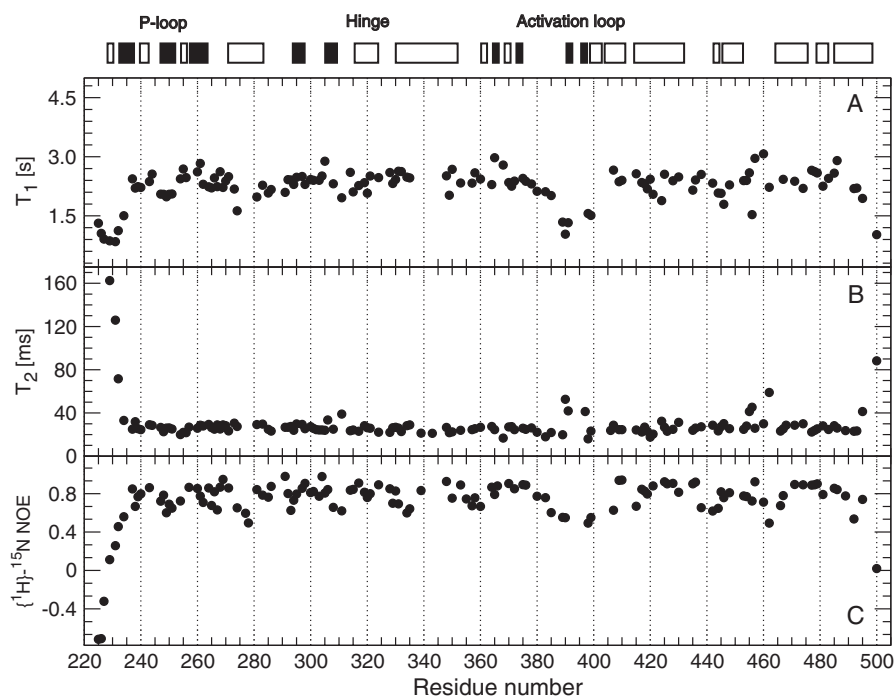


FIGURE 7. ^{15}N relaxation data of the ABL-imatinib complex. T_1 (A), T_2 (B), and hetero- $\{^1\text{H}\}$ - ^{15}N NOE (C) values are shown along the primary sequence. Boxes at the top indicate secondary structure elements: β -sheets (black) and helices (white). For clarity, error bars were omitted. Typical errors propagated from the experimental noise by Monte Carlo estimates are as follows: $T_1 \sim \pm 25$ ms; $T_2 \sim \pm 3$ ms; and NOE $\sim \pm 0.03$.

In the fast chemical exchange regime, the RDC values of an averaging ensemble of structures are given by the average over the RDCs of the different conformations. Using an error estimate of 3 Hz for both experimental and predicted data, we calculate from the large differences and, in some cases, the different sign between measured and predicted data (e.g. for Phe³⁸² and Gly³⁸³) that we would detect an inactive imatinib conformation if it were populated by more than $\sim 15\%$. However, our experimental data are fully compatible with the ABL-dasatinib complex being exclusively in the active conformation.

The crystal structures of the PD180970 and AFN941 complexes show that these inhibitors also bind to ABL in the active conformation, although the path of the activation loop in these complexes is more variable than in the dasatinib complex. Thus, the PD180970 complex is in the active conformation with respect to most of the activation loop, but the crystal structure for this complex shows the DFG motif in a conformation in which the Asp side chain is flipped over to form a hydrogen bond with a main chain carbonyl group. This conformation does not support the binding of ATP, yet is completely different from the DFG-out conformation (11). The experimental RDC values for the dasatinib complex strongly deviate in the DFG region (positions 381–383) from the predicted RDC values (Fig. 6E). Therefore, we conclude that the PD180970 crystal DFG conformation is not highly populated in solution by the dasatinib complex. In contrast, the predictions from the ABL-AFN941 complex (Fig. 6D), which has a typical active DFG-in conformation similar to 2GQG, are closer to those from the dasatinib structure and the experimental RDCs in the activation loop. However, a strong deviation is observed for Gly²⁵⁴ in the P-loop region, probably because the crystal structure shows

disorder in this region and only part of the P-loop could be seen in the electron density.

In summary, all RDC data indicate that for the imatinib and nilotinib complexes, the ensembles of solution conformations are very close to the static structures observed in the crystal. The RDC data also unambiguously show that the conformational ensemble of the dasatinib complex in solution clusters around the active DFG-in conformation observed in the crystal and that inactive DFG-out conformations are not sampled to a significant extent.

Backbone Dynamics—The backbone of the ABL-imatinib complex was characterized by ^{15}N relaxation experiments (Fig. 7). Decreases in ^{15}N T_1 and $\{^1\text{H}\}$ - ^{15}N NOE values and an increase in T_2 values at the N terminus before Met²³⁷ and at the C terminus after Phe⁴⁹³ indicate high nanosecond mobility at both termini. A high rigidity throughout

most of the remaining residues is evident from rather uniform ^{15}N T_1 and T_2 values and $\{^1\text{H}\}$ - ^{15}N NOE values close to 0.8. A clear exception is the activation loop, which has $\{^1\text{H}\}$ - ^{15}N NOE values close to 0.5, together with decreased T_1 and increased T_2 values, indicating large amplitude motions on the subnanosecond time scale. Further regions of higher mobility can be identified around Glu⁴⁶² and close to Lys²⁷⁴, which are both located in turns and have very high temperature factors in the crystal structures. Notably, the P-loop (Lys²⁴⁷–Glu²⁵⁵) of the ABL-imatinib complex does not show pronounced variations in ^{15}N T_1 , T_2 , or $\{^1\text{H}\}$ - ^{15}N NOE values, which would indicate high subnanosecond mobility. For the dasatinib complex, disorder was detected in this region by x-ray crystallography (19). Because of the low concentrations and the selective labeling, quantitative ^{15}N relaxation data on the latter complex are currently missing. However, stronger conformational exchange for the dasatinib complex than for the imatinib complex may be inferred from the detected line broadening in this region (see below).

A quantitative evaluation of the relaxation data was carried out by the program TENSOR2 (32), yielding the isotropic rotational correlation time (τ_c), the Lipari-Szabo (38), subnanosecond order parameters (S^2), and exchange contributions to transverse relaxation from conformational exchange on the micro- to millisecond range (supplemental Fig. 1). The resulting value for τ_c of 21 ns is slightly larger than expected for a molecule of the size of ABL tumbling in aqueous solution at 20 °C. Presumably, this larger value is caused by the onset of aggregation, which occurs at the concentration of 0.4 mM used for the experiments. However, further dilution was not attempted because of the reduced sensitivity that would result.

Solution Conformations of ABL in Complex with Inhibitors

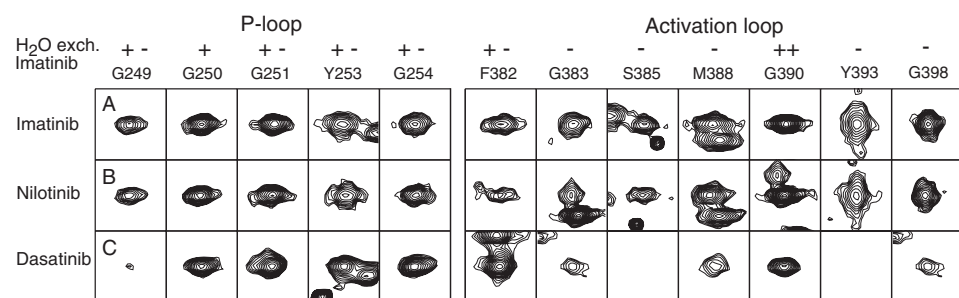


FIGURE 8. Evidence of chemical exchange broadening in the P-loop and activation loop of ABL kinase complexes. The panels show small regions of the ¹H-¹⁵N HSQC spectra of selectively FGMY-labeled ABL extracted at sequential residue positions of the P-loop and activation loop for the imatinib (A), nilotinib (B), and dasatinib (C) complexes. Empty boxes indicate absence of the resonance in the respective spectrum. The information at the top indicates the strength of amide proton exchange (exch.) with water as detected by the intensity of cross-peaks to water in the ¹⁵N-edited NOE spectrum.

The higher mobility in the region of the activation loop is reflected in the TENSOR2 analysis both by a reduced order parameter (S^2) of ~ 0.6 – 0.7 and contributions from chemical exchange broadening on the microsecond time scale of up to 25–30 Hz.

Such chemical exchange effects are clearly visible as a broadening and weakening of the resonance lines in the ¹H-¹⁵N HSQC spectra. Fig. 8 shows the respective peaks of the activation loop and P-loop for the imatinib, nilotinib, and dasatinib complexes. For imatinib, prominent line broadening is observed for Met³⁸⁸ and Tyr³⁹³, which cannot be attributed to hydrogen exchange with water because exchange peaks with water are absent in the ¹⁵N-edited NOE spectra. Hence, the line broadening is caused by conformational exchange on the microsecond time scale of chemical shifts. The broadening of Tyr³⁹³ is particularly interesting because this residue becomes phosphorylated in the activated complex. Similar line broadening is observed for the nilotinib complex. However, much more pronounced broadening occurs for the dasatinib complex, e.g. Ser³⁸⁵, and Tyr³⁹³ could not be detected at all. This clearly indicates a differing dynamic behavior of the activation loop in the dasatinib complex.

Within the P-loop, weak exchange broadening is observed for Gly²⁴⁹ and Tyr²⁵³ in the case of imatinib and nilotinib (Fig. 8). Again, much stronger exchange broadening occurs for Gly²⁴⁹ in the dasatinib complex. Thus, also the P-loop appears more mobile in the case of the active state inhibitor dasatinib.

These relaxation results on the ABL-inhibitor complexes are significant because they directly show the presence of dynamic processes in several regions of the protein, including the activation loop. It should be noted, however, that although the effects of line broadening are considerable, they do not necessarily imply that the populations of the other conformations, which are exchanging with main species, are large. Admixtures of populations on the order of 1% can lead to significant broadening effects (39). Thus, the detection of such motions by line broadening is not contradicting the finding from the RDC analysis that the major part of the solution ensemble is close to the crystal structure.

DISCUSSION

Our results compose the first detailed structural characterization of protein kinase-inhibitor complexes in solution. They

are based on the availability of the almost complete assignment of the backbone resonances of the ABL-imatinib complex and partial resonance assignments of the ABL-nilotinib, ABL-dasatinib, and ABL-PD180970 complexes. This allowed an analysis of chemical shift perturbations, RDC, and ¹⁵N relaxation data. For the imatinib and nilotinib complexes, the ensemble of solution conformations closely resembles the static inactive DFG-out structure determined in the crystal, although residual mobility

of the activation loop can be detected from ¹⁵N relaxation data and the line broadening of some resonances.

For the dasatinib complex, the RDC data clearly show that the ensemble of solution conformations is close to the active DFG-in structure. However, line broadening effects around Ser³⁸⁵ and Tyr³⁹³ in the activation loop and Gly²⁴⁹ in the P-loop indicate the presence of microsecond to millisecond motions in these regions. Relaxation dispersion can reveal the time scale and chemical shift differences of the species involved in the exchange (39). Because of the low solubility (≤ 0.2 mM) of the dasatinib complex and the weak intensity of the signals, such experiments were not attempted. Nevertheless, based on the close agreement between measured RDCs and the prediction according to the active state conformation, the amplitude of these microsecond to millisecond motions and/or the populations of the exchanging minor conformations should be small. Such minor conformations may be the result of small rearrangements of the backbone or the side chains or variations in hydrogen bond geometries.

Based on molecular modeling and molecular dynamics studies, it has been hypothesized that dasatinib can bind to both the active DFG-in and inactive DFG-out conformations of ABL (19, 20). This notion has become widespread despite the absence of supportive experimental evidence and the fact that the x-ray structure of the ABL-dasatinib complex shows only the active ABL conformation. This study is the first to actually assess the extent of DFG-out conformation in the ABL-dasatinib complex in solution. No significant admixture of the DFG-out conformation is detectable from the measured RDC values. In a further experiment (data not shown), we displaced imatinib by adding dasatinib in high excess to the ABL-imatinib complex rather than adding dasatinib directly to unliganded ABL. Even when offering the preformed inactive DFG-out state to dasatinib in this manner, the resulting ensemble of conformations is indistinguishable from the ensemble observed when adding dasatinib to unliganded ABL.

This study was performed with non-phosphorylated protein, which was necessary to allow the protein to also adopt the inactive DFG-out conformation. The phosphorylation of Tyr³⁹³ in the activation loop stabilizes the active conformation of the protein by forming interactions with neighboring side chains (40). It can be expected that this will reduce the flexibility of the dasatinib complex, thereby narrowing the ensemble of the

active DFG-in conformations even further. For the same reason, any propensity of the ABL-dasatinib complex to adopt the inactive DFG-out conformation should be even more reduced.

It is generally believed that there is only one active state, which satisfies the requirement to have all essential elements correctly orientated for efficient catalysis. In contrast, multiple inactive states may exist. This is supported by experimental findings for both ABL and SRC kinases in complex with various ligands (41). Our findings of very similar inactive states for the imatinib and nilotinib complexes do not contradict this notion; although multiple ligand-dependent inactive states may exist, we have shown that the inactive state with a particular ligand is well defined and resembles closely the conformation observed in the crystal. However, these inactive conformations are not completely rigid because we still observed high nanosecond backbone flexibility within the activation loop despite the fact that the RDC values indicate that the ensemble average is close to the x-ray structure.

The observed flexibility seen in both the active and inactive states is likely to be an intrinsic requirement for catalytic activity and for the transition between the active and inactive conformations of ABL as well as other kinases (42). Indeed, molecular dynamics calculations have shown that the various inactive ABL kinase conformations may be necessary intermediates in this transition (41).

Considering the tendency of different kinases to adopt different inactive states, e.g. either DFG-in or DFG-out conformations, the free energy differences between these states appear to vary significantly between kinases. For example, whereas DFG-out conformations seem extraordinarily stable for certain ABL complexes, such conformations, although possible, have a high thermodynamic penalty in SRC complexes (43). The reason for this strikingly differing behavior is unknown and cannot be attributed to a few individual differing amino acids. The role of such residues in the neighborhood of the DFG motif or elsewhere and the overall energetics of the different states cannot be determined from the crystal structures alone. NMR-derived dynamic information, as obtained here, should lead us to better understand these systems and to comprehend why certain inactive conformations are more or less favorable in some kinases relative to others (21).

To understand how point mutations cause patient resistance to imatinib and eventual relapse is also of crucial importance because the efficacy of inhibitors may be related to the free energy landscape of the various inactive states (43). A comprehensive description of this behavior is a fundamental prerequisite for a more rational design of potent new drugs.

Acknowledgments—We thank Drs. Sonja Alexandra Dames, Sebastian Meier, and Martin Allan for help during the initial phase of the project.

REFERENCES

- Melo, J. V., and Barnes, D. J. (2007) *Nat. Rev. Cancer* **7**, 441–453
- Ren, R. (2005) *Nat. Rev. Cancer* **5**, 172–183
- Huse, M., and Kuriyan, J. (2002) *Cell* **109**, 275–282
- Mol, C. D., Dougan, D. R., Schneider, T. R., Skene, R. J., Kraus, M. L., Scheibe, D. N., Snell, G. P., Zou, H., Sang, B. C., and Wilson, K. P. (2004) *J. Biol. Chem.* **279**, 31655–31663
- Griffith, J., Black, J., Faerman, C., Swenson, L., Wynn, M., Lu, F., Lippke, J., and Saxena, K. (2004) *Mol. Cell* **13**, 169–178
- Schindler, T., Bornmann, W., Pellicena, P., Miller, W. T., Clarkson, B., and Kuriyan, J. (2000) *Science* **289**, 1938–1942
- Hubbard, S. R., Wei, L., Elis, L., and Hendrickson, W. A. (1994) *Nature* **372**, 746–754
- Wan, P. T. C., Garnett, M. J., Roe, S. M., Lee, S., Niculescu-Duvaz, D., Good, V. M., Jones, C. M., Marshall, C. J., Springer, C. J., Barford, D., and Marais, R. (2004) *Cell* **116**, 855–867
- Pargellis, C., Tong, L., Churchill, L., Cirillo, P. F., Gilmore, T., Graham, A. G., Grob, P. M., Hickey, E. R., Moss, N., Pav, S., and Regan, J. (2002) *Nat. Struct. Biol.* **9**, 268–272
- Druker, B. J., Guilhot, F., O'Brien, S. G., Gathmann, I., Kantarjian, H., Gattermann, N., Deininger, M. W. N., Silver, R. T., Goldman, J. M., Stone, R. M., Cervantes, F., Hochhaus, A., Powell, B. L., Gabrilove, J. L., Rousselot, P., Reiffers, J., Cornelissen, J. J., Hughes, T., Agis, H., Fischer, T., Verhoef, G., Shepherd, J., Saglio, G., Gratwohl, A., Nielsen, J. L., Radich, J. P., Simonsson, B., Taylor, K., Baccarani, M., So, C., Letvak, L., and Larson, R. A. (2006) *N. Engl. J. Med.* **355**, 2408–2417
- Cowan-Jacob, S. W., Fendrich, G., Floersheimer, A., Furet, P., Liebetanz, J., Rummel, G., Rheinberger, P., Centeleghe, M., Fabbro, D., and Manley, P. W. (2007) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **63**, 80–93
- Cowan-Jacob, S. W., Guez, V., Fendrich, G., Griffin, J. D., Fabbro, D., Furet, P., Liebetanz, J., Mestan, J., and Manley, P. W. (2004) *Mini-Rev. Med. Chem.* **4**, 285–299
- Nagar, B., Hantschel, O., Young, M. A., Scheffzek, K., Veach, D., Bornmann, V., Clarkson, B., Superti-Furga, G., and Kuriyan, J. (2003) *Cell* **112**, 859–871
- Nagar, B., Bornmann, W. G., Pellicena, P., Schindler, T., Veach, D. R., Miller, W. T., Clarkson, B., and Kuriyan, J. (2002) *Cancer Res.* **62**, 4236–4243
- Manley, P. W., Cowan-Jacob, S. W., Buchdunger, E., Fabbro, D., Fendrich, G., Furet, P., Meyer, T., and Zimmermann, J. (2002) *Eur. J. Cancer* **38**, S19–S27
- Weisberg, E., Manley, P. W., Breitenstein, W., Bruggen, J., Cowan-Jacob, S. W., Ray, A., Huntly, B., Fabbro, D., Fendrich, G., Hall-Meyers, E., Kung, A. L., Mestan, J., Daley, G. Q., Callahan, L., Catley, L., Cavazzall, C., Azam, M., Neuberger, D., Wright, R. D., Gilliland, G., and Griffin, J. D. (2005) *Cancer Cell* **7**, 129–141
- Weisberg, E., Manley, P., Mestan, J., Cowan-Jacob, S., Ray, A., and Griffin, J. D. (2006) *Br. J. Cancer* **94**, 1765–1769
- Das, J., Chen, P., Norris, D., Padmanabha, R., Lin, J., Moquin, R. V., Shen, Z. Q., Cook, L. S., Doweiko, A. M., Pitt, S., Pang, S. H., Shen, D. R., Fang, Q., de Fex, H. F., McIntyre, K. W., Shuster, D. J., Gillooly, K. M., Behnia, K., Schieven, G. L., Wityak, J., and Barrish, J. C. (2006) *J. Med. Chem.* **49**, 6819–6832
- Tokarski, J. S., Newitt, J. A., Chang, C. Y. J., Cheng, J. D., Wittekind, M., Kiefer, S. E., Kish, K., Lee, F. Y. F., Borzillieri, R., Lombardo, L. J., Xie, D. L., Zhang, Y. Q., and Klei, H. E. (2006) *Cancer Res.* **66**, 5790–5797
- Verkhivker, G. A. (2007) *Biopolymers* **85**, 333–348
- Wissing, J., Godl, K., Brehmer, D., Blencke, S., Weber, M., Habenberger, P., Stein-Gerlach, M., Missio, A., Cotten, M., Muller, S., and Daub, H. (2004) *Mol. Cell. Proteomics* **3**, 1181–1193
- Wisniewski, D., Lambek, C. L., Liu, C. Y., Strife, A., Veach, D. R., Nagar, B., Young, M. A., Schindler, T., Bornmann, W. G., Bertino, J. R., Kuriyan, J., and Clarkson, B. (2002) *Cancer Res.* **62**, 4244–4255
- Kraker, A. J., Hartl, B. G., Amar, A. M., Barvian, M. R., Showalter, H. D. H., and Moore, C. W. (2000) *Biochem. Pharmacol.* **60**, 885–898
- Wiesner, S., Wybenga-Groot, L. E., Warner, N., Lin, H., Pawson, T., Forman-Kay, J. D., and Sicheri, F. (2006) *EMBO J.* **25**, 4686–4696
- Vogtherr, M., Saxena, K., Hoelder, S., Grimme, S., Betz, M., Schieborr, U., Pescatore, B., Robin, M., Delarbre, L., Langer, T., Wendt, K. U., and Schwalbe, H. (2006) *Angew. Chem. Int. Ed. Engl.* **45**, 993–997
- Strauss, A., Bitsch, F., Fendrich, G., Graff, P., Knecht, R., Meyhack, B., and Jahnke, W. (2005) *J. Biomol. NMR* **31**, 343–349
- Strauss, A., Bitsch, F., Cutting, B., Fendrich, G., Graff, P., Liebetanz, J., Zurini, M., and Jahnke, W. (2003) *J. Biomol. NMR* **26**, 367–372

Solution Conformations of ABL in Complex with Inhibitors

28. Cowan-Jacob, S. W., Fendrich, G., Manley, P. W., Jahnke, W., Fabbro, D., Liebetanz, J., and Meyer, T. (2005) *Structure* **13**, 861–871
29. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR* **6**, 277–293
30. Johnson, B. A., and Blevins, R. A. (1994) *J. Biomol. NMR* **4**, 603–614
31. Ottiger, M., Delaglio, F., and Bax, A. (1998) *J. Magn. Reson.* **131**, 373–378
32. Dosset, P., Hus, J. C., Blackledge, M., and Marion, D. (2000) *J. Biomol. NMR* **16**, 23–28
33. Grzesiek, S., and Bax, A. (1993) *J. Biomol. NMR* **3**, 185–204
34. Vajpai, N., Strauss, A., Fendrich, G., Cowan-Jacob, S. W., Manley, P. W., Jahnke, W., and Grzesiek, S. (2008) *Biomol. NMR Assign.* 10.1007/s12104-008-9079-7
35. Hansen, M. R., Mueller, L., and Pardi, A. (1998) *Nat. Struct. Biol.* **5**, 1065–1074
36. Tjandra, N., and Bax, A. (1997) *Science* **278**, 1111–1113
37. Sass, J., Cordier, F., Hoffmann, A., Cousin, A., Omichinski, J. G., Lowen, H., and Grzesiek, S. (1999) *J. Am. Chem. Soc.* **121**, 2047–2055
38. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc.* **104**, 4546–4559
39. Korzhnev, D. M., Salvatella, X., Vendruscolo, M., Di Nardo, A. A., Davidson, A. R., Dobson, C. M., and Kay, L. E. (2004) *Nature* **430**, 586–590
40. Young, M. A., Shah, N. P., Chao, L. H., Seeliger, M., Milanov, Z. V., Biggs, W. H., Treiber, D. K., Patel, H. K., Zarrinkar, P. P., Lockhart, D. V. J., Sawyers, C. L., and Kuriyan, J. (2006) *Cancer Res.* **66**, 1007–1014
41. Levinson, N. M., Kuchment, O., Shen, K., Young, M. A., Koldobskiy, M., Karplus, M., Cole, P. A., and Kuriyan, J. (2006) *PLoS Biol.* **4**, 753–767
42. Eisenmesser, E. Z., Millet, O., Labeikovsky, W., Korzhnev, D. M., Wolf-Watz, M., Bosco, D. A., Skalicky, J. J., Kay, L. E., and Kern, D. (2005) *Nature* **438**, 117–121
43. Seeliger, M. A., Nagar, B., Frank, F., Cao, X., Henderson, M. N., and Kuriyan, J. (2007) *Structure* **15**, 299–311



**SOLUTION CONFORMATIONS AND DYNAMICS OF ABL KINASE INHIBITOR
COMPLEXES DETERMINED BY NMR SUBSTANTIATE THE DIFFERENT BINDING
MODES OF IMATINIB/NILOTINIB AND DASATINIB**

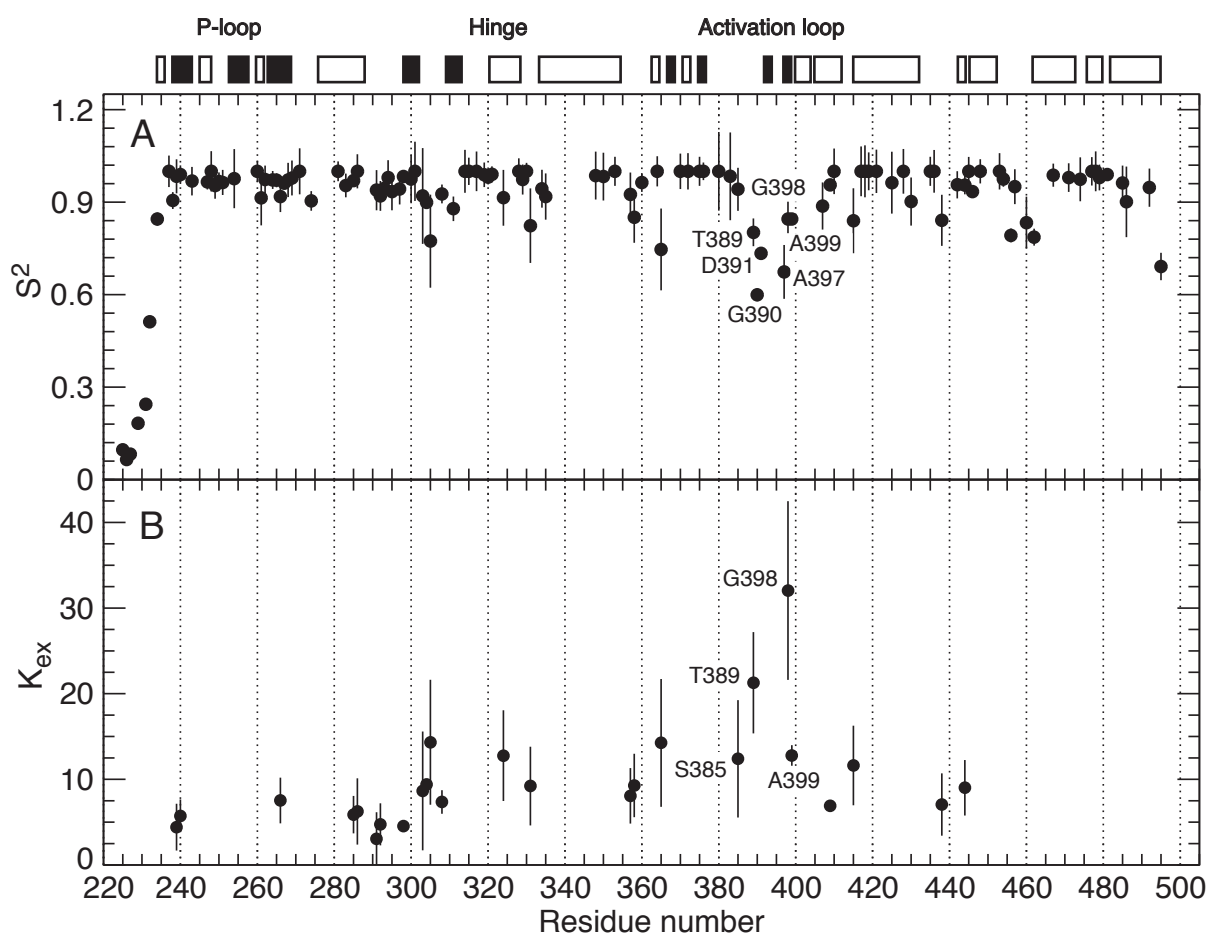
**Navratna Vajpai², André Strauss¹, Gabriele Fendrich¹, Sandra W. Cowan-Jacob¹, Paul W.
Manley¹, Stephan Grzesiek², Wolfgang Jahnke¹**

¹Novartis Institutes for Biomedical Research, Basel, Switzerland, ²Biozentrum, University of Basel,
Basel, Switzerland.

Correspondence should be addressed to
WJ (wolfgang.jahnke@novartis.com) or SG (stephan.grzesiek@unibas.ch).

Running head: Dynamic activation loop of ABL in complex with inhibitor.

Supplementary information



Supplemental Figure 1: Lipari-Szabo model-free analysis of ^{15}N -relaxation data of the ABL-imatinib complex. A) Order parameter, S^2 and B) chemical exchange, K_{ex} . For the calculation of the isotropic overall diffusion tensor, amide resonances with $\{^1\text{H}\}$ - ^{15}N NOE < 0.65 were removed from the analysis. Data points corresponding to amides within the activation loop are labeled with residue information. Boxes at the top indicate secondary structure elements corresponding to beta-sheets (filled) and helices (open). Error bars indicate statistical errors derived by Monte Carlo analysis using experimental noise values.

Chapter 3:

Conformational studies of unstructured oligopeptides by residual dipolar couplings

Abstract

The description of (ϕ, φ) space sampling of an unstructured polypeptide chain is of considerable interest for the understanding of protein folding. During the last decade, novel NMR techniques such as paramagnetic relaxation enhancement (PRE) and residual dipolar couplings (RDCs) have become powerful tools to characterize unfolded polypeptides in solution. Here, we have used RDCs to directly monitor average net orientations and order parameters of individual bonds in unfolded polypeptides.

In section 3.1, the systematic investigation of the effect of various single and multiple amino acid substitutions on the conformation of short peptides EGAAXAASS is reported. $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$ and $^1\text{H}^{\alpha}\text{-}^{13}\text{C}^{\alpha}$ RDCs were detected at natural abundance of ^{15}N and ^{13}C in strained polyacrylamide gels. The data show a specific dependence on the substitution X that can be attributed to steric or hydrophobic interactions with adjacent amino acids. Reasonable agreement with $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$ RDCs derived from the statistical coil model implies that local preferences are closely related to the torsion angle distribution of non- α and non- β structures in folded proteins.

In section 3.2, all-atom molecular dynamics (MD) simulations were performed for the quantitative investigation of nonapeptides EGAAXAASS. Ensemble averaged RDCs were calculated from the MD simulations and compared to the experimental RDCs. Significant differences were obtained between the predictions and the experimental data, which may be attributed to specific deficiencies in the MD simulations, such as inadequate sampling of polypeptide conformational space in the time used for the simulations.

Background

The ability of a highly dynamic structural ensemble of an unfolded polypeptide to quickly fold to its native state structure has given rise to a very important interdisciplinary problem of the current time, namely, the ‘protein folding problem’. Although major advances have been made in elucidating the folding mechanism of several proteins, detailed understanding of the molecular events that occur during the folding process remains limited.

Understanding the process of protein folding is of fundamental interest in biology. As a prerequisite to achieve this goal, an accurate structural and dynamic characterization of all the species along the folding pathways is required. This includes the fully denatured states, the partially folded intermediates and the native state. To date, extensive structural and dynamic studies have been reported on the folded states. However, a proper description of the partially folded or unfolded state ensemble remains challenging, since only a limited number of measurable parameters are available to describe the very large space of the unstructured conformations (Meier *et al.*, 2008). Thus, a thorough understanding of the structure and dynamics of unfolded states or denatured polypeptides is of great importance for understanding protein folding.

Unfolded proteins are not only important from the viewpoint of protein folding. Unfolded and partly folded proteins play important roles in numerous cellular processes and signaling events. Many unfolded proteins that adopt structure upon binding target proteins have now been found to act as molecular switches that regulate gene expression, both in transcription and in translation. In particular, trans-activation domains of transcriptional activators tend to be unstructured by themselves, but fold upon binding co-activators (Fuxreiter *et al.*, 2004, Fink, 2005, Dyson & Wright, 2005). There are also other classes of natively unstructured proteins that function without significant ordering, such as elastin, whose conformational disorder plays a central role in its elastomeric function (Rauscher *et al.*, 2006), or the molten globular domain of clusterin, which appears to act as a proteinaceous detergent for cell remodeling (Dunker *et al.*, 2001). It became evident only recently that a large fraction of eukaryotic proteins, possibly as

many as 30%, are either completely or partially unstructured (Fink, 2005); in cancer-associated or signaling proteins, the predictions are even higher: 80% and 67%, respectively (Iakoucheva *et al.*, 2002). Aggregation of these proteins plays an important role in the onset of neurodegenerative diseases, such as Parkinson's (Uversky *et al.*, 2001) and Alzheimer's diseases (Mandelkow & Mandelkow, 1998), motivating studies of the role of transient, fluctuating structure in aggregation (Dedmon *et al.*, 2005, Mukrasch *et al.*, 2005, Marsh *et al.*, 2006). The normal physiological role of many unstructured proteins in the mediation of protein interactions (Dunker *et al.*, 2005) has also led to investigations of the influence of dynamic states on recognition mechanisms (Dyson & Wright, 2005).

Unfolded or denatured proteins are a challenging subject for structural investigations because of their heterogeneity and conformational flexibility (Mittag & Forman-Kay, 2007, Tanford *et al.*, 1966). In addition, unfolded states are typically only weakly populated under non-denaturing conditions. Therefore, different methods are employed to destabilize folded proteins to enable studies of denatured states. These include extremes of temperature (Dill & Shortle, 1991), addition of denaturants such as guanidinium chloride (Tanford *et al.*, 1966, Greene & Pace, 1974, Shortle, 1996a), urea (Greene & Pace, 1974, Shortle, 1996a, Shortle & Ackerman, 2001) or acid (Goto *et al.*, 1990, Jeng *et al.*, 1990, Kamatari *et al.*, 2004, Dill & Shortle, 1991), methionine oxidation (Chugha *et al.*, 2006), pressure (Kamatari *et al.*, 2004), truncation (Flanagan *et al.*, 1992) and mutation (Shortle *et al.*, 1992).

Many solution spectroscopic and other experimental techniques have been applied to characterize unfolded or denatured states (Mittag & Forman-Kay, 2007). Studies based on intrinsic viscosity (Tanford *et al.*, 1966), hydrodynamic radii (Wilkins *et al.*, 1999), and small-angle scattering experiments (SAXS) (Millett *et al.*, 2002, Kohn *et al.*, 2004) have provided evidence that natively unfolded or chemically denatured proteins exhibit biophysical properties consistent with random-coil distributions. These observations have led to the random-coil model being regarded as the standard reference state for interpretation of experimental data from unfolded proteins, and the starting point for most theoretical considerations of the folding process (McCarney *et al.*, 2005). In this model,

the random coil was assumed to be a state in which there were no nonlocal interactions along the polypeptide chain and descriptions of the (ϕ, ψ) populations of each residue in the protein were taken from the distribution of (ϕ, ψ) torsion angles in the protein database. The random coil model was quantitatively validated by $^3J_{\text{HN-H}\alpha}$ scalar couplings (Schwalbe *et al.*, 1997).

In contrast, other studies, mainly NMR and SAXS, clearly demonstrate that significant secondary structure and long-range hydrophobic clusters persist in unfolded proteins, even under strongly denaturing conditions (Evans *et al.*, 1991, Neri *et al.*, 1992, Alexandrescu & Shortle, 1994, Shortle, 1996a, Hodsdon & Frieden, 2001, Kazmirski *et al.*, 2001, Klein-Seetharaman *et al.*, 2002, Meier *et al.*, 2007b, Meier *et al.*, 2008, Garcia *et al.*, 2001). Under such extreme conditions, the observed residual structures represent local preferences and suggest that the surface of the folding energy landscape has an inherent bias to overcome Levinthal's paradox, (Zwanzig *et al.*, 1992, Shortle, 1996b) which states that if a protein were to fold by sequentially sampling all possible conformations, it would take an astronomical amount of time to do so, even if the conformations were sampled within nano- or even pico-seconds (Levinthal, 1968).

Quantitative characterization of unfolded states by NMR spectroscopy

In recent years, high-resolution solution NMR spectroscopy has shown particular promise for the investigation of the structure and dynamics of unfolded or denatured states (Dyson & Wright, 2001, Dyson & Wright, 2005, Mittag & Forman-Kay, 2007, Abragam, 1983). While the chemical shift dispersion for unfolded states is very narrow (especially for protons and aliphatic carbons) due to conformational averaging, the internal motions on fast timescales enhance the sensitivity of NMR spectroscopy. Several experimental observables can be used to gain detailed information about the heterogeneous ensembles at atomic level. The most widely used observables include chemical shifts, for which deviation from random coil values may indicate local conformational propensities (Bundi & Wuthrich, 1979, Wishart & Sykes, 1994, Merutka *et al.*, 1995, Schwarzingger *et al.*, 2001), ^{15}N transverse magnetization relaxation rates, which yield information about long-range interactions (Klein-Seetharaman *et al.*, 2002); and conformational exchange within the ensemble (Tollinger *et al.*, 2001, Korzhnev *et al.*, 2004). Elements of residual

structure can be inferred from nuclear Overhauser effect (NOE) data (Neri *et al.*, 1992, Zhang *et al.*, 1997b, Zhang *et al.*, 1997a, Mok *et al.*, 1999, Abragam, 1983); three-bond J-couplings provide information about torsion angles (Serrano, 1995, Smith *et al.*, 1996, Schwalbe *et al.*, 1997). High pressure NMR (Kitahara & Akasaka, 2003, Akasaka, 2003, Kamatari *et al.*, 2004) provides information about the effective volumes. Amide proton hydrogen exchange (Baum *et al.*, 1989, Hughson *et al.*, 1990) reports protection of hydrogen bonds against exchange with solvent, and diffusion-based methods are used to determine hydrodynamic radius (Jones *et al.*, 1997, Pan *et al.*, 1997). Other approaches include 2D photo-CIDNP (chemically induced dynamic nuclear polarization), which allows the determination of differential accessibility of aromatic side-chains involved in hydrophobic clustering (Schlorb *et al.*, 2006); the use of spin labels (paramagnetic relaxation enhancement; PRE) enables the determination of residual long-range order (Gillespie & Shortle, 1997b, Gillespie & Shortle, 1997a, Bertoncini *et al.*, 2005, Kristjansdottir *et al.*, 2005, Dedmon *et al.*, 2005, Lindorff-Larsen *et al.*, 2004), and residual dipolar couplings (RDCs), which report on residual short- (Shortle & Ackerman, 2001, Ohnishi & Shortle, 2003) and long-range order (Ohnishi *et al.*, 2004).

In contrast to many conventional observables mentioned above, PREs and RDCs have proven to be much more valuable in the quantitative characterization of global and local order in unstructured state of proteins. This is due to their simple analytical dependence on the average over the electron-nucleus distant vector (PREs) or geometrical dependence on internuclear vector orientations (RDCs). Recent technical advances in NMR that allow easy measurement of PREs and RDCs have led to these techniques becoming increasingly popular. A brief description of these two observables is given in the following section.

Paramagnetic relaxation enhancement

Spin labels in NMR can be used to obtain information on long-range distances in both folded and unfolded proteins. The applicability of spin label is based on the increase in the relaxation rates of its neighboring protons caused by the dipolar interaction with the paramagnetic center, so called paramagnetic relaxation enhancement (PRE). Like NOEs, the magnitude of the PREs is proportional to r^{-6} , where r is the distance between a proton

and the spin label. Due to the large gyromagnetic ratio of the electron, proton relaxation rates in the vicinity of spin label are significantly enhanced. This effect can be precisely detected over a large distance. PREs can provide unique long-range distance information in the 15-25 Å range, thus making them useful for the characterization of unfolded or denatured states. However, the r^{-6} distance dependence leads to compact conformers in the ensemble dominating the PRE effect.

The most commonly used spin labels are nitroxide spin labels, such as MTSL [(1-oxy-2,2,5,5-tetramethyl-3-pyrroline-3-methyl) methanesulfonate] or PROXYL [N-(1-oxy-2,2,5,5-tetramethyl-3-pyrrolidinyl) iodoacetamide], which are covalently attached to a single cysteine in the protein. In the absence of a cysteine in the amino-acid sequence, a single cysteine residue is engineered in the protein by site-directed mutagenesis, thus making the protein amenable for PRE studies. A disadvantage of these spin labels is that the need to introduce single cysteine at several positions is a time- and labor-intensive process and might perturb the ensemble conformations and populations especially for the unfolded polypeptides. Another limitation of using such labels is the possibility to form hydrophobic clusters (Card *et al.*, 2005) particularly for denatured proteins with solvent-accessible sites. Other alternatives that can be used as spin labels are amino-terminal Cu^{2+} - Ni^{2+} -binding (ATCUN) motif, which binds paramagnetic Cu^{2+} with very high affinity (Donaldson *et al.*, 2001) and metal ions such as Fe^{2+} or Mn^{2+} chelated to EDTA (Iwahara *et al.*, 2003, Iwahara & Clore, 2006).

Residual Dipolar Couplings

As RDCs have been discussed in detail in the first chapter, this section focuses on their application to unfolded proteins and model peptides. Further, their interpretation with statistical models is described.

RDCs as a tool for the study of unfolded or denatured states

In recent years, RDCs have characterized residual structure in natively unfolded proteins under different alignment media and conditions. For example, RDCs have revealed β -turn propensities above the melting transition of the T4 fibritin foldon β -hairpin (Guthe *et al.*, 2004, Meier *et al.*, 2004). Dependence of salt or temperature on the destabilization of

first α -helix was reported for α -helical ribonuclease S-peptide (Alexandrescu & Kammerer, 2003) as observed by gradual decrease in the size of RDCs. Bertoncini *et al.* have characterized long-range interactions and dynamics in monomeric α -synuclein (α S) that play a role to inhibit oligomerization and aggregation in the Parkinson's disease (Bertoncini *et al.*, 2005). Local regions of enhanced flexibility or chain compaction were characterized in urea-denatured apomyoglobin by a decrease in the magnitude of the residual dipolar couplings (Mohana-Borges *et al.*, 2004).

Alongside natively unstructured proteins, several denatured proteins and model peptides have been investigated by RDCs. The initial report by Shortle and Ackerman showed well observable RDCs for the Δ 131 Δ fragment of staphylococcal nuclease under strongly denaturing conditions in 8 M urea (Shortle & Ackerman, 2001, Ackerman & Shortle, 2002b, Ackerman & Shortle, 2002a). This study argued that denatured proteins retain some native-like topology in the denatured state. However, more recent trends to interpret experimental RDCs by theoretical models (discussed in the following section) may result in for some revision of this initial interpretation. Later, the Shortle group characterized the long-range structure that persists in the urea-denatured form of the 70-residue protein eglin C (Ohnishi *et al.*, 2004). Recently, a native-like local structure in the N-terminal β -hairpin (Meier *et al.*, 2007b) have been shown. This study was based on previous reports on long-range dipolar couplings in deuterated ubiquitin (Wu & Bax, 2002, Meier *et al.*, 2003).

In contrast to the above studies, very small RDCs measured for the low pH, thermally unfolded state of protein GB1 showed no evidence of any native-like structure (Ding *et al.*, 2004). In another study, non-zero RDCs for the unfolded state of apomyoglobin have been argued in favor of local conformational propensities (Mohana-Borges *et al.*, 2004). For short peptides, small but non-zero RDCs have been observed and interpreted as local stiffness of the backbone (Ohnishi & Shortle, 2003).

The interpretation of RDC data from unstructured polypeptides is complicated due to heterogeneity in the conformational ensemble. A few theoretical models have been developed in order to provide a solid foundation for data analysis: 1) RDCs of unfolded

polypeptides have been described theoretically by polymer random flight models (Louhivuori *et al.*, 2003, Louhivuori *et al.*, 2004), and more recently by 2) two similar statistical models derived from coil subsets of the Protein Data Bank (Jha *et al.*, 2005, Bernado *et al.*, 2005). At present, the most successful model for prediction of RDCs is this statistical coil model, which reproduces experimental data (especially $^1D_{NH}$) to a high degree of agreement (Bernado *et al.*, 2005, Meier *et al.*, 2007b). These statistical models use random sampling of distinct amino-acid-specific (ϕ, φ) propensities to generate an ensemble of conformers of the protein of interest. The amino-acid-specific conformational energy basins were derived from the coil regions, which are classified as non-alpha helix and non-beta sheet regions, of high-resolution X-ray structures in the protein data base (Fitzkee & Rose, 2004).

The two statistical models differ slightly in their approach. In the case of Jha *et al.*, nearest neighbour effects were explicitly taken into account, which led to a significant improvement in the agreement between calculated and experimental RDCs. In the case of Bernado *et al.*, steric overlap was avoided by residue-specific volume exclusion and reasonable agreement with experimental data was obtained. Differences between experimental and calculated RDCs can be interpreted as a valuable source of information on secondary structure and tertiary contacts (Mukrasch *et al.*, 2005).

Some of the previous reports have demonstrated that unfolded or denatured states retain small, but significant, local conformational preferences along the amino-acid sequence. In that case, substitutions of specific amino acids should have substantial influence on the RDC patterns. Prior to this thesis, it had not been studied in detail. In the following section, a systematic investigation of conformational preferences of individual amino acids and/or their side-chains on the local or global order is carried out.

Section 3.1:

Conformational preferences of individual amino acids in short peptides revealed by residual dipolar couplings

The initial report on RDCs of short peptides by Ohnishi and Shortle mainly interpreted the RDCs as a function of local stiffness of the backbone (Ohnishi & Shortle, 2003). Their studies, however, lacked details of amino-acid-specific effects. In this section, a systematic study of the influence of amino acid substitutions X on the model peptide EGAXXAASS is described, as monitored by backbone RDCs.

The sequence used for the study was designed with the aim of providing neutral next neighbor alanine residues for X and making the peptides water-soluble by hydrophilic residues at the N- and C-terminal ends. $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs were detected at natural abundance of ^{15}N and ^{13}C in strained polyacrylamide gels (Sass *et al.*, 2000, Tycko *et al.*, 2000). Overall, 14 single amino acids, and a β -turn sequence KNGE were substituted for investigation.

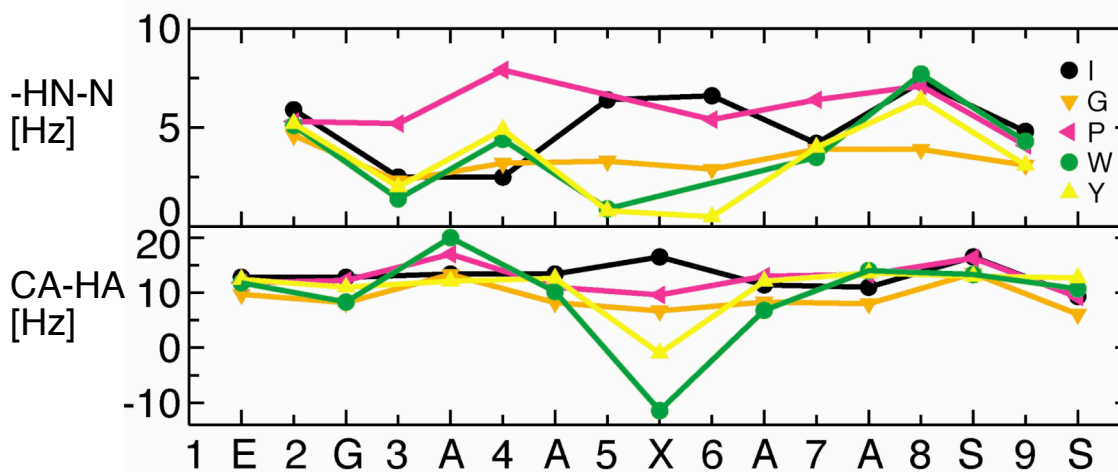


Figure 3.1.1 Residual order in model peptide EGAXXAASS. $^1D_{NH}$ (above) and $^1D_{C\alpha H\alpha}$ (lower) RDCs are shown for five typical amino acid substitutions. For X = G substitution, only one $^1H^\alpha$ resonance was observable. For clarity, error bars were omitted. Statistical errors from repeated experiments were ± 0.8 Hz for $^1D_{NH}$ and ± 1.5 Hz for $^1D_{C\alpha H\alpha}$.

The data show an amino-acid-specific dependence on the substitution X that correlates to steric or hydrophobic interactions with adjacent residues. In particular, smaller or stronger RDCs for glycine and proline substitutions indicate less or more order, respectively, than other amino acids. Substitution of amino acids with aromatic side-chains, such as Trp and Tyr, gives evidence of a kink in the peptide backbone (Figure 3.1.1). The substitution by the β -turn sequence KNGE shows differences in the RDC pattern from the single amino acid substitution results. Predictions from the statistical models (Bernado *et al.*, 2005) of unfolded polypeptides reproduced the overall $^1D_{NH}$ RDC pattern for most substitutions. The predictions for $^1D_{C\alpha H\alpha}$ RDCs show strong deviations from the experimental data that may be related to imperfect modeling of the side-chains.

The study reported here clearly shows the influence of individual amino acids and their interactions on the orientational preferences in polypeptides. This work opens the possibilities for the rigorous experimental characterization of the influence of individual amino acids in on the unfolded ensemble.

The detailed description of this work has been published in the article by Dames *et al.* 2006.

Original Publication

Dames S.A, Aregger R., **Vajpai N.**, Bernado P., Blackledge M., and Grzesiek S.

Residual dipolar couplings in short peptides reveal systematic conformational preferences of individual amino acids

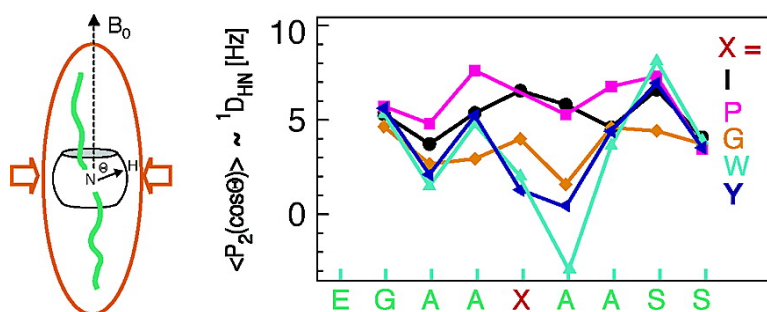
J Am Chem Soc 2006 **128**: 13508-13514

Residual Dipolar Couplings in Short Peptides Reveal Systematic Conformational Preferences of Individual Amino Acids

Sonja Alexandra Dames, Regula Aregger, Navratna Vajpai,
 Pau Bernado, Martin Blackledge, and Stephan Grzesiek

J. Am. Chem. Soc., **2006**, 128 (41), 13508-13514 • DOI: 10.1021/ja063606h • Publication Date (Web): 26 September 2006

Downloaded from <http://pubs.acs.org> on March 31, 2009



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 4 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

[View the Full Text HTML](#)

Residual Dipolar Couplings in Short Peptides Reveal Systematic Conformational Preferences of Individual Amino Acids

Sonja Alexandra Dames,[†] Regula Aregger,[†] Navratna Vajpai,[†] Pau Bernado,[‡]
Martin Blackledge,[‡] and Stephan Grzesiek^{*†}

Contribution from the Biozentrum, University of Basel, Switzerland, and
Institut de Biologie Structurale Jean-Pierre Ebel, Grenoble, France

Received May 23, 2006; E-mail: stephan.grzesiek@unibas.ch

Abstract: Residual dipolar couplings (RDCs) observed by NMR in solution under weak alignment conditions can monitor average net orientations and order parameters of individual bonds. By their simple geometrical dependence, RDCs bear particular promise for the quantitative characterization of conformations in partially folded or unfolded proteins. We have systematically investigated the influence of amino acid substitutions X on the conformation of unfolded model peptides EGAAAXAASS as monitored by their $^1\text{H}^{\alpha}$ – ^{15}N and $^1\text{H}^{\alpha}$ – $^{13}\text{C}^{\alpha}$ RDCs detected at natural abundance of ^{15}N and ^{13}C in strained polyacrylamide gels. In total, 14 single amino acid substitutions were investigated. The RDCs show a specific dependence on the substitution X that correlates to steric or hydrophobic interactions with adjacent amino acids. In particular, the RDCs for the glycine and proline substitutions indicate less or more order, respectively, than the other amino acids. The RDCs for aromatic substitutions tryptophane and tyrosine give evidence of a kink in the peptide backbone. This effect is also observable for orientation by Pf1 phages and corroborated by variations in $^{13}\text{C}^{\alpha}$ secondary shifts and $^3J_{\text{HNH}\alpha}$ scalar couplings in isotropic samples. RDCs for a substitution with the β -turn sequence KNGE differ from single amino acid substitutions. Terminal effects and next neighbor effects could be demonstrated by further specific substitutions. The results were compared to statistical models of unfolded peptide conformations derived from PDB coil subsets, which reproduce overall trends for $^1\text{H}^{\alpha}$ – ^{15}N RDCs for most substitutions, but deviate more strongly for $^1\text{H}^{\alpha}$ – $^{13}\text{C}^{\alpha}$ RDCs. The outlined approach opens the possibility to obtain a systematic experimental characterization of the influence of individual amino acid/amino acid interactions on orientational preferences in polypeptides.

Introduction

Weak alignment of molecules dissolved in anisotropic liquid phases¹ has become a powerful tool to directly monitor average net orientations and order parameters of individual bonds by residual dipolar couplings (RDCs). RDCs are proportional to the ensemble average $\langle 3 \cos^2(\Theta) - 1 \rangle / 2$, where Θ is the angle between the internuclear vector and the magnetic field in the laboratory frame. Similar to other applications in physical chemistry,² $\langle 3 \cos^2(\Theta) - 1 \rangle / 2$ can be interpreted as a local order parameter S of the internuclear vector relative to an external director. S adopts a value of 1, if there is perfect alignment of the bond along the magnetic field, $-1/2$ if there is perfect alignment perpendicular to the magnetic field, and 0 if for example all orientations are equally probable or if there is perfect alignment along the magic angle $\Theta = 54.7^\circ$. By this geometrical dependence and because usually many different RDCs can be determined, RDCs bear particular promise for a detailed,

quantitative characterization of conformations in partially folded or unfolded proteins.

Thus RDCs have revealed residual structure in urea-denatured forms of staphylococcal nuclease³ and natively unfolded alpha-synuclein,⁴ α -helix propensities in the unfolded S-peptide,⁵ the acyl-coenzyme A binding protein⁶ and myoglobin,⁷ and β -turn propensities above the melting transition of the T4 fibritin foldon β -hairpin.⁸ For shorter peptides, modest RDCs have been observed and interpreted as a local stiffness of the backbone.⁹ RDCs of unfolded polypeptides have been described theoretically by polymer random flight models^{10–12} and more recently

- (3) Shortle, D.; Ackerman, M. S. *Science* **2001**, *293*, 487–489.
- (4) Bertocini, C. W.; Jung, Y. S.; Fernandez, C. O.; Hoyer, W.; Griesinger, C.; Jovin, T. M.; Zweckstetter, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 1430–1435.
- (5) Alexandrescu, A. T.; Kammerer, R. A. *Protein Sci.* **2003**, *12*, 2132–2140.
- (6) Fieber, W.; Kristjansdottir, S.; Poulsen, F. M. *J. Mol. Biol.* **2004**, *339*, 1191–1199.
- (7) Mohana-Borges, R.; Goto, N. K.; Kroon, G. J.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2004**, *340*, 1131–1142.
- (8) Mejer, S.; Guthe, S.; Kiefhaber, T.; Grzesiek, S. *J. Mol. Biol.* **2004**, *344*, 1051–1069.
- (9) Ohnishi, S.; Shortle, D. *Proteins* **2003**, *50*, 546–551.
- (10) Louhivuori, M.; Paakkonen, K.; Fredriksson, K.; Permi, P.; Lounila, J.; Annala, A. *J. Am. Chem. Soc.* **2003**, *125*, 15647–15650.

[†] University of Basel.

[‡] Institut de Biologie Structurale Jean-Pierre Ebel.

(1) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111–1114.

(2) Doruker, P.; Mattice, W. L. *J. Phys. Chem. B* **1999**, *103*, 178–183.

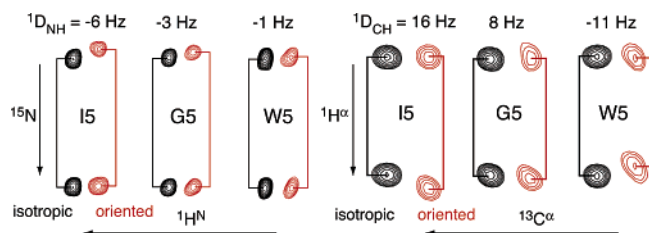


Figure 1. Detection of amino acid specific order in peptides EGAXAASS from RDCs. ${}^1D_{\text{NH}}$ and ${}^1D_{\text{CAHA}}$ RDCs are obtained from the difference in doublet splittings of nondecoupled natural abundance ${}^1\text{H}-{}^{15}\text{N}$ (left) and ${}^1\text{H}-{}^{13}\text{C}$ HSQCs (right) of peptides EGAXAASS ($X = \text{I, G, W}$) under isotropic (black) and anisotropic (red) conditions in mechanically strained polyacrylamide gels. ${}^1\text{H}$ (${}^{13}\text{C}$) decoupling was omitted during the ${}^{15}\text{N}$ (${}^1\text{H}^\alpha$) evolution period.

by statistical models derived from coil subsets of the Protein Data Bank.^{13,14} At present, a systematic experimental characterization of the influence of individual amino acids on the RDC-derived local and global order of polypeptides is lacking.

Here, we have systematically investigated the effect of various single and multiple amino acid substitutions on the conformation of short peptides as monitored by ${}^1\text{H}^{\text{N}}-{}^{15}\text{N}$ and ${}^1\text{H}^\alpha-{}^{13}\text{C}^\alpha$ RDCs detected at natural abundance of ${}^{15}\text{N}$ and ${}^{13}\text{C}$ in strained polyacrylamide gels.^{15,16} The RDCs show specific dependencies on the amino acid substitutions that for the investigated cases correlate to steric or hydrophobic interactions with adjacent amino acids. The determination of RDC-derived order parameters in conjunction with a systematic variation of the amino sequence opens the possibility for a rigorous experimental characterization of individual amino acid/amino acid interactions in polypeptides.

Results

The peptides used in this study were all derived from the sequence EGAXAASS where X is the amino acid under investigation. This sequence was based on the rationale of providing neutral next neighbor alanine residues for X and making the peptides water-soluble by hydrophilic residues at their N- and C-terminal ends. In total, 14 amino acids X were investigated; they comprise G, V, L, I, P as aliphatic; T, N, Q as polar; K, D, E as charged; and Y, W, H as aromatic residues.

${}^1D_{\text{NH}}$ and ${}^1D_{\text{CAHA}}$ RDCs of amide ${}^1\text{H}-{}^{15}\text{N}$ and ${}^1\text{H}^\alpha-{}^{13}\text{C}^\alpha$ internuclear vectors were determined as the difference of the respective doublet splittings in nondecoupled ${}^1\text{H}-{}^{15}\text{N}$ and ${}^1\text{H}-{}^{13}\text{C}$ HSQCs of anisotropic and isotropic samples. Figure 1 shows typical examples for the peptides $X = \text{I, G, W}$. Despite a very similar overall alignment (see below), the RDCs of the three peptides vary significantly in the vicinity of amino acid X , e.g., ${}^1D_{\text{CAHA}}$ equals 16 Hz for I5 but -11 Hz for W5.

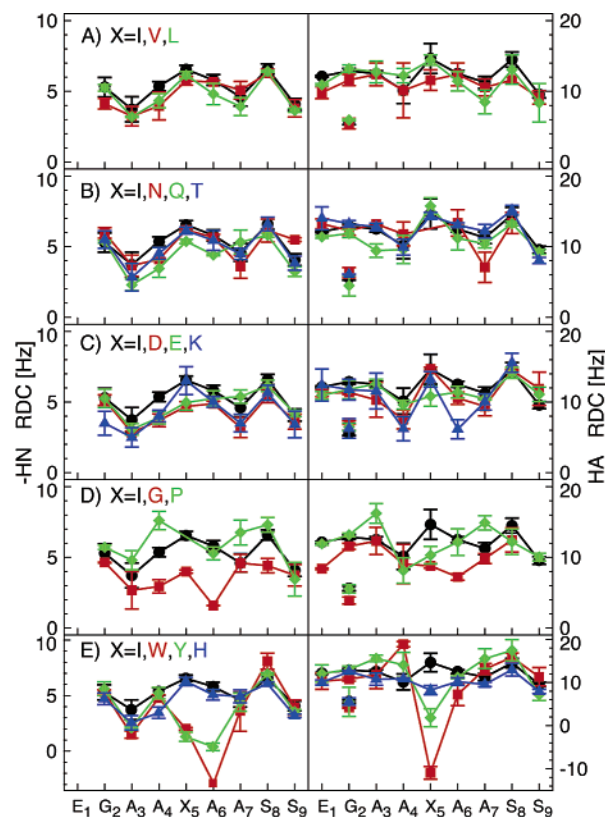


Figure 2. Sequential ${}^1D_{\text{NH}}$ (left) and ${}^1D_{\text{CAHA}}$ (right) RDCs in oriented peptides EGAXAASS. Experimental RDCs are given for aliphatic (A), hydrophilic (B), charged (C), G and P (D), and aromatic (E) substitutions of the residue X . The specific amino acid substitutions are marked. For comparison, the behavior of the $X = \text{I}$ substitution is shown in all panels. For G2, the separately observable ${}^1D_{\text{CAHA}}$ RDCs of both ${}^1\text{H}^\alpha$ protons are shown. For the $X = \text{G}$ substitution (D) only one ${}^1\text{H}^\alpha$ resonance is observable. The corresponding average ${}^1D_{\text{CAHA}}$ is shown (see text). Error bars indicate statistical errors from repeated experiments.

Sequential RDCs. Sequential ${}^1D_{\text{NH}}$ and ${}^1D_{\text{CAHA}}$ RDCs for the 14 investigated peptides EGAXAASS are shown in Figure 2. It is evident that the aliphatic side chains of the amino acids I, V, and L at position X5 (Figure 2A) all lead to very similar RDC profiles and thus indicate similar average orientations of the ${}^1\text{H}^\alpha-{}^{13}\text{C}^\alpha$ and ${}^1\text{H}^{\text{N}}-{}^{15}\text{N}$ internuclear vectors. The absolute values of ${}^1D_{\text{NH}}$ show a bell-shaped, almost 2-fold increase in the center of the peptide (A4, X5, A6). This increase is equivalent to an increase in the average $\langle 3 \cos^2(\Theta) - 1 \rangle / 2$ for the respective N-H bond vectors, which may be caused by a stiffening and/or a kink of the backbone due to the larger side chain at position X5. Toward both peptide termini the ${}^1D_{\text{NH}}$ RDCs decrease. However, the penultimate residues G2 and S8 show increased ${}^1D_{\text{NH}}$ RDCs, which are analyzed in more detail below. The increase for these residues deviates from the expected profile for a random chain polymer,¹¹ where terminal fraying causes a continuous bell-shaped decrease of RDCs from the peptide center toward the termini.

In comparison to the ${}^1D_{\text{NH}}$ RDCs, the variation of ${}^1D_{\text{CAHA}}$ RDCs along the peptide chain is less pronounced. A weak (about 20%) increase in ${}^1D_{\text{CAHA}}$ is also observed at position X5 for the case of $X = \text{I}$ and L, but it is undetectable for $X = \text{V}$. It is remarkable that in all investigated cases the two G2 ${}^1\text{H}^\alpha$ nuclei showed distinguishable resonances and strongly differing ${}^1D_{\text{CAHA}}$ values. This indicates that the averages of the two ${}^1\text{H}^\alpha-$

- Louhivuori, M.; Fredriksson, K.; Paakkonen, K.; Permi, P.; Annala, A. *J. Biomol. NMR* **2004**, *29*, 517–524.
- Fredriksson, K.; Louhivuori, M.; Permi, P.; Annala, A. *J. Am. Chem. Soc.* **2004**, *126*, 12646–12650.
- Jha, A. K.; Colubri, A.; Freed, K. F.; Sosnick, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099–13104.
- Bernado, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17002–17007. Epub 12005 Nov 17011.
- Tycko, R.; Blanco, F. J.; Ishii, Y. *J. Am. Chem. Soc.* **2000**, *122*, 9340–9341.
- Sass, H. J.; Musco, G.; Stahl, S. J.; Wingfield, P. T.; Grzesiek, S. *J. Biomol. NMR* **2000**, *18*, 303–309.

$^{13}\text{C}^\alpha$ directions are not identical on the time scale of the inverse of the RDC value (hundreds of milliseconds) and hence that backbone flexibility at the position of G2 is restricted. Unfortunately, an experimental stereoassignment of the $^1\text{H}^{\alpha 2}$ and $^1\text{H}^{\alpha 3}$ nuclei ($\text{H}^{\alpha 2}$ corresponds to H^α in nonglycine amino acids) was not possible, since both showed very similar ROEs and $^3J_{\text{HNH}\alpha}$ coupling constants.

The substitution of residue X by polar residues N, Q, and T (Figure 2B) or by the charged residues D, E, and K (Figure 2C) does not alter the observed profile of sequential $^1D_{\text{NH}}$ and $^1D_{\text{CAHA}}$ RDCs in a significant way. For all of these substitutions, there is a very similar increase of $^1D_{\text{NH}}$ RDCs in the center of the peptide as well as at G2 and S8. Likewise, less pronounced variations are observed for the $^1D_{\text{CAHA}}$ RDCs.

A significantly different profile is observed for the substitutions $X = \text{G}, \text{P}$ (Figure 2D). When the side chain at position X5 is replaced by the hydrogen of a glycine, the (absolute) increase of $^1D_{\text{NH}}$ RDC values for residues 5, 6, and 8 is abolished and the $^1D_{\text{CAHA}}$ RDC profile becomes even flatter. Due to averaging or symmetry, the two $^1\text{H}^\alpha$ resonances of G5 are completely indistinguishable even at 800 MHz ^1H frequency, and Figure 2D shows only their average $^1D_{\text{CAHA}}$. In strong contrast to glycine, the proline substitution induces markedly elevated (absolute) $^1D_{\text{NH}}$ RDCs for the preceding residue A4 and for residue A7. In addition, there is a significant increase of $^1D_{\text{CAHA}}$ RDCs for A3 and A7. The observed behavior is consistent with the expectation that substitution by the least sterically hindered glycine should lead to a loss of local order, whereas the proline ring should induce additional orientational order by a stiffening of the polypeptide chain.¹¹ A comparable increase of local RDCs around prolines has also been observed in the case of the thermally unfolded β -hairpin of foldon.⁸

The most significant variations in RDC profile were observed for the aromatic substitutions $X = \text{W}, \text{Y}$ (Figure 2E). As compared to the adjacent amino acids, X5 and A6 $^1D_{\text{NH}}$ as well as X5 $^1D_{\text{CAHA}}$ RDCs are strongly reduced and even change sign to slightly [$X5 = \text{Y}$: $^1D_{\text{NH}}(\text{A6})$] or even pronounced [$X5 = \text{W}$: $^1D_{\text{CAHA}}(\text{W5})$, $^1D_{\text{NH}}(\text{A6})$] negative values. The abrupt changes of RDCs along the polypeptide chain indicate a strong kinking or bulging of the peptide backbone around residues X5 and A6.¹¹ No such pronounced behavior was observed for the substitution of X5 by histidine (Figure 2E), which displays a profile similar to that of isoleucine albeit with a reduced $^1D_{\text{CAHA}}$ RDC. Unfortunately, it was not possible to test the substitution by phenylalanine, since this peptide was not sufficiently water-soluble for natural abundance ^{13}C and ^{15}N detection.

$^{13}\text{C}^\alpha$ Secondary Shifts and $^3J_{\text{HNH}\alpha}$ Scalar Couplings. To obtain additional insights into the intrinsic conformational preferences of the investigated peptides, their $^{13}\text{C}^\alpha$ secondary shifts were also analyzed (Figure 3, left side). In general, the $^{13}\text{C}^\alpha$ secondary shifts for all peptides are very similar and close to random coil values. In all cases, N- and C-terminal $^{13}\text{C}^\alpha$ shifts are altered by -0.6 ppm and $+2$ ppm, respectively, due to the terminal $-\text{NH}_3^+$ and $-\text{COO}^-$ substitutions. As expected, the presence of the proline ring for the substitution $X5 = \text{P}$ shifts the $^{13}\text{C}^\alpha$ resonance of the preceding A4 by -2 ppm (Figure 3D). It seems remarkable that, in most cases (with the exception of G, N, H), the X5 residues have moderately negative (-0.5 ppm) $^{13}\text{C}^\alpha$ secondary shifts, whereas their A4 and A6 neighbors have $^{13}\text{C}^\alpha$ secondary shifts that are zero or moderately positive. This

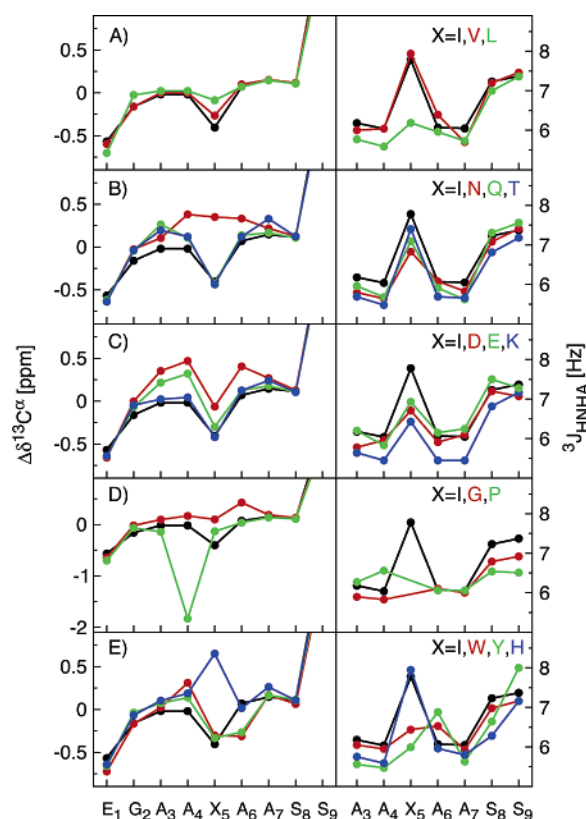


Figure 3. $^{13}\text{C}^\alpha$ secondary shifts ($\Delta\delta^{13}\text{C}^\alpha$) and backbone $^3J_{\text{HNH}\alpha}$ scalar couplings in peptides EGAAXAASS. $\Delta\delta^{13}\text{C}^\alpha$ values (left) are calculated as the difference of the observed shift and the random coil shift. $^3J_{\text{HNH}\alpha}$ values (right) were determined from resolved splittings of the $^1\text{H}^\text{N}$ resonances in $^1\text{H}^\text{N}$ - ^{15}N HSQCs at natural abundance of ^{15}N . Data are shown for aliphatic (A), hydrophilic (B), charged (C), G and P (D), and aromatic (E) substitutions of the residue X. For comparison, data of the $X = \text{I}$ substitution are shown in all panels.

may indicate that the X5 residues adopt a slightly more extended conformation than their neighbors.¹⁷ A further remarkable feature is that the $^{13}\text{C}^\alpha$ secondary shifts of A6 are reduced to about -0.3 ppm for the aromatic substitutions $X5 = \text{W}, \text{Y}$ (Figure 3E). No such shift is observed for any of the other peptides. Thus, these aromatic substitutions entail a specific interaction on the C^α position of the next residue, which can be either a conformational preference or a ring current shift. Similarly, the observed variations of the X5 $^{13}\text{C}^\alpha$ secondary shifts for $X5 = \text{H}$ or N must be related to specific interactions of the respective side chains.

In addition to $^{13}\text{C}^\alpha$ shifts also three-bond $^1\text{H}^\text{N}$ - $^1\text{H}^\alpha$ J -couplings ($^3J_{\text{HNH}\alpha}$) were analyzed as an independent indication for phi angle preferences (Figure 3, right side). Consistent with earlier studies and a random coil sampling of conformational space,¹⁸ the observed $^3J_{\text{HNH}\alpha}$ values are in the range from 6 to 8 Hz for residue X5, whereas the adjacent alanine residues A3, A4, A6, and A7 have smaller values ($< \sim 6$ Hz) for all substitutions of X5 with the exception of tyrosine and tryptophane. According to the Karplus relation for $^3J_{\text{HNH}\alpha}$, this indicates that these alanines have less extended phi angles than X5. In contrast, alanine residues A6 for the aromatic substitu-

(17) Spera, S.; Bax, A. *J. Am. Chem. Soc.* **1991**, *113*, 5490–5492.

(18) Schwalbe, H.; Fiebig, K. M.; Buck, M.; Jones, J. A.; Grimshaw, S. B.; Spencer, A.; Glaser, S. J.; Smith, L. J.; Dobson, C. M. *Biochemistry* **1997**, *36*, 8977–8991.

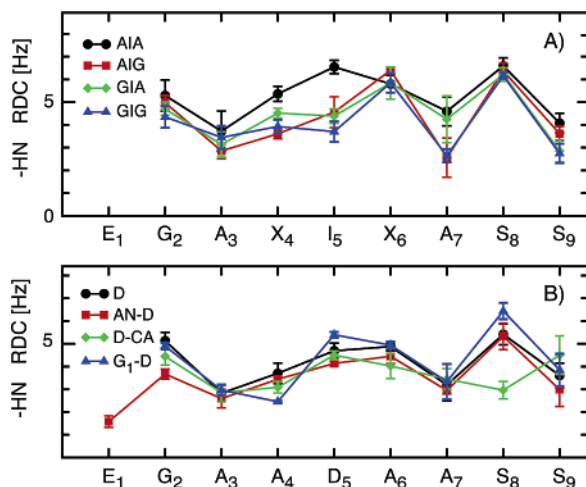


Figure 4. Next neighbor (A) and terminal effects (B) on sequential peptide RDCs. (A) Variation of the central sequence AIA of EGAAIAASS to AIG, GIA, and GIG. (B) Variation of the termini of the peptide EGAADAASS [D] to amidated C-terminus [D-CA], acetylated N-terminus [AN-D], and substitution EIG [G₁-D].

tions X5 = Y and W have larger $^3J_{\text{HNH}\alpha}$ values (~ 6.5 Hz) than X5 (Figure 3E). This signifies that the phi angle of A6 is more extended than in the case of the nonaromatic substitutions. More extended conformations cause negative $^{13}\text{C}^\alpha$ secondary shifts,¹⁷ consistent with the observed values of about -0.3 ppm. Thus both $^{13}\text{C}^\alpha$ shifts and $^3J_{\text{HNH}\alpha}$ values for X5 = Y and W show that the backbone conformation differs from the nonaromatic substitutions.

Next Neighbor Effects. The continuous increase of RDCs around the peptide center for most investigated peptides indicates that orientational preferences are transmitted between neighboring residues. To test the effect of the neighbors next to residue X5, we have replaced alanine residues 4 and 6 by glycines. Figure 4A shows a comparison of the sequential $^1D_{\text{NH}}$ RDCs of the peptide EGAAIAASS and variations of the form EGAGIAASS, EGAAIGASS, and EGAGIGASS. Whereas the overall alignment is still very similar for the GIA, AIG, and GIG substitutions, it is evident that both a preceding or a following glycine strongly reduce the $^1D_{\text{NH}}$ RDC of I5 in the peptide center. In addition, the substitutions A6G (AIG and GIG) reduce the $^1D_{\text{NH}}$ RDC of A7. These observations show that the next neighbors have a profound influence on the local backbone propensities and that the larger side chain of alanine induces larger RDCs for the neighboring residues as compared to the smaller glycine. When interpreted as a dynamic effect, this indicates that alanine restricts the neighboring residue more than glycine.

Terminal Effects. To gain insight into the reason for the elevated RDCs at the penultimate residues G2 and S8, we have replaced the C-terminal carboxylate by carboxamide ($-\text{CO}-\text{NH}_2$, $-\text{CA}$) and the N-terminal ammonium by acetamide ($\text{H}_3\text{C}-\text{CO}-\text{NH}-$, AN-). Figure 4B shows a comparison of the X5 = D peptide with unmodified termini and its AN- and $-\text{CA}$ substitutions. The C-terminal carboxamide abolishes the increased $^1D_{\text{NH}}$ RDC for residue S8. Therefore the increased orientation of the S8 N-H vector can be clearly related to interactions between residue S8 and the C-terminal carboxylate group. For the N-terminal AN- substitution only a minor reduction of the G2 $^1D_{\text{NH}}$ RDC is observed. A replacement of

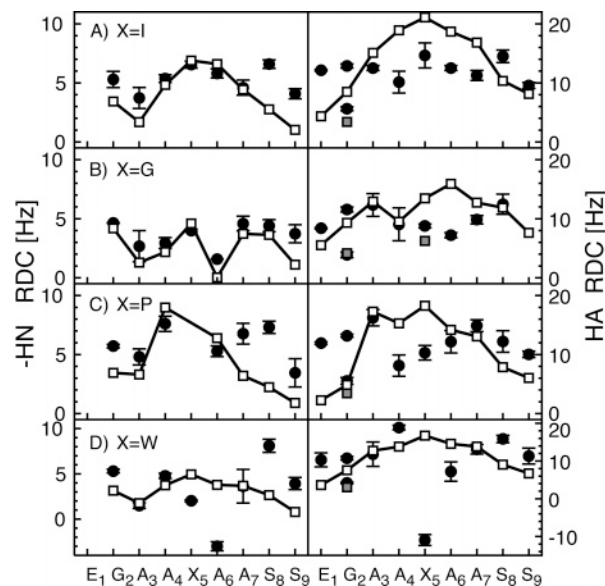


Figure 5. Comparison of experimental RDCs in oriented peptides EGAXAASS and the predictions from the statistical coil model. Experimental $^1D_{\text{NH}}$ (left) and $^1D_{\text{CAHA}}$ (right) RDCs are indicated as filled circles, and predicted RDCs, as open squares. (A–D) Single amino acid substitutions X = I, G, P, and W, respectively. For G2 and G5, predictions of $^1D_{\text{CAHA}2}$ and $^1D_{\text{CAHA}3}$ are shown as open and gray squares, respectively. For each peptide, RDCs of the coil model were scaled by eye such that the $^1D_{\text{NH}}$ RDCs best fitted the experimental data. The same scaling factor was used for $^1D_{\text{NH}}$ and $^1D_{\text{CAHA}}$ RDCs.

the large N-terminal residue glutamic acid by glycine (EIG) did not cause any reduction either (Figure 4B). Hence the reason for the increased G2 $^1D_{\text{NH}}$ RDC is less obvious. However, together with the strongly differing $^1D_{\text{CAHA}}$ RDCs for the two $^1\text{H}^\alpha$ protons, it indicates a particular dynamic behavior or conformational preference for G2. This may be caused either by an uncharged N-terminal effect or by an interaction with the following two alanines (see next paragraph).

Comparison to Random Coil Models. Recently, trends in sequential RDCs of chemically denatured proteins have been reproduced by statistical models derived from backbone conformational frequencies in coil subsets of the Protein Data Bank.^{13,14} This implies that local preferences in denatured proteins are closely related to the torsion angle distribution of non- α , non- β structures in folded proteins. Predictions according to such a statistical model¹⁴ are shown in Figure 5 for the cases X5 = I, G, P, W. For I, G, and P, the $^1D_{\text{NH}}$ RDCs are reproduced quite well from residue G2 to A7. Apparently, the statistical model also predicts an increase of $^1D_{\text{NH}}$ RDCs at residue G2. In the model calculations, the termini were represented by truncated peptide bonds. Therefore the increase at G2 can only be caused by normal sequential interactions. The coil model fails to predict the increase of the $^1D_{\text{NH}}$ RDC at residue S8. This is not surprising, since the C-terminal carboxylate group is not represented properly in the calculations. When the van der Waals radius of the C-terminal carbonyl carbon atom was artificially increased to 2.0 Å, an increase of the $^1D_{\text{NH}}$ RDC of S8 relative to the values at A7 and S9 was observed (data not shown). This corroborates the experimental result that the stronger $^1D_{\text{NH}}$ and $^1D_{\text{CAHA}}$ RDCs at residue S8 are caused by interactions with the C-terminal carboxylate group.

For the X5 = I, G, P substitutions, the agreement between the coil model and the experimental values for $^1D_{\text{CAHA}}$ is

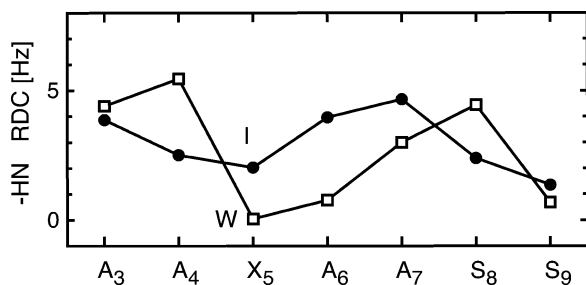


Figure 6. Sequential $^1D_{\text{NH}}$ RDCs of the peptide EGAXAASS observed in Pf1 phages. Data for substitutions $X = \text{I}$ are shown as solid circles, and those for $X = \text{W}$, as open rectangles.

considerably worse than that for $^1D_{\text{NH}}$ (Figure 5). The reason is currently unclear but may be related to the simplified representation of the side chains by a pseudoatom at the position of the β -carbon. Indeed, variations in the radius of this pseudoatom in the calculations lead to significant variations of the predicted $^1D_{\text{CAHA}}$ values (M. Blackledge in preparation).

The statistical model also fails to predict the strong effect of the tryptophane and tyrosine substitutions for both $^1D_{\text{NH}}$ and $^1D_{\text{CAHA}}$ RDCs around residue X5 (Figure 5D). Apparently, this discrepancy indicates genuine differences between the ensemble of conformations in the gel-oriented peptide solution and the statistical coil ensemble derived from folded proteins.

Orientation by Phages. To investigate whether specific interactions with the orienting medium contribute to the strong variations of the RDCs for the aromatic substitutions, the X5 = I and W peptides were also oriented in a suspension of Pf1 phages¹⁹ (Figure 6). Orientation by the phages is caused by both electrostatic and steric interactions with the highly negatively charged phage surface. Therefore the alignment tensor and the RDCs obtained in phage suspensions usually differ from strained gels¹⁶ or DMPC/DHPC bicelles,¹ which predominantly induce steric alignment. Due to aggregation problems of the phages at the pH of 4.5 used in the gel experiment, the pH had to be raised to 6.0. This rendered the $^1\text{H}^{\text{N}}$ resonances of G2 unobservable because of chemical exchange with water protons.

The sequential $^1D_{\text{NH}}$ RDCs for the X5 = I peptide in the phage suspension (Figure 6) differ to some extent from the gel orientation in Figure 2A. The profile appears shifted by one residue toward the C-terminus, such that the maximum is located around residues A6 and A7 and the initial decrease occurs from residue A3 to A4 instead of from G2 to A3. The different RDC pattern may indicate either true conformational differences of the peptide in the two different media caused by specific interactions or a changed orientation tensor from the additional electrostatic interactions with the phage. We prefer the latter explanation since (1) the gel yields very similar RDC profiles for almost all peptides (besides G,P,Y,W), which makes specific interactions of the isoleucine peptide with the gel unlikely and (2) the highly negatively charged phages are not expected to interact strongly with the rather hydrophobic isoleucine peptide.

For the X5 = W peptide the $^1D_{\text{NH}}$ RDCs in phages (Figure 6) closely resemble the $^1D_{\text{NH}}$ RDCs in the strained gels (Figure 2E). Specifically there is also a very sharp decrease of the RDCs after residue A4 to very small values at residues W5 and A6

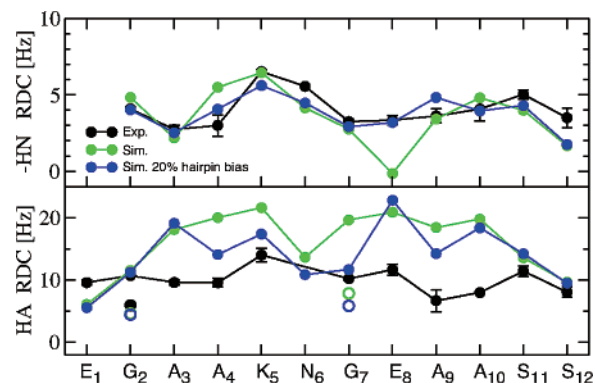


Figure 7. Sequential $^1D_{\text{NH}}$ RDCs in the oriented peptide EGAAKNGEAASS containing the foldon KNGE β -turn sequence. Black, experimental data; green, prediction from unbiased coil PDB subset; blue, prediction from a PDB subset biased by 20% toward KNGE β -turns. Predictions for $^1D_{\text{CAHA}2}$ and $^1D_{\text{CAHA}3}$ of G2 and G5 are shown as closed and open circles, respectively.

followed by an increase at residue A7. Since this abrupt change in $^1D_{\text{NH}}$ RDCs between A4 and W5 is observed for both alignment media, it is implausible that the evident kink or bulge in the backbone at this position is caused by specific interactions with the medium.

Hairpin Substitution. The characterization of orientational preferences in designed peptides by RDCs may ultimately make it possible to follow the peptide folding process depending on sequence and conditions from slight local preferences to more extended stable local structures such as α -helices or β -hairpins. Indeed recently, the gradual loss in local order during thermal denaturation of the foldon β -hairpin could be followed from RDCs.⁸ To test whether the foldon β -turn sequence KNGE induces more extended changes in local order within the peptide EGAXAASS, the sequence KNGE was substituted at position X (Figure 7). The resulting pattern of $^1D_{\text{NH}}$ RDCs shows a pronounced maximum at the first two hairpin residues K5 and N6 and a flat profile for most other residues with the exception of the increased values at the penultimate residues G2 and S11 that are also observed for the other peptides. Also similar to the other peptides, the variations in $^1D_{\text{CAHA}}$ RDCs are much less pronounced.

The behavior of the $^1D_{\text{NH}}$ RDCs is well reproduced by the statistical coil model and even better when the coil model is biased by 20% toward KNGE β -turn configurations (Figure 7). In contrast and again similar to the single amino acid substitutions, the coil model deviates significantly for the $^1D_{\text{CAHA}}$ RDCs. The model considerably overestimates $^1D_{\text{CAHA}}$ and predicts a more varied pattern with a dip around residue N6. As mentioned, a possible cause for these deviations is the improper modeling of the side chains by the C^β pseudoatoms.

At the present, the limited number of experimental observations (two RDCs per residue) precludes a more quantitative interpretation of the β -turn data. However, the reasonable agreement between $^1D_{\text{NH}}$ RDCs and the β -turn-biased coil model is compatible with a significant population of β -turn conformations in the experimental ensemble.

Conclusion

In summary, the reported $^1D_{\text{NH}}$ and $^1D_{\text{CAHA}}$ RDCs give evidence that single (or multiple) amino acid substitutions alter significantly local structural preferences in short polypeptide

(19) Hansen, M. R.; Mueller, L.; Pardi, A. *Nat. Struct. Biol.* **1998**, *5*, 1065–1074.

chains. The observed changes for certain substitutions can be rationalized by specific interactions with neighboring amino acids or terminal groups. For most substitutions besides X5 = W and Y, the observed $^1D_{\text{NH}}$ RDCs can be reproduced to a reasonable extent by the statistical coil model. The coil model fails to reproduce $^1D_{\text{CAHA}}$ RDCs in a satisfactory way. This may be related to the imperfect modeling of the side chains.

The observed strong variations of RDCs along the polypeptide chain for the tryptophane and tyrosine substitutions must be the result of a kink or bulge of the peptide backbone between residues X5 and A6.¹¹ Similar strong RDC variations are also observed in phage suspensions used as a second orienting medium. Furthermore, the profiles of $^{13}\text{C}^\alpha$ secondary shifts and $^3J_{\text{HNH}\alpha}$ values obtained under isotropic conditions also deviate from the nonaromatic peptides and point to a more extended conformation of the A6 phi angle. Therefore the observed kink or bulge does not result from specific interactions with the medium. Rather, it must be caused by local interactions of the aromatic side chains with the neighboring alanine amino acids. Indeed, an ROE can be observed between the W5- $^1\text{H}^\epsilon$ and the A6- $^1\text{H}^\alpha$ protons (data not shown). The statistical coil model does not reproduce these strong RDC variations for the aromatic substitutions. Thus there are genuine differences between the conformational ensemble of the unfolded peptides in solution and the current statistical coil model. In contrast, preliminary results from molecular dynamics simulations on fully hydrated peptides are in closer agreement with the observed RDC pattern (N. Vajpai personal observations).

In principle, it should be possible to obtain strong bounds or even entire probability distributions for the torsion angles of an unfolded polypeptide chain from a sufficiently large number of measured RDCs. Currently, this problem is underdetermined since only two parameters per amino acid ($^1D_{\text{NH}}$ and $^1D_{\text{CAHA}}$) were measured at natural abundance of ^{13}C and ^{15}N . Previously, we have shown that a very larger number ($> \sim 10$ per residue) of short-range and long-range RDCs can be detected by the combined use of ^{13}C - and ^{15}N -labeling and perdeuteration.^{20,21} The use of other orienting media with different alignment tensors yields additional, independent RDCs of similar number. Such an enormous amount of information has previously yielded new insights on collective slow motions in the folded structure of protein G.²² When applied to unfolded polypeptides, a similar extensive collection of RDCs should make it possible to overcome the parameter underdetermination and derive distributions for a large number of torsion angles. This would allow a high-resolution thermodynamic description of the structural ensemble of an unfolded polypeptide chain. Current efforts in our laboratories are directed to such a goal.

Experimental Section

Sample Preparation. HPLC-grade, chemically synthesized peptides EGAXAASS without isotopic enrichment were obtained from a commercial source and used without further purification. Isotropic samples were prepared as solutions of typically 3–6 mM peptide dissolved in 10 mM sodium acetate-*d*₅, pH 4.5, 95/5% H₂O/D₂O.

Anisotropic conditions were achieved as described²³ by introducing the buffered peptide solutions (final concentration 3 mM) into acrylamide gels (10% w/v) and horizontal compression (aspect ratio 2.9) in NEW-ERA sample tubes. Residual alignment by Pf1 phages¹⁹ was achieved at 20 mg/mL phage concentration, pH 6.0.

NMR Experiments. All NMR experiments were carried out at 25 °C on a Bruker DMX800 MHz spectrometer equipped with a TCI cryoprobe. Peptide resonances were assigned from a combination of ROESY, TOCSY, and natural abundance $^1\text{H}^\alpha$ - $^{13}\text{C}^\alpha$ and $^1\text{H}^{\text{N}}$ - ^{15}N HSQC spectra. $^1D_{\text{NH}}$ and $^1D_{\text{CH}}$ RDCs were obtained as the difference in ^1H - ^{15}N and ^1H - ^{13}C doublet splittings observed under anisotropic and isotropic conditions. The splittings were determined from natural abundance ^1H - ^{15}N (^1H -coupled) and ^1H (^{13}C -coupled)- ^{13}C HSQCs where ^1H or ^{13}C decoupling had been omitted during the respective evolution periods. Total experimental times were 6/17 h (isotropic/anisotropic) for the coupled ^1H - ^{15}N and 1.5/5 h for the coupled ^1H - ^{13}C HSQCs. Each experiment was carried out at least twice, and the reported values for the RDCs and the estimates for their statistical errors refer to mean and standard deviations derived from such repeated experiments.

Generation of the Conformational Ensemble. A recently developed algorithm, *Flexible-Meccano*, was used to sample conformational space efficiently. This uses conformational sampling based on amino acid propensity and side chain volume. Consecutive peptide planes and C^α tetrahedral junctions are constructed¹⁴ in the inverse direction to the primary sequence, starting from the C-terminal residue to the N-terminus for each peptide. The position of the peptide plane (*i*) is defined from the C^α and C' atoms of plane (*i* + 1), the selected ϕ/ψ combination, and the tetrahedral angle. Amino acid specific ϕ/ψ combinations are randomly extracted from a database of loop structures, built from 500 high-resolution X-ray structures (resolutions < 1.8 Å and B factors < 30 Å²)²⁴ from which all residues in α -helices and β -sheets were removed. Residues preceding prolines were considered as an additional amino acid type due to their specific sampling properties.²⁵ Amino acid specific volumes were represented by spheres placed at C^β (or C^α for Gly).²⁶ In the case of steric clash with another amino acid of the chain, the ϕ/ψ pair is rejected and another set of ϕ/ψ dihedral angles is selected, until no overlap was found. All ensembles comprise 100 000 conformers, and simulated properties are averaged over all members.

In addition to random sampling of the residue specific ϕ/ψ distribution, additional conformational wells with a specific width and center of the ϕ/ψ distribution can also be sampled. Random conformations are then selected from these distributions at the specified rate (for example 1 in 5 conformers for 20% of the population). In the case of the β -turn, dihedral angles from the NMR-determined structure of foldon were used to define the center of the distribution, and a standard deviation of $\pm 20^\circ$ was used to define the width. The presence of the structural element is cooperative; that is to say if residue 5 adopts a β -turn conformation, residues 6–8 also adopt this conformation. The remaining amino acids follow the standard residue specific sampling described above. In 20% of the conformers residues 5–8 were constrained to β -turns, and 80% follow the residue specific sampling.

Prediction of RDCs from the Conformational Ensemble

The alignment tensor for each member of the ensemble was calculated based on the assumption of steric alignment using the program PALES.²⁷ RDCs were calculated with respect to

(20) Meier, S.; Haussinger, D.; Jensen, P.; Rogowski, M.; Grzesiek, S. *J. Am. Chem. Soc.* **2003**, *125*, 44–45.
(21) Jensen, P.; Sass, H. J.; Grzesiek, S. *J. Biomol. NMR* **2004**, *30*, 443–450.
(22) Bouvignies, G.; Bernado, P.; Meier, S.; Cho, K.; Grzesiek, S.; Bruschweiler, R.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13885–13890. Epub 12005 Sep 13819.

(23) Chou, J. J.; Gaemers, S.; Howder, B.; Louis, J. M.; Bax, A. *J. Biomol. NMR* **2001**, *21*, 377–382.
(24) Lovell, S. C.; Davis, I. W.; Arendall, W. B., III; de Bakker, P. I.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. *Proteins* **2003**, *50*, 437–450.
(25) MacArthur, M. W.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218*, 397–412.
(26) Levitt, M. *J. Mol. Biol.* **1976**, *104*, 59–107.
(27) Zweckstetter, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 3791–3792.

this tensor using the relationship:

$$D_{ij}(\theta, \phi) = -\frac{\gamma_i \gamma_j \mu_0 h}{16\pi^3 r_{ij, \text{eff}}^3} \left[A_a (3 \cos^2 \theta - 1) + \frac{3}{2} A_r \sin^2 \theta \cos 2\phi \right]$$

where A_a and A_r are the axial and rhombic components of the alignment tensor, and θ and ϕ are the polar angles

of the vector with respect to the tensor principal axes. Effective RDCs were averaged over all values from the 100 000 conformers.

Acknowledgment. This work was supported by SNF Grants 31-61757 and 31-109712 (S.G.) and a fellowship from the Treubel Fonds (S.A.D.).

JA063606H

Section 3.2:

Residual dipolar couplings of nonapeptides as predicted from molecular dynamic simulations

Navratna Vajpai and Stephan Grzesiek

Biozentrum, University of Basel, Basel, Switzerland

Introduction

In the previous section, experimental RDCs in oriented peptides EGAAXAASS were compared to the RDCs predicted from the statistical coil model (Bernado *et al.*, 2005). In this case, RDCs for each member of the statistical ensemble were predicted assuming steric alignment (Zweckstetter & Bax, 2000) and then averaged over all values from the 100 000 structures. For most amino acid substitutions for X, a reasonable agreement is obtained for $^1\text{D}_{\text{NH}}$ RDCs; however, $^1\text{D}_{\text{C}\alpha\text{H}\alpha}$ RDCs show a strong deviation, especially for aromatic substitutions (Dames *et al.*, 2006). The reason for this discrepancy is currently unclear but may be related to the simplified representation of the amino acid side chains by a pseudo atom at the β -carbon.

Since the theoretical predictions according to the statistical coil model cannot reproduce the sequential profile of $^1\text{D}_{\text{C}\alpha\text{H}\alpha}$ RDCs, alternatively all-atom molecular dynamics (MD) simulations can be applied. There, the sequential RDC profile can be calculated by averaging RDCs of single snapshots over the entire MD trajectory. The resulting statistical ensemble averages are equal to time averages of the system (ergodic hypothesis).

In recent years, MD simulations have provided detailed information on the fluctuations and conformational changes of unfolded proteins or model peptides in order to gain insight into the folding process (Scheraga *et al.*, 2007). The advances in molecular simulations of protein folding have been discussed recently (Gnanakaran *et al.*, 2003, Swope *et al.*, 2004, Snow *et al.*, 2005, Munoz, 2007). Arguably, it is the only method capable of describing protein dynamics in atomic detail (Karplus & Kuriyan, 2005).

In this study, I have used a MD-generated ensemble of conformations of nonapeptides to investigate local preferences and/or neighboring effects quantitatively. MD simulations were run for a few fully hydrated EGAAXAASS nonapeptides, over a period of 10 ns. These simulations were used to assist in the interpretation of NMR results.

Materials and Methods

Extended chain coordinates and topology files of the EGAAXAASS nonapeptide were generated with the program XPLOR 3.1 (Brunger, A) and then energy minimized. Each nonapeptide was solvated with the TIP3P explicit water model in a ~ 40 Å cubic box (which corresponds to a hydration shell of about 6 Å thickness). Periodic boundary conditions and Particle-Mesh-Ewald electrostatics were used in combination with a nonbonded interaction cutoff of 1 nm. All simulations were performed with NAMD 2.5 (Phillips *et al.*, 2005) using CHARMM c31b1 force field parameters. MD production runs were carried out with time steps of 2 fs after a 2000-step energy minimization and a 200 ps equilibration. Each trajectory spanned 10 ns under constant temperature (300 K) and pressure (1 atm). Snapshots of the trajectory were taken every 200 fs. RDC values were calculated at each time point of the snapshot assuming shape-based steric alignment using the program PALES (Zweckstetter & Bax, 2000), and then ensemble averaged over the trajectory of 10 ns.

Results and Discussion

MD simulations were run for four substitutions G, I, P and W of X in the EGAAXAASS nonapeptide. These amino acids were chosen on the basis of their experimental data that show a significant variation in their RDCs, despite a very similar overall alignment (Dames *et al.*, 2006). Sequential $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs were determined for each nonapeptide as mentioned above (see Materials and Methods).

Comparison with experimental RDC data

Both experimental and MD-derived $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs for the four investigated peptides are shown in Figure 3.2.2. For X=W, results from the MD simulations reproduced to some extent the abrupt decrease in the RDC value at residue position five for both $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs. This decrease indicates a kink or bulge at the centre of

the peptide possibly because of an interaction of aromatic side-chain with neighboring residues. In the VMD (Humphrey *et al.*, 1996) MD trajectory of this nonapeptide, such an interaction can be seen during a major part of the simulation (Figure 3.2.1). However, MD simulations fail to reproduce the strong effect of the aromatic side-chain on the neighboring residue, experimental $^1D_{NH}$ data show strong effect on Ala at position six. For X=P, the experimental RDC values clearly show an induced orientational order or a strong preference for the stiffening of the polypeptide backbone; in contrast, the RDC profile derived from MD simulations is rather flat and does not agree with the experimental data. The reason for this discrepancy is currently unclear. For X=G and X=I, a reasonable agreement, to some extent, between calculated and experimental RDCs was observed along the polypeptide chain. The MD simulations also predicted an increase of RDC values at the middle of the polypeptide chain for X=I. This increase in the RDCs may be caused by the stiffening of the backbone due to the larger side chain. For X=G, a flat RDC profile is obtained from the MD simulations as expected. Furthermore, MD simulations failed to predict an increase in RDCs for the penultimate residues G2 and S8. For these residues, a preferential orientation was earlier shown as interactions with the C-terminal carboxylate for S8 or interactions with neighboring residues for G2. On analyzing the MD trajectory, the termini were found to interact with each other during a significant part of the simulations. Possibly, this led to averaging of RDCs at the termini for the time scale used in the simulations.

In summary, comparison between statistical ensemble averaged RDCs derived from MD simulations and the experimentally determined RDCs show that the overall agreement was far from perfect. Apparently, these differences indicate inaccurate description by MD simulations that may be attributed to improper hydration shell, non-optimal force field or insufficient sampling time used in the MD simulations. The most plausible reason for the discrepancies between simulation and experiment may be insufficient sampling of protein conformational space. This hypothesis corroborates the non-reproducibility of the predicted RDCs for repeated MD simulations of the nonapeptides. Thus, increase in the sampling time may lead us to better reproducibility of the experimental data by an ensemble-averaged observables.

Conclusions

The discrepancy in reproducibility of experimental data by MD simulations indicate genuine deficiencies in the *in silico* predictions. Very likely, this may be attributed to inadequate sampling of the conformational space in the time used for the simulations of nonapeptides. This study has shown that even though the MD derived data does not reproduce the entire profile as obtained from the experiments, the differences in the predictions from MD simulations for investigated substitutions, X= W, P, G and I, in the nonapeptide indicate influence of local amino acid based interactions.

Acknowledgements

I would like to thank Prof. M. Meuwly for introducing me to MD simulations and Dr. Hans Juergen Sass for stimulating discussions and for his help in the set up of MD simulations.

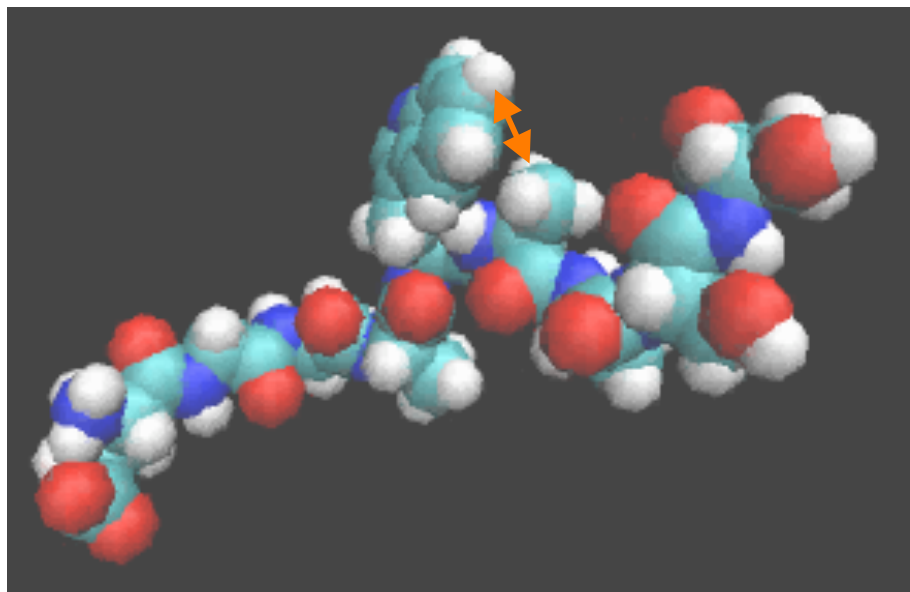


Figure 3.2.1: Snapshot from the MD simulations of X=W in EGAAXAASS. An interaction (marked by an orange arrow) can be seen between W5 and A6 during the simulations. This interaction can be seen for a significant part of the MD simulations.

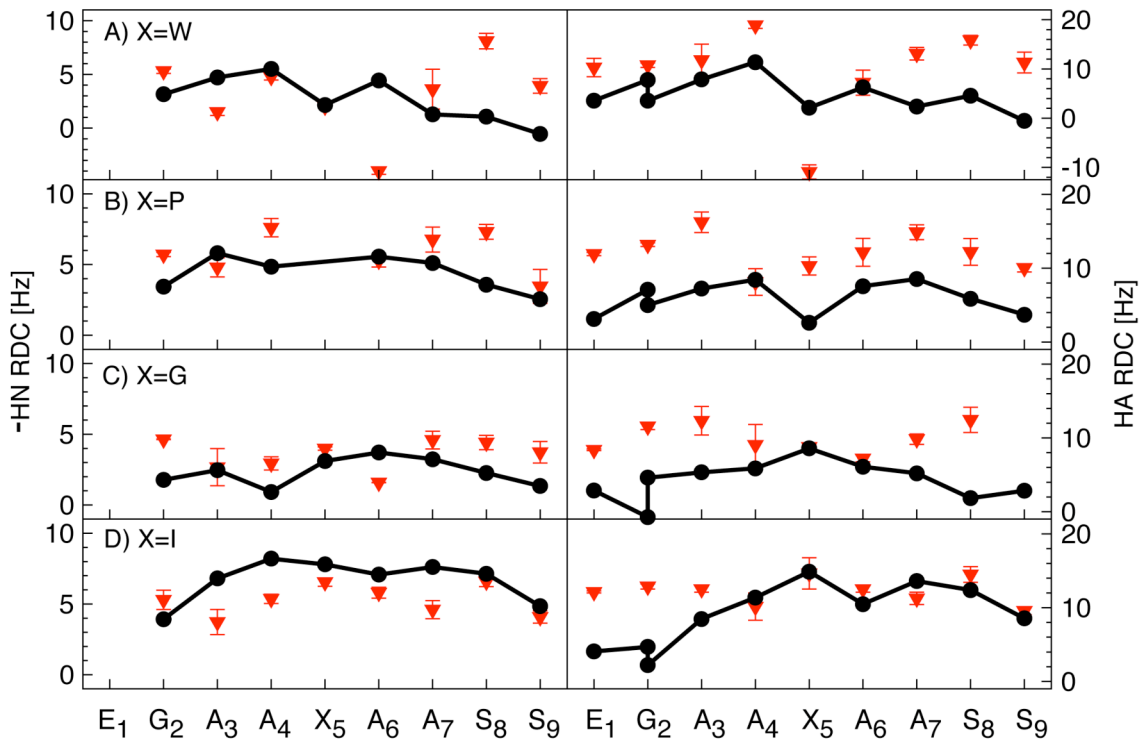


Figure 3.2.2: Comparison of experimental RDCs in oriented peptides EGAAXAASS and predictions from MD simulations. Experimental $^1D_{\text{NH}}$ (left) and $^1D_{\text{C}\alpha\text{H}\alpha}$ (right) RDCs are indicated as filled triangles, and predicted RDCs are shown as filled circles. Predicted RDCs were calculated using PALES (Zweckstetter & Bax, 2000) over the ensemble average of the snapshots taken every 200 fs throughout the trajectory of 10 ns. Errorbars for experimental RDCs indicate the statistical errors from repeated experiments. Statistical ensemble averaged RDCs were calculated from a very large distribution of predicted values. These errors are not shown for clarity reasons.

Chapter 4:

Side-chain χ_1 conformations in urea-denatured proteins: a study by 3J coupling constants and residual dipolar couplings

A detailed, quantitative description of the conformational ensemble of backbone and side-chains of unfolded state is prerequisite for understanding of protein folding. During recent years, backbone dynamics and the residual structure of unfolded or denatured states have been studied in some detail by NMR spectroscopy, reviewed by (Meier *et al.*, 2008), but the information on side-chains of unstructured polypeptides is rather sparse. The studies on side-chains are mainly limited due to extensive conformational averaging of side-chains leading to poor dispersion of signals. For this reason, to date, there are only a few NMR investigations performed on the side-chains of unfolded polypeptides (Mathieson *et al.*, 1999, West & Smith, 1998, Hennig *et al.*, 1999, Choy *et al.*, 2003). These initial studies have provided insight into the side-chain torsion angle distributions and have determined side-chain dynamics in methyl-containing residues in unfolded proteins (Choy *et al.*, 2003).

In a previous study done on urea-denatured hen egg white lysozyme, Hennig *et al.* performed quantitative J -correlation experiments to obtain $^3J_{C'C_\gamma}$ and $^3J_{NC_\gamma}$ coupling constants, which enabled them to show that individual residues in the denatured protein have distinct preferences for certain rotamers that reflect the steric and/or electrostatic properties of the side-chain. These preferences showed good agreement with the statistical coil model (Serrano, 1995, Smith *et al.*, 1996, Schwalbe *et al.*, 1997, Fiebig *et al.*, 1996), derived from the database of non α -helix and non β -sheet structures, indicating that mostly only local interactions persist in this denatured state. Aromatic residues, however, were found to deviate from the predicted populations, a feature attributed to the persistence of non-local hydrophobic clusters in the polypeptide chain, even in the presence of high concentrations of denaturant (Hennig *et al.*, 1999). The above report suggests that for the side-chains of unfolded states, the statistical coil model provides the framework for interpretation of NMR data. The precision in this analysis

was limited by the lack of precise Karplus coefficients for ${}^3J_{NC\gamma}$ and ${}^3J_{C\gamma}$ and the fact that the three populations $p_{-60^\circ, 60^\circ, 180^\circ}$ (two independent parameters because $p_{-60^\circ} + p_{60^\circ} + p_{180^\circ} = 1$) were determined from only two experimental values.

In order to improve the description of χ_1 torsional angle distribution, an extensive set of up to six three-bond scalar couplings (${}^3J_{NH\beta}$, ${}^3J_{C'H\beta}$ and ${}^3J_{H\alpha H\beta}$) and two ${}^1D_{C\beta H\beta}$ residual dipolar couplings (RDCs) were measured (Figure 4.1). Original versions of quantitative J -correlation experiments, ${}^3J_{NH\beta}$ -HNHB, ${}^3J_{C'H\beta}$ -HN(CO)HB, and ${}^3J_{H\alpha H\beta}$ -HAHB(CACO)NH experiments have been optimized for their applicability to unfolded proteins. ${}^1D_{C\beta H\beta}$ RDCs for χ_1 angle information were determined from an HBHA(CO)NH experiment, where an IPAP detection scheme was introduced into the mixed constant time ${}^1H^{\alpha/\beta}$ evolution/ ${}^1H^{\alpha/\beta} \rightarrow {}^{13}C^{\alpha/\beta}$ transfer period.

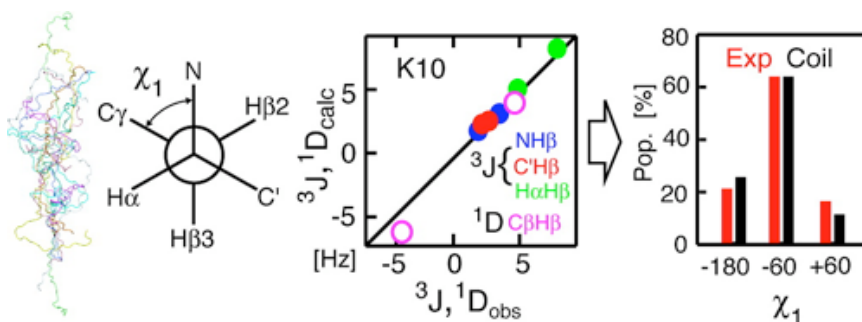


Figure 4.1: General scheme of the study. Data and analysis is shown for a typical residue K10 in urea-denatured protein G. The shown assignments are according to the IUPAC nomenclature. A plot of experimental versus calculated (3J coupling constants and RDCs) is shown. A weighted averaged fit of the data leads to the populations shown in the extreme right window.

The analyses were based on the assumption that only the three staggered rotamers around the side-chain have significant populations. Previously determined amino-acid-specific Karplus coefficients were used to predict ${}^3J_{NH\beta}$, ${}^3J_{C'H\beta}$ and ${}^3J_{H\alpha H\beta}$ couplings for single χ_1 staggered rotamers ($\chi_1 = +60, 180, -60$ degrees). A weighted average fit of these predicted values to the experimental data, enabled us to obtain the stereoassignments of most methylene protons and the populations of the three rotamers. For most residues, the precision of individual χ_1 rotamer populations is better than 2 %. As found in earlier studies (Hennig *et al.*, 1999), the rotamer populations are in vicinity of predictions

according to the statistical coil model. However, individual variations from these averages of up to 40 % are highly significant and indicate sequence- and residue-specific interactions.

Independent analysis of $^1D_{C\beta H\beta}$ RDCs obtained in polyacrylamide gels show good correlation with the RDCs predicted from the χ_1 populations obtained from the 3J data and a coil model ensemble of 50000 conformers according to the coil library backbone angle distribution. Theoretical alignment tensors were generated assuming steric exclusion. These data agree well with the distributions derived from the 3J data and coil library χ_1 distributions. This agreement validates the coil model as a good first approximation of the unfolded state. Furthermore, an analysis of chemical shift dependence on the rotamer distribution was carried out.

The detailed description of this work has been published in the article by Vajpai *et al.* 2010.

Original Publication

Vajpai N., Gentner M., Huang J.-r., Blackledge M., and Grzesiek S.

Side-Chain χ_1 Conformations in Urea-Denatured Ubiquitin and Protein G from 3J Coupling Constants and Residual Dipolar Couplings

J Am Chem Soc 2010 **132 (9)**: 3196-3203

Side-Chain χ_1 Conformations in Urea-Denatured Ubiquitin and Protein G from 3J Coupling Constants and Residual Dipolar Couplings

Navratna Vajpai,[†] Martin Gentner,[†] Jie-rong Huang,[†] Martin Blackledge,[‡] and Stephan Grzesiek^{*†}

Biozentrum, University of Basel, Klingelbergstrasse 50/70, 4056 Basel, Switzerland, and Institut de Biologie Structurale Jean-Pierre Ebel, CEA, CNRS, UJF UMR 5075, 41 Rue Jules Horowitz, Grenoble 38027, France

Received December 7, 2009; E-mail: stephan.grzesiek@unibas.ch

Abstract: Current NMR information on side-chain conformations of unfolded protein states is sparse due to the poor dispersion particularly of side-chain proton resonances. We present here optimized schemes for the detection of $^3J_{\text{H}\alpha\text{H}\beta}$, $^3J_{\text{NH}\beta}$, and $^3J_{\text{C}^{\text{H}}\text{H}\beta}$ scalar and $^1D_{\text{C}\beta\text{H}\beta}$ residual dipolar couplings (RDCs) in unfolded proteins. For urea-denatured ubiquitin and protein G, up to six 3J -couplings to $^1\text{H}\beta$ are detected, which define the χ_1 angle at very high precision. Interpretation of the 3J couplings by a model of mixed staggered χ_1 rotamers yields excellent agreement and also provides stereoassignments for $^1\text{H}\beta$ methylene protons. For all observed amino acids with the exception of leucine, the chemical shift of $^1\text{H}\beta^3$ protons was found downfield from $^1\text{H}\beta^2$. For most residues, the precision of individual χ_1 rotamer populations is better than 2%. The experimental χ_1 rotamer populations are in the vicinity of averages obtained from coil regions in folded protein structures. However, individual variations from these averages of up to 40% are highly significant and indicate sequence- and residue-specific interactions. Particularly strong deviations from the coil average are found for serine and threonine residues, an effect that may be explained by a weakening of side-chain to backbone hydrogen bonds in the urea-denatured state. The measured $^1D_{\text{C}\beta\text{H}\beta}$ RDCs correlate well with predicted RDCs that were calculated from a sterically aligned coil model ensemble and the 3J -derived χ_1 rotamer populations. This agreement supports the coil model as a good first approximation of the unfolded state. Deviations between measured and predicted values at certain sequence locations indicate that the description of the local backbone conformations can be improved by incorporation of the RDC information. The ease of detection of a large number of highly precise side-chain RDCs opens the possibility for a more rigorous characterization of both side-chain and backbone conformations in unfolded proteins.

Introduction

A detailed, quantitative description of the unfolded state of proteins is crucial for understanding protein folding,¹ protein misfolding and aggregation in amyloidogenic diseases such as Alzheimer's and Parkinson's,² and function of intrinsically disordered proteins.^{3,4} Such a description is both experimentally and theoretically highly challenging, because only a limited number of measurable parameters are available to describe the vast space of possible unstructured conformations.

During recent years, the conformations of the backbone of unfolded proteins have been described in some detail using paramagnetic relaxation enhancements PREs^{5,6} and RDCs.^{7,8}

These two parameters report on well-defined ensemble averages of the long- and short-range backbone geometry and are thus more amenable to a rigorous quantitative interpretation than chemical shifts, NOE, or relaxation data. Both PREs and RDCs have revealed long-range contacts and residual structure in many denatured proteins showing that such states contain structural bias that may drive them toward a folded structure. The correct prediction of such structural propensities of denatured states from the amino acid sequence may be an important step toward solving the protein folding problem.

RDCs offer particular advantages for the characterization of unfolded states, because they do not require additional chemical labeling and can be detected with ease for a large number of internuclear vectors; for example, a recent study showed that up to seven RDCs per peptide unit could be determined for urea-unfolded ubiquitin.⁹ For a number of unfolded proteins, trends of backbone RDCs along the polypeptide sequence could be reproduced in structural ensembles created according to the

[†] University of Basel.

[‡] Institut de Biologie Structurale Jean-Pierre Ebel.

(1) Shortle, D. *FASEB J.* **1996**, *10*, 27–34.

(2) Dobson, C. M. *Nature* **2003**, *426*, 884–890.

(3) Dunker, A. K.; Silman, I.; Uversky, V. N.; Sussman, J. L. *Curr. Opin. Struct. Biol.* **2008**, *18*, 756–764.

(4) Wright, P. E.; Dyson, H. J. *Curr. Opin. Struct. Biol.* **2009**, *19*, 31–38.

(5) Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997**, *268*, 170–184.

(6) Mittag, T.; Forman-Kay, J. *Curr. Opin. Struct. Biol.* **2007**, *17*, 3–14.

(7) Shortle, D.; Ackerman, M. *Science* **2001**, *293*, 487–489.

(8) Meier, S.; Blackledge, M.; Grzesiek, S. *J. Chem. Phys.* **2008**, *128*, 052204.

(9) Meier, S.; Grzesiek, S.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 9799–9807.

amino-acid-specific phi/psi angle propensities in non-alpha, non-beta conformations of PDB structures (PDB coil libraries).^{8,10,11} This indicates that the so-called coil model^{12,13} is a good, first approximation of the unfolded state ensemble. In turn, deviations from the coil model point to residual order within the unfolded state. Such deviations have revealed highly populated turn conformations in the natively unfolded Tau protein¹⁴ and have shown that urea binding drives the backbone to more extended conformations for ubiquitin.⁹ Additional long-range RDCs between amide protons have given evidence for a remaining, significant (10–20%) population of the first β -hairpin (residues 1–18) in (8 M) urea-denatured ubiquitin.¹⁵

In contrast to the backbone of unstructured polypeptides, experimental information on side chains is rather sparse. Such investigations are severely hampered by the poor dispersion of side-chain signals resulting from the conformational averaging. A small number of $^3J_{\text{H}\alpha\text{H}\beta}$ couplings have been determined in shorter unfolded peptides without ^{13}C labeling,^{16,17} which correlated with predictions from χ_1 coil distributions. However, no stereoassignments of methylene $\text{H}^{\beta 2}$ and $\text{H}^{\beta 3}$ resonances were obtained. A more advanced study¹⁸ determined heteronuclear $^3J_{\text{NC}\gamma}$ and $^3J_{\text{C}\gamma}$ couplings in urea-denatured lysozyme. Assuming staggered χ_1 rotamers, estimates for their populations were derived for approximately 50 amino acids, which also showed correlations to coil model predictions for most amino acids with the exception of aromatics. Precision in this analysis was limited by the lack of precise Karplus coefficients for $^3J_{\text{NC}\gamma}$ and $^3J_{\text{C}\gamma}$ and the fact that the three populations $p_{-60^\circ, 60^\circ, 180^\circ}$ (two independent parameters because $p_{-60^\circ} + p_{60^\circ} + p_{180^\circ} = 1$) were determined from only two experimental values.

In the present study, we have improved the description of χ_1 conformations in unfolded proteins by optimized heteronuclear experiments involving β -protons, which are able to resolve most methylene $\text{H}^{\beta 2}$ and $\text{H}^{\beta 3}$ pairs. Stereoassignments and χ_1 angle information could be obtained for the predominant part of residues in urea-denatured ubiquitin and protein G from an extensive set of up to six three-bond scalar couplings ($^3J_{\text{NH}\beta 2,3}$, $^3J_{\text{CH}\beta 2,3}$, and $^3J_{\text{H}\alpha\text{H}\beta 2,3}$). A combined analysis of all 3J couplings according to the staggered conformer model yields individual populations with a maximal error of 2%. This analysis is corroborated by independent $^1D_{\text{C}\beta\text{H}\beta 2,3}$ RDC data detected in strained polyacrylamide gels.^{19,20} These side-chain RDCs agree well with theoretical RDCs calculated from the 3J -derived χ_1 conformer distribution and a coil model ensemble of backbone conformations generated by the program *Flexible-Meccano*.¹¹ The obtained χ_1 conformer populations cluster around coil model

averages, but individual variations in particular for serines and threonines of up to 40% are significant and indicate sequence- and residue-specific preferences.

Materials and Methods

Sample Preparation and NMR Spectroscopy. $^{15}\text{N}/^{13}\text{C}$ -labeled human ubiquitin and protein G (GB1 sequence $^1\text{M}^1\text{QYKLLINGK}^1\text{TLKGETTTEA}^2\text{VDAATAEKVF}^3\text{KQYANDNGVD}^4\text{GEW-TYDDATK}^5\text{TFTVTE}$) were prepared according to standard protocols.²¹ Ubiquitin NMR samples contained 1.0 (0.6) mM $^{15}\text{N}/^{13}\text{C}$ -labeled protein in 10 mM glycine, 8 M urea, pH 2.5, 95/5% $\text{H}_2\text{O}/\text{D}_2\text{O}$ for measurement under isotropic (anisotropic) conditions. Protein G samples contained 0.6 mM $^{15}\text{N}/^{13}\text{C}$ -labeled protein in 10 mM glycine, 7.4 M urea, pH 2.0, 95/5% $\text{H}_2\text{O}/\text{D}_2\text{O}$. Residual alignment of urea-denatured proteins was achieved by introducing the protein solutions into 7% (w/v) polyacrylamide gels and horizontal compression (aspect ratio 2.9:1) in NEW-ERA sample tubes²² yielding maximal $^1D_{\text{NH}}$ RDCs of about 13 Hz for both proteins.

All NMR experiments were carried out at 298 K on a Bruker Avance DRX 800 spectrometer equipped with a TCI cryoprobe. Spectra were processed with NMRPipe²³ and evaluated with NMRView²⁴ and PIPP.²⁵

Assignments. Assignments of urea-denatured ubiquitin (BMRB entry 4375)²⁶ and protein G²⁷ were transferred to our sample preparations and extended by a combination of CBCA(CO)NH,²⁸ HNC0,²⁹ HBHA(CO)NH,²⁸ and HNHB³⁰ experiments. To obtain higher resolution, the constant time ^{15}N acquisition period was increased in these experiments to about 40 ms. Note that for CBCA(CO)NH, HNC0, and HBHA(CO)NH, this still achieves a transfer of about 95% via the $^1J_{\text{NC}'}$ (~15 Hz) coupling. Almost complete assignments of all $^1\text{H}^{\text{N}}$, ^{15}N , $^{13}\text{C}'$, $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, $^1\text{H}^\alpha$, and $^1\text{H}^\beta$ resonances were obtained from this procedure. Missing assignments mainly comprise amino acids preceding proline or were due to signal degeneracy of some geminal protons. The obtained chemical shifts are close to the published data with the exception of residues in the vicinity of the mutated T2Q site in protein G. They also extend the previous data by the stereoassignments of β -methylene protons and the $^{13}\text{C}'$ chemical shifts (protein G). The assignments are deposited in the BMRB data bank under accession numbers 16626 (ubiquitin) and 16627 (protein G).

Determination of Scalar and Residual Dipolar Coupling Constants. 3J scalar couplings carrying information on the χ_1 angle of the denatured proteins were obtained from modified versions of quantitative $^3J_{\text{NH}\beta}$ -HNHB,³⁰ $^3J_{\text{CH}\beta}$ -HN(CO)HB,³¹ and $^3J_{\text{H}\alpha\text{H}\beta}$ -HAHB-(CACO)NH³² experiments. $^1D_{\text{C}\beta\text{H}\beta}$ RDCs for χ_1 angle information were determined from an HBHA(CO)NH²⁸ experiment, where an IPAP detection scheme³³ was introduced into the mixed constant

- (10) Jha, A.; Colubri, A.; Freed, K.; Sosnick, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099–13104.
- (11) Bernado, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17002–17007.
- (12) Serrano, L. *J. Mol. Biol.* **1995**, *254*, 322–333.
- (13) Smith, L.; Bolin, K.; Schwalbe, H.; MacArthur, M.; Thornton, J.; Dobson, C. *J. Mol. Biol.* **1996**, *255*, 494–506.
- (14) Mukrasch, M.; Markwick, P.; Biernat, J.; Bergen, M.; Bernado, P.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 5235–5243.
- (15) Meier, S.; Strohmeier, M.; Blackledge, M.; Grzesiek, S. *J. Am. Chem. Soc.* **2007**, *129*, 754–755.
- (16) West, N. J.; Smith, L. *J. Mol. Biol.* **1998**, *280*, 867–877.
- (17) Mathieson, S. I.; Penkett, C. J.; Smith, L. *J. Pacific Symp. Biocomputing* **1999**, 542–553.
- (18) Hennig, M.; Bermel, W.; Spencer, A.; Dobson, C. M.; Smith, L. J.; Schwalbe, H. *J. Mol. Biol.* **1999**, *288*, 705–723.
- (19) Sass, H.; Musco, G.; Stahl, S.; Wingfield, P.; Grzesiek, S. *J. Biomol. NMR* **2000**, *18*, 303–309.
- (20) Tycko, R.; Blanco, F.; Ishii, Y. *J. Am. Chem. Soc.* **2000**, *122*, 9340–9341.

- (21) Sass, J.; Cordier, F.; Hoffmann, A.; Cousin, A.; Omichinski, J.; Lowen, H.; Grzesiek, S. *J. Am. Chem. Soc.* **1999**, *121*, 2047–2055.
- (22) Chou, J.; Gaemers, S.; Howder, B.; Louis, J.; Bax, A. *J. Biomol. NMR* **2001**, *21*, 377–382.
- (23) Delaglio, F.; Grzesiek, S.; Vuister, G.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277–293.
- (24) Johnson, B.; Blevins, R. *J. Biomol. NMR* **1994**, *4*, 603–614.
- (25) Garrett, D.; Powers, R.; Gronenborn, A.; Clore, G. *J. Magn. Reson.* **1991**, *95*, 214–220.
- (26) Peti, W.; Smith, L.; Redfield, C.; Schwalbe, H. *J. Biomol. NMR* **2001**, *19*, 153–165.
- (27) Frank, M. K.; Clore, G. M.; Gronenborn, A. M. *Protein Sci.* **1995**, *4*, 2605–2615.
- (28) Grzesiek, S.; Bax, A. *J. Biomol. NMR* **1993**, *3*, 185–204.
- (29) Grzesiek, S.; Bax, A. *J. Magn. Reson.* **1992**, *96*, 432–440.
- (30) Archer, S.; Ikura, M.; Torchia, D.; Bax, A. *J. Magn. Reson.* **1991**, *95*, 636–641.
- (31) Grzesiek, S.; Ikura, M.; Clore, G.; Gronenborn, A.; Bax, A. *J. Magn. Reson.* **1992**, *96*, 215–221.
- (32) Lohr, F.; Schmidt, J.; Ruterjans, H. *J. Am. Chem. Soc.* **1999**, *121*, 11821–11826.
- (33) Ottiger, M.; Delaglio, F.; Bax, A. *J. Magn. Reson.* **1998**, *131*, 373–378.

time $^1\text{H}^{\alpha\beta}$ evolution/ $^1\text{H}^{\alpha\beta} \rightarrow ^{13}\text{C}^{\alpha\beta}$ transfer period. $^1D_{\text{C}\beta\text{H}\beta}$ RDCs were calculated as the difference in couplings observed under anisotropic conditions in strained polyacrylamide gels and a second experiment under isotropic conditions. Details of these experiments are given in Supporting Information Figures S1–S4. Similar to the assignment experiments, acquisition times in the indirect ^{15}N and $^1\text{H}^{\alpha\beta}$ dimensions were set to 40 and 30 ms, respectively, to obtain sufficient resolution. Each experiment for scalar and dipolar couplings was carried out twice, and the reported coupling constants and error estimates refer to mean and standard deviations from such repeated experiments. The J -coupling constants were not corrected for effects of scalar relaxation of the second kind,³⁴ because these are expected to be small due to the fast effective correlation times in unfolded proteins.

Analysis of 3J Coupling Constants and RDCs. Analysis of the 3J data according to a model of staggered χ_1 rotamers (see main text) was carried out using in-house written Matlab (MathWorks, Inc.) routines and its QUADPROG function for constrained linear minimization. For the analysis of RDC data, theoretical RDCs from steric alignment were calculated for the three staggered rotamers by an in-house written C program³⁵ from an ensemble of 50 000 unfolded protein structures generated by the *Flexible-Meccano* program.¹¹ Experimental RDCs were then compared and fitted to these predicted values by the same staggered rotamer model (see main text).

Coil Library Rotamers. The experimentally derived χ_1 rotamer populations were also compared to average populations from a protein coil library. This library was downloaded as version 20080310_pc20_res1.6_R0.25 generated on 3/10/2008 from the Rose lab server³⁶ and contained 16 856 protein fragments from nonhomologous proteins of X-ray structures with resolution better than 1.6 Å in non-alpha, non-beta conformations.

Results and Discussion

χ_1 Torsion Angle Information from Scalar Couplings and RDCs. The NMR analysis of unfolded proteins in solution is made difficult by the low spectral dispersion resulting from conformational averaging. In particular, this applies to side-chain resonances. Thus, so far, the potential of side-chain protons has not fully been used as a source of conformational information. Here, we have probed the χ_1 torsion angles in urea-denatured ubiquitin and protein G by 3J scalar and residual dipolar couplings involving H^β protons. Scalar couplings were detected by modified versions of the quantitative- $^3J_{\text{NH}\beta}$ HNHB,³⁰ $^3J_{\text{C}\beta\text{H}\beta}$ HN(CO)HB,³¹ and $^3J_{\text{H}\alpha\text{H}\beta}$ HAHB(CACO)NH³² experiments (Supporting Information). These experiments overcome the problems of low dispersion in both backbone and side-chain resonances by making use of the long transverse relaxation times in the unfolded state to achieve maximal frequency resolution. Acquisition times of 30–35 ms for $^1\text{H}^\alpha$ and $^1\text{H}^\beta$ resonances and of 40 ms for ^{15}N in the indirect dimensions proved to be sufficient to resolve most of the overlap in the side-chain experiments. Additional information from similarly optimized 3D CBCA(CO)NH, HBHA(CO)NH experiments was used to establish sequential assignments. Figure 1 shows the good resolution of the side-chain $^1\text{H}^\beta$ resonances in the quantitative- J spectra for residues L7 and K10 of urea-denatured protein G. The intensity ratios of the resonances in these spectra yield information on the χ_1 torsion angles and the stereospecific assignments for β -methylene protons. The high population of -60° χ_1 conformations is obvious from the higher intensities

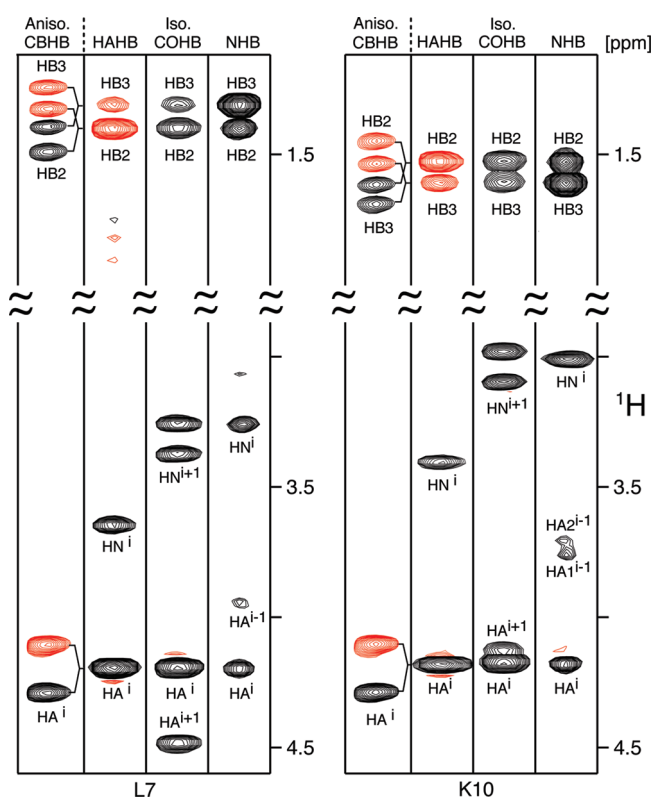


Figure 1. Strip plots of 3D spectra used for determination of $^3J_{\text{H}\alpha\text{H}\beta}$, $^3J_{\text{NH}\beta}$, $^3J_{\text{C}\beta\text{H}\beta}$ coupling constants and $^1D_{\text{C}\beta\text{H}\beta}$ RDCs in urea-denatured protein G. Data are shown for two typical residues L7 (left) and K10 (right) from the quantitative $^3J_{\text{H}\alpha\text{H}\beta}$ -HAHB(CACO)NH (HAHB), $^3J_{\text{C}\beta\text{H}\beta}$ -HN(CO)HB (COHB), and $^3J_{\text{NH}\beta}$ -HNHB (NHB) experiments recorded under isotropic conditions as well as from the $^1D_{\text{C}\beta\text{H}\beta}$ -IPAP-HBHA(CO)NH (CBHB) experiment recorded under anisotropic conditions. Resonances are labeled with assignment information. Amide proton signals are folded in the indirect proton dimension to reduce experimental time. These signals are split by the $^1J_{\text{NH}}$ coupling for the HN(CO)HB experiment. For the quantitative- J HAHB, COHB, and NHB experiments, resonances shown in red are negative signals. For the IPAP-HBHA(CO)NH experiment, red is used to distinguish the $^1\text{H}^{\alpha\beta}$ upfield from the downfield (black) components.

of the $^1\text{H}^{\beta 2}$ resonances relative to the $^1\text{H}^{\beta 3}$ resonances in the quantitative- $^3J_{\text{H}\alpha\text{H}\beta}$ HAHB(CACO)NH experiment, which corresponds to $^3J_{\text{H}\alpha\text{H}\beta 2} > ^3J_{\text{H}\alpha\text{H}\beta 3}$, from the inverse situation in the quantitative- $^3J_{\text{NH}\beta}$ HNHB experiment, and from more equal intensities in the quantitative- $^3J_{\text{C}\beta\text{H}\beta}$ HN(CO)HB experiment.

In addition to 3J couplings, also RDCs yield information on the χ_1 conformations. The detection of $^1D_{\text{C}\beta\text{H}\beta}$ RDCs induced in strained polyacrylamide gels proved particularly easy from an IPAP-HBHA(CO)NH experiment (Supporting Information). Figure 1 shows as an example the data of this experiment on the oriented urea-denatured protein G. The quality of the spectra is excellent and allowed the unambiguous determination of 91 (61) $^1D_{\text{C}\beta\text{H}\beta}$ RDCs in unfolded ubiquitin (protein G).

In total for ubiquitin (protein G), 353 (246) $^3J_{\text{NH}\beta}$, $^3J_{\text{C}\beta\text{H}\beta}$, $^3J_{\text{H}\alpha\text{H}\beta}$, and $^1D_{\text{C}\beta\text{H}\beta}$ couplings (Supporting Information Table S1) could be derived from the quantitative analysis of the spectra. These data cover 82% of all side chains with variable χ_1 angle, that is, 55 out of 68 (ubiquitin) and 40 out of 46 (protein G) non-(Gly, Ala) residues.

Analysis of 3J Coupling Constants by Staggered χ_1 Rotamer Populations. Our analysis of the 3J couplings in terms of side-chain conformations assumes as a first approximation that the conformations are a population mixture of three staggered χ_1 (-60° , $+60^\circ$, 180°) rotamers. Previously deter-

(34) Bax, A.; Vuister, G. W.; Grzesiek, S.; Delaglio, F.; Wang, A. C.; Tschudin, R.; Zhu, G. *Methods Enzymol.* **1994**, *239*, 79–105.

(35) Huang, J.-r.; Grzesiek, S. *J. Am. Chem. Soc.* **2010**, *132*, 694–705.

(36) Fitzkee, N.; Fleming, P.; Rose, G. *Proteins* **2005**, *58*, 852–854.

mined amino-acid-specific Karplus coefficients³⁷ were used to predict theoretical ${}^3J_{\text{H}\alpha\text{H}\beta}$, ${}^3J_{\text{NH}\beta}$, and ${}^3J_{\text{C}'\text{H}\beta}$ coupling constants for these rotamers according to the Karplus relation:

$${}^3J_{ij}^{\text{calc}}(\chi_1) = C_{ij}^0 + C_{ij}^1 \cos(\vartheta_{ij}(\chi_1)) + C_{ij}^2 \cos(2\vartheta_{ij}(\chi_1)) \quad (1)$$

where $\vartheta_{ij}(\chi_1)$ is the intervening dihedral angle between the nuclei i and j in a ${}^3J_{ij}$ coupling for a specific side-chain torsion angle χ_1 . For all side chains, only single sets of resonances were observed. Thus, the side chains are in fast exchange on the time scale of the chemical shift, that is, faster than milliseconds. Accordingly, the observed coupling constants should be population averages over the individual rotamers:^{38–40}

$$\langle {}^3J_{ij}^{\text{calc}} \rangle = \sum_{\chi_1=-60^\circ, 60^\circ, 180^\circ} p_{\chi_1} {}^3J_{ij}^{\text{calc}}(\chi_1) \quad (2)$$

where $p_{-60^\circ, 60^\circ, 180^\circ}$ are the individual populations.

To derive these populations from the experimental ${}^3J^{\text{exp}}$ couplings, their deviation from the calculated average $\langle {}^3J^{\text{calc}} \rangle$ was minimized with respect to p_{-60° , p_{60° , and p_{180° by a constrained linear least-squares fit of the target function

$$\chi^2 = \frac{1}{N} \sum_{ij} \left(\frac{{}^3J_{ij}^{\text{exp}} - \langle {}^3J_{ij}^{\text{calc}} \rangle}{\sigma_{ij}} \right)^2 \quad (3)$$

under the conditions $p_{-60^\circ} + p_{60^\circ} + p_{180^\circ} = 1$ and $0 \leq p_{-60^\circ, 60^\circ, 180^\circ} \leq 1$. In eq 3, σ presents the statistical, experimental error of the coupling constant obtained from a repetition of the experiment, the summation runs over all individual nuclei i and j , for which a 3J coupling could be determined for an individual side chain, and N indicates the total number of measured 3J values.

The stereospecific assignments of geminal $\text{H}^{\beta 2}$ and $\text{H}^{\beta 3}$ protons in urea-denatured ubiquitin and protein G were not known prior to the current analysis. This information was also derived from the 3J couplings by carrying out the fit procedure for both possible stereo assignments and using the assignment that corresponded to the lower χ^2 value. Typically these χ^2 values were about 10–20 times smaller than the values for the swapped assignment, such that discrimination was achieved easily. The stereospecific assignments obtained by this method comprise 82% of all β -methylene protons, corresponding to 41 and 25 residues in ubiquitin and protein G, respectively.

Figure 2 shows the experimental ${}^3J_{\text{NH}\beta}$, ${}^3J_{\text{C}'\text{H}\beta}$, and ${}^3J_{\text{H}\alpha\text{H}\beta}$ coupling constants and their values according to the fit of eq 3 for a number of residues in ubiquitin and protein G (the complete data for both proteins are shown in Supporting Information Figures S5 and S6). For most residues, the agreement between experimental and predicted data is excellent with average RMSDs between measured and predicted 3J -values of less than 0.3 Hz. This indicates not only a high precision of both experimental 3J data and Karplus coefficients, but also validates the staggered rotamer model as a reasonable approximation for the side-chain conformations. In total, 53 (39) χ_1 rotamer populations could be derived for those amino acids in ubiquitin (protein G) for which at least two 3J coupling constants had

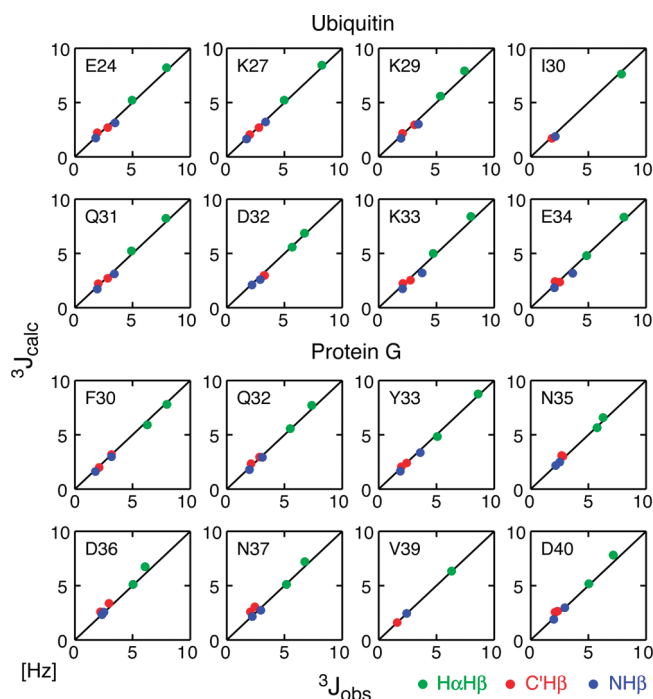


Figure 2. Comparison of experimental (obs) ${}^3J_{\text{XH}\beta}$ ($X = \text{C}', \text{N}, \text{HA}$) constants to values derived (calc) according to the χ_1 rotamer population fit of eq 3 for selected residues of urea-denatured ubiquitin (top) and protein G (bottom).

been measured (Supporting Information Table S2). The constrained linear fit also provides error estimates for the p_{-60° , p_{60° , and p_{180° populations derived by propagation from the statistical experimental error. These errors range between 0.01 and 0.02, indicating a very high precision of the population estimates.

Analysis of RDC Data. To make use of the experimental ${}^1D_{\text{C}\beta\text{H}\beta}$ RDCs for the analysis of χ_1 conformations, theoretical RDC estimates for all staggered rotamers were obtained from large simulated ensembles of unfolded ubiquitin or protein G structures. For both proteins, ensembles of 50 000 unfolded structures were generated by the *Flexible-Meccano* program¹¹ according to the amino-acid-specific phi/psi angle propensities in non-alpha, non-beta conformations of PDB structures (PDB coil library) and omitting structures with sterical clashes. Such ensembles have previously been shown to reproduce the trends of backbone RDCs along the polypeptide sequence.^{8,10,11} Because the *Flexible-Meccano* algorithm represents side chains only by a pseudoatom at the C^β position, full coordinates for the C^β and H^β atoms for all staggered rotamers were generated from the N, C^α , and C' positions using idealized tetrahedral geometry. The alignment tensor for each member k of the ensemble was then calculated on the basis of the assumption of steric exclusion^{41,42} using an efficient in-house written algorithm.³⁵ In brief, this algorithm calculates the maximal extension of the molecule for each direction of the unit sphere. The probability for finding the molecule in a certain orientation is then derived as the volume that can be occupied by the molecule between two infinitely extended, parallel planes relative to the total volume between the planes. The alignment tensor then corresponds to the average over all orientations of

(37) Pérez, C.; Löhr, F.; Rüterjans, H.; Schmidt, J. M. *J. Am. Chem. Soc.* **2001**, *123*, 7081–7093.

(38) Pachler, A. *Spectrochim. Acta* **1963**, *19*, 2085–2092.

(39) Pachler, A. *Spectrochim. Acta* **1964**, *20*, 581–587.

(40) Dzakuła, Z.; Westler, W.; Edison, A.; Markley, J. L. *J. Am. Chem. Soc.* **1992**, *114*, 6195–6199.

(41) Zweckstetter, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 3791–3792.

(42) van Lune, F.; Manning, L.; Dijkstra, K.; Berendsen, H. J.; Scheek, R. M. *J. Biomol. NMR* **2002**, *23*, 169–179.

second rank spherical harmonics weighted by this probability. Theoretical RDC values were derived for each χ_1 rotamer and in each individual structure as

$$D_{k,ij}^{\text{calc}}(\chi_1) = -\frac{\gamma_i \gamma_j \hbar \mu_0}{4\pi^2} \sqrt{\frac{4\pi}{5}} \sum_{m=-2}^2 S_{k,m}^* \frac{Y_{2m}(\Theta_{k,ij}(\chi_1), \Phi_{k,ij}(\chi_1))}{r_{k,ij}^3(\chi_1)} \quad (4)$$

where $D_{k,ij}^{\text{calc}}$ represents the RDC between nuclei i and j for ensemble member k with individual alignment tensor $S_{k,m}$ (written in irreducible form⁴³), Y_{2m} are spherical harmonics, $r_{k,ij}$, $\Theta_{k,ij}$, $\Phi_{k,ij}$ are the polar coordinates of the internuclear vector, and $\gamma_{i,j}$ are the nuclear gyromagnetic ratios. These RDC values for the individual structures were then averaged over all N members of the ensemble to obtain an estimate D_{ij}^{calc} for the RDC in the unfolded protein:

$$D_{ij}^{\text{calc}}(\chi_1) = \frac{1}{N} \sum_{k=1}^N D_{k,ij}^{\text{calc}}(\chi_1) \quad (5)$$

Because the absolute size of the alignment tensor $S_{k,m}$ is difficult to predict from the experimental conditions, an additional common overall scaling was used such that the mean square difference between measured and predicted average backbone ${}^1D_{\text{HN}}$ was minimized. Analogously to eq 2, the population average over the individual χ_1 rotamers was then calculated as

$$\langle D_{ij}^{\text{calc}} \rangle = \sum_{\chi_1=-60^\circ, 60^\circ, 180^\circ} p_{\chi_1} D_{ij}^{\text{calc}}(\chi_1) \quad (6)$$

Figure 3 shows the experimental ${}^1D_{\text{C}\beta\text{H}\beta}$ RDCs and their values predicted by eq 6 from the 3J -derived χ_1 rotamer populations (red ●) for the same residues in ubiquitin and protein G as in Figure 2. For comparison, measured and predicted 3J -couplings (blue ●) are also shown. The complete data for both proteins are given in Supporting Information Figures S7 and S8. For many residues, such as E24, K29, I30, Q31, and E34 in ubiquitin and F30, Q32, N35, N37, V39, and D40 in protein G, measured and predicted RDCs agree within about 5 Hz. This agreement is very reasonable when compared to the full variation of about 20–30 Hz of observed RDCs. For other residues like D32 in ubiquitin or D36 in protein G, the deviations from the predictions are clearly larger, but still agree with the trends of the predictions. The correlations between all measured and predicted ${}^1D_{\text{C}\beta\text{H}\beta}$ RDCs have a Pearson's correlation coefficient of 0.86 and 0.70 for ubiquitin and protein G, respectively (Figure 4, ●). The correlation can be improved to some extent, when the ${}^1D_{\text{C}\beta\text{H}\beta}$ RDCs are also included into the fit of the χ_1 rotamer populations by extending the χ^2 function of eq 3 to the differences between measured and predicted D_{ij} couplings (Figures 3 and 4, ○). In this case, the χ_1 rotamer populations only change by a few percent (not shown), but the correlation coefficient increases to 0.92 (0.76) for ubiquitin and protein G, respectively.

Considering the crudeness of the assumptions for the coil model and the steric alignment of the unfolded model ensemble, the agreement between measured and predicted ${}^1D_{\text{C}\beta\text{H}\beta}$ RDCs is surprisingly good. This provides an independent confirmation for both the backbone coil model as implemented by the *Flexible-Meccano* algorithm¹¹ as well as the staggered χ_1

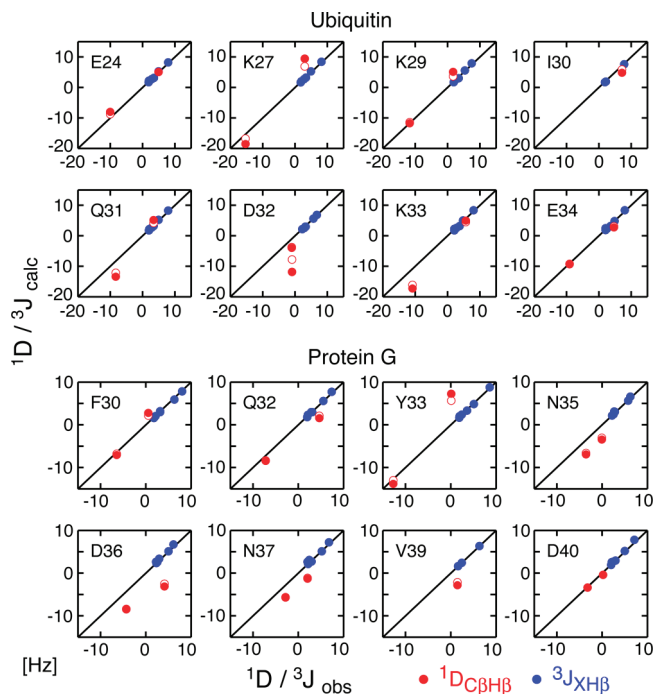


Figure 3. Comparison of experimental (obs) and predicted (calc) ${}^1D_{\text{C}\beta\text{H}\beta}$ RDCs (red ●) based on the χ_1 rotamer populations derived from the fit of eq 3 of ${}^3J_{\text{XH}\beta}$ constants and the coil model ensemble. For comparison, experimental and predicted ${}^3J_{\text{XH}\beta}$ couplings are also indicated (blue ●). The same residues are shown as in Figure 2 for urea-denatured ubiquitin (top) and protein G (bottom). Red ○ indicate predictions for ${}^1D_{\text{C}\beta\text{H}\beta}$ RDCs, when these RDC data were also included in the rotamer population fit of eq 3 (see text).

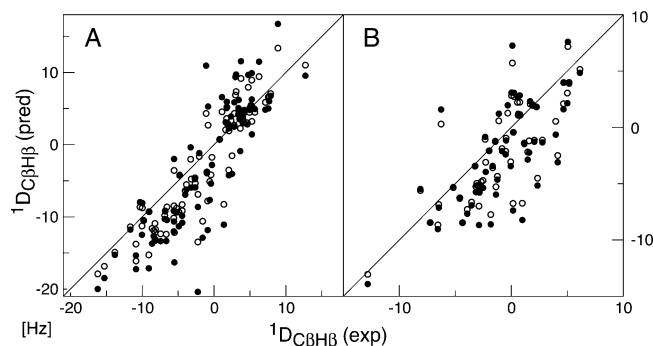


Figure 4. Correlation of all observed experimental (obs) and predicted (calc) ${}^1D_{\text{C}\beta\text{H}\beta}$ RDCs (●) based on the χ_1 rotamer populations derived from the fit of eq 3 of ${}^3J_{\text{XH}\beta}$ constants and the coil model ensemble for urea-denatured ubiquitin (A) and protein G (B). The Pearson's correlation coefficient is 0.86 and 0.70 for ubiquitin and protein G, respectively. The "○" indicate predictions for ${}^1D_{\text{C}\beta\text{H}\beta}$ RDCs, when these RDC data were also included in the rotamer population fit of eq 3. In this case, the correlation coefficient increases to 0.92 and 0.76 for ubiquitin and protein G, respectively.

rotamer model. However, the deviations, which exceed the experimental errors, also clearly indicate shortcomings of this interpretation. Because the agreement of the staggered rotamer predictions is almost perfect for the 3J -couplings, it is very likely that the RDC deviations result from inaccuracies of the local backbone geometry predicted by the coil ensemble and from the unknown microscopic details of the alignment interaction, which may not be adequately covered by the simple steric alignment model. In principle, the inaccuracies of the backbone geometry may be reduced by refining the backbone conformations using information from additional backbone RDCs. Such

(43) Moltke, S.; Grzesiek, S. *J. Biomol. NMR* **1999**, *15*, 77–82.

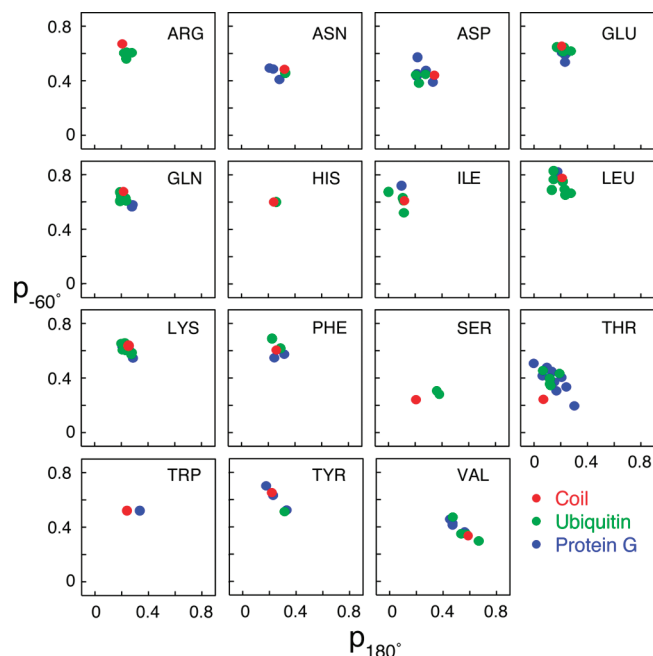


Figure 5. p_{-60° and p_{180° χ_1 rotamer populations obtained from the combined fit of ${}^3J_{\text{NH}\beta}$, ${}^3J_{\text{C}\text{H}\beta}$, and ${}^3J_{\text{H}\alpha\text{H}\beta}$ couplings of eq 3. Populations are shown separately for ubiquitin (green), protein G (blue), and the average of the PDB coil structures (red).

an approach is currently pursued by constrained molecular dynamics ensemble calculations, which include all of the RDCs as restraints.³⁵ The ease of detection of a large number of highly precise RDCs on side-chain nuclei may thus make it possible to increase the accuracy of the description of side-chain and backbone conformations in unfolded ensembles beyond the simple coil model.

Comparison to Coil χ_1 Populations. Previous analyses of the χ_1 rotamer conformations in unfolded proteins using more limited 3J data^{16–18} have concluded that their populations correlate to the average rotamer populations in the PDB coil conformations. Figure 5 shows the p_{-60° and p_{180° χ_1 rotamer populations according to the combined fit of ${}^3J_{\text{NH}\beta}$, ${}^3J_{\text{C}\text{H}\beta}$, and ${}^3J_{\text{H}\alpha\text{H}\beta}$ couplings (Figure 2) together with the respective coil averages for all amino acids in ubiquitin and protein G, for which populations could be derived. The populations are indeed in the vicinity of the coil values. However, in some cases, they deviate by more than 30% from the coil values (Figure 6) and also vary around their mean by approximately 10%. These variations are significant considering that the errors of the individual populations are only about 1–2% as estimated from error propagation of the linear least-squares fit. Thus, they must reflect sequence-specific preferences of the side chains along the unfolded polypeptide chain (see below).

As remarked earlier,¹⁸ many amino acids prefer the -60° χ_1 rotamer both in the coil model predictions and in 3J -derived populations due to the repulsion of substituents at the γ^1 position from the main chain. In our analysis (Figure 5), this is the case for all unbranched amino acids (R, N, D, E, Q, H, L, K, F, W, Y) as well as for the branched isoleucine. These residues have J -derived and coil average p_{-60° values of 40–80%, where the shortest side-chain asparagine and aspartic acid residues have the lowest and leucine the highest -60° preference. Valines have similar populations for the -60° (30–50%) and $+180^\circ$

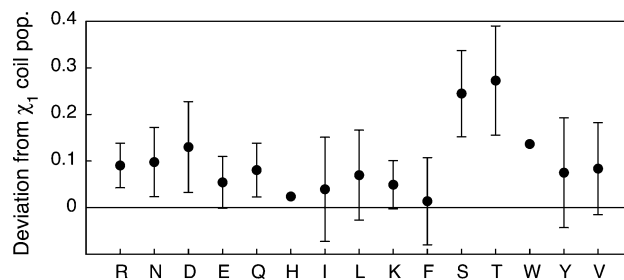


Figure 6. Deviations of the experimentally (3J -only) derived χ_1 rotamer populations in urea-denatured proteins from the average of the PDB coil structures. The deviations are calculated as $|\bar{p}_{\text{exp}} - \bar{p}_{\text{coil}}|$ with $\bar{p} = (p_{-60^\circ, 60^\circ, 180^\circ})$. Average and standard deviations of $|\bar{p}_{\text{exp}} - \bar{p}_{\text{coil}}|$ are shown for all amino acids for which rotamer populations could be obtained in urea-denatured ubiquitin and protein G.

(40–70%) χ_1 rotamers, which on average places both γ -methyl groups at the furthest distance from the backbone carbonyl.

Exceptions to this behavior are found for serines and threonines. For these residues, the coil model predicts p_{-60° populations of only 24% (S, T) and much larger p_{60° populations of 55% (S) and 68% (T). It has been speculated¹⁸ that this is caused by favorable polar interactions between the side-chain hydroxyl and the main-chain amide group in the $+60^\circ$ χ_1 rotamer. Among all amino acid types, the experimentally derived rotamer populations for serines and threonines in urea-denatured ubiquitin and protein G show the strongest deviations from these coil predictions of folded protein structures. Thus, the J -derived p_{60° populations amount to only about 35% for serine and 40–50% for threonine, whereas p_{-60° and p_{+180° are correspondingly higher (Figure 5). The deviations are found for all serine and threonine residues, and no particular correlation to specific locations in the sequence is evident (Supporting Information Figure S9).

It is unlikely that this behavior is an experimental artifact connected to the particular ${}^{13}\text{C}^\beta$ chemical shift of these two amino acids, because the ${}^{13}\text{C}^\beta$ nuclei are not involved in the magnetization pathways of the quantitative ${}^3J_{\text{NH}\beta}$ -HNHB, ${}^3J_{\text{C}\text{H}\beta}$ -HN(CO)HB, and ${}^3J_{\text{H}\alpha\text{H}\beta}$ -HAHB(CACO)NH experiments. To further test for systematic errors from the ${}^3J_{\text{H}\alpha\text{H}\beta}$ -HAHB(CACO)NH experiment, we have also fitted the serine and threonine χ_1 populations by using only the ${}^3J_{\text{NH}\beta}$ - and ${}^3J_{\text{C}\text{H}\beta}$ -couplings (not shown). The deviations of the resulting populations from the all- 3J -value results are in most cases smaller than 4%, that is, much smaller than the deviations from the coil values. It should be noted that the fit uses specific Karplus coefficients for serines and threonines.³⁷ Thus, particular effects of the side-chain oxygen on the size of the J -couplings are corrected. Deviations from the χ_1 coil populations for serines and threonines are also evident from the experimental RDCs, since their agreement with predictions from J -derived χ_1 populations is considerably better than with predictions from the coil populations (not shown).

We attribute this genuine difference of serine and threonine χ_1 populations from the coil average to a destabilization of the side-chain hydroxyl/amide interaction in the urea-denatured state, where urea or water could form hydrogen bonds to both groups. Interestingly, also deviations in backbone conformations have been found for these residues in a recent study using RDCs.⁴⁴ Such deviations had not been detected in the earlier studies of χ_1 conformations in unfolded proteins,^{16–18} possibly due to the more limited precision. In contrast, stronger deviations were observed for aromatic residues in urea-denatured

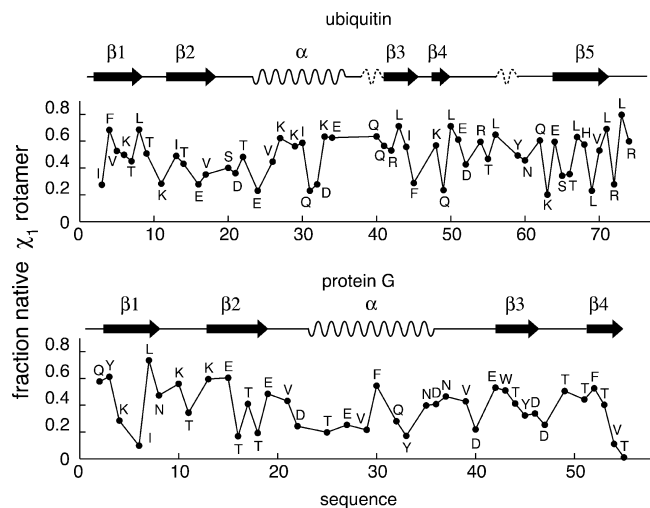


Figure 7. Fraction of native-state χ_1 angles contained in 3J -derived rotamer populations of urea-denatured ubiquitin (top) and protein G (bottom). Native-state χ_1 angles are approximated as staggered rotamers and taken from the first entry of the folded native NMR structure (PDB code 1d3z)⁴⁶ and the 1.1 Å X-ray structure of protein G (PDB code 1igd).⁴⁷ The native-state secondary structure of both proteins is shown at the top of the respective panels.

lysozyme.¹⁸ This effect is not clearly visible for the 10 aromatic residues in urea-denatured ubiquitin and protein G (Figure 5), where the deviations from the coil predictions and their individual variations are not stronger than for many other amino acids such as, for example, valines and leucines.

It is expected that the observed variations in the χ_1 populations of the urea-denatured state correspond to sequence-specific structural preferences, such as the recently detected 10–20% population of the native-state, first β -hairpin (residues M1–V17) in ubiquitin.¹⁵ However, the exact extent of the structural preferences is currently difficult to establish, because proper random coil “baseline” χ_1 populations are not known with sufficient precision in solution to be able to interpret population differences on the order of 10% with confidence. Nevertheless, it is clear that a high degree of native-state χ_1 conformations is contained in the observed χ_1 populations of the urea-denatured states. Figure 7 shows this fraction of native-state side-chain conformations in the χ_1 populations when native-state χ_1 angles are approximated by staggered rotamers. High native rotamer populations (>50%) are found for a number of residues in ubiquitin’s β -strands β_1 , β_3 , β_4 , and β_5 , and in its α -helix, as well as in all β -strands of protein G. Average native rotamer populations are 49% for ubiquitin and 38% for protein G. This indicates that the transition to a fully formed structure does not require a particularly high entropic cost.

Stereospecific Methylene H^β Chemical Shifts. The availability of the stereospecific assignments for the $^1H^{\beta 2/3}$ resonances allows one to analyze their chemical shift behavior. Average chemical shift differences, $\delta H^{\beta 2} - \delta H^{\beta 3}$, are shown in Figure 8 for all amino acids with distinct β -methylene resonances that could be observed in both proteins. For all observed amino acids (D, E, F, H, K, L, N, P, Q, R, S, W, Y) with the exception of leucine

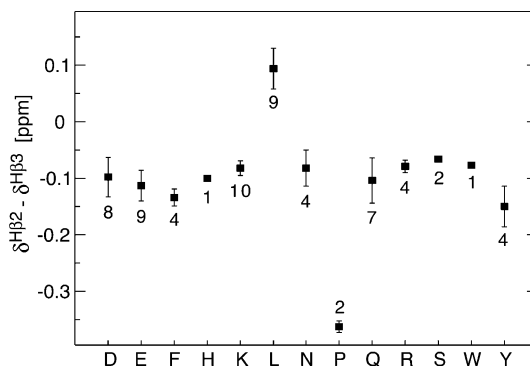


Figure 8. Differences of β -methylene 1H chemical shifts in urea-denatured proteins. Average chemical shift differences of stereospecifically assigned $H^{\beta 2}$ and $H^{\beta 3}$ resonances are shown for all amino acids that could be observed in ubiquitin and protein G. Error bars indicate standard deviations. Data are labeled by the number of observations.

and proline, $\delta H^{\beta 2}$ is smaller than $\delta H^{\beta 3}$ by about 0.1 ppm. Thus, $H^{\beta 2}$ is usually upfield from $H^{\beta 3}$ (see also Figure 1, K10). Variations of individual amino acids around the mean shift difference are about 0.05 ppm. For prolines, the upfield shift of $H^{\beta 2}$ is stronger (0.36 ppm). In contrast, for leucine, $H^{\beta 2}$ is found downfield of $H^{\beta 3}$ with a mean chemical shift difference of 0.09 ppm. This behavior is evident in Figure 1, where the intensity patterns of the upfield and downfield H^β protons are inverted for L10 relative to K10 in the two quantitative $^3J_{NH\beta}$ and $^3J_{H\alpha H\beta}$ experiments. We speculate that this unusual behavior is related to the extreme bulkiness of the two leucine δ -methyl groups, which restricts the entire side-chain conformations in folded structures to only two strongly populated classes ($\chi_1/\chi_2 = -60^\circ/180^\circ$ or $180^\circ/+60^\circ$) in folded structures⁴⁵ and also causes the extremely high (70–80%) p_{-60} and low (1–17%) p_{+60} values in the unfolded proteins (Figure 5).

Conclusion

In summary, we have presented optimized detection schemes for side-chain $^1H^\beta$ resonances in unfolded proteins that yield highly precise structural information about the χ_1 angle from up to six $^3J_{H\alpha H\beta}$, $^3J_{NH\beta}$, and $^3J_{CH\beta}$ coupling constants and up to two $^1D_{H\beta C\beta}$ RDCs. Interpretation of the detected 3J couplings in urea-denatured ubiquitin and protein G by a model of staggered χ_1 rotamers^{38–40} and previously published Karplus coefficients³⁷ provides stereoassignments of $^1H^\beta$ methylene protons and yields excellent agreement. This corroborates both the staggered rotamer model and the high precision of the Karplus coefficients. For most residues, the precision of individual rotamer populations is better than 2% as estimated from error propagation. As found in earlier studies,¹⁸ the rotamer populations are in the vicinity of averages obtained from coil regions of folded protein structures. However, individual variations from these averages of up to 40% are highly significant and must originate from sequence- and residue-specific interactions. Particularly strong deviations from the coil average are found for serine and threonine residues, an effect that may be explained by a weakening of side-chain to backbone hydrogen bonds in the urea-denatured state.

The measured $^1D_{H\beta C\beta}$ RDCs correlate well with predicted RDCs based on steric alignment of a coil model ensemble of the unfolded state generated by the program *Flexible-Meccano*,

(44) Nodet, G.; Salmon, L.; Ozanne, V.; Meier, S.; Jensen, M.; Blackledge, M. *J. Am. Chem. Soc.* **2009**, *131*, 17908–17918.

(46) Cornilescu, G.; Marquardt, J.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 6836–6837.

(47) Derrick, J.; Wigley, D. *J. Mol. Biol.* **1994**, *243*, 906–918.

(45) Ponder, J. W.; Richards, F. M. *J. Mol. Biol.* **1987**, *193*, 775–791.

where the side-chain conformations had been adjusted according to the J -derived χ_1 rotamer populations. This agreement validates the coil model as a good first approximation of the unfolded state. However, deviations between the measured and predicted values also indicate that the local backbone geometries may be improved by incorporation of the additional RDC information. The ease of detection of a large number of highly precise side-chain RDCs should make it possible to obtain such a more accurate description of backbone and side-chain conformations in unfolded states.

Acknowledgment. We thank C. Howald, K. Rathgeb-Szabo, and M. Rogowski for sample preparation, and we also thank H.-J.

Sass and S. Kasprzak for helpful discussions. This work was supported by SNF Grant 31-109712.

Supporting Information Available: Details of quantitative $^3J_{\text{NH}\beta\text{-HNHB}}$, $^3J_{\text{C}^{\text{H}}\beta\text{-HN(CO)HB}}$, $^3J_{\text{H}\alpha\text{H}\beta\text{-HAHB(CACO)NH}}$, and IPAP-HBHA(CO)NH experiments, figures showing comparisons of experimental and calculated 3J and RDC values for all observed residues in urea-denatured ubiquitin and protein G, and tables of measured 3J and RDC values as well as of the fitted χ_1 rotamer populations. This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA910331T

Side chain χ_1 conformations in urea-denatured ubiquitin and protein G from 3J coupling constants and residual dipolar couplings

Supporting Information

Navratna Vajpai¹, Martin Gentner¹, Jie-rong Huang¹, Martin Blackledge² and Stephan Grzesiek¹

¹Biozentrum, University of Basel, Klingelbergstrasse 50/70, 4056 Basel, Switzerland

²Institut de Biologie Structurale Jean-Pierre Ebel, CEA, CNRS, UJF UMR 5075, 41 Rue Jules Horowitz, Grenoble 38027, France

To whom correspondence should be addressed (stephan.grzesiek@unibas.ch)

file: Supmat_2_1.doc
last saved: 1/13/10 6:25 PM

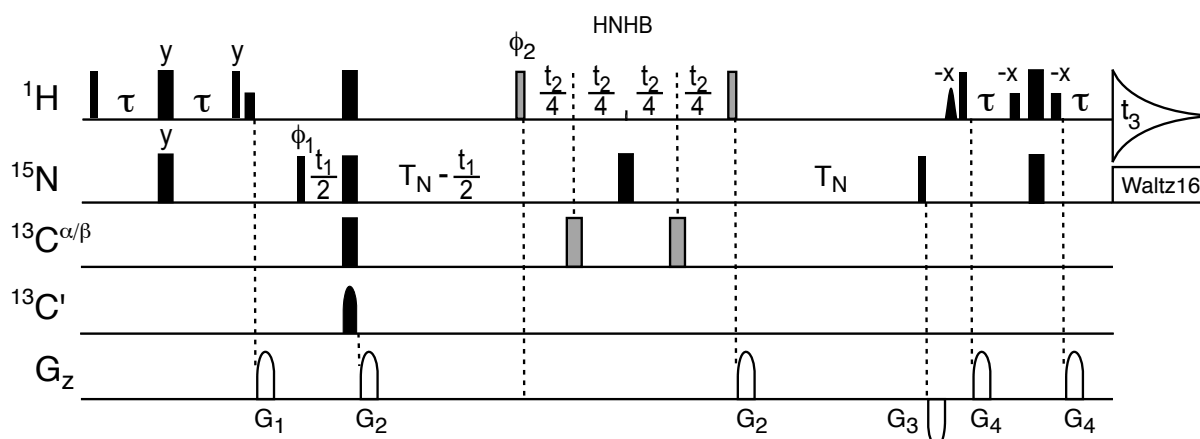


Figure S1: Pulse scheme of the quantitative-J HNHB experiment for determination of $^3J_{\text{NHB}}$ coupling constants. Narrow and wide pulses have flip angles of 90° and 180° , respectively and are applied with phase x unless specified otherwise. The ^1H , ^{15}N , $^{13}\text{C}'$, and $^{13}\text{C}^{\alpha/\beta}$ carriers are set to 4.7 (H_2O), 116.5, 177, and 46 ppm, respectively. ^1H pulses have an RF field strength of 28 kHz with the exception of the ~ 2 ms sinc water-selective 90° pulse applied before the reverse ^{15}N - ^1H INEPT transfer and the water-selective, low-power, rectangular ~ 1 ms 90° pulses. The ^{15}N pulses have an RF field strength of 6.25 kHz. ^{15}N WALTZ-16 decoupling during acquisition is applied at an RF field strength of 1.5 kHz. $^{13}\text{C}^{\alpha/\beta}$ 180° decoupling pulses have an RF field strength of 15.1 kHz. The $^{13}\text{C}'$ 180° decoupling pulse has a sinc shape and a duration of 220 μs . Gradient durations (z-direction, sine bell shaped; 30 G/cm at center): $G_{1,2,3,4} = 3.0, 1.5, 2.0, 0.4$ ms. Delays: $\tau = 2.25$ ms, $T_N = 27.0$ ms. Phases: $\phi_1 = x, -x$; $\phi_2 = 45^\circ, 45^\circ, 225^\circ, 225^\circ$; $\phi_{\text{rec}} = x, -x, -x, x$. Quadrature detection in the indirect dimensions was achieved by incrementing phases ϕ_1 (^{15}N) and ϕ_2 ($^1\text{H}^\beta$) in the usual States-TPPI manner. Data matrices were recorded as $1024 * (^1\text{H}^{\text{N}}) \times 120 * (^1\text{H}^\beta) \times (52/47) * (^{15}\text{N})$ complex points (ubiquitin/protein G) with acquisition times of 85.2, 30.2, and 51.3 ms, respectively. The total experimental times using 4 scans were 29/26 h for ubiquitin/protein G.

For determination of $^3J_{\text{NHB}}$ coupling constants, a separate reference 2D experiment was recorded, where the transfer from ^{15}N to $^1\text{H}^\beta$ is suppressed.¹ This is achieved by omitting the grey shaded pulses and the t_2 evolution period and changing the receiver phase to $\phi_{\text{rec}} = x, -x, x, -x$. Data were analyzed by fitting the resonances with the NLINLS program of NMRPipe. The intensity of the 2D reference experiment was scaled appropriately to its 3D counterpart,¹ and the $^3J_{\text{NHB}}$ coupling constants were determined from the ratios of cross (A_C) and reference peak (A_R) heights according to $^3J_{\text{NHB}} = (1/(\pi T_N)) * \text{asin}(\text{sqrt}(|A_C/A_R|))$.

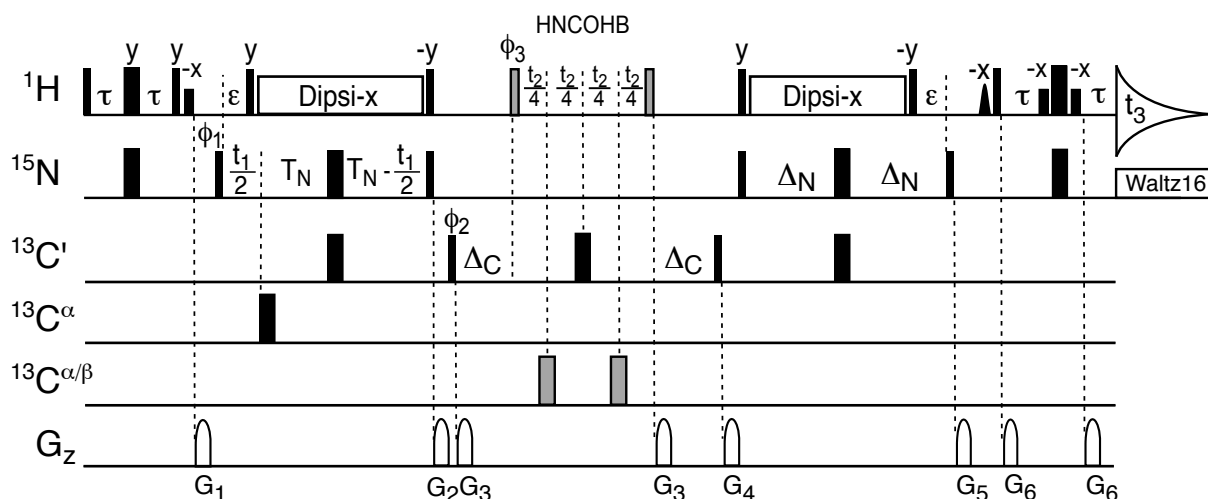


Figure S2: Pulse scheme of the quantitative-J HN(CO)HB experiment for determination of $^3J_{\text{COHB}}$ coupling constants. Narrow and wide pulses have flip angles of 90° and 180° , respectively and are applied with phase x unless specified otherwise. The ^1H , ^{15}N , $^{13}\text{C}'$, $^{13}\text{C}^\alpha$, and $^{13}\text{C}^{\alpha/\beta}$ carriers are set to 4.7 (H_2O), 116.5, 177, 56, and 46 ppm, respectively. ^1H pulses have an RF field strength of 28 kHz with the exception of the ~ 2 ms sinc water-selective 90° pulse applied before the reverse ^{15}N - ^1H INEPT transfer, the water-selective, low-power, rectangular ~ 1 ms 90° pulses, and the 6.25 kHz DIPSII-2 decoupling sequence (applied along the x -axis). The ^{15}N pulses have an RF field strength of 6.25 kHz. ^{15}N WALTZ-16 decoupling during acquisition is applied at an RF field strength of 1.5 kHz. $^{13}\text{C}'$, $^{13}\text{C}^\alpha$, and $^{13}\text{C}^{\alpha/\beta}$ pulses have an RF field strength of 6.3, 6.3, and 14.3 kHz, respectively. Gradient durations (z -direction, sine bell shaped; 30 G/cm at center): $G_{1,2,3,4,5,6} = 1.0, 1.5, 2.35, 1.5, 1.0, 0.4$ ms. Delays: $\tau = 2.25$ ms, $T_N = 20.0$ ms, $\Delta_N = 13.5$ ms, $\Delta_C = 26.85$ ms. Phases: $\phi_1 = x$; $\phi_2 = x, -x$; $\phi_3 = 45^\circ, 45^\circ, 225^\circ, 225^\circ$; $\phi_{\text{rec}} = x, -x, -x, x$. Quadrature detection in the indirect dimensions was achieved by incrementing phases ϕ_1 (^{15}N) and ϕ_3 ($^1\text{H}^\beta$) in the usual States-TPPI manner. Data matrices were recorded as $1024 * (^1\text{H}^\text{N}) \times 120 * (^1\text{H}^\beta) \times (40/36) * (^{15}\text{N})$ complex points (ubiquitin/protein G) with acquisition times of 85.2, 30.2, and 39.4 ms, respectively. The total experimental times using 4 scans were 24/20 h (ubiquitin/protein G).

For determination of $^3J_{\text{COHB}}$ coupling constants, a separate reference 2D experiment was recorded, where the transfer from $^{13}\text{C}'$ to $^1\text{H}^\beta$ is suppressed.¹ This is achieved by omitting the grey shaded pulses and the t_2 evolution period and changing the receiver phase to $\phi_{\text{rec}} = x, -x, x, -x$. Data were analyzed by fitting the resonances with the NLINLS program of NMRPipe. The intensity of the 2D reference experiment was scaled appropriately to its 3D counterpart,¹ and the $^3J_{\text{NHB}}$ coupling constants were determined from the ratios of cross (A_C) and reference peak (A_R) heights according to $^3J_{\text{COHB}} = (1/(\pi\Delta_C)) * \text{asin}(\text{sqrt}(|A_C/A_R|))$.

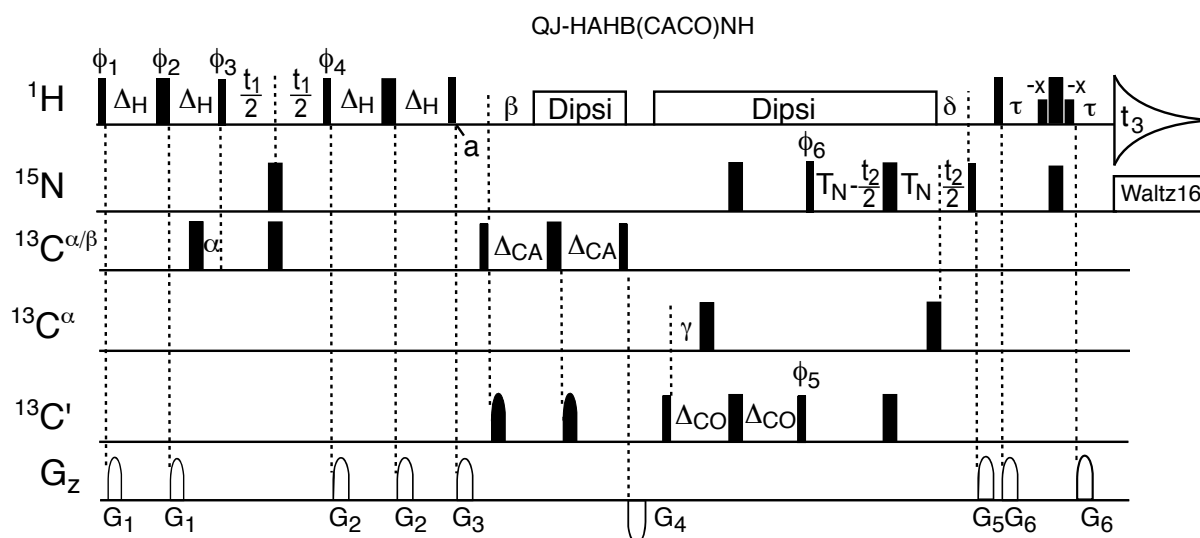


Figure S3: Pulse scheme of the quantitative-J HAHB(CACO)NH experiment for determination of $^3J_{\text{HAHB}}$ coupling constants. Narrow and wide pulses have flip angles of 90° and 180° , respectively and are applied with phase x unless specified otherwise. The ^{15}N , $^{13}\text{C}'$, $^{13}\text{C}^{\alpha/\beta}$, and $^{13}\text{C}^\alpha$ carriers are set to 116.5, 177, 46, and 56 ppm, respectively. During the first part of the sequence (HAHB J-correlation), the ^1H carrier is set to 2.7 ppm, whereas it is switched to 4.7 ppm (H_2O) at point a. ^1H pulses have an RF field strength of 28 kHz with the exception of the water-selective, low-power, rectangular ~ 1 ms 90° pulses, and the 6.25 kHz DIPSII-2 decoupling sequence. The ^{15}N pulses have an RF field strength of 6.25 kHz. ^{15}N WALTZ-16 decoupling during acquisition is applied at an RF field strength of 1.5 kHz. The RF strength of ^{13}C pulses is adjusted to minimize cross excitation between the $^{13}\text{C}^{\alpha/\beta}$ ($^{13}\text{C}^\alpha$) and the $^{13}\text{C}'$ regions: $^{13}\text{C}^{\alpha/\beta}$ $90^\circ/180^\circ$: 6.8/15.1 kHz; $^{13}\text{C}^\alpha$ 180° : 14.0 kHz; $^{13}\text{C}'$ 180° shaped pulses have a sinc profile and a duration of 220 μs ; $^{13}\text{C}'$ rectangular pulses: 6.3 kHz. Gradient durations (z-direction, sine bell shaped; 30 G/cm at center): $G_{1,2,3,4,5,6} = 0.45, 0.2, 3.61, 1.4, 2.5, 0.4$ ms. Delays: $\Delta_{\text{H}} = 10.0$ ms, $\alpha = 1.78$ ms, $\Delta_{\text{CA}} = 3.72$ ms, $\beta = 3.57$ ms, $\Delta_{\text{CO}} = 15.0$ ms, $\gamma = 4.5$ ms, $T_{\text{N}} = 20.0$ ms, $\delta = 5.37$ ms, $\tau = 2.25$ ms. Phases: $\phi_1 = 4(y), 4(-y)$; $\phi_2 = x$; $\phi_3 = y$; $\phi_4 = y, y, -y, -y$; $\phi_5 = 48^\circ$ (to compensate the Bloch-Siegert effect); $\phi_6 = x, -x$; $\phi_{\text{rec}} = x, -x, x, -x, -x, x, -x, x$. Quadrature detection in the indirect dimensions was achieved by incrementing phases ϕ_{1-3} ($^1\text{H}^{\alpha/\beta}$) and ϕ_6 (^{15}N) in the usual States-TPPI manner. Data matrices were recorded as $1024 * (^1\text{H}^{\text{N}}) \times 105 * (^1\text{H}^{\text{B}}) \times (40/36) * (^{15}\text{N})$ complex points (ubiquitin/protein G) with acquisition times of 85.2, 28.7, and 39.4 ms, respectively. Total experimental times using 8 scans were 42/38 hrs (ubiquitin/protein G). The $^3J_{\text{HAHB}}$ coupling constants were determined from the ratios of cross (A_{C}) and diagonal peak (A_{R}) heights according to $^3J_{\text{HAHB}} = (1/(2\pi\Delta_{\text{H}})) * \text{atan}(\text{sqrt}(|A_{\text{C}}/A_{\text{R}}|))$.

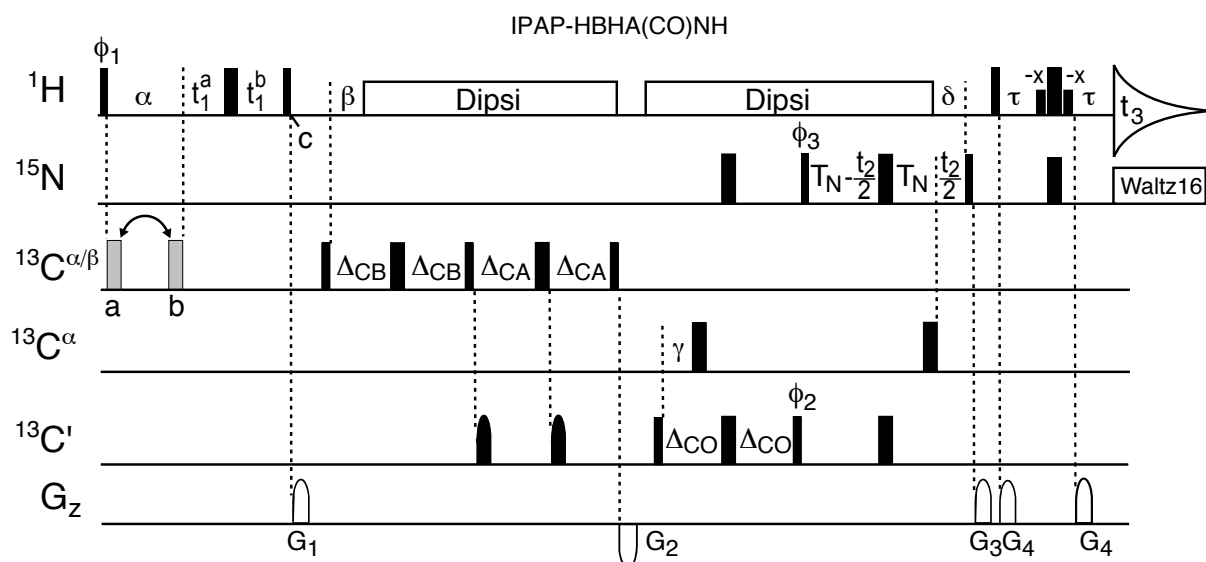


Figure S4: Pulse scheme of the IPAP-HBHA(CO)NH experiment for determination of $^1D_{\text{C}\beta\text{H}\beta}$ and $^1D_{\text{C}\alpha\text{H}\alpha}$ RDCs. The pulse sequence is derived from the standard semi-constant time HBHA(CO)NH experiment² and enhanced by an IPAP element³ for the $^1\text{H}^{\alpha/\beta}$ dimension. Narrow and wide pulses have flip angles of 90° and 180° , respectively and are applied with phase x unless specified otherwise. The ^{15}N , $^{13}\text{C}'$, $^{13}\text{C}^{\alpha/\beta}$ and $^{13}\text{C}^\alpha$ carriers are set to 116.5, 177, 46 and 56 ppm, respectively. At the beginning of the pulse sequence, the ^1H carrier is set to 2.7 ppm, whereas it is switched to 4.7 ppm (H_2O) at point c. ^1H pulses have an RF field strength of 28 kHz with the exception of the water-selective, low-power, rectangular ~ 1 ms 90° pulses, and the 6.25 kHz DIPSI-2 decoupling sequence. The ^{15}N pulses have an RF field strength of 6.25 kHz. ^{15}N WALTZ-16 decoupling during acquisition is applied at an RF field strength of 1.5 kHz. The RF strength of ^{13}C pulses is adjusted to minimize cross excitation between the $^{13}\text{C}^{\alpha/\beta}$ ($^{13}\text{C}^\alpha$) and the $^{13}\text{C}'$ regions: $^{13}\text{C}^{\alpha/\beta}$ $90^\circ/180^\circ$: 6.8/15.1 kHz; $^{13}\text{C}^\alpha$ 180° : 14.0 kHz; $^{13}\text{C}'$ 180° shaped pulses have a sinc profile and a duration of 220 μs ; $^{13}\text{C}'$ rectangular pulses: 6.3 kHz. Gradient durations (z-direction, sine bell shaped; 30 G/cm at center): $G_{1,2,3,4} = 2.0, 1.4, 2.5, 0.4$ ms. Delays: $\alpha = 1.79$ ms, $\beta = 2.1$ ms, $\Delta_{\text{CB}} = 3.32$ ms, $\Delta_{\text{CA}} = 3.72$ ms, $\Delta_{\text{CO}} = 15.0$ ms, $\gamma = 4.5$ ms, $T_{\text{N}} = 20.0$ ms, $\delta = 5.37$ ms, $\tau = 2.25$ ms. Phases: $\phi_1 = 45^\circ$; $\phi_2 = 48^\circ$ (to compensate the Bloch-Siegert effect); $\phi_3 = x, -x$; $\phi_{\text{rec}} = x, -x$. Quadrature detection in the indirect dimensions was achieved by incrementing phases ϕ_1 ($^1\text{H}^{\alpha/\beta}$) and ϕ_2 (^{15}N) in the usual States-TPPI manner. The initial delays for the semi-constant time proton evolution period were set to $t_{1,0}^{\text{a}} = 4$ μs , $t_{1,0}^{\text{b}} = t_{1,0}^{\text{a}} + \alpha$. t_1^{b} is decremented by $\Delta t_1^{\text{b}} = -t_{1,0}^{\text{b}} / (\text{N}(^1\text{H}^{\alpha/\beta}) + 1)$, where $\text{N}(^1\text{H}^{\alpha/\beta})$ is the number of complex increments in the $^1\text{H}^{\alpha/\beta}$ dimension. t_1^{a} is incremented by $\Delta t_1^{\text{a}} = 1/\text{sw}(^1\text{H}^{\alpha/\beta}) + \Delta t_1^{\text{b}}$, where $\text{sw}(^1\text{H}^{\alpha/\beta})$ is the sweep width in the $^1\text{H}^{\alpha/\beta}$ dimension. The anti-phase (AP) and in-phase (IP) spectra are recorded in an interleaved manner by switching the position of the 180° pulse (grey) on the $^{13}\text{C}^{\alpha/\beta}$ channel between positions a and b, respectively. Data matrices were recorded as $1024 * (^1\text{H}^{\text{N}}) \times 136 * (^1\text{H}^{\alpha/\beta}) \times (40/36) * (^{15}\text{N})$ complex points (ubiquitin/protein G) with acquisition times of 85.2, 34.2, and 39.4 ms, respectively. Total experimental times using 2 scans were 27/23 h (ubiquitin/protein G).

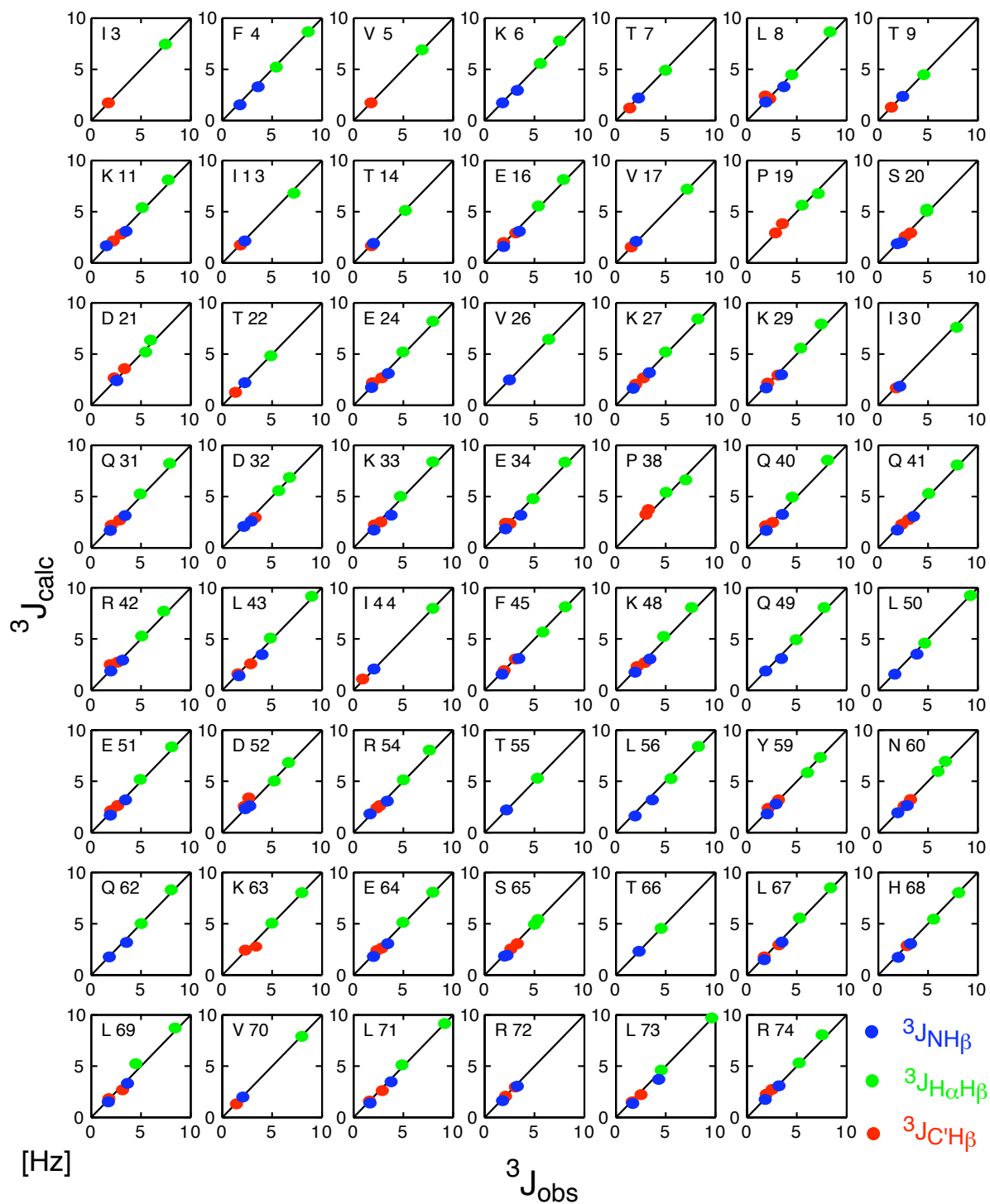


Figure S5: Urea-denatured ubiquitin: comparison of experimental (obs) $^3J_{X\text{H}\beta}$ ($X = \text{C}', \text{N}, \text{HA}$) constants to values derived (calc) according to the χ_1 rotamer population fit of Eq. 3 for all residues, for which populations could be obtained.

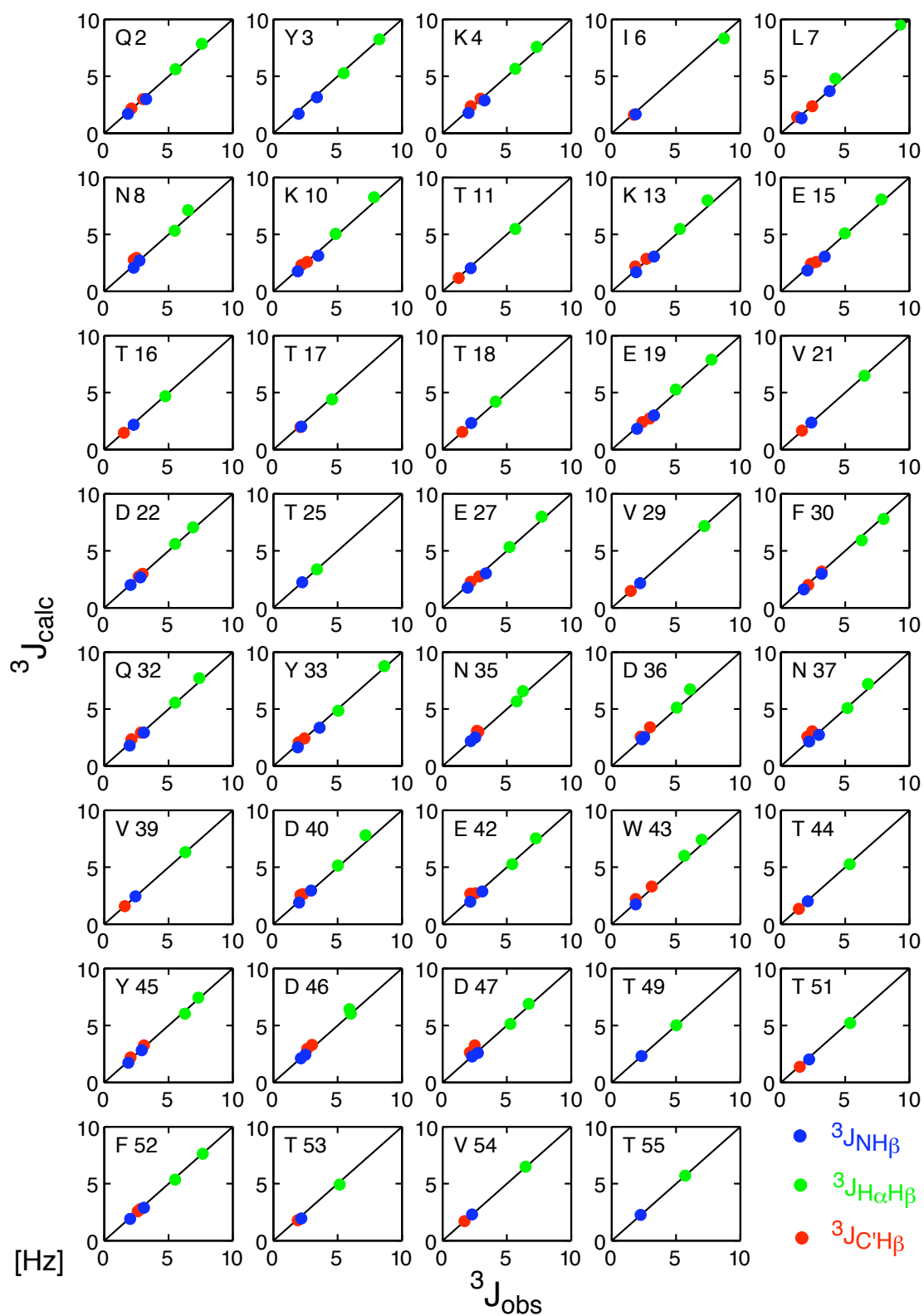


Figure S6: Same as Figure S5 for urea-denatured protein G.

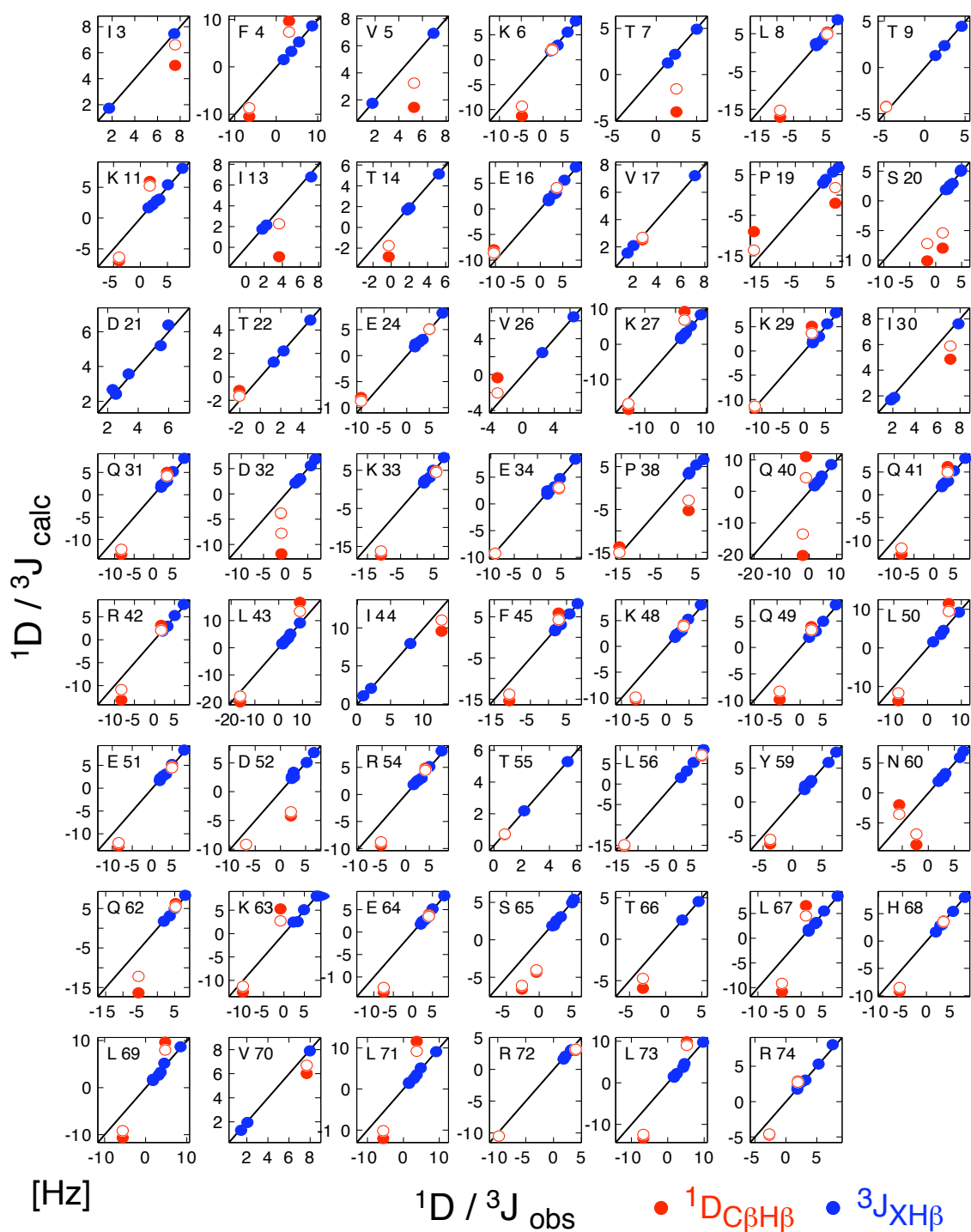


Figure S7: Urea-denatured ubiquitin: comparison of experimental (obs) and predicted (calc) $^1\text{D}_{\text{C}\beta\text{H}\beta}$ RDCs (red filled circles) based on the χ_1 rotamer populations derived from the fit of Eq. 3 of $^3\text{J}_{\text{XH}\beta}$ constants and the coil model ensemble. For comparison experimental and predicted $^3\text{J}_{\text{XH}\beta}$ couplings are also indicated (blue filled circles). Data are shown for all residues, for which χ_1 rotamer populations could be derived. Red open circles indicate predictions for $^1\text{D}_{\text{C}\beta\text{H}\beta}$ RDCs, when these RDC data were also included into the rotamer population fit of Eq. 3 (see text).

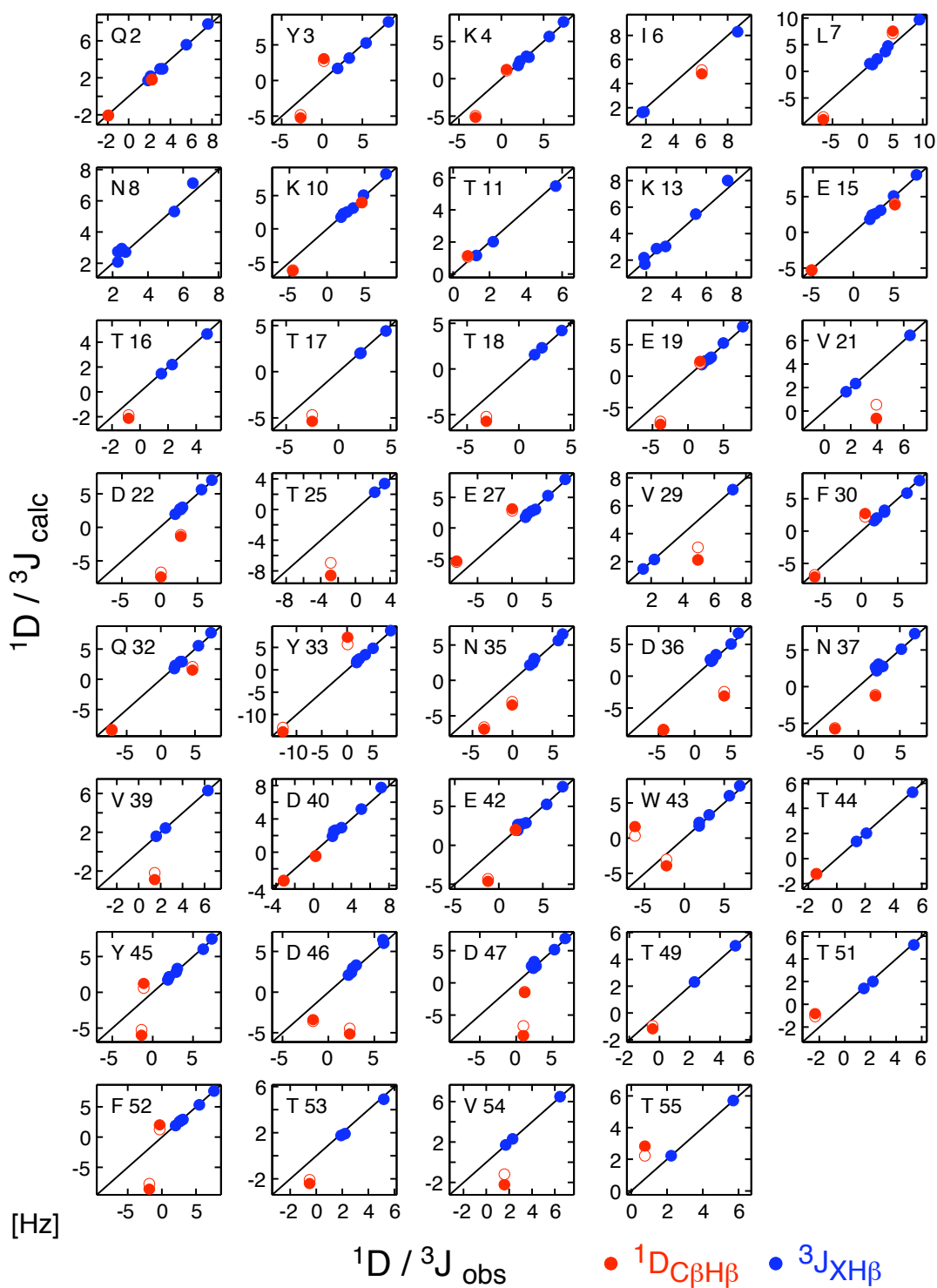


Figure S8: Same as Figure S7 for urea-denatured protein G.

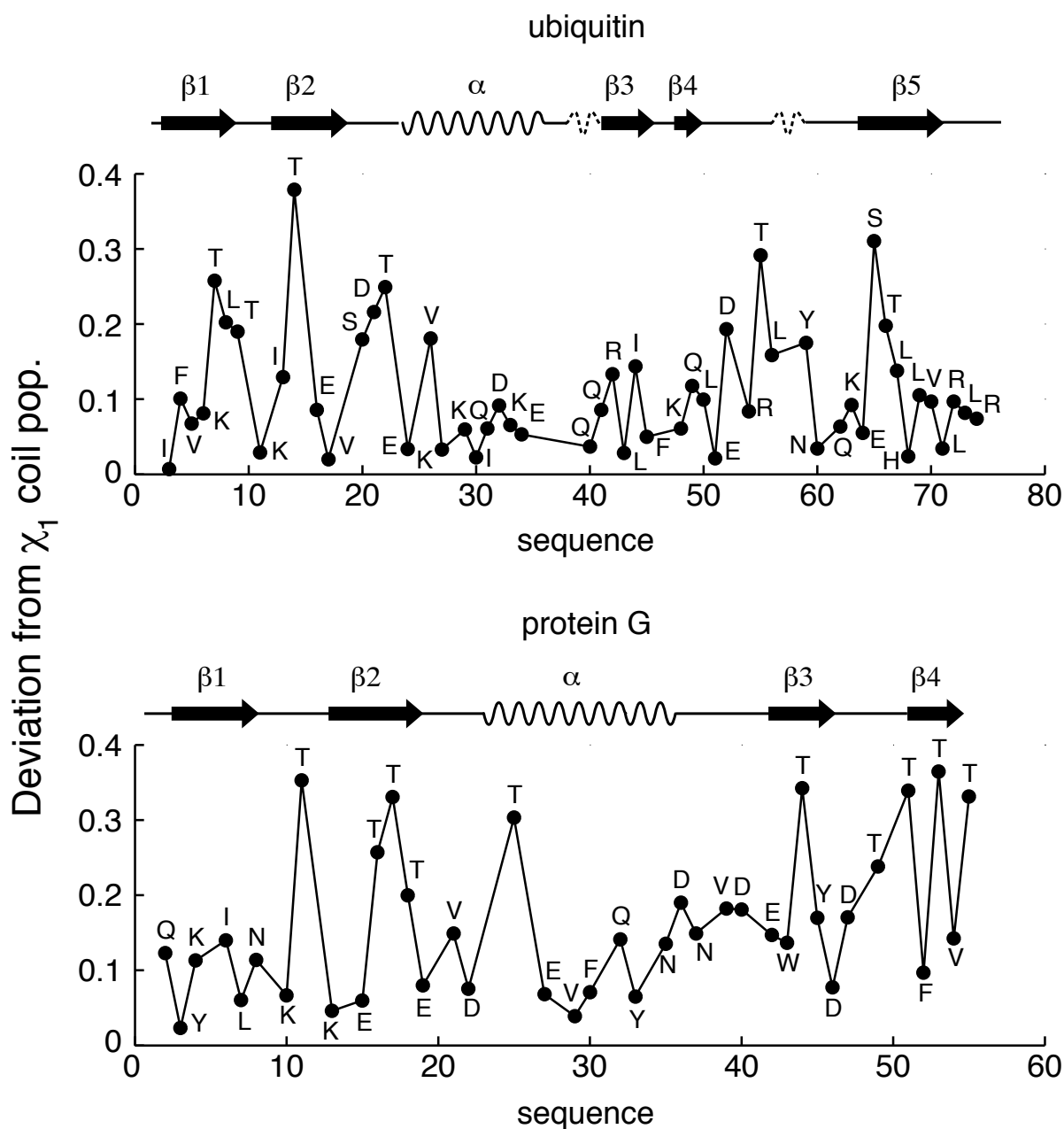


Figure S9: Deviations of the experimentally (3J -only) derived χ_1 rotamer populations in urea-denatured ubiquitin (top) and protein G (bottom) from the average of the PDB coil structures as a function of the residue number. The deviations are calculated as $\|\vec{p}_{\text{exp}} - \vec{p}_{\text{coil}}\|$ with $\vec{p} = (p_{-60^\circ}, p_{+60^\circ}, p_{180^\circ})$.

38	P	J	HA	HB2	5.05	0.15	52	D	D	CB	HB2	1.66	1.00
38	P	J	HA	HB3	7.03	0.15	52	D	D	CB	HB3	-6.07	1.62
40	Q	J	N	HB2	1.92	0.10	54	R	J	N	HB2	1.66	0.22
40	Q	J	N	HB3	3.55	0.10	54	R	J	N	HB3	3.40	0.10
40	Q	J	CO	HB2	2.58	0.10	54	R	J	CO	HB2	2.67	0.10
40	Q	J	CO	HB3	1.89	0.10	54	R	J	CO	HB3	2.33	0.10
40	Q	J	HA	HB2	8.07	0.15	54	R	J	HA	HB2	7.62	0.15
40	Q	J	HA	HB3	4.52	0.15	54	R	J	HA	HB3	4.99	0.15
40	Q	D	CB	HB2	0.40	2.48	54	R	D	CB	HB2	3.64	1.12
40	Q	D	CB	HB3	-2.75	1.06	54	R	D	CB	HB3	-4.56	1.22
41	Q	J	N	HB2	1.95	0.10	55	T	J	N	HB	2.19	0.10
41	Q	J	N	HB3	3.58	0.10	55	T	J	HA	HB	5.30	0.15
41	Q	J	CO	HB2	3.03	0.10	55	T	D	CB	HB	0.73	1.00
41	Q	J	CO	HB3	2.37	0.13	56	L	J	N	HB2	1.95	0.10
41	Q	J	HA	HB2	7.91	0.15	56	L	J	N	HB3	3.65	0.10
41	Q	J	HA	HB3	5.07	0.15	56	L	J	HA	HB2	8.29	0.15
41	Q	D	CB	HB2	1.39	3.41	56	L	J	HA	HB3	5.54	0.15
41	Q	D	CB	HB3	-7.24	1.57	56	L	D	CB	HB2	8.75	1.41
42	R	J	N	HB2	1.98	0.10	56	L	D	CB	HB3	-11.82	3.81
42	R	J	N	HB3	3.18	0.10	59	Y	J	N	HB2	2.00	0.10
42	R	J	CO	HB2	2.64	0.30	59	Y	J	N	HB3	2.89	0.10
42	R	J	CO	HB3	1.90	0.18	59	Y	J	CO	HB2	3.18	0.10
42	R	J	HA	HB2	7.29	0.15	59	Y	J	CO	HB3	2.12	0.25
42	R	J	HA	HB3	5.08	0.15	59	Y	J	HA	HB2	7.35	0.15
42	R	D	CB	HB2	2.63	1.46	59	Y	J	HA	HB3	6.07	0.25
42	R	D	CB	HB3	-7.32	2.63	59	Y	D	CB	HB3	-3.84	1.00
43	L	J	N	HB2	1.69	0.10	60	N	J	N	HB2	1.98	0.10
43	L	J	N	HB3	4.01	0.10	60	N	J	N	HB3	2.88	0.10
43	L	J	CO	HB2	2.87	0.10	60	N	J	CO	HB2	3.22	0.10
43	L	J	CO	HB3	1.57	0.10	60	N	J	CO	HB3	2.61	0.10
43	L	J	HA	HB2	8.96	0.15	60	N	J	HA	HB2	6.76	0.15
43	L	J	HA	HB3	4.84	0.15	60	N	J	HA	HB3	6.03	0.15
43	L	D	CB	HB2	8.72	1.07	60	N	D	CB	HB2	-2.54	5.87
43	L	D	CB	HB3	-13.67	3.74	60	N	D	CB	HB3	-0.99	3.13
44	I	J	N	HB	2.05	0.10	62	Q	J	N	HB2	1.82	0.10
44	I	J	CO	HB	0.90	0.50	62	Q	J	N	HB3	3.54	0.10
44	I	J	HA	HB	7.97	0.15	62	Q	J	HA	HB2	8.03	0.15
44	I	D	CB	HB	9.36	6.79	62	Q	J	HA	HB3	5.04	0.15
45	F	J	N	HB2	1.75	0.10	62	Q	D	CB	HB2	3.18	2.97
45	F	J	N	HB3	3.44	0.10	62	Q	D	CB	HB3	-6.79	2.41
45	F	J	CO	HB2	3.10	0.10	63	K	J	CO	HB2	3.38	0.45
45	F	J	CO	HB3	1.97	0.10	63	K	J	CO	HB3	2.32	0.24
45	F	J	HA	HB2	8.09	0.15	63	K	J	HA	HB2	8.00	0.15
45	F	J	HA	HB3	5.82	0.15	63	K	J	HA	HB3	4.96	0.15
45	F	D	CB	HB2	3.31	1.00	63	K	D	CB	HB2	0.30	1.69
45	F	D	CB	HB3	-11.41	1.00	63	K	D	CB	HB3	-7.63	3.72
48	K	J	N	HB2	1.94	0.10	64	E	J	N	HB2	2.03	0.10
48	K	J	N	HB3	3.40	0.10	64	E	J	N	HB3	3.46	0.10
48	K	J	CO	HB2	2.89	0.10	64	E	J	CO	HB2	2.85	0.10
48	K	J	CO	HB3	2.13	0.10	64	E	J	CO	HB3	2.38	0.10
48	K	J	HA	HB2	7.61	0.15	64	E	J	HA	HB2	7.95	0.15
48	K	J	HA	HB3	4.78	0.15	64	E	J	HA	HB3	4.97	0.15
48	K	D	CB	HB2	4.15	1.00	64	E	D	CB	HB2	4.37	1.00
48	K	D	CB	HB3	-6.39	1.02	64	E	D	CB	HB3	-7.05	1.00
49	Q	J	N	HB2	1.91	0.10	65	S	J	N	HB2	2.02	0.10
49	Q	J	N	HB3	3.44	0.10	65	S	J	N	HB3	2.24	0.10
49	Q	J	HA	HB2	7.76	0.15	65	S	J	CO	HB2	3.29	0.10
49	Q	J	HA	HB3	4.96	0.15	65	S	J	CO	HB3	2.65	0.10
49	Q	D	CB	HB2	2.34	1.00	65	S	J	HA	HB2	5.02	0.15
49	Q	D	CB	HB3	-3.78	2.28	65	S	J	HA	HB3	5.34	0.15
50	L	J	N	HB2	1.70	0.12	65	S	D	CB	HB2	-2.64	1.00
50	L	J	N	HB3	3.92	0.30	65	S	D	CB	HB3	-1.02	1.00
50	L	J	HA	HB2	9.29	0.15	66	T	J	N	HB	2.33	0.10
50	L	J	HA	HB3	4.68	0.15	66	T	J	HA	HB	4.56	0.15
50	L	D	CB	HB2	5.26	2.67	66	T	D	CB	HB	-3.36	1.00
50	L	D	CB	HB3	-7.69	1.47	67	L	J	N	HB2	1.74	0.10
51	E	J	N	HB2	1.92	0.10	67	L	J	N	HB3	3.51	0.10
51	E	J	N	HB3	3.46	0.10	67	L	J	CO	HB2	3.22	0.10
51	E	J	CO	HB2	2.68	0.10	67	L	J	CO	HB3	1.73	0.10
51	E	J	CO	HB3	1.99	0.10	67	L	J	HA	HB2	8.41	0.24
51	E	J	HA	HB2	8.08	0.15	67	L	J	HA	HB3	5.30	0.15
51	E	J	HA	HB3	4.96	0.27	67	L	D	CB	HB2	1.03	1.00
51	E	D	CB	HB2	3.65	2.61	67	L	D	CB	HB3	-3.73	1.34
51	E	D	CB	HB3	-6.97	2.50	68	H	J	N	HB2	2.01	0.10
52	D	J	N	HB2	2.30	0.10	68	H	J	N	HB3	3.26	0.10
52	D	J	N	HB3	2.74	0.10	68	H	J	CO	HB2	2.93	0.10
52	D	J	CO	HB2	2.22	0.16	68	H	J	HA	HB2	8.07	0.15
52	D	J	CO	HB3	2.63	0.69	68	H	J	HA	HB3	5.58	0.15
52	D	J	HA	HB2	6.63	0.15	68	H	D	CB	HB2	1.64	4.39
52	D	J	HA	HB3	5.22	0.15	68	H	D	CB	HB3	-6.24	1.54

69	L	J	N	HB2	1.72	0.10
69	L	J	N	HB3	3.67	0.10
69	L	J	CO	HB2	3.17	0.10
69	L	J	CO	HB3	1.75	0.10
69	L	J	HA	HB2	8.46	0.15
69	L	J	HA	HB3	4.48	0.15
69	L	D	CB	HB2	3.67	1.69
69	L	D	CB	HB3	-5.04	1.00
70	V	J	N	HB	2.05	0.10
70	V	J	CO	HB	1.42	0.16
70	V	J	HA	HB	7.98	0.15
70	V	D	CB	HB	6.75	1.77
71	L	J	N	HB2	1.65	0.10
71	L	J	N	HB3	3.75	0.10
71	L	J	CO	HB2	2.90	0.10
71	L	J	CO	HB3	1.62	0.10
71	L	J	HA	HB2	9.15	0.15
71	L	J	HA	HB3	4.89	0.15
71	L	D	CB	HB2	3.68	1.00
71	L	D	CB	HB3	-5.56	1.04
72	R	J	N	HB2	1.81	0.10
72	R	J	N	HB3	3.29	0.10
72	R	J	CO	HB2	3.08	0.10
72	R	J	CO	HB3	2.08	0.10
72	R	D	CB	HB2	3.76	1.30
72	R	D	CB	HB3	-8.54	2.15
73	L	J	N	HB2	1.70	0.10
73	L	J	N	HB3	4.29	0.10
73	L	J	CO	HB2	2.52	0.10
73	L	J	CO	HB3	1.63	0.10
73	L	J	HA	HB2	9.65	0.15
73	L	J	HA	HB3	4.56	0.15
73	L	D	CB	HB2	4.99	1.00
73	L	D	CB	HB3	-5.90	1.25
74	R	J	N	HB2	1.85	0.10
74	R	J	N	HB3	3.20	0.10
74	R	J	CO	HB2	2.51	0.10
74	R	J	CO	HB3	1.94	0.10
74	R	J	HA	HB2	7.54	0.15
74	R	J	HA	HB3	5.26	0.15
74	R	D	CB	HB2	3.19	2.85
74	R	D	CB	HB3	-2.80	1.00

Table S1B: Detected ${}^3J_{\text{NH}\beta}$, ${}^3J_{\text{C}'\text{H}\beta}$, ${}^3J_{\text{H}\alpha\text{H}\beta}$ couplings and ${}^1D_{\text{C}\beta\text{H}\beta}$ RDCs in urea-denatured protein G (in Hz).

Nr.	Res.	Type	Nuc.1	Nuc.2	Coupling	Error
2	Q	J	N	HB2	1.85	0.10
2	Q	J	N	HB3	3.27	0.10
2	Q	J	CO	HB2	3.01	0.10
2	Q	J	CO	HB3	2.12	0.10
2	Q	J	HA	HB2	7.61	0.15
2	Q	J	HA	HB3	5.56	0.15
2	Q	D	CB	HB2	1.71	1.00
2	Q	D	CB	HB3	-2.36	1.00
3	Y	J	N	HB2	1.96	0.10
3	Y	J	N	HB3	3.40	0.10
3	Y	J	HA	HB2	8.25	0.15
3	Y	J	HA	HB3	5.47	0.15
3	Y	D	CB	HB2	1.20	1.36
3	Y	D	CB	HB3	-0.94	3.55
4	K	J	N	HB2	2.01	0.10
4	K	J	N	HB3	3.28	0.13
4	K	J	CO	HB2	2.97	0.10
4	K	J	CO	HB3	2.19	0.10
4	K	J	HA	HB2	7.33	0.15
4	K	J	HA	HB3	5.66	0.15
4	K	D	CB	HB2	0.36	1.00
4	K	D	CB	HB3	-2.65	1.36
6	I	J	N	HB	1.86	0.10
6	I	J	CO	HB	1.73	0.10
6	I	J	HA	HB	8.73	0.23
6	I	D	CB	HB	6.22	1.00
7	L	J	N	HB2	1.61	0.10
7	L	J	N	HB3	3.79	0.10
7	L	J	CO	HB2	2.42	0.10
7	L	J	CO	HB3	1.26	0.10
7	L	J	HA	HB2	9.44	0.19
7	L	J	HA	HB3	4.26	0.28
7	L	D	CB	HB2	4.58	1.00
7	L	D	CB	HB3	-5.95	1.11
8	N	J	N	HB2	2.28	0.11
8	N	J	N	HB3	2.74	0.10
8	N	J	CO	HB2	2.30	0.10
8	N	J	CO	HB3	2.50	0.10
8	N	J	HA	HB2	6.54	0.15
8	N	J	HA	HB3	5.50	0.15
10	K	J	N	HB2	1.89	0.10
10	K	J	N	HB3	3.51	0.10
10	K	J	CO	HB2	2.65	0.10
10	K	J	CO	HB3	2.18	0.10
10	K	J	HA	HB2	7.84	0.15
10	K	J	HA	HB3	4.83	0.15
10	K	D	CB	HB2	3.97	1.19
10	K	D	CB	HB3	-4.63	1.00
11	T	J	N	HB	2.21	0.10
11	T	J	CO	HB	1.27	0.10
11	T	J	HA	HB	5.65	0.15
11	T	D	CB	HB	0.96	1.00
13	K	J	N	HB2	1.90	0.10
13	K	J	N	HB3	3.28	0.10
13	K	J	CO	HB2	2.68	0.12
13	K	J	CO	HB3	1.85	0.10
13	K	J	HA	HB2	7.43	0.15
13	K	J	HA	HB3	5.30	0.15
13	K	D	CB	HB2	2.25	1.68
13	K	D	CB	HB3	-3.43	1.64
15	E	J	N	HB2	2.07	0.10
15	E	J	N	HB3	3.40	0.10
15	E	J	CO	HB2	2.75	0.12
15	E	J	CO	HB3	2.34	0.10
15	E	J	HA	HB2	7.81	0.16
15	E	J	HA	HB3	4.97	0.16
15	E	D	CB	HB2	4.12	2.39

15	E	D	CB	HB3	-3.91	2.24
16	T	J	N	HB	2.29	0.10
16	T	J	CO	HB	1.53	0.10
16	T	J	HA	HB	4.77	0.15
16	T	D	CB	HB	-0.54	1.00
17	T	J	N	HB	2.18	0.10
17	T	J	CO	HB	2.11	0.12
17	T	J	HA	HB	4.56	0.15
17	T	D	CB	HB	-2.26	1.00
18	T	J	N	HB	2.24	0.10
18	T	J	CO	HB	1.52	0.10
18	T	J	HA	HB	4.13	0.15
18	T	D	CB	HB	-3.67	1.00
19	E	J	N	HB2	1.96	0.10
19	E	J	N	HB3	3.31	0.10
19	E	J	CO	HB2	2.95	0.10
19	E	J	CO	HB3	2.37	0.10
19	E	J	HA	HB2	7.74	0.15
19	E	J	HA	HB3	5.00	0.15
19	E	D	CB	HB2	2.87	2.37
19	E	D	CB	HB3	-3.46	1.00
21	V	J	N	HB	2.37	0.10
21	V	J	CO	HB	1.65	0.10
21	V	J	HA	HB	6.49	0.15
21	V	D	CB	HB	3.04	2.37
22	D	J	N	HB2	2.05	0.10
22	D	J	N	HB3	2.80	0.10
22	D	J	CO	HB2	3.00	0.10
22	D	J	CO	HB3	2.69	0.11
22	D	J	HA	HB2	6.89	0.15
22	D	J	HA	HB3	5.53	0.15
22	D	D	CB	HB2	1.59	1.89
22	D	D	CB	HB3	-0.16	1.00
25	T	J	N	HB	2.25	0.10
25	T	J	HA	HB	3.37	0.15
25	T	D	CB	HB	-1.57	1.96
27	E	J	N	HB2	1.97	0.10
27	E	J	N	HB3	3.39	0.10
27	E	J	CO	HB2	2.85	0.10
27	E	J	CO	HB3	2.19	0.10
27	E	J	HA	HB2	7.70	0.15
27	E	J	HA	HB3	5.22	0.15
27	E	D	CB	HB2	0.95	1.93
27	E	D	CB	HB3	-7.02	1.81
29	V	J	N	HB	2.19	0.10
29	V	J	CO	HB	1.49	0.10
29	V	J	HA	HB	7.19	0.15
29	V	D	CB	HB	4.57	3.19
30	F	J	N	HB2	1.78	0.10
30	F	J	N	HB3	3.18	0.10
30	F	J	CO	HB2	3.18	0.10
30	F	J	CO	HB3	2.12	0.10
30	F	J	HA	HB2	7.98	0.15
30	F	J	HA	HB3	6.28	0.15
30	F	D	CB	HB2	2.02	2.30
30	F	D	CB	HB3	-3.76	4.72
32	Q	J	N	HB2	1.97	0.10
32	Q	J	N	HB3	3.10	0.10
32	Q	J	CO	HB2	2.86	0.10
32	Q	J	CO	HB3	2.11	0.10
32	Q	J	HA	HB2	7.38	0.15
32	Q	J	HA	HB3	5.51	0.15
32	Q	D	CB	HB2	4.11	1.38
32	Q	D	CB	HB3	-7.02	1.00
33	Y	J	N	HB2	1.89	0.10
33	Y	J	N	HB3	3.60	0.10
33	Y	J	CO	HB2	2.42	0.10
33	Y	J	CO	HB3	1.97	0.10
33	Y	J	HA	HB2	8.61	0.15
33	Y	J	HA	HB3	5.07	0.35
33	Y	D	CB	HB2	3.32	5.94
33	Y	D	CB	HB3	-8.57	8.35
35	N	J	N	HB2	2.18	0.10
35	N	J	N	HB3	2.54	0.10
35	N	J	CO	HB2	2.79	0.10

35 N	J	CO	HB3	2.69	0.16	47 D	D	CB	HB3	-0.88	3.44
35 N	J	HA	HB2	6.26	0.15	49 T	J	N	HB	2.30	0.10
35 N	J	HA	HB3	5.76	0.15	49 T	J	HA	HB	5.02	0.15
35 N	D	CB	HB2	0.59	1.81	49 T	D	CB	HB	-1.87	2.63
35 N	D	CB	HB3	-1.68	3.44	51 T	J	N	HB	2.20	0.10
36 D	J	N	HB2	2.36	0.10	51 T	J	CO	HB	1.48	0.10
36 D	J	N	HB3	2.52	0.10	51 T	J	HA	HB	5.39	0.15
36 D	J	CO	HB2	2.23	0.10	51 T	D	CB	HB	-2.62	1.00
36 D	J	CO	HB3	2.97	0.10	52 F	J	N	HB2	2.03	0.10
36 D	J	HA	HB2	6.10	0.15	52 F	J	N	HB3	3.09	0.10
36 D	J	HA	HB3	5.05	0.15	52 F	J	CO	HB2	2.82	0.10
36 D	D	CB	HB2	4.34	1.00	52 F	J	CO	HB3	2.63	0.10
36 D	D	CB	HB3	-3.54	1.34	52 F	J	HA	HB2	7.66	0.15
37 N	J	N	HB2	2.21	0.10	52 F	J	HA	HB3	5.52	0.15
37 N	J	N	HB3	2.97	0.10	52 F	D	CB	HB2	-0.96	2.32
37 N	J	CO	HB2	2.06	0.18	52 F	D	CB	HB3	-3.25	2.28
37 N	J	CO	HB3	2.43	0.20	53 T	J	N	HB	2.19	0.10
37 N	J	HA	HB2	6.77	0.15	53 T	J	CO	HB	1.91	0.10
37 N	J	HA	HB3	5.18	0.15	53 T	J	HA	HB	5.17	0.15
37 N	D	CB	HB2	1.83	1.00	53 T	D	CB	HB	-0.84	1.00
37 N	D	CB	HB3	-2.90	1.00	54 V	J	N	HB	2.30	0.10
39 V	J	N	HB	2.44	0.10	54 V	J	CO	HB	1.70	0.10
39 V	J	CO	HB	1.59	0.10	54 V	J	HA	HB	6.46	0.15
39 V	J	HA	HB	6.32	0.15	54 V	D	CB	HB	1.76	1.00
39 V	D	CB	HB	0.08	3.04	55 T	J	N	HB	2.24	0.10
40 D	J	N	HB2	2.00	0.10	55 T	J	HA	HB	5.71	0.15
40 D	J	N	HB3	2.96	0.10	55 T	D	CB	HB	0.45	1.00
40 D	J	CO	HB2	2.32	0.10	56 E	J	N	HB2	2.04	0.10
40 D	J	CO	HB3	2.12	0.10	56 E	J	N	HB3	3.39	0.10
40 D	J	HA	HB2	7.15	0.15						
40 D	J	HA	HB3	5.04	0.15						
40 D	D	CB	HB2	1.52	2.19						
40 D	D	CB	HB3	-2.67	1.00						
42 E	J	N	HB2	2.17	0.10						
42 E	J	N	HB3	3.10	0.10						
42 E	J	CO	HB2	2.60	0.10						
42 E	J	CO	HB3	2.16	0.19						
42 E	J	HA	HB2	7.25	0.15						
42 E	J	HA	HB3	5.42	0.15						
42 E	D	CB	HB2	1.58	1.07						
42 E	D	CB	HB3	-2.74	2.14						
43 W	J	N	HB2	1.89	0.10						
43 W	J	CO	HB2	3.12	0.27						
43 W	J	CO	HB3	1.86	0.10						
43 W	J	HA	HB2	6.98	0.15						
43 W	J	HA	HB3	5.64	0.15						
43 W	D	CB	HB2	-4.35	3.42						
43 W	D	CB	HB3	-2.01	1.00						
44 T	J	N	HB	2.10	0.10						
44 T	J	CO	HB	1.40	0.10						
44 T	J	HA	HB	5.35	0.15						
44 T	D	CB	HB	-1.84	1.00						
45 Y	J	N	HB2	1.87	0.10						
45 Y	J	N	HB3	2.92	0.10						
45 Y	J	CO	HB2	3.08	0.10						
45 Y	J	CO	HB3	2.05	0.10						
45 Y	J	HA	HB2	7.32	0.15						
45 Y	J	HA	HB3	6.28	0.15						
45 Y	D	CB	HB2	0.90	4.26						
45 Y	D	CB	HB3	-1.79	1.00						
46 D	J	N	HB2	2.16	0.10						
46 D	J	N	HB3	2.49	0.10						
46 D	J	CO	HB2	3.02	0.10						
46 D	J	CO	HB3	2.64	0.10						
46 D	J	HA	HB2	5.93	0.15						
46 D	J	HA	HB3	6.03	0.15						
46 D	D	CB	HB2	0.07	3.85						
46 D	D	CB	HB3	4.06	3.31						
47 D	J	N	HB2	2.32	0.10						
47 D	J	N	HB3	2.73	0.10						
47 D	J	CO	HB2	2.11	0.21						
47 D	J	CO	HB3	2.49	0.33						
47 D	J	HA	HB2	6.71	0.15						
47 D	J	HA	HB3	5.27	0.15						
47 D	D	CB	HB2	1.96	1.54						

Table S2A: χ_1 rotamer populations $p_{-60^\circ, 60^\circ, 180^\circ}$ and errors $\Delta p_{-60^\circ, 60^\circ, 180^\circ}$ derived by fitting detected $^3J_{\text{NH}\beta}$, $^3J_{\text{C}^{\text{H}}\beta}$, $^3J_{\text{H}\alpha\text{H}\beta}$ to staggered rotamer model for urea-denatured ubiquitin.

Nr.	Res.	N ^a	p_{180°	Δp_{180°	p_{-60°	Δp_{-60°	p_{+60°	Δp_{+60°
3	I	2	0.115	0.000 ^b	0.608	0.000	0.277	0.000
4	F	4	0.230	0.018	0.688	0.016	0.082	0.028
5	V	2	0.535	0.000	0.349	0.000	0.116	0.000
6	K	4	0.276	0.018	0.570	0.016	0.153	0.028
7	T	3	0.116	0.029	0.400	0.018	0.484	0.029
8	L	6	0.132	0.017	0.690	0.016	0.178	0.026
9	T	3	0.124	0.017	0.345	0.018	0.532	0.029
11	K	6	0.252	0.012	0.616	0.014	0.132	0.014
13	I	3	0.116	0.017	0.521	0.033	0.364	0.029
14	T	3	0.191	0.023	0.432	0.018	0.377	0.029
16	E	6	0.276	0.012	0.617	0.014	0.107	0.014
17	V	3	0.573	0.023	0.347	0.066	0.080	0.024
20	S	6	0.364	0.012	0.338	0.017	0.297	0.020
21	D	6	0.228	0.012	0.383	0.014	0.389	0.014
22	T	3	0.118	0.023	0.393	0.018	0.489	0.029
24	E	6	0.227	0.013	0.627	0.015	0.145	0.025
26	V	2	0.474	0.000	0.472	0.000	0.054	0.000
27	K	6	0.227	0.012	0.657	0.014	0.116	0.014
29	K	6	0.277	0.014	0.587	0.018	0.135	0.018
30	I	3	0.106	0.017	0.630	0.018	0.264	0.028
31	Q	6	0.231	0.012	0.628	0.014	0.141	0.014
32	D	6	0.276	0.016	0.446	0.015	0.278	0.023
33	K	6	0.201	0.012	0.652	0.014	0.147	0.014
34	E	6	0.173	0.015	0.646	0.015	0.182	0.023
40	Q	6	0.190	0.012	0.673	0.014	0.137	0.014
41	Q	6	0.238	0.012	0.608	0.014	0.154	0.016
42	R	6	0.238	0.016	0.563	0.015	0.199	0.020
43	L	6	0.212	0.012	0.756	0.014	0.032	0.014
44	I	3	0.000	0.080	0.675	0.025	0.325	0.033
45	F	6	0.291	0.012	0.617	0.014	0.092	0.014
48	K	6	0.233	0.012	0.606	0.014	0.161	0.014
49	Q	4	0.191	0.018	0.608	0.016	0.201	0.028
50	L	4	0.146	0.018	0.765	0.018	0.089	0.033
51	E	6	0.224	0.014	0.646	0.014	0.130	0.014
52	D	6	0.206	0.015	0.442	0.016	0.352	0.027
54	R	6	0.220	0.012	0.604	0.014	0.175	0.015
55	T	2	0.066	0.000	0.453	0.000	0.481	0.000
56	L	4	0.235	0.018	0.654	0.016	0.111	0.028
59	Y	6	0.315	0.014	0.512	0.015	0.174	0.023
60	N	6	0.328	0.012	0.457	0.014	0.216	0.014
62	Q	4	0.199	0.018	0.642	0.016	0.159	0.028
63	K	4	0.209	0.018	0.605	0.019	0.186	0.038
64	E	6	0.218	0.012	0.609	0.014	0.173	0.014
65	S	6	0.385	0.012	0.323	0.015	0.291	0.014
66	T	2	0.119	0.000	0.355	0.000	0.527	0.000
67	L	6	0.274	0.012	0.665	0.019	0.061	0.014
68	H	5	0.259	0.013	0.601	0.016	0.140	0.028
69	L	6	0.230	0.012	0.695	0.014	0.075	0.014
70	V	3	0.667	0.018	0.297	0.029	0.036	0.026
71	L	6	0.215	0.012	0.751	0.014	0.034	0.014
72	R	6	0.278	0.016	0.605	0.027	0.117	0.015
73	L	6	0.148	0.012	0.828	0.014	0.024	0.014
74	R	6	0.239	0.012	0.609	0.014	0.152	0.014

^anumber of detected J-coupling constants.

^ban error of 0.000 is given for situations where not enough ($N < 3$) couplings are detected to carry out a meaningful error propagation for the experimental errors.

Table S2B: χ_1 rotamer populations $p_{-60^\circ, 60^\circ, 180^\circ}$ and errors $\Delta p_{-60^\circ, 60^\circ, 180^\circ}$ derived by fitting detected $^3J_{\text{NH}\beta}$, $^3J_{\text{C}^{\text{H}}\beta}$, $^3J_{\text{H}\alpha\text{H}\beta}$ to staggered rotamer model for urea-denatured protein G.

Nr.	Res.	N ^a	p_{180°	Δp_{180°	p_{-60°	Δp_{-60°	p_{+60°	Δp_{+60°
2	Q	6	0.281	0.012	0.579	0.014	0.141	0.014
3	Y	4	0.232	0.018	0.633	0.016	0.134	0.028
4	K	6	0.287	0.012	0.545	0.015	0.168	0.014
6	I	3	0.092	0.017	0.721	0.025	0.187	0.029
7	L	6	0.171	0.014	0.822	0.017	0.007	0.014
8	N	6	0.241	0.012	0.485	0.014	0.273	0.014
10	K	6	0.207	0.012	0.633	0.014	0.160	0.014
11	T	3	0.101	0.017	0.478	0.018	0.421	0.029
13	K	6	0.262	0.013	0.600	0.014	0.138	0.014
15	E	6	0.212	0.013	0.608	0.015	0.181	0.014
16	T	3	0.155	0.017	0.370	0.018	0.475	0.029
17	T	3	0.246	0.021	0.335	0.018	0.418	0.029
18	T	3	0.171	0.017	0.306	0.018	0.522	0.029
19	E	6	0.236	0.012	0.587	0.014	0.176	0.014
21	V	3	0.475	0.018	0.429	0.028	0.096	0.017
22	D	6	0.282	0.012	0.475	0.014	0.243	0.015
25	T	2	0.306	0.000 ^b	0.196	0.000	0.498	0.000
27	E	6	0.244	0.012	0.598	0.014	0.158	0.014
29	V	3	0.568	0.018	0.365	0.028	0.067	0.017
30	F	6	0.320	0.012	0.575	0.014	0.106	0.014
32	Q	6	0.274	0.012	0.562	0.014	0.164	0.014
33	Y	6	0.180	0.015	0.702	0.014	0.118	0.014
35	N	6	0.288	0.012	0.410	0.015	0.302	0.019
36	D	6	0.214	0.012	0.432	0.014	0.354	0.014
37	N	6	0.213	0.015	0.493	0.015	0.294	0.021
39	V	3	0.455	0.018	0.458	0.028	0.087	0.017
40	D	6	0.221	0.012	0.573	0.014	0.206	0.014
42	E	6	0.236	0.012	0.537	0.015	0.226	0.020
43	W	5	0.336	0.016	0.520	0.017	0.143	0.014
44	T	3	0.135	0.017	0.449	0.018	0.417	0.029
45	Y	6	0.334	0.012	0.527	0.014	0.138	0.014
46	D	6	0.336	0.012	0.390	0.014	0.274	0.014
47	D	6	0.218	0.016	0.451	0.015	0.331	0.025
49	T	2	0.066	0.000	0.415	0.000	0.518	0.000
51	T	3	0.139	0.017	0.443	0.018	0.418	0.029
52	F	6	0.246	0.012	0.548	0.014	0.206	0.014
53	T	3	0.209	0.017	0.404	0.018	0.386	0.029
54	V	3	0.476	0.018	0.416	0.028	0.108	0.017
55	T	2	0.000	0.000	0.506	0.000	0.494	0.000

^anumber of detected J-coupling constants.

^ban error of 0.000 is given for situations where not enough ($N < 3$) couplings are detected to carry out a meaningful error propagation for the experimental errors.

References

- (1) Zhu, G.; Bax, A. *Journal of Magnetic Resonance, Series A* **1993**, *104*, 353-357.
- (2) Grzesiek, S.; Bax, A. *J Biomol NMR* **1993**, *3*, 185-204.
- (3) Ottiger, M.; Delaglio, F.; Bax, A. *Journal of Magnetic Resonance* **1998**, *131*, 373-378.

Chapter 5:

Backbone resonance assignment and homology modeling of the 31 kDa protein dimer of HES1: a transcriptional repressor protein in the Notch signaling pathway

Navratna Vajpai¹, Romel Bobby¹, Alessandro Pintar² and Stephan Grzesiek¹

¹*Biozentrum, University of Basel, Switzerland*

²*International Centre of Genetic Engineering and Biotechnology, Trieste, Italy*

Abstract

The Notch signaling pathway is a conserved intercellular signaling mechanism that is essential for proper embryonic development in numerous metazoan organisms. HES1 acts as an effector of Notch signaling by repressing the expression of target genes that include tissue-specific transcriptional activators. It has a basic helix-loop-helix (bHLH) motif and an Orange domain, and is highly conserved within the respective family.

Here, I report the partial assignment of backbone resonances of the homo-dimer of HES1 in the apo-state. All the unassigned resonances were broadened beyond detection by NMR. Studies based on these assignments are currently in progress. To the best of my knowledge, this is the first structural study performed on any of the later component in the Notch pathway.

Background

The evolutionarily conserved Notch signaling pathway controls cell fate in metazoans through local cell-cell interactions. Specific intercellular contacts activate this highly complex signaling cascade, leading to down-regulation or inhibition of cell-type-specific transcriptional activators. Cells are thus forced to take on a secondary fate or remain undifferentiated while awaiting later inductive signals. Analyses of loss- and gain-of-function mutants of Notch in vertebrates and invertebrates have demonstrated that these repressive Notch functions are remarkably conserved throughout species (Greenwald, 1998, Egan *et al.*, 1998, Artavanis-Tsakonas *et al.*, 1999).

Notch signaling controls an extraordinarily broad spectrum of cell fates and developmental processes in organisms ranging from sea urchins to humans. This resulted in an increasingly large number of Notch-related studies in the past two decades (Miele, 2006). Four different Notch receptors (Notch1-4) and a few ligands, DSL (Delta-like, Serrate, Lag 2), and Jagged-1 (JAG1) and -2 (JAG2) have been characterized in mammalian cells. The Notch receptors and ligands are single-pass transmembrane proteins, and have large extracellular domains that consist primarily of epidermal growth factor (EGF)-like repeats (Wharton *et al.*, 1985). They are expressed in different combinations in most, if not all, cell types. Notch receptors participate in a signaling pathway that regulates many aspects of morphogenesis in multicellular animals through diverse effects on differentiation, proliferation, and cell survival. Notch signaling enables neighboring cells to acquire distinct phenotypes through lateral inhibition (Ronojoy & Claire, 2001).

Overview of Notch signaling

The Notch receptors are pre-cleaved in the Golgi apparatus and are targeted subsequently to the plasma membrane where they interact with transmembrane ligands of the DSL or the JAG class, located on neighboring cells. Notch receptors undergo a complex set of proteolytic processing events in response to the conformational change induced by ligand interaction. The proteolytic activity requires two proteases, namely the ADAM-family metalloproteases (tumor necrosis factor- α converting enzyme or TACE) and γ -secretase, an enzyme complex that contains presenilin, nicastrin, PEN2 and APh1. The first

cleavage, mediated by TACE, releases the Notch extracellular domain (NECD), which continues to interact with the ligand. The ligand–NECD complex is then endocytosed by the ligand-expressing cell. There may be signaling effects in the ligand-expressing cell; this part of Notch signaling is a topic of active research. The second cleavage, mediated by the γ -secretase complex, enables the Notch intracellular domain (NICD) to translocate to the nucleus to activate Notch target genes (Figure 5.1) (Weinmaster, 1998, Mumm *et al.*, 2000). Thus, inhibiting γ -secretase function can prevent the final cleavage of the Notch receptor, blocking Notch signal transduction.

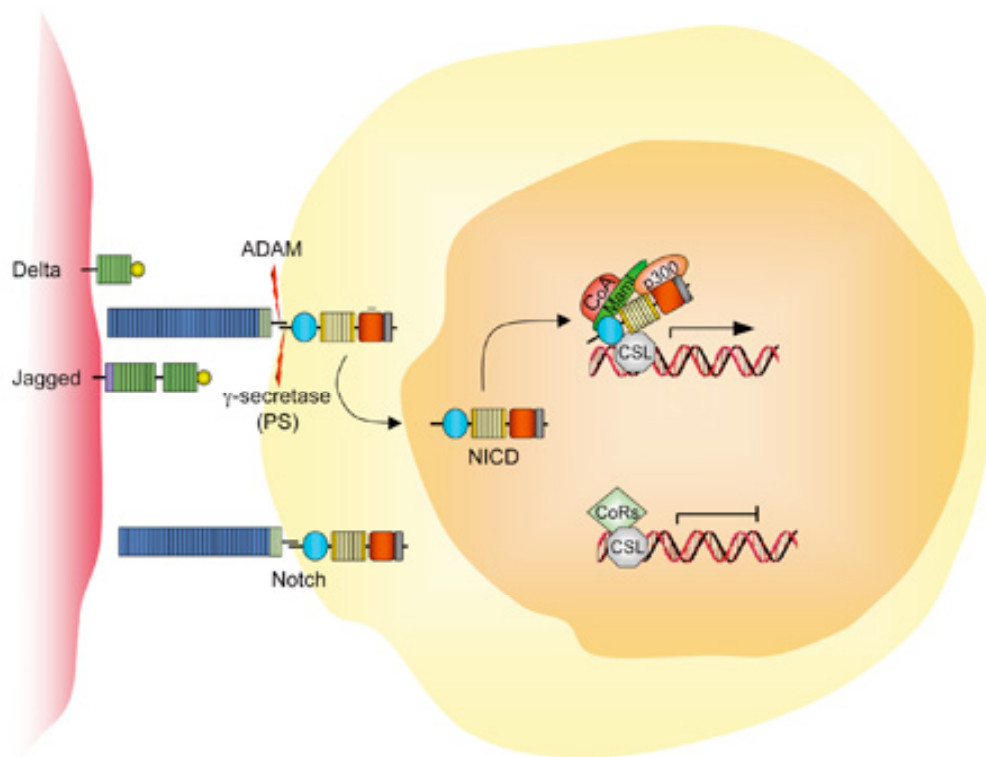


Figure 5.1: Mechanism of Notch signaling. Notch signaling is triggered by ligand-receptor interaction, which induces two sequential proteolytic cleavages: the first one in the extracellular domain mediated by ADAM proteases, and the second one within the transmembrane domain mediated by γ -secretase. The NICD is released after the second cleavage and translocates into the nucleus to activate Notch target genes (Taken from http://www.isrec.ch/research/groups/research_groups_detail_eid_3263_lid_2.htm; group of Prof. Radtke Freddy)

In the absence of NICD cleavage, transcription of Notch target genes in the nucleus is inhibited by a repressor complex mediated by a protein of the RBP-J κ family (also known as CSL for CBF1, Suppressor of Hairless, and Lag-1) (Weinmaster, 1998, Greenwald, 1998, Egan *et al.*, 1998, Artavanis-Tsakonas *et al.*, 1999, Mumm *et al.*, 2000). NICD has a transcriptional activation domain, but no DNA binding domain of its own. When NICD enters the nucleus, it disrupts the repressor complex, binds to RBP-J κ , and activates transcription from its DNA binding site (Ling *et al.*, 1994, Kao *et al.*, 1998). The NICD-RBP-J κ complex ultimately leads to conversion of RBP-J κ protein from transcription repressor to transcriptional activator. This complex subsequently up-regulates expression of primary target genes of Notch signaling, such as HESs (for Hairy/Enhancer of Split genes) family and HEYs family (Hesr/Hey family). While the Notch signaling pathway is deceptively simple, the consequences of Notch activation on cell fate are complex and context-dependent. The manner in which other signaling pathways cross-talk with Notch signaling appears to be particularly complicated.

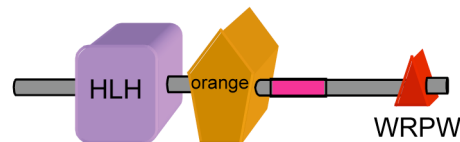
The Notch pathway plays crucial roles in the development of most organs. Mutations of receptors and ligands in Notch pathway lead to abnormalities in many tissues, including vessels, thymus, craniofacial region, limb, central nervous system, heart, kidney as well as hematopoietic cells (Swiatek *et al.*, 1994, Conlon *et al.*, 1995, de la Pompa *et al.*, 1997, Hrabe de Angelis *et al.*, 1997, Sidow *et al.*, 1997, Xue *et al.*, 1999, Krebs *et al.*, 2000, McCright *et al.*, 2001, Dunwoodie *et al.*, 2002). Disruption of Notch has been implicated in multiple tumor types. Evidence from in vitro experiments, mouse models and human tumor samples indicates that Notch plays a predominantly oncogenic role in breast cancer and interacts with other pathways involved in tumorigenesis. In addition, Notch signaling is required for physiological angiogenesis and may promote tumor angiogenesis. A variety of strategies for blocking Notch signaling, in particular γ -secretase inhibition, are being considered as potential therapies for breast cancer and tumor angiogenesis (Kopan & Goate, 2000, Curry *et al.*, 2005, van Es *et al.*, 2005).

In the current study, the focus is on the characterization of structure and function of one of the later component in the Notch pathway, namely the Notch effector, HES1.

Overview of HES/E(spl) family:

HES (in mammals) and its homologues *hairy* and E(spl) in *Drosophila* proteins contain a basic domain, which determines DNA binding specificity, and a helix-loop-helix domain, which allows the proteins to form homo- or hetero-dimers (Murre *et al.*, 1989, Blackwell *et al.*, 1993, Ferre-D'Amare *et al.*, 1993). Dimers of HES/E(spl) suppress expression of downstream target genes such as tissue-specific transcriptional activators *e.g.* MASH1 (mammalian *achaete-scute* homolog 1) and neurogenin (Ohsako *et al.*, 1994, Ishibashi *et al.*, 1995, Chen *et al.*, 1997). HES/E(spl) has three unique, evolutionary conserved features: (i) an invariant proline at a specific position within the DNA-binding basic domain, (ii) an Orange domain in corresponding regions carboxy-terminus to bHLH region, and (iii) a carboxy-terminal tetrapeptide WRPW motif (Figure 5.2).

A)



B)

```

1      10      20      30      40      50      60      70
sp|Q14469|HES1_HUMAN  KPKTASEHRKSSKPIMEKRRRARINESLSQLKTLILDALKKDDSSRHSKLEKADILEMTVKHLRNLQRAQM
sp|P35428|HES1_MOUSE  KPKTASEHRKSSKPIMEKRRRARINESLSQLKTLILDALKKDDSSRHSKLEKADILEMTVKHLRNLQRAQM
sp|Q04666|HES1_RAT    KPKTASEHRKSSKPIMEKRRRARINESLSQLKTLILDALKKDDSSRHSKLEKADILEMTVKHLRNLQRAQM
sp|Q3ZBG4|HES1_BOVIN  KPKTASEHRKSSKPIMEKRRRARINESLSQLKTLILDALKKDDSSRHSKLEKADILEMTVKHLRNLQRAQM
sp|O57337|HES1_CHICK  KPRSASEHRKSSKPIMEKRRRARINESLSQLKMLILDALKKDDSSRHSKLEKADILEMTVKHLRNLQRAQM
consensus>50        KPktASEHRKSSKPIMEKRRRARINESLSQlKtLILDALKKDDSSRHskLEKADILEMTvKHLrNLQRAQM

80      90      100     110     120     130
sp|Q14469|HES1_HUMAN  TAALSIDPSVLGKYRAGFSECMNEVTRFLSTCEGVNTEVRRRLLGHLANCMTQINAMTYPGQ
sp|P35428|HES1_MOUSE  TAALSIDPSVLGKYRAGFSECMNEVTRFLSTCEGVNTEVRRRLLGHLANCMTQINAMTYPGQ
sp|Q04666|HES1_RAT    TAALSIDPSVLGKYRAGFSECMNEVTRFLSTCEGVNTEVRRRLLGHLANCMTQINAMTYPGQ
sp|Q3ZBG4|HES1_BOVIN  TAALSIDPSVLGKYRAGFSECMNEVTRFLSTCEGVNTEVRRRLLGHLANCMTQINAMTYPGQ
sp|O57337|HES1_CHICK  TAALSADPSVLGKYRAGFSECMNEVTRFLSTCEGVNADVRRRLLGHLsAClGQIVAMNylpP
consensus>50        tAALstDPSVLGKYRAGfSECMNEVTRfLSTCEGVnt#vRtRLLGHLanC$tQInAmTyPpgQ

```

Figure 5.2: A) Modular structure of human HES1. B) Sequence alignment of the HES1 construct used in our study, which include basic helix-loop-helix (bHLH) domain and the Orange domain. High conservation of residues can be seen for different species.

Mechanisms of transcriptional repression by HES/E(spl) have been extensively studied. The three mechanisms that have been proposed for transcriptional repression by HES/E(spl) are as follows:

The first mechanism is DNA-binding-dependent transcriptional repression, also known as active repression (Kageyama & Nakanishi, 1997, Kageyama, 2000). HES proteins form a homodimer and bind class C (CACGNG) or N (CACNAG) consensus DNA sites (Sasai *et al.*, 1992, Tietze *et al.*, 1992, Oellers *et al.*, 1994, Van Doren *et al.*, 1994). They recruit the corepressor TLE (for transducin-like enhancer of split in mammals) or its *Drosophila melanogaster* homologue Groucho via the C-terminal WRPW motif (Paroush *et al.*, 1994, Fisher *et al.*, 1996, Grbavec & Stifani, 1996). These studies have shown that WRPW is both necessary and sufficient to confer repression when expressed as a fusion protein with a heterologous DNA binding domain of Gal4.

The second mechanism is a passive mechanism (Sasai *et al.*, 1992, Hirata *et al.*, 2000) involving protein sequestration. For example, HES1 can form a non-functional heterodimer with other bHLH factors such as E47, a common heterodimer partner of tissue-specific bHLH proteins like MyoD and MASH1. This disrupts the formation of functional heterodimers (*e.g.* MyoD-E47 and MASH1-E47).

The third mechanism is mediated by the Orange domain, helix-3 and helix-4 in HES1 (Castella *et al.*, 2000). The Orange domain is essential to repress transcription of its own (HES1) promoter as well as p21^{WAF} promoter (Castella *et al.*, 2000). This ability of Orange domain is dependent on the presence of the DNA-binding bHLH domain. An important role of Orange domain has been demonstrated in a sex determination assay in *Drosophila* (Dawson *et al.*, 1995). Castella and colleagues proposed that the Orange domain is necessary for either the direct recruitment of an unknown corepressor and/or stabilization of the WRPW-mediated repression function through intra- or intermolecular interaction (Castella *et al.*, 2000).

DNA binding site specificity and target genes for HES

While the *hairy* and E(spl) proteins in *Drosophila* have been reported to bind mainly class B core site (CACGTG) by in vitro random oligonucleotide binding site selection, a

weaker binding has also been seen for N box (CACGAG) and class C site (CACGCG) (Jennings *et al.*, 1999). Further, in vivo significance of binding class B site have been confirmed by the observation that even subtle changes within class B core or flanking bases have dramatic consequences for lacZ reporter gene expression in transgenic flies (Jennings *et al.*, 1999). Mammalian homologue HES proteins have been shown to bind N box (Akazawa *et al.*, 1992, Sasai *et al.*, 1992, Ishibashi *et al.*, 1995, Hirata *et al.*, 2000).

Several target genes have been proposed for HES, but in vivo target genes have been established mainly for HES1 (Akazawa *et al.*, 1992, Sasai *et al.*, 1992, Ishibashi *et al.*, 1995, Hirata *et al.*, 2000). HES1 negatively regulates its own promoter activity, which was confirmed by the observation that mutation in N box sequences in the HES1 promoter diminished the negative auto-regulation (Takebayashi *et al.*, 1994). Overexpressed HES1 functions as negative regulator of neurogenesis by directly repressing the proneural gene MASH1 (Ishibashi *et al.*, 1995). Overexpression of HES1 also leads to N-box-dependent repression of the CD4 promoter as well as downregulation of endogenous CD4 expression in CD4⁺CD8⁻ T_H cells (Kim & Siu, 1998). The acid α -glucosidase promoter is also repressed by HES1 in a class C site (CACGCG)-dependent manner in hepatoma-derived Hep G2 cells (Yan *et al.*, 2001). Collectively, these data indicate that class C sites and N boxes are likely to be critical in vivo binding sites for HES1 in mammals.

Structural studies of HES1

Until now, structural studies have been restricted to the Notch receptors and their ligands (Iakoucheva *et al.*, 2002, Zweifel *et al.*, 2003, Popovic *et al.*, 2006, Pintar *et al.*, 2007, Kelly *et al.*, 2007), whereas no structural studies have been reported on the later components of Notch pathway, such as the Notch effectors, HES/E(spl). Such information is important if we are to understand the molecular mechanisms that govern the transcriptional repression by HES/E(spl).

NMR is a widely established technique for characterizing proteins at atomic level. Therefore, it was sought to characterize the structure and activity of HES1 by solution NMR studies. The main focus of the study is to understand 1) its DNA binding mechanism, and 2) what confers the specificity of function to different domains in HES1.

To enable such studies, backbone resonance assignment of HES1 (residues M27-Q158) was performed using state-of-the-art heteronuclear NMR experiments. Partial backbone resonance assignments have been achieved. The unassigned resonances broadened beyond detection by NMR. These assignments will aid in the characterization of different domains of HES1. Work in this direction is under way.

Materials and Methods

Uniformly ^{15}N - and $^{13}\text{C}/^{15}\text{N}$ -labeled HES1 was obtained from the collaborators in the group of Dr. Alessandro Pintar at International Centre of Genetic Engineering and Biotechnology, Trieste, Italy. The construct (M27-Q158; molecular weight 15.5 kDa) used in the studies includes the bHLH domain and the Orange domain.

NMR samples and experiments:

Uniformly ^{15}N - and $^{13}\text{C}/^{15}\text{N}$ -labeled HES1 samples were prepared as 0.65 and 0.4 mM solutions in 280 μl of 95% H_2O and 5% D_2O , 25 mM potassium phosphate buffer and 10 mM TCEP at pH 5.85, respectively. NMR spectra were recorded at 303 K on Bruker DRX 800 MHz equipped with a TCI cryoprobe. Backbone assignment followed standard triple-resonance strategies, which includes HNCO, HNCA, HN(CO)CA, ^{15}N -edited ^1H - ^1H NOESY and CBCA-type experiments (Grzesiek & Bax, 1993). All NMR data were processed using the NMRPipe suite of programs (Delaglio *et al.*, 1995) and analyzed with NMRView (Johnson & Blevins, 1994).

NMR relaxation experiments and analysis: ^{15}N T_1 and T_2 relaxation measurements were performed on the uniformly ^{15}N -labeled HES1 sample at 600 MHz equipped with TXI probe. Analysis was done for a few representative peaks along the amino acid sequence only to estimate the correlation time of the protein. T_1 and T_2 decay curves were fitted by an in-house written routine implemented in Matlab (MathWorks, Inc.) using a simplex search minimization and Monte Carlo estimation of errors. Lipari-Szabo model-free analysis of ^{15}N relaxation data was performed using the TENSOR2 program (Dosset *et al.*, 2000).

Results

Backbone resonance assignment of HES1

The achieved assignments comprise partial (~75 %) backbone $^1\text{H}^{\text{N}}$, ^{15}N , ^{13}CO and $^{13}\text{C}^{\alpha}$ and (~40 %) $^{13}\text{C}^{\beta}$ assignments for the homo-dimer of HES1 in its apo-state (Figure 5.3). The resonance assignments have been performed on protonated samples with non-TROSY versions of backbone CA- and CBCA-type experiments, and an ^{15}N -edited ^1H - ^1H NOESY spectrum. For CBCA-type experiments, which would have yielded highly distinctive amino-acid type information, only 40% of all expected resonances could be observed. Missing assignments mainly cluster in the bHLH domain (Figure 5.4B). These resonances were not visible in the NMR spectra; presumably these residues are mobile and NMR signals broadened beyond detection by intermediate exchange (*i.e.* μs to ms -dynamics). HES1 is known to undergo dimerization, thus, exchange contribution to line broadening may be attributed to monomer-dimer equilibrium.

Preliminary qualitative analysis of ^{15}N relaxation data and the following TENSOR2 analysis confirms that HES1 exists as dimer in solution (data not shown). Possibly, this led to slow tumbling of the molecule and consequently, the short relaxation time of the observable nuclei.

For the construct of HES1 used in the studies, secondary structure predictions with PSIPRED (V2.6; David Jones) show four large helices and one short helix. While, helices 1 and 2, and the short helix were predicted in the bHLH domain, helices 3 and 4 were predicted in the Orange domain (Figure 5.4A). Secondary chemical shifts observed by NMR agree well with the predicted secondary structure pattern. These results indicate stronger helices in the Orange domain (Figure 5.4B) compared to those present in bHLH domain. The break in the sequential assignments, for *e.g.* amino acid stretch of R47-L58 in the helix-1 and residues preceding T85 in the helix-2 might be because of highly flexible helices in the bHLH domain (Figure 5.4B). In contrast, the missing assignments in the helix-3 and -4 are mainly due to signal overlap resulting in the ambiguity in the resonance assignments. This ambiguity could not be resolved by the combination of HNCA, HN(CO)CA and ^{15}N -edited NOESY.

This backbone resonance assignment is a prerequisite for the future studies on HES1. Currently, these assignments are being used to understand the DNA binding mechanism and monomer-dimer equilibrium.

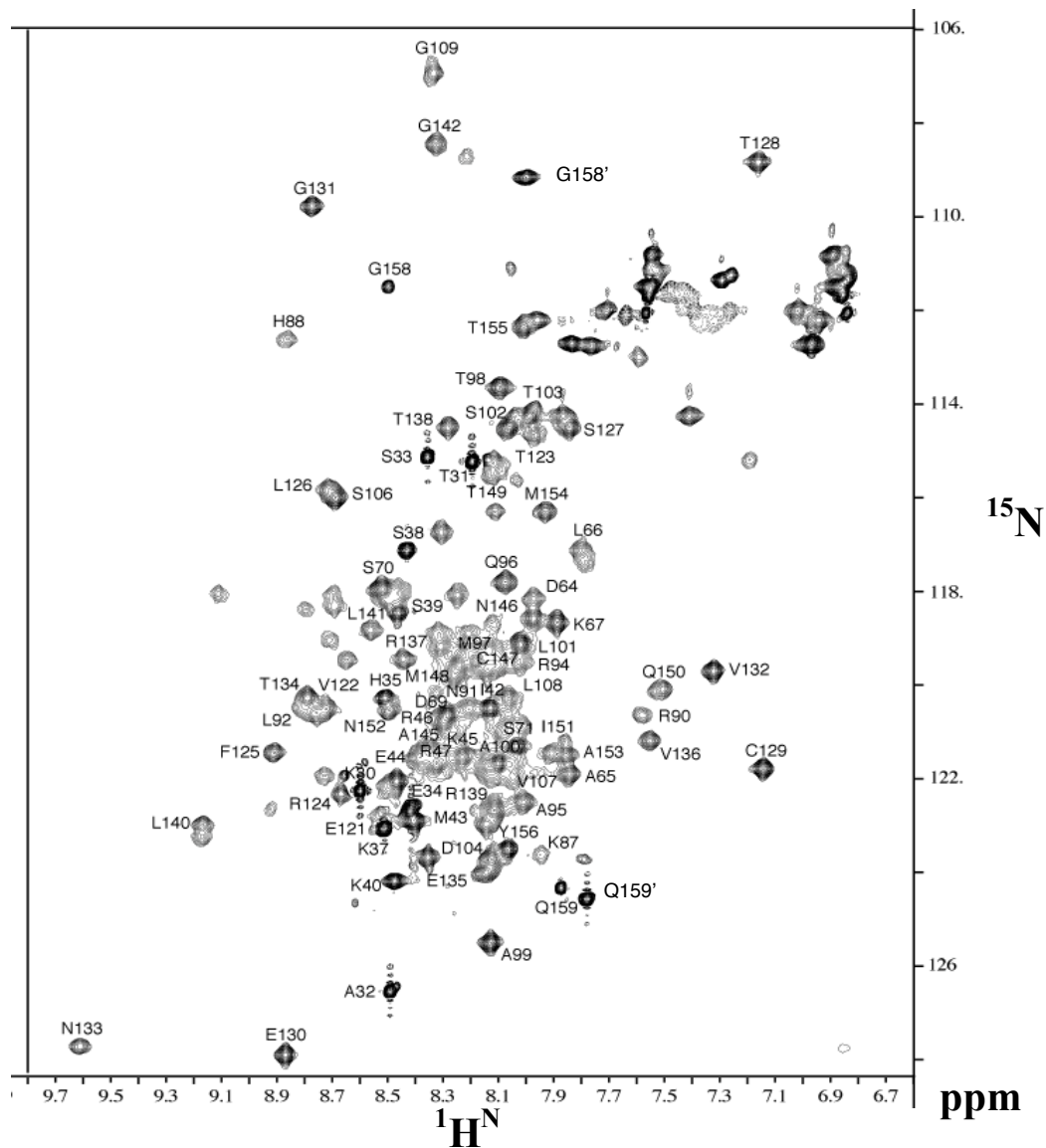


Figure 5.3: ^{15}N -HSQC spectrum of homodimer of apo-HES1 for the construct M27-Q158 with assignment of resonances. The unassigned resonances are broadened beyond detection in the 3D experiments.

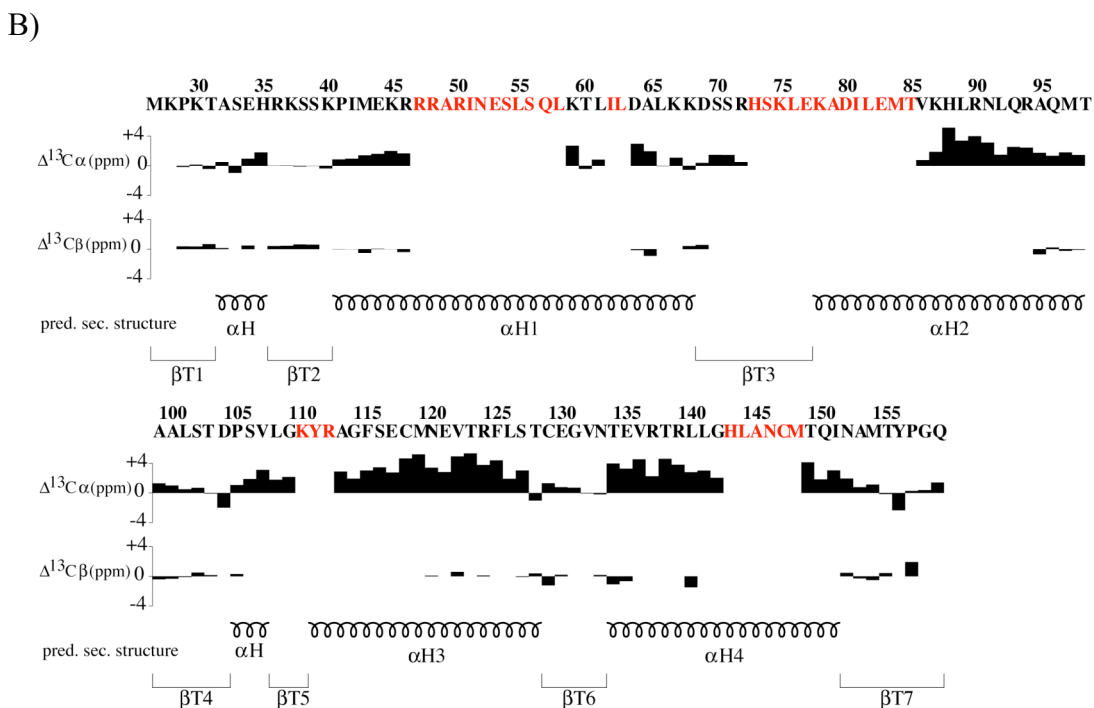
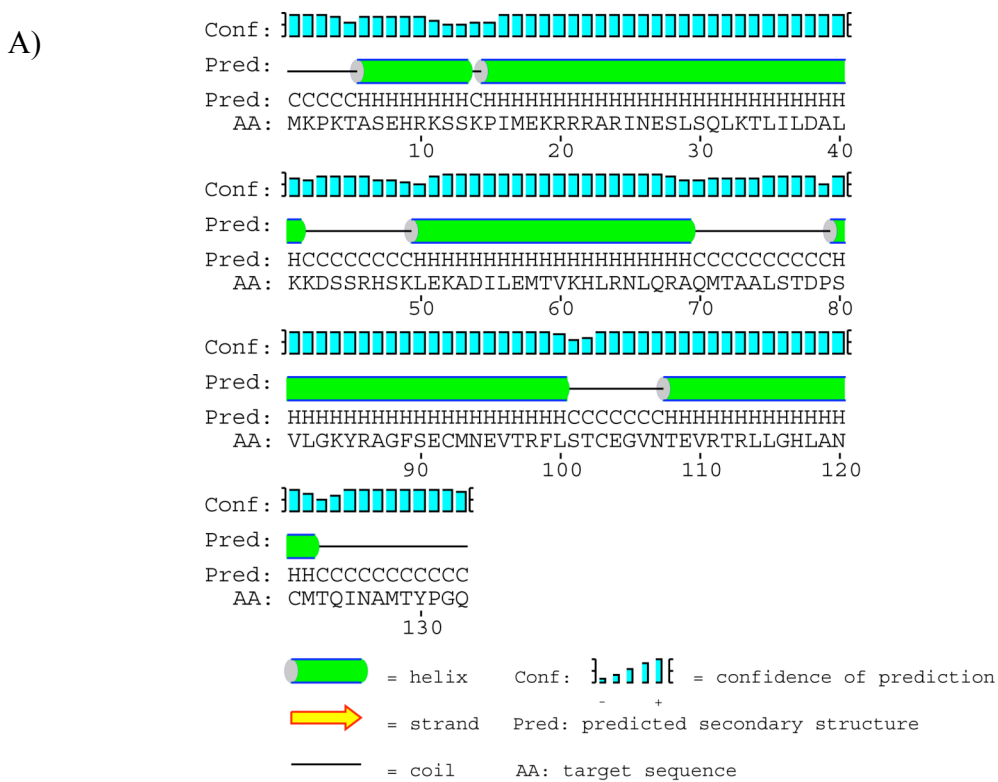


Figure 5.4: A) Secondary structure predicted using PSIPRED (V2.6; David Jones). B) Secondary chemical shift and predicted secondary structure match in the HES1. Amino acids lacking backbone resonance assignments are shown in red. Random coil shifts are taken from (Spera & Bax, 1991).

Homology modeling of the Orange domain of HES1

(Together with R. Bobby)

Alongside NMR experiments, homology modeling was performed with the Swiss Modeling Server (An Automated Comparative Protein Modeling Server; <http://swissmodel.expasy.org/SWISS-MODEL.html>) to produce a model structure of HES1 Orange domain. The obtained structure has high degree of agreement (r.m.s.d for C α atoms 0.431 Å; sequence homology of 44%) with the 1.9 Å crystal structure (PDB entry: 2DB7) of a hypothetical protein, MS0332 described in the PDB. The crystal structure of 2DB7 was reported as dimer, and it corresponds to the Orange domain in a HES1 related protein called Hairy/E(spl) with conserved YRPW motif 1. Structural analysis of 2DB7 confirmed that the dimerization was due to Cys22 that forms a disulfide bridge between the two monomers. This feature could only be an artifact since the protein inside the cell is under highly reducing conditions, and therefore, cannot form disulfide bridges. Nevertheless, there were also indications of inter-monomer salt bridges (e.g. E34 – R45) that could be responsible for stabilizing the dimer (Figure 5.5). Thus, the modeling studies suggest that Orange domain of HES1 exist as dimer in solution; possibly, it mediates the dimerization of HES1 to undergo DNA binding.

Further investigations on the HES1 are currently in the process.

A)

```

                                1      10      20
Q14469 | HES1_HUMAN           .....LGKYRAGFSECMNEVTRFLSTCEGVNT..
2DB7_A | PDBID | CHAIN | SEQUENCE SGGYFDAHALAMDYRSLGFRERCLAEVARYLSIIIEGLDASD

                                30      40
Q14469 | HES1_HUMAN           EVRTRLGHLANCMTOI..
2DB7_A | PDBID | CHAIN | SEQUENCE PLRVRLVSHLNNYASCREA

```

B)

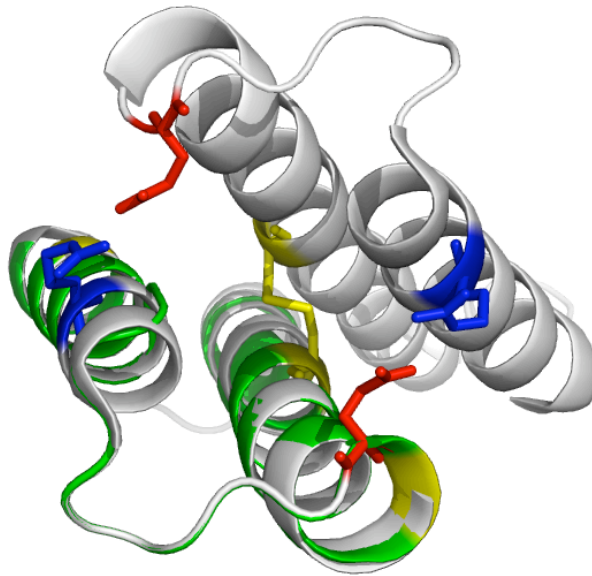


Figure 5.5: A) Sequence alignment of the Orange domain of HES1 (Homo sapiens) to the Orange domain of the crystal structure (PDB entry: 2DB7). Cys22, which forms a disulphide bridge, is conserved in HES1 (Cys11). Residues marked in red boxes are conserved in two sequences.

B) Homology model of the Orange domain of HES1 (green) aligned to the crystal structure (2DB7) of homologous Orange domain of Hairy/E(spl) (grey). Clearly, crystal structure undergoes dimerization due to disulphide bond shown (yellow). Close proximity between E34 (red) and R45 (blue) indicates inter-monomer electrostatic interaction.

Chapter 6:

Conclusions and perspectives

In this thesis, residual dipolar couplings, measured under conditions of partial molecular alignment, in combination with other high-resolution solution NMR techniques have been used to characterize folded and unfolded states of polypeptides. Following are the conclusions and context for each chapter of my thesis:

Solution NMR studies of ABL kinase in complex with three clinical inhibitors

Chapter two demonstrated characterization of the structure and dynamics of ABL kinase complexes under solution conditions. The findings of this study showed that in solution, the activation loop for imatinib and nilotinib complexes adopts the inactive conformation, whereas the dasatinib complex preserves the active conformation, contrary to the predictions based on molecular modeling. This study also indicates more conformational plasticity in the active state of the kinase complex compared to the inactive state complexes. The results of this thesis enhance our understanding of multiple inactive conformations observed in some kinases and may also shed light on how point mutations in BCR-ABL lead to drug resistance, and enable the rational design of more potent inhibitors.

Until now, structural studies reported on ABL kinase have all been performed in the presence of inhibitors and information regarding biologically more important apo state is missing. X-ray studies based on these kinase-inhibitor complexes proposed that in solution, the apo state of the kinase exists in dynamic equilibrium between open and closed conformations of the activation loop (Nagar *et al.*, 2002). However, to date, experimental data to support this hypothesis is missing. The obstacle to performing such a study is the production of ABL in its apo-state due to its poor solubility. This work is currently in progress in collaboration with Novartis Pharma, Basel. It is hoped that the findings of this thesis open a possibility for the experimental characterization of the apo state of the kinase in both phosphorylated and non-phosphorylated forms. A comprehensive understanding of the apo-state may allow rational modifications to the inhibitors in the hopes of producing drugs for improved leukemia therapy.

Conformational studies of unstructured polypeptides by residual dipolar couplings

Chapter three showed investigation of conformational preferences of individual amino acids as monitored by backbone $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs. Our results showed that the presence of larger or aromatic side-chains causes stiffness or a kink in the polypeptide backbone.

In **section 3.1**, the experimental data was compared with the predictions according to the statistical coil model (Bernado *et al.*, 2005). Overall, the statistical coil model failed to reproduce the experimental data to a high degree of agreement, especially for the $^1D_{C\alpha H\alpha}$ RDCs. This implies that there are genuine differences between the ensemble of conformations in the gel-oriented peptide solution and the statistical coil ensemble derived from folded proteins. Improvement in the modeling of side-chains in the statistical model may improve the reproducibility of the simulated data.

In **section 3.2**, the experimental data was compared to an ensemble averaged RDCs obtained from all-atom molecular dynamics simulations. A poor reproducibility of the experimental data indicates deficiencies in the MD simulations. The non-reproducibility of the predicted RDCs for repeated MD simulations indicates inadequate sampling of protein conformational space in the time used for the simulations. Employing longer sampling time for MD simulations may improve the reproducibility of the simulated data.

In summary, this study showed the possibility of a rigorous experimental characterization of individual amino acid/amino acid interactions in unfolded polypeptides.

Side-chains conformations in urea-denatured proteins: a study by 3J scalar couplings and residual dipolar couplings

Chapter four provided a detailed analysis of side-chain conformations in two urea-denatured proteins, ubiquitin and protein G. The two observables used were 3J scalar couplings and residual dipolar couplings involving the two $^1H^{\beta}$ atoms. The presented data clearly show that for most residues, the precision of individual χ_1 rotamer populations is better than 2 % and are in vicinity of predictions obtained from the protein coil library, which is a database containing fragments of protein data bank (PDB) structures which cannot be either classified as α -helix or β -strand. However, individual variations from

these averages of up to 40 % are highly significant and indicate sequence- and residue-specific interactions. Independent analysis of $^1D_{C\beta H\beta}$ RDCs obtained in polyacrylamide gels show good correlation with the RDCs predicted from the χ_1 populations obtained from the 3J data and a coil model ensemble of 50000 conformers according to the coil library backbone angle distribution. The presented study improved our understanding of unfolded states of polypeptides and opened possibilities for the rigorous characterization of side-chains in the unfolded proteins.

Backbone resonance assignment of the 31 kDa of homodimer of apo-HES1

Chapter five reported the chemical shift assignments and preliminary homology modeling results of HES1, an effector in the Notch signaling pathway. This pathway is highly conserved in all metazoans. Currently, the obtained partial backbone assignments are used to characterize different domains in the HES1 as well as the monomer-dimer equilibrium. It is hoped that the structural studies on HES1 will improve our understanding of molecular repression by HES1.

Bibliography

- Abragam, A., (1983) *The Principles of Nuclear Magnetism*. Clarendon Press.
- Ackerman, M. S. & D. Shortle, (2002a) Molecular alignment of denatured states of staphylococcal nuclease with strained polyacrylamide gels and surfactant liquid crystalline phases. *Biochemistry* **41**: 3089-3095.
- Ackerman, M. S. & D. Shortle, (2002b) Robustness of the long-range structure in denatured staphylococcal nuclease to changes in amino acid sequence. *Biochemistry* **41**: 13791-13797.
- Adrian, F. J., Q. Ding, T. Sim, A. Velentza, C. Sloan, Y. Liu, G. Zhang, W. Hur, S. Ding, P. Manley, J. Mestan, D. Fabbro & N. S. Gray, (2006) Allosteric inhibitors of Bcr-abl-dependent cell proliferation. *Nat Chem Biol* **2**: 95-102.
- Akasaka, K., (2003) Highly fluctuating protein structures revealed by variable-pressure nuclear magnetic resonance. *Biochemistry* **42**: 10875-10885.
- Akazawa, C., Y. Sasai, S. Nakanishi & R. Kageyama, (1992) Molecular characterization of a rat negative regulator with a basic helix-loop-helix structure predominantly expressed in the developing nervous system. *J Biol Chem* **267**: 21879-21885.
- Al-Hashimi, H. M. & D. J. Patel, (2002) Residual dipolar couplings: synergy between NMR and structural genomics. *J Biomol NMR* **22**: 1-8.
- Alexandrescu, A. T. & R. A. Kammerer, (2003) Structure and disorder in the ribonuclease S-peptide probed by NMR residual dipolar couplings. *Protein Sci* **12**: 2132-2140.
- Alexandrescu, A. T. & D. Shortle, (1994) Backbone dynamics of a highly disordered 131 residue fragment of staphylococcal nuclease. *J Mol Biol* **242**: 527-546.
- Artavanis-Tsakonas, S., M. D. Rand & R. J. Lake, (1999) Notch signaling: cell fate control and signal integration in development. *Science* **284**: 770-776.
- Atreya, H. S. & T. Szyperski, (2005) Rapid NMR data collection. *Methods Enzymol* **394**: 78-108.
- Balbach, J. J., A. T. Petkova, N. A. Oyler, O. N. Antzutkin, D. J. Gordon, S. C. Meredith & R. Tycko, (2002) Supramolecular structure in full-length Alzheimer's beta-amyloid fibrils: evidence for a parallel beta-sheet organization from solid-state nuclear magnetic resonance. *Biophys J* **83**: 1205-1216.
- Ban, N., P. Nissen, J. Hansen, P. B. Moore & T. A. Steitz, (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**: 905-920.
- Baum, J., C. M. Dobson, P. A. Evans & C. Hanley, (1989) Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig alpha-lactalbumin. *Biochemistry* **28**: 7-13.
- Baumeister, W., (2002) Electron tomography: towards visualizing the molecular organization of the cytoplasm. *Curr Opin Struct Biol* **12**: 679-684.
- Bax, A., (1994) Multidimensional nuclear magnetic resonance methods for protein studies. *Curr. Opin. Struct. Biol.* **4**: 738-744.
- Bax, A., (2003) Weak alignment offers new NMR opportunities to study protein structure and dynamics. *Protein Sci* **12**: 1-16.
- Beraud, S., B. Bersch, B. Brutscher, P. Gans, F. Barras & M. Blackledge, (2002) Direct structure determination using residual dipolar couplings: reaction-site conformation of methionine sulfoxide reductase in solution. *J Am Chem Soc* **124**: 13709-13715.
- Bernado, P., L. Blanchard, P. Timmins, D. Marion, R. W. Ruigrok & M. Blackledge, (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**: 17002-17007.

- Bertoncini, C. W., C. O. Fernandez, C. Griesinger, T. M. Jovin & M. Zweckstetter, (2005) Familial Mutants of $\{\alpha\}$ -Synuclein with Increased Neurotoxicity Have a Destabilized Conformation. *J. Biol. Chem.* **280**: 30649-30652.
- Blackwell, T. K., J. Huang, A. Ma, L. Kretzner, F. W. Alt, R. N. Eisenman & H. Weintraub, (1993) Binding of myc proteins to canonical and noncanonical DNA sequences. *Mol Cell Biol* **13**: 5216-5224.
- Bothner-By, A., P. Domaille, C. Gayathri & (1981) Ultra-high field NMR spectroscopy: observation of proton-proton dipolar coupling in paramagnetic bis[tolyltris(pyrazolyl)borato]cobalt(II). *J. Am. Chem. Soc.* **103**: 5602-5603.
- Bouvignies, G., S. Meier, S. Gzesiek & M. Blackledge, (2006) Ultrahigh-Resolution Backbone Structure of Perdeuterated Protein GB1 Using Residual Dipolar Couplings from Two Alignment Media. *Angewandte Chemie* **118**: 8346-8349.
- Bundi, A. & K. Wuthrich, (1979) ^1H -nmr parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. *Biopolymers* **18**: 285-297.
- Card, P. B., P. J. A. Erbel & K. H. Gardner, (2005) Structural Basis of ARNT PAS-B Dimerization: Use of a Common Beta-sheet Interface for Hetero- and Homodimerization. *Journal of Molecular Biology* **353**: 664-677.
- Castella, P., S. Sawai, K. Nakao, J. A. Wagner & M. Caudy, (2000) HES-1 repression of differentiation and proliferation in PC12 cells: role for the helix 3-helix 4 domain in transcription repression. *Mol Cell Biol* **20**: 6170-6183.
- Chase, A. & N. C. Cross, (2006) Signal transduction therapy in haematological malignancies: identification and targeting of tyrosine kinases. *Clin Sci (Lond)* **111**: 233-249.
- Chen, H., A. Thiagalingam, H. Chopra, M. W. Borges, J. N. Feder, B. D. Nelkin, S. B. Baylin & D. W. Ball, (1997) Conservation of the Drosophila lateral inhibition pathway in human lung cancer: a hairy-related protein (HES-1) directly represses achaete-scute homolog-1 expression. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 5355-5360.
- Chen, J., V. A. Mandelshtam & A. J. Shaka, (2000) Regularization of the two-dimensional filter diagonalization method: FDM2K. *J Magn Reson* **146**: 363-368.
- Chou, J. J., S. Gaemers, B. Howder, J. M. Louis & A. Bax, (2001) A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles. *J Biomol NMR* **21**: 377-382.
- Choy, W. Y., D. Shortle & L. E. Kay, (2003) Side chain dynamics in unfolded protein states: an NMR based ^2H spin relaxation study of $\Delta 131\Delta$. *J Am Chem Soc* **125**: 1748-1758.
- Cierpicki, T. & J. H. Bushweller, (2004) Charged gels as orienting media for measurement of residual dipolar couplings in soluble and integral membrane proteins. *J Am Chem Soc* **126**: 16259-16266.
- Clore, G. M., M. R. Starich & A. M. Gronenborn, (1998) Measurement of Residual Dipolar Couplings of Macromolecules Aligned in the Nematic Phase of a Colloidal Suspension of Rod-Shaped Viruses. *J. Am. Chem. Soc.* **120**: 10571-10572.
- Coffin, J. M., S. H. Hughes & H. Varmus, (1999) *Retroviruses*. CSHL Press, 1997.
- Cohen, G. B., R. Ren & D. Baltimore, (1995) Modular binding domains in signal transduction proteins. *Cell* **80**: 237-248.
- Collett, M. S., A. F. Purchio & R. L. Erikson, (1980) Avian sarcoma virus-transforming protein, pp60src shows protein kinase activity specific for tyrosine. *Nature* **285**: 167-169.
- Conlon, R. A., A. G. Reaume & J. Rossant, (1995) Notch1 is required for the coordinate segmentation of somites. *Development* **121**: 1533-1545.
- Cowan-Jacob, S. W., V. Guez, G. Fendrich, J. D. Griffin, D. Fabbro, P. Furet, J. Liebetanz, J. Mestan & P. W. Manley, (2004) Imatinib (STI571) resistance in

- chronic myelogenous leukemia: molecular basis of the underlying mechanisms and potential strategies for treatment. *Mini Rev Med Chem* **4**: 285-299.
- Curry, C. L., L. L. Reed, T. E. Golde, L. Miele, B. J. Nickoloff & K. E. Foreman, (2005) Gamma secretase inhibitor blocks Notch activation and induces apoptosis in Kaposi's sarcoma tumor cells. *Oncogene* **24**: 6333-6344.
- Cutting, B., A. Strauss, G. Fendrich, P. W. Manley & W. Jahnke, (2004) NMR resonance assignment of selectively labeled proteins by the use of paramagnetic ligands. *J Biomol NMR* **30**: 205-210.
- Dames, S. A., R. Aregger, N. Vajpai, P. Bernado, M. Blackledge & S. Grzesiek, (2006) Residual dipolar couplings in short peptides reveal systematic conformational preferences of individual amino acids. *J Am Chem Soc* **128**: 13508-13514.
- Dawson, S. R., D. L. Turner, H. Weintraub & S. M. Parkhurst, (1995) Specificity for the hairy/enhancer of split basic helix-loop-helix (bHLH) proteins maps outside the bHLH domain and suggests two separable modes of transcriptional repression. *Mol Cell Biol* **15**: 6923-6931.
- de Groot, H. J., (2000) Solid-state NMR spectroscopy applied to membrane proteins. *Curr Opin Struct Biol* **10**: 593-600.
- de la Pompa, J. L., A. Wakeham, K. M. Correia, E. Samper, S. Brown, R. J. Aguilera, T. Nakano, T. Honjo, T. W. Mak, J. Rossant & R. A. Conlon, (1997) Conservation of the Notch signalling pathway in mammalian neurogenesis. *Development* **124**: 1139-1148.
- Dedmon, M. M., K. Lindorff-Larsen, J. Christodoulou, M. Vendruscolo & C. M. Dobson, (2005) Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations. *J Am Chem Soc* **127**: 476-477.
- Delaglio, F., (2000) Protein structure determination using molecular fragment replacement and NMR dipolar couplings. *J. Am. Chem. Soc.* **122**: 2142.
- Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer & A. Bax, (1995) nmrPipe - a multidimensional spectral processing system based on unix pipes. *J. Biomol. NMR* **6**: 277-293.
- Delaglio, F., G. Kontaxis & A. Bax, (2000) Protein Structure Determination Using Molecular Fragment Replacement and NMR Dipolar Couplings. *J. Am. Chem. Soc.* **122**: 2142-2143.
- Dill, K. A. & D. Shortle, (1991) Denatured states of proteins. *Annu Rev Biochem* **60**: 795-825.
- Ding, K., J. M. Louis & A. M. Gronenborn, (2004) Insights into conformation and dynamics of protein GB1 during folding and unfolding by NMR. *J Mol Biol* **335**: 1299-1307.
- Donaldson, L. W., N. R. Skrynnikov, W. Y. Choy, D. R. Muhandiram, B. Sarkar, J. D. Forman-Kay & L. E. Kay, (2001) Structural Characterization of Proteins with an Attached ATCUN Motif by Paramagnetic Relaxation Enhancement NMR Spectroscopy. *J. Am. Chem. Soc.* **123**: 9843-9847.
- Dosset, P., J. C. Hus, M. Blackledge & D. Marion, (2000) Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J Biomol NMR*. **16**: 23-28.
- Drohat, A. C., N. Tjandra, D. M. Baldisseri & D. J. Weber, (1999) The use of dipolar couplings for determining the solution structure of rat apo-S100B(beta-beta). *Protein Sci* **8**: 800-809.
- Dunker, A. K., M. S. Cortese, P. Romero, L. M. Iakoucheva & V. N. Uversky, (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J* **272**: 5129-5148.
- Dunker, A. K., J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C.

- Garner & Z. Obradovic, (2001) Intrinsically disordered protein. *J Mol Graph Model* **19**: 26-59.
- Dunwoodie, S. L., M. Clements, D. B. Sparrow, X. Sa, R. A. Conlon & R. S. Beddington, (2002) Axial skeletal defects caused by mutation in the spondylocostal dysplasia/pudgy gene *Dll3* are associated with disruption of the segmentation clock within the presomitic mesoderm. *Development* **129**: 1795-1806.
- Dyson, H. J. & P. E. Wright, (2001) Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states. *Methods Enzymol* **339**: 258-270.
- Dyson, H. J. & P. E. Wright, (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**: 197-208.
- Egan, S. E., B. St-Pierre & C. C. Leow, (1998) Notch receptors, partners and regulators: from conserved domains to powerful functions. *Curr Top Microbiol Immunol* **228**: 273-324.
- Eggert, U. S. & G. Superti-Furga, (2008) Drugs in action. *Nat Chem Biol* **4**: 7-11.
- Evans, P. A., K. D. Topping, D. N. Woolfson & C. M. Dobson, (1991) Hydrophobic clustering in nonnative states of a protein: Interpretation of chemical shifts in NMR spectra of denatured states of lysozyme. *Proteins: Structure, Function, and Genetics* **9**: 248-266.
- Ferre-D'Amare, A. R., G. C. Prendergast, E. B. Ziff & S. K. Burley, (1993) Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* **363**: 38-45.
- Fiaux, J., E. B. Bertelsen, A. L. Horwich & K. Wuthrich, (2002) NMR analysis of a 900K GroEL GroES complex. *Nature* **418**: 207-211.
- Fiebig, K. M., H. Schwalbe, M. Buck, L. J. Smith & C. M. Dobson, (1996) Toward a Description of the Conformations of Denatured States of Proteins. Comparison of a Random Coil Model with NMR Measurements. *J. Phys. Chem.* **100**: 2661-2666.
- Fink, A. L., (2005) Natively unfolded proteins. *Curr Opin Struct Biol* **15**: 35-41.
- Fischer, M. W., J. A. Losonczi, J. L. Weaver & J. H. Prestegard, (1999) Domain orientation and dynamics in multidomain proteins from residual dipolar couplings. *Biochemistry* **38**: 9013-9022.
- Fisher, A. L., S. Ohsako & M. Caudy, (1996) The WRPW motif of the hairy-related basic helix-loop-helix repressor proteins acts as a 4-amino-acid transcription repression and protein-protein interaction domain. *Mol Cell Biol* **16**: 2670-2677.
- Fitzkee, N. C. & G. D. Rose, (2004) Reassessing random-coil statistics in unfolded proteins. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 12497-12502.
- Flanagan, J. M., M. Kataoka, D. Shortle & D. M. Engelman, (1992) Truncated staphylococcal nuclease is compact but disordered. *Proceedings of the National Academy of Sciences of the United States of America* **89**: 748-752.
- Frydman, L., A. Lupulescu & T. Scherf, (2003) Principles and features of single-scan two-dimensional NMR spectroscopy. *J Am Chem Soc* **125**: 9204-9217.
- Frydman, L., T. Scherf & A. Lupulescu, (2002) The acquisition of multidimensional NMR spectra within a single scan. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 15858-15862.
- Fuxreiter, M., I. Simon, P. Friedrich & P. Tompa, (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J Mol Biol* **338**: 1015-1026.
- Garcia, P., L. Serrano, D. Durand, M. Rico & M. Bruix, (2001) NMR and SAXS characterization of the denatured state of the chemotactic protein CheY: implications for protein folding initiation. *Protein Sci* **10**: 1100-1112.
- Gillespie, J. R. & D. Shortle, (1997a) Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. paramagnetic relaxation

- enhancement by nitroxide spin labels. *Journal of Molecular Biology* **268**: 158-169.
- Gillespie, J. R. & D. Shortle, (1997b) Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *Journal of Molecular Biology* **268**: 170-184.
- Gnanakaran, S., H. Nymeyer, J. Portman, K. Y. Sanbonmatsu & A. E. Garcia, (2003) Peptide folding simulations. *Curr. Opin. Struct. Biol.* **13**: 168-174.
- Gorre, M. E., M. Mohammed, K. Ellwood, N. Hsu, R. Paquette, P. N. Rao & C. L. Sawyers, (2001) Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. *Science* **293**: 876-880.
- Goto, Y., L. J. Calciano & A. L. Fink, (1990) Acid-induced folding of proteins. *Proceedings of the National Academy of Sciences of the United States of America* **87**: 573-577.
- Grbavec, D. & S. Stifani, (1996) Molecular interaction between TLE1 and the carboxyl-terminal domain of HES-1 containing the WRPW motif. *Biochem Biophys Res Commun* **223**: 701-705.
- Greene, R. F., Jr. & C. N. Pace, (1974) Urea and Guanidine Hydrochloride Denaturation of Ribonuclease, Lysozyme, agr-Chymotrypsin, and {beta}-Lactoglobulin. *J. Biol. Chem.* **249**: 5388-5393.
- Greenwald, I., (1998) LIN-12/Notch signaling: lessons from worms and flies. *Genes Dev* **12**: 1751-1762.
- Grzesiek, S. & A. Bax, (1993) Amino-Acid Type Determination in the Sequential Assignment Procedure of Uniformly C-13/N-15-Enriched Proteins. *J. Biomol. NMR* **3**: 185-204.
- Guilhot, F., J. Apperley, D. W. Kim, E. O. Bullorsky, M. Bacarani, G. J. Roboz, S. Amadori, C. A. de Souza, J. H. Lipton, A. Hochhaus, D. Heim, R. A. Larson, S. Branford, M. C. Muller, P. Agarwal, A. Gollerkeri & M. Talpaz, (2007) Dasatinib induces significant hematologic and cytogenetic responses in patients with imatinib-resistant or -intolerant chronic myeloid leukemia in accelerated phase. *Blood* **109**: 4143-4150.
- Guthe, S., L. Kapinos, A. Moglich, S. Meier, S. Grzesiek & T. Kiefhaber, (2004) Very fast folding and association of a trimerization domain from bacteriophage T4 fibrin. *J Mol Biol* **337**: 905-915.
- Hanks, S. K. & T. Hunter, (1995) Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J* **9**: 576-596.
- Hanks, S. K., A. M. Quinn & T. Hunter, (1988) The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science* **241**: 42-52.
- Hansen, M. R., L. Mueller & A. Pardi, (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol* **5**: 1065-1074.
- Hennig, M., W. Bermel, A. Spencer, C. M. Dobson, L. J. Smith & H. Schwalbe, (1999) Side-chain conformations in an unfolded protein: chi1 distributions in denatured hen lysozyme determined by heteronuclear ¹³C, ¹⁵N NMR spectroscopy. *J Mol Biol* **288**: 705-723.
- Hirata, H., T. Ohtsuka, Y. Bessho & R. Kageyama, (2000) Generation of structurally and functionally distinct factors from the basic helix-loop-helix gene Hes3 by alternative first exons. *J Biol Chem* **275**: 19083-19089.
- Hodsdon, M. E. & C. Frieden, (2001) Intestinal Fatty Acid Binding Protein: The Folding Mechanism As Determined by NMR Studies. *Biochemistry* **40**: 732-742.
- Honndorf, V. S., N. Coudeville, S. Laufer, S. Becker & C. Griesinger, (2008) Dynamics in the p38alpha MAP kinase-SB203580 complex observed by liquid-state NMR spectroscopy. *Angew Chem Int Ed Engl* **47**: 3548-3551.

- Hrabe de Angelis, M., J. McIntyre, 2nd & A. Gossler, (1997) Maintenance of somite borders in mice requires the Delta homologue Dll1. *Nature* **386**: 717-721.
- Hu, H., A. A. De Angelis, V. A. Mandelshtam & A. J. Shaka, (2000) The multidimensional filter diagonalization method. *J Magn Reson* **144**: 357-366.
- Hughson, F. M., P. E. Wright & R. L. Baldwin, (1990) Structural characterization of a partly folded apomyoglobin intermediate. *Science* **249**: 1544-1548.
- Humphrey, W., A. Dalke & K. Schulten, (1996) VMD: visual molecular dynamics. *J Mol Graph* **14**: 33-38, 27-38.
- Hunter, T. & B. M. Sefton, (1980) Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. *Proceedings of the National Academy of Sciences of the United States of America* **77**: 1311-1315.
- Hus, J. C., D. Marion & M. Blackledge, (2001) Determination of protein backbone structure using only residual dipolar couplings. *J. Am. Chem. Soc.* **123**: 1541-1542.
- Huse, M. & J. Kuriyan, (2002) The conformational plasticity of protein kinases. *Cell* **109**: 275-282.
- Iakoucheva, L. M., C. J. Brown, J. D. Lawson, Z. Obradovic & A. K. Dunker, (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol* **323**: 573-584.
- Ishibashi, M., S. L. Ang, K. Shiota, S. Nakanishi, R. Kageyama & F. Guillemot, (1995) Targeted disruption of mammalian hairy and Enhancer of split homolog-1 (HES-1) leads to up-regulation of neural helix-loop-helix factors, premature neurogenesis, and severe neural tube defects. *Genes Dev* **9**: 3136-3148.
- Iwahara, J., D. E. Anderson, E. C. Murphy & G. M. Clore, (2003) EDTA-Derivatized Deoxythymidine as a Tool for Rapid Determination of Protein Binding Polarity to DNA by Intermolecular Paramagnetic Relaxation Enhancement. *J. Am. Chem. Soc.* **125**: 6634-6635.
- Iwahara, J. & G. M. Clore, (2006) Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature* **440**: 1227-1230.
- Jelsch, C., M. M. Teeter, V. Lamzin, V. Pichon-Pesme, R. H. Blessing & C. Lecomte, (2000) Accurate protein crystallography at ultra-high resolution: valence electron distribution in crambin. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 3171-3176.
- Jeng, M. F., S. W. Englander, G. A. Elove, A. J. Wand & H. Roder, (1990) Structural description of acid-denatured cytochrome c by hydrogen exchange and 2D NMR. *Biochemistry* **29**: 10433-10437.
- Jennings, B. H., D. M. Tyler & S. J. Bray, (1999) Target specificities of Drosophila enhancer of split basic helix-loop-helix proteins. *Mol Cell Biol* **19**: 4600-4610.
- Jha, A. K., A. Colubri, K. F. Freed & T. R. Sosnick, (2005) Statistical coil model of the unfolded state: resolving the reconciliation problem. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 13099-13104.
- Johnson, B. A. & R. A. Blevins, (1994) Nmr View - a Computer-Program for the Visualization and Analysis of Nmr Data. *J. Biomol. NMR* **4**: 603-614.
- Jones, D. H. & S. J. Opella, (2004) Weak alignment of membrane proteins in stressed polyacrylamide gels. *J Magn Reson* **171**: 258-269.
- Jones, J. A., D. K. Wilkins, L. J. Smith & C. M. Dobson, (1997) Characterisation of protein unfolding by NMR diffusion measurements. *J. Biomol. NMR* **10**: 199-203.
- Kageyama, R., (2000) [Regulation of neuronal differentiation by bHLH factors]. *Tanpakushitsu Kakusan Koso* **45**: 1605-1611.
- Kageyama, R. & S. Nakanishi, (1997) Helix-loop-helix factors in growth and differentiation of the vertebrate nervous system. *Curr Opin Genet Dev* **7**: 659-665.

- Kamatari, Y. O., R. Kitahara, H. Yamada, S. Yokoyama & K. Akasaka, (2004) High-pressure NMR spectroscopy for characterizing folding intermediates and denatured states of proteins. *Methods* **34**: 133-143.
- Kao, H. Y., P. Ordentlich, N. Koyano-Nakagawa, Z. Tang, M. Downes, C. R. Kintner, R. M. Evans & T. Kadesch, (1998) A histone deacetylase corepressor complex regulates the Notch signal transduction pathway. *Genes Dev* **12**: 2269-2277.
- Karplus, M. & J. Kuriyan, (2005) Molecular dynamics and protein function. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 6679-6685.
- Kay, L. E., (2001) Nuclear magnetic resonance methods for high molecular weight proteins: a study involving a complex of maltose binding protein and beta-cyclodextrin. *Methods Enzymol* **339**: 174-203.
- Kazmirski, S. L., K. B. Wong, S. M. Freund, Y. J. Tan, A. R. Fersht & V. Daggett, (2001) Protein folding from a highly disordered denatured state: the folding pathway of chymotrypsin inhibitor 2 at atomic resolution. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 4349-4354.
- Kelly, D. F., R. J. Lake, T. Walz & S. Artavanis-Tsakonas, (2007) Conformational variability of the intracellular domain of Drosophila Notch and its interaction with Suppressor of Hairless. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 9591-9596.
- Kendrew, J. C., G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff & D. C. Phillips, (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**: 662-666.
- Kim, H. K. & G. Siu, (1998) The notch pathway intermediate HES-1 silences CD4 gene expression. *Mol Cell Biol* **18**: 7166-7175.
- Kim, S. & T. Szyperski, (2003) GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information. *J Am Chem Soc* **125**: 1385-1393.
- Kitahara, R. & K. Akasaka, (2003) Close identity of a pressure-stabilized intermediate with a kinetic intermediate in protein folding. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 3167-3172.
- Klein-Seetharaman, J., M. Oikawa, S. B. Grimshaw, J. Wirmer, E. Duchardt, T. Ueda, T. Imoto, L. J. Smith, C. M. Dobson & H. Schwalbe, (2002) Long-range interactions within a nonnative protein. *Science* **295**: 1719-1722.
- Kohn, J. E., I. S. Millett, J. Jacob, B. Zagrovic, T. M. Dillon, N. Cingel, R. S. Dothager, S. Seifert, P. Thiyagarajan, T. R. Sosnick, M. Z. Hasan, V. S. Pande, I. Ruczinski, S. Doniach & K. W. Plaxco, (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 12491-12496.
- Kopan, R. & A. Goate, (2000) A common enzyme connects notch signaling and Alzheimer's disease. *Genes Dev* **14**: 2799-2806.
- Korzhnev, D. M., K. Kloiber, V. Kanelis, V. Tugarinov & L. E. Kay, (2004) Probing slow dynamics in high molecular weight proteins by methyl-TROSY NMR spectroscopy: application to a 723-residue enzyme. *J Am Chem Soc* **126**: 3964-3973.
- Krebs, L. T., Y. Xue, C. R. Norton, J. R. Shutter, M. Maguire, J. P. Sundberg, D. Gallahan, V. Closson, J. Kitajewski, R. Callahan, G. H. Smith, K. L. Stark & T. Gridley, (2000) Notch signaling is essential for vascular morphogenesis in mice. *Genes Dev* **14**: 1343-1352.
- Kristjansdottir, S., K. Lindorff-Larsen, W. Fieber, C. M. Dobson, M. Vendruscolo & F. M. Poulsen, (2005) Formation of Native and Non-native Interactions in Ensembles of Denatured ACBP Molecules from Paramagnetic Relaxation Enhancement Studies. *Journal of Molecular Biology* **347**: 1053-1062.
- Kupce, E. & R. Freeman, (2003a) Fast multi-dimensional Hadamard spectroscopy. *J Magn Reson* **163**: 56-63.

- Kupce, E. & R. Freeman, (2003b) Frequency-domain Hadamard spectroscopy. *J Magn Reson* **162**: 158-165.
- Kupce, E. & R. Freeman, (2003c) Projection-reconstruction of three-dimensional NMR spectra. *J Am Chem Soc* **125**: 13958-13959.
- Kupce, E. & R. Freeman, (2004) Projection-reconstruction technique for speeding up multidimensional NMR spectroscopy. *J Am Chem Soc* **126**: 6429-6440.
- Kupce, E. & R. Freeman, (2005) Resolving ambiguities in two-dimensional NMR spectra: the 'TILT' experiment. *J Magn Reson* **172**: 329-332.
- Lauterbur, P. C., (1973) Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance. *Nature* **242**: 190-191.
- Levinthal, C., (1968) Are there pathways for protein folding? *Journal de chimie physique et de physico-chimie biologique* **65**: 44.
- Lin, H. Y., X. F. Wang, E. Ng-Eaton, R. A. Weinberg & H. F. Lodish, (1992) Expression cloning of the TGF-beta type II receptor, a functional transmembrane serine/threonine kinase. *Cell* **68**: 775-785.
- Lin, Y., (1999) NMR studies of active-site structures of adenylate kinase. In.: Purdue University, pp.
- Lindorff-Larsen, K., S. Kristjansdottir, K. Teilum, W. Fieber, C. M. Dobson, F. M. Poulsen & M. Vendruscolo, (2004) Determination of an Ensemble of Structures Representing the Denatured State of the Bovine Acyl-Coenzyme A Binding Protein. *J. Am. Chem. Soc.* **126**: 3291-3299.
- Ling, P. D., J. J. Hsieh, I. K. Ruf, D. R. Rawlins & S. D. Hayward, (1994) EBNA-2 upregulation of Epstein-Barr virus latency promoters and the cellular CD23 promoter utilizes a common targeting intermediate, CBF1. *J Virol* **68**: 5375-5383.
- Lipsitz, R. S. & N. Tjandra, (2004) Residual dipolar couplings in NMR structure analysis. *Annu Rev Biophys Biomol Struct* **33**: 387-413.
- Louhivuori, M., K. Fredriksson, K. Paakkonen, P. Permi & A. Annala, (2004) Alignment of chain-like molecules. *J Biomol NMR* **29**: 517-524.
- Louhivuori, M., K. Paakkonen, K. Fredriksson, P. Permi, J. Lounila & A. Annala, (2003) On the origin of residual dipolar couplings from denatured proteins. *J Am Chem Soc* **125**: 15647-15650.
- Maguire, Y., I. L. Chuang, S. Zhang & N. Gershenfeld, (2007) Ultra-small-sample molecular structure detection using microslot waveguide nuclear spin resonance. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 9198-9203.
- Mandelkow, E. M. & E. Mandelkow, (1998) Tau in Alzheimer's disease. *Trends Cell Biol* **8**: 425-427.
- Mandelshtam, V. A., (2000) The multidimensional filter diagonalization method. *J Magn Reson* **144**: 343-356.
- Manley, P. W., S. W. Cowan-Jacob, G. Fendrich, A. Strauss, N. Vapai, S. Grzesiek & W. Jahnke, (2006) Bcr-Abl Binding Modes of Dasatinib, Imatinib and Nilotinib: An NMR Study. *Blood, ASH Annual Meeting Abstracts* **108**: 747-.
- Marion, D., (2005) Fast acquisition of NMR spectra using Fourier transform of non-equispaced data. *J Biomol NMR* **32**: 141-150.
- Marsh, J. A., V. K. Singh, Z. Jia & J. D. Forman-Kay, (2006) Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci* **15**: 2795-2804.
- Mathews, L. S. & W. W. Vale, (1991) Expression cloning of an activin receptor, a predicted transmembrane serine kinase. *Cell* **65**: 973-982.
- Mathieson, S. I., C. J. Penkett & L. J. Smith, (1999) Characterisation of side-chain conformational preferences in a biologically active but unfolded protein. *Pac Symp Biocomput*: 542-553.
- Matsuzaki, K., J. Xu, F. Wang, W. L. McKeehan, L. Krummen & M. Kan, (1993) A widely expressed transmembrane serine/threonine kinase that does not bind

- activin, inhibin, transforming growth factor beta, or bone morphogenic factor. *J Biol Chem* **268**: 12719-12723.
- McCarney, E. R., J. E. Kohn & K. W. Plaxco, (2005) Is there or isn't there? The case for (and against) residual structure in chemically denatured proteins. *Crit Rev Biochem Mol Biol* **40**: 181-189.
- McCright, B., X. Gao, L. Shen, J. Lozier, Y. Lan, M. Maguire, D. Herzlinger, G. Weinmaster, R. Jiang & T. Gridley, (2001) Defects in development of the kidney, heart and eye vasculature in mice homozygous for a hypomorphic Notch2 mutation. *Development* **128**: 491-502.
- Meier, S., M. Blackledge & S. Grzesiek, (2008) Conformational distributions of unfolded polypeptides from novel NMR techniques. *J Chem Phys* **128**: 052204.
- Meier, S., S. Grzesiek & M. Blackledge, (2007a) Mapping the conformational landscape of urea-denatured ubiquitin using residual dipolar couplings. *J Am Chem Soc* **129**: 9799-9807.
- Meier, S., S. Guthe, T. Kiefhaber & S. Grzesiek, (2004) Foldon, the natural trimerization domain of T4 fibrin, dissociates into a monomeric A-state form containing a stable beta-hairpin: atomic details of trimer dissociation and local beta-hairpin stability from residual dipolar couplings. *J Mol Biol* **344**: 1051-1069.
- Meier, S., D. Haussinger & S. Grzesiek, (2002) Charged acrylamide copolymer gels as media for weak alignment. *J Biomol NMR* **24**: 351-356.
- Meier, S., D. Haussinger, P. Jensen, M. Rogowski & S. Grzesiek, (2003) High-accuracy residual 1HN-13C and 1HN-1HN dipolar couplings in perdeuterated proteins. *J Am Chem Soc* **125**: 44-45.
- Meier, S., M. Strohmeier, M. Blackledge & S. Grzesiek, (2007b) Direct observation of dipolar couplings and hydrogen bonds across a beta-hairpin in 8 M urea. *J Am Chem Soc* **129**: 754-755.
- Merutka, G., H. J. Dyson & P. E. Wright, (1995) 'Random coil' 1H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG. *J Biomol NMR* **5**: 14-24.
- Miele, L., (2006) Notch signaling. *Clin Cancer Res* **12**: 1074-1079.
- Millett, I. S., S. Doniach & K. W. Plaxco, (2002) Toward a taxonomy of the denatured state: small angle scattering studies of unfolded proteins. *Adv Protein Chem* **62**: 241-262.
- Mittag, T. & J. D. Forman-Kay, (2007) Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol* **17**: 3-14.
- Miyazono, K., P. ten Dijke, H. Yamashita & C. H. Heldin, (1994) Signal transduction via serine/threonine kinase receptors. *Semin Cell Biol* **5**: 389-398.
- Mohana-Borges, R., N. K. Goto, G. J. A. Kroon, H. J. Dyson & P. E. Wright, (2004) Structural Characterization of Unfolded States of Apomyoglobin using Residual Dipolar Couplings. *Journal of Molecular Biology* **340**: 1131-1142.
- Mok, Y.-K., C. M. Kay, L. E. Kay & J. Forman-Kay, (1999) NOE data demonstrating a compact unfolded state for an SH3 domain under non-denaturing conditions. *Journal of Molecular Biology* **289**: 619-638.
- Mukrasch, M. D., J. Biernat, M. von Bergen, C. Griesinger, E. Mandelkow & M. Zweckstetter, (2005) Sites of tau important for aggregation populate {beta}-structure and bind to microtubules and polyanions. *J Biol Chem* **280**: 24978-24986.
- Mumm, J. S., E. H. Schroeter, M. T. Saxena, A. Griesemer, X. Tian, D. J. Pan, W. J. Ray & R. Kopan, (2000) A ligand-induced extracellular cleavage regulates gamma-secretase-like proteolytic activation of Notch1. *Mol Cell* **5**: 197-206.
- Munoz, V., (2007) Conformational Dynamics and Ensembles in Protein Folding. *Annu. Rev. Biophys. Biomolec. Struct.* **36**: 395-412.

- Murre, C., P. S. McCaw & D. Baltimore, (1989) A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* **56**: 777-783.
- Nagar, B., W. G. Bornmann, P. Pellicena, T. Schindler, D. R. Veach, W. T. Miller, B. Clarkson & J. Kuriyan, (2002) Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). *Cancer Res* **62**: 4236-4243.
- Neri, D., M. Billeter, G. Wider & K. Wuthrich, (1992) NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science* **257**: 1559-1563.
- Oellers, N., M. Dehio & E. Knust, (1994) bHLH proteins encoded by the Enhancer of split complex of *Drosophila* negatively interfere with transcriptional activation mediated by proneural genes. *Mol Gen Genet* **244**: 465-473.
- Ohnishi, S., A. L. Lee, M. H. Edgell & D. Shortle, (2004) Direct Demonstration of Structural Similarity between Native and Denatured Eglin C. *Biochemistry* **43**: 4064-4070.
- Ohnishi, S. & D. Shortle, (2003) Observation of residual dipolar couplings in short peptides. *Proteins* **50**: 546-551.
- Ohsako, S., J. Hyer, G. Panganiban, I. Oliver & M. Caudy, (1994) Hairy function as a DNA-binding helix-loop-helix repressor of *Drosophila* sensory organ formation. *Genes Dev* **8**: 2743-2755.
- Orengo, C. A., F. M. Pearl, J. E. Bray, A. E. Todd, A. C. Martin, L. Lo Conte & J. M. Thornton, (1999) The CATH Database provides insights into protein structure/function relationships. *Nucleic Acids Res* **27**: 275-279.
- Overhauser, A. W., (1953) Polarization of Nuclei in Metals. *Physical Review* **92**: 411.
- Pan, B., B. Li, S. J. Russell, J. Y. Tom, A. G. Cochran & W. J. Fairbrother, (2002) Solution structure of a phage-derived peptide antagonist in complex with vascular endothelial growth factor. *J Mol Biol* **316**: 769-787.
- Pan, H., G. Barany & C. Woodward, (1997) Reduced BPTI is collapsed. A pulsed field gradient NMR study of unfolded and partially folded bovine pancreatic trypsin inhibitor. *Protein Sci* **6**: 1985-1992.
- Paroush, Z., R. L. Finley, Jr., T. Kidd, S. M. Wainwright, P. W. Ingham, R. Brent & D. Ish-Horowicz, (1994) Groucho is required for *Drosophila* neurogenesis, segmentation, and sex determination and interacts directly with hairy-related bHLH proteins. *Cell* **79**: 805-815.
- Pawson, T., (1994a) SH2 and SH3 domains in signal transduction. *Adv Cancer Res* **64**: 87-110.
- Pawson, T., (1994b) Tyrosine kinase signalling pathways. *Princess Takamatsu Symp* **24**: 303-322.
- Pawson, T., (1995) Protein modules and signalling networks. *Nature* **373**: 573-580.
- Pervushin, K., R. Riek, G. Wider & K. Wuthrich, (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 12366-12371.
- Peti, W., J. Meiler, R. Bruschweiler & C. Griesinger, (2002) Model-Free Analysis of Protein Backbone Motion from Residual Dipolar Couplings. *J. Am. Chem. Soc.* **124**: 5822-5833.
- Petkova, A. T., Y. Ishii, J. J. Balbach, O. N. Antzutkin, R. D. Leapman, F. Delaglio & R. Tycko, (2002) A structural model for Alzheimer's beta -amyloid fibrils based on experimental constraints from solid state NMR. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 16742-16747.
- Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale & K. Schulten, (2005) Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26**: 1781-1802.

- Pintar, A., A. De Biasio, M. Popovic, N. Ivanova & S. Pongor, (2007) The intracellular region of Notch ligands: does the tail make the difference? *Biol Direct* **2**: 19.
- Popovic, M., M. Coglievina, C. Guarnaccia, G. Verdone, G. Esposito, A. Pintar & S. Pongor, (2006) Gene synthesis, expression, purification, and characterization of human Jagged-1 intracellular region. *Protein Expr Purif* **47**: 398-404.
- Prestegard, J. H., C. M. Bougault & A. I. Kishore, (2004) Residual dipolar couplings in structure determination of biomolecules. *Chem Rev* **104**: 3519-3540.
- Prosser, R. S., J. S. Hwang & R. R. Vold, (1998) Magnetically aligned phospholipid bilayers with positive ordering: a new model membrane system. *Biophys J* **74**: 2405-2418.
- Ramirez, R. & A. Bax, (1998) Modulation of the Alignment Tensor of Macromolecules Dissolved in a Dilute Liquid Crystalline Medium. *J. Am. Chem. Soc.* **120**: 9106-9107.
- Rauscher, S., S. Baud, M. Miao, F. W. Keeley & R. Pomes, (2006) Proline and glycine control protein self-organization into elastomeric or amyloid fibrils. *Structure* **14**: 1667-1676.
- Rohl, C. A. & D. Baker, (2002) De novo determination of protein backbone structure from residual dipolar couplings using Rosetta. *J Am Chem Soc* **124**: 2723-2729.
- Ronojoy, G. & T. Claire, (2001) Lateral Inhibition through Delta-Notch Signaling: A Piecewise Affine Hybrid Model. In: Proceedings of the 4th International Workshop on Hybrid Systems: Computation and Control. Springer-Verlag, pp.
- Rovnyak, D., D. P. Frueh, M. Sastry, Z. Y. Sun, A. S. Stern, J. C. Hoch & G. Wagner, (2004) Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction. *J Magn Reson* **170**: 15-21.
- Sasai, Y., R. Kageyama, Y. Tagawa, R. Shigemoto & S. Nakanishi, (1992) Two mammalian helix-loop-helix factors structurally related to Drosophila hairy and Enhancer of split. *Genes Dev* **6**: 2620-2634.
- Sass, H. J., G. Musco, S. J. Stahl, P. T. Wingfield & S. Grzesiek, (2000) Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *J Biomol NMR* **18**: 303-309.
- Sass, J., F. Cordier, A. Hoffmann, A. Cousin, J. G. Omichinski, H. Lowen & S. Grzesiek, (1999) Purple membrane induced alignment of biological macromolecules in the magnetic field. *J. Am. Chem. Soc.* **121**: 2047-2055.
- Saupe, A. & G. Englert, (1963) High-Resolution Nuclear Magnetic Resonance Spectra of Orientated Molecules. *Physical Review Letters* **11**: 462.
- Scheraga, H. A., M. Khalili & A. Liwo, (2007) Protein-Folding Dynamics: Overview of Molecular Simulation Techniques. *Annual Review of Physical Chemistry* **58**: 57-83.
- Schindler, T., W. Bornmann, P. Pellicena, W. T. Miller, B. Clarkson & J. Kuriyan, (2000) Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* **289**: 1938-1942.
- Schlorb, C., S. Mensch, C. Richter & H. Schwalbe, (2006) Photo-CIDNP Reveals Differences in Compaction of Non-Native States of Lysozyme. *J. Am. Chem. Soc.* **128**: 1802-1803.
- Schwalbe, H., K. M. Fiebig, M. Buck, J. A. Jones, S. B. Grimshaw, A. Spencer, S. J. Glaser, L. J. Smith & C. M. Dobson, (1997) Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry* **36**: 8977-8991.
- Schwarzinger, S., G. J. Kroon, T. R. Foss, J. Chung, P. E. Wright & H. J. Dyson, (2001) Sequence-dependent correction of random coil NMR chemical shifts. *J Am Chem Soc* **123**: 2970-2978.

- Serrano, L., (1995) Comparison between the phi distribution of the amino acids in the protein database and NMR data indicates that amino acids have various phi propensities in the random coil conformation. *J Mol Biol* **254**: 322-333.
- Shah, N. P., C. Tran, F. Y. Lee, P. Chen, D. Norris & C. L. Sawyers, (2004) Overriding imatinib resistance with a novel ABL kinase inhibitor. *Science* **305**: 399-401.
- Shortle, D., (1996a) The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J* **10**: 27-34.
- Shortle, D. & M. S. Ackerman, (2001) Persistence of native-like topology in a denatured protein in 8 M urea. *Science* **293**: 487-489.
- Shortle, D., H. S. Chan & K. A. Dill, (1992) Modeling the effects of mutations on the denatured states of proteins. *Protein Sci* **1**: 201-215.
- Shortle, D. R., (1996b) Structural analysis of non-native states of proteins by NMR methods. *Curr Opin Struct Biol* **6**: 24-30.
- Shuker, S. B., P. J. Hajduk, R. P. Meadows & S. W. Fesik, (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **274**: 1531-1534.
- Sidow, A., M. S. Bulotsky, A. W. Kerrebrock, R. T. Bronson, M. J. Daly, M. P. Reeve, T. L. Hawkins, B. W. Birren, R. Jaenisch & E. S. Lander, (1997) Serrate2 is disrupted in the mouse limb-development mutant syndactyly. *Nature* **389**: 722-725.
- Simon, B. & M. Sattler, (2002) De novo structure determination from residual dipolar couplings by NMR spectroscopy. *Angew Chem Int Ed Engl* **41**: 437-440.
- Smith, L. J., K. M. Fiebig, H. Schwalbe & C. M. Dobson, (1996) The concept of a random coil. Residual structure in peptides and denatured proteins. *Fold Des* **1**: R95-106.
- Snow, C. D., E. J. Sorin, Y. M. Rhee & V. S. Pande, (2005) How well can simulation predict protein folding kinetics and thermodynamics? *Annu. Rev. Biophys. Biomolec. Struct.* **34**: 43-69.
- Spera, S. & A. Bax, (1991) Empirical correlation between protein backbone conformation and C.alpha. and C.beta. ¹³C nuclear magnetic resonance chemical shifts. *J. Am. Chem. Soc.* **113**: 5490-5492.
- Sprangers, R. & L. E. Kay, (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* **445**: 618-622.
- Stehling, M. K., R. Turner & P. Mansfield, (1991) Echo-planar imaging: magnetic resonance imaging in a fraction of a second. *Science* **254**: 43-50.
- Strauss, A., F. Bitsch, B. Cutting, G. Fendrich, P. Graff, J. Liebetanz, M. Zurini & W. Jahnke, (2003) Amino-acid-type selective isotope labeling of proteins expressed in Baculovirus-infected insect cells useful for NMR studies. *J Biomol NMR* **26**: 367-372.
- Strauss, A., F. Bitsch, G. Fendrich, P. Graff, R. Knecht, B. Meyhack & W. Jahnke, (2005) Efficient uniform isotope labeling of Abl kinase expressed in Baculovirus-infected insect cells. *J Biomol NMR* **31**: 343-349.
- Swiatek, P. J., C. E. Lindsell, F. F. del Amo, G. Weinmaster & T. Gridley, (1994) Notch1 is essential for postimplantation development in mice. *Genes Dev* **8**: 707-719.
- Swope, W. C., J. W. Pitera & F. Suits, (2004) Describing protein folding kinetics by molecular dynamics simulations. 1. Theory. *J. Phys. Chem. B* **108**: 6571-6581.
- Takebayashi, K., Y. Sasai, Y. Sakai, T. Watanabe, S. Nakanishi & R. Kageyama, (1994) Structure, chromosomal locus, and promoter analysis of the gene encoding the mouse helix-loop-helix factor HES-1. Negative autoregulation through the multiple N box elements. *J Biol Chem* **269**: 5150-5156.
- Tanford, C., K. Kawahara & S. Lapanje, (1966) Proteins in 6 m Guanidine Hydrochloride. Demonstration of random coil behavior. *J. Biol. Chem.* **241**: 1921-1923.

- Taylor, S. S., E. Radzio-Andzelm & T. Hunter, (1995) How do protein kinases discriminate between serine/threonine and tyrosine? Structural insights from the insulin receptor protein-tyrosine kinase. *FASEB J* **9**: 1255-1266.
- ten Dijke, P., P. Franzen, H. Yamashita, H. Ichijo, C. H. Heldin & K. Miyazono, (1994) Serine/threonine kinase receptors. *Prog Growth Factor Res* **5**: 55-72.
- ten Dijke, P., H. Ichijo, P. Franzen, P. Schulz, J. Saras, H. Toyoshima, C. H. Heldin & K. Miyazono, (1993) Activin receptor-like kinases: a novel subclass of cell-surface receptors with predicted serine/threonine kinase activity. *Oncogene* **8**: 2879-2887.
- Tietze, K., N. Oellers & E. Knust, (1992) Enhancer of splitD, a dominant mutation of Drosophila, and its use in the study of functional domains of a helix-loop-helix protein. *Proceedings of the National Academy of Sciences of the United States of America* **89**: 6152-6156.
- Tjandra, N. & A. Bax, (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**: 1111-1114.
- Tokarski, J. S., J. A. Newitt, C. Y. Chang, J. D. Cheng, M. Wittekind, S. E. Kiefer, K. Kish, F. Y. Lee, R. Borzilleri, L. J. Lombardo, D. Xie, Y. Zhang & H. E. Klei, (2006) The structure of Dasatinib (BMS-354825) bound to activated ABL kinase domain elucidates its inhibitory activity against imatinib-resistant ABL mutants. *Cancer Res* **66**: 5790-5797.
- Tollinger, M., N. R. Skrynnikov, F. A. Mulder, J. D. Forman-Kay & L. E. Kay, (2001) Slow dynamics in folded and unfolded states of an SH3 domain. *J Am Chem Soc* **123**: 11341-11352.
- Tolman, J. R., (2001) Dipolar couplings as a probe of molecular dynamics and structure in solution. *Curr Opin Struct Biol* **11**: 532-539.
- Tolman, J. R., J. M. Flanagan, M. A. Kennedy & J. H. Prestegard, (1995) Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proceedings of the National Academy of Sciences of the United States of America* **92**: 9279-9283.
- Tolman, J. R., J. M. Flanagan, M. A. Kennedy & J. H. Prestegard, (1997) NMR evidence for slow collective motions in cyanometmyoglobin. *Nat Struct Biol* **4**: 292-297.
- Tracy, S., P. van der Geer & T. Hunter, (1995) The receptor-like protein-tyrosine phosphatase, RPTP alpha, is phosphorylated by protein kinase C on two serines close to the inner face of the plasma membrane. *J Biol Chem* **270**: 10587-10594.
- Tugarinov, V., P. M. Hwang, J. E. Ollerenshaw & L. E. Kay, (2003) Cross-correlated relaxation enhanced ¹H-¹³C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J Am Chem Soc* **125**: 10420-10428.
- Tugarinov, V. & L. E. Kay, (2003a) Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *J Am Chem Soc* **125**: 13868-13878.
- Tugarinov, V. & L. E. Kay, (2003b) Quantitative NMR studies of high molecular weight proteins: application to domain orientation and ligand binding in the 723 residue enzyme malate synthase G. *J Mol Biol* **327**: 1121-1133.
- Tugarinov, V. & L. E. Kay, (2003c) Side chain assignments of Ile delta 1 methyl groups in high molecular weight proteins: an application to a 46 ns tumbling molecule. *J Am Chem Soc* **125**: 5701-5706.
- Tugarinov, V., L. E. Kay, I. Ibraghimov & V. Y. Orekhov, (2005) High-resolution four-dimensional ¹H-¹³C NOE spectroscopy using methyl-TROSY, sparse data acquisition, and multidimensional decomposition. *J Am Chem Soc* **127**: 2767-2775.
- Tugarinov, V., R. Muhandiram, A. Ayed & L. E. Kay, (2002) Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase g. *J Am Chem Soc* **124**: 10025-10035.

- Tugarinov, V., R. Sprangers & L. E. Kay, (2004) Line narrowing in methyl-TROSY using zero-quantum ^1H - ^{13}C NMR spectroscopy. *J Am Chem Soc* **126**: 4921-4925.
- Tycko, R., F. J. Blanco & Y. Ishii, (2000) Alignment of Biopolymers in Strained Gels: A New Way To Create Detectable Dipole-Dipole Couplings in High-Resolution Biomolecular NMR. *J. Am. Chem. Soc.* **122**: 9340-9341.
- Uversky, V. N., J. Li & A. L. Fink, (2001) Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *J Biol Chem* **276**: 10737-10744.
- Vajpai, N., A. Strauss, G. Fendrich, S. Cowan-Jacob, P. Manley, W. Jahnke & S. Grzesiek, (2008a) Backbone NMR resonance assignment of the Abelson kinase domain in complex with imatinib. *Biomolecular NMR Assignments* **2**: 41-42.
- Vajpai, N., A. Strauss, G. Fendrich, S. W. Cowan-Jacob, P. W. Manley, S. Grzesiek & W. Jahnke, (2008b) Solution conformations and dynamics of ABL kinase-inhibitor complexes determined by NMR substantiate the different binding modes of imatinib/nilotinib and dasatinib. *J Biol Chem* **283**: 18292-18302.
- van der Geer, P., T. Hunter & R. A. Lindberg, (1994) Receptor protein-tyrosine kinases and their signal transduction pathways. *Annu Rev Cell Biol* **10**: 251-337.
- Van Doren, M., A. M. Bailey, J. Esnayra, K. Ede & J. W. Posakony, (1994) Negative regulation of proneural gene activity: hairy is a direct transcriptional repressor of achaete. *Genes Dev* **8**: 2729-2742.
- van Es, J. H., M. E. van Gijn, O. Riccio, M. van den Born, M. Vooijs, H. Begthel, M. Cozijnsen, S. Robine, D. J. Winton, F. Radtke & H. Clevers, (2005) Notch/gamma-secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells. *Nature* **435**: 959-963.
- Venters, R. A., C. C. Huang, B. T. Farmer, 2nd, R. Trolard, L. D. Spicer & C. A. Fierke, (1995) High-level $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ labeling of proteins for NMR studies. *J Biomol NMR* **5**: 339-344.
- Vigneri, P. & J. Y. Wang, (2001) Induction of apoptosis in chronic myelogenous leukemia cells through nuclear entrapment of BCR-ABL tyrosine kinase. *Nat Med* **7**: 228-234.
- Vogtherr, M., K. Saxena, S. Hoelder, S. Grimme, M. Betz, U. Schieborr, B. Pescatore, M. Robin, L. Delarbre, T. Langer, K. U. Wendt & H. Schwalbe, (2006) NMR characterization of kinase p38 dynamics in free and ligand-bound forms. *Angew Chem Int Ed Engl* **45**: 993-997.
- Weinmaster, G., (1998) Notch signaling: direct or what? *Curr Opin Genet Dev* **8**: 436-442.
- Weisberg, E., P. W. Manley, W. Breitenstein, J. Bruggen, S. W. Cowan-Jacob, A. Ray, B. Huntly, D. Fabbro, G. Fendrich, E. Hall-Meyers, A. L. Kung, J. Mestan, G. Q. Daley, L. Callahan, L. Catley, C. Cavazza, M. Azam, D. Neuberg, R. D. Wright, D. G. Gilliland & J. D. Griffin, (2005) Characterization of AMN107, a selective inhibitor of native and mutant Bcr-Abl. *Cancer Cell* **7**: 129-141.
- West, N. J. & L. J. Smith, (1998) Side-chains in native and random coil protein conformations. Analysis of NMR coupling constants and χ_1 torsion angle preferences. *J Mol Biol* **280**: 867-877.
- Wharton, K. A., K. M. Johansen, T. Xu & S. Artavanis-Tsakonas, (1985) Nucleotide sequence from the neurogenic locus notch implies a gene product that shares homology with proteins containing EGF-like repeats. *Cell* **43**: 567-581.
- Wiesner, S., G. Stier, M. Sattler & M. J. Macias, (2002) Solution structure and ligand recognition of the WW domain pair of the yeast splicing factor Prp40. *J Mol Biol* **324**: 807-822.
- Wiesner, S., L. E. Wybenga-Groot, N. Warner, H. Lin, T. Pawson, J. D. Forman-Kay & F. Sicheri, (2006) A change in conformational dynamics underlies the activation of Eph receptor tyrosine kinases. *EMBO J* **25**: 4686-4696.

- Wilkins, D. K., S. B. Grimshaw, V. Receveur, C. M. Dobson, J. A. Jones & L. J. Smith, (1999) Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **38**: 16424-16431.
- Williamson, M. P., T. F. Havel & K. Wuthrich, (1985) Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ¹H nuclear magnetic resonance and distance geometry. *J Mol Biol* **182**: 295-315.
- Wishart, D. S. & B. D. Sykes, (1994) Chemical shifts as a tool for structure determination. *Methods Enzymol* **239**: 363-392.
- Wu, Z. & A. Bax, (2002) Measurement of long-range ¹H-¹H dipolar couplings in weakly aligned proteins. *J Am Chem Soc* **124**: 9672-9673.
- Xue, Y., X. Gao, C. E. Lindsell, C. R. Norton, B. Chang, C. Hicks, M. Gendron-Maguire, E. B. Rand, G. Weinmaster & T. Gridley, (1999) Embryonic lethality and vascular defects in mice lacking the Notch ligand Jagged1. *Hum Mol Genet* **8**: 723-730.
- Yan, B., J. Heus, N. Lu, R. C. Nichols, N. Raben & P. H. Plotz, (2001) Transcriptional regulation of the human acid alpha-glucosidase gene. Identification of a repressor element and its transcription factors Hes-1 and YY1. *J Biol Chem* **276**: 1789-1793.
- Zhang, O., J. D. Forman-Kay, D. Shortle & L. E. Kay, (1997a) Triple-resonance NOESY-based experiments with improved spectral resolution: applications to structural characterization of unfolded, partially folded and folded proteins. *J Biomol NMR* **9**: 181-200.
- Zhang, O., L. E. Kay, D. Shortle & J. D. Forman-Kay, (1997b) Comprehensive NOE characterization of a partially folded large fragment of staphylococcal nuclease Delta131Delta, using NMR methods with improved resolution. *J Mol Biol* **272**: 9-20.
- Zwanzig, R., A. Szabo & B. Bagchi, (1992) Levinthal's paradox. *Proceedings of the National Academy of Sciences of the United States of America* **89**: 20-22.
- Zweckstetter, M. & A. Bax, (2000) Prediction of Sterically Induced Alignment in a Dilute Liquid Crystalline Phase: Aid to Protein Structure Determination by NMR. *J. Am. Chem. Soc.* **122**: 3791-3792.
- Zweckstetter, M. & A. Bax, (2001) Single-Step Determination of Protein Substructures Using Dipolar Couplings: Aid to Structural Genomics. *J. Am. Chem. Soc.* **123**: 9490-9491.
- Zweifel, M. E., D. J. Leahy, F. M. Hughson & D. Barrick, (2003) Structure and stability of the ankyrin domain of the Drosophila Notch receptor. *Protein Sci* **12**: 2622-2632.

CURRICULUM VITAE

Name NAVRATNA VAJPAI
Date of Birth 20th March 1980
Designation Post Doctoral Fellow
Address Landskronstrasse 65
CH 4056, Basel
Switzerland
Phone +41-615341360 (Res.) +41-612672080 (Office)
+41-787334485 (Mobile)
Email navratna.vajpai@unibas.ch ; navratna.vajpai@gmail.com
Nationality Indian

Academic qualifications

- Ph.D in Structural Biology at Biozentrum, University of Basel, Switzerland
Oct 2004 – Oct 2008.
Supervisor: Prof. Stephan Grzesiek
Dissertation Title: “**Structural characterization of the leukemia drug target ABL kinase and unfolded polypeptides by novel solution NMR techniques**”
- Master of Science in Chemistry at Indian Institute of Technology (IIT), Madras, India in 2004.
Supervisor: Prof. N. Chandrakumar
Dissertation Title: “**Spin Evolution during Shaped Pulse Excitation**”
- Bachelor of Science (Chemistry, Physics, Mathematics) at Brahma Nand Degree College, C.S.J.M University, Kanpur (U.P.), India in 2001.

Computer Skills

- MS Office, Basic Programming in Matlab, Tcl/Tk, Unix, Awk
- Worked with Molecular Dynamics Simulation package: CHARMM, NAMD, GROMACS and NMR softwares

Work Experience

- Post Doctoral Fellow at Biozentrum. University of Basel
Oct 2008 – Apr 2010
Project Title: “**NMR studies of unfolded polypeptides and investigation of protein properties under high pressure**”.
Supervisor: Prof. Stephan Grzesiek
Prof. J. P. Waltho, University of Sheffield, UK
Prof. Harald Schwalbe, University of Frankfurt
- Selected summer trainee in the Visiting Student Research Program (**VSRP**) at the Department of Chemical Sciences, Tata Institute of Fundamental Research, Mumbai.
(Project Title: “**Triplet state of Naphthalene Studied by Electron Paramagnetic Resonance (EPR)**”)
Supervisor: Dr. Ranjan Das

Academic Awards/Prizes

- Recipient of Swiss National Science Foundation Fellowship (SNSF) for Postdoctoral research in United States.
- Recipient of Travel Award for ISMAR-2007 conference held in Kenting, Taiwan.
- Received Fellowship as a Merit Scholar during full duration of the Master's Degree.
- Received Fellowship as a part of Visiting Student Research Program at Tata Institute of Fundamental Research (TIFR), Mumbai, India
- Received Travel Award for ICMRBS-2006 conference held at Goettingen, Germany
- Qualified Joint CSIR-UGC examination for Junior Research Fellowship (JRF) and eligibility for Lecturer ship (NET) held in December 2003 and June 2004 (twice)

Scientific Publications:

- Dames S.A., Aregger R., **Vajpai N.**, Bernado P., Blackledge M., Grzesiek S.
Residual dipolar couplings in short peptides reveal systematic conformational preferences of individual amino acids.
J Am Chem Soc (2006) **128**: 13508 – 13514
- P.W. Manley, S.W. Cowan-Jacob, G. Fendrich, A. Strauss, **N. Vajpai**, S. Grzesiek, W. Jahnke
Bcr-Abl binding modes of dasatinib, imatinib and nilotinib: an NMR study.
Blood **108**: 747 (ASH meeting abstract 2006)
- **Vajpai N.**, Strauss A., Fenderich G., Manley P.W., Jacob S., Grzesiek S, Jahnke W.
Solution conformations and dynamics of ABL kinase inhibitor complexes determined by NMR substantiate the different binding modes of imatinib/nilotinib and dasatinib.
J Biol Chem (2008) **283**: 18292-18302
* This paper is selected as the JBC **"Paper of the week"** and also featured as the cover page article of the issue released on 27th June 2008
* This paper is reviewed in the Faculty of 1000 Biology
- **Vajpai N.**, Strauss A., Fenderich G., Manley P.W., Jacob S., Jahnke W., Grzesiek S.
Backbone resonance assignment for the Abelson kinase domain in complex with Imatinib. *Biomol NMR Assgn* (2008) **2**: 41-42
- **Vajpai N.**, Gentner M., Huang J-r., Blackledge M., Grzesiek S.
Side-chain conformations in urea-denatured proteins studied by ³J scalar couplings and residual dipolar couplings.
J Am Chem Soc (2010) **132**: 3196 – 3203
- J. Zhang, F.J. Adrián, W. Jahnke, S.W. Cowan-Jacob, A.G. Li, R.E. Iacob, T. Sim, J. Powers, C. Dierks, F. Sun, G-R. Guo, Q. Ding, B. Okram, Y. Choi, A. Wojciechowski, X. Deng, G. Liu, G. Fendrich, A. Strauss, **N. Vajpai**, S. Grzesiek, T. Tuntland, Y. Liu, B. Bursulaya, M. Azam, P.W. Manley, J.R. Engen, G.Q. Daley, M. Warmuth, N.S. Gray
Targeting wild-type and T315I Bcr-Abl by combining allosteric with ATP-site inhibitors.
Nature (2010) v**463** n7280, 501-506

Extra Curricular Activities

Oral/Poster presentations:

- International Centre of Genetic Engineering and Biotechnology, New Delhi, India March 2010 (Invited oral presentation)
- NMRS 2010 held at Lucknow, India (Poster)
- Keystone Symposium Conference, Feb, 2009 at Santa Fe, New Mexico, USA (Poster)
- Biophysics and Structural Biology Symposium, Biozentrum Basel, July 2008 (Oral presentation)
- Novartis Institutes of Biomedical Research, PSU, Basel, April 2007 (Invited oral presentation)
- ISMAR-2007 conference held at Kenting, Taiwan (Poster)
- ICMRBS-2006 conference held at Goettingen, Germany (Poster).
- Selected participant in EMBO-2006 course in NMR (theory) held at Il Ciocco, Italy.
- Selected participant in EMBO-2005 course in NMR (practical) held at Biozentrum, Basel, Switzerland
- Chemist Meet-2004 held at IIT Madras, Chennai, India (Poster)
- Active participant and Student coordinator at College cultural programs.

Languages

- Hindi Write and speak
- English Write and speak

Hobbies and Interest

- Outdoor sports
- Traveling and meeting people and learning their culture
- Listening music

Positions held

- Class representative during the Master's program at IIT Madras
- Coordinator of Cultural fest "Resonance 2003 and 2004" at IIT Madras.

Signature and Declaration

I hereby declare that all the given information are true and can be presented whenever asked.

Date : March 2010

Place : Basel

(Navratna Vajpai)