UNIVERSITY OF LEEDS

Article:
Zhuravleva, A and Korzhnev, DM (2017) Protein Folding by NMR. Progress in Nuclear Magnetic Resonance Spectroscopy, 100. pp. 52-77. ISSN 0079-6565

https://doi.org/10.1016/j.pnmrs.2016.10.002

Accepted Manuscript

Protein Folding by NMR

Anastasia Zhuravleva, Dmitry M. Korzhnev

PII: S0079-6565(16)30028-0
DOI: http://dx.doi.org/10.1016/j.pnmrs.2016.10.002
Reference: JPNMRS 1430

To appear in: *Progress in Nuclear Magnetic Resonance Spectroscopy*

Received Date: 9 July 2016
Revised Date: 17 October 2016
Accepted Date: 17 October 2016

Please cite this article as: A. Zhuravleva, D.M. Korzhnev, Protein Folding by NMR, *Progress in Nuclear Magnetic Resonance Spectroscopy* (2016), doi: http://dx.doi.org/10.1016/j.pnmrs.2016.10.002

Protein Folding by NMR

Anastasia Zhuravleva[1*] and Dmitry M. Korzhnev[2*]

[1] Astbury Centre for Structural Molecular Biology and Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, United Kingdom

[2] Department of Molecular Biology and Biophysics, University of Connecticut Health Center, Farmington, CT 06030, USA

* Corresponding authors: Anastasia Zhuravleva (a.zhuravleva@leeds.ac.uk)
Dmitry M. Korzhnev (korzhniev@uchc.edu)

Abstract

Protein folding is a highly complex process proceeding through a number of disordered and partially folded nonnative states with various degrees of structural organization. These transiently and sparsely populated species on the protein folding energy landscape play crucial roles in driving folding toward the native conformation, yet some of these nonnative states may also serve as precursors for protein misfolding and aggregation associated with a range of devastating diseases, including neuro-degeneration, diabetes and cancer. Therefore, *in vivo* protein folding is often reshaped co- and post-translationally through interactions with the ribosome, molecular chaperones and/or other cellular components. Owing to developments in instrumentation and methodology, solution NMR spectroscopy has emerged as the central experimental approach for the detailed characterization of the complex protein folding processes *in vitro* and *in vivo*. NMR relaxation dispersion and saturation transfer methods provide the means for a detailed characterization of protein folding kinetics and thermodynamics under native-like conditions, as well as modeling high-resolution structures of weakly populated short-lived conformational states on the protein folding energy landscape. Continuing development of isotope labeling strategies and NMR methods to probe high molecular weight protein assemblies, along with advances of in-cell NMR, have recently allowed protein folding to be studied in the context of ribosome-nascent chain complexes and molecular chaperones, and even inside living cells. Here we review solution NMR approaches to investigate the protein folding energy landscape, and discuss selected applications of NMR methodology to studying protein folding *in vitro* and *in vivo*. Together, these examples highlight a vast potential of solution NMR in providing atomistic insights into molecular mechanisms of protein folding and homeostasis in health and disease.

Keywords

Nonnative protein states; folding intermediates; protein quality control; molecular chaperones; ribosome-nascent chain; in-cell NMR

Glossary of abbreviations

ADP - adenosine diphosphate
ATP -adenosine triphosphate
BSA - bovine serum albumin
CD - circular dichroism
CEST - chemical exchange saturation transfer
CPMG - Carr Purcell Meiboom Gill
CryoEM - cryo electron microscopy
CW - continuous wave
DEST - dark-state exchange saturation transfer
DQ - double quantum
DOSY - diffusion-ordered spectroscopy
HMQC - heteronuclear multiple quantum correlation
HSQC - heteronuclear single quantum correlation
HX - hydrogen exchange
IDP - intrinsically disordered protein
IDPR - intrinsically disordered protein region
IPAP – in-phase anti-phase
MD - molecular dynamics
MQ - multiple quantum
NAC - non-amyloid-β component SQ -single-quantum
NBD - nucleotide-binding domain
NC - nascent chain
NMR – nuclear magnetic resonance
NOE - nuclear Overhauser effect
PQC - protein quality control
PRE - paramagnetic relaxation enhancement
RCI – random coil index
RCSA – residual chemical shift anisotropy,
RD - relaxation dispersion
RDC - residual dipolar coupling
RF - radiofrequency
RNC - ribosome-nascent chain
SAXS – small angle X-ray scattering
SBD - substrate-binding domain
SSP - secondary structure propensity
ST - saturation transfer
SUV - small unilamellar liposome/vesicle
TF - trigger factor
TROSY - transverse relaxation optimized spectroscopy
ZQ - zero quantum

Table of contents

1. Introduction

The protein folding problem is concerned with how the protein sequence determines three-dimensional shape and how proteins attain their shapes [1]. While the wealth of structural data for folded proteins has led to remarkable advances in the prediction of native structures from sequence alone or from limited experimental data [2, 3], details of the process by which proteins fold remain elusive. Transiently and sparsely populated nonnative states on the protein folding energy landscape play crucial roles in driving folding toward the native conformer [4, 5]. Although protein function is commonly attributed to the natively folded state, there is increasing evidence that thermally accessible nonnative species may also play important functional roles [6, 7]. On the other hand, some nonnative states may serve as precursors of aggregation and formation of β-sheet amyloids, deposits of which are associated with over 40 pathologies [8-11]. Some of these conditions, such as Alzheimer's and Parkinson's diseases, are caused by aggregation of disordered peptides. Many others, however, stem from misfolding and aggregation of structured globular proteins. A number of proteins unrelated to disease can also form β-sheet amyloids under certain conditions, suggesting that protein folding and aggregation form different aspects of the same general problem.

Protein folding proceeds through a number of partially ordered states, from denatured ensembles [12-14] to intermediate [13, 15-17] and transition states [13, 18-20] that have various degrees of structural organization. Studying these nonnative states is complicated because of their transient nature and low populations under non-denaturating conditions. Weakly populated short-lived folding intermediates are not readily detected, and, if detected, are often studied indirectly due to the difficulty of their isolation [13, 15, 16]. Denatured states are evanescent under native conditions and are difficult to characterize because the methods used to destabilize native proteins can alter their structural ensembles [12].

Proteins are synthesized on the ribosome as linear polymers of amino acids. The nascent polypeptide chain emerging from the ribosome often cannot adopt its native conformation until the full protein or its modular domain is synthesized. This represents an eminent threat, since an incomplete polypeptide chain containing exposed hydrophobic groups is prone to aggregation. Therefore, cells have evolved sophisticated systems of protein quality control that facilitate co-translational and post-translational protein folding [21]. These complex multi-component protein folding machines are extremely difficult to characterize because of their large molecular weights and complex rearrangements and dynamics.

The past two decades have brought the development of new NMR methods that extend both the molecular weights of protein systems and the complexity of processes amenable to study by solution NMR spectroscopy. Thus, transverse relaxation optimized spectroscopy (TROSY) NMR methods [22, 23] allow, for the first time, probing of dynamics and interactions of large (up to 1 MDa) protein assemblies [24-26], opening an avenue for deciphering molecular

mechanisms of their action. Relaxation dispersion (RD) [27-31] and saturation transfer (ST) [31-33] methods provide a detailed look into the pathways of biomolecular processes, allowing studies of transiently populated intermediates of protein folding [29, 34, 35], enzyme catalysis [36, 37] and binding [38]. This review provides an overview of these NMR methods and key questions about protein folding that can be addressed by NMR spectroscopy, as well as a description of selected recent applications of NMR to studies of protein folding *in vitro and in vivo*.

2. NMR tools to probe the protein folding energy landscape

Characteristic times for protein folding range from ~1 μs up to seconds and slower [39], with a large number of single-domain proteins and small protein domains reported to fold on the ms time-scale [40, 41]. Under physiological conditions, proteins and/or their modular domains reside in the lowest energy native state, but also undergo excursions on the folding time-scale to sparsely populated nonnative conformers that correspond to the higher energy local minima on the folding energy landscape. These intermediates [13, 15-17] and/or denatured states [12-14] exhibit various degrees of structural organization and play important roles in folding [34, 42, 43], as well as misfolding and aggregation [44, 45]. The short life-times (often μs-ms) and low populations render these minor states invisible to most modern biophysical techniques, including NMR spectroscopy where the presence of such states can only be inferred from line-broadening of the dominant native state, or, otherwise, from exchange contributions (R_{ex}) to transverse relaxation rates $R_{2,eff}$ of NMR-active nuclei in the native state [29, 35]. These minor nonnative states are connected to the ground native state via energy barriers (transition states [13, 18-20]) whose properties determine folding kinetics.

A range of NMR methods for studying μs-ms protein dynamics (reviewed in section 2.1) enable characterization of both folding kinetics and structural organization of the minor nonnative states on the folding energy landscape, including the relaxation dispersion (RD) and saturation transfer (ST) methods reviewed here. The two major types of RD experiments include (i) Carr-Purcell-Meiboom-Gill (CPMG) methods [31, 46, 47] that exploit modulation of R_{ex} by a sequence of evenly spaced refocusing pulses, and (ii) rotating-frame $R_{1\rho}$ relaxation experiments [30, 31] that utilize modulation of R_{ex} by an on- or off-resonance continuous wave (CW) radiofrequency (RF) field. The ST experiments exploit modulation of R_{ex} by a weak RF field [31-33], and are conceptually similar to off-resonance $R_{1\rho}$ measurements [31]. Transient nonnative protein states can be investigated by other types of NMR measurements, such as paramagnetic relaxation enhancement (PRE) reviewed in detail elsewhere [32, 48]. A number of NMR tools are also available for studies of protein folding occurring with characteristic times of seconds and slower. Generally, such dynamics can be investigated by monitoring changes in NMR spectra in real-time, as reviewed in [49, 50]. Longitudinal magnetization (ZZ) exchange experiments can be used to probe interconversions between states with comparable populations (free-energy difference ΔG close to 0) occurring on the time scale of seconds, as exemplified by studies of folding/unfolding transitions of the drkN SH3 domain [51-53]. Furthermore, hydrogen exchange (HX) measurements have been instrumental for studies of protein folding, making possible the detection of evanescent populations of disordered nonnative states transiently sampled by proteins under native-like conditions, as reviewed in detail elsewhere [15, 54].

In the cellular context, protein folding involves different components of co-translational and post-translational folding machinery such as ribosomes, molecular chaperones and other protein quality-control enzymes [21]. Characterization of the highly dynamic, multicomponent networks and molecular machines that facilitate protein folding in a cellular environment is an extremely challenging task. The advantage of modern solution NMR is its ability to obtain detailed atomic-level information about such complex molecular systems, allowing characterization of the highly-populated ground states as well as transient, low-populated functional conformations of large (up to 1 MDa) and multi-component biological systems (for details see recent reviews [24-26]). Importantly, these studies can be carried out under native-like conditions (e.g. physiological buffers), near cellular or cellular environments and physiological concentrations (as low as sub-μM). As a result, solution NMR has been successfully used for characterization of protein folding in the context of multi-component folding machines [55], including studies of protein folding on the ribosome [56-60], studies of ATP-dependent molecular chaperone systems and their interactions with protein substrates (e.g., Hsp60 chaperones GroEL/GroES [61-67] and TRiC [68]; Hsp90 [69-80] and Hsp70 [81-93] systems), studies of ATP-independent chaperones (e.g., the trigger factor [94, 95], small HSPs [96-98]), and studies of protein degradation machines such as AAA+ ATPase chaperonin ClpB [99, 100] and the proteasome [101-104]. TROSY NMR methods and isotope labeling to probe dynamics and interactions of large multi-protein assemblies involved in assisted protein folding are briefly reviewed in section 2.2.

2.1 Relaxation dispersion and saturation transfer methods to probe folding landscape

2.1.1 CPMG-type experiments

2.1.1.1 Basics of CPMG methods: CPMG RD experiments build upon the classic spin-echo pulse-sequence proposed in 1950 [105]: 90^{o}_{y}-τ-180^{o}_{x}-τ-acquisition (Fig. 1A; n=1). The first 90^{o}_{y} pulse of this experiment puts the magnetization of a nucleus along the X-axis of the rotating frame where it starts precessing about the direction of static magnetic field with characteristic angular frequency ω, acquiring phase $\phi_1=\omega\tau$ during the period τ. The following 180^{o}_{x} pulse inverts the phase of magnetization, so that the phases ϕ_1 and ϕ_2 acquired before and after the pulse cancel at the end of the period 2τ, resulting in full refocusing of magnetization along the X-axis referred to as a *spin-echo* (Fig. 1B). If, during the period 2τ, a nucleus undergoes exchange between states with different resonance frequencies, the phases acquired by magnetization before and after the 180^{o}_{x} pulse are not equal, resulting in incomplete refocusing of magnetization and leading to a so-called exchange enhancement R_{ex} of the transverse relaxation rate R_2: $R_{2,eff}=R_2+R_{ex}$. The same principle applies to the multiple-echo CPMG pulse-sequence [46, 47], 90^{o}_{y}-$(\tau$-180^{o}_{x}-$\tau)_{2n}$-acquisition (Fig. 1A), which consists of a variable number of evenly spaced 180^{o}_{x} refocusing pulses. The dependence of $R_{2,eff}$ rates on the frequency of application of 180^{o}

refocusing pulses in CPMG sequence (termed CPMG frequency, $\nu_{CPMG}=1/(4\tau)$) is called *CPMG relaxation dispersion* (Fig. 1C). In the absence of conformational/chemical exchange RD profiles $R_{2,eff}(\nu_{CPMG})$ measured in CPMG experiment are flat. If exchange takes place between states with different resonance frequencies (chemical shifts), $R_{2,eff}$ rates decrease with increasing frequency of application of the refocusing pulses, exhibiting characteristic dependence on ν_{CPMG} (Fig. 1C).

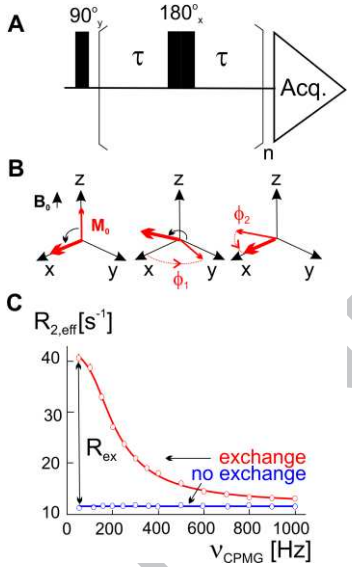


Figure 1. Spin-echo (n=1) and CPMG (even n) pulse-schemes for an isolated nucleus (**A**). Magnetization evolution during spin-echo experiment (**B**), RD profiles measured in CPMG experiment (**C**).

2.1.1.2 CPMG-type experiments for different nuclei and spin-coherences: The distinct advantage of NMR is that separate resonances in NMR spectra corresponding to individual ^{1}H, ^{13}C and ^{15}N nuclei provide a large number of site-specific probes of protein dynamics. Moreover, the possibility to create different spin-coherences in multi-spin systems sensitive to conformational exchange adds greatly to the number of independent reporters of protein dynamics. Since the pioneering work of Loria and Palmer in 1999 [106] that describes a relaxation-compensated ^{15}N CPMG scheme for studying μs-ms dynamics at the protein backbone amide positions, CPMG RD experiments have been proposed for every type of NMR active nucleus of the protein backbone, including ^{15}N [52, 106, 107], $^{1}H^{N}$ [108, 109], $^{13}C'$ [110-113], $^{13}C^{\alpha}$ [112, 114] and $^{1}H^{\alpha}$ [114, 115], as well as for ^{1}H nuclei of side-chain methyl groups (CH_3, CHD_2) [116, 117], ^{13}C nuclei of side-chain methylene (CH_2) [118], methyl (CH_3) [119, 120] and carbonyl (C') [121] groups, and ^{15}N nuclei of side-chain amino groups (NH_2, NH_3^+) [122, 123]. Different versions of these experiments are available, as, for example, relaxation compensated ^{15}N CPMG [52, 106], ^{15}N CPMG with ^{1}H CW decoupling during relaxation period [107], ^{15}N TROSY [124, 125] and anti-TROSY [125] CPMG schemes for ^{15}N RD measurements for the backbone amide groups. Most of the above pulse-sequences exploit single-quantum (SQ) coherences and, hence, are called SQ CPMG experiments. There are a number of spin-transitions in protein multi-spin

systems beyond SQ that are sensitive to μs-ms dynamics, and which can be used to design separate CPMG-type RD experiments. For example, ^{15}N and $^{1}H^{N}$ SQ CPMG pulse sequences for backbone amide ^{15}NH groups were complemented by CPMG-type schemes that exploit zero-quantum (ZQ) [109], double-quantum (DQ) [109] and multiple-quantum (MQ) [126] coherences.

CPMG-type RD experiments are designed to minimize undesirable effects arising from hetero- and homo-nuclear couplings and cross-correlated relaxation that may cause artifacts in measured CPMG RD profiles. However, some unwanted effects, e.g. associated with homo-nuclear couplings, cannot be fully removed in uniformly $^{1}H/^{13}C$ enriched protein samples by means of pulse-sequence design, necessitating the use of fractional and/or specific ^{13}C labeling and deuteration in ^{13}C and ^{1}H CPMG experiments. Robust protocols for incorporation of NMR active isotopes at key positions in proteins have been proposed to be suitable for different types of CPMG RD measurements [127]. For example, uniform U-^{15}N labeling is sufficient for ^{15}N CPMG measurements; a set of $^{1}H^{N}$, ^{15}N and $^{13}C'$ CPMG profiles can be recorded for a U-$^{15}N,^{13}C,^{2}H$ labeled protein expressed in *E. Coli* using 100% $^{2}H_2O$ M9 medium with added $^{15}NH_4Cl$ and [U-$^{2}H,^{13}C$]-glucose; $^{13}C^{\alpha}$ CPMG experiments can be performed for a selectively $^{13}C^{\alpha}$ labeled protein expressed using [2-^{13}C]-glucose as the carbon source, which leads to uniform ^{13}C incorporation at C^{α} positions without concomitant labeling of C' and C^{β} sites [128]; $^{1}H^{\alpha}$ CPMG experiments require U-^{13}C,partially-^{2}H samples expressed in 50%H_2O,50%$^{2}H_2O$ M9 medium using [U-$^{2}H,^{13}C$]-glucose, which results in up to 90% deuteration at C^{β} positions helping suppress $^{1}H^{\alpha}$ - $^{1}H^{\beta}$ scalar coupling [115]. The detailed requirements for isotope labeling and recipes for protein sample preparation can be found in papers that describe specific CPMG-type RD experiments.

2.1.1.3 Analysis of CPMG RD data: RD profiles measured in CPMG experiment allow extraction of populations of exchanging states, p_i, exchange rate constants, k_{ij}, and absolute values of frequency differences between states, $|\Delta\omega_{ij}|$, (or chemical shift differences $|\Delta\varpi_{ij}|=|\Delta\omega_{ij}/\omega_o|$, where ω_o is Larmor frequency) for each considered nucleus, so long as the minor state population(s) are 0.5% and higher and the rates of interconversion between states are on the ms time-scale. These parameters provide thermodynamic and kinetic characterization of the exchange process, along with structural information about the exchanging states. The exchange parameters can be obtained by least-square fits of experimental CPMG RD profiles $R_{2,eff}(\nu_{CPMG})$ to theoretical models [34, 129]. Closed form analytical expressions were derived that describe $R_{2,eff}(\nu_{CPMG})$ as a function of parameters of 2-state conformational exchange [130-134], including a general Carver-Richards equation valid in a wide range of exchange regimes [135, 136]. The Carver-Richards formula can be used to analyze both SQ and DQ/ZQ CPMG data; a modification of this equation has also been derived to analyze RD data obtained in MQ CPMG experiments [120, 126]. Typically, all the parameters that describe 2-state conformational exchange of a nucleus between a ground state A and a minor state B can be extracted from SQ CPMG data

recorded at two or more magnetic field strengths provided that minor state population p_B is ~0.5-10% and exchange kinetics is in the intermediate regime ($k_{ex}=k_{AB}+k_{BA}$ ~ $|\Delta\omega_{AB}|$, typically a few hundred to several thousand 1/s). On the other hand, only certain combinations of exchange parameters can be obtained in fast exchange $k_{ex}>>|\Delta\omega_{AB}|$ and high field $2\pi\nu_{CPMG}>>|\Delta\omega_{AB}|$ limits ($p_Ap_B\Delta\omega_{AB}^2$ and k_{ex}), as well as in slow exchange regime $k_{ex}<<|\Delta\omega_{AB}|$ ($k_{ex}p_B$ and $|\Delta\omega_{AB}|$) [52, 137]. Alternatively, in the cases of 2-site and multiple-site conformational exchange, calculation of theoretical $R_{2,eff}(\nu_{CPMG})$ values can be done by modeling the evolution of magnetization during a CPMG sequence by numerical solution of Bloch-McConnell equations [34, 138, 139].

2.1.2 Rotating-frame relaxation $R_{1\rho}$ methods

2.1.2.1 Basics of $R_{1\rho}$ methods: Rotating-frame relaxation $R_{1\rho}$ experiments provide a powerful complement to CPMG RD methods, potentially allowing extraction of the same set of μs-ms conformational exchange parameters [30, 140, 141]. In this class of experiment, the rotating-frame relaxation rate $R_{1\rho}$ is measured as a function of strength, ω_1, and/or offset, Ω, of the applied RF field (spin-lock) for magnetization aligned along the direction of the effective field $\boldsymbol{\omega_e}=(\omega_1,0,\Omega)$ (Fig. 2A,B). The offset dependence of $R_{1\rho}$ is given by $R_1\cos^2\theta+(R_2+R_{ex})\sin^2\theta$, where $\theta=\text{arccot}(\Omega/\omega_1)$ is the angle between the effective field $\boldsymbol{\omega_e}$ and static magnetic field $\mathbf{B_0}$ (Z axis), while R_1 and R_2 are longitudinal and transverse relaxation rates, respectively. In the absence of exchange, the offset dependence of $R_{1\rho}$ is symmetric with respect to zero offset, exhibiting a maximum of R_2 at $\Omega=0$ and approaching R_1 off-resonance (Fig. 2C). If μs-ms exchange takes place, R_2 is enhanced by an R_{ex} contribution that depends on the exchange parameters, RF field strength ω_1 and offset Ω.

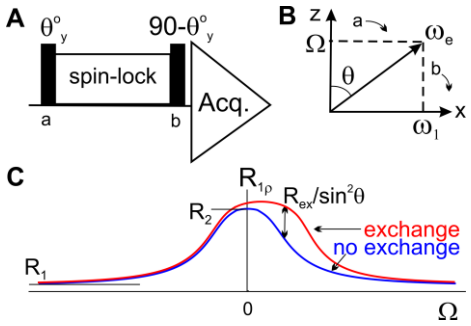


Figure 2. Rotating-frame $R_{1\rho}$ pulse-sequences for an isolated nucleus (**A**), in which magnetization is aligned along the effective field $\boldsymbol{\omega_e}=(\omega_1,0,\Omega)$, where ω_1 spin-lock field strengths and Ω is resonance offset from the spin-lock carrier frequency. (**B**). Offset dependence of the rotating-frame $R_{1\rho}$ rates (**C**).

2.1.2.2 Experimental aspects of $R_{1\rho}$ measurements: Historically, $R_{1\rho}$ measurements were considered as the method of choice for studies of fast μs conformational exchange processes [30]. The time-scale of processes that can be effectively studied by $R_{1\rho}$ and CPMG RD methods

is determined by the range of effective fields, ω_e, and CPMG frequencies, ν_{CPMG}, accessible in corresponding experiments. For example, relatively high RF fields $\omega_1/(2\pi)$ ~ 1-2 kHz and offsets Ω within ~ $\pm 3 \cdot \omega_1$ that are typically used in ^{15}N $R_{1\rho}$ experiments [30, 142-145] allow studies of 2-state conformational exchange with $k_{ex} \sim \omega_e = (\omega_1^2 + \Omega^2)^{1/2}$ of up to ~40,000/s, where k_{ex} is the sum of forward and reverse rates, while ^{15}N CPMG RD experiments that typically employ $\nu_{CPMG} = 1/(4\tau)$ (where 2τ is the delay between the refocusing pulses) from 25 to 1000 Hz are better suited for characterization of the processes with $k_{ex} \sim 2\pi\nu_{CPMG}$ of few hundred to several thousands 1/s. The maximal pulse repetition rate in CPMG experiments implied by limitations on pulse length and duty cycle and the maximal spin-lock field strengths in $R_{1\rho}$ experiments are determined by hardware limitations and depend on individual RF probes.

NMR pulse-sequences were described for on- and off-resonance $R_{1\rho}$ measurements for ^{15}N [142-145], $^{1}H^{N}$ [146, 147], $^{13}C'$ [148] and $^{13}C^{\alpha}$ [149] nuclei of the protein backbone that utilize high spin-lock field strengths $\omega_1/(2\pi)$ of one to several kHz. In these experiments, the alignment of magnetizations for different nuclei along the directions of their respective effective fields (which are different for different nuclei) is readily achieved by adiabatic pulses [143] or specialized pulse / free-precession elements [150-152] applied prior to and following the spin-lock period. Hence, these experiments enable $R_{1\rho}$ measurements for multiple nuclei at the same time. Accurate alignment of magnetization for multiple nuclei becomes increasingly difficult at lower spin-lock field strengths, although a 2D ^{15}N $R_{1\rho}$ pulse scheme was proposed that can be used at spin-lock field strengths as low as 150 Hz [153]. An alternative class of selective 1D $R_{1\rho}$ pulse-schemes was proposed, allowing $R_{1\rho}$ measurements for individual protein nuclei at low spin-lock field strengths $\omega_1/(2\pi)$ of 25Hz and higher [141, 154, 155], in which magnetization alignment along the effective field is achieved by hard pulses with suitably adjusted flip angles (Fig. 2A). These experiments can effectively probe ms time-scale processes accessible to CPMG RD methods. To date, selective 1D $R_{1\rho}$ schemes were described for ^{15}N [141], $^{1}H^{N}$ [154], $^{13}C^{\alpha}$ [155], $^{1}H^{\alpha}$ [154] and methyl ^{13}C [156] nuclei that complement regular 2D $R_{1\rho}$ pulse-sequences employing high spin-lock field strengths [30, 144].

2.1.2.3 Analysis of $R_{1\rho}$ data: On-resonance (measured as a function of spin-lock field strength ω_1) or off-resonance (measured as a function of offset Ω) $R_{1\rho}$ RD profiles can be least-square fitted to theoretical models to extract conformational exchange parameters, including the relative populations of exchanging states, p_i, exchange rate constants, k_{ij}, and chemical shift differences between states, $\Delta\varpi_{ij}$. Several analytical expressions that describe the dependence of $R_{1\rho}$ rates on the parameters of 2-state conformational exchange have been proposed [136, 157]; of special note is a general formula derived by Trott and Palmer [140], which is valid in a wide range of exchange regimes and related equations [158-161]. Alternatively, theoretical $R_{1\rho}$ rates can be obtained by modeling magnetization evolution during a spin-lock period using numerical

solution of Bloch-McConnell equations [138, 139]. By analogy with CPMG RD data, in the fast exchange limit $k_{ex}>>|\Delta\omega_{AB}|$ and in the high field limit $\omega_1>>|\Delta\omega_{AB}|$ (which is generally the case for 2D $R_{1\rho}$ employing kHz spin-lock fields) one can only extract from $R_{1\rho}$ a combination of $p_A p_B \Delta\omega_{AB}^2$ and k_{ex}, while for weak spin-lock field strengths in an intermediate regime one can obtain from $R_{1\rho}$ data all the parameters of 2-state exchange [30, 140, 141]. An important distinction between $R_{1\rho}$ and CPMG RD data is that off-resonance $R_{1\rho}$ profiles measured as a function of spin-lock offset Ω are sensitive to the signs of chemical shift differences between states $\Delta\varpi = \Delta\omega/\omega_o$ [140] and, therefore, can be used to determine the sign of of $\Delta\omega_{BA}=\Omega_B-\Omega_A$ [154]. Thus, the R_{ex} contribution to the transverse relaxation rate R_2 in the expression $R_{1\rho} = R_1\cos^2\theta+(R_2+R_{ex})\sin^2\theta$ exhibits a maximum at the offset Ω_B corresponding to the resonance frequency of the minor state [140], leading to an asymmetry in the off-resonance $R_{1\rho}$ profile with respect to the frequency of the ground state Ω_A (Fig 2C) and sometime resulting in off-resonance $R_{1\rho}$ profiles that display two maxima at distinct offsets [141]. At weak spin-lock field strengths ($\omega_1/(2\pi)$ of tenths to hundreds of Hz) and at $\omega_1 \sim k_{ex}$ it is possible to extract from off-resonance $R_{1\rho}$ data measured at one static magnetic field strengths *all* the parameters that describe 2-state ms time-scale conformational exchange, including k_{ex}, p_B and signed $\Delta\varpi_{BA}$, provided that $R_{1\rho}$ values are well sampled around Ω_B [30, 140, 141].

2.1.3 Saturation transfer methods

2.1.3.1 Basics of saturation transfer methods: In saturation transfer (ST) experiments [31-33], the longitudinal magnetization of the major state A is subjected to a weak RF field applied at different resonance offsets Ω for the time period T, followed by measurement of a relative intensity of the major state resonance I(T)/I(0). The ST methods exploit the fact that a weak RF field applied at the resonance frequency Ω_B of the minor state B will also affect magnetization of the ground state A due to saturation transfer resulting from conformational exchange between the two states. Conceptually, the ST approach represents a variation on the off-resonance $R_{1\rho}$ measurement at low-spin lock field strengths [31], and hence can be described in terms of the schematic in Fig. 2 that illustrates the basics of the $R_{1\rho}$ method. Both the ST and $R_{1\rho}$ methods quantify evolution of magnetization that is aligned with $\boldsymbol{\omega_e}$ during the spin-lock period. However, the essential difference between the two is that, in contrast to $R_{1\rho}$ measurements, in an ST experiment an off-resonance ω_1 field is applied to longitudinal magnetization without aligning the magnetization along the effective field $\boldsymbol{\omega_e}=(\omega_1,0,\Omega)$ (flip angle $\theta^o_y=0$ in Fig. 2A); this causes signal attenuation by a factor of $\cos^2\theta$ in ST experiments relative to $R_{1\rho}$ experiments, due to projection of the magnetization to $\boldsymbol{\omega_e}$ and subsequently back to the Z-axis. Therefore, I(T)/I(0) ratios measured

in ST experiments are related to rotating-frame relaxation rates obtained in $R_{1\rho}$ experiments by $I(T)/I(0) = \cos^2\theta \cdot \exp(-R_{1\rho}T)$ [31].

2.1.3.2 Saturation transfer experiments and data analysis: The idea of ST experiments dates back to the 1960s [162], but in the past several years this type of experiment has received renewed attention due to its applications to studies of ms time-scale conformational exchange in proteins. Two ST type approaches were described: dark-state exchange ST (DEST) [33, 163, 164] and chemical exchange ST (CEST) [165-171], based on irradiation of longitudinal magnetization of the ground state by a weak off-resonance RF field. DEST experiments were intended to quantify exchange between states with significantly different intrinsic R_2 rates (such as exchange between free and large aggregate bound forms of a protein) [33, 163, 164], while CEST methods were designed to probe slow to intermediate (k_{ex} of tens to hundreds 1/s) exchange processes accompanied by changes in chemical shifts (such as unfolding transitions of small proteins and protein domains) [165-171]. To date, a range of ST experiments have been proposed, including those for protein ^{15}N [165], $^{13}C^{\alpha}$ [169, 170], $^{13}C^{\beta}$ [170], $^{13}C'$ [168], $^{13}CH_3$ [167] and $^{13}CHD_2$ groups [171], allowing measurements of exchange parameters and chemical shifts in so-called 'invisible' conformational states. These parameters can be extracted by least-square fits of experimental I(T)/I(0) ST profiles to theoretical values calculated as $\cos^2\theta \cdot \exp(-R_{1\rho}T)$, or obtained by numerical solution of the Bloch-McConnell equations [138, 139]. In analogy with off-resonance $R_{1\rho}$ measurements (see above), CEST experiments recorded at low-spin lock fields $\omega_1/(2\pi)$ of ten to several hundred Hz are suitable for studies of conformational exchange processes with $k_{ex} \sim \omega_1$.

2.2. NMR approaches to probing supramolecular assemblies that assist protein folding

2.2.1 Isotope-labeling schemes for NMR studies of large proteins: Protein NMR becomes increasingly difficult with increasing molecular weight of the protein, mainly due to fast relaxation of ^1H nuclei caused by efficient dipole-dipole interactions with a large number of neighboring protons. High degrees of protein deuteration [172] along with selective protonation of certain groups (e.g. back-protonation of labile backbone amide groups, or biosynthetic incorporation methyl C^1H_3 groups) allows one to significantly reduce ^1H transverse relaxation rates and drastically increase NMR sensitivity for large systems [24-26]. Extensive protein deuteration with subsequent back-protonation of the backbone ^{15}NH groups is essential and often sufficient for characterization of protein systems with molecular weights in the range of 30-200 kDa, while selective incorporation of ^{13}C-labeled protonated side-chain methyl groups against a highly deuterated background enables methyl NMR studies of supramolecular protein assemblies with molecular weights of up to 1 MDa [24-26]. High-quality methyl-methyl NMR spectra can be obtained even for very large and complicated protein systems because methyl groups are

intrinsically dynamic, i.e. exhibit rapid stochastic rotations about the CH_3 symmetry axis (also resulting in methyl ^1H chemical shift degeneracy). Selective $^{13}CH_3$ labels introduced in one or several types methyl containing side-chains, including Ala $C^{\beta}H_3$, Met $C^{\varepsilon}H_3$, Ile $C^{\delta 1}H_3$ and $C^{\gamma 2}H_3$, Leu $C^{\delta}H_3$ and Val $C^{\gamma}H_3$ groups [173-176], provide versatile probes for conformational dynamics and interactions of large protein systems [24-26]. Methyl labels can be incorporated in proteins by over-expressing proteins in *E. coli* using appropriate precursors (e.g. α-ketobutyrate for Ile, α-ketoisovalerate for Val and Leu) or labeled amino acids (Ala and Met) added to the growth media prior to induction of protein expression; protocols for recombinant expression of deuterated methyl labeled proteins are described in several recent reviews [173-177].

2.2.2 TROSY methods: To achieve maximum sensitivity, protein deuteration and selective isotope labeling are combined with TROSY methods for amide ^{15}NH [22] and methyl $^{13}CH_3$ [23] groups. TROSY NMR experiments exploit favorable relaxation properties of certain transitions in ^{15}NH (AX) and $^{13}CH_3$ (AX_3) spin systems. Thus, slow-relaxing components of ^1H and ^{15}N doublets (Fig. 3A) are selected in the amide TROSY experiment, whose favorable relaxation properties stem from the effect of cross-correlation between ^1H-^{15}N dipole-dipole (DD) and ^1H or ^{15}N chemical shift anisotropy (CSA) relaxation mechanisms [22]. On the other hand, the methyl-TROSY (^1H-^{13}C HMQC) experiment exploits slowly-relaxing components of $^{13}CH_3$ multiplets whose relaxation properties result from interference between intra-methyl ^1H–^1H and ^1H–^{13}C dipole-dipole relaxation mechanisms [23] (Fig. 3B). Overall, amide and methyl TROSY experiments are designed to select and preserve slowly relaxing multiplet components, resulting in improved spectral resolution and sensitivity for large protein systems relative to HSQC experiments, where slowly- and fast-relaxing magnetization components are interchanged in the course of the pulse-sequence.

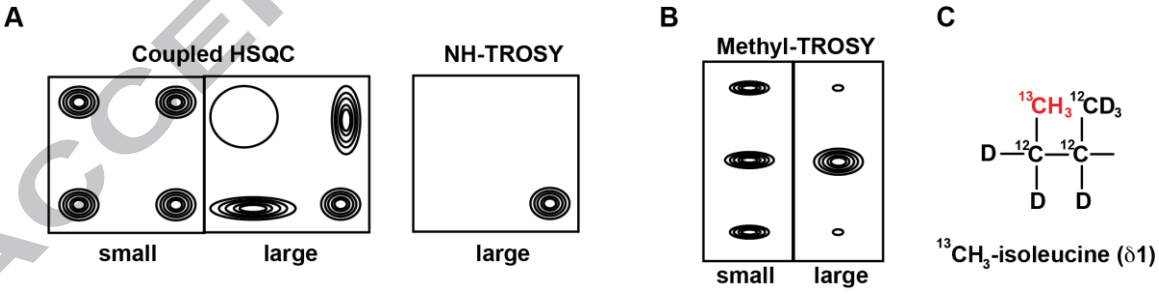


Figure 3. TROSY-based approach for NMR characterization of large systems: (A) Schematics of ^1H-^{15}N-multiplet structures for small and large (ca. >30 KDa) proteins in coupled ^1H-^{15}N HSQC and TROSY spectra. 2D ^1H-^{15}N TROSY-type correlation experiments are employed to obtain pure absorptive spectrum containing only the most slowly relaxing component of the 2D multiplets [22]. (B) Schematics of methyl-TROSY (^1H-^{13}C HMQC) spectra showing ^{13}C-multiplet structure of a methyl group in a small molecule with three ^{13}C MQ multiplet lines visible (left), and in large (ca. >30-40 kDa) highly deuterated protein with intensities of the outer lines reduced as compared to the central line (right). It was show that for proteins with overal rotation correlation times greater than ca. 40 ns, the standard ^1H-^{13}C HMQC pulse sequence is effectively optimized for methyl TROSY spectroscopy without modification [23]. (C) Schematic of U[^2H, ^{12}C], δ_1-[$^{13}CH_3$] Ile with δ_1 methyl group labeled (highlighted in red).

A number of NMR techniques that are commonly used for studies of small proteins, including those discussed in the previous section, have been adapted for studies of large proteins using the TROSY approach (reviewed in [178, 179]). These methods include, for example, ^{15}NH TROSY-based triple-resonance NMR experiments for the backbone resonance assignment [180] and methyl-TROSY based experiments for side-chain assignment [181] of large proteins, 3D and 4D methyl NOESY pulse-schemes [182], ^{15}N TROSY CPMG [124] and $R_{1\rho}$ [145] experiments as well as ^{13}C methyl-TROSY CPMG- [183, 184] and CEST-type [167] experiments for studying µs-ms conformational exchange, longitudinal (ZZ) exchange experiments for studying slow dynamics on the time-scale of seconds [185, 186], and methods for measurement of methyl paramagnetic relaxation enhancements (PREs) [99, 186] and residual dipolar couplings (RDCs) [187] data to characterize structural features of large protein assemblies. These methods are now widely used for characterization of structural properties and dynamics of protein machines assisting protein folding [61, 62, 82, 84, 85, 88-90, 98, 99, 186].

2.2.3 'Divide-and conquer' approach: Residue-specific backbone and side-chain methyl resonance assignment is a bottleneck during structural and dynamic studies of large, multi-component folding machines. Although TROSY-based experiments for the backbone and side-chain NMR resonance assignment of large proteins are available [180, 188, 189], it is often advantageous to simplify complex and large protein systems. This 'divide-and conquer' approach [190] aims to dissect the quaternary multi-component complexes or multi-domain proteins into smaller sized fragments more amenable to NMR resonance assignment using a conventional approach [191]. This strategy has been successfully used for characterization of homo- and hetero-oligomeric folding machines [100, 101, 103, 104, 186], and for studies of multi-domain chaperones [69, 72, 81, 89] and chaperone complexes [88, 99]. Mutations and truncations are widely used to isolate monomeric subunits and individual domains without causing significant perturbation to the structures (or NMR spectra) of these building blocks [173]. Site-directed mutagenesis, selective amino acid-type labeling (or unlabeling), and/or segmental isotope labeling are often used to facilitate transferring NMR assignments from the isolated fragments to the full-size systems (approaches for selective and segmental labeling are discussed in a recent review by Michel & Allain [192]).

2.2.4 Synergy of NMR and other structural methods: Despite recent dramatic improvements, NMR characterization of large, multi-component folding machines remains a challenging and time-consuming task. Moreover, large protein systems are usually not amenable to collection of extensive sets of NMR data, as is required, for example, to solve high-resolution structure of multidomain proteins and large protein complexes. Instead, solution NMR provides a number of site-specific probes to monitor functionally important changes in local protein structure

and conformational dynamics for such large systems. A number of recent studies have successfully demonstrated that a combination of sparse NMR data with other structural biology tools and computational methods provides a powerful means to address challenging, yet exciting questions in the protein-folding field. For example, a combination of NMR, small angle X-ray scattering (SAXS) and X-ray crystallography data was used to obtain the high resolution model of the Hsp90 - substrate complex with Tau [193]; NMR, X-ray crystallography and computational modeling data were combined to characterize individual steps of the Hsp70 functional cycle [81, 83]; a combination of NMR, X-ray crystallography and cross-linking mass spectrometry provided a structural model of a substrate recognition by Eukaryotic chaperonin TRiC/CCT [68]; NMR and cryo electron microscopy (CryoEM) were utilized to explain the mechanism of stress-induced activity of a small heat shock protein HspB5 [96]. These elegant examples highlight a significant yet unexploited potential of NMR methods in characterization of complex and multistep protein folding processes.

3. Folding landscape of small single-domain proteins

Studying folding mechanisms of small single-domain proteins and protein domains is one of the most prominent applications of RD and ST NMR, uncovering the full potential of these methods. These techniques can readily be applied to studies of equilibrium unfolding/refolding interconversions of marginally stable proteins and destabilized protein mutants with unfolding free energies of ~1-3 kcal/mol and folding rates on the μs-ms time-scale [29]. Under these conditions the fractional population of nonnative (intermediate and/or denatured) states is above ~0.5%, so that exchange with these low-populated states accompanied by changes in chemical shifts influences transverse relaxation in the ground native state. Protein stability can be fine-tuned by a suitable choice of mutations that bring it to the range accessible by RD and/or ST methods, which, therefore, are applicable to studies folding energy landscape of a broad range proteins and modular domains with folding rates in the μs-ms range.

RD/ST NMR studies of protein folding can address a range of problems inaccessible to other experimental methods [29]. First, with these NMR methods it is possible to examine delicate intermediate and denatured states under native-like conditions without the use of denaturants, low pH or high temperature to perturb the unfolding/refolding equilibrium. Second, site-specific NMR RD/ST data that report on the folding process are available for a large number of independent probes, in contrast to global measurements such as those provided by circular dichroism (CD) or fluorescence. Analysis of such extensive data sets provides an effective way for studying multi-state folding [34, 129], making RD/ST NMR methods exquisitely sensitive to the presence of transiently populated folding intermediates. Third, RD/ST NMR methods are unique in their capacity to obtain information about the local structure and topology of the minor states in

the form of chemical shifts [112] and, in the case of CPMG RD, residual dipolar couplings (RDCs) [125, 194] and residual chemical shift anisotropies (RCSAs) [195], thus providing the means to explore structural ensembles of nonnative states populated under non-denaturing conditions. Here we discuss the type of questions about protein folding that can be addressed by RD/ST NMR (section 3.1) and briefly discuss a representative example of NMR studies of the folding energy landscape of the FF domain from HYPA/FBP11 (section 3.2).

To date, CPMG RD methods have been widely used for studies of the unfolding/refolding transitions in marginally stable proteins and destabilized mutants [29, 34, 35], yet $R_{1\rho}$ and ST techniques are similarly capable of characterizing such protein systems, and, in fact, many of those techniques were developed and tested on proteins undergoing unfolding/refolding transitions (or their variants) that had previously been characterized by CPMG RD methods. Therefore, in this section we discuss the general capacities of RD/ST NMR methods. We also emphasize how RD/ST methods can be combined with and complemented by other NMR tools (e.g. hydrogen exchange [15, 54]), protein engineering approaches such as Φ-value analysis [1, 196]), and computational methods for modeling protein structures/ensembles from limited NMR data [12, 197-201].

3.1. What can we learn about small protein folding by NMR?

3.1.1 Detection of folding intermediates and unraveling folding mechanisms

Protein folding is often a multi-step process that proceeds through intermediate states, rather than a 2-state interconversion between the two reaction endpoints [15, 16, 29, 34]. The mechanism of the folding process is not known *a priori* and generally needs to be established in the course of data analysis. The sensitivity of RD/ST NMR data for a single nucleus may not be enough for discrimination between 2-state and more complex exchange models [129]. Therefore, a combined analysis of RD/ST NMR data for *multiple probes* (nuclei and/or spin-coherences that report on the same physical process) within the frame of *global exchange models* followed by model selection using statistical criteria is often required to establish the multi-state nature of the processes [34, 129]. Folding is a global event accompanied by large ^{1}H, ^{15}N, and ^{13}C chemical shift changes, which results in significant exchange contributions (R_{ex}) to R_2 for most of the protein nuclei; therefore, it can be queried by multiple independent NMR probes. Availability of a large number of probes and high sensitivity of the data to even small changes of exchange parameters make RD NMR very efficient for the discrimination of 2-state and multi-state folding models and exquisitely sensitive to the presence of weakly populated intermediates with μs-ms lifetimes [29]. Despite the fact RD NMR deals with *equilibrium* interconversions, it is possible to establish the order of states along the folding pathway because, along with kinetic and thermodynamic parameters that describe unfolding-refolding process (exchange rate constants

and populations), NMR chemical shift differences are obtained that provide unique spectroscopic signatures of the exchanging states.

There is increasing evidence that small proteins and protein domains can populate metastable intermediates along their folding pathways [13, 15, 16, 29, 34]. Although conventional chevron analysis of unfolding-refolding kinetics as a function of denaturant concentration using stop-flow technique is often consistent with a 2-state folding mechanism [1], it is clear that some transiently populated states on the folding energy landscape have escaped experimental detection [29]. RD/ST NMR methods [27-30] as well as native state hydrogen exchange (HX) measurements [15, 54] are uniquely capable of detecting such hidden states, thus offering a more detailed view of protein folding. For example, CPMG RD studies revealed the presence of a transient on-pathway folding intermediate for SH3 domains [29, 34, 202], which had for a long time been considered classic examples of small modules that fold by a 2-state mechanism. Denatured, intermediate and transition states on the folding energy landscape capture various stages of the folding process, and, therefore, their characterization holds a great promise for uncovering underlying mechanisms of protein folding.

3.1.2 Properties of nonnative states from unfolding-refolding kinetics

Another advantage of the global analysis of RD/ST data obtained using multiple probes is that very accurate populations of exchanging states, p_i, and exchange rate constants, k_{ij}, can be extracted that provide a description of the process kinetics and thermodynamics [29]. While populations depend on free-energy differences between minor and ground states, ΔG_{ij}, the rates of interconversion depend on activation free energies, $\Delta G^{\dagger}_{i\rightarrow j}$, and report on energy barriers (transition states), so that the extracted parameters can be recast in terms of free-energy profiles along the folding pathway. The errors in ΔG and $\Delta G^{\dagger}$ that ensue from analyzing global NMR RD/ST data are often ~0.05 kcal/mol or less [29, 34, 129], making it possible to explore the effect on stabilities of nonnative sates (denatured, intermediate, transition) of various parameters such as temperature [34, 202-204], pressure [205-207], denaturant concentration [208], side-chain deuteration [209], and mutations [210]. These studies can address different aspects of the organization of nonnative protein states *indirectly* [29], providing insights into (i) changes in polypeptide chain hydration and packing and (ii) formation of native-like and nonnative interactions in the course of protein folding; thus, these studies can help in answering important questions about key folding events, e.g.: At what stage does secondary structure form? At what point do side-chains collapse to form a loosely packed core, and when do their conformations tighten? When is water expelled from the protein core?

3.1.2.1 Temperature: Temperature dependencies of populations and exchange rate constants measured by RD/ST NMR can be recast in terms of enthalpy (ΔH), entropy (ΔS), [34,

202-204] and in favorable cases heat capacity (ΔC_P), [203] profiles along the protein folding pathway. Relatively small changes in free-energy ΔG upon protein folding are typically the result of large compensating enthalpic ΔH and entropic $-T\Delta S$ terms [1]. This entropy-enthalpy compensation upon folding results from (i) a significant loss in configurational entropy of the polypeptide chain accompanied by a decrease in enthalpy due to the formation of favorable contacts between protein groups, and (ii) an increase in solvent entropy caused by contraction of the protein hydration shell containing immobilized water molecules accompanied by an enthalpy increase due to the elimination of favorable interactions in the solvent shell. Thus, ΔH and $-T\Delta S$ profiles along the folding pathway obtained from temperature dependent RD/ST data can be interpreted in terms of changes in polypeptide chain hydration and packing. In favorable cases, it is possible to detect non-linearity of $\Delta G(T)$ (the temperature dependence of ΔG) [203] that reports on heat capacity change ΔC_P. Protein folding is accompanied by negative ΔC_P, since higher heat capacity correlates with increased protein chain hydration [1].

3.1.2.2 Pressure: Pressure is another important variable that can be used to probe volumetric properties of nonnative protein states [29, 205-207, 211], providing additional information on the polypeptide chain hydration and packing. Changes in partial molar volume ΔV and, in favorable cases, isothermal compressibility $\Delta\kappa_T$, can be obtained, respectively, from the slope $(\partial\Delta G/\partial P)_T$ and curvature $-(\partial^2\Delta G/\partial P^2)_T$ of $\Delta G(P)$ [211]. Typically, protein stability decreases with the application of increasing pressure, because the protein native state occupies a larger volume than do less ordered hydrated nonnative states [211]. The increase in partial molar volume upon folding is attributed to (i) the release of water molecules from a densely packed protein solvent shell into the bulk water, and (ii) the presence of packing defects in the native state contributing to its increased *intrinsic* volume as compared to random-coil, which, by definition, does not have packing defects. Protein folding is also accompanied by an increase in isothermal compressibility κ_T [206, 207, 211] that correlates with the size of the protein solvent shell, which is less compressible than the bulk water, and with the presence of packing defects, which make the collapsed protein states more compressible than the disordered ensembles. Note that due to the high accuracy of RD/ST NMR derived populations and rate constants, it is possible to use relatively low pressures of ~100-200 bar [205, 206] that minimize perturbations to the delicate nonnative ensembles.

3.1.2.3 Denaturants: The rate constants and populations obtained from RD/ST data measured as a function of denaturant concentration C can be recast in terms of m-value profiles along the protein folding pathway [1, 208]. The m-value is defined as a slope of free-energy *vs.* denaturant concentration, $m=-d\Delta G/dC$, and provides a measure of change in accessible surface area [1]. Hence, m-value profiles can be interpreted in terms of the polypeptide chain compaction in the course of protein folding. The advantage of RD NMR is that relatively small free-energy

changes can be accurately quantified, allowing the use of mild denaturant concentrations (e.g. up to ~0.5 mM urea or GdnHCl) [208].

3.1.2.4 Deuterium isotope effects: Substitution of protein or solvent protons with deuterons differentially affects stabilities of protein states with different degrees of structural organization, and, therefore, it can be used, in combination with RD/ST NMR, to probe structure formation in various states on the folding energy landscape. For example, the fact that van der Waals interactions involving deuterons are weaker as compared to those of protons means that side-chain deuteration destabilizes protein states where a considerable number of side-chain contacts are formed [209, 212]. Therefore, studies of unfolding-refolding kinetics as a function of the degree of side-chain deuteration can be used to probe the extent of formation of (specific and/or non-specific) side-chain contacts in the nonnative states along the folding pathway [29, 209]. The degree of side-chain deuteration can be altered by introducing protonated aromatic residues (Phe, Tyr, Trp) [209] or protonated Ile, Val, Leu methyl groups in a highly deuterated background [174, 176], achieved by expressing protein in *E. Coli* using ^2H_2O based M9 medium supplemented with protonated precursors (see section 2.2 above). In principle, RD/ST measurements can be also performed in buffers with different ^2H_2O content. Protein stability is known to increase when H_2O is replaced by ^2H_2O, primarily due to strengthening hydrogen bonds involving deuterons as compared to protons [213]. The change in protein stability measured as a function of increasing ^2H_2O content, therefore, can be used to probe the extent of hydrogen bond formation in various states along the folding pathway.

3.1.2.5 Mutations: Structure formation along the protein folding pathway can be studied through the effect of point mutations on unfolding-refolding kinetics (that can be probed by RD/ST NMR) using a protein engineering approach called Φ-value analysis [1, 196]. In this method contacts formed by a residue in nonnative state A (transition or intermediate) are evaluated through the effect of mutating this residue on stabilities of state A and the native state N relative to the denatured state D by calculating the $\Phi_A=\Delta\Delta G_{AD}/\Delta\Delta G_{ND}$ value (here $\Delta\Delta G_{AD}$ and $\Delta\Delta G_{ND}$ are changes in free energy differences $\Delta G_{AD} = G_A-G_D$ and $\Delta G_{ND} = G_N-G_D$ caused by a mutation). $\Phi_A=1$ suggests that native-like contacts are fully formed in state A at the site of mutation, while $\Phi_A=0$ indicates that the site of mutation is likely disordered in state A. Traditional Φ-value analysis of 2-state folding transition states (using stop-flow experiments) has significantly advanced our understanding of protein folding mechanisms [18]. The method can also be applied to studies of multi-state protein folding [210, 214], in which case populations of nonnative states and the rates of their interconversion are converted into Φ-value profiles along the folding pathway that report on both native-like and nonnative structure formation. Another protein engineering approach called double mutant cycles [1, 215] explores the effect of mutating a pair of residues on the stabilities of various states, and can be used to probe the formation of pairwise interactions between residues along the folding pathway.

3.1.3 NMR probes of nonnative state structure

RD/ST NMR data allow extraction of chemical shift differences between states, $\Delta\varpi_{ij}$, that are sensitive to changes in protein structure and flexibility [216, 217]. In addition, a set of CPMG RD experiments performed under conditions of weak protein alignment allow measurements of residual dipolar coupling (RDC) [125, 194] and residual chemical shift anisotropy (RCSA) [195] differences between states that report on changes in orientations of protein structure elements. Chemical shifts obtained by RD/ST NMR can be used for characterization of secondary structure (e.g. by TALOS [218, 219], or δ2D [220] suitable for structure prediction of disordered states) and conformational flexibility (by RCI [221]) of low-populated nonnative states on the folding energy landscape. Furthermore, computational protocols such as CS-Rosetta [197], Cheshire [198] and CS23D [199] enable structure determination of well ordered states using the backbone and $^{13}C^{\beta}$ chemical shifts as primary input data. Computational methods such as restrained molecular dynamics (MD) simulations are also available that employ chemical shifts and other sparse NMR restraints for modeling ensembles of conformationally heterogeneous states [12, 200, 201, 222]. Therefore, the availability of RD/ST NMR derived chemical shifts and additional RDC/RCSA data, along with computational methods for protein structure prediction [197-200], enable modeling of high-resolution structures and/or ensembles of minor states transiently visited in the course of protein folding [43, 223].

3.1.3.1 Minor state chemical shifts: The resonances of minor protein states with populations of a few percent and ms life-times are not observed directly in NMR spectra [29]. Yet NMR spectra of these states can be reconstructed from resonance frequencies of the major species, ϖ_A, and chemical shift differences between states, $\Delta\varpi_{BA}$, derived from RD/ST NMR data: $\varpi_B=\varpi_A+\Delta\varpi_{BA}$ (Fig. 4). By using RD/ST experiments for different nuclei (as outlined above) one can obtain nearly complete backbone and side-chain resonance assignments of a low-population protein state [43, 112, 223]. NMR chemical shifts provide a fingerprint of protein structure containing information that is often sufficient for protein structure generation without the use of any additional data [197-199, 216, 217]. Analysis of CPMG RD data provides absolute values of chemical shift differences, $|\Delta\varpi|$, while $\Delta\varpi$ signs required for reconstruction of NMR spectra of a low-populated state need to be determined from additional experiments [154, 155, 224]. One approach based on the comparison of peak positions in the indirect dimension of HSQC and HMQC spectra recorded at several magnetic fields [154, 224] can provide $\Delta\varpi$ signs for ^{15}N, $^{13}C^{\alpha}$, $^{13}C'$ and methyl ^{13}C nuclei [43, 223]. Another method based on the analysis of off-resonance $R_{1\rho}$ data obtained at low spin-lock field strength using selective $R_{1\rho}$ experiments [141, 154, 155] can be used for measurement of $\Delta\varpi$ signs for ^{15}N, $^{1}H^{N}$, $^{13}C^{\alpha}$ and $^{1}H^{\alpha}$ nuclei [43, 223]. Relative signs of $^{1}H^{N}$ and ^{15}N $\Delta\varpi$ can be determined from ZQ and DQ CPMG experiments [109], so that only one

of the ^{15}N and $^{1}H^{N}$ $\Delta\varpi$ signs needs to be measured independently. In contrast to CPMG RD data, off-resonance $R_{1\rho}$ and CEST profiles obtained at low spin-lock field strengths allow extraction of signed $\Delta\varpi$ values (see above) that can be used to reconstruct NMR spectra of invisible nonnative protein states without the need of additional experiments.

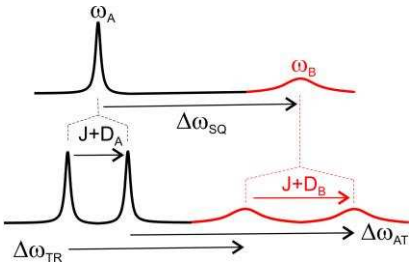


Figure 4. Calculation of ^{15}N chemical shifts and ^{15}N-$^{1}H^{N}$ RDCs in excited states from $\Delta\omega$ values measured in SQ, TROSY (TR) and anti-TROSY (AT) CPMG RD experiments.

3.1.3.1 Minor state RDCs and RCSAs: Residual dipolar couplings (RDCs) and residual chemical shift anisotropies (RCSAs) report on orientations of internuclear vectors and principal axes of CSA tensors, respectively, that can be used as long-range restraints for protein structure calculation [225, 226]. Dipole-dipole interactions between a pair of spins are averaged to zero in isotropic solution due to molecular overall rotation, but can be re-introduced to NMR spectra by placing a molecule in a weakly aligning anisotropic medium such as bicelles [227], bacteriophage solutions [228], polyethylene glycol (PEG) C_nE_m/*n*-alcohol mixtures [229] or stretched polyacrylamide gels [230]. A variety of NMR experiments for RDC measurements in protein ground states were proposed that are routinely used during protein structure determination [226]. More recently, new methods have been described for RDC and RCSA measurements in transient low-populated states that utilize spin-state selective CPMG RD experiments [194, 195, 231]. For example, ^{15}NH RDCs can be obtained from measuring splittings between the components of ^{15}N doublets measured with and without partial alignment, which, in the ground state, can be obtained using a ^{15}NH IPAP experiment [232, 233]. Splittings between doublet components in a low-population invisible state can be obtained by measuring frequency differences between states for individual doublet components using spin-state selective ^{15}N TROSY and anti-TROSY CPMG schemes, thereby providing RDCs for the backbone amide groups in a minor state [231] (Fig. 4). Beyond ^{15}NH groups, CPMG-type RD methods have been proposed for measurements of $^{1}H^{N}$-$^{13}C^{O}$, $^{13}C^{\alpha}$-$^{1}H^{\alpha}$ [194] and methyl $^{13}CH_3$ [234] RDCs in the minor states. Additionally, frequency differences between states for $^{13}C'$ nuclei obtained from CPMG RD data collected with and without partial protein alignment can be analyzed to extract minor state RCSAs for carbonyl carbons [195].

3.1.4 Modeling structure and dynamics of transient nonnative states

3.1.4.1 Chemical shift based structure determination: NMR chemical shifts for protein ^{1}H, ^{15}N and ^{13}C nuclei provide sensitive probes of protein local structure [216, 217]. Yet structural information contained in chemical shifts is not readily accessible because of their complex dependence on a number of factors, which makes chemical shifts difficult to predict and interpret in a quantitative manner. The development of robust empirical methods for chemical shift prediction such as SPARTA [235] [236], SHIFTX [237], SHIFTS [238] and CAMSHIFT [239] that build upon a wealth of chemical shift information available for proteins with known structure has led to a breakthrough in chemical shift based protein structure determination. The fragment-based approaches for protein structure generation such as CS-Rosetta [197], Cheshire [198] and CS23D [199] work by assembly of protein structure from short fragments compatible with experimental NMR chemical shifts. For example, the program CS-Rosetta [197] guides selection of 3- and 9-residue structural fragments from a database by comparing experimental ^{15}N, $^{1}H^{N}$, $^{13}C^{\alpha}$, $^{1}H^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$ chemical shifts with those predicted by the program SPARTA [235] [236]. The fragments are then used to generate three-dimensional models of a protein using the program Rosetta [3]. It has been demonstrated that CS-Rosetta [197] and a related protocol based on the program Cheshire [198] can generate accurate high-resolution models of small proteins and modular domains of up to ~15 kDa with the backbone rmsd to experimental structures of ~1-2Å. The robustness of chemical shift based protocols can be enhanced by the use of long-range restraints such as provided by RDCs and RCSAs [43, 223, 240]. Improved chemical shift based protocols have been recently proposed that make use of additional RDC and RCSA restraints [43, 223, 240].

3.1.4.2 Dynamics from chemical shifts: It has been shown that model-free order parameters, S^2, for the backbone amide groups, which report on the angular amplitude of ^{15}NH vector motions [241], can be predicted with reasonable accuracy from backbone and $^{13}C^{\beta}$ chemical shifts using the Random Coil Index (RCI) approach [221]. This method provides a unique means to probe flexibility of low-population protein states [43, 223] where ^{15}N NMR relaxation measurements commonly used to estimate S^2 of fast ps-ns motions are not available. The protocols for chemical shift based protein structure prediction described above [197-199] generally perform well for structured proteins, while the presence of flexible loops/termini adversely affects the convergence of structure calculation. At first glance it may seem that studies of low-populated folding intermediates would necessarily require their description in terms of ensembles of interconverting conformations, limiting the use of CS-Rosetta [197] and Cheshire [198] protocols for their structure calculation. However, recent studies of the FF domain from HYPA/FBP11 [43, 223] and a mutational variant of Fyn SH3 domain [44] have shown that intermediates may consist of ordered regions with native-like and/or nonnative secondary and tertiary structure well described by a single conformation. The RCI approach for prediction of

conformational flexibility from NMR chemical shifts [221] can be used to map such ordered regions of folding intermediates amenable to structure determination using chemical shift based protocols.

3.1.4.3 Modeling conformationally heterogeneous ensembles: Nonnative high-energy states sampled in the course of protein folding exhibit various degrees of conformational flexibility and often need to be described as ensembles of interconverting conformations [12, 200, 201, 222]. Modeling structural ensembles of conformationally heterogeneous states is a challenging task because only limited experimental data are typically available for such states, yet the number of data needed to adequately describe the ensemble of conformations is much greater than that required to determine a single structure. One potentially general method for modeling heterogeneous states is based on the use of various types of experimental data, including but not limited to NMR chemical shifts and RDCs, as restraints in molecular dynamics (MD) simulations [200, 222]. Since experimental data report on ensemble averages over a range of conformations, the rigorous approach involves imposing restraints on an ensemble of conformations sampled in one MD trajectory (time-averaged restraints) [242] or in parallel MD trajectories for a number of molecule replicas (ensemble-averaged restraints) [200, 222, 243]. For example, MD simulations were employed for the characterization of folding intermediates using chemical shifts and HX protection factors [244], and transition states using protein engineering Φ-values as restraints [245]. Another approach that was used for modeling disordered protein states involves generation of a large pool of conformers and assignment of their statistical weights to satisfy experimental restraints [12, 201], as, for example, implemented in ENSEMBLE [246, 247] or ASTEROIDS [248, 249] algorithms.

3.1.4.4 Probing structural organization of nonnative states: Chemical shifts, RDCs and RCSAs obtained by RD/ST NMR provide unique experimental data that enable determination of the structure and dynamics of sparsely populated nonnative states on the folding energy landscape. With this information it is possible to address many questions about structural organization of these obscure species that are often beyond the reach of other experimental techniques. About denatured states - what is the extent of their residual native-like and/or nonnative structure, and what is this residual structure? What is the role of residual structure (if any) in driving the polypeptide chain toward downstream conformational states en route to the native state? About folding intermediates - how structured are they, and what is their structure? What interactions are responsible for their stability, and what is the nature of barriers that separate them from native and denatured states? What is the extent of nonnative interactions, and what is their role in determining folding rates? Are intermediates molten-globule-like collapsed states with partially formed secondary structure held together by non-specific side-chain interactions, or do they consist of well structured regions with packed cores lacking side-chain fluidity?

Many of these questions can be answered from RD/ST derived chemical shifts even before attempting to determine structure or model ensembles of transient intermediates and denatured states. As outlined above, the backbone chemical shifts can be used to assess conformational dynamics (using RCI [221]) and predict secondary structure (using TALOS [218, 219] or δ2D [220]) of nonnative states on the folding energy landscape. The later analysis, for example, can be used to identify nonnative secondary structure elements. In cases where continuous ordered regions of protein sequence with high RCI order parameters S^2 are found, the structure of such regions can be modeled by CS-Rosetta [197, 240] and/or Cheshire [198] based on the backbone chemical shift, RDC and RCSA restraints [43, 223]. This situation may likely happen for protein folding intermediates. If a well-defined rigid structure is lacking, which might be expected for denatured states, MD simulations with ensemble-averaged restraints or other methods for modeling ensembles of conformationally heterogeneous states [12, 200, 201, 222] can be employed.

3.1.5 Isolation and further analysis of nonnative states on the folding energy landscape

Structural models of low-populated nonnative states obtained using RD/ST NMR data in combination with computational approaches can help design rational strategies for isolation of such states [43, 223]. Specifically, these models can be used to identify amino-acid residues and/or whole regions of protein structure that do not contribute significantly to the nonnative state stability, but are critical for structural integrity of the native protein. Mutation of such residues and/or truncation of the identified regions, therefore, can selectively destabilize the ground native state, making a nonnative state (e.g. folding intermediate) the most stable species on the folding energy landscape. This strategy, for example, has been successfully applied to isolation of the folding intermediate of the FF domain from HYPA/FBP11 based on a high-resolution model of this state obtained by RD NMR [43, 223]. A similar approach has also been also employed to trap folding intermediates of apo-cytochrome b_{562} [250], engrailed homeodomain En-HD [251] and Im7 protein [252], albeit using a different strategy for mutant design.

Once isolated, nonnative protein states can be further investigated using conventional NMR approaches and other biophysical methods. For example, characterization of their structure using a standard NOE-based NMR approach [253] and ps-ns dynamics using NMR spin-relaxation measurements [241] can help corroborate the accuracy of a structural model based on chemical shifts, RDCs and RCSAs obtained by RD/ST NMR. Furthermore, it might be possible to explore transitions to higher-energy states on the protein energy landscape by using RD/ST NMR [27-30] and/or hydrogen exchange (HX) experiments [15, 54] performed on an isolated species. Folding intermediates have long been considered as branching points of the protein folding pathways that, beyond their roles in protein self-assembly, initiate the formation of molecular

aggregates [8, 9]. Therefore, exploring the tendency of isolated intermediates to form oligomeric species can shed light on early events of protein aggregation.

3.2. Folding energy landscape of HYPA/FBP11 FF domain by RD NMR: a case study

To date, μs-ms unfolding-refolding transitions leading to formation of transient nonnative conformational states have been studied using the RD/ST NMR approach (primarily by CPMG RD) for a number of small proteins and protein domains, e.g. Fyn and Abp1p SH3 domains [34, 44, 202] (reviewed in [29]), FF domain from HYPA/FBP11 [43, 208, 223, 254-256], redesigned apo-cytochrome b_{562} [203, 206, 250, 257], the activation domain of human procarboxypeptidase A2 (ADA2h) [204], cold shock proteins B (CspB) [258], Im7 protein [252, 259], gpW protein [260], KIX protein [261, 262], PBX homeodomain [263, 264], villin headpiece domain [265-267] and apomyoglobin [268]. For proteins such as Fyn SH3 domain (reviewed in [29]) and the FF domain from HYPA/FBP11 (reviewed here), the RD NMR approach resulted in a remarkably detailed picture of the folding energy landscape, as well as structure determination of the transiently populated on-pathway folding intermediates [43, 44, 223]. A case study of the FF domain folding described in section 3.2.1 illustrates all steps of the RD/ST NMR analysis of transient nonnative states on the folding energy landscape, from probing the details of the folding pathway based on unfolding-refolding kinetics to structure determination of the transiently populated folding intermediate, its isolation and further analysis.

3.2.1 Details of the FF domain folding pathway from unfolding-refolding kinetics: The first FF domain from human HYPA/FBP11 is a small 71 residue four-helix module with α-α-3_{10}-α topology that provides a model system for studying protein folding mechanisms [43, 208, 223, 269-271] (Fig. 5). Previous studies have shown that the FF domain folds by a three-state mechanism $U \leftrightarrow I \rightarrow F$ (U is fully unfolded, I is intermediate and F is folded) *via* an on-pathway intermediate that is close in energy (~2 kcal/mol) to the natively folded state [270, 271]. The intermediate forms rapidly on the μs time-scale and rearranges to the native state within a few milliseconds [43, 270, 271] (Fig. 5). The folding behavior of HYPA/FBP11 FF domain has been examined to a considerable level of detail by RD NMR methods [43, 208, 223]. ^{15}N CPMG RD measurements performed for the wild-type (WT) FF domain revealed that in the temperature range of 20-35^{o}C the folding intermediate is populated at 1-5%, while the unfolded state is below the level of detection. For the destabilized mutants A17G and Q19G, on the other hand, ^{15}N CPMG RD data allowed detection of both intermediate and unfolded states [208]. The analysis of ^{15}N CPMG RD data for the WT, A17G and Q19G FF domains recorded as function of temperature and urea concentration provided the populations of the exchanging states and exchange rate constants that were recast into free-energy, entropy, enthalpy (temperature dependent data) and m-value (urea dependent data) profiles along the FF domain folding

pathway (Fig. 6A,B). These profiles suggest that the FF domain folding proceeds in several consecutive structure formation / compaction / dehydration steps accompanied by a gradual decrease in enthalpy, entropy and m-value (Fig. 6A,B), and that the folding intermediate exhibits a level of structural organization between those of the folded and the fully unfolded states [208]. The initial analysis of CPMG RD derived ^{15}N chemical shifts in the intermediate state, as well as protection factors for the backbone amide groups obtained from HX measurements (Fig. 6C), revealed that the region corresponding to the first three native helices of the FF domain is relatively well structured in the intermediate state while the C-terminal region corresponding to the fourth native helix is more disordered [208]. Subsequent RD NMR studies have also measured the FF domain unfolding-refolding kinetics as a function of viscosity, providing important insights into the impact of solvent friction on the FF domain folding rate [255, 256].

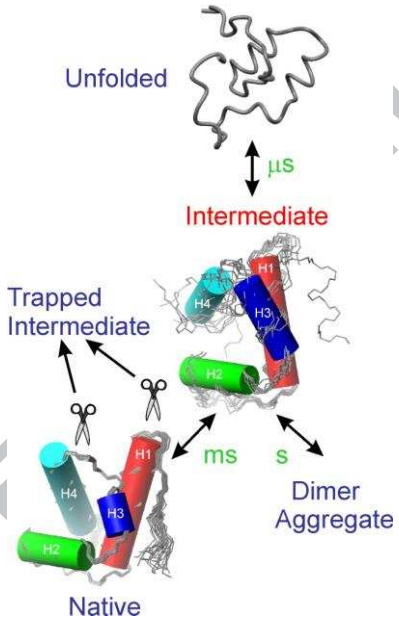


Figure 5. Folding pathway of the WT FF domain from HYPA/FBP11, structures of the native and intermediate states, and strategy for isolation of the intermediate. Adapted with permission from reference [43].

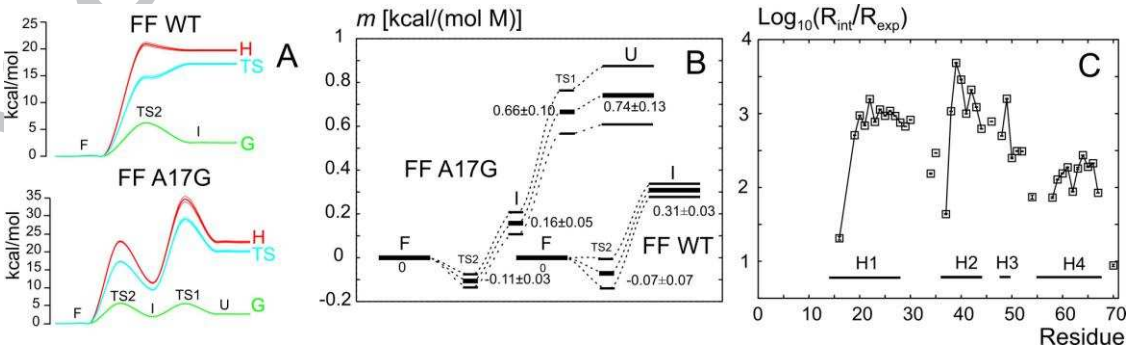


Figure 6. (A) Free-energy G, entropic TS and enthalpic H contribution to G along the folding pathway of the WT and A17G FF domain obtained from temperature dependent ^{15}N CPMG RD data. (B) m-value profiles along the folding pathway of WT and A17G FF from urea-dependent ^{15}N CPMG RD data. (C) Protection factors for the backbone amide groups of the WT FF domain obtained from HX data, Adapted from reference [208].

3.2.2 Structure of the FF folding intermediate by RD NMR: Structural analysis of the folding intermediate of the WT FF domain (Fig. 5) was based on a nearly complete set of CPMG RD derived backbone ^{15}N, $^{1}H^{N}$, $^{13}C^{O}$, $^{13}C^{\alpha}$ and $^{1}H^{\alpha}$ chemical shifts and ^{15}N-^{1}H RDCs measured for the domain weakly aligned in a PEG/hexanol mixture [43]. Chemical shifts obtained from the CPMG RD data were used initially to assess secondary structure and dynamics of the folding intermediate using, respectively, TALOS+ [218] and RCI [221]. Order parameters S^2 for the backbone amide groups predicted by RCI approach revealed that the folding intermediate is remarkably well-structured, and rather represents an alternative less favorable fold of the FF domain. Secondary structure prediction revealed that, similar to the FF domain native state, the folding intermediate consists of the four helices (H1-H4 in Fig. 5), with the first two helices having identical boundaries with corresponding helices of the native state. The third helix H3 of the intermediate is nonnative (blue in Fig 5), spanning the region that forms the 3_{10}-helix H3, the H3-H4 loop, and the beginning of α-helix H4 of the native protein. Additionally, the intermediate has a poorly defined nonnative helix H4 in the C-terminal part (cyan in Fig. 5). Following the initial characterization, the three-dimensional structure of the folding intermediate was modeled using a modification of the CS-Rosetta protocol using RD NMR derived backbone chemical shift and RDC data [43], (Fig 5). Although the native-like topology is preserved in the intermediate state, the obtained high-resolution model reveals nonnative secondary and tertiary structure that must be broken on way to the native state (Fig. 5), slowing down protein folding by about two orders of magnitude. The obtained structure offered a simple route to isolation of the folding intermediate by truncation of the C-terminal part of the domain [43], which is at least partially disordered in the intermediate, but forms an α-helix H4 in the native protein critical for structural integrity of the FF domain (Fig. 5). The truncation, therefore, selectively destabilizes the native state and traps the domain in the intermediate state, which becomes the most stable species on the folding energy landscape.

3.2.3 Nonnative interactions in the FF domain folding pathway: It is increasingly recognized that protein folding is not just a progressive accumulation of the native-like structure, but a complex process involving the formation of nonnative interactions that may speed up the folding by lowering energy barriers [272-274] or trap the molecules in misfolded conformers [275]. The folding intermediate of the FF domain provides an example of such a trap stabilized by nonnative contacts that creates a rate-limiting barrier on the way to the native fold (Fig. 5). To investigate the role of nonnative interactions in the folding of FF domain, CPMG RD studies have been extended to its L24A mutant, which has an overall folding rate an order of magnitude slower than the WT domain and an unusual Φ-value of 1.3 in the folding transition state [271], consistent with the formation of nonnative interactions. The structure of the folding intermediate of the L24A

FF domain determined by RD NMR methods revealed an even more extensive network of nonnative interactions than that of the WT intermediate (Fig. 7). Thus, a nonnative interface between helices H1 and H3 in the folding intermediate of the L24A FF domain is additionally stabilized by nonnative contacts formed by residue Y49 from H3, which in the WT FF intermediate is located on the interface between helices H2 and H3 (Fig. 7). These findings suggest a key role for nonnative interactions in determining the FF domain folding rate [223].

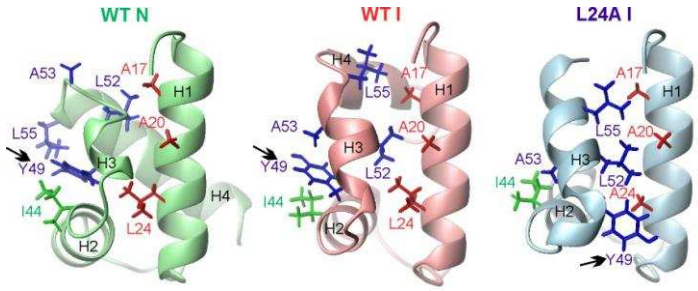


Figure 7. Nonnative interactions in the WT and L24A intermediates (**I**) *vs.* native (**N**) FF fold. Extra nonnative contacts in L24A **I** are formed by Y49. Adapted with permission from reference [223].

3.2.4 Intermediate as a branching point in the FF folding pathway: The formation of partially folded intermediates has long been thought of as a primary step in the onset of aggregation leading to the formation of cytotoxic molecular species [8, 9]. However, due to their transient nature and low populations, the involvement of folding intermediates in aggregation has been extremely difficult to verify. NMR studies of the folding intermediate of the FF domain provided a rare opportunity to explore the role of such states in the formation of higher-order molecular species (Fig. 5). Thus, a ^{1}H-^{15}N HSQC spectrum of a variant FF_{1-60} (residues 1-60) with the truncated C-terminal part [43] that mimics intermediate of the full-length FF_{1-71} (residues 1-71) displays two sets of signals, revealing the presence of two conformations of the protein in slow exchange with one another. The interconversion between the two conformers was studied by ^{15}N longitudinal exchange (ZZ) experiments [276], a dilution series and ^{15}N relaxation measurements [241], revealing that the two sets of signals correspond to monomer and dimer forms of the isolated folding intermediate. Analysis of the backbone ^{1}H, ^{15}N and ^{13}C chemical shifts of the two forms of FF_{1-60} assigned using standard triple-resonance NMR experiments [191] suggests that dimerization involves the flexible N-terminal part of the domain preceding the α-helix H1. Truncation of this part resulted in another mimic of the folding intermediate - FF_{11-60} (residues 11-60) that predominantly exists in the monomeric form (Fig. 5). This study established that the folding intermediate serves as a branching point of the FF domain folding pathway, illustrating how incomplete folding can result in the formation of higher-order molecular assemblies [254].

3.2.5 Cross-validation of the RD NMR based structure determination: The structure of the folding intermediate of the WT FF domain is the first high-resolution model of a nonnative protein state with population of a few percent and a ms lifetime to have been determined by RD NMR. Therefore, its corroboration by other methods could provide an important validation of the RD NMR based approach for studying transient nonnative protein states. The isolated folding intermediate obtained by truncation of the C-terminal part of the domain (FF_{11-60}) has provided an important means for such validation. The structure of the monomeric truncated variant FF_{11-60} that mimics the intermediate of the full-length domain has been determined using a standard NOE-based NMR approach, and has confirmed the set of nonnative contacts observed in RD NMR derived model of the intermediate [277].

4. Protein folding in a living cell

While the detailed characterization of protein folding *in vitro* can provide insights into structural, kinetic and thermodynamic characteristics of interconverting states on the protein folding energy landscape, the protein folding process in living cells is significantly more complex than protein folding *in vitro* [278]. For the majority of *in vitro* studies, refolding is initiated from a full-length protein, either by dilution of a denatured protein into a refolding buffer or by using metastable variants of the wild-type protein that populate both the folded state and significant amounts of fully and/or partially unfolded conformers. Because protein folding *in vivo* occurs under very different conditions, growing evidence suggests that *in vivo* folding is significantly altered [279], both co- and post-translationally. Furthermore, because protein synthesis is energetically very expensive and misfolded and aggregated species are often extremely cytotoxic, cells have evolved robust and efficient protein quality control (PQC) networks in the cytoplasm and organelles to ensure that proteins stay in the correct conformations during their entire lifetime [21]. Interactions with molecular chaperones and other PQC components create yet another level of complexity, allowing protein homeostasis to adapt in complicated ways to physiological and pathological changes in different aspects of cellular environment [280]: e.g. fluxes of metabolites and ions, interactions with other macromolecules, stresses, viral infections, etc. The highly dynamic and partially redundant nature of the PQC networks is essential for the plasticity and efficiency of *de novo* protein folding and repair of damaged proteins in a complex cellular environment [278]. In turn, an imbalance in cellular folding capacity is associated with many devastating diseases, including neuro-degeneration, diabetes and cancer [280]. The mechanistic understanding of *in vivo* protein folding is a central, yet extremely challenging, area in modern molecular biology and biomedicine. Advances in bio-NMR reviewed in section 2 have recently allowed characterization of the key aspects *of in vivo* protein folding at atomic resolution, as well providing high-resolution snapshots of the folding process in the presence of molecular

chaperones and ribosomes. In this section we overview recent examples of NMR studies that successfully address these fundamental questions.

4.1 De Novo Protein folding

It has been postulated that for many proteins *in vivo* folding starts while a polypeptide chain is still tethered to the ribosome [279]. Indeed, during translation, the N-terminal end of the polypeptide chain is available for folding before the C-terminal end has been synthesized. Gradual release of the nascent polypeptide chain from the ribosome, as well as interaction with the ribosome-associated chaperones, constrains the available conformational space and alters the folding pathways observed *in vitro* [278]. On the one hand, the vectorial nature of translation might prevent nonspecific interactions between the N- and C-terminal sequences, potentially reducing the chance of misfolding and aggregation for some proteins. On the other hand, co-translational folding may be time-constrained because the long-range contacts most likely occur outside of the ribosomes and require the release of large segments of the nascent chain [279], suggesting that the rate of translation might fine-tune co-translational folding [281, 282]. In other words, the translation pattern might kinetically trap folding intermediates and determine preferential routes of the folding process, either driving or preventing off-pathway misfolding reactions.

Detailed structural and dynamic studies of protein folding on the ribosome are complicated by the high molecular weight of the polypeptide chain-ribosome complex (~2.5 MDa), limited lifetime of the complex and relatively low working concentrations (ca. 100 nM - 10 μM). Advances in NMR methodology for studying large protein systems have recently allowed many of these challenges to be overcome, making possible NMR characterization of three small global proteins, FLN5 [56], barnase [57] and an SH3 domain [58] tethered to the ribosome. These studies demonstrated the power of solution NMR to probe conformation and dynamics of the nascent chain during translation (as discussed in detail in a recent review by Waudby et al. [283]). Here, we overview the examples of two most recent studies by Christodoulou and colleagues (published in 2016), elucidating the mechanisms by which interactions with the ribosome [59] and ribosome-associated molecular chaperones [60] shape the folding landscape of the nascent polypeptide chain.

4.1.1 Nascent chain interactions with the ribosome surface: The first study [59] investigated the structural ensemble of an *E. coli* ribosome-nascent chain (RNC) complex. This nascent chain (NC) included sequences of two immunoglobulin-like domains: the N-terminal 105-residue FLN5 that can adopt a folded conformation and a folding-incompetent C-terminal 110-residue FLN6 (Fig. 8A). The NC was tethered to the ribosome and translationally arrested using the SecM motif. Signals from the folded and unfolded parts of the NC were observed using ^{1}H-^{15}N

SOFAST-HMQC experiments in H_2O with U{^{15}N,^{2}H} labelled RNC samples, which detect amide group signals from unfolded regions of the NC, and ^{1}H-^{13}C SOFAST-HMQC experiments in D_2O with RNC samples selectively labeled with $^{13}C^{1}H_3$ methyl groups at all Ile δ-methyl groups, which detect well-dispersed signals from the folded FLN5 domain. FLN5 and FLN6 chemical shifts were used as replica-averaged structural restrains for modeling the conformational ensemble of the RNC complex using molecular dynamics (MD) simulations. The generated ensemble (Fig. 8B) revealed that N-terminal folding-competent FLN5 adopts a folded conformation and has some limited access to the ribosome surface. On the other hand, C-terminal folding-incompetent FLN6 adopts a compact disordered conformation with transiently formed inter-residue contacts and fractionally populated native-like secondary structure elements (ca. 20% on average). As expected, FLN6 also transiently interacts with several ribosomal proteins.

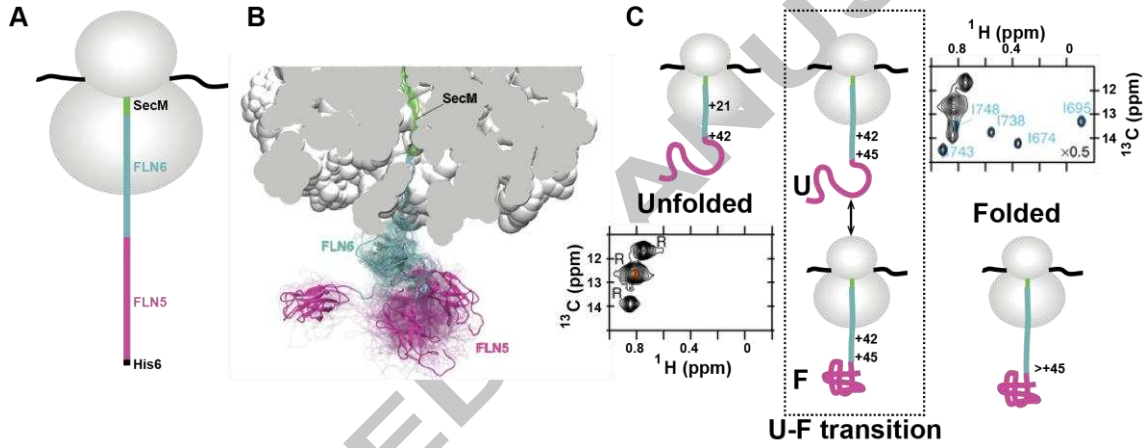


Figure 8. Nascent chain folding on the ribosome: (A) Design of the SecM translationally arrested RNC used for NMR characterization, in which N-terminal FLN5 domain (magenta) is tethered to the ribosome through C-terminal FLN6 domain (cyan). (B) Three representative conformations of the nascent-chain ensemble of FLN5+FLN6 RNC obtained from MD simulations using NMR chemical shifts as replica-average structural restrains (reproduced with permission from [59]). (C) FLN5 folding is sterically delayed by interaction of the FNL6 tether with the ribosome: The FNL5 folded and unfolded states observed for RNCs with different lengths of truncated FLN6 (from 21 to 110 residues) as monitored by ^{1}H-^{13}C correlation spectra of Ileδ1-[$^{13}CH_3$] labeled FLN5 RNCs (black); the spectra of RNC overlaid with isolated folded (cyan) FLN5 and unfolded (orange) truncated FLN5; resonances marked R arise from background labeling of ribosomal proteins (reproduced with permission from [59]).

To investigate how interactions with the ribosome affect the FLN5 folding, the original NC construct, which included folding-competent FLN5 and 110-residue folding-incompetent FLN6 (as a tether), was gradually truncated from 110 to 21 C-terminal FLN6 tether residues. Peak intensities in amide and methyl 2D NMR spectra were used to monitor how the length of the tether (the C-terminal folding-incompetent FLN6 fragment) affects the population of folded FLN5 (Fig. 8C). Interestingly, the midpoint of the unfolded-to-folded transition of FNL5 was observed only when the FLN6 tether was around 42 to 45 residues long (Fig. 8C), while the first 31 residues of the FLN6 tether were sufficient for the entire FLN5 sequence to emerge from the ribosome (as reported by PEGylation of a Cys residue located on the boundary between FLN5 and FLN6). These results evidently demonstrated that emergence of the entire FLN5 sequence

from the ribosome tunnel is not sufficient for FLN5 folding. Indeed, when the tether (truncated C-terminal FLN6) is short enough (31-45 residues) to permit interactions of as yet unfolded FLN5 with the ribosome surface, FLN5 folding is sterically destabilized and delayed. In turn, a longer tether (over 42-45 residue long) sterically decreases contacts between as yet unfolded FLN5 and the ribosome and thus enables FLN5 folding.

4.1.2 Nascent chain interactions with the ribosome-associated chaperones: In vivo, a nascent chain tethered to the ribosome has considerable access to ribosome-associated molecular chaperones such as the ATP-independent trigger factor and ATP-dependent Hsp70 system in bacteria, and the ATP-independent nascent polypeptide-associated complex and GimC/Prefolding, ATP-dependent Hsp70 and Hsp60 (TRiC/CCT) chaperones in eukaryotes [279, 282, 284]. The second study by Christodoulou and colleagues [60] investigated how interaction with the ribosome-associated trigger factor (TF) affects the RNC conformation.

As a model system to probe such interactions, the authors designed an NC construct that consists of intrinsically disordered α-synuclein translationally arrested by the SecM motif. There is substantial previous evidence that the TF has moderate (μM) affinity for the ribosome, as well as weak affinity for the ribosome-tethered α-synuclein. In this study [60], different isotope labeling schemes for α-synuclein and TF allowed the authors to depict and characterize complex individual interactions between TF, NC, and the ribosome: ^{1}H-^{15}N SOFAST-HMQC was used to monitor ^{15}N-labeled α-synuclein NC, while ^{1}H-^{13}C SOFAST-HMQC was used to monitor methyl (^{2}H, $^{13}CH_3$-ILV) labeled TF. In agreement with the previously detected TF-ribosome association, a significant reduction in peak intensities (ca. six fold) was observed in ^{1}H-^{13}C SOFAST-HMQC of TF in the presence of unlabeled ribosomes. In turn, inhomogeneous line broadening in ^{1}H-^{15}N SOFAST-HMQC of the α-synuclein RNC complex suggested that α-synuclein tethered to the ribosome transiently interacts with the ribosome surface (Fig. 9A). The addition of TF to the α-synuclein RNC complex resulted in further non-uniform reduction in intensities of α-synuclein amide peaks in ^{1}H-^{15}N SOFAST-HMQC, reporting on weak interaction between the ribosome-tethered α-synuclein and TF.

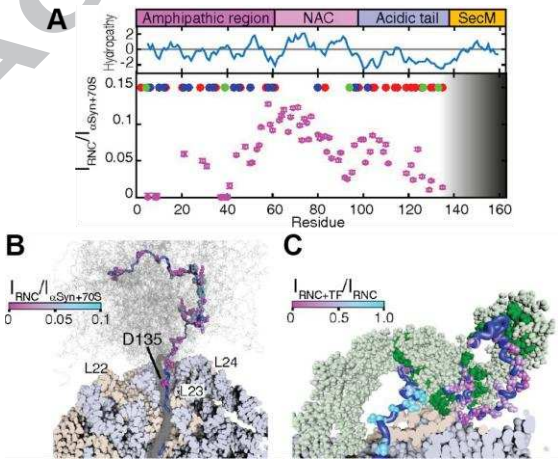


Figure 9. α-synuclein nascent chain interactions with the ribosome and ribosome-associated trigger factor: (A) Inhomogeneous line broadening in ^{1}H-^{15}N SOFAST-HMQC of the α-synuclein RNC complex suggesting α-synuclein nascent chain with the ribosome: Cross-peak intensities in the α-synuclein RNC (I_{RNC}) relative to isolated α-synuclein in the presence of ribosomes ($I_{\alpha Syn+70S}$); the gray-shaded area depicts the approximate length of the exit tunnel. A Kyte and Doolittle hydropathy plot and an amino acid classification of the α-synuclein sequence in positively charged (blue), negatively charged (red), and aromatic (green) residues are also shown (reproduced with permission from [60]). (B) MD-generated structural modeling of the α-synuclein RNC: Cross-section of simulated αSyn RNC showing the nascent chain ensemble (gray) with a representative nascent chain structure (blue), ribosomal proteins (beige), and RNA (blue-

gray) highlighted. Observed NMR resonances (relative to isolated αSyn in the presence of 70S, Fig. 9A) are shown as spheres colored according to their relative intensity (reproduced with permission from [60]). (C) The MD-generated α-synuclein RNC model with trigger factor monomer (light green) and substrate binding pockets (dark green); the observed NC resonances are shown as spheres colored according to relative intensities of the α-synuclein RNC (I_{RNC}) following addition of the trigger factor (I_{RNC+TF}) (reproduced with permission from [60]).

Next, to develop structural models of the α-synuclein RNC complex in the presence and absence of TF, the authors utilized coarse-grained molecular dynamic (MD) simulations. As a starting point, a previously obtained cryo-EM model of SecM-stalled translating ribosome and an NMR model of the TF in complex with a disordered protein substrate PhoA were used. Remarkably, the MD-generated structural ensembles of RNC (Fig. 9B,C) agreed very well with non-uniform reduction in amide peak intensities in ^{1}H-^{15}N SOFAST-HMQC spectra of α-synuclein observed upon TF association and/or tethering to the ribosome. The generated structural ensembles provided novel insights into interactions of the disordered NC with the ribosome. Specifically, (i) these interactions are found to be transient and apparently driven both by charged and by aromatic NC residues; and (ii) these transient NC-ribosome interactions do not directly affect the disordered NC conformation. Taken together, these findings by Christodoulou and colleagues provide solid experimental evidence that NC interactions with the ribosome surface and TF play an important 'shielding' role for as yet unfolded NCs, altering the protein folding landscape (Fig. 10) and apparently protecting folding-competent NCs from nonnative intra- and intermolecular contacts and misfolding.

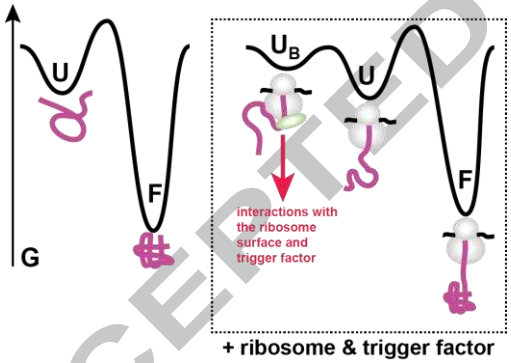


Figure 10. Co-translational modulations of the protein folding energy landscape: (Left) folding free-energy diagram for an isolated protein, showing the difference in free energy (G) between the folded (F) and unfolded (U) states. (Right) free-energy diagram for a protein tethered to the ribosome, showing a ribosome- and trigger-factor bound state (U_B). The ribosome and ribosome-associated trigger factor may alter this landscape and inhibit nascent-chain folding, as indicated by red arrow.

4.2 The role of protein quality control network in protein folding

Protein quality control (PQC) includes a sophisticated network of molecular chaperones, degradation enzymes, and other highly effective protein machines that facilitate correct folding of nascent protein chains and prevent accumulation of misfolded and aggregated protein species [21]. This network maintains protein homeostasis under physiological and/or pathological conditions [285, 286]; any imbalance in its capacity increases the risk of diseases associated with protein misfolding and aggregation [285, 287].

Molecular chaperones, as the key PQC players, facilitate protein folding, prevent and reverse misfolding, and target damaged proteins for degradation. To perform their functions,

chaperones recognize and bind to their protein substrates in a cyclic manner. Many chaperones (called ATP-dependent molecular chaperones, e.g., Hsp90s, Hsp70s, Hsp60s) use energy from ATP hydrolysis to control their substrate binding and release cycle [288, 289]. Other chaperones (called ATP-independent chaperones, e.g., trigger factor, small HSPs, periplasmic bacterial chaperones, etc.) perform their functions without the use of ATP energy [290, 291]. Despite significant recent progress in the field, molecular mechanisms underlying the complex dynamic action of molecular chaperones still remain elusive. Advances in structural and dynamic studies of chaperone machines by solution NMR are summarized in recent reviews [55, 193, 292]. Here, we give an overview of recent NMR studies of Hsp70 molecular chaperones addressing the following fundamental questions: How do Hsp70 bind to a variety of different substrates, regardless of their conformations and shapes? How do Hsp70s fine-tune the folding landscape of their substrate proteins?

4.2.1 The role of Hsp70 conformational flexibility in substrate binding: An important feature of Hsp70 molecular chaperones is their ability to interact with a variety of different protein substrate conformations, including completely extended nascent chains tethered to the ribosome, partially not-yet-folded or/and misfolded intermediates populated along the protein folding pathway, misfolded or partially unfolded species accumulated under stress conditions, as well as large insoluble aggregates [293-295]. Moreover, all Hsp70 substrates share only one common feature, an Hsp70 recognition motif [296]. This motif comprises five solvent-exposed preferentially hydrophobic amino acids usually framed by positively charged residues. Because this motif occurs very frequently in the hydrophobic core of globular proteins, Hsp70s recognize a wide variety of different protein substrates, and solvent accessibility of the recognition motif (e.g., upon protein unfolding or misfolding) is usually sufficient for interaction with Hsp70, regardless of the protein substrate size, shape or other conformational characteristics.

Hsp70s bind and release their substrates in a highly coordinated manner, regulated by ATP binding and hydrolysis. All Hsp70s consist of two domains, an N-terminal 40 kDa nucleotide-binding domain (NBD) that binds and hydrolyses ATP and a 30 kDa substrate-binding domain (SBD) that recognizes and binds protein substrates. Recent NMR [81] and X-ray studies [297] [298] of bacterial Hsp70 revealed that the complex mechanism of controllable substrate binding and release relies on drastic changes in domain arrangement between the ADP- and ATP-bound states. When bound to ADP, the two Hsp70 domains, SBD and NBD, behave independently, resulting in a high substrate binding affinity of SBD. By contrast, in the ATP-bound state, the two domains dock tightly onto each other, resulting in dramatic changes in SBD conformation, apparently incompatible with the substrate binding [297, 298]. In turn, NMR characterization of the Hsp70 functional cycle [81] revealed that, in the presence of ATP and substrate, Hsp70 adopts a partially docked, highly dynamic conformation with a low substrate binding affinity.

Intriguingly, ATP binding not only results in significant structural changes, but also dramatically increases conformational dynamics around the substrate binding site, as evidenced by the large B-factors in X-ray structures [297, 298] and a significant line-broadening in amide TROSY spectra [81, 92]. These findings raise the following question: How important is this conformational flexibility for substrate binding and release? A recent characterization of bacterial Hsp70 employed solution NMR and MD simulations to address this intriguing question [83]. In this study, key Hsp70 allosteric sites that enable fine-tuning of the Hsp70 conformational transitions were perturbed using site-directed mutagenesis and protein truncations. Long-range effects upon these allosteric perturbations were monitored by chemical shift changes in amide TROSY and HNCO spectra, as well as conformational and dynamic perturbations in the MD trajectories [83]. These NMR and MD data revealed that conformational flexibility of the substrate-binding site is tightly coupled with the domain rearrangements (Fig. 11). ATP-induced domain docking leads to a dramatic enhancement of conformational flexibility in the substrate-binding site. In turn, enhanced dynamics in the substrate-binding site (e.g. upon destabilizing mutations) favors domain docking. While 'rigid' X-ray structures suggested that the domain-docked, ATP-bound Hsp70 state is incompatible with substrate binding [297, 298], the conformational flexibility in the substrate-binding site provides the mechanism that enables substrate binding to this open and flexible SBD conformation (Fig. 11). Moreover, this flexibility apparently allows the substrate-binding site to sample a wide variety of different conformations and, thus, to adjust to protein substrates with different sizes and shapes.

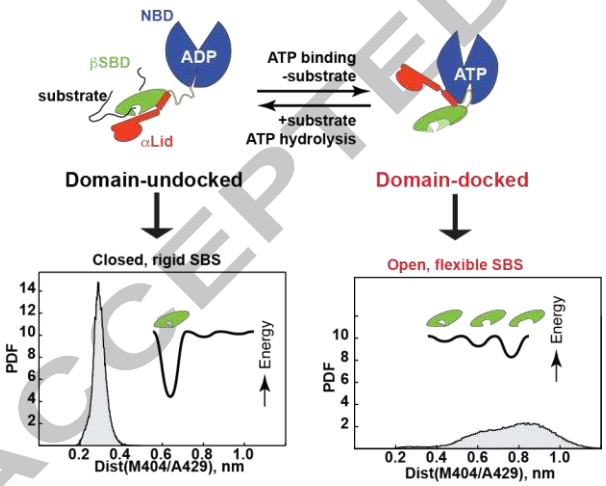


Figure 11. ATP binding/hydrolysis controls flexibility of the Hsp70 substrate-binding site: (Top) Two 'end-point' Hsp70 conformations (ADP-bound, domain-undocked and ATP-bound, domain-docked), schematically showing the NBD in blue and SBD in green (for βSBD) and red (for αLid). (Bottom) Probability density functions (PDFs) for the distance distribution between residues Met404 and Phe429 forming the substrate-binding sites are obtained from MD trajectories of the domain-undocked βSBD and domain-docked βSBD (reproduced with permission from [83]); cartoons show the smooth and rough βSBD conformational landscapes for domain-undocked (left) and domain-docked (right) Hsp70 conformations respectively.

4.2.2 Hsp70 preferentially interacts with unfolded protein conformations: It has been widely accepted that Hsp70 chaperones affect the thermodynamics and kinetics of interconversion between states on the folding energy landscape of their protein substrates [299]; however, the exact mechanism by which Hsp70s promote protein folding remains enigmatic. Recent studies by the Cavagnero and Kay groups [84, 85, 300] employed solution NMR to address this challenging question. Cavagnero and colleagues [300] characterized conformational

changes in the N-terminal SH3 domain of the *Drosophila melanogaster* adaptor protein drk (drkN SH3) upon its interactions with ADP-bound bacterial Hsp70. In the absence of Hsp70, the drkN SH3 domain is thermodynamically unstable [51-53]. As a result, two sets of peaks (corresponding to the folded and unfolded conformations, interconverting slowly on the NMR time-scale) are detected in ^{1}H-^{15}N HSQC of ^{15}N-labeled drkN SH3. In the presence of Hsp70, significant changes in drkN SH3 ^{1}H-^{15}N HSQC point to the interaction between Hsp70 and the domain. 2D diffusion-ordered spectroscopy (DOSY) revealed that only the unfolded SH3 conformation interacts with ADP-bound Hsp70 [300]. In full agreement with the DOSY data, many ^{1}H-^{15}N HSQC peaks corresponding to the unfolded state of the drkN SH3 domain were affected by Hsp70 binding, showing an increase or decrease in intensity. By contrast, uniform reductions in peak intensities for all folded SH3 domain peaks were observed in ^{1}H-^{15}N HSQC upon addition of Hsp70, suggesting that in the presence of Hsp70, the population of folded drkN SH3 was significantly reduced (from ca. 62% in the absence of Hsp70 to ca. 14% in the presence of Hsp70). In other words, Hsp70 interaction with the unfolded state of the drkN SH3 domain shifts thermodynamic equilibrium toward the unfolded state (Fig. 12A). In turn, the analysis of Trp side-chain peaks in ^{1}H-^{15}N HSQC revealed the presence of multiple Hsp70-bound SH3 conformations. These multiple conformations interconvert slowly on the NMR time-scale, and have distinct chemical shifts that, however, are very similar to the chemical shifts of the globally unfolded state of the SH3 domain observed in the absence of Hsp70.

To identify the regions of drkN SH3 domain that directly interact with Hsp70, the authors employed phase-modulated CLEAN chemical exchange experiments. They found that a relatively large central portion of the unfolded SH3 domain becomes less solvent exposed in the presence of Hsp70. Importantly, this region is significantly larger than a single Hsp70 recognition motif comprising only 5-8 nonpolar residues, suggesting that different Hsp70 molecules interact with several distinct regions of the unfolded SH3 domain [300].

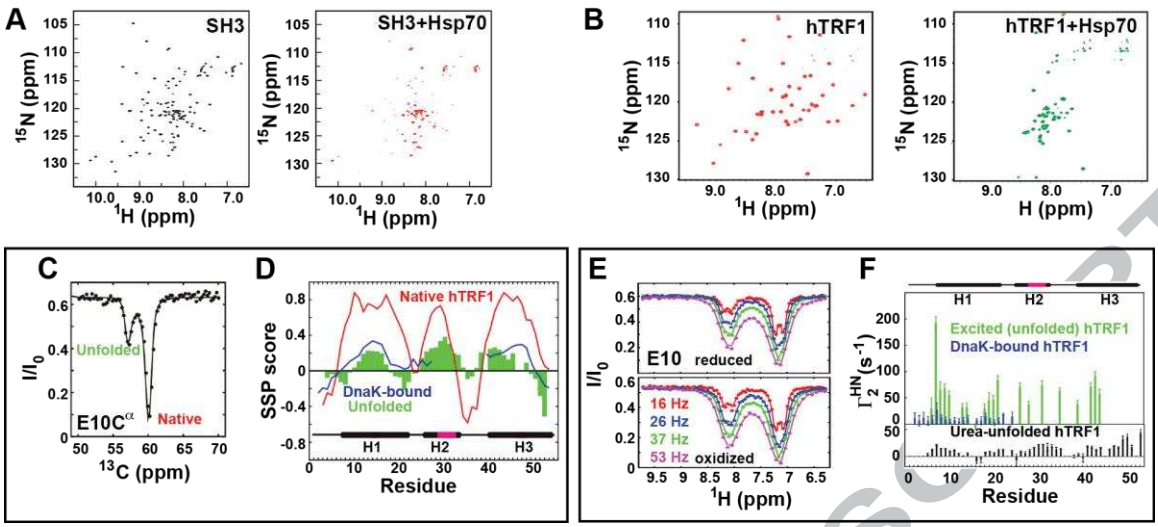


Figure 12. Conformational changes in unfolded proteins upon binding to molecular chaperone Hsp70: (A and B) Hsp70 binding stabilizes the unfolded states of drkN SH3 domain (A) and hTRF1 (B) as monitored by characteristic peak patterns in the amide correlation spectra (reproduced with permission from [300] and [84]). (C and D) Hsp70 binding does not perturb hTRF1 secondary structure propensities as monitored by CEST experiments: (C) The CEST profile for the residue E10 showing peak intensity (I) as a function of the carrier frequency of a weak rf field, normalized to the peak intensity in the absence of the rf irradiation. (D) SSP scores for the unfolded state (green) calculated from CEST-derived chemical shifts for ^{15}N, $^{13}C^{\alpha}$, and $^{13}C'$ nuclei; the red and blue lines are the SSP scores for the native and ADP-Hsp70 bound states of hTRF1, respectively (reproduced with permission from [84]). (E and F) Hsp70 binding alters hTRF1 long-range contacts as monitored by a ^{1}H CEST-based method for measuring PREs: (E) CEST profile of the residue E10 of hTRF1 with the nitroxide spin label at position 52 in reduced and oxidized forms; ratios of intensities of cross-peaks in the presence (I) and absence (I_0) of delay T_{EX} are plotted along the y axis as a function of the carrier of the weak ^{1}H B_1 field. (reproduced with permission from [85]). (F) PRE values in the Hsp70-bound hTRF1 (blue), Hsp70-unbound unfolded (excited) hTRF1 (green), and urea-unfolded hTRF1 (black).

Intriguingly, the Hsp70-induced stabilization of the unfolded substrate states is not unique to the drkN SH3 domain. In excellent agreement with the SH3 domain results [300], NMR characterization of another metastable protein, a three-helix bundle member of the homeodomain family, hTRF1, revealed that the Hsp70 binding similarly shifts the conformational equilibrium for hTRF1 toward the globally unfolded conformation [84]. Thus, at 35^{o}C the population of the unfolded hTRF1 state is only about 4%. By contrast, in the presence of Hsp70, ^{1}H-^{15}N TROSY spectroscopy of ^{15}N labeled hTRF1 revealed that the protein became predominantly unfolded (Fig. 12B). To validate that unfolded hTRF1 is a specific Hsp70 substrate and, thus, occupies the Hsp70 substrate-binding site, methyl-TROSY spectra of perdeuterated, I,L,V,M methyl-labeled ADP-bound Hsp70 were used to monitor characteristic chemical shift fingerprints corresponding to the substrate-bound Hsp70 conformation. Titrations of methyl-labeled Hsp70 with unlabeled hTRF1 revealed chemical shift changes for the methyl groups proximal to the Hsp70 substrate-binding site, suggesting that hTRF1 indeed binds to this site with μM affinity.

4.2.3 Hsp70 binding directly affects the unfolded protein ensemble: While studies of drkN SH3 [300] and hTRF1 [84] suggested that bacterial Hsp70 predominantly interacts with the globally unfolded forms of its protein substrates, and these interactions allow SH3 and hTRF1 to sample a heterogeneous ensemble of unfolded conformations, the question arises: Does Hsp70

alter this unfolded ensemble and directly affect the folding pathways of the substrate proteins? To address this question, and structurally characterize Hsp70-bound hTRF1 conformations, Kay and colleagues performed a secondary structure propensity (SSP) analysis of $^{13}C^{\alpha}$ and $^{13}C'$ chemical shifts [84]. In the absence of Hsp70, CEST experiments were employed to obtain $^{13}C^{\alpha}$ and $^{13}C'$ chemical shifts for the low populated (~4%) globally unfolded hTRF1 conformation (Fig. 12C). In the presence of Hsp70, $^{13}C^{\alpha}$ and $^{13}C'$ chemical shifts were obtained directly from HNCO and HNCA spectra. The CSP analysis revealed that in the presence and absence of Hsp70, unfolded hTRF1 adopts very similar globally unfolded conformations. Moreover, Hsp70 binding does not notably perturb residual secondary structure (up to 40% helical structure for some regions), transiently populated by unfolded hTRF1.

To probe whether Hsp70 binding affects long-range interactions in the unfolded hTRF1, Kay and colleagues [85] employed ^{1}H paramagnetic relaxation enhancement (PRE) measurements. In the presence of Hsp70, ^{1}H PREs were obtained using amide proton transverse relaxation measurements on Hsp70-bound K52C-tempol hTRF1. To measure PREs for low population unfolded hTRF1 (~4%) in the absence of Hsp70, a new ^{1}H-based CEST experiment was developed to allow ^{1}H PREs to be probed for low population unfolded states (Fig. 12E,F). The analysis of PREs obtained for Hsp70-unbound and -unbound unfolded hTRF1 revealed that Hsp70 binding prevents the formation of long-range contacts within the unfolded hTRF1 chain [85].

Despite the fact that the ‘unfoldase’ activity of Hsp70 chaperones has been suggested in previous work [299], the molecular mechanism of this process remains poorly understood. NMR studies of interactions of the Hsp70 molecular chaperone with two protein substrates, drkN SH3 [300] and hTRF1 [84, 85, 300], have indeed provided detailed molecular insights into how Hsp70 molecular chaperones fine-tune the folding landscape of their substrate proteins. These studies demonstrated that Hsp70 chaperones not only shield ‘sticky’, aggregation-prone regions by binding to the hydrophobic stretches of the protein sequence, but also prevent the formation of nonnative long-range contacts. As a result, Hsp70s actively destabilize (‘unfold’) off-pathway contacts, converting misfolded off-pathway intermediates into extended unfolded conformations that contain only local secondary structure (Fig. 13). Moreover, this ‘unfoldase’ activity is not a unique property of the Hsp70 chaperone family. A recent NMR study of GroEL [62], an ATP-dependent chaperone from the unrelated Hsp60 family, revealed that, similarly to Hsp70, GroEL binding to a metastable mutant of the Fyn SH3 domain significantly alters the SH3 domain folding landscape, stabilizing the unfolded intermediate and accelerating interconversion between the folded SH3 domain and the intermediate [62].

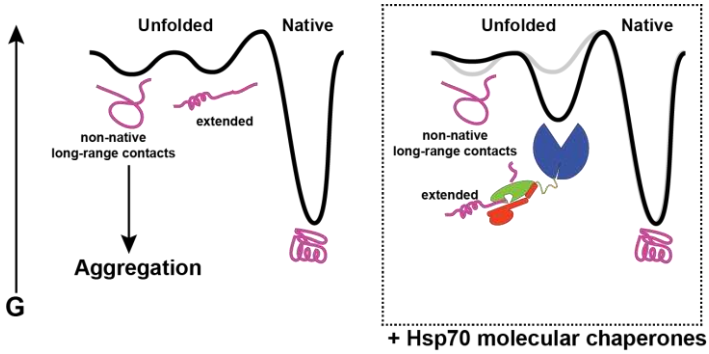


Figure 13. Modulation of a protein folding landscape by Hsp70: (Left) schematic of a free energy landscape for an isolated protein, showing the difference in free energy (G) between the natively folded state, an unfolded ensemble of extended conformations and a nonnative off-pathway intermediate. (Right) Schematic of a free-energy diagram of a protein in the presence of molecular chaperones that stabilize the extended unfolded conformations, alter the protein folding landscape and inhibit aggregation.

To efficiently perform their functions, Hsp70s and the majority of other chaperones interact with a variety of co-chaperones and collaborate with a number of other PQC members [288, 301-303], adding yet another level of complexity to the chaperone-protein substrate relationship. The majority of these interactions are extremely dynamic and transient and involve the formation of large multicomponent complexes. As a result, characterization of this complex and multicomponent chaperone network is a very challenging task. However, in the last decade solution NMR has become one of the most powerful techniques allowing detailed studies of such challenging systems. For further examples, interested readers are referred to a recent review [292] that described NMR studies of the Hsp70/ClpB-assisted disaggregation cascade.

4.3 Conformational landscape of intrinsically disordered proteins in vivo

Intrinsically disordered proteins (IDPs) and disordered protein regions (IDPRs) represent an important paradigm in molecular and cell biology [304]. Approximately 30% of all mammalian proteins (and 75% of all signaling proteins) are IDPs or contain extended IDPRs [305]. IDPs do not adopt stable secondary and tertiary structure (at least in certain environments); instead, they can be described as highly dynamic, heterogeneous ensembles of interconverting conformations [306-308]. IDPs and IDPRs are components of the majority of signaling pathways and are associated with many devastating diseases, including neurodegenerative disorders, cancers, diabetes, cardiovascular and metabolic diseases [309]. IDP conformation strongly depends on a complex intracellular environment and even minor perturbations of the intracellular environment are thought to be critical for the conformational (and functional) properties of IDPs [304]. As a result, despite significant progress in characterization of IDP/IDPR conformational ensembles *in vitro*, how IDPs behave and function in the living cell remains unclear. Do IDPs maintain their disordered conformation inside eukaryotic cells? How does the cellular environment affect an IDP's conformational ensemble (e.g., local and global fold, long-range interactions, compactness, etc.) and function, or cause dysfunctions (e.g., spontaneous aggregation or degradation)? Recent

developments in NMR instrumentation and methodology have provided the means to address these questions by allowing detailed molecular characterization of protein behavior *in vivo*. For more technical information about in-cell NMR, the reader is referred to comprehensive recent reviews [310, 311] that describe modern NMR approaches, as well as some of their applications to *in vivo* protein folding, maturation and post-translational modifications. In this section, we discuss two recent examples of the in-cell NMR characterization of an intrinsically disordered protein α-synuclein in different human cell types [312, 313].

4.3.1 Previous NMR studies of α-synuclein in vitro and in vivo: The formation of amyloid fibrils by α-synuclein in the cytoplasm of dopaminergic cells is a hallmark of Parkinson's disease. This devastating disease affects about 1% of people older than 65 years of age and is characterized by the progressive degeneration of dopaminergic neurons in the *substantia nigra* and other brain regions [314]. Several *in vitro* NMR and biophysical studies have characterized conformational variability of α-synuclein and elucidated its aggregation mechanism(s). The details of these studies can be found in a comprehensive review by Alderson & Markley [315]. Previous *in vitro* data demonstrated that the N-terminal domain (residues 1-65) of α-synuclein is largely disordered, but adopts an α-helical conformation upon interactions with biological membrane [316]. By contrast, the C-terminal region (residues 96-140), enriched in acidic and Pro residues, is fully disordered [314, 317]. Despite numerous *in vitro* studies, the behavior of α-synuclein and its mechanism of toxicity *in vivo* still remain obscure. Indeed, a number of factors, such as interactions with lipids [318] [319], post-translational modifications [320] [321] [322], and the presence of certain metabolites and metal ions [323], have been suggested to significantly affect the formation of secondary structure elements and long-range contacts, as well as aggregation propensity of this protein *in vivo* [324].

NMR characterization of α-synuclein in *E. coli* cells revealed a significant broadening of resonances from its N-terminal terminal domain. The observed line broadening in NMR spectra suggests that this region weakly and transiently interacts with the cell membrane and/or macromolecules in cytoplasm. By contrast, no changes in the secondary structure can be inferred from NMR chemical shifts of recombinant α-synuclein *in vitro* or in *E. coli* cells [325] [326]. Moreover, the lack of significant chemical shift perturbations in the in-cell NMR HSQC spectrum suggested that α-synuclein adopts a monomeric, intrinsically disordered conformation in bacterial and eukaryotic cells [322, 325]. A very similar conformation was observed for an isolated recombinant α-synuclein *in vitro*. By contrast, analytical ultracentrifugation, scanning transmission electron microscopy and crosslinking experiments suggest that native, mammalian-cell-derived α-synuclein predominantly forms a dynamic tetramer [327]. These controversial findings suggest that α-synuclein behavior (and potentially its aggregation) apparently depend on experimental conditions and/or cell type and on physiological fitness of the cell.

4.3.2 In-cell NMR characterization of α-synuclein in non-neuronal and neuronal cells: To address some of these findings and investigate α-synuclein behavior in physiologically relevant cell types (non-neuronal and neuronal cells) and concentrations (c.a. 15 μM), Selenko and colleagues [312] employed NMR and electron paramagnetic resonance (EPR) spectroscopy. To deliver ^{15}N labeled α-synuclein into mammalian cells, the authors used an electroporation protocol that allows tight control of the cellular α-synuclein concentration and results in a uniform distribution of the protein in cytoplasm, with minimal perturbations in cell viability [328]. 2D ^{1}H-^{15}N correlation NMR spectra were used to monitor α-synuclein conformation in different cell lines, including neuronal B65 and SK-N-SH cells and RCSN-3 cells, directly derived from rat *substantia nigra* neurons. Several common features of α-synuclein were revealed for all these cell lines: (i) The intracellular environment does not lead to major conformational rearrangement of conformation of the monomeric disordered α-synuclein; (ii) the N-terminus of α-synuclein undergoes post-translational acetylation in cells; (iii) α-synuclein does not form stable interactions with cellular membranes; however, (iv) N- and C-termini of α-synuclein transiently interact with the cytoplasmic components and/or membrane.

Selenko and colleagues [312] performed ^{15}N relaxation measurements to investigate transient interactions *in vivo* resulting in a significant line broadening in ^{1}H-^{15}N NMR spectra for resonances from N- and C-terminal residues. ^{15}N R_1 and R_2 relaxation measurements were utilized to discriminate between line broadening due to μs-ms conformational dynamics (giving rise to exchange contributions R_{ex} to transverse relaxation rates R_2) and due to viscosity-driven increases in the overall rotation correlation time. To separate the effects of molecular crowding, unspecific interactions with other cytoplasmic proteins and transient membrane binding, in-cell NMR data were complemented by *in vitro* NMR experiments using isolated acetylated α-synuclein in the presence of (i) Ficoll (an inert macromolecular crowding agent), (ii) two globular proteins, bovine serum albumin (or BSA, a 69 kDa, negatively charged protein) and egg white lysozyme (a 14 kDa positively charged protein), and (iii) small unilamellar liposomes/vesicles (SUVs) (Fig. 14A).

Electrostatic interactions were shown to be responsible for transient contacts between the negatively charged C-terminal domain of α-synuclein and cytoplasmic components. Indeed, the enhanced μs-ms dynamics (large R_{ex} terms) for the C-terminal domain were observed inside the cells as well as in the presence of the positively charged lysozyme *in vitro*. However, high salt concentrations (300 mM) resulted in drastic reduction of R_{ex} contributions (Fig. 14A). On the other hand, hydrophobic interactions are shown to play a key role in transient contacts between the N-terminal domain of α-synuclein with the membrane and/or cytoplasmic proteins. Enhanced μs-ms dynamics for the N-terminal domain were observed inside the cells, as well as for an isolated α-synuclein in the presence of BSA and SUVs *in vitro*. In turn, substitution of the two N-terminal

aromatic residues Phe4 and Tyr39 to Ala resulted in a significant decrease of N-terminal μs-ms dynamics in cells and *in vitro* (Fig. 14A).

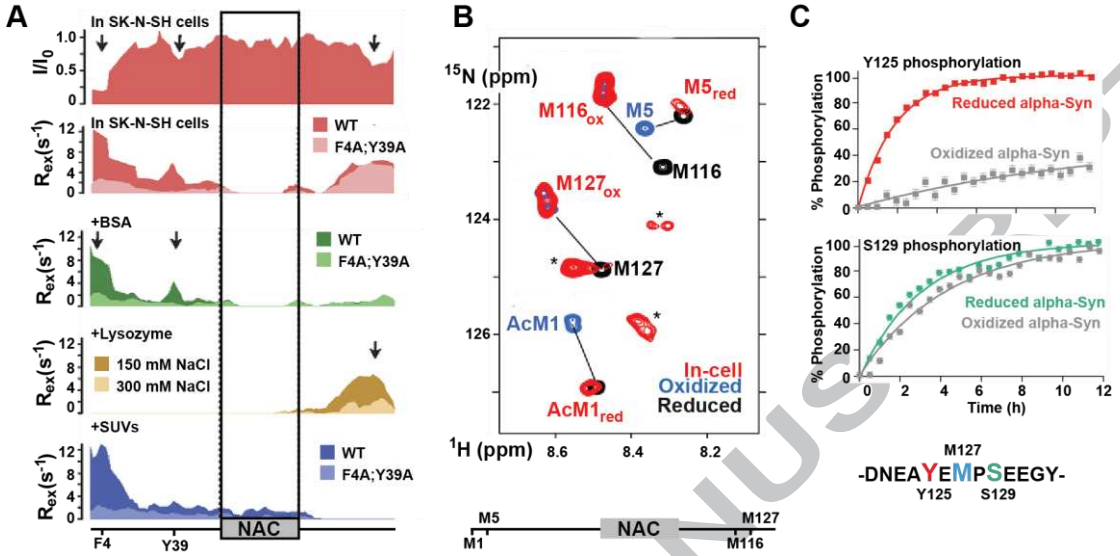


Figure 14. Conformational variability of α-synuclein in cells: (A) α-synuclein transiently interacts with other cytoplasmic proteins and membrane: Residue-resolved NMR signal intensity ratios I/I_0, where I and I_o are signal intensities for α-synuclein in cells and for an isolated N-terminally acetylated α-synuclein, and exchange contributions (R_{ex}) to ^{15}N transverse relaxation for α-synuclein in SK-N-SH cells and for an isolated N-terminally acetylated α-synuclein in the presence of two globular proteins, bovine serum albumin (or BSA, a 69 kDa, negatively charged protein) and egg white lysozyme (a 14 kDa positively charged protein), and small unilamellar liposomes/vesicles (SUVs). The α-synuclein sequence with highlighted non-amyloid-β component (NAC) region (residues 61-95) and the two N-terminal aromatic residues Phe4 and Tyr39 is schematically shown on the bottom. (reproduced with permission from [312]) (B) Oxidized N-terminal Met1 and Met5 (but not Met116 and Met127) are repaired inside the cell: Overlay of 2D in-cell NMR spectra of Met-^{15}N isotope-enriched α-synuclein in cells (red), with reference spectra of the reduced (black) and oxidized (blue) isolated protein. Asterisks denote metabolite background signals in cells. The α-synuclein sequence with the [NAC] region and four Met residues are schematically shown on the bottom. (reproduced with permission from [313]) (C) C-terminal methionine-oxidation affects phosphorylation of α-synuclein: Real-time NMR profiles of Tyr125 and S129 phosphorylation kinetics of the reduced and Met-oxidized, N-terminally acetylated α-synuclein in reconstituted kinase reactions with recombinant kinases: Fyn for Y125 and Plk3 for S129. Amino-acid sequence of α-synuclein residues flanking Met127 is shown on the bottom (reproduced with permission from [313]).

To investigate how cellular environment affects long-range contacts and compactness of the α-synuclein chain, Selenko and colleagues [312] performed NMR PRE and EPR double-electron resonance measurements, showing that α-synuclein adopts a more compact conformation *in vivo* with the average radius of gyration of 3.6 nm inside the cell *vs.* 4.0 nm in the buffer. A very similar level of compaction was observed for an isolated α-synuclein *in vitro* in the presence of crowding agents (the average radius of gyration of 3.4 and 3.6 nm in the presence of Ficoll and BSA, respectively), suggesting that macromolecular crowding is predominantly responsible for the greater level of compaction observed *in vivo.* It was previously demonstrated that the non-amyloid-β component (NAC) region (residues 61-95), located between the N-terminal domain and the charged C-terminus of α-synuclein, is responsible for self-association and aggregation processes [49, 329]. Hypothetically, this compact conformation of α-synuclein observed *in vivo* can play a role in shielding the NAC region and, thus, protecting α-synuclein from aggregation under normal physiological conditions.

4.3.3 In-cell NMR characterization of α-synuclein conformation under oxidative stress: Cellular oxidative stress is suggested to be a key triggering factor for inducing α-synuclein aggregation. To elucidate how oxidative stress can alter the fate of α-synuclein inside the living cell, Selenko and colleagues [313] produced ^{15}N-labeled, N-terminally acetylated, methionine-(Met1, Met5, Met116 and Met127)-oxidized α-synuclein. Chemical shift perturbation and secondary structure propensity analyses, performed for an isolated α-synuclein *in vitro,* revealed that methionine oxidation reduces α-helicity of the N-terminal domain, but does not significantly affect the disordered conformation of the C-terminus. Next, ^{15}N-labeled, N-terminally acetylated, methionine-(Met1, Met5, Met116 and Met127)-oxidized α-synuclein was delivered into non-neuronal and neuronal cells (Fig. 14B). Time-resolve in-cell NMR demonstrated that inside the cells oxidized N-terminal Met1 and Met5 are repaired by the methionine sulfoxide reductases in a step-wise manner (with Met5 preceding Met1). However, no repair of the C-terminal Met116 and Met127 was detected. Intriguingly, Met127 is located in close proximity to several phosphorylation sites of α-synuclein. While phosphorylation of Ser129, Tyr133 and Tyr136 was not affected by the C-terminal methionine-oxidation, Tyr125 phosphorylation was severely impaired, as monitored by time-resolved NMR experiments *in vitro* (Fig. 14C). Consequently, these findings suggest that alteration of the cellular environment (e.g. upon physiological or pathological stresses) can selectivity affect post-translational modifications of α-synuclein and, thus, potentially control its conformation, conformational landscape, and aggregation propensity (Fig. 15).

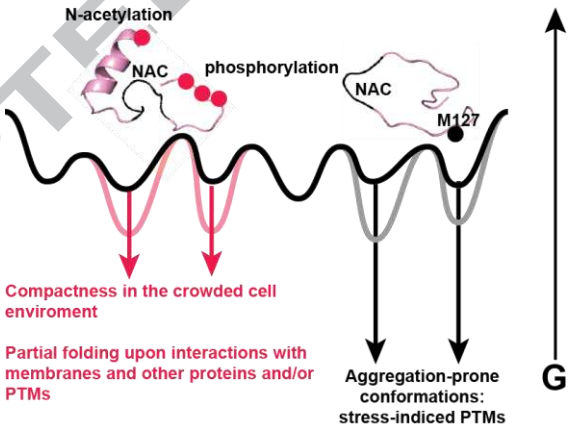


Figure 15. Folding landscape of an IDP in a living cell: Schematic of a free-energy diagram for an intrinsically disordered protein (e.g., α-synuclein), showing relatively small differences in free energy (G) among different conformers in the IDP ensemble. The cellular environment may stabilize some conformations of this ensemble, altering protein folding energy landscape, function and aggregation-propensity. For example, post-translational N-terminal acetylation of α-synuclein in cells stabilizes an α-helical formation of the N-terminal region [330]; C-terminal phosphorylation, e.g., at Ser129 inhibits its fibrillation [320]; and cellular crowding stabilizes a more compact α-synuclein conformation, shielding the aggregation-prone NAC region [312].

IDP toxicity is highly integrated with several key cellular stress pathways [317, 331, 332, 333]. Moreover, accumulation of non-functional, toxic, aggregation-prone IDP/IDPR conformers is

associated with many devastating diseases, including cancer, cardiovascular diseases, neurodegenerative diseases, and diabetes [309]. Despite recognition that IDPs and IDPRs undergo function-related conformational changes in the continuously changing cellular environment [306, 334, 335], how exactly IDPs respond to different physiological and pathological conditions remains poorly understood. More studies will be required to unravel the detailed mechanisms of function, aggregation and toxicity of IDPs *in vivo*, as well as exact roles of their posttranslational modifications and interactions with membranes, other proteins and different components of PQC networks in the cytoplasm [314] and in organelles such as the endoplasmic reticulum [317] and mitochondria [336]. Solution NMR spectroscopy has emerged at the forefront of experimental methods for high-resolution characterization of structural and dynamic features of complex, multicomponent IDP ensembles under native conditions, i.e. in the presence of multiple interacting partners, PQC enzymes and inside the cells.

5. Concluding Remarks

Advances in the development of NMR methods for studying protein transitions to low-population conformational states have positioned NMR as the primary experimental tool to probe the properties of nonnative states on the protein folding energy landscape, as well as to elucidate the roles these states play in protein folding, misfolding, aggregation and function. On the other hand, continuing development of NMR approaches and isotope labeling strategies for studying high molecular weight protein assemblies has recently allowed elucidation of the mechanisms by which protein folding occurs in the context of a complex and multicomponent cellular environment, providing new insights into ribosome-nascent chain complexes, folding in the presence of components of protein quality control networks, and multicomponent IDP and IDPR ensembles. It is highly anticipated that solution NMR, alone and/or in combination with other structural and biophysical methods such as X-ray crystallography, cryo-electron microscopy and computational modeling, will continue to play a major role in elucidating protein folding *in vitro* and *in vivo*.

Acknowledgements

We would like to thank Prof. Eugenia Clerico for stimulating discussion. AZ acknowledges funding from the Biotechnology and Biological Sciences Research Council (BB/M021874/1). Research in DMK lab is supported by NSF MCB (1615866).

References

[1] A. Fersht, Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding, W.H. Freeman, New York, 1999.

[2] D. Baker, A. Sali, Protein structure prediction and structural genomics, Science, 294 (2001) 93-96.

[3] R. Das, D. Baker, Macromolecular modeling with Rosetta, Annual Review of Biochemistry, 77 (2008) 363-382.

[4] K.A. Dill, H.S. Chan, From Levinthal to pathways to funnels, Nat Struct Biol, 4 (1997) 10-19.

[5] T.R. Jahn, S.E. Radford, The Yin and Yang of protein folding, FEBS J, 272 (2005) 5962-5970.

[6] A.J. Baldwin, L.E. Kay, NMR spectroscopy brings invisible protein states into focus, Nat Chem Biol, 5 (2009) 808-814.

[7] K. Henzler-Wildman, D. Kern, Dynamic personalities of proteins, Nature, 450 (2007) 964-972.

[8] C.M. Dobson, Protein folding and misfolding, Nature, 426 (2003) 884-890.

[9] F. Chiti, C.M. Dobson, Protein misfolding, functional amyloid, and human disease, Annu Rev Biochem, 75 (2006) 333-366.

[10] F. Chiti, C.M. Dobson, Amyloid formation by globular proteins under native conditions, Nat Chem Biol, 5 (2009) 15-22.

[11] T.P. Knowles, M. Vendruscolo, C.M. Dobson, The amyloid state and its association with protein misfolding diseases, Nat Rev Mol Cell Biol, 15 (2014) 384-396.

[12] T. Mittag, J.D. Forman-Kay, Atomic-level characterization of disordered protein ensembles, Curr Opin Struct Biol, 17 (2007) 3-14.

[13] T.R. Sosnick, D. Barrick, The folding of single domain proteins - have we reached a consensus?, Curr Opin Struct Biol, 21 (2011) 12-24.

[14] K.A. Dill, D. Shortle, Denatured states of proteins, Annu Rev Biochem, 60 (1991) 795-825.

[15] S.W. Englander, L. Mayne, M.M.G. Krishna, Protein folding and misfolding: mechanism and principles, Quart Rev Biophys, 40 (2007) 287-326.

[16] D.J. Brockwell, S.E. Radford, Intermediates: ubiquitous species on folding energy landscapes?, Curr Opin Struct Biol, 17 (2007) 30-37.

[17] M. Vendruscolo, C.M. Dobson, Structural biology: Protein self-assembly intermediates, Nat Chem Biol, 9 (2013) 216-217.

[18] S.E. Jackson, How do small single-domain proteins fold?, Fold Des, 3 (1998) R81-R91.

[19] A.R. Fersht, V. Daggett, Protein folding and unfolding at atomic resolution, Cell, 108 (2002) 573-582.

[20] C. Clementi, H. Nymeyer, J.N. Onuchic, Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins, J Mol Biol, 298 (2000) 937-953.

[21] R.M. Vabulas, S. Raychaudhuri, M. Hayer-Hartl, F.U. Hartl, Protein folding in the cytoplasm and the heat shock response, Cold Spring Harb Perspect Biol, 2 (2010) a004390.

[22] K. Pervushin, R. Riek, G. Wider, K. Wuthrich, 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 U S A, 94 (1997) 12366-12371.

[23] V. Tugarinov, P.M. Hwang, J.E. Ollerenshaw, L.E. Kay, Cross-correlated relaxation enhanced 1H[bond]13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes, J Am Chem Soc, 125 (2003) 10420-10428.

[24] R. Sprangers, A. Velyvis, L.E. Kay, Solution NMR of supramolecular complexes: providing new insights into function, Nat Methods, 4 (2007) 697-703.

[25] S.R. Tzeng, M.T. Pai, C.G. Kalodimos, NMR studies of large protein systems, Methods Mol Biol, 831 (2012) 133-140.

[26] D. Sheppard, R. Sprangers, V. Tugarinov, Experimental approaches for NMR studies of side-chain dynamics in high-molecular-weight proteins, Prog Nucl Magn Reson Spectrosc, 56 (2010) 1-45.

[27] A.G. Palmer, C.D. Kroenke, J.P. Loria, Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules, Nuclear Magnetic Resonance of Biological Macromolecules, Pt B, 339 (2001) 204-238.

[28] D.F. Hansen, P. Vallurupalli, L.E. Kay, Using relaxation dispersion NMR spectroscopy to determine structures of excited, invisible protein states, J Biomol NMR, 41 (2008) 113-120.

[29] D.M. Korzhnev, L.E. Kay, Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: An application to protein folding, Acc Chem Res, 41 (2008) 442-451.

[30] A.G. Palmer, F. Massi, Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy, Chem Rev, 106 (2006) 1700-1719.

[31] A.G. Palmer, 3rd, Chemical exchange in biomacromolecules: past, present, and future, J Magn Reson, 241 (2014) 3-17.

[32] N.J. Anthis, G.M. Clore, Visualizing transient dark states by NMR spectroscopy, Q Rev Biophys, 48 (2015) 35-116.

[33] N.L. Fawzi, J. Ying, D.A. Torchia, G.M. Clore, Probing exchange kinetics and atomic resolution dynamics in high-molecular-weight complexes using dark-state exchange saturation transfer NMR spectroscopy, Nat Protoc, 7 (2012) 1523-1533.

[34] D.M. Korzhnev, X. Salvatella, M. Vendruscolo, A.A. Di Nardo, A.R. Davidson, C.M. Dobson, L.E. Kay, Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR, Nature, 430 (2004) 586-590.

[35] P. Neudecker, P. Lundstrom, L.E. Kay, Relaxation dispersion NMR spectroscopy as a tool for detailed studies of protein folding, Biophys J, 96 (2009) 2045-2054.

[36] D.D. Boehr, H.J. Dyson, P.E. Wright, An NMR perspective on enzyme dynamics, Chem Rev, 106 (2006) 3055-3079.

[37] G.P. Lisi, J.P. Loria, Using NMR spectroscopy to elucidate the role of molecular motions in enzyme function, Prog Nucl Magn Reson Spectrosc, 92-93 (2016) 1-17.

[38] K. Sugase, H.J. Dyson, P.E. Wright, Mechanism of coupled folding and binding of an intrinsically disordered protein, Nature, 447 (2007) 1021-1025.

[39] J. Kubelka, J. Hofrichter, W.A. Eaton, The protein folding 'speed limit', Curr Opin Structural Biol, 14 (2004) 76-88.

[40] D.N. Ivankov, S.O. Garbuzynskiy, E. Alm, K.W. Plaxco, D. Baker, A.V. Finkelstein, Contact order revisited: Influence of protein size on the folding rate, Protein Sci, 12 (2003) 2057-2062.

[41] K.L. Maxwell, D. Wildes, A. Zarrine-Afsar, M.A. de los Rios, A.G. Brown, C.T. Friel, L. Hedberg, J.C. Horng, D. Bona, E.J. Miller, A. Vallee-Belisle, E.R.G. Main, F. Bemporad, L.L. Qiu, K. Teilum, N.D. Vu, A.M. Edwards, I. Ruczinski, F.M. Poulsen, B.B. Kragelund, S.W. Michnick, F. Chiti, Y.W. Bai, S.J. Hagen, L. Serrano, M. Oliveberg, D.P. Raleigh, P. Wittung-Stafshede, S.E. Radford, S.E. Jackson, T.R. Sosnick, S. Marqusee, A.R. Davidson, K.W. Plaxco, Protein folding: Defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins, Protein Sci, 14 (2005) 602-616.

[42] D.M. Korzhnev, T.L. Religa, W. Banachewicz, A.R. Fersht, L.E. Kay, A Transient and Low-Populated Protein-Folding Intermediate at Atomic Resolution, Science, 329 (2010) 1312-1316.

[43] D.M. Korzhnev, T.L. Religa, W. Banachewicz, A.R. Fersht, L.E. Kay, A transient and low-populated protein folding intermediate at atomic resolution, Science, 329 (2010) 1312-1316.

[44] P. Neudecker, P. Robustelli, A. Cavalli, P. Walsh, P. Lundstrom, A. Zarrine-Afsar, S. Sharpe, M. Vendruscolo, L.E. Kay, Structure of an intermediate state in protein folding and aggregation, Science, 336 (2012) 362-366.

[45] D.M. Korzhnev, T.L. Religa, L.E. Kay, Transiently populated intermediate functions as a branching point of the FF domain folding pathway, Proc Natl Acad Sci U S A, 109 (2012) 17777-17782.

[46] H.Y. Carr, E.M. Purcell, Effects of diffusion on free precession in nuclear magnetic resonance experiments, Phys Rev, 94 (1954) 630-638.

[47] S. Meiboom, D. Gill, Modified spin-echo method for measuring nuclear relaxation times, Rev Sci Instrum, 29 (1958) 688-691.

[48] G.M. Clore, J. Iwahara, Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes, Chem Rev, 109 (2009) 4108-4139.

[49] D. Ghosh, P.K. Singh, S. Sahay, N.N. Jha, R.S. Jacob, S. Sen, A. Kumar, R. Riek, S.K. Maji, Structure based aggregation studies reveal the presence of helix-rich intermediate during α-Synuclein aggregation, Sci Rep, 5 (2015) 9228.

[50] E. Rennella, B. Brutscher, Fast real-time NMR methods for characterizing short-lived molecular states, Chem Phys Chem, 14 (2013) 3059-3070.

[51] O. Zhang, J.D. Forman-Kay, Structural characterization of folded and unfolded states of an SH3 domain in equilibrium in aqueous buffer, Biochemistry, 34 (1995) 6784-6794.

[52] M. Tollinger, N.R. Skrynnikov, F.A.A. Mulder, J.D. Forman-Kay, L.E. Kay, Slow dynamics in folded and unfolded states of an SH3 domain, J Am Chem Soc, 123 (2001) 11341-11352.

[53] I. Bezsonova, D.M. Korzhnev, R.S. Prosser, J.D. Forman-Kay, L.E. Kay, Hydration and packing along the folding pathway of SH3 domains by pressure-dependent NMR, Biochemistry, 45 (2006) 4711-4719.

[54] Y.W. Bai, T.R. Sosnick, L. Mayne, S.W. Englander, Protein folding intermediates - native-state hydrogen-exchange, Science, 269 (1995) 192-197.

[55] B.M. Burmann, S. Hiller, Chaperones and chaperone-substrate complexes: Dynamic playgrounds for NMR spectroscopists, Prog Nucl Magn Reson Spectrosc, 86-87 (2015) 41-64.

[56] L.D. Cabrita, S.T. Hsu, H. Launay, C.M. Dobson, J. Christodoulou, Probing ribosome-nascent chain complexes produced in vivo by NMR spectroscopy, Proc Natl Acad Sci U S A, 106 (2009) 22239-22244.

[57] A. Rutkowska, M. Beerbaum, N. Rajagopalan, J. Fiaux, P. Schmieder, G. Kramer, H. Oschkinat, B. Bukau, Large-scale purification of ribosome-nascent chain complexes for biochemical and structural studies, FEBS Lett, 583 (2009) 2407-2413.

[58] C. Eichmann, S. Preissler, R. Riek, E. Deuerling, Cotranslational structure acquisition of nascent polypeptides monitored by NMR spectroscopy, Proc Natl Acad Sci U S A, 107 (2010) 9111-9116.

[59] L.D. Cabrita, A.M. Cassaignau, H.M. Launay, C.A. Waudby, T. Wlodarski, C. Camilloni, M.E. Karyadi, A.L. Robertson, X. Wang, A.S. Wentink, L.S. Goodsell, C.A. Woolhead, M. Vendruscolo, C.M. Dobson, J. Christodoulou, A structural ensemble of a ribosome-nascent chain complex during cotranslational protein folding, Nat Struct Mol Biol, (2016).

[60] A. Deckert, C.A. Waudby, T. Wlodarski, A.S. Wentink, X. Wang, J.P. Kirkpatrick, J.F. Paton, C. Camilloni, P. Kukic, C.M. Dobson, M. Vendruscolo, L.D. Cabrita, J. Christodoulou, Structural characterization of the interaction of α-synuclein nascent chains with the ribosomal surface and trigger factor, Proc Natl Acad Sci U S A, 113 (2016) 5012-5017.

[61] D.S. Libich, N.L. Fawzi, J. Ying, G.M. Clore, Probing the transient dark state of substrate binding to GroEL by relaxation-based solution NMR, Proc Natl Acad Sci U S A, 110 (2013) 11361-11366.

[62] D.S. Libich, V. Tugarinov, G.M. Clore, Intrinsic unfoldase/foldase activity of the chaperonin GroEL directly demonstrated using multinuclear relaxation-based NMR, Proc Natl Acad Sci U S A, 112 (2015) 8817-8823.

[63] E. Koculi, R. Horst, A.L. Horwich, K. Wuthrich, Nuclear magnetic resonance spectroscopy with the stringent substrate rhodanese bound to the single-ring variant SR1 of the E. coli chaperonin GroEL, Protein Sci, 20 (2011) 1380-1386.

[64] N. Nishida, F. Motojima, M. Idota, H. Fujikawa, M. Yoshida, I. Shimada, K. Kato, Probing dynamics and conformational change of the GroEL-GroES complex by ^{13}C NMR spectroscopy, J Biochem, 140 (2006) 591-598.

[65] R. Horst, E.B. Bertelsen, J. Fiaux, G. Wider, A.L. Horwich, K. Wuthrich, Direct NMR observation of a substrate protein bound to the chaperonin GroEL, Proc Natl Acad Sci U S A, 102 (2005) 12748-12753.

[66] J. Fiaux, E.B. Bertelsen, A.L. Horwich, K. Wuthrich, NMR analysis of a 900K GroEL GroES complex, Nature, 418 (2002) 207-211.

[67] Z. Wang, H. Feng, S.J. Landry, J. Maxwell, L.M. Gierasch, Basis of substrate binding by the chaperonin GroEL, Biochemistry, 38 (1999) 12537-12546.

[68] L.A. Joachimiak, T. Walzthoeni, C.W. Liu, R. Aebersold, J. Frydman, The structural basis of substrate recognition by the eukaryotic chaperonin TRiC/CCT, Cell, 159 (2014) 1042-1055.

[69] G.E. Karagoz, A.M. Duarte, H. Ippel, C. Uetrecht, T. Sinnige, M. van Rosmalen, J. Hausmann, A.J. Heck, R. Boelens, S.G. Rudiger, N-terminal domain of human Hsp90 triggers binding to the cochaperone p23, Proc Natl Acad Sci U S A, 108 (2011) 580-585.

[70] H. Zhang, C. Zhou, W. Chen, Y. Xu, Y. Shi, Y. Wen, N. Zhang, A dynamic view of ATP-coupled functioning cycle of Hsp90 N-terminal domain, Sci Rep, 5 (2015) 9542.

[71] B.K. Zierer, M. Weiwad, M. Rubbelke, L. Freiburger, G. Fischer, O.R. Lorenz, M. Sattler, K. Richter, J. Buchner, Artificial accelerators of the molecular chaperone Hsp90 facilitate rate-limiting conformational transitions, Angew Chem Int Ed Engl, 53 (2014) 12257-12262.

[72] G.E. Karagoz, A.M. Duarte, E. Akoury, H. Ippel, J. Biernat, T. Moran Luengo, M. Radli, T. Didenko, B.A. Nordhues, D.B. Veprintsev, C.A. Dickey, E. Mandelkow, M. Zweckstetter, R. Boelens, T. Madl, S.G. Rudiger, Hsp90-Tau complex reveals molecular basis for specificity in chaperone action, Cell, 156 (2014) 963-974.

[73] T. Didenko, A.M. Duarte, G.E. Karagoz, S.G. Rudiger, Hsp90 structure and function studied by NMR spectroscopy, Biochim Biophys Acta, 1823 (2012) 636-647.

[74] S.J. Park, M. Kostic, H.J. Dyson, Dynamic Interaction of Hsp90 with Its Client Protein p53, J Mol Biol, 411 (2011) 158-173.

[75] T.O. Street, L.A. Lavery, D.A. Agard, Substrate binding drives large-scale conformational changes in the Hsp90 molecular chaperone, Mol Cell, 42 (2011) 96-105.

[76] M. Retzlaff, F. Hagn, L. Mitschke, M. Hessling, F. Gugel, H. Kessler, K. Richter, J. Buchner, Asymmetric activation of the hsp90 dimer by its cochaperone aha1, Mol Cell, 37 (2010) 344-354.

[77] S. Sreeramulu, H.R. Jonker, T. Langer, C. Richter, C.R. Lancaster, H. Schwalbe, The human Cdc37.Hsp90 complex studied by heteronuclear NMR spectroscopy, J Biol Chem, 284 (2009) 3885-3896.

[78] D.M. Jacobs, T. Langer, B. Elshorst, K. Saxena, K.M. Fiebig, M. Vogtherr, H. Schwalbe, NMR backbone assignment of the N-terminal domain of human HSP90, J Biomol NMR, 36 Suppl 1 (2006) 52.

[79] S.J. Park, B.N. Borin, M.A. Martinez-Yamout, H.J. Dyson, The client protein p53 adopts a molten globule-like state in the presence of Hsp90, Nat Struct Mol Biol, 18 (2011) 537-541.

[80] M.A. Martinez-Yamout, R.P. Venkitakrishnan, N.E. Preece, G. Kroon, P.E. Wright, H.J. Dyson, Localization of sites of interaction between p23 and Hsp90 in solution, J Biol Chem, 281 (2006) 14457-14464.

[81] A. Zhuravleva, E.M. Clerico, L.M. Gierasch, An Interdomain Energetic Tug-of-War Creates the Allosterically Active State in Hsp70 Molecular Chaperones, Cell, 151 (2012) 1296-1307.

[82] A. Zhuravleva, L.M. Gierasch, Allosteric signal transmission in the nucleotide-binding domain of 70-kDa heat shock protein (Hsp70) molecular chaperones, Proc Natl Acad Sci U S A, 108 (2011) 6987-6992.

[83] A. Zhuravleva, L.M. Gierasch, Substrate-binding domain conformational dynamics mediate Hsp70 allostery, Proc Natl Acad Sci U S A, 112 (2015) E2865-E2873.

[84] A. Sekhar, R. Rosenzweig, G. Bouvignies, L.E. Kay, Mapping the conformation of a client protein through the Hsp70 functional cycle, Proc Natl Acad Sci U S A, 112 (2015) 10395-10400.

[85] A. Sekhar, R. Rosenzweig, G. Bouvignies, L.E. Kay, Hsp70 biases the folding pathways of client proteins, Proc Natl Acad Sci U S A, 113 (2016) E2794-2801.

[86] T.R. Alderson, J.H. Kim, K. Cai, R.O. Frederick, M. Tonelli, J.L. Markley, The specialized Hsp70 (HscA) interdomain linker binds to its nucleotide-binding domain and stimulates ATP hydrolysis in both cis and trans configurations, Biochemistry, 53 (2014) 7148-7159.

[87] X. Li, S.R. Srinivasan, J. Connarn, A. Ahmad, Z.T. Young, A.M. Kabza, E.R. Zuiderweg, D. Sun, J.E. Gestwicki, Analogs of the Allosteric Heat Shock Protein 70 (Hsp70) Inhibitor, MKT-077, as Anti-Cancer Agents, ACS Med Chem Lett, 4 (2013).

[88] A. Ahmad, A. Bhattacharya, R.A. McDonald, M. Cordes, B. Ellington, E.B. Bertelsen, E.R. Zuiderweg, Heat shock protein 70 kDa chaperone/DnaJ cochaperone complex employs an unusual dynamic interface, Proc Natl Acad Sci U S A, 108 (2011) 18966-18971.

[89] E.B. Bertelsen, L. Chang, J.E. Gestwicki, E.R. Zuiderweg, Solution conformation of wild-type E. coli Hsp70 (DnaK) chaperone complexed with ADP and substrate, Proc Natl Acad Sci U S A, 106 (2009) 8471-8476.

[90] Y. Zhang, E.R. Zuiderweg, The 70-kDa heat shock protein chaperone nucleotide-binding domain in solution unveiled as a molecular machine that can reorient its functional subdomains, Proc Natl Acad Sci U S A, 101 (2004) 10272-10277.

[91] M. Pellecchia, D.L. Montgomery, S.Y. Stevens, C.W. Vander Kooi, H.P. Feng, L.M. Gierasch, E.R. Zuiderweg, Structural insights into substrate binding by the molecular chaperone DnaK, Nat Struct Biol, 7 (2000) 298-303.

[92] J.F. Swain, G. Dinler, R. Sivendran, D.L. Montgomery, M. Stotz, L.M. Gierasch, Hsp70 chaperone ligands control domain association via an allosteric mechanism mediated by the interdomain linker, Mol Cell, 26 (2007) 27-39.

[93] J.F. Swain, E.G. Schulz, L.M. Gierasch, Direct comparison of a stable isolated Hsp70 substrate-binding domain in the empty and substrate-bound states, J Biol Chem, 281 (2006) 1605-1611.

[94] T. Saio, X. Guan, P. Rossi, A. Economou, C.G. Kalodimos, Structural basis for protein antiaggregation activity of the trigger factor chaperone, Science, 344 (2014) 1250494.

[95] S.T. Hsu, C.M. Dobson, ^{1}H, ^{15}N and ^{13}C assignments of the dimeric ribosome binding domain of trigger factor from Escherichia coli, Biomol NMR Assign, 3 (2009) 17-20.

[96] P. Rajagopal, E. Tse, A.J. Borst, S.P. Delbecq, L. Shi, D.R. Southworth, R.E. Klevit, A conserved histidine modulates HSPB5 structure to trigger chaperone activity in response to stress-related acidosis, Elife, 4 (2015).

[97] S. Jehle, B.S. Vollmar, B. Bardiaux, K.K. Dove, P. Rajagopal, T. Gonen, H. Oschkinat, R.E. Klevit, N-terminal domain of αB-crystallin provides a conformational switch for multimerization and structural heterogeneity, Proc Natl Acad Sci U S A, 108 (2011) 6409-6414.

[98] A.J. Baldwin, P. Walsh, D.F. Hansen, G.R. Hilton, J.L. Benesch, S. Sharpe, L.E. Kay, Probing dynamic conformations of the high-molecular-weight αB-crystallin heat shock protein ensemble by NMR spectroscopy, J Am Chem Soc, 134 (2012) 15343-15350.

[99] R. Rosenzweig, S. Moradi, A. Zarrine-Afsar, J.R. Glover, L.E. Kay, Unraveling the mechanism of protein disaggregation through a ClpB-DnaK interaction, Science, 339 (2013) 1080-1083.

[100] R. Rosenzweig, P. Farber, A. Velyvis, E. Rennella, M.P. Latham, L.E. Kay, ClpB N-terminal domain plays a regulatory role in protein disaggregation, Proc Natl Acad Sci U S A, 112 (2015) E6872-6881.

[101] R. Sprangers, L.E. Kay, Quantitative dynamics and binding studies of the 20S proteasome by NMR, Nature, 445 (2007) 618-622.

[102] V. Tugarinov, R. Sprangers, L.E. Kay, Probing side-chain dynamics in the proteasome by relaxation violated coherence transfer NMR spectroscopy, J Am Chem Soc, 129 (2007) 1743-1750.

[103] M.P. Latham, A. Sekhar, L.E. Kay, Understanding the mechanism of proteasome 20S core particle gating, Proc Natl Acad Sci U S A, 111 (2014) 5532-5537.

[104] A.M. Ruschak, L.E. Kay, Proteasome allostery as a population shift between interchanging conformers, Proc Natl Acad Sci U S A, 109 (2012) E3454-3462.

[105] E.L. Hahn, Spin echoes, Phys Rev, 80 (1950) 580-594.

[106] J.P. Loria, M. Rance, A.G. Palmer, A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy, J Am Chem Soc, 121 (1999) 2331-2332.

[107] D.F. Hansen, P. Vallurupalli, L.E. Kay, An improved 15N relaxation dispersion experiment for the measurement of millisecond time-scale dynamics in proteins, J Phys Chem B, 112 (2008) 5898-5904.

[108] R. Ishima, D.A. Torchia, Extending the range of amide proton relaxation dispersion experiments in proteins using a constant-time relaxation-compensated CPMG approach, J Biomol NMR, 25 (2003) 243-248.

[109] V.Y. Orekhov, D.M. Korzhnev, L.E. Kay, Double- and zero-quantum NMR relaxation dispersion experiments sampling millisecond time scale dynamics in proteins, J Am Chem Soc, 126 (2004) 1886-1891.

[110] F.A.A. Mulder, M. Akke, Carbonyl ^{13}C transverse relaxation measurements to sample protein backbone dynamics, Magn Res Chem, 41 (2003) 853-865.

[111] R. Ishima, J. Baber, J.M. Louis, D.A. Torchia, Carbonyl carbon transverse relaxation dispersion measurements and ms-mu s timescale motion in a protein hydrogen bond network, J Biomol NMR, 29 (2004) 187-198.

[112] D.F. Hansen, P. Vallurupalli, P. Lundström, P. Neudecker, L.E. Kay, Probing chemical shifts of invisible states of proteins with relaxation dispersion NMR spectroscopy: How well can we do?, J Am Chem Soc, 130 (2008) 2667-2675.

[113] P. Lundstöm, D.F. Hansen, L.E. Kay, Measurement of carbonyl chemical shifts of excited protein states by relaxation dispersion NMR spectroscopy: comparison between uniformly and selectively C-13 labeled samples, J Biomol NMR, 42 (2008) 35-47.

[114] P. Vallurupalli, D.F. Hansen, P. Lundstrom, L.E. Kay, CPMG relaxation dispersion NMR experiments measuring glycine $^{1}H^{\alpha}$ and $^{13}C^{a}$ chemical shifts in the 'invisible' excited states of proteins, J Biomol NMR, 45 (2009) 45-55.

[115] P. Lundstöm, D.F. Hansen, P. Vallurupalli, L.E. Kay, Accurate measurement of α proton chemical shifts of excited protein states by relaxation dispersion NMR spectroscopy, J Am Chem Soc, 131 (2009) 1915-1926.

[116] V. Tugarinov, L.E. Kay, Separating degenerate ^{1}H transitions in methyl group probes for single-quantum 1H CPMG relaxation dispersion NMR spectroscopy, J Am Chem Soc, 129 (2007) 9514-9521.

[117] A.J. Baldwin, T.L. Religa, D.F. Hansen, G. Bouvignies, L.E. Kay, $^{13}CHD_2$ methyl group probes of millisecond time scale exchange in proteins by 1H relaxation dispersion: An application to proteasome gating residue dynamics, J Am Chem Soc, 132 (2010) 10992-10995.

[118] P. Lundström, H. Lin, L.E. Kay, Measuring $^{13}C^{\beta}$ chemical shifts of invisible excited states in proteins by relaxation dispersion NMR spectroscopy, J Biomol NMR, 44 (2009) 139-155.

[119] N.R. Skrynnikov, F.A.A. Mulder, B. Hon, F.W. Dahlquist, L.E. Kay, Probing slow time scale dynamics at methyl-containing side chains in proteins by relaxation dispersion NMR measurements: Application to methionine residues in a cavity mutant of T4 lysozyme, J Am Chem Soc, 123 (2001) 4556-4566.

[120] D.M. Korzhnev, K. Kloiber, V. Kanelis, V. Tugarinov, L.E. Kay, Probing slow dynamics in high molecular weight proteins by methyl-TROSY NMR spectroscopy: Application to a 723-residue enzyme, J Am Chem Soc, 126 (2004) 3964-3973.

[121] A.L. Hansen, L.E. Kay, Quantifying millisecond time-scale exchange in proteins by CPMG relaxation dispersion NMR spectroscopy of side-chain carbonyl groups, J Biomol NMR, 50 (2011) 347-355.

[122] F.A.A. Mulder, N.R. Skrynnikov, B. Hon, F.W. Dahlquist, L.E. Kay, Measurement of slow (us-ms) time scale dynamics in protein side chains by ^{15}N relaxation dispersion NMR spectroscopy: Application to Asn and Gln residues in a cavity mutant of T4 lysozyme, J Am Chem Soc, 123 (2001) 967-975.

[123] A. Esadze, D.W. Li, T.Z. Wang, R. Bruschweiler, J. Iwahara, Dynamics of lysine side-chain amino groups in a protein studied by heteronuclear ^{1}H-^{15}N NMR spectroscopy, J Am Chem Soc, 133 (2011) 909-919.

[124] J.P. Loria, M. Rance, A.G. Palmer, A TROSY CPMG sequence for characterizing chemical exchange in large proteins, J Biomol NMR, 15 (1999) 151-155.

[125] P. Vallurupalli, D.F. Hansen, E. Stollar, E. Meirovitch, L.E. Kay, Measurement of bond vector orientations in invisible excited states of proteins, Proc Natl Acad Sci U S A, 104 (2007) 18473-18477.

[126] D.M. Korzhnev, K. Kloiber, L.E. Kay, Multiple-quantum relaxation dispersion NMR spectroscopy probing millisecond time-scale dynamics in proteins: Theory and application, J Am Chem Soc, 126 (2004) 7320-7329.

[127] P. Lundström, P. Vallurupalli, D.F. Hansen, L.E. Kay, Isotope labeling methods for studies of excited protein states by relaxation dispersion NMR spectroscopy, Nature Protocols, 4 (2009) 1641-1648.

[128] P. Lundström, K. Teilum, T. Carstensen, I. Bezsonova, S. Wiesner, F. Hansen, T.L. Religa, M. Akke, L.E. Kay, Fractional ^{13}C enrichment of isolated carbons using [1-^{13}C]- or [2-^{13}C]-glucose facilitates the accurate measurement of dynamics at backbone Ca and side-chain methyl positions in proteins, J Biomol NMR, 38 (2007) 199-212.

[129] D.M. Korzhnev, P. Neudecker, A. Mittermaier, V.Y. Orekhov, L.E. Kay, Multiple-site exchange in proteins studied with a suite of six NMR relaxation dispersion experiments: An application to the folding of a Fyn SH3 domain mutant, J Am Chem Soc, 127 (2005) 15602-15611.

[130] Z. Luz, S. Meiboom, Nuclear magnetic resonance study of the protolysis of trimethylammonium ion in aqueous solution - order of the reaction with respect to solvent, J Chem Phys, 39 (1963) 366-370.

[131] M. Bloom, L.W. Reeves, E.J. Wells, Spin echoes and chemical exchange, J Chem Phys, 42 (1965) 1615-1624.

[132] A. Allerhand, H.S. Gutowsky, Spin-echo studies of chemical exchange. II. Closed formulas for two sites, J Chem Physs, 42 (1965) 1587-1599.

[133] A. Allerhand, E. Thiele, Analysis of Carr-Pursell spin-echo NMR experiments on multiple-spin systems. II. The effect of chemical exchange, J Chem Phys, 45 (1966) 902-916.

[134] J. Jen, Chemical Exchange and NMR T_2 Relaxation - Multisite Case, J Magn Reson, 30 (1978) 111-128.

[135] J.P. Carver, R.E. Richards, General two-site solution for the chemical exchange produced dependence of T_2 upon the Carr-Purcell pulse separation, J Magnc Reson (1969-1992), 6 (1972) 89-105.

[136] D.G. Davis, M.E. Perlman, R.E. London, Direct measurements of the dissociation rate constant for inhibitor-enzyme complexes via the $T_{1\rho}$ and T_2 (CPMG) methods, J Magn Reson B, 104 (1994) 266-275.

[137] R. Ishima, D.A. Torchia, Estimating the time scale of chemical exchange of proteins from measurements of transverse relaxation rates in solution, J Biomol NMR, 14 (1999) 369-372.

[138] F. Bloch, Nuclear induction Phys Rev, 70 (1946) 460-474.

[139] H.M. McConnell, Reaction rates by nuclear magnetic resonance, J Chem Phys, 28 (1958) 430-431.

[140] O. Trott, A.G. Palmer, $R_{1\rho}$ relaxation outside of the fast-exchange limit, J Magn Reson, 154 (2002) 157-160.

[141] D.M. Korzhnev, V.Y. Orekhov, L.E. Kay, Off-resonance $R_{1\rho}$ NMR studies of exchange dynamics in proteins with low spin-lock fields: An application to a fyn SH3 domain, J Am Chem Soc, 127 (2005) 713-721.

[142] M. Akke, A.G. Palmer, Monitoring macromolecular motions on microsecond to millisecond time scales by R(1)rho-R(1) constant relaxation time NMR spectroscopy, J Am Chem Soc, 118 (1996) 911-912.

[143] F.A.A. Mulder, R.A. de Graaf, R. Kaptein, R. Boelens, An off-resonance rotating frame relaxation experiment for the investigation of macromolecular dynamics using adiabatic rotations, J Magn Reson, 131 (1998) 351-357.

[144] D.M. Korzhnev, N.R. Skrynnikov, O. Millet, D.A. Torchia, L.E. Kay, An NMR experiment for the accurate measurement of heteronuclear spin-lock relaxation rates, J Am Chem Soc, 124 (2002) 10743-10753.

[145] T.I. Igumenova, A.G. Palmer, 3rd, Off-resonance TROSY-selected R1rho experiment with improved sensitivity for medium- and high-molecular-weight proteins, J Am Chem Soc, 128 (2006) 8110-8111.

[146] P. Lundstrom, M. Akke, Off-resonance rotating-frame amide proton spin relaxation experiments measuring microsecond chemical exchange in proteins, J Biomol NMR, 32 (2005) 163-173.

[147] C. Eichmuller, N.R. Skrynnikov, A new amide proton R1rho experiment permits accurate characterization of microsecond time-scale conformational exchange, J Biomol NMR, 32 (2005) 281-293.

[148] F.A.A. Mulder, M. Akke, Carbonyl 13C transverse relaxation measurements to sample protein backbone dynamics., Magn Reson Chem, 41 (2003) 853–865.

[149] P. Lundstrom, M. Akke, Microsecond protein dynamics measured by $^{13}C^{\alpha}$ rotating-frame spin relaxation, Chem Biochem, 6 (2005) 1685-1692.

[150] C. Griesinger, R.R. Ernst, Frequency Offset Effects and Their Elimination in NMR Rotating-Frame Cross-Relaxation Spectroscopy, J. Magn. Reson., 75 (1987) 261-271.

[151] T. Yamazaki, R. Muhandiram, L.E. Kay, Nmr Experiments for the Measurement of Carbon Relaxation Properties in Highly Enriched, Uniformly ^{13}C, ^{15}N-Labeled Proteins - Application to $^{13}C^{\alpha}$ Carbons, J Am Chem Soc, 116 (1994) 8266-8278.

[152] D.F. Hansen, L.E. Kay, Improved magnetization alignment schemes for spin-lock relaxation experiments, J Biomol NMR, 37 (2007) 245-255.

[153] F. Massi, E. Johnson, C.Y. Wang, M. Rance, A.G. Palmer, NMR $R_{1\rho}$ rotating-frame relaxation with weak radio frequency fields, J Am Chem Soc, 126 (2004) 2247-2256.

[154] R. Auer, D.F. Hansen, P. Neudecker, D.M. Korzhnev, D.R. Muhandiram, R. Konrat, L.E. Kay, Measurement of signs of chemical shift differences between ground and excited protein states: a comparison between H(S/M)QC and $R_{1\rho}$ methods, J Biomol NMR, 46 (2010) 205-216.

[155] R. Auer, P. Neudecker, D.R. Muhandiram, P. Lundstrom, D.F. Hansen, R. Konrat, L.E. Kay, Measuring the signs of $^{1}H^{\alpha}$ chemical shift differences between ground and excited protein

states by off-resonance spin-lock $R_{1\rho}$ NMR spectroscopy, J Am Chem Soc, 131 (2009) 10832-10833.

[156] A.J. Baldwin, L.E. Kay, Measurement of the signs of methyl ^{13}C chemical shift differences between interconverting ground and excited protein states by $R_{1\rho}$: an application to αB-crystallin, J Biomol NMR, 53 (2012) 1-12.

[157] C. Deverell, R.E. Morgan, J.H. Strange, Studies of chemical exchange by nuclear magnetic relaxation in the rotating frame, Mol Phys, 18 (1970) 553-559.

[158] O. Trott, D. Abergel, A.G. Palmer, An average analysis of R1r relaxation outside of the fast-exchange limit, J Magn Reson, 101 (2003) 753-763.

[159] O. Trott, A.G. Palmer, 3rd, Theoretical study of $R_{1\rho}$ rotating-frame and R_2 free-precession relaxation in the presence of n-site chemical exchange, J Magn Reson, 170 (2004) 104-112.

[160] V.Z. Miloushev, A.G. Palmer, 3rd, $R_{1\rho}$ relaxation for two-site chemical exchange: general approximations and some exact solutions, J Magn Reson, 177 (2005) 221-227.

[161] A.J. Baldwin, L.E. Kay, An $R_{1\rho}$ expression for a spin in chemical exchange between two sites with unequal transverse relaxation rates, J Biomol NMR, 55 (2013) 211-218.

[162] S. Forsén, R.A. Hoffman, Study of moderately rapid chemical exchange reactions by means of nuclear magnetic double resonance, J Chem Phys, 39 (1963) 2892-2901.

[163] N.L. Fawzi, J. Ying, D.A. Torchia, G.M. Clore, Kinetics of amyloid beta monomer-to-oligomer exchange by NMR relaxation, J Am Chem Soc, 132 (2010) 9948-9951.

[164] N.L. Fawzi, J. Ying, R. Ghirlando, D.A. Torchia, G.M. Clore, Atomic-resolution dynamics on the surface of amyloid-beta protofibrils probed by solution NMR, Nature, 480 (2011) 268-272.

[165] P. Vallurupalli, G. Bouvignies, L.E. Kay, Studying "invisible" excited protein states in slow exchange with a major state conformation, J Am Chem Soc, 134 (2012) 8148-8161.

[166] G. Bouvignies, L.E. Kay, Measurement of proton chemical shifts in invisible states of slowly exchanging protein systems by chemical exchange saturation transfer, J Phys Chem B, 116 (2012) 14311-14317.

[167] G. Bouvignies, L.E. Kay, A 2D ^{13}C-CEST experiment for studying slowly exchanging protein systems using methyl probes: an application to protein folding, J Biomol NMR, 53 (2012) 303-310.

[168] P. Vallurupalli, L.E. Kay, Probing slow chemical exchange at carbonyl sites in proteins by chemical exchange saturation transfer NMR spectroscopy, Angew Chem Int Ed Engl, 52 (2013) 4156-4159.

[169] A.L. Hansen, G. Bouvignies, L.E. Kay, Probing slowly exchanging protein systems via $^{13}C^{\alpha}$-CEST: monitoring folding of the Im7 protein, J Biomol NMR, 55 (2013) 279-289.

[170] D. Long, A. Sekhar, L.E. Kay, Triple resonance-based $^{13}C^{\alpha}$ and $^{13}C^{b}$ CEST experiments for studies of ms timescale dynamics in proteins, J Biomol NMR, 60 (2014) 203-208.

[171] E. Rennella, R. Huang, A. Velyvis, L.E. Kay, $^{13}CHD_2$-CEST NMR spectroscopy provides an avenue for studies of conformational exchange in high molecular weight proteins, J Biomol NMR, 63 (2015) 187-199.

[172] M. Sattler, S.W. Fesik, Use of deuterium labeling in NMR: overcoming a sizeable problem, Structure, 4 (1996) 1245-1249.

[173] M.J. Pevin, J. Boisbouvier, Royal Society of Chemistry (Great Britain), Chapter 1:Isotope-Labelling of Methyl Groups for NMR Studies of Large Proteins
From the book: Recent developments in biomolecular NMR, in: J.P. Marius Clore (Ed.) RSC biomolecular sciences,, Royal Society of Chemistry,, Cambridge, U.K., 2012, pp. 1 online resource (xv, 347 p.

[174] V. Tugarinov, V. Kanelis, L.E. Kay, Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy, Nature Protocols, 1 (2006) 749-754.

[175] S.-R. Tzeng, M.-T. Pai, C.G. Kalodimos, NMR studies of large protein systems, Methods in molecular biology (Clifton, N.J.), 831 (2012) 133-140.

[176] N.K. Goto, K.H. Gardner, G.A. Mueller, R.C. Willis, L.E. Kay, A robust and cost-effective method for the production of Val, Leu, Ile (delta 1) methyl-protonated ^{15}N, ^{13}C, ^{2}H-labeled proteins, J Biomol NMR, 13 (1999) 369-374.

[177] V. Tugarinov, V. Kanelis, L.E. Kay, Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy, Nature Protocols, 1 (2006) 749-754.

[178] L.E. Kay, Solution NMR spectroscopy of supra-molecular systems, why bother? A methyl-TROSY view, J Magn Reson, 210 (2011) 159-170.

[179] G.P. Lisi, J. Patrick Loria, Using NMR spectroscopy to elucidate the role of molecular motions in enzyme function, Prog Nucl Magn Reson Spectrosc, 92-93 (2016) 1-17.

[180] M. Salzmann, G. Wider, K. Pervushin, H. Senn, K. Wuthrich, TROSY-type triple-resonance experiments for sequential NMR assignments of large proteins, J Am Chem Soc, 121 (1999) 844-848.

[181] V. Tugarinov, P.M. Hwang, J.E. Ollerenshaw, L.E. Kay, 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 (2003) 10420-10428.

[182] S. Hiller, R.G. Garces, T.J. Malia, V.Y. Orekhov, M. Colombini, G. Wagner, Solution structure of the integral human membrane protein VDAC-1 in detergent micelles, Science, 321 (2008) 1206-1210.

[183] D.M. Korzhnev, K. Kloiber, V. Kanelis, V. Tugarinov, L.E. Kay, Probing slow dynamics in high molecular weight proteins by methyl-TROSY NMR spectroscopy: application to a 723-residue enzyme, J Am Chem Soc, 126 (2004) 3964-3973.

[184] R. Sprangers, A. Gribun, P.M. Hwang, W.A. Houry, L.E. Kay, Quantitative NMR spectroscopy of supramolecular complexes: dynamic side pores in ClpP are important for product release, Proc Natl Acad Sci U S A, 102 (2005) 16678-16683.

[185] M.J. Audin, G. Dorn, S.A. Fromm, K. Reiss, S. Schutz, M.K. Vorlander, R. Sprangers, The archaeal exosome: identification and quantification of site-specific motions that correlate with cap and RNA binding, Angew Chem Int Ed Engl, 52 (2013) 8312-8316.

[186] T.L. Religa, R. Sprangers, L.E. Kay, Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR, Science, 328 (2010) 98-102.

[187] N. Sibille, B. Bersch, J. Coves, M. Blackledge, B. Brutscher, Side chain orientation from methyl 1H-1H residual dipolar couplings measured in highly deuterated proteins, J Am Chem Soc, 124 (2002) 14616-14625.

[188] V. Tugarinov, L.E. Kay, 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 (2003) 13868-13878.

[189] V. Tugarinov, L.E. Kay, 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 (2003) 5701-5706.

[190] I. Gelis, A.M. Bonvin, D. Keramisanou, M. Koukaki, G. Gouridis, S. Karamanou, A. Economou, C.G. Kalodimos, Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR, Cell, 131 (2007) 756-769.

[191] M. Sattler, J. Schleucher, C. Griesinger, Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients, Prog Nucl Magn Reson Spectrosc, 34 (1999) 93-158.

[192] E. Michel, F.H. Allain, Selective amino acid segmental labeling of multi-domain proteins, Methods Enzymol, 565 (2015) 389-422.

[193] G.E. Karagoz, S.G. Rudiger, Hsp90 interaction with clients, Trends Biochem Sci, 40 (2015) 117-125.

[194] D.F. Hansen, P. Vallurupalli, L.E. Kay, Quantifying two-bond (HN) ^{1}H - (CO) ^{13}C and one-bond $^{1}H^{\alpha}$-$^{13}C^{\alpha}$ dipolar couplings of invisible protein states by spin-state selective relaxation dispersion NMR spectroscopy, J Am Chem Soc, 130 (2008) 8397-8405.

[195] P. Vallurupalli, D.F. Hansen, L.E. Kay, Probing structure in invisible protein states with anisotropic NMR chemical shifts, J Am Chem Soc, 130 (2008) 2734-2735.

[196] A. Matouschek, J.T. Kellis, L. Serrano, A.R. Fersht, Mapping the transition-state and pathway of protein folding by protein engineering, Nature, 340 (1989) 122-126.

[197] Y. Shen, O. Lange, F. Delaglio, P. Rossi, J.M. Aramini, G.H. Liu, A. Eletsky, Y.B. Wu, K.K. Singarapu, A. Lemak, A. Ignatchenko, C.H. Arrowsmith, T. Szyperski, G.T. Montelione, D.

Baker, A. Bax, Consistent blind protein structure generation from NMR chemical shift data, Proc Natl Acad Sci U S A, 105 (2008) 4685-4690.

[198] A. Cavalli, X. Salvatella, C.M. Dobson, M. Vendruscolo, Protein structure determination from NMR chemical shifts, Proc Natl Acad Sci U S A, 104 (2007) 9615-9620.

[199] D.S. Wishart, D. Arndt, M. Berjanskii, P. Tang, J. Zhou, G. Lin, CS23D: a web server for rapid protein structure generation using NMR chemical shifts and sequence data, Nucleic Acids Res, 36 (2008) W496-W502.

[200] M. Vendruscolo, Determination of conformationally heterogeneous states of proteins, Curr Opin Struct Biol, 17 (2007) 15-20.

[201] M.R. Jensen, R.W. Ruigrok, M. Blackledge, Describing intrinsically disordered proteins at atomic resolution by NMR, Curr Opin Struct Biol, 23 (2013) 426-435.

[202] D.M. Korzhnev, P. Neudecker, A. Zarrine-Afsar, A.R. Davidson, L.E. Kay, Abp1p and Fyn SH3 domains fold through similar low-populated intermediate states. Biochemistry, 45 (2006) 10175-10183.

[203] W.Y. Choy, Z. Zhou, Y.W. Bai, L.E. Kay, An ^{15}N NMR spin relaxation dispersion study of the folding of a pair of engineered mutants of apocytochrome b562, J Am Chem Soc, 127 (2005) 5066-5072.

[204] Y. Pustovalova, P. Kukic, M. Vendruscolo, D.M. Korzhnev, Probing the Residual Structure of the Low Populated Denatured State of ADA2h under Folding Conditions by Relaxation Dispersion Nuclear Magnetic Resonance Spectroscopy, Biochemistry, 54 (2016) 4611-4622.

[205] I. Bezsonova, D.M. Korzhnev, R.S. Prosser, J.D. Forman-Kay, L.E. Kay, Hydration and packing along the folding pathway of a pair of SH3 domains by pressure dependent NMR, Biochemistry, 45 (2006) 4711-4719.

[206] D.M. Korzhnev, I. Bezsonova, F. Evanics, N. Taulier, Z. Zhou, Y.W. Bai, T.V. Chalikian, R.S. Prosser, L.E. Kay, Probing the transition state ensemble of a protein folding reaction by pressure-dependent NMR relaxation dispersion, J Am Chem Soc, 128 (2006) 5262-5269.

[207] V. Tugarinov, D.S. Libich, V. Meyer, J. Roche, G.M. Clore, The energetics of a three-state protein folding system probed by high-pressure relaxation dispersion NMR spectroscopy, Angew Chem Int Ed Engl, 54 (2015) 11157-11161.

[208] D.M. Korzhnev, T.L. Religa, P. Lundström, A.R. Fersht, L.E. Kay, The folding pathway of an FF domain: Characterization of an on-pathway intermediate state under folding conditions by ^{15}N, $^{13}C^{\alpha}$ and ^{13}C-methyl relaxation dispersion and $^{1}H/^{2}H$-exchange NMR spectroscopy. J Mol Biol, 37 (2007) 497-512.

[209] A. Mittermaier, D.M. Korzhnev, L.E. Kay, Side-chain interactions in the folding pathway of a Fyn SH3 domain mutant studied by relaxation dispersion NMR spectroscopy, Biochemistry, 44 (2005) 15430-15436.

[210] P. Neudecker, A. Zarrine-Afsar, A.R. Davidson, L.E. Kay, F-value analysis of a three-state protein folding pathway by NMR relaxation dispersion spectroscopy, Proc Natl Acad Sci U S A, 104 (2007) 15717-15722.

[211] T.V. Chalikian, Volumetric properties of proteins, Annu Rev Biophys Biomol Struct, 32 (2003) 207-235.

[212] M. Turowski, N. Yamakawa, J. Meller, K. Kimata, T. Ikegami, K. Hosoya, N. Tanaka, E.R. Thornton, Deuterium isotope effects on hydrophobic interactions: The importance of dispersion interactions in the hydrophobic phase, J Am Chem Soc, 125 (2003) 13836-13849.

[213] L.S. Itzhaki, P.A. Evans, Solvent isotope effects on the refolding kinetics of hen egg-white lysozyme, Protein Sci, 5 (1996) 140-146.

[214] A.P. Capaldi, C. Kleanthous, S.E. Radford, Im7 folding mechanism: misfolding on a path to the native state, Nat Struct Biol, 9 (2002) 209-216.

[215] G. Schreiber, A.R. Fersht, Energetics of protein-protein interactions - analysis of the barnase-barstar interface by single mutations and double mutant cycles, J Mol Biol, 248 (1995) 478-486.

[216] D.S. Wishart, D.A. Case, Use of chemical shifts in macromolecular structure determination, Methods in Enzymology, Nuclear Magnetic Resonance of Biological Macromolecules, Pt A, 338 (2001) 3-34.

[217] D.S. Wishart, Interpreting protein chemical shift data, Prog Nucl Magn Reson Spectrosc, 58 (2011) 62-87.

[218] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, TALOS plus: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts, J Biomol NMR, 44 (2009) 213-223.

[219] Y. Shen, A. Bax, Protein structural information derived from NMR chemical shift with the neural network program TALOS-N, Methods Mol Biol, 1260 (2015) 17-32.

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

[221] M.V. Berjanskii, D.S. Wishart, A simple method to predict protein flexibility using secondary chemical shifts, J Am Chem Soc, 127 (2005) 14970-14971.

[222] B. Fu, M. Vendruscolo, Structure and Dynamics of Intrinsically Disordered Proteins, Adv Exp Med Biol, 870 (2015) 35-48.

[223] D.M. Korzhnev, R.M. Vernon, T.L. Religa, A.L. Hansen, D. Baker, A.R. Fersht, L.E. Kay, Nonnative interactions in the FF domain folding pathway from an atomic resolution structure of a sparsely populated intermediate: An NMR relaxation dispersion study, J Am Chem Soc, 133 (2011) 10974-10982.

[224] N.R. Skrynnikov, F.W. Dahlquist, L.E. Kay, Reconstructing NMR spectra of "invisible" excited protein states using HSQC and HMQC experiments, J Am Chem Soc, 124 (2002) 12352-12360.

[225] J.H. Prestegard, H.M. Al-Hashimi, J.R. Tolman, NMR structures of biomolecules using field oriented media and residual dipolar couplings, Quart Rev Biophys, 33 (2000) 371-424.

[226] J.H. Prestegard, C.M. Bougault, A.I. Kishore, Residual dipolar couplings in structure determination of biomolecules, Chem Rev, 104 (2004) 3519-3540.

[227] N. Tjandra, A. Bax, Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium, Science, 278 (1997) 1111-1114.

[228] M.R. Hansen, L. Mueller, A. Pardi, Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions, Nat Struct Biol, 5 (1998) 1065-1074.

[229] M. Ruckert, G. Otting, Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments, J Am Chem Soc, 122 (2000) 7793-7797.

[230] J.J. Chou, J.D. Kaufman, S.J. Stahl, P.T. Wingfield, A. Bax, Micelle-induced curvature in a water-insoluble HIV-1 Env peptide revealed by NMR dipolar coupling measurement in stretched polyacrylamide gel, J Am Chem Soc, 124 (2002) 2450-2451.

[231] P. Vallurupalli, D.F. Hansen, L.E. Kay, Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy, Proc Natl Acad Sci U S A, 105 (2008) 11766-11771.

[232] M. Ottiger, F. Delaglio, A. Bax, Measurement of J and Dipolar Couplings from Simplified Two-Dimensional NMR Spectra, J Magn Reson 131 (1998) 373-378.

[233] L. Yao, J. Ying, A. Bax, Improved accuracy of ^{15}N–^{1}H scalar and residual dipolar couplings from gradient-enhanced IPAP-HSQC experiments on protonated proteins, J. Biomol. NMR, 43 (2009) 161-170.

[234] A.J. Baldwin, D.F. Hansen, P. Vallurupalli, L.E. Kay, Measurement of methyl axis orientations in invisible, excited states of proteins by relaxation dispersion NMR spectroscopy, J Am Chem Soc, 131 (2009) 11939-11948.

[235] Y. Shen, A. Bax, Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology, J Biomol NMR, 38 (2007) 289-302.

[236] Y. Shen, A. Bax, SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network, J Biomol NMR, 48 (2010) 13-22.

[237] S. Neal, A.M. Nip, H.Y. Zhang, D.S. Wishart, Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts, J Biomol NMR, 26 (2003) 215-240.

[238] X.P. Xu, D.A. Case, Automated prediction of ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and $^{13}C'$ chemical shifts in proteins using a density functional database, J Biomol NMR, 21 (2001) 321-333.

[239] K.J. Kohlhoff, P. Robustelli, A. Cavalli, X. Salvatella, M. Vendruscolo, Fast and accurate predictions of protein NMR chemical shifts from interatomic distances, J Am Chem Soc, 131 (2009) 13894-13895.

[240] S. Raman, O.F. Lange, P. Rossi, M. Tyka, X. Wang, J. Aramini, G.H. Liu, T.A. Ramelot, A. Eletsky, T. Szyperski, M.A. Kennedy, J. Prestegard, G.T. Montelione, D. Baker, NMR structure determination for larger proteins using backbone-only data, Science, 327 (2010) 1014-1018.

[241] D.M. Korzhnev, M. Billeter, A.S. Arseniev, V.Y. Orekhov, NMR studies of Brownian tumbling and internal motions in proteins, Prog Nucl Magn Reson Spectrosc, 38 (2001) 197-266.

[242] A.E. Torda, R.M. Scheek, W.F. VanGunsteren, Time-dependent distance restraints in molecular dynamics simulations, Chem Phys Letts, 157 (1989) 289-294.

[243] C.J. Francis, K. Lindorff-Larsen, R.B. Best, M. Vendruscolo, Characterization of the residual structure in the unfolded state of the Delta 131 Delta fragment of staphylococcal nuclease, Proteins-Struct Funct Bioinformatics, 65 (2006) 145-152.

[244] J. Gsponer, H. Hopearuoho, S.B.M. Whittaker, G.R. Spence, G.R. Moore, E. Paci, S.E. Radford, M. Vendruscolo, Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7, Proc Natl Acad Sci U S A, 103 (2006) 99-104.

[245] X. Salvatella, C.M. Dobson, A.R. Fersht, M. Vendruscolo, Determination of the folding transition states of barnase by using Phi(I)-value-restrained simulations validated by double mutant Phi(IJ)-values, Proc Natl Acad Sci U S A, 102 (2005) 12389-12394.

[246] J.A. Marsh, C. Neale, F.E. Jack, W.Y. Choy, A.Y. Lee, K.A. Crowhurst, J.D. Forman-Kay, Improved structural characterizations of the drkN SH3 domain unfolded state suggest a compact ensemble with native-like and non-native structure, J Mol Biol, 367 (2007) 1494-1510.

[247] J.A. Marsh, J.D. Forman-Kay, Ensemble modeling of protein disordered states: experimental restraint contributions and validation, Proteins, 80 (2012) 556-572.

[248] G. Nodet, L. Salmon, V. Ozenne, S. Meier, M.R. Jensen, M. Blackledge, Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings, J Am Chem Soc, 131 (2009) 17908-17918.

[249] M.R. Jensen, L. Salmon, G. Nodet, M. Blackledge, Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts, J Am Chem Soc, 132 (2010) 1270-1272.

[250] H.Q. Feng, Z. Zhou, Y.W. Bai, A protein folding pathway with multiple folding intermediates at atomic resolution, Proc Natl Acad Sci U S A, 102 (2005) 5026-5031.

[251] T.L. Religa, J.S. Markson, U. Mayor, S.M.V. Freund, A.R. Fersht, Solution structure of a protein denatured state and folding intermediate, Nature, 437 (2005) 1053-1056.

[252] S.B.M. Whittaker, G.R. Spence, J.G. Grossmann, S.E. Radford, G.R. Moore, NMR analysis of the conformational properties of the trapped on-pathway folding intermediate of the bacterial immunity protein Im7, J Mol Biol, 366 (2007) 1001-1015.

[253] K. Wüthrich, NMR of proteins and nucleic acids, Wiley, New York, 1986.

[254] D.M. Korzhnev, T.L. Religa, L.E. Kay, Transiently populated intermediate functions as a branching point of the FF domain folding pathway, Proc Natl Acad Sci U S A, 109 (2012) 17777-17782.

[255] A. Sekhar, P. Vallurupalli, L.E. Kay, Folding of the four-helix bundle FF domain from a compact on-pathway intermediate state is governed predominantly by water motion, Proc Natl Acad Sci U S A, 109 (2012) 19268-19273.

[256] A. Sekhar, M.P. Latham, P. Vallurupalli, L.E. Kay, Viscosity-dependent kinetics of protein conformational exchange: microviscosity effects and the need for a small viscogen, J Phys Chem B, 118 (2014) 4546-4551.

[257] X. Ai, Z. Zhou, Y. Bai, W.Y. Choy, 15N NMR spin relaxation dispersion study of the molecular crowding effects on protein folding under native conditions, J Am Chem Soc, 128 (2006) 3916-3917.

[258] M. Zeeb, J. Balbach, NMR spectroscopic characterization of millisecond protein folding by transverse relaxation dispersion measurements, J Am Chem Soc, 127 (2005) 13207-13212.

[259] S.B. Whittaker, N.J. Clayden, G.R. Moore, NMR characterisation of the relationship between frustration and the excited state of Im7, J Mol Biol, 414 (2011) 511-529.

[260] C. Sanchez-Medina, A. Sekhar, P. Vallurupalli, M. Cerminara, V. Munoz, L.E. Kay, Probing the free energy landscape of the fast-folding gpW protein by relaxation dispersion NMR, J Am Chem Soc, 136 (2014) 7444-7451.

[261] M. Tollinger, K. Kloiber, B. Agoston, C. Dorigoni, R. Lichtenecker, W. Schmid, R. Konrat, An isolated helix persists in a sparsely populated form of KIX under native conditions, Biochemistry, 45 (2006) 8885-8893.

[262] P. Schanda, B. Brutscher, R. Konrat, M. Tollinger, Folding of the KIX domain: characterization of the equilibrium analog of a folding intermediate using 15N/13C relaxation dispersion and fast 1H/2H amide exchange NMR spectroscopy, J Mol Biol, 380 (2008) 726-741.

[263] P. Farber, H. Darmawan, T. Sprules, A. Mittermaier, Analyzing protein folding cooperativity by differential scanning calorimetry and NMR spectroscopy, J Am Chem Soc, 132 (2010) 6214-6222.

[264] P.J. Farber, J. Slager, A.K. Mittermaier, Local folding and misfolding in the PBX homeodomain from a three-state analysis of CPMG relaxation dispersion NMR data, J Phys Chem B, 116 (2012) 10317-10329.

[265] M.J. Grey, Y. Tang, E. Alexov, C.J. McKnight, D.P. Raleigh, A.G. Palmer, 3rd, Characterizing a partially folded intermediate of the villin headpiece domain under non-denaturing conditions: contribution of His41 to the pH-dependent stability of the N-terminal subdomain, J Mol Biol, 355 (2006) 1078-1094.

[266] N.E. O'Connell, M.J. Grey, Y. Tang, P. Kosuri, V.Z. Miloushev, D.P. Raleigh, A.G. Palmer, 3rd, Partially folded equilibrium intermediate of the villin headpiece HP67 defined by 13C relaxation dispersion, J Biomol NMR, 45 (2009) 85-98.

[267] J.H. Cho, N. O'Connell, D.P. Raleigh, A.G. Palmer, 3rd, Phi-value analysis for ultrafast folding proteins by NMR relaxation dispersion, J Am Chem Soc, 132 (2010) 450-451.

[268] D.W. Meinhold, P.E. Wright, Measurement of protein unfolding/refolding kinetics and structural characterization of hidden intermediates by NMR relaxation dispersion, Proc Natl Acad Sci U S A, 108 (2011) 9078-9083.

[269] M. Allen, A. Friedler, O. Schon, M. Bycroft, The structure of an FF domain from human HYPA/FBP11, J Mol Biol, 323 (2002) 411-416.

[270] P. Jemth, S. Gianni, R. Day, B. Li, C.M. Johnson, V. Daggett, A.R. Fersht, Demonstration of a low-energy on-pathway intermediate in a fast-folding protein by kinetics, protein engineering, and simulation, Proc Natl Acad Sci U S A, 101 (2004) 6450-6455.

[271] P. Jemth, R. Day, S. Gianni, F. Khan, M. Allen, V. Daggett, A.R. Fersht, The structure of the major transition state for folding of an FF domain from experiment and simulation, J Mol Biol, 350 (2005) 363-378.

[272] A.A. Di Nardo, D.M. Korzhnev, P.J. Stogios, A. Zarrine-Afsar, L.E. Kay, A.R. Davidson, Dramatic acceleration of protein folding by stabilization of a nonnative backbone conformation, Proc Natl Acad Sci U S A, 101 (2004) 7954-7959.

[273] C. Clementi, S.S. Plotkin, The effects of nonnative interactions on protein folding rates: Theory and simulation, Protein Sci, 13 (2004) 1750-1766.

[274] L. Li, L.A. Mirny, E.I. Shakhnovich, Kinetics, thermodynamics and evolution of non-native interactions in a protein folding nucleus, Nat Struct Biol, 7 (2000) 336-342.

[275] T.R. Sosnick, L. Mayne, R. Hiller, S.W. Englander, The barriers in protein-folding, Nat Struct Biol, 1 (1994) 149-156.

[276] N.A. Farrow, O.W. Zhang, J.D. Formankay, L.E. Kay, A Heteronuclear correlation experiment for simultaneous determination of ^{15}N longitudinal decay and chemical-exchange rates of systems in slow equilibrium, J Biomol NMR, 4 (1994) 727-734.

[277] J. Barette, A. Velyvis, T.L. Religa, D.M. Korzhnev, L.E. Kay, Cross-validation of the structure of a transiently formed and low populated FF domain folding intermediate determined by relaxation dispersion NMR and CS-Rosetta, J Phys Chem B, 116 (2012) 6637-6644.

[278] A. Gershenson, L.M. Gierasch, Protein folding in the cell: challenges and progress, Curr Opin Struct Biol, 21 (2011) 32-41.

[279] S. Pechmann, F. Willmund, J. Frydman, The ribosome as a hub for protein quality control, Mol Cell, 49 (2013) 411-421.

[280] N. Gregersen, P. Bross, S. Vang, J.H. Christensen, Protein misfolding and human disease, Annu Rev Genomics Hum Genet, 7 (2006) 103-124.

[281] G. Zhang, Z. Ignatova, Folding at the birth of the nascent chain: coordinating translation with co-translational folding, Curr Opin Struct Biol, 21 (2011) 25-31.

[282] F. Gloge, A.H. Becker, G. Kramer, B. Bukau, Co-translational mechanisms of protein maturation, Curr Opin Struct Biol, 24 (2014) 24-33.

[283] C.A. Waudby, H. Launay, L.D. Cabrita, J. Christodoulou, Protein folding on the ribosome studied using NMR spectroscopy, Prog Nucl Magn Reson Spectrosc, 74 (2013) 57-75.

[284] O. Brandman, R.S. Hegde, Ribosome-associated protein quality control, Nat Struct Mol Biol, 23 (2016) 7-15.

[285] F.U. Hartl, M. Hayer-Hartl, Converging concepts of protein folding in vitro and in vivo, Nat Struct Mol Biol, 16 (2009) 574-581.

[286] M.S. Hipp, S.H. Park, F.U. Hartl, Proteostasis impairment in protein-misfolding and -aggregation diseases, Trends Cell Biol, 24 (2014) 506-514.

[287] S. Wickner, M.R. Maurizi, S. Gottesman, Posttranslational quality control: folding, refolding, and degrading proteins, Science, 286 (1999) 1888-1893.

[288] H. Saibil, Chaperone machines for protein folding, unfolding and disaggregation, Nat Rev Mol Cell Biol, 14 (2013) 630-642.

[289] J.C. Young, J.M. Barral, F. Ulrich Hartl, More than folding: localized functions of cytosolic chaperones, Trends Biochem Sci, 28 (2003) 541-547.

[290] D. Reichmann, Y. Xu, C.M. Cremers, M. Ilbert, R. Mittelman, M.C. Fitzgerald, U. Jakob, Order out of disorder: working cycle of an intrinsically unfolded chaperone, Cell, 148 (2012) 947-957.

[291] O. Suss, D. Reichmann, Protein plasticity underlines activation and function of ATP-independent chaperones, Front Mol Biosci, 2 (2015) 43.

[292] R. Rosenzweig, L.E. Kay, Solution NMR Spectroscopy Provides an Avenue for the Study of Functionally Dynamic Molecular Machines: The Example of Protein Disaggregation, J Am Chem Soc, 138 (2016) 1466-1477.

[293] M.P. Mayer, R. Kityk, Insights into the molecular mechanism of allostery in Hsp70s, Front Mol Biosci, 2 (2015) 58.

[294] E.R. Zuiderweg, E.B. Bertelsen, A. Rousaki, M.P. Mayer, J.E. Gestwicki, A. Ahmad, Allostery in the Hsp70 chaperone proteins, Top Curr Chem, 328 (2013) 99-153.

[295] E.M. Clerico, J.M. Tilitsky, W. Meng, L.M. Gierasch, How hsp70 molecular machines interact with their substrates to mediate diverse physiological functions, J Mol Biol, 427 (2015) 1575-1588.

[296] S. Rudiger, A. Buchberger, B. Bukau, Interaction of Hsp70 chaperones with substrates, Nat Struct Biol, 4 (1997) 342-349.

[297] R. Kityk, J. Kopp, I. Sinning, M.P. Mayer, Structure and dynamics of the ATP-bound open conformation of Hsp70 chaperones, Mol Cell, 48 (2012) 863-874.

[298] R. Qi, E.B. Sarbeng, Q. Liu, K.Q. Le, X. Xu, H. Xu, J. Yang, J.L. Wong, C. Vorvis, W.A. Hendrickson, L. Zhou, Q. Liu, Allosteric opening of the polypeptide-binding site when an Hsp70 binds ATP, Nat Struct Mol Biol, 20 (2013) 900-907.

[299] R.U. Mattoo, P. Goloubinoff, Molecular chaperones are nanomachines that catalytically unfold misfolded and alternatively folded proteins, Cell Mol Life Sci, 71 (2014) 3311-3325.

[300] J.H. Lee, D. Zhang, C. Hughes, Y. Okuno, A. Sekhar, S. Cavagnero, Heterogeneous binding of the SH3 client protein to the DnaK molecular chaperone, Proc Natl Acad Sci U S A, 112 (2015) E4206-4215.

[301] A. Bracher, J. Verghese, The nucleotide exchange factors of Hsp70 molecular chaperones, Front Mol Biosci, 2 (2015) 10.

[302] N.B. Nillegoda, B. Bukau, Metazoan Hsp70-based protein disaggregases: emergence and mechanisms, Front Mol Biosci, 2 (2015) 57.

[303] S.L. Dekker, H.H. Kampinga, S. Bergink, DNAJs: more than substrate delivery to HSPA, Front Mol Biosci, 2 (2015) 35.

[304] F.X. Theillet, A. Binolfi, T. Frembgen-Kesner, K. Hingorani, M. Sarkar, C. Kyne, C. Li, P.B. Crowley, L. Gierasch, G.J. Pielak, A.H. Elcock, A. Gershenson, P. Selenko, Physicochemical properties of cells and their effects on intrinsically disordered proteins (IDPs), Chem Rev, 114 (2014) 6661-6714.

[305] R. van der Lee, M. Buljan, B. Lang, R.J. Weatheritt, G.W. Daughdrill, A.K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D.T. Jones, P.M. Kim, R.W. Kriwacki, C.J. Oldfield, R.V. Pappu, P. Tompa, V.N. Uversky, P.E. Wright, M.M. Babu, Classification of intrinsically disordered regions and proteins, Chem Rev, 114 (2014) 6589-6631.

[306] V.N. Uversky, Dancing Protein Clouds: The Strange Biology and Chaotic Physics of Intrinsically Disordered Proteins, J Biol Chem, 291 (2016) 6681-6688.

[307] H.J. Dyson, Expanding the proteome: disordered and alternatively folded proteins, Q Rev Biophys, 44 (2011) 467-518.

[308] H.J. Dyson, Making Sense of Intrinsically Disordered Proteins, Biophys J, 110 (2016) 1013-1016.

[309] V.N. Uversky, C.J. Oldfield, A.K. Dunker, Intrinsically disordered proteins in human diseases: introducing the D2 concept, Annu Rev Biophys, 37 (2008) 215-246.

[310] M.J. Smith, C.B. Marshall, F.X. Theillet, A. Binolfi, P. Selenko, M. Ikura, Real-time NMR monitoring of biological activities in complex physiological environments, Curr Opin Struct Biol, 32 (2015) 39-47.

[311] E. Luchinat, L. Banci, A Unique Tool for Cellular Structural Biology: In-cell NMR, J Biol Chem, 291 (2016) 3776-3784.

[312] F.X. Theillet, A. Binolfi, B. Bekei, A. Martorana, H.M. Rose, M. Stuiver, S. Verzini, D. Lorenz, M. van Rossum, D. Goldfarb, P. Selenko, Structural disorder of monomeric α-synuclein persists in mammalian cells, Nature, 530 (2016) 45-50.

[313] A. Binolfi, A. Limatola, S. Verzini, J. Kosten, F.X. Theillet, H. May Rose, B. Bekei, M. Stuiver, M. van Rossum, P. Selenko, Intracellular repair of oxidation-damaged α-synuclein fails to target C-terminal modification sites, Nat Commun, 7 (2016) 10251.

[314] S.N. Witt, Molecular chaperones, α-synuclein, and neurodegeneration, Mol Neurobiol, 47 (2013) 552-560.

[315] T.R. Alderson, J.L. Markley, Biophysical characterization of α-synuclein and its controversial structure, Intrinsically Disord Proteins, 1 (2013) 18-39.

[316] I. Dikiy, D. Eliezer, Folding and misfolding of α-synuclein on membranes, Biochim Biophys Acta, 1818 (2012) 1013-1018.

[317] T. Wang, J.C. Hay, A-synuclein Toxicity in the Early Secretory Pathway: How It Drives Neurodegeneration in Parkinsons Disease, Front Neurosci, 9 (2015) 433.

[318] D. Eliezer, E. Kutluay, R. Bussell, Jr., G. Browne, Conformational properties of α-synuclein in its free and lipid-associated states, J Mol Biol, 307 (2001) 1061-1073.

[319] A.S. Maltsev, J. Chen, R.L. Levine, A. Bax, Site-specific interaction between α-synuclein and membranes probed by NMR-observed methionine oxidation rates, J Am Chem Soc, 135 (2013) 2943-2946.

[320] K.E. Paleologou, A.W. Schmid, C.C. Rospigliosi, H.Y. Kim, G.R. Lamberto, R.A. Fredenburg, P.T. Lansbury, Jr., C.O. Fernandez, D. Eliezer, M. Zweckstetter, H.A. Lashuel, Phosphorylation at Ser-129 but not the phosphomimics S129E/D inhibits the fibrillation of α-synuclein, J Biol Chem, 283 (2008) 16895-16905.

[321] I. Dikiy, D. Eliezer, N-terminal acetylation stabilizes N-terminal helicity in lipid- and micelle-bound α-synuclein and increases its affinity for physiological membranes, J Biol Chem, 289 (2014) 3652-3665.

[322] B. Fauvet, M.B. Fares, F. Samuel, I. Dikiy, A. Tandon, D. Eliezer, H.A. Lashuel, Characterization of semisynthetic and naturally Nα-acetylated α-synuclein in vitro and in intact cells: implications for aggregation and cellular properties of α-synuclein, J Biol Chem, 287 (2012) 28243-28262.

[323] J. Bai, K. Cheng, M. Liu, C. Li, Impact of the α-Synuclein Initial Ensemble Structure on Fibrillation Pathways and Kinetics, J Phys Chem B, 120 (2016) 3140-3147.

[324] T.K. Karamanos, A.P. Kalverda, G.S. Thompson, S.E. Radford, Mechanisms of amyloid formation revealed by solution NMR, Prog Nucl Magn Reson Spectrosc, 88-89 (2015) 86-104.

[325] A. Binolfi, F.X. Theillet, P. Selenko, Bacterial in-cell NMR of human α-synuclein: a disordered monomer by nature?, Biochem Soc Trans, 40 (2012) 950-954.

[326] C.A. Waudby, C. Camilloni, A.W. Fitzpatrick, L.D. Cabrita, C.M. Dobson, M. Vendruscolo, J. Christodoulou, In-cell NMR characterization of the secondary structure populations of a disordered conformation of α-synuclein within E. coli cells, PLoS One, 8 (2013) e72286.

[327] T. Bartels, J.G. Choi, D.J. Selkoe, α-Synuclein occurs physiologically as a helically folded tetramer that resists aggregation, Nature, 477 (2011) 107-110.

[328] D.I. Freedberg, P. Selenko, Live cell NMR, Annu Rev Biophys, 43 (2014) 171-192.

[329] A.M. Bodles, D.J. Guthrie, B. Greer, G.B. Irvine, Identification of the region of non-Abeta component (NAC) of Alzheimer's disease amyloid responsible for its aggregation and toxicity, J Neurochem, 78 (2001) 384-395.

[330] T. Bartels, N.C. Kim, E.S. Luth, D.J. Selkoe, N-α-acetylation of α-synuclein increases its helical folding propensity, GM1 binding specificity and resistance to aggregation, PLoS One, 9 (2014) e103727.

[331] E. Colla, P.H. Jensen, O. Pletnikova, J.C. Troncoso, C. Glabe, M.K. Lee, Accumulation of toxic α-synuclein oligomer within endoplasmic reticulum occurs in α-synucleinopathy in vivo, J Neurosci, 32 (2012) 3301-3305.

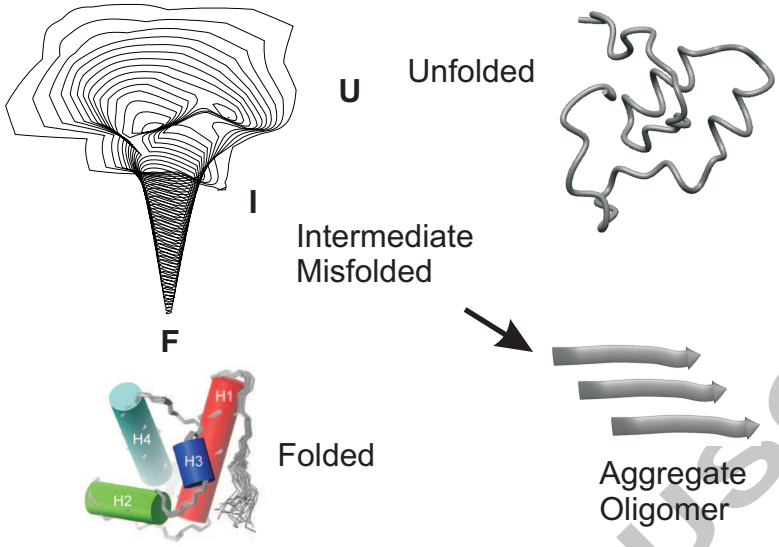
[332] P. Valdes, G. Mercado, R.L. Vidal, C. Molina, G. Parsons, F.A. Court, A. Martinez, D. Galleguillos, D. Armentano, B.L. Schneider, C. Hetz, Control of dopaminergic neuron survival by the unfolded protein response transcription factor XBP1, Proc Natl Acad Sci U S A, 111 (2014) 6804-6809.

[333] D.R. Jones, S. Moussaud, P. McLean, Targeting heat shock proteins to modulate α-synuclein toxicity, Ther Adv Neurol Disord, 7 (2014) 33-51.

[334] U. Jakob, R. Kriwacki, V.N. Uversky, Conditionally and transiently disordered proteins: awakening cryptic disorder to regulate protein function, Chem Rev, 114 (2014) 6779-6805.

[335] N.S. Latysheva, T. Flock, R.J. Weatheritt, S. Chavali, M.M. Babu, How do disordered regions achieve comparable functions to structured domains?, Protein Sci, 24 (2015) 909-922.

[336] K. Nakamura, α-Synuclein and mitochondria: partners in crime?, Neurotherapeutics, 10 (2013) 391-399.



Highlights

NMR emerged as a powerful experimental method to probe protein folding *in vitro* and *in vivo*

NMR RD and ST methods provided detailed insights into folding kinetics and thermodynamics

NMR revealed properties of structural ensembles of nonnative states on the folding energy landscape

NMR provided unique means to explore co-translational and post-translational protein folding

NMR revealed details of how protein folding occurs inside the living cells