

Investigation of Enzyme Catalysis and
Protein-Ligand Interaction Using
Solution-State NMR Spectroscopy

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Marianne Schulte

aus

Troisdorf

Jülich, Dezember 2017

aus dem Institut für Physikalische Biologie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. Dieter Willbold
Koreferent: Prof. Dr. Jörg Pietruszka

Tag der mündlichen Prüfung: 2. Februar 2018

Eidesstattliche Erklärung

Ich, Marianne Schulte, versichere an Eides statt, dass die Dissertation von mir selbst verfasst und ohne unzulässige Hilfe unter Beachtung der Grundsätze zur Sicherung der guten wissenschaftlichen Praxis an der Heinrich-Heine-Universität Düsseldorf erstellt worden ist. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Bei Ergebnissen, welche in Zusammenarbeit mit anderen produziert wurden, habe ich meinen Beitrag genannt.

Ferner erkläre ich, dass ich in keinem anderen Dissertationsverfahren mit oder ohne Erfolg versucht habe, diese Dissertation einzureichen.

Summary

Proteins are intrinsically flexible and their function depends on the three-dimensional structure as well as dynamics. Protein motions are characterized by a number of conformations that constantly undergo exchange processes. Defining the intrinsic dynamics of proteins provides valuable insight into their mode of action. Nuclear magnetic resonance (NMR) spectroscopy, complemented with molecular dynamics (MD) simulations, is a powerful technique to study protein structure and dynamics, at atomic resolution. In this work, a combined NMR and MD simulation study was employed to study the effect of structure and dynamics on the protein function, using two different protein systems.

2-Deoxyribose-5-phosphate aldolase (DERA) catalyzes the reversible conversion of 2-deoxyribose-5-phosphate (dR5P) into acetaldehyde and glyceraldehyde-3-phosphate (G3P) and is involved in nucleotide catabolism. Its unique ability to catalyze C-C bond formation between two aldehydes and a broad acceptor-substrate tolerance makes it an attractive biocatalyst. The reaction mechanism of the *E. coli* DERA (*ecDERA*), a symmetric homodimer with 259 amino acids per monomeric unit, has been proposed based on crystal structures of *ecDERA* with reaction intermediates and mutagenesis studies [1]. Although numerous biochemical studies have highlighted the importance of the C-terminal tyrosine (Y259) in the catalytic reaction, the structural basis of its participation remained elusive due to missing electron density of the last 8 to 9 C-terminal residues (C-terminal tail) in the available crystal structures. Using NMR spectroscopy on a monomer variant of the *ecDERA* and its mutants, transient sampling of catalytically-relevant conformations of the C-terminal tail, where Y259 enters the active site (closed state), was observed in the absence of the substrate. A significantly populated closed state gave rise to nuclear Overhauser effect (NOE) cross-peaks in NOE spectroscopy (NOESY) experiments, based on which the first solution structure of the closed state ensemble of the C-terminal tail was derived. Extensive MD simulations, complementing the NMR results, were further used to characterize the transient interactions stabilizing the C-terminal tail in the closed state conformation. Moreover, hydrogen/deuterium exchange experiments revealed that Y259 side chain is necessary for efficient substrate-proton abstraction and thus, may facilitate the imine

to enamine transition in the *ec*DERA reaction mechanism. Furthermore, besides the known phosphate binding site identified in *ec*DERA crystal structures, a previously unknown auxiliary binding site was identified, located at the C-terminal tail, using chemical shift perturbation analysis. This phosphate binding site may dock onto the substrate's phosphate group, thereby stabilizing the optimal position of Y259 during the catalytic reaction.

Human Nedd4-1 (hNedd4-1) (Neuronal precursor cell expressed developmentally down-regulated gene 4-1) is an E3 ubiquitin ligase, that mediates the transfer of ubiquitin from the cognate E2 ubiquitin-conjugating enzyme to the substrate. The first recognized target of Nedd4-1 was the human epithelial Na⁺ channel (hENaC). The characteristic architecture of Nedd4 family proteins comprises three to four WW domains sandwiched between an N-terminal C2 domain and a C-terminal HECT domain. Substrate specificity and affinity is defined by the protein-protein interaction domains; the WW domains. WW domains of hNedd4-1 recognize the PPxY motif from hENaC (PY motif) and have high sequence similarity amongst themselves. However, the affinities to the PY motif vary significantly, with the third WW domain (WW3*) showing the highest affinity. Using NMR spectroscopy, WW3* domain was found to exist in an equilibrium between a folded, binding-competent, and a random coil-like state in solution. The PY motif binds to the binding-competent state, thereby shifting the folding equilibrium towards the folded state. Hence, WW3*-PY motif interaction occurs through coupled folding and binding equilibria.

The highest sequence diversity amongst the four hNedd4-1 WW domains lies within the loop connecting the first two β -strands (loop I). Loop I of WW3* adopts a type I β -turn. Type I turns are known to form within 100 ns, thereby allowing fast folding. A highly statistically preferred proline is found at $i+1$ position in loop I of WW3*, but not the other three hNedd4-1 WW domains. Characterization of loop I stability and dynamics was accomplished using a combination of NMR spectroscopy and MD simulations. Exchange of the loop I proline from WW3* domain into the corresponding residue from WW4 (threonine) resulted in a dynamic seven residue loop rather than the wild-type type I β -turn. Peptide binding, however, locks this loop in a type I β -turn for both, wild-type and P433T mutant of WW3*. Moreover, the P to T mutation in loop I not only alters loop I stability but also the overall stability of the apo-WW3* P433T mutant, evident from a 6 °C decrease in melting temperature. Therefore, presence of proline in loop I enables presence of a stable type I β -turn, which may affect the affinity towards the α -hENaC peptide.

Zusammenfassung

Proteine sind intrinsisch flexibel und ihre Funktion hängt sowohl von der dreidimensionalen Struktur als auch von der Dynamik ab. Proteinbewegungen sind durch eine Anzahl von Konformationen gekennzeichnet, die ständige Austauschprozesse durchlaufen. Die Charakterisierung der intrinsischen Dynamik von Proteinen gibt wertvolle Einblicke in ihre Wirkungsweise. Kernspinresonanz (NMR, engl.: Nuclear Magnetic Resonance)-Spektroskopie, ergänzt durch Moleküldynamik (MD)-Simulationen, ist eine leistungsfähige Methode zur Untersuchung der Proteinstruktur als auch ihrer Dynamik auf atomarer Ebene. In dieser Arbeit wurde eine kombinierte NMR- und MD-Simulations Studie verwendet, um den Effekt von Struktur und Dynamik unter Verwendung von zwei unterschiedlichen Proteinsystemen auf die Proteinfunktion zu untersuchen.

2-Desoxyribose-5-phosphat Aldolase (DERA) katalysiert die reversible Umwandlung von 2-Desoxyribose-5-phosphat in Acetaldehyd und Glyceraldehyd-3-phosphat (G3P) und ist im Nukleotidkatabolismus beteiligt. Die einzigartige Fähigkeit der DERA, die C-C-Bindungsreaktion zwischen zwei Aldehyden zu katalysieren, und eine breite Akzeptor-Substrat Toleranz, machen es zu einem attraktiven Biokatalysator. Der Reaktionsmechanismus der *E. coli* DERA (*ecDERA*), welche als symmetrischer Homodimer mit einer Größe von 259 Aminosäuren pro Monomer vorliegt, wurde basierend auf Kristallstrukturen des Proteins mit Reaktionsintermediaten und Mutagenese Studien postuliert [1]. Obwohl in zahlreichen Studien die Bedeutung des C-terminalen Tyrosins (Y259) in der katalytischen Reaktion hervorgehoben wurde, blieb die strukturelle Basis der Beteiligung dieses Restes aufgrund fehlender Elektronendichte der letzten 8 bis 9 C-terminalen Reste (C-terminal Tail) verborgen. Unter Verwendung von NMR-Spektroskopie an einer Monomervariante der *ecDERA* und ihrer Mutanten wurde das Sampling einer katalytisch relevanten Konformation des C-terminalen Tails, wobei Y259 ins aktive Zentrum hineinreicht (geschlossener Zustand), in Abwesenheit des Substrats beobachtet. Ein signifikant populierter geschlossener Zustand führte zu Kern-Overhauser-Effekt (NOE, engl.: Nuclear Overhauser Effect)-Kreuzpeaks in NOE-Spektroskopie (NOESY) Experimenten, auf deren Grundlage die erste Struktur des C-terminalen Tails im geschlossenen-Zustands Ensemble berechnet wurde. Umfangreiche

MD-Simulationen, die die NMR Ergebnisse ergänzen, wurden verwendet, um die transienten Wechselwirkungen näher zu charakterisieren, welche den C-terminalen Tail im geschlossenen Zustand stabilisieren. Darüber hinaus zeigten Wasserstoff/Deuterium-Austausch Experimente, dass die Y259 Seitenkette für eine effiziente Substrat-Proton-Abstraktion notwendig ist und somit den Imin-Enamin-Übergang im *ecDERA*-Reaktionsmechanismus beschleunigen könnte. Des Weiteren wurde neben der bekannten Phosphatbindungsstelle, die in *ecDERA*-Kristallstrukturen identifiziert wurde, eine zuvor unbekannte Hilfsbindungsstelle, welche sich am C-terminalen Tail befindet, anhand von Änderungen der chemischen Verschiebung mittels einer Phosphat Titration identifiziert. Diese Phosphatbindestelle kann die Phosphatgruppe des Substrats koordinieren, wodurch möglicherweise die optimale Position von Y259 während der katalytischen Reaktion stabilisiert wird.

Humanes Nedd4-1 (hNedd4-1) (engl.: Neuronal precursor cell expressed developmentally down-regulated gene 4-1) ist eine E3-Ubiquitin-Ligase, die den Ubiquitintransfer des entsprechenden E2-Enzyms auf das Substrat vermittelt. Das erste bekannte Target von Nedd4-1 war der humane epitheliale Natriumkanal (hENaC, engl.: epithelial Na⁺ channel). Der charakteristische Aufbau der Nedd4-Proteine umfasst drei bis vier WW-Domänen, die zwischen einer N-terminalen C2-Domäne und einer C-terminalen HECT-Domäne liegen. Die Substratspezifität wird durch die Protein-Protein-Interaktionsdomänen, den WW-Domänen, definiert. WW-Domänen von Nedd4-1 erkennen das PPxY-Motiv des hENaC (PY-Motiv) und weisen eine hohe Sequenzähnlichkeit untereinander auf. Die Affinität zum PY-Motiv variiert jedoch signifikant, wobei die dritte WW-Domäne (WW3*) die höchste Affinität zeigt. Mit Hilfe von NMR-Spektroskopie konnte gezeigt werden, dass die WW3*-Domäne in einem Gleichgewicht zwischen einem gefalteten bindungskompetenten, und einem ungefalteten Zustand in Lösung existiert. Das PY-Motiv bindet an den bindungskompetenten Zustand und verschiebt dadurch das Gleichgewicht zum gefalteten Zustand. Die Wechselwirkung der WW3*-Domäne mit dem PY-Motiv tritt daher durch gekoppelte Faltungs- und Bindungsgleichgewichte auf.

Die größte Sequenzdiversität unter den vier hNedd4-1 WW-Domänen liegt innerhalb des Loops, welcher die ersten beiden β -Faltblätter verbindet (Loop I). Loop I der WW3*-Domäne nimmt die Struktur eines Type I β -Turns an. Type I β -Turns bilden sich innerhalb von 100 ns und ermöglichen somit eine schnelle Faltung. Ein in Type I β -Turns statistisch bevorzugtes Prolin befindet sich an der $i+1$ Position von Loop I der WW3*-Domäne, allerdings nicht in den anderen drei hNedd4-1 WW-Domänen. Die Charakterisierung der Stabilität und der Dynamik von Loop I wurde mit einer Kombination aus NMR-Spektroskopie und MD-Simulationen durchgeführt. Der Austausch des Prolins an der $i+1$ Position in Loop I der WW3*-Domäne durch den entsprechenden

Rest in WW4 (Threonin) führt zu einem dynamischen, aus sieben Resten bestehenden Loop I anstelle eines Type I β -Turns. Die Bindung des hENaC-Peptids stabilisiert den Type I β -Turn in Loop I, sowohl für den Wildtyp als auch für die P433T Mutante der WW3*-Domäne. Darüber hinaus verändert die P zu T Mutation in Loop I nicht nur die Stabilität von Loop I, sondern auch die Gesamtstabilität der WW3* P433T Mutante, was aus einer Abnahme der Schmelztemperatur um 6 °C ersichtlich ist. Somit ermöglicht ein Prolin in $i+1$ Position in Loop I einen stabilen Type I β -Turn, was die Affinität zum α -hENaC Peptid beeinflussen könnte.

Acknowledgements

I would like to thank everyone who has contributed, in one way or the other, to this work:

Firstly, I would like to express my sincere gratitude to my advisor Dieter Willbold for the continuous support of my PhD study, motivation, and for helping me to realize my own ideas and for giving me access to excellent laboratory and NMR facilities.

I would also like to thank my mentor Jörg Pietruszka for insightful comments and his immense knowledge in the field of biocatalysis.

My sincere thanks goes to my enthusiastic supervisor Vineet Panwalkar for the continuous support, patience (!), motivation and for cheering me up in difficult times. He introduced me to biology lab work and taught me how efficient, headache-free protein expression and purification is done. Moreover, he gave me a practical introduction and guidance in the field of NMR spectroscopy, even when an experiment had to be set up at 11 p.m. on a Friday evening.

I thank my colleges at the ICS-6 and the IBOC for interesting discussions and a very nice work environment, which made working an enjoyable time. In particular, I would like to thank my present and former office mates Christina, Icaro, Pauline, Tim and Irina for a nice office atmosphere and many funny moments that made Jülich a nicer place. And I also thank Nici for assistance with the cloning.

Also many thanks to the IT-team Justin, Fritz and Volker and the NMR-team Rudi, Kevin and Matthias for providing constant support, without which a smooth work would be impossible. Also thanks for helping me tackling small or big everyday problems, such as getting new NMR experiments running, re-writing pulse-sequences or helping me to install LaTeX on my computer, so that I could write this thesis.

I would further like to thank Dusan for the great collaboration, and many nice discussions at the ICS-6, and now, via skype, to Uppsala.

I thank Andrew Dingley and Philipp Neudecker for instructive collaborations and helpful advice.

I would like to acknowledge the financial support from the Ministry of Innovation, Science and Research of the German federal state of North Rhine-Westphalia and the Heinrich-Heine-Universität Düsseldorf (scholarship within the CLIB Graduate Cluster Industrial Biotechnology).

Last but not least, I thank my parents, my sister Sophia and my brother Ulrich for their constant support during my education, which has been quite a long journey.

Contents

List of Figures	xv
List of Tables	xvii
List of Abbreviations	xix
1 Introduction	1
1.1 Solution NMR Spectroscopy to Study Proteins	1
1.1.1 A Short Introduction to Protein NMR	1
1.1.2 Principles of NMR	1
1.1.3 A First Orientation: Sequence-Specific Chemical Shift Assignments	3
1.1.4 Structure Determination using NMR	7
1.1.5 Power of NMR to Study Protein Dynamics	10
1.1.6 Combination of MD Simulation and NMR Spectroscopy	12
1.2 Protein-Ligand Interactions	14
1.2.1 Overview of Protein-Ligand Interactions	14
1.2.2 Techniques for Quantifying Protein-Ligand Interactions	16
1.3 Enzyme catalysis	19
1.3.1 Introduction to Enzyme Catalysis	19
1.3.2 Enzyme Kinetics and Thermodynamics	20
1.4 2-Deoxyribose-5-phosphate Aldolase from <i>E. coli</i>	23
1.5 Human Nedd4-1 WW3* High Affinity Domain	26
2 Aims	31
3 Scientific publications	33
3.1 ^1H , ^{13}C , and ^{15}N backbone and sidechain resonance assignments of a monomeric variant of <i>E. coli</i> deoxyribose-5-phosphate aldolase	35
3.2 Conformational Sampling of the Intrinsically Disordered C-terminal Tail of DERA is Important for Enzyme Catalysis	42
3.3 The Nedd4-1 WW Domain Recognizes the PY Motif Peptide through Coupled Folding and Binding Equilibria	72

3.4	Data describing the solution structure of the WW3* domain from human Nedd4-1	113
3.5	Proline Restricts Loop I Conformation of the High Affinity WW Domain from Human Nedd4-1 to a Ligand Binding-Competent Type I β -Turn .	123
4	Conclusions and Outlook	159
A		161
A.1	Material	161
A.1.1	Bacterial Strains	161
A.1.2	Chemicals and Biochemicals	161
A.1.3	Kits	162
A.2	Methods	163
A.2.1	Cloning of pET15b-kombi-P_deoC_K58EY96W	163
A.2.2	Transformation of Vector DNA into <i>E. coli</i>	163
A.2.3	Protein Expression and Purification	163
A.2.4	NMR Spectroscopy	165
A.2.5	Databases	166
A.2.6	Software	166
	Presentations and Awards	169
	Bibliography	171

List of Figures

1.1	Magnetization transfer in HNCA and CBCA(CO)NH	5
1.2	Sequential backbone walk	6
1.3	Restraints used for NMR structure calculation	8
1.4	Timescale of motions	10
1.5	Conformational selection versus induced fit model	15
1.6	NMR lineshape analysis used for studying protein-ligand interactions .	18
1.7	Progress of a catalyzed and uncatalyzed reaction	21
1.8	Michaelis-Menten model	22
1.9	DERA reaction <i>in vivo</i>	23
1.10	<i>ec</i> DERA crystal structure with the Schiff base intermediate	24
1.11	Close-up of the <i>ec</i> DERA active site	25
1.12	Modular organization of hNedd4-1	26
1.13	hNedd4-1 WW3* domain with the α -hENaC peptide	27
1.14	Sequence alignment of the four hNedd4-1 WW domains	28
3.1	TOC graphic Paper II	42
3.2	TOC graphic Paper V	123
A.1	DERAm size-exclusion chromatogram	164
A.2	DERAm purification confirmed by SDS-PAGE	165

List of Tables

1.1	Standard Triple-Resonance Experiments used for Backbone Assignments	4
A.1	Overview of used bacterial strains	161
A.2	Used chemicals and biochemicals	162
A.3	Commercially obtained Kits	162
A.4	NMR spectrometers	166

List of Abbreviations

α-hENaC	PPxY motif of the human α -subunit of the epithelial sodium channel
ADP	adenosine diphosphate
AMP	adenosine monophosphate
ATP	adenosine triphosphate
BMRB	Biological Magnetic Resonance Data Bank
CEST	chemical exchange saturation transfer
COSY	correlation spectroscopy
CPMG	Carr-Purcell-Meiboom-Gill
cryo-EM	cryo-electron microscopy
CSP	chemical shift perturbation
DERA	2-deoxyribose-5-phosphate aldolase
ecDERA	<i>E. coli</i> 2-deoxyribose-5-phosphate aldolase
DERAm	monomer variant of <i>E. coli</i> 2-deoxyribose-5-phosphate aldolase, carrying K58E and Y96W mutations
DHAP	dihydroxyacetone phosphate
DHF	dihydrofolate
DHFR	dihydrofolate reductase
DNA	deoxyribonucleic acid
dR	deoxyribose
dR5P	2-deoxyribose-5-phosphate
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
EC	Enzyme Commission
EDTA	ethylenediaminetetraacetic acid
FBPA	fructose-1,6-bisphosphate aldolase
FID	free induction decay

FRET	Förster resonance energy transfer
G3P	glyceraldehyde-3-phosphate
H/D	hydrogen/deuterium
HECT	homologous to E6-AP carboxyl terminus
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hetNOE	heteronuclear nuclear Overhauser effect
HIV	human immunodeficiency virus
HSQC	heteronuclear single quantum coherence
IDP	intrinsically disorderd protein
IDR	intrinsically disorderd region
ITC	isothermal titration calorimetry
MD	molecular dynamics
NADPH	reduced nicotinamide adenine dinucleotide phosphate
Nedd4-1	neuronal precursor cell expressed developmentally down-regulated gene 4-1
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
PDB	Protein Data Bank
Pin1	peptidyl-prolyl cis/trans isomerase
PRE	paramagnetic relaxation enhancement
PY	PPxY motif
RDC	residual dipolar coupling
RNA	ribonucleic acid
SPR	surface plasmon resonance
THF	tetrahydrofolate
TIM	triosephosphate isomerase
TMS	tetramethylsilane
TOCSY	total correlation spectroscopy
UV	ultraviolet
XFEL	fixed-target X-ray free-electron laser

Chapter 1

Introduction

1.1 Solution NMR Spectroscopy to Study Proteins

1.1.1 A Short Introduction to Protein NMR

Nuclear magnetic resonance (NMR) spectroscopy is a powerful technique to obtain structural, dynamic and functional information of biomolecules at atomic resolution. Unlike X-ray crystallography or cryo-electron microscopy (cryo-EM), that also allow detailed structural analysis of biomacromolecules, proteins can be studied in solution with physiological buffer and temperature, resembling their natural environment. Typical NMR experiments, however, require non-natural protein purity at very high concentrations (typically 50 μM - 1 mM) resulting in experimental conditions which still slightly differ from their native environment, lacking e.g. macromolecular crowding or possible interactions with other cellular molecules. The emerging field of in-cell NMR allows to study proteins in living cells at atomic resolution [2] and is particularly interesting when studying a protein's native network of interactions or post-translational modifications. However, in-cell NMR is particularly challenging due to large and complex systems. Another advantageous feature of NMR spectroscopy is the inwardly labeling technique used for classical NMR experiments. Enriching proteins with ^{13}C or ^{15}N isotopes does not affect their structure or dynamics, unlike other methods (e.g. Förster resonance energy transfer (FRET)), that require the covalent attachment of sizable labels to the protein.

1.1.2 Principles of NMR

NMR is a powerful technique to study structure and dynamics of biomolecules at atomic resolution by exploiting magnetic properties of atomic nuclei. NMR active nuclei are characterized by a total nuclear spin (I) different from zero requiring an odd number of either neutrons, protons or both. These nuclei with nonzero spin angular momen-

tum also have nuclear magnetic moments. The main elements found in proteins are hydrogen, carbon, nitrogen and oxygen out of which only the most abundant isotopes of hydrogen (^1H , 99.98 %) and nitrogen (^{14}N , 99.6 %) are NMR active. However, ^{14}N , with a spin of $I = 1$, is unattractive for studying biological systems using NMR due to strong resonance broadening caused by fast quadrupolar relaxation. The natural abundance of NMR active isotopes of the other elements is marginal: ^{13}C (1.1 %) and ^{17}O (0.04 %). Though, advances in sample preparation, in particular isotopic labeling, have made these elements (i.e. N, C, and O) accessible to NMR spectroscopic studies.

A spin is an intrinsic property of a particle and is a form of the angular momentum, that has a total of $2I + 1$ angular momentum states which are degenerate. In presence of a magnetic field (B) the intrinsic magnetic moment interacts with the magnetic field, resulting in equally spaced energy levels, known as Zeeman levels, with energies given by:

$$E = -\mu * B \quad (1.1)$$

where μ represents the nuclear magnetic moment, which is proportional to the gyromagnetic ratio γ . For nuclei with $\gamma > 0$ the magnetic energy is minimal when the spin angular momentum is parallel to the magnetic field and is maximized when it is anti-parallel and only a single transition between the energy levels exist. When aligned to the external magnetic field (either parallel or anti-parallel), the nuclear spins precess around the magnetic field at the Larmor frequency, ω , which depends on the type of nucleus and the magnetic field strength, that is mainly dominated by the external magnetic field (typically 11.7 T - 23.6 T), but also depends on the local environment of the nucleus:

$$\omega = \gamma B \quad (1.2)$$

In thermal equilibrium, the relative populations of these energy states are described by the Boltzmann distribution. The population difference depends mainly on the type of nucleus (γ) and the magnetic field (B). As an example, the population difference is on the order of 2 in 10^5 for ^1H spins at 14.1 T and room temperature. Since the sensitivity of NMR spectroscopy depends on the population differences between the Zeeman states, NMR is a rather insensitive spectroscopic technique. To increase sensitivity, magnets with higher field are being constructed (today available field strengths < 24 T) and nuclei with a large gyromagnetic ratio are preferentially used, such as ^1H .

During an NMR experiment, the equilibrium, in which nuclear spins precess around the magnetic field, gets perturbed by a radio frequency (RF) pulse, that is resonant with the precession of a given nucleus – at the Larmor frequency. The pulse, usually at the order of several μs in duration, displaces the bulk magnetic moment (vector sum of μ of individual nuclei) by generating a transition between the energy levels.

When the RF field is turned off, the bulk magnetic moment precesses with ω in a plane perpendicular to the magnetic field, inducing an oscillating electric current known as the *free-induction decay* (FID). This time-dependent current, which is induced and also detected in the receiver coil of the NMR spectrometer, results from all precessing nuclei in the sample. Thus, the observed FID is the sum of all individual FIDs for each nucleus, which oscillates at a characteristic frequency. After the FID is acquired, it is Fourier transformed from a time domain into a frequency domain, yielding the three primary NMR observables (chemical shift δ , intensity I and the linewidth λ) for each NMR signal, where δ is defined as:

$$\delta(\text{ppm}) = \frac{\nu - \nu_{ref}}{\nu_{ref}} \times 10^6 \quad (1.3)$$

with the resonance frequency ν of a nucleus in the protein and ν_{ref} represents the resonance frequency of a reference substance (e.g. tetramethylsilane (TMS) or 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) [3]).

After a pulse, the spins return to equilibrium through two different relaxation processes, thereby describing the decay of the FID. The longitudinal relaxation time corresponds to the process of re-establishing the equilibrium distribution in the magnetic field of the spin states and is characterized by T_1 . The loss of coherence of the transverse nuclear spin magnetization is described by the time constant T_2 . For small molecules in solution, T_1 and T_2 are typically several seconds, whereas in proteins T_2 is usually on the order of milliseconds leading to a rapid decay of the FID and hence, broad linewidths in the NMR spectrum, as T_2 is inversely proportional to the linewidth. The transverse relaxation rate R_2 ($= 1/T_2$) and thus the homogeneous linewidth (in absence of chemical exchange processes or aggregation) is proportional to the rotational correlation time τ_c , that essentially depends on the molecular size and shape of a protein, with larger proteins having larger τ_c and thus broader linewidths. Consequently, solution-state NMR is predominantly applied for the structural analyses of smaller systems (< 80 kDa). The largest monomeric protein structure solved using solution-state NMR till date is the one of the enzyme malate synthase G with a size of 81.4 kDa [4] exploiting advances in NMR techniques along with specific isotope labeling methods in order to increase the sensitivity and reduce spectral crowding.

1.1.3 A First Orientation: Sequence-Specific Chemical Shift Assignments

NMR spectra contain a wealth of structural and dynamic information. Protein-protein interactions, for example, can be easily observed by titrating one compound to the other using simple 2D spectra and analyzing the resulting chemical shift changes. In

a fingerprint ^1H - ^{15}N HSQC spectrum, there should be in principle one resonance per amino acid (apart from prolines) representing the amide groups, two additional peaks for glutamine and asparagine side chain NH_2 , and one additional peak for each tryptophan indole group. A brief look at a ^1H - ^{15}N HSQC spectrum directly reveals whether the protein is well folded resulting in well dispersed peaks, or unfolded, indicated by a poor chemical shift dispersion with the majority of resonances located between 8 - 9 ppm in proton dimension. However, for further analyses, e.g. protein-ligand interaction studies, structure determination or dynamics analyses, it is necessary to identify the amino acid represented by the given HSQC cross-peak. To obtain sequence-specific chemical shift assignments, multiple-resonance experiments have been developed, exploiting interactions between spins of NMR active nuclei. There are two common spin-spin interactions: through bonds and through space. Through-bond interactions are known as J -couplings, that provide information about e.g. the connectivity of nuclei and dihedral angles according to the Karplus equation [5]. The nuclear Overhauser effect (NOE) allows estimation of distances between nuclei (usually up to 5 Å), that are not directly covalently attached, providing useful distance-restraints, particularly for structure determination of proteins. For small unlabeled molecules and peptides (< 15 amino acids) homonuclear 2D experiments such as *correlation spectroscopy* (COSY) connecting neighboring protons and *total correlation spectroscopy* (TOCSY) connecting all protons of an individual amino acid using J -couplings may be sufficient [6]. However, for larger proteins, higher number of protons create severe spectral overlap, which can hamper this approach. In heteronuclear multi-dimensional NMR experiments the spectral overlap is significantly reduced by use of additional nuclei, i.e. ^{13}C and ^{15}N , thereby extending information of individual 2D spectra into additional dimensions. For obtaining backbone chemical shift assignments, a set of standard triple-resonance experiments (s. Table 1.1), exclusively relying on J -couplings over one or two bonds ($^1J/{}^2J$), is recorded on a ^{13}C and ^{15}N isotropically enriched protein sample.

Table 1.1: Standard Triple-Resonance Experiments used for Backbone Assignments

Experiment	Nuclei observed
HNCO [7–9]	$\text{H}(i)$, $\text{N}(i)$, $\text{C}'(i - 1)$
HNCA [7, 8, 10]	$\text{H}(i)$, $\text{N}(i)$, $\text{C}_\alpha(i)$, $\text{C}_\alpha(i - 1)$
HN(CO)CA [8, 11]	$\text{H}(i)$, $\text{N}(i)$, $\text{C}_\alpha(i - 1)$
HN(CA)CO [12]	$\text{H}(i)$, $\text{N}(i)$, $\text{C}'(i)$
CBCA(CO)NH [13]	$\text{H}(i)$, $\text{N}(i)$, $\text{C}_\alpha(i - 1)$, $\text{C}_\beta(i - 1)$
CBCANH [8]	$\text{H}(i)$, $\text{N}(i)$, $\text{C}_\alpha(i)$, $\text{C}_\beta(i)$, $\text{C}_\alpha(i - 1)$, $\text{C}_\beta(i - 1)$
HNCACB [8]	$\text{H}(i)$, $\text{N}(i)$, $\text{C}_\alpha(i)$, $\text{C}_\beta(i)$, $\text{C}_\alpha(i - 1)$, $\text{C}_\beta(i - 1)$

As the 1J -couplings are quite large between ^{13}C and ^{15}N nuclei ($11\text{ Hz} < J < 55\text{ Hz}$) as well as between their attached protons ($90\text{ Hz} < J < 140\text{ Hz}$) and the 2J -coupling between ^{15}N and $^{13}\text{C}_\alpha(i-1)$ ($^2J = 7\text{ Hz}$), they allow fast coherence transfer which can keep up with the loss of magnetization during the pulses sequence. This is particularly important for large proteins ($> 25\text{ kDa}$) due to their short T_2 relaxation times. Using deuterium labeling techniques, the relaxation times can be significantly increased [14–16], thus making these experiments feasible for larger proteins with molecular masses well above 25 kDa.

Two common examples for triple-resonance experiments correlating sequential amino acids are the HNCA and CBCA(CO)NH experiment. The magnetization transfer in these experiments is illustrated in Figure 1.1.

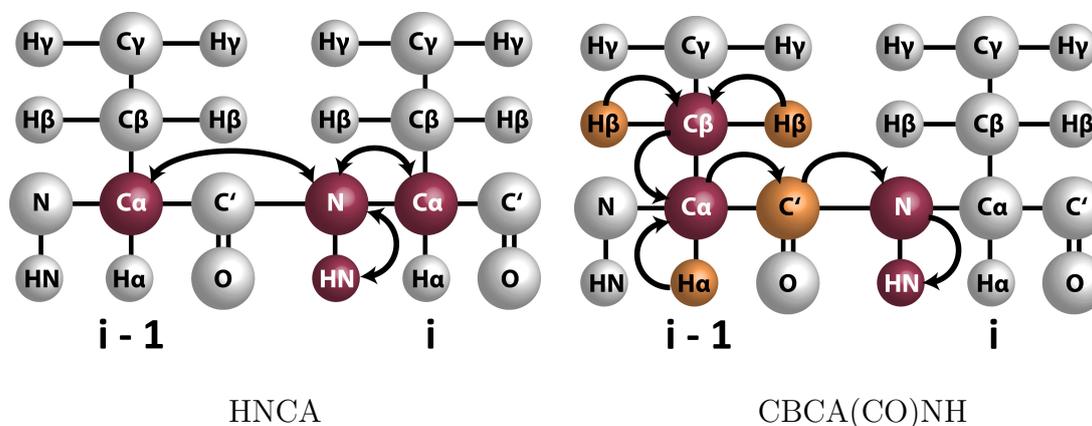


Figure 1.1: Magnetization transfer in HNCA and CBCA(CO)NH experiments, typically used for backbone resonance assignment. Nuclei colored in maroon are observed in the resulting spectra, whereas the ones colored in orange are only used for magnetization transfer. Figure adapted from www.protein-nmr.org.uk.

In the HNCA, the magnetization is transferred from $^1\text{H}_\text{N}$ to the amide nitrogen ^{15}N and then to the $^{13}\text{C}_\alpha$ of the same amino acid (i) and the preceding one ($i-1$). From the $^{13}\text{C}_\alpha$ the magnetization is transferred back to $^1\text{H}_\text{N}$ for detection (*out-and-back* experiment). Because the ^{15}N couples with the $^{13}\text{C}_\alpha$ of the same amino acid and the preceding one, both resonances will be visible in each $^1\text{H}_\text{N}$ - ^{15}N strip. However, since the $^{15}\text{N}(i) - ^{13}\text{C}_\alpha(i)$ 1J -coupling is larger (11 Hz) compared to the $^{15}\text{N}(i) - ^{13}\text{C}_\alpha(i-1)$ 2J -coupling (7 Hz), the peak representing the $^{13}\text{C}_\alpha(i)$ will be more intense in the spectrum. In the CBCA(CO)NH the magnetization is transferred from the $^1\text{H}_\alpha$ and the $^1\text{H}_\beta$ s to their directly attached carbon, $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$, respectively, and from the $^{13}\text{C}_\beta$ to the $^{13}\text{C}_\alpha$. Via the $^{13}\text{C}'$ and ^{15}N the magnetization is transferred to the amide proton $^1\text{H}_\text{N}$ for detection. In the resulting spectrum each $^1\text{H}_\text{N}$ - ^{15}N strip contains two peaks representing the $^{13}\text{C}_\alpha$ and the $^{13}\text{C}_\beta$ of the preceding amino acid.

A combination of triple-resonance experiments (Table 1.1) is typically recorded to allow a sequential backbone walk illustrated in Figure 1.2. As starting points, usually alanine residues are used due to their characteristic $^{13}\text{C}_\beta$ chemical shift (~ 20 ppm).

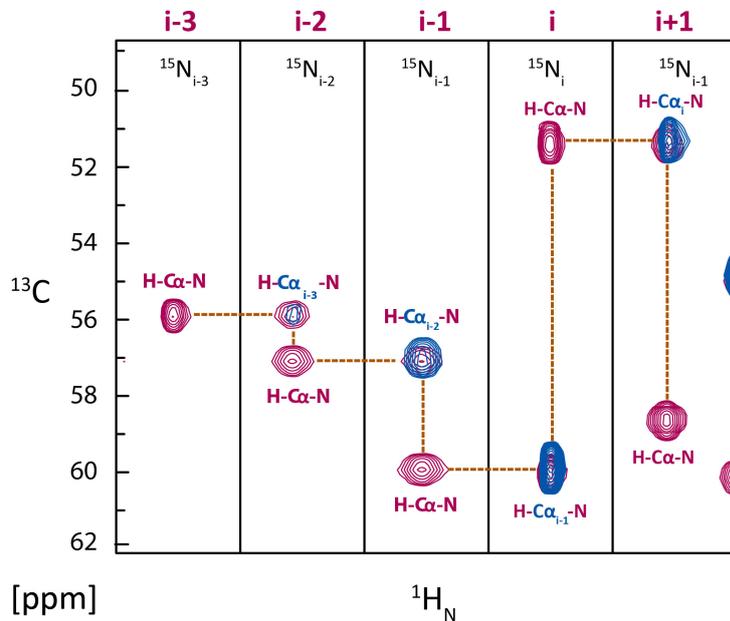


Figure 1.2: General example of a backbone walk. ^{15}N -strips from a HNCA showing the resonances of the $^{13}\text{C}_\alpha$ of residue (i) and the preceding residue ($i-1$) in purple. Resonances of the CBCA(CO)NH representing the $^{13}\text{C}_\alpha$ of only the preceding residue are colored blue. A combination of both spectra can be used to obtain the backbone chemical shifts by sequentially connecting the backbone atoms.

Once the backbone ($^1\text{H}_N$, ^{15}N , $^{13}\text{C}_\alpha$, $^{13}\text{C}'$) and $^{13}\text{C}_\beta$ chemical shifts are obtained, the side chain resonances can be assigned using, e.g. HBHANH, CC(CO)NH, H(CCO)NH and ^{13}C or ^{15}N edited TOCSY experiments. For the aromatic side chain assignment, typically (HB)CB(CGCD)HD, (HB)CB(CGCDCE)HE and ^1H - ^{13}C NOESY-HSQC of the aromatic region are employed.

Recording the above described multiple-resonance experiments can be NMR time-intensive. Since increments in every dimension have to be recorded, the experimental times for 3D experiments typically lie between 2 - 6 days, whereas these times significantly increase for 4D or 5D experiments. To tackle this issue, non-uniform sampling [17, 18] of the data can be employed, thereby reducing the total experimental time to $\sim 1 - 10$ %. Also, time-optimized pulse sequences such as BEST [19] and SOFAST [20] experiments allow acquisition of multi-dimensional NMR experiments in a shorter time.

1.1.4 Structure Determination using NMR

Structure and dynamics of proteins are key determinants for their function, thus making them a central aspect of structural biology. For determination of protein structures, NMR spectroscopy, X-ray diffraction and also cryo-EM can be used. As described in section 1.1.2, solution-state NMR spectroscopy studies are limited to small systems (< 80 kDa), whereas X-ray diffraction can be used for considerably larger systems, however, requiring non-physiological crystallization conditions. The emerging field of single-particle cryo-EM has been shown to be a valuable tool for solving structures of large systems, e.g. integral membrane proteins [21, 22], microtubules [23] and large assemblies such as transcription initiation complexes [24] at nearly atomic resolution, while only requiring very small amounts of sample compared to X-ray diffraction and especially NMR spectroscopy. Cryo-EM reports on structures of single particles which may, however, be challenging for heterogeneous samples, whereas NMR reports on time and molecule averaged quantities complementing cryo-EM data, particularly useful for structure validation purposes. Moreover, NMR allows studies of small proteins that may not be feasible for cryo-EM.

The process of obtaining a solution NMR structure is lengthy and expensive and therefore usually only applied if a protein doesn't crystallize or the solution structure is expected to have features not visible or detectable in the crystal structure. Also, NMR does not directly provide atomic coordinates as such, but it provides local geometric information, that can be used to calculate the three-dimensional protein structure. Structure determination via NMR involves firstly, the production of ^{13}C , ^{15}N (and ^2H for proteins > 20 kDa) labeled protein samples at high concentration and purity and, secondly, a complete backbone and side chain resonance assignment (s. section 1.1.3). For large proteins or proteins with intrinsically disordered regions (IDRs), a special labeling technique (e.g. Ile-Leu-Val [25]) may be necessary to reduce spectral overlap. After majority of resonances is assigned, conformational restraints such as backbone torsion angles (Φ , Ψ), hydrogen bonds and NOE based distance restraints need to be defined. Using these restraints, a three-dimensional structure can be calculated and subsequently evaluated and refined.

Backbone torsion angles (Φ , Ψ) and certain side chain rotamers (e.g. χ_1) can be easily predicted based on $^1\text{H}_N$, ^{15}N , $^{13}\text{C}_\alpha$, $^1\text{H}_\alpha$ and $^{13}\text{C}'$ chemical shifts, which are dominated by the local protein secondary structure, using programs such as TALOS-N provided by Ad Bax's group [26] or DANGLE [27]. It is further possible to determine torsion angles experimentally by measuring the 3J -couplings that correlate to the corresponding dihedral angle according to the Karplus equation [5]. However, coupling constants for some 3J -couplings are small (< 10 Hz), thus hampering efficient magneti-

zation transfer during the pulse sequence and therefore not suitable for larger proteins (> 20 kDa) with short T_2 times.

Hydrogen bonds in proteins can be identified using either direct or indirect methods. Direct experimental approaches allow identification of the hydrogen bond donor and acceptor by exploiting the partial covalent nature of hydrogen bonds, thus giving rise to a weak J -coupling [28]. The rate of hydrogen/deuterium (H/D) exchange of the amide $^1\text{H}_N$ is slower if a particular proton is involved in hydrogen bonding, thereby indirectly identifying presence of hydrogen bonds. Moreover, the magnitude of the temperature coefficient describing the dependence of the chemical shift on temperature is smaller for hydrogen-bonded protons than non-bonded ones. Therefore, the temperature coefficient can also be used for indirect detection of hydrogen bonds. The indirect methods for detecting hydrogen bonds are usually simple experiments, however, they only provide information about the donor $^1\text{H}_N$ - ^{15}N group, not the acceptor. For obtaining a more reliable indicator of hydrogen bonds the two indirect methods can be combined [29].

The most routinely used distance restraints between non-bonded atoms are the NOE distances obtained by a NOESY experiment. Unlike other experiments that depend on spin-spin couplings, the NOE occurs through space and not through bonds (i.e. covalent, hydrogen bonds). It is a cross-relaxation effect and the intensity of the cross peaks connecting protons in close proximity to each other depends mainly on the distance between them with a distance dependence of r^{-6} . Therefore, NOE derived distance restraints are usually within 5 Å and can be estimated using an internal standard (e.g. atoms within a phenylalanine or tyrosine ring) for calibration.

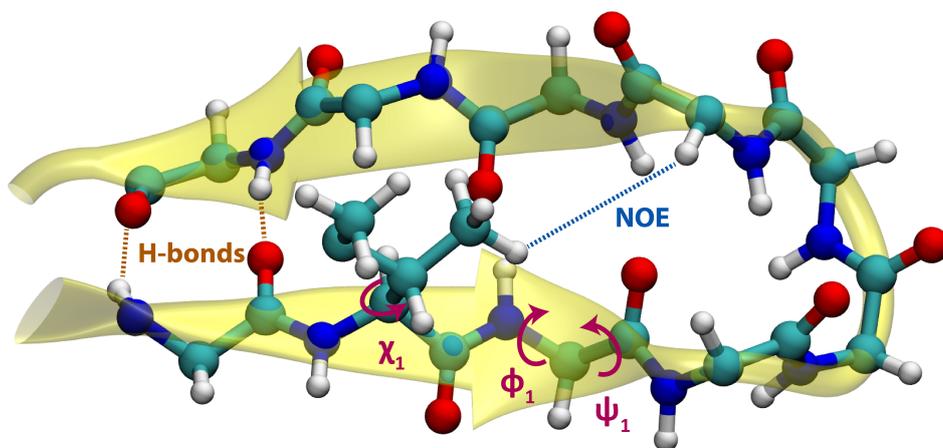


Figure 1.3: Illustration of commonly used structural restraints including hydrogen bond, NOE and dihedral angle restraints used for structure calculation.

NOE derived interproton distance restraints supplemented by torsion angle restraints and hydrogen bond information is usually sufficient for the study of globular proteins, given the large number of short proton-proton distances between amino acids which are far apart in the protein sequence. However, long-range distance information is at times indispensable, especially, when studying challenging systems like multi-domain complexes, proteins with IDRs or macromolecular complexes. There are two ways to obtain long-range distance information: residual dipolar couplings (RDCs) and the application of paramagnetic relaxation enhancement (PRE).

RDCs are measured in weakly anisotropic alignment media (e.g. phages, bicelles, polyacrylamide gels) and it allows the determination of a heteronuclear vector (e.g. ^1H - ^{15}N or ^1H - ^{13}C) with respect to the molecular axis. This is particularly useful for determining the orientation of multiple domains relative to each other in a multi-domain complex [30], where there is usually only limited short-range distance information available at the domain interface.

The PRE effect arises from the interaction of magnetic dipoles between unpaired electrons of a paramagnetic agent and a nucleus, causing an increase in relaxation rates and this appears in form of strong line-broadening in the NMR spectrum. Application of PRE can provide distance information up to ~ 35 Å since the magnetic moment of an unpaired electron is quite large. By covalently attaching a paramagnetic spin label (e.g. based on nitroxide free radicals [31] or metal ions chelated to EDTA [32]), usually via disulfide chemistry to the protein, resonances of residues will be line-broadened, depending on their distance to the spin label. In addition to this application, it is also possible to use solvent PREs, which do not require any chemical modifications of the protein, to obtain structural information. Here, different concentrations of a soluble and inert paramagnetic agent (e.g. gadolinium diethylenetriamine pentaacetic acid-bismethylamide [Gd(DTPA-BMA)] [33]) is added to the solvent, enhancing the relaxation rates of nearby nuclei in a concentration depended manner. Nuclei at the surface are affected the most, whereas nuclei located in the protein core will only show paramagnetic relaxation enhancement at higher concentrations of the solvent PRE agent. By measuring the concentration dependent relaxation rates, information about the solvent accessibility of nuclei can be obtained, aiding in structure refinement [34], in particular, in identifying domain interfaces [35] or protein-protein interaction sites in protein-ligand complexes [36].

Once a combination of the aforementioned restraints is obtained, the solution structure can be calculated using various softwares like ARIA [37], CYANA [38] or Xplor-NIH [39].

1.1.5 Power of NMR to Study Protein Dynamics

As described in section 1.1.4, it is possible to solve protein solution-structures using NMR spectroscopy methods, however, its real strength is its unique ability to identify and quantify protein motion which is crucial for protein function, at atomistic resolution. Protein motions are characterized by a number of conformations that constantly undergo exchange processes. Defining the intrinsic dynamics (conformational energy landscape) of proteins therefore provides valuable insight into their mode of action. Functional processes and thus, also the accompanying protein dynamics happen on various timescales ranging from picoseconds to several hours [40,41]. Figure 1.4 illustrates the timescale of important dynamical processes and NMR experiments, that can be used to characterize them.

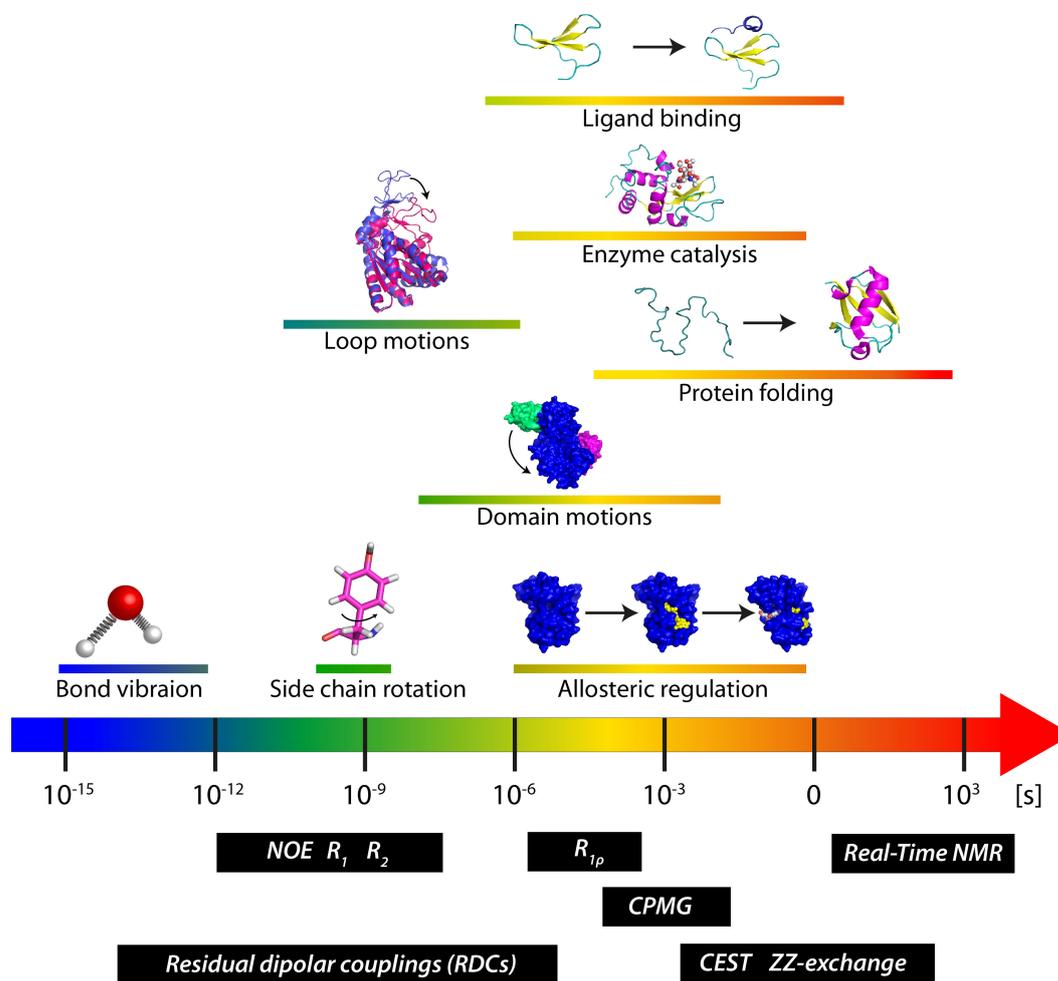


Figure 1.4: Timescale of motions detectable using NMR spectroscopy. Protein dynamics range from pico seconds to several hours.

Processes requiring large conformational changes such as protein folding happen on a relatively slow timescale ranging from ms to hours. Conformational changes involved in, e.g. domain motions, ligand binding, enzyme catalysis or to transfer information in

allosteric regulation processes happens on a timescale between μs and ms . Loop motions, side chain rotations and local atomic fluctuations are on the ps to μs timescale.

NMR spectroscopy is uniquely suited to study protein dynamics because it provides techniques allowing the investigation of motions over a wide range of timescales at atomic resolution. Internal dynamics on the ps to ns timescale and rotational tumbling of the protein affects the relaxation processes of the NMR active nuclei. The longitudinal and transversal relaxation rates (R_1 , R_2) as well as the heteronuclear NOE (hetNOE) are sensitive on this timescale and thus report on atomic fluctuations and side chain rotations. Experiments probing nuclear relaxation processes usually focus on the amide bond vector and report on backbone mobility.

Conformational exchange processes on the μs to ms timescale, which have been directly linked to protein function [42] (e.g. enzyme catalysis or ligand recognition and allostery), have been in the focus of recent dynamics studies and can be characterized by analyzing the lineshape of NMR resonances. Conformational exchange (thereby changing the chemical environment of a given nucleus) can have an effect on the lineshape depending on the population of different conformations (sub-states), their exchange rate, k_{ex} , and the difference in chemical shift, $\Delta\omega$. In case of slow chemical exchange ($k_{ex} \ll \Delta\omega$) separate peaks for each conformation are visible. A faster exchange rate in the regime of $k_{ex} \approx \Delta\omega$ results in line-broadening and a distinction between different states becomes challenging. In the fast exchanging regime ($k_{ex} \gg \Delta\omega$) a decrease in line-broadening is observed and instead of multiple peaks only one average peak is visible representing the different states. For quantification of the exchange parameters (population of the sub-states, k_{ex} , $\Delta\omega$), typically the Carr-Purcell-Meiboom-Gill (CPMG) experiment is used. In this experiment, the line-broadening caused by μs - ms exchange processes is reduced by applying a variable number of refocusing pulses (i.e. CPMG pulse train) to the magnetization when it evolves under the influence of a stochastically varying chemical shift, thereby providing information about the nature of the exchange process [43]. The first CPMG experiments were employed in the 1950s and 1960s to quantify conformational exchange processes in small molecules [44–46] and was further improved by A. G. Palmer III and coworkers [47] allowing the characterization of μs - ms chemical exchange of the protein backbone. Since then, CPMG relaxation dispersion experiments have been constantly improved and its application expanded to all backbone nuclei [48] and several side chains [49–53]. In addition to characterizing the exchange between different conformations in a protein with a minor state population as low as $\sim 0.5\%$, which cannot be directly observed in NMR spectra, it is also possible to calculate the structure of the *invisible* state using the chemical shifts differences ($\Delta\omega$) [54]. Excited states are usually low populated, according to

the Boltzmann distribution, but they can be extremely important for the function or dysfunction of the protein. For instance, using CPMG relaxation dispersion experiments, a low populated folding intermediate of the SH3 domain could be characterized that has been shown to be implicated in amyloid fibril formation [54]. Further, the relevance of minor states in enzyme catalysis has been well established [55–60]. Wright and coworkers characterized catalytically relevant conformations of *Escherichia coli* (*E. coli*) dihydrofolate reductase (DHFR) and found that the catalytic turnover rate is governed by the exchange between ground and excited states [55].

CPMG relaxation dispersion experiments are usually applicable to characterize exchange rates of $\sim 200 - 2000 \text{ s}^{-1}$ and a minor population $> 0.5 \%$ [61]. For slower interconversion rates of $\sim 20 - 200 \text{ s}^{-1}$ between a major and an invisible minor state (population $> 0.5 \%$) chemical exchange saturation transfer (CEST) experiments can be used to characterize the exchange process and detect the invisible, excited protein conformation [61]. In this method, a weak B_1 field is applied, irradiating various regions of a spectrum and monitoring the effect on the major peak. If B_1 is applied at the position of the invisible minor peak the resulting perturbation is transferred to the exchanging major peak and visible as a decrease in the intensity [61]. Using ZZ-exchange experiments, even slower exchange processes between a major and minor state in the range of $0.1 - 10 \text{ s}^{-1}$ ($k_{ex} \sim R_1$) can be characterized if the minor-state population is large enough to generate resonance signals in conventional spectra [62]. In this experiment, the exchange of longitudinal magnetization between the two exchanging states is monitored [62].

1.1.6 Combination of MD Simulation and NMR Spectroscopy

A multidisciplinary approach to study the structure and dynamics of proteins comprising experimental techniques such as NMR spectroscopy and X-ray diffraction along with molecular dynamics (MD) simulation has become common practice. Getting a complete picture of a biological system using experimental techniques is often hampered by e.g. spectral overlap or severe linebroadening using NMR or missing electron density or dynamic information using X-ray crystallography. To fill in the gaps and facilitate the interpretation of experimental data, MD simulations can be a powerful tool providing insight into time-resolved protein motions at atomic level. MD simulations strongly depend on the so called force field that defines a parameter set used to calculate the potential energy of a system of particles in the simulation. Different force fields are suitable for different systems and it can be a challenge choosing the appropriate parameter set for a given system [63], in particular for IDPs [64,65]. Thus, it is crucial to validate results obtained by MD simulations using experimental data in order to draw meaningful and reliable conclusions. For validation, average properties derived

by experiments can be compared to simulations, provided that the conformational distribution obtained by MD simulation reflects the dominant conformations observed in the experiment [66]. Preferably, properties that are directly measurable are used for validation [66], such as J -couplings, relaxation times (T_1 , T_2) and, most commonly, chemical shifts. Chemical shifts are sensitive to backbone and side chain torsion angle fluctuations, hydrogen bonds and aromatic ring positions, and thus provide a complex description of the chemical environment of each nucleus. Moreover, chemical shift assignments are relatively straight forward to obtain (s. section 1.1.3) and are necessary for any structural or dynamic NMR study, and thus, usually available.

MD simulations provide a trajectory of three-dimensional protein structures over time (typically ns - μ s). There are different methodologies available to predict chemical shifts from three-dimensional structures ranging from quantum mechanical to empirical ones [67]. Since the protein systems that are typically studied are large (\gg 100 amino acids) and trajectories easily consist of more than 100,000 structures, the computationally expensive quantum mechanical approaches are not feasible, and hence, empirical methods are routinely employed. The most accurate and time efficient prediction methods for backbone and $^{13}\text{C}_\beta$ chemical shifts employ a combination of machine learning techniques and empirical methods, namely SPARTA+ [68] and SHIFTX2 [69].

1.2 Protein-Ligand Interactions

1.2.1 Overview of Protein-Ligand Interactions

Macromolecular interactions, typically protein-ligand interactions, form the basis of response to stimuli of all living organisms. Most common protein-ligand interactions include enzyme-substrate, protein-lipids, protein-nucleic acids and protein-protein interactions. Understanding the nature of these interactions is key to understanding how biological systems function and particularly useful for e.g. drug design.

In a simple protein (P) ligand (L) interaction model, the interaction can be described as:



with the rate constants of association and dissociation k_{on} and k_{off} , respectively. The equilibrium dissociation constant K_D is defined as:

$$K_D = \frac{[P][L]}{[PL]} = \frac{k_{off}}{k_{on}} \quad (1.5)$$

where $[P]$ and $[L]$ are the concentrations of the free protein and ligand, respectively, and $[PL]$ is the concentration of the protein-ligand complex. For protein-ligand interactions the K_D is typically in the nm - mM range.

Protein function depends mainly on their three-dimensional structure. Thus, obtaining the protein structure is a necessity for understanding how proteins function on a molecular level. The first protein structure solved by X-ray diffraction was the one of the enzyme hen egg-white lysozyme in 1965 by C. C. F. Blake and coworkers [70] and has paved the way for structural biology. The first solution-state NMR structure of proteinase inhibitor IIA followed in 1985 [71] laying the ground work for protein structure determination by NMR spectroscopy. Since then, well more than 100,000 protein and protein-complex structures have been solved by these two methods (<https://www.rcsb.org/pdb/statistics/holdings.do>) allowing extensive studies. Investigation of protein structures in the ligand-free and ligand-bound states, revealing conformational changes, rapidly discarded Fischer's *lock-and-key* model as mechanistic explanation for molecular recognition, which postulates that the conformation of a protein in the free or ligand-bound state is essentially the same [72]. In 1958 Koshland proposed the *induced fit* hypothesis, stating, that different conformations of these two states result upon binding the ligand, which drives the protein towards a novel conformation [73]. In both cases, proteins are considered to exist in a single and stable conformation. Proteins, however, are inherently dynamic and sample a large ensemble of different conformations and therefore, conformations other than the lowest-energy conformation may play a role in molecular recognition [74].

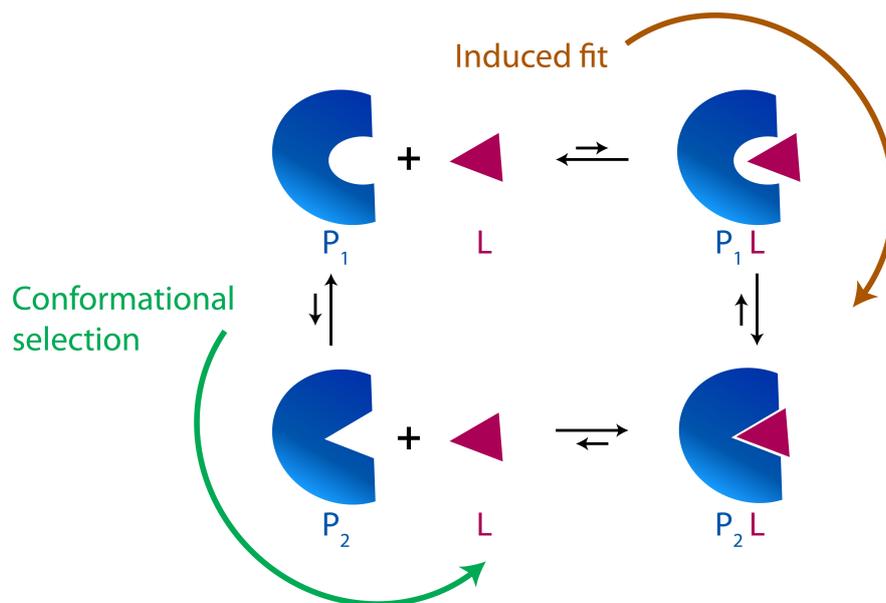


Figure 1.5: Conformational selection versus induced fit model. In the induced fit mechanism, binding of the ligand (L) drives conformational changes, whereas in conformational selection the binding-competent state (P_2) already exists in equilibrium with the ground state (P_1) which is then *selectively* bound by the ligand. Figure is adapted from [56].

This *conformational selection* model, derived from the energy landscape theory of protein structure and dynamics [74–77], considers the conformational diversity and proposes that low populated, high energy conformations are responsible for the recognition and binding to ligands with a subsequent shift of the population towards this conformer [56]. The main difference between the two extreme models, induced fit model and conformational selection, is thus, that in the induced fit model binding of the ligand induces conformational changes towards the bound conformation, whereas this is pre-sampled in the conformational selection mechanism in absence of the ligand. A simplified conformational selection and induced fit process are shown in Figure 1.5. According to this model, the conformational selection process will dominate if the concentration of the high energy, binding-competent state [P_2] is higher than the induced fit intermediate [P_1L] [78]. A study of the recognition dynamics of ubiquitin revealed that conformational selection suffices to explain structural adaptations of the ubiquitin backbone upon complex formation with different binding partners observed in crystal structures [79]. Though, subsequent small conformational changes, in particular side chain rearrangements, may be induced after the binding event [79].

Conformational changes related to enzyme function are typically on the μs - ms timescale and are therefore well suited for NMR studies (s. Figure 1.4) combined with MD simulations. For instance, extensive studies combining these two methods were

carried out on TIM, identifying and quantifying active-site loop motions that were found to have an implication on its catalytic function [80–82]. Furthermore, multi-temperature and fixed-target X-ray free-electron laser (XFEL) crystallography have been used to map the conformational landscape of cyclophilin A (Cyp A) [83] and hence, also presents a valuable tool for the investigation of conformational changes related to protein-ligand interactions.

1.2.2 Techniques for Quantifying Protein-Ligand Interactions

Various techniques can be used to characterize protein-ligand interactions, with the most common ones being isothermal titration calorimetry (ITC), surface plasmon resonance (SPR), fluorescence spectroscopy and NMR spectroscopy.

An isothermal titration calorimeter consists of two identical cells, a reference cell containing the buffer, and a sample cell containing the reaction buffer and the protein. Ligand-aliquots are titrated into the sample cell causing a change in temperature depending on the nature of the reaction, and the power required to maintain equal temperature in both, the reference and the sample cell, is measured. ITC allows the determination of the binding affinity ($K_A = 1/K_D$), stoichiometry (n) and enthalpy (ΔH). From these measurements, the changes in Gibbs free energy (ΔG) and entropy (ΔS) can be determined, thus providing a complete thermodynamic characterization. This approach is feasible for K_D s in the nM to μ M range and requires no chemical modification of the protein or the ligand, has no limitations in molecular weight and allows a variety of experimental conditions to study binding thermodynamics in solution. However, it usually requires high amounts of sample and does not report on binding kinetics.

SPR is an optical technique for measuring the refractive index of thin layers of material absorbed on a metal surface (typically gold). The binding event changes the refraction index of the material, which translates into the SPR signal. The experimental setup for determination of protein-ligand interaction using SPR involves the immobilization of the target protein (or ligand) onto a surface (typically dextran) located at the straight side of half a circular prism coated with a thin metal film between the glass and dextran layer. A solution containing the ligand is injected over the layer of the immobilized interaction partner, binding the protein, and thereby increasing the SPR signal during the association time. After the steady-state is reached, a solution without the ligand (buffer only) is injected resulting in the dissociation of the ligand and hence, in a decrease of the SPR signal. This approach provides kinetic information such as the association and dissociation rate constants k_{on} and k_{off} , respectively, and thus also on the K_D . Small quantities of protein (\sim ng quantities) required for this

technique and the possibility to establish high-throughput measurements make this method very attractive for obtaining binding kinetics of protein-ligand interactions, whereas thermodynamic information are not obtained. However, immobilization of one component typically requires the chemical modification of this component to couple it to the dextran matrix, which may also sterically hinder ligand or protein binding. Moreover, it prevents free diffusion of one component and SPR is only applicable to limited experimental conditions that do not interfere with the optical measuring technique.

Multiple fluorescence methods can be used to characterize protein ligand interactions [84], many of them, however, require the covalent attachment of a chromophore to the protein. A very simple, label-free technique to obtain the K_D of a protein-ligand binding interaction is to follow changes in ligand fluorescence or the intrinsic protein fluorescence sensitive to ligand binding. The intrinsic protein fluorescence mainly arises from tryptophan and tyrosine side chains and can be observed in the ultraviolet (UV) spectrum. Relative changes in the fluorescence intensity or wavelength that reflect ligand binding can be quantified as a function of ligand concentration to extract the dissociation constant. Hence, a basic requirement for the application of this method is naturally the presence of tryptophan and/or tyrosine residues in the protein or the ligand that are affected by the binding event. Moreover, no kinetic information can be obtained using this technique, however, it is a fast and straight-forward approach to determine the binding affinity that requires only small amounts of protein and is feasible for high-throughput screening.

Characterization of protein-ligand binding via NMR can be accomplished by following changes in two-dimensional spectra (typically ^1H - ^{15}N HSQC spectra) of the protein upon titration of a ligand. These titration experiments are typically straight forward to conduct, but information-rich and capable of simultaneously characterizing binding thermodynamics (K_D), kinetics (k_{on} , k_{off}), along with structural aspects of the binding event via changes in the chemical shift (δ). Chemical shift changes of non-overlapping resonances can be used to extract residue-specific dissociation constants. Moreover, the appearance of NMR resonances (i.e. the exchange regime: slow, intermediate or fast) in a titration experiment depends on the exchange rate (k_{ex}) between the apo and complex forms as well as the chemical shift difference ($\Delta\omega$) between them, provided that the exchange is on the μs - ms timescale (s. section 1.1.5). The exchange regime is strongly affected by the ligand binding affinity, as a tighter binding yields a longer-lived protein-ligand complex and thus slower exchange between free (P) and bound (PL) states, whereas weak binding consequently leads to fast exchange. Schematic linewidth changes in the extreme cases of slow, intermediate, and fast exchange processes upon ligand binding are illustrated in Figure 1.6.

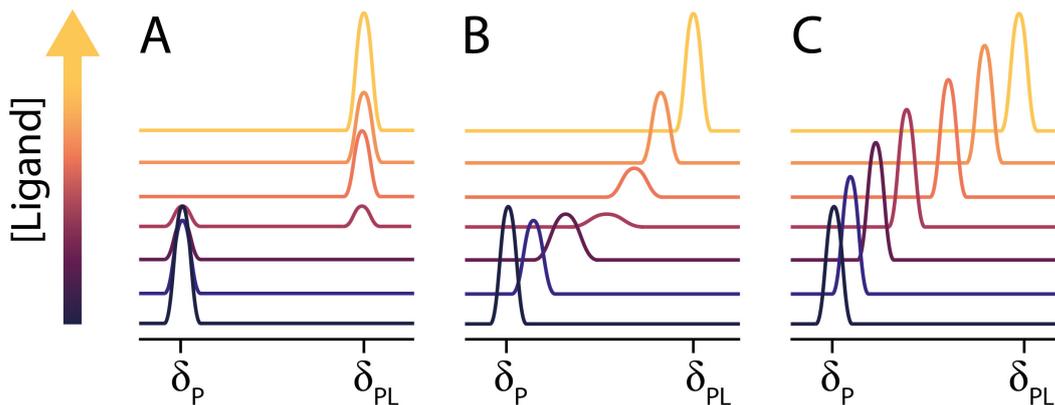


Figure 1.6: NMR line shape analysis used for studying protein-ligand interactions. The chemical exchange regime is affected by the ligand binding affinity, K_A , with a strong affinity leading to a more stable complex and thus in slower exchange. Linewidth changes upon ligand binding for slow (A), intermediate (B) and fast exchange (C) are schematically illustrated as 1D trace.

Since the exchange rate k_{ex} depends on the association and dissociation rate constants according to

$$k_{ex} = k_{on}[L] + k_{off} \quad (1.6)$$

it is possible to extract kinetic information based on changes in linewidth, e.g. with the program TITration ANalysis (TITAN) [85]. Considering that the on-rate for protein-ligand interactions is usually diffusion-limited, the K_D is typically determined by the off-rate.

Unlike the other methods mentioned above, NMR provides direct structural, residue-specific information about the protein-ligand interaction, provided that a residue specific resonance assignment is available. This method is label-free, yet, requires isotopic labeling of the protein. As for other NMR experiments, only small protein systems (< 80 kDa) can typically be studied and a high sample concentration ($> \mu\text{M}$) is required. NMR is particularly suited to characterize medium to weak protein-ligand binding in the μM to mM range. Since no thermodynamic information other than the K_D can be obtained with this method, it can be complemented with ITC to obtain a complete characterization of the protein-ligand binding interaction. When comparing these two methods, they may, however, result in non-identical K_D values. For instance, if the protein is in equilibrium between a binding-competent and low populated, non-binding-competent conformation, then the NMR and ITC derived K_D values may differ. ITC does not differentiate between multiple conformations of a reaction partner, whereas specific K_D s can be obtained from NMR chemical shifts of a given conformation. Thus, the ITC derived K_D can be amended to account for the minor, non-binding-competent population. (e.g. using CPMG relaxation dispersion experiments) [86].

1.3 Enzyme catalysis

1.3.1 Introduction to Enzyme Catalysis

Enzymes are biomacromolecules that are able to catalyze chemical reactions which would otherwise not occur on biologically relevant timescales. They are typically proteins, however, in very few cases nucleic acids can act like enzymes, so called ribozymes, which catalyze e.g. the hydrolysis of RNA [87]. In living organisms enzymes are involved in numerous metabolic pathways and thus crucial for survival. Understanding how enzymes function on a molecular level is not only essential to understand natural processes, but also with respect to medical applications. For example, L-Asparaginase, that catalyzes the hydrolysis of L-asparagine to L-aspartic acid and ammonia, is used in leukemia treatment exploiting that tumor cells require high amounts of L-asparagine that they usually cannot synthesize themselves [88]. Some enzymes function without any additives, whereas others require cofactors such as small organic molecules or metal ions.

Moreover, it has been shown that enzymes can be used within living cells or outside to catalyze reactions and thus have been applied in several processes as biocatalysts (e.g. in food production such as beer [89], wine [90] and cheese [91]). The commercial application of biocatalysis has been enormously extended over the last century and is now established in various industrial sectors such as food production, therapeutics, chemicals, detergents and many more [92, 93], due to the many advantages of enzyme catalysis namely high catalytic efficiency, stereospecificity or environmental factors.

Enzymes are classified based on the chemical reaction they catalyze into six main groups (Enzyme Commission (EC) numbers 1-6) according to the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) [94, 95]:

EC 1 **Oxidoreductases** catalyze oxidation and reduction reactions involving electron transfers. A well studied member of this group is the dihydrofolate reductase (DHFR) which reduces dihydrofolate (DHF) to tetrahydrofolate (THF) using reduced nicotinamide adenine dinucleotide phosphate (NADPH) as electron donor.

EC 2 **Transferases** catalyze the transfer of a functional group (e.g. amino or phosphate groups) from a donor to an acceptor. An example of phosphotransferase is the adenylate kinase catalyzing the reversible conversion of adenine nucleotides ATP and AMP into two ADP molecules.

EC 3 **Hydrolases** catalyze the hydrolysis of a substrate, thereby splitting it into multiple compounds by addition of water. A well-known member of the sub-class

aspartic proteases is the human immunodeficiency virus (HIV) protease cleaving viral proteins at specific sites crucial for HIV replication.

EC 4 **Lyases** catalyze reactions involving the elimination of a double bond by addition of a functional group or the reverse way. A prominent example is Fructose-1,6-bisphosphate aldolase (FBPA), that catalyzes the conversion of fructose 1,6-bisphosphate into dihydroxyacetone phosphate (DHAP) and glyceraldehyde 3-phosphate (G3P). Also, the 2-deoxyribose-5-phosphate aldolase (DERA), which is further described in section 1.4, belongs to this enzyme class.

EC 5 **Isomerases** catalyze the conversion of an isomer of a compound to another involving intra-molecular rearrangements. Triosephosphate isomerase (TIM), that catalyzes the reversible interconversion of dihydroxyacetone phosphate into its isomer G3P, is one of the most studied members of this group.

EC 6 **Ligases** catalyse the formation of covalent bonds to ligate multiple substrates such as the DNA ligase that catalyzes the formation of a phosphodiester bond thereby facilitating the linkage of DNA strands.

The mechanism through which enzyme-substrate interactions proceed could be either induced fit, conformational selection or a combination of both [56] as described in section 1.2.1. For several enzymes including adenylate kinase [60,96], dihydrofolate reductase [55,97] and RNase A [59], the conformational selection process has been shown to be implicated in their catalytic function. NMR studies on *E. coli* and *Aquifex aeolicus* adenylate kinase revealed that the enzyme fluctuates between an open and closed conformation and that the release of the product presumably depends on the transition between the open and closed state, due to a strong correlation of the catalytic turnover number (k_{cat}) and the exchange rate between the two states [60,96].

1.3.2 Enzyme Kinetics and Thermodynamics

According to the transition state theory developed in 1935 by Eyring, Evans and Polanyi [98], the conversion from reactants to products occurs via a transition state which has the highest free energy along the reaction coordinate. Catalysts (i.e. enzymes) reduce the activation energy (ΔG^\ddagger) thereby dramatically increasing the reaction rate as shown in Figure 1.7. For instance, urease, catalyzing the hydrolysis of urea into carbon dioxide and ammonia, reduces the activation enthalpy from 32.1 kcal/mol (non-catalyzed) to 9.9 kcal/mol thereby speeding up the reaction by a factor of 10^{16} or chorismate mutase that catalyzes the conversion of chorismate to prephenate, essential in the production of phenylalanine and tyrosine, decreases the enthalpy from

24.4 kcal/mol to 8.6 kcal/mol, which translates into an enhancement by a factor of 10^{12} [99]. Rate acceleration by enzymes can go up to a factor of more than 10^{27} , which is accomplished, for example, by the alkaline phosphatase [100, 101].

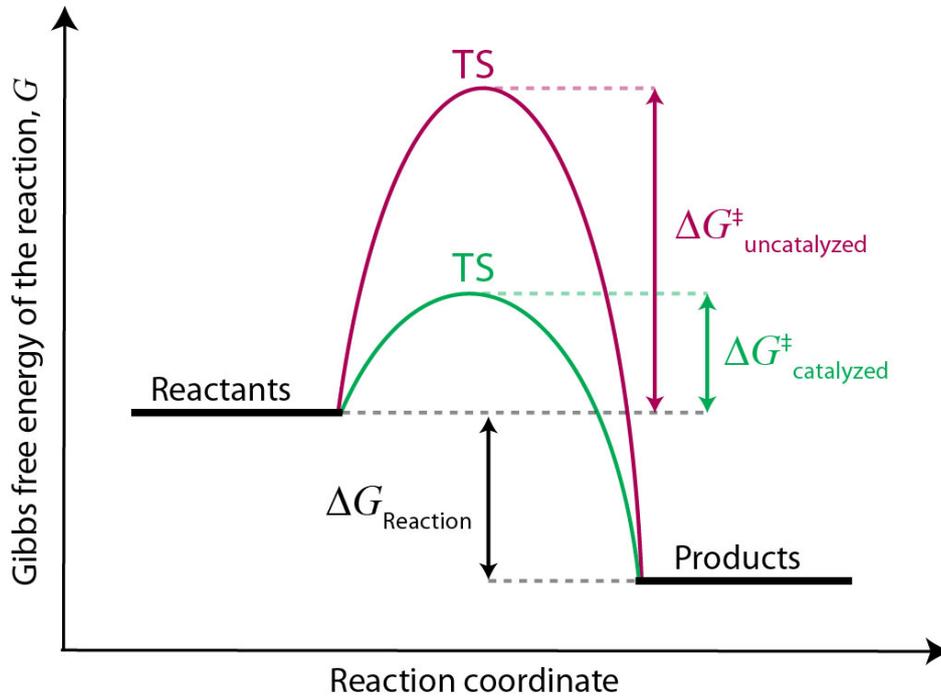
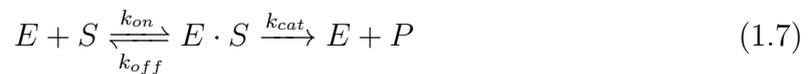


Figure 1.7: Reaction coordinate diagram in an uncatalyzed (purple) and a catalyzed (green) reaction, respectively. TS denotes the transition state. The activation (ΔG^{\ddagger}) as well as the Gibbs free energy of reaction ($\Delta G_{\text{Reaction}}$) are illustrated.

To describe and quantify enzyme-catalyzed reaction rates, the Michaelis-Menten kinetics model [102], relating the rate of reaction, v , with substrate concentration $[S]$, is commonly employed. In a reaction, the enzyme (E) forms a complex with the substrate (S) with the association rate, k_{on} , and the dissociation rate, k_{off} , of the enzyme-substrate complex ($E \cdot S$), respectively and subsequently releases the product (P) and the unaltered enzyme:



where k_{cat} is the catalytic rate constant describing the rate of product formation. The rate of the catalytic reaction, v , can be defined as product formation per time unit:

$$v = \frac{d[P]}{dt} \quad (1.8)$$

with $[P]$ denoting the product concentration. At an initial substrate concentration, $[S]_0$, the rate of reaction, v , is proportional to the total enzyme concentration, $[E]_0$. At

a given $[E]_0$ and low $[S]_0$, v is proportional to $[S]_0$, whereas at high $[S]_0$ the reaction rate is independent from $[S]_0$ and approaches the maximum velocity, V_{max} :

$$v = V_{max} \frac{[S]}{K_M + [S]} \quad (1.9)$$

with the Michaelis constant, K_M , i.e. the substrate concentration at $1/2 V_{max}$, illustrated in Figure 1.8.

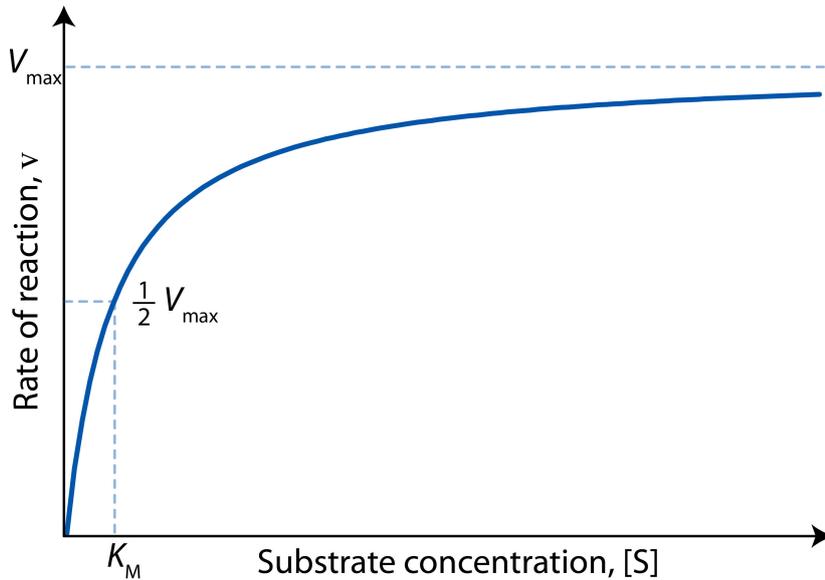


Figure 1.8: Michaelis Menten saturation curve describing the dependence of the reaction rate, v , on the substrate concentration, $[S]_0$, in an enzyme catalyzed reaction.

$$K_M = \frac{k_{off} + k_{cat}}{k_{on}} \quad (1.10)$$

and the maximum velocity V_{max} :

$$V_{max} = k_{cat}[E]_0 \quad (1.11)$$

k_{cat} , the so called turnover number, denotes the number of products formed in a catalytic cycle per unit of time, typically s^{-1} . The catalytic efficiency, ϵ , of an enzyme is defined as the ratio of k_{cat} and K_M :

$$\epsilon = \frac{k_{cat}}{K_M} = \frac{k_{on}k_{cat}}{k_{off} + k_{cat}} \quad (1.12)$$

ϵ is maximal when $k_{cat} \gg k_{off}$. Since k_{on} is the rate constant of the formation of the enzyme-substrate complex with both, enzyme and substrate freely diffusing in solution, the maximum efficiency is diffusion limited.

1.4 2-Deoxyribose-5-phosphate Aldolase from *E. coli*

2-Deoxyribose-5-phosphate aldolase (DERA, EC 4.1.2.4) is an acetaldehyde-dependent, homodimeric class I aldolase, that catalyzes the reversible conversion of acetaldehyde and G3P to generate 2-deoxyribose-5-phosphate (dR5P), illustrated in Figure 1.9, through a Schiff base intermediate [103]. Due to its ability to catalyze carbon-carbon cleavage or formation, DERA is classified as a lyase and can be assigned to the first subclass (E.C. 4.1). Aldolases are typically classified into two groups according to their mechanism: class I aldolases have a conserved active-site lysine residue that forms a Schiff base intermediate with the donor substrate, whereas class II aldolases utilize divalent metal ions (Zn^{2+}) in their catalytic mechanism.

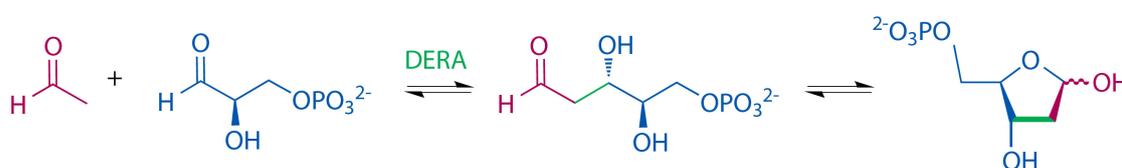


Figure 1.9: DERA reaction *in vivo*. DERA catalyzes the reversible aldol reaction between the donor acetaldehyde, and the acceptor G3P producing dR5P.

DERA plays an important role in key metabolic pathways by providing crucial metabolic intermediates namely acetaldehyde and G3P, and is involved in the nucleotide catabolism [104]. Acetaldehyde generated by DERA along with other enzymes (e.g. alcohol dehydrogenase), is important for synthesis of acetyl co-enzyme A which participates in the citric acid cycle. Additionally, G3P is involved in glycolysis and the pentose phosphate pathway. Thus, DERA is crucial for bacterial survival and considered as a potential pharmaceutical target against several human pathogens [105, 106]. The *E. coli* DERA (*ecDERA*) is particularly interesting to study in this aspect given that it is highly sequence identical (> 94 %) to DERAs of several human pathogens such as *Shigella flexneri*, *Salmonella typhimurium* and *Cronobacter sakazakii*, but shares only 38 % sequence identity with human DERA (*hDERA*).

ecDERA has the typical $(\alpha/\beta)_8$ TIM barrel fold, which is a common structural feature among class I aldolases, and consists of 259 residues per monomeric unit [1]. Heine et al. [1] proposed a reaction mechanism, based on *ecDERA* crystal structures with carbinolamine and Schiff base intermediates (s. Figure 1.10) along with site-directed mutagenesis studies, which states that the reversible aldol reaction of acetaldehyde and G3P generating dR5P proceeds through Schiff base formation with the active-site K167. Active-site residues K201, D102 and a conserved water molecule are proposed to perturb the pK_a of K167 and serve as a proton shuffling system in the multistage reaction mechanism.

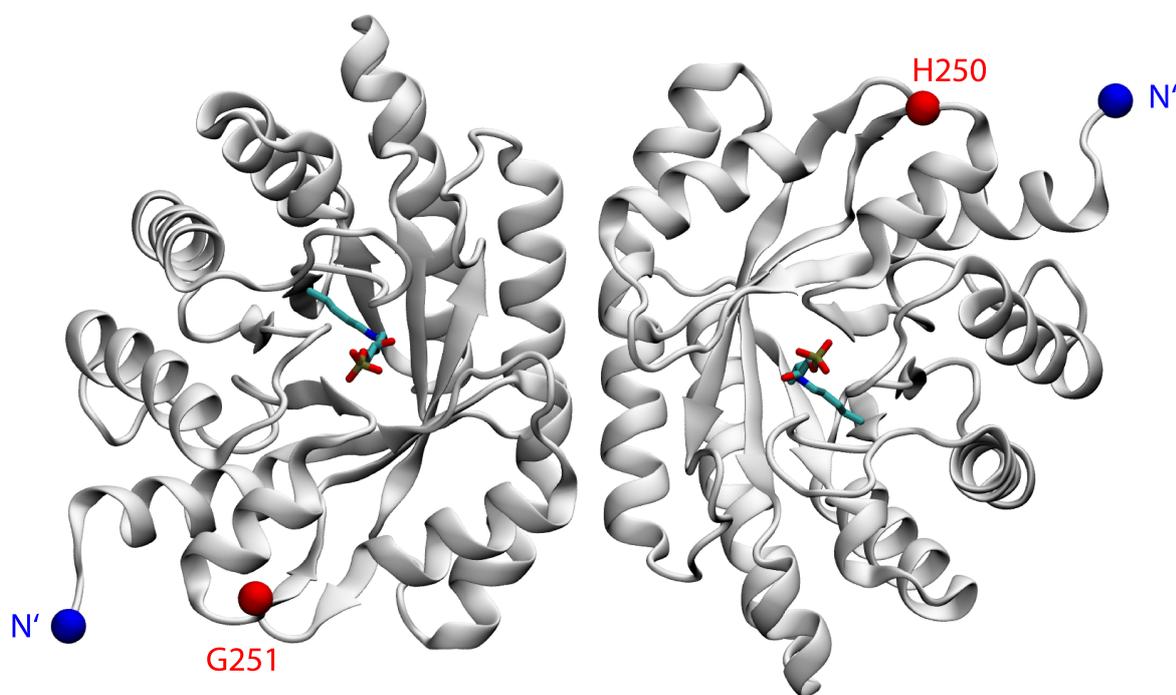


Figure 1.10: *ecDERA* crystal structure with the Schiff base intermediate (pdb: 1JCJ) [1]. The reaction intermediate with the active-site residue K167 is shown as sticks. N' and C' terminal residues are indicated by blue and red spheres, respectively. Due to a lack of electron density, the last 8 to 9 C-terminal residues are missing from the crystal structure, and the last visible residues G251 (Monomer A) and H250 (Monomer B) are indicated.

The active site of *ecDERA*, shown in Figure 1.11, can be sub-divided into two regions: the aldehyde binding region and the phosphate binding region. The aldehyde binding region comprises the Schiff base-forming K167 and residues K201 and D102 involved in the proton shuffling system. Backbone amide groups of G205 and S238 directly interact with the phosphate group and, via a water molecule, also K172 side chain and G204, V206, S239 and G171 backbone amide groups according to [1, 107].

Also the C-terminal Y259 of *ecDERA* has been shown to be important for catalysis since the Y259F mutation leads to a drastic drop (~ 200 fold) in catalytic efficiency [1]. However, investigation of its implication in the catalytic reaction mechanism has been hampered due to missing electron density of the last 8 to 9 C-terminal residues in all available *ecDERA* crystal structures [1, 107, 108], probably due to intrinsic flexibility of the C-terminal tail [107].

In addition to its biological function, DERA is unique in catalyzing the carbon-carbon bond formation between two aldehydes and has a broad acceptor substrate tolerance, thus it is being utilized as a “green” biocatalyst for the synthesis of, e.g. chiral

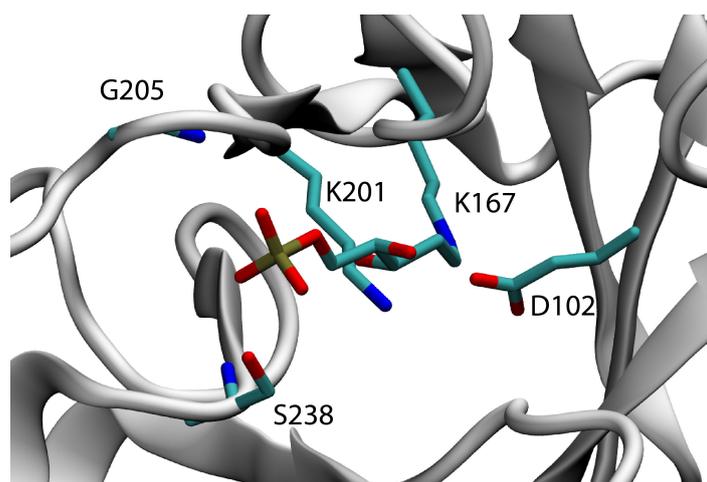


Figure 1.11: Close-up of the *ecDERA* active site with the Schiff base intermediate (pdb: 1JCL and 1JCJ). Active-site residues K167 (in complex with dR5P), D102 and K201 as well as phosphate binding residues S238 and G205 are shown as stick representation.

synthons for the synthesis of statins [109–111]. Kinetic studies, however, have shown that DERA is strictly dependent on phosphorylated acceptor substrates, demonstrated by a severe loss in activity when comparing dR5P and deoxyribose (dR) as acceptor substrates [112], hence limiting its application as efficient biocatalyst to phosphorylated aldehydes.

1.5 Human Nedd4-1 WW3* High Affinity Domain

Ubiquitination of cellular proteins is a post-translational modification process in eukaryotes that can mark proteins for proteasome degradation [113] and regulates numerous other cellular activities [114, 115]. The functional diversity arises due to the large variety of E3 ubiquitin ligases responsible for mediating the transfer of ubiquitin from an E2 ubiquitin-conjugating enzyme to the substrate. The modular composition of these ligases mediates substrate specificity, affinity and thus ligase activity. The human Nedd4 (neuronal precursor cell expressed developmentally down-regulated gene 4) family of HECT-type E3 ubiquitin ligases consists of nine members comprising NEDD4 (Nedd4-1), NEDD4-2, ITCH, WWP1, WWP2, NEDL1, NEDL1, SMURF1 and SMURF2 [116]. The founding member of this family, Nedd4-1, was originally identified as a developmentally down-regulated mouse gene in the central nervous system [117]. The first recognized target of Nedd4-1 was the human epithelial Na⁺ channel (hENaC) [118, 119]. Human Nedd4-1 (hNedd4-1) has further been shown to regulate the activity of signaling receptors [120] such as the insulin-like growth factor-1 [121]. Furthermore, hNedd4-1 is associated in cancer development, supported by an overexpression in prostate and bladder cancer [122], and is thus considered as a pharmaceutical target [123].

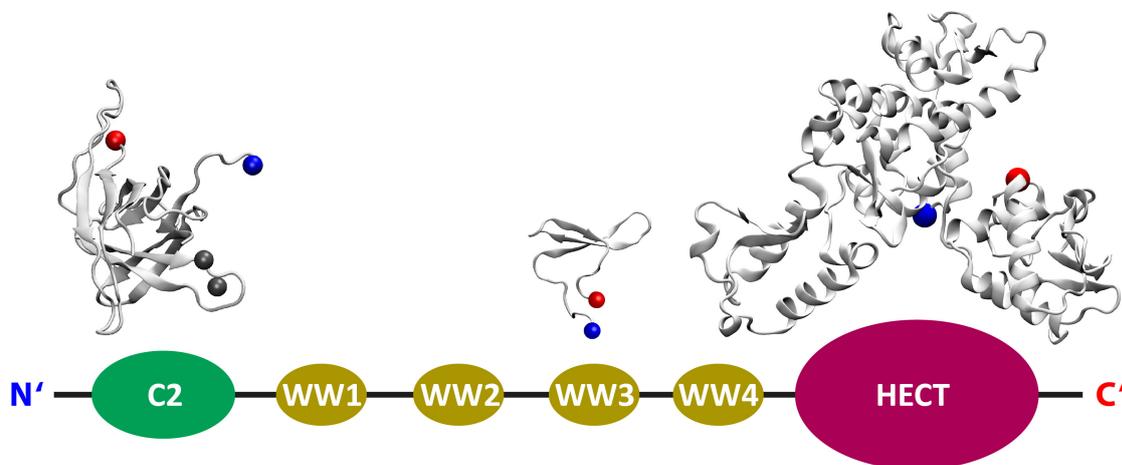


Figure 1.12: Modular organization of hNedd4-1. hNedd4-1 contains an N-terminal C2 domain (green), a central region comprising four WW domains (weat-colored) and a C-terminal HECT domain (magenta). For hNedd4-1, crystal structures of the C2 domain (pdb: 3B7Y), WW3 domain (pdb: 4N7F) and the HECT-domain (pdb: 2XBF) are available and illustrated above the schematic representation with blue and red spheres indicating the N-terminal and C-terminal region of each domain, respectively.

The characteristic domain architecture of Nedd4 family proteins is illustrated in Figure 1.12 for hNedd4-1. It includes presence of an N-terminal C2 domain (Ca²⁺-

dependent lipid binding domain), three to four WW domains (protein-protein interaction domains) and a catalytic, C-terminal HECT domain, which mediates the transfer of ubiquitin from the cognate E2 to the substrate [116]. Nedd4 proteins exist in an autoinhibited conformation through an interdomain interaction between the C2 and HECT domain, which is disrupted by binding of Ca^{2+} , thus activating E3 ligase activity [124]. Additionally, this autoinhibited state can also be unlocked by adapter proteins interacting with the WW domains of the protein [125]. Hence, WW domains are involved in both, regulating the activity as well as substrate recognition of Nedd4 proteins.

WW domains are compact interaction modules consisting of 38 to 40 amino acids and exhibit two highly conserved tryptophan residues and an invariant proline [126,127] illustrated for the third domain of hNedd4-1 in Figure 1.13.

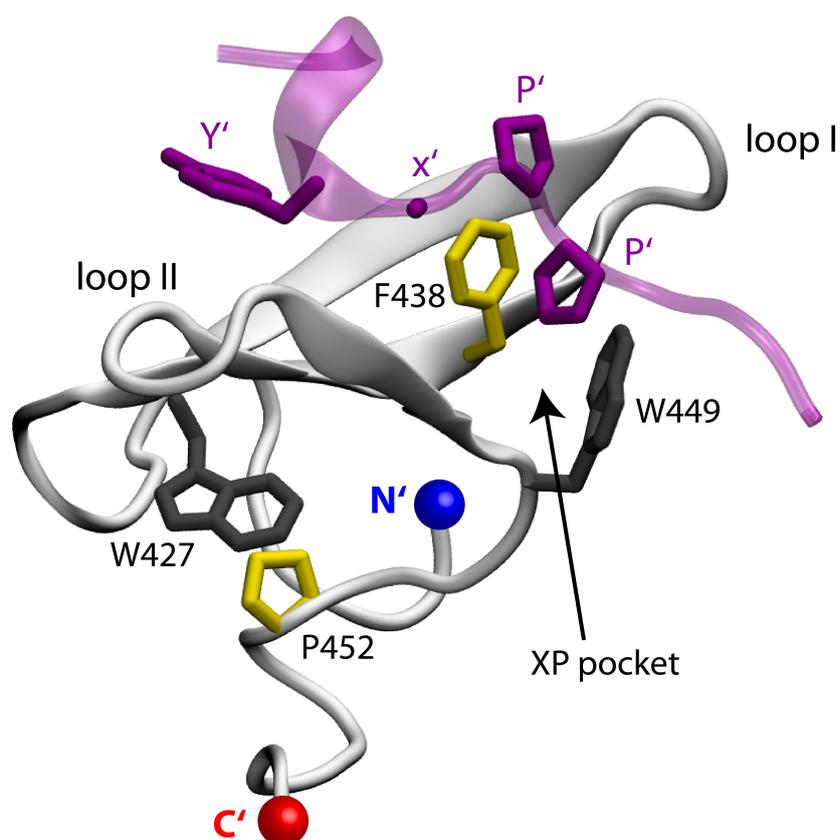


Figure 1.13: hNedd4-1 WW3* domain with the α -hENaC peptide (purple, pdb: 2M3O) containing the PPxY motif. Side chains of the conserved tryptophan residues, the invariant proline as well as F438 that forms the XP pocket with W449 are shown as sticks representations.

Structures of WW domains in complex with their cognate peptide reveal that the first tryptophan residue and the invariant proline are crucial for domain stability,

whereas the second tryptophan is important for peptide recognition [128–132]. The second tryptophan, along with one other conserved aromatic residue (typically phenylalanine), forms the “XP pocket” that binds a conserved proline in proline-rich motifs that are recognized by WW domains. WW domains adopt a characteristic three-stranded antiparallel β -sheet fold and can be subdivided into five different classes according to the proline-rich sequence they recognize. The WW domains of Nedd4 family members bind the PPxY (PY) motif-peptides and are thus members of the class I WW domains [133].

The WW domains of hNedd4-1 show a high sequence similarity, however, they bind the α -hENaC PY motif with different affinities [134,135] as shown in Figure 1.14. WW1 showed no physiologically relevant affinity, whereas the highest affinity was reported for WW3 (WW3*, the asterisk designates a high-affinity domain) with a K_D of $\sim 5 \mu\text{M}$. WW2 and WW4 showed binding, however with a significantly lower binding affinity (s. Figure 1.14). Furthermore, a tandem construct of WW2, WW3* and WW4 of hNedd4-1 did not show increased peptide binding affinity towards the α -hENaC peptide, demonstrating the WW3* domain suffices for high-affinity binding [134].

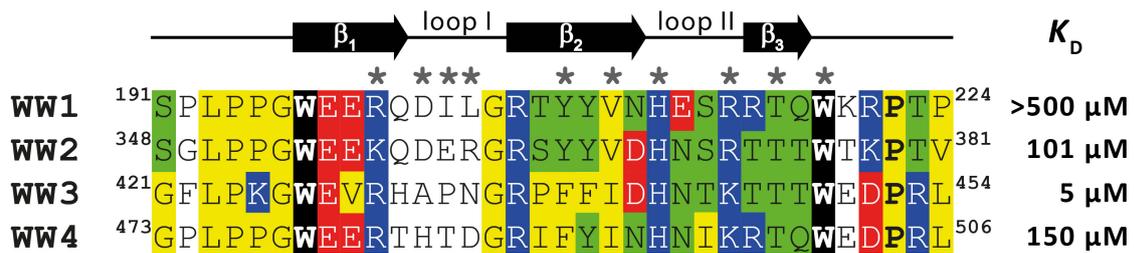


Figure 1.14: Sequence homology between the four WW domains is highlighted for nonpolar (yellow), polar (green), acidic (red) and basic (blue) amino acids and the two conserved tryptophans (black). The invariant proline along with the two tryptophan residues is highlighted in bold letters. Sequence numbering is according to NCBI entry (NM_006153.3) and CLUSTALW [136] was used for sequence alignment. Additionally, the binding affinity of the individual WW domains towards the α -hENaC PY motif [134] is presented. Directly interacting residues are highlighted with a gray asterisk. Adapted from [86].

Sequence alignment of the four hNedd4-1 WW domains illustrates that most residues involved in direct α -hENaC peptide interaction (s. Figure 1.14) are sequence identical or at least have a conserved substitution. The high degree of sequence similarity and differing binding affinities from Nedd4 extends to other species, including mouse, rat, *Drosophila* and *Xenopus* [135]. The highest sequence diversity among the hNedd4-1 WW domains is observed in loop I (s. Figure 1.14) that adopts a type I β -turn connecting the first two β -sheets. In this type I β -turn of the WW3* domain, three

residues interact with the α -hENaC peptide [137], indicating that peptide-binding affinity of the WW domains may be defined by the loop I sequence. Mutagenesis studies of Nedd4 have illustrated that exchanging loop I residues affects the peptide binding affinity [135, 138]. Moreover, studies of the human peptidyl-prolyl isomerase (Pin1) WW domain showed that, in addition to the amino acid composition, loop I flexibility affects the peptide-binding affinity [139]. Hence, chemical recognition as well as dynamics appear to play a role in determining the peptide binding affinity of WW domains.

Chapter 2

Aims

The objective of this thesis was to obtain a deeper understanding of the structural and dynamic determinants of protein function using a combination of NMR spectroscopy and MD simulations. To this end, two different protein systems namely *ecDERA* and WW3* domain of hNedd4-1, were studied.

Previous studies of *ecDERA* highlighted the importance of the C-terminal tyrosine (Y259), however its involvement in the catalytic reaction remained elusive. Missing electron density of the last 8 to 9 C-terminal residues in *ecDERA* crystal structures have hampered structural analyses of this region. Thus, the aim for this work was to understand the involvement of Y259 in the catalytic reaction and to obtain a structural and dynamic characterization of the C-terminal tail. Due to the size of the homodimeric *ecDERA* (56 kDa), an active, monomeric variant, carrying the K58E and Y96W mutations (DERAm) [108], was used for these studies.

Target binding specificity and affinity of hNedd4-1 is defined by the protein-protein interaction modules, the WW-domains, with the third domain showing the highest binding affinity to the PPxY motif of the α -ENaC peptide. The objective of this project was to characterize the structure and dynamics of the high affinity third WW domain (WW3*) from hNedd4-1, in the apo-state and in complex with the α -ENaC peptide, to understand the molecular basis of the tight binding. Considering the high sequence diversity in loop I amongst the four hNedd4-1 WW domains, and its uniqueness in WW3* domain, loop I stability could be a key aspect defining the high affinity binding. In particular, the effect of the exchange of the highly preferred proline in $i+1$ position to a threonine, to loop I and the overall WW3* domain stability was investigated.

Chapter 3

Scientific publications

1. **^1H , ^{13}C , and ^{15}N backbone and sidechain resonance assignments of a monomeric variant of *E. coli* deoxyribose-5-phosphate aldolase**
M. Schulte, M. Stoldt, P. Neudecker, J. Pietruszka, D. Willbold and V. Panwalkar
Biomolecular NMR Assignments **2017**, 11(2), 197-201.

Complete execution of experimental procedures and data evaluation including backbone and side chain resonance assignment of *E.coli* DERA monomer, 90 % composition of the manuscript.

2. **Conformational Sampling of the Intrinsically Disordered C-terminal Tail of DERA is Important for Enzyme Catalysis**
M. Schulte, D. Petrović, P. Neudecker, R. Hartmann, J. Pietruszka, S. Willbold, D. Willbold and V. Panwalkar
ACS Catalysis **2018**, 8, 3971-3984

Protein expression, purification and execution of all NMR experiments (except ^{31}P NMR) and subsequent processing and analysis of the data. Further, I carried out all enzyme activity assays. The calculation of DERA_m with the C-terminal tail in the closed state using Xplor-NIH was carried out by P. Neudecker and the MD simulations were setup and analyzed by D. Petrović. My overall contribution to the manuscript is about 75 %.

3. **The Nedd4-1 WW Domain Recognizes the PY Motif Peptide through Coupled Folding and Binding Equilibria**
V. Panwalkar, P. Neudecker, M. Schmitz, J. Lecher, M. Schulte, K. Medini, M. Stoldt, M. A. Brimble, D. Willbold and A. J. Dingley
Biochemistry **2016**, 55(4), 659-674.

Complete MD simulation setup and trajectory analysis, 10 % composition of the manuscript.

4. **Data describing the solution structure of the WW3* domain from human Nedd4-1**

V. Panwalkar, M. Schulte, J. Lecher, M. Stoldt, D. Willbold and A. J. Dingley
Data in Brief **2016**, 8, 605-612.

Complete MD simulation setup and trajectory analysis, 25 % composition of the manuscript.

5. **Proline Restricts Loop I Conformation of the High Affinity WW Domain from Human Nedd4-1 to a Ligand Binding-Competent Type I β -Turn**

M. Schulte, V. Panwalkar, S. Freischem, D. Willbold and A. J. Dingley
The Journal of Physical Chemistry B **2018**, 122(15), 4219-4230

Complete MD simulation setup and trajectory analysis, protein expression and purification and thermal melt analysis. NMR measurements and analysis was equally carried out by me and V. Panwalkar, and S. Freischem helped with protein expression and purification. My overall contribution to the manuscript is about 75 %.

3.1 ^1H , ^{13}C , and ^{15}N backbone and sidechain resonance assignments of a monomeric variant of *E. coli* deoxyribose-5-phosphate aldolase

Summary

Before any structural or dynamic NMR analyses can be performed, chemical shift assignments need to be obtained. In this publication, we provide a near complete set of backbone and side chain chemical shift assignments of a monomer variant of *ecDERA* (DERAm) in HEPES buffer (BMRB accession number: 27048). Moreover, an expression and purification protocol of this construct is provided. TALOS-N [26] secondary structure prediction, based on the chemical shifts, reveals presence of eight β -strands and eleven α -helices. These predictions are in accordance with the TIM barrel fold and in close agreement with the X-ray structure of DERAm [108]. We further show, that presence of a C-terminal His₆-tag, used in other biochemical studies of *ecDERA*, alters the chemical environment of active-site residues and its presence drastically reduces the catalytic activity of DERAm.

Acknowledgments Reprint Permission

This section contains a complete reprint of the publication published in Biomolecular NMR Assignments (<https://link.springer.com/journal/12104>). Reprinted by permission from Springer Customer Service Center GmbH: Springer, *Biomol NMR Assign* **2017**, 11(2), 197-201, doi:[10.1007/s12104-017-9747-6](https://doi.org/10.1007/s12104-017-9747-6) (^1H , ^{13}C , and ^{15}N backbone and sidechain resonance assignments of a monomeric variant of *E. coli* deoxyribose-5-phosphate aldolase, M. Schulte, M. Stoldt, P. Neudecker, J. Pietruszka, D. Willbold and V. Panwalkar), Copyright 2017.

^1H , ^{13}C , and ^{15}N backbone and sidechain resonance assignments of a monomeric variant of *E. coli* deoxyribose-5-phosphate aldolase

Marianne Schulte^{1,2} · Matthias Stoldt^{1,2} · Philipp Neudecker^{1,2} · Jörg Pietruszka^{3,4} · Dieter Willbold^{1,2} · Vineet Panwalkar^{1,2} 

Received: 28 March 2017 / Accepted: 29 May 2017 / Published online: 30 May 2017
© Springer Science+Business Media Dordrecht 2017

Abstract Deoxyribose-5-phosphate aldolase (DERA) catalyses the reversible conversion of 2-deoxyribose-5-phosphate (dR5P) into glyceraldehyde-3-phosphate (G3P) and acetaldehyde. For industrial applications, this enzyme is used in organic synthesis for aldol reactions between acetaldehyde as a donor and a wide range of aldehydes as acceptors. Here, we present a near complete set of sequence-specific ^1H , ^{13}C and ^{15}N resonance assignments of a 28 kDa monomeric variant of the *Escherichia coli* DERA. These assignments provide the basis for ongoing structural and dynamic analysis of DERA substrate specificity.

Keywords NMR assignments · Pentose-phosphate-pathway · DERA · Enzyme catalysis · Non-uniform sampling

Biological context

Deoxyribose-5-phosphate aldolase (DERA, E.C. 4.1.2.4) is an important enzyme participating in key metabolic

pathways, like nucleotide catabolism, pentose-phosphate pathway and glycolysis, in living organisms (Tozzi et al. 2006). It is responsible for deoxyribose phosphate aldolase activity in bacteria where it catalyses the conversion of 2-deoxy-D-ribose-5-phosphate (dR5P) into glyceraldehyde-3-phosphate (G3P) and acetaldehyde (Machajewski and Wong 2000). DERA belongs to class I aldolases whose reaction mechanism involves formation of a Schiff base intermediate with a conserved active-site lysine residue (Hoffee et al. 1974; Rose and O'Connell 1969). Because of its role in the production of key metabolic intermediates crucial for bacterial survival, DERA is also being considered as an interesting pharmaceutical target against several bacterial pathogens (Cao et al. 2016; Tonkin et al. 2015).

In addition to its role in metabolic pathways, DERA is also able to catalyse the carbon–carbon bond formation under mild conditions with stereospecific control and irrespective of substrate structure, thus showing an abundant potential as an environment-friendly biocatalyst (Windle et al. 2014). DERA-catalysed reactions are already being used to synthesize crucial pharmaceutical precursors like statins (Greenberg et al. 2004). The *Escherichia coli* DERA (DERA_{EC}) is of particular interest because it is unique in catalysing cross-aldol condensation between two aldehydes (Barbas et al. 1990) along with having high protein stability and expression yields. However, DERA_{EC} shows poor tolerance towards high concentrations of acetaldehyde, the enzyme is irreversibly inactivated at acetaldehyde concentrations of ~300 mM (Dick et al. 2016a; Hoffee et al. 1965). Considerable success has been achieved in modifying DERAs for accelerated enzyme catalysis (Britton et al. 2016) and higher acetaldehyde tolerance to increase product yield (Dick et al. 2016a).

Besides its strict dependence on acetaldehyde as a donor molecule, DERA_{EC} is also heavily dependent on

✉ Vineet Panwalkar
v.panwalkar@fz-juelich.de

¹ ICS-6 (Strukturbiochemie), Forschungszentrum Jülich, 52425 Jülich, Germany

² Institut für Physikalische Biologie, Heinrich-Heine-Universität, 40225 Düsseldorf, Germany

³ Institute of Bioorganic Chemistry, Heinrich-Heine-Universität im Forschungszentrum Jülich, 40225 Düsseldorf, Germany

⁴ Institute of Bio- and Geosciences IBG-1: Biotechnology, Forschungszentrum Jülich, 52425 Jülich, Germany

phosphorylated substrate molecules (Barbas et al. 1990). Furthermore, the C-terminal tyrosine residue of DERA_{EC} is crucial for enzyme catalysis (Heine et al. 2001; Hoffee et al. 1974). However, the exact mechanism of the underlying two processes is poorly understood. Although several high resolution crystal structures of DERA_{EC} are available (Dick et al. 2016a; Heine et al. 2001, 2004), the mechanistic role of its C-terminus remains elusive because the electron density for the last ~10 C-terminal amino acids is poorly defined, indicating significant conformational mobility. Thus we are using NMR spectroscopy to elucidate the structural and dynamic determinants of DERA_{EC} function. Wild type DERA_{EC} is a symmetric homodimer with a combined molecular weight of 56 kDa. This high molecular weight makes the use of NMR spectroscopy to provide atomic-resolution structural and dynamic information challenging. Therefore, we have used a monomeric variant of DERA_{EC}, a K58E-Y96W double mutant (DERA_{EC} monomer), which exhibits a virtually identical tertiary structure as observed by X-ray crystallography and similar enzymatic activity (Dick et al. 2016a).

As a crucial first step in the investigation of the structure and dynamics of DERA_{EC} monomer by NMR spectroscopy in order to understand the substrate specificity and the role of the mobile C-terminus towards the enzyme monomer activity, we have obtained sequence-specific chemical shift resonance assignments of the backbone and sidechain ¹H, ¹³C and ¹⁵N nuclei. As a prerequisite for NMR analysis of the structure and dynamics of DERA_{EC}, these resonances will therefore add to the ongoing development of DERA_{EC} variants with enhanced or novel enzyme activity.

Materials and experiments

Protein expression and purification

The gene encoding DERA_{EC} monomer was amplified by PCR and inserted into pET15b-Kombi-P expression vector carrying an amino-terminal His₆-tag and a PreScission protease cleavage site. The recombinant plasmid was transformed into chemically competent *E. coli* BL21(DE3) T1^R (Sigma Aldrich Chemie GmbH, Munich, Germany) for expression. Bacterial cells were grown at 37 °C in 100 ml LB medium with ampicillin overnight and then transferred into M9 minimal medium containing ¹⁵NH₄Cl and ¹³C-glucose as exclusive nitrogen and carbon sources, respectively. At the OD₆₀₀ of 0.8, the protein expression was induced with isopropyl-β-thiogalactoside (IPTG) at a final concentration of 0.25 mM and the cells were grown for an additional 16 h at 27 °C. The cells were harvested by centrifugation for 20 min at 5000g and resuspended in lysis buffer containing 20 mM potassium phosphate (pH

7.0). The cells were lysed using a cell-disruptor (I&L Biosystems GmbH, Königswinter, Germany) and the soluble lysate was subsequently loaded onto a Ni²⁺-affinity chromatography column (GE Healthcare, Freiburg, Germany). Column-bound His₆-tagged protein was washed with the lysis buffer containing an additional 10 mM imidazole and then was eluted with 250 mM imidazole. Fractions containing the DERA_{EC} monomer were pooled and treated with PreScission protease (GE Healthcare, Freiburg, Germany) overnight to remove the N-terminal His₆-tag, leaving two exogenous residues (GP) at the N-terminus. The tagless protein was further purified and buffer exchanged into 50 mM HEPES buffer (pH 6.8) using size-exclusion chromatography (HiLoad 16/60 Superdex 75 prep-grade, GE Healthcare, Freiburg, Germany). The final NMR samples were concentrated up to ~850 μM using a 3.0 kDa molecular weight cut-off filter in an Amicon chamber (Merck Millipore, Darmstadt, Germany), freeze-dried and stored at -20 °C. The protein concentration was estimated from the absorbance at 280 nm using an extinction coefficient of 18,450 M⁻¹ cm⁻¹.

NMR spectroscopy

The NMR samples of DERA_{EC} monomer contained [U-¹³C, ¹⁵N] or [U-¹⁵N] DERA_{EC} K58E Y96W in 50 mM HEPES buffer (pH 6.8) with 0.03% (w/v) NaN₃ and 1 mM 2,2,3,3-d(4)-3-(tetramethylsilyl)propionic acid sodium salt (TSP, Thermo Fisher Scientific, Karlsruhe, Germany) in a 90/10% (v/v) H₂O/D₂O mixture.

NMR spectra were recorded at 25.0 °C on Bruker Avance III HD (600 and 700 MHz) and Varian VNMRs (800 and 900 MHz) NMR spectrometers equipped with cryogenically cooled z-gradient probes. The NMR sample temperature was calibrated with a perdeuterated methanol sample as described by Findeisen et al. (2007). Sequence specific backbone and sidechain chemical shift assignments of DERA_{EC} monomer were accomplished using standard multidimensional experiments (Sattler et al. 1999), namely 2D ¹H-¹⁵N HSQC, 2D ¹H-¹³C CT HSQC, 2D ¹H-¹³C CT HSQC (aromatic region), 3D HNCO, 3D HNCACB, 3D HNCA, 3D CBCA(CO)NH, 3D CC(CO)NH, 3D H(CCO)NH, 2D (HB)CB(CGCD)HD, 2D (HB)CB(CGCDCE)HE, 3D HBHA(CBCACO)NH, 3D HBHANH, 3D HCCH-TOCSY, 3D ¹H-¹⁵N TOCSY-HSQC and 3D ¹H-¹³C NOESY-HSQC (aromatic region, mixing time: 140 ms). The triple resonance experiments, except for the 3D ¹H-¹³C NOESY-HSQC, were recorded using non-uniform sampling of the NMR data based on a 10% Poisson gap sampling schedule (Hyberts et al. 2012) and subsequently reconstructed using the iterative soft threshold method (Hyberts et al. 2013). Proton chemical shifts were referenced to TSP, whereas ¹⁵N and ¹³C chemical shifts were

indirectly referenced according to the ratios given by Wishart et al. (Wishart et al. 1995). The NMR data was processed using the NMRPipe package (Delaglio et al. 1995) and analysed with CcpNMR Analysis (Vranken et al. 2005).

Activity essays

To determine the enzymatic activity, consumption of NADH (Sigma Aldrich Chemie GmbH, Munich, Germany) in a retroaldol reaction with 0.02–1.5 mM dR5P in 50 mM HEPES (pH 6.8) at 19 °C was monitored at 340 nm using a coupled enzyme assay as described previously (Dick et al. 2016a; Nicholas 1988).

Resonance assignment, data deposition and chemical shift analysis

As seen in Fig. 1, DERA_{EC} monomer shows a very well-dispersed ¹H-¹⁵N HSQC spectrum in 50 mM HEPES, pH 6.8, in the absence of any His₆-tag. A recent combined NMR and biochemical analysis on DERA_{EC} monomer variant, in a buffer containing 50 mM potassium phosphate pH 6.8, employed a C-terminal His₆-tagged protein (^{His}DERA_{EC} monomer) (Dick et al. 2016a, b) for investigation of the enzyme inactivation at high acetaldehyde concentrations (Dick et al. 2016a). Since *E. coli* DERA has a phosphate binding site (Heine et al. 2001), we decided to switch from the previously used potassium phosphate buffer to a non-phosphate buffer (50 mM HEPES, pH 6.8). Comparison of a ¹H-¹⁵N HSQC spectrum of the tagless

DERA_{EC} monomer with the spectrum of ^{His}DERA_{EC} monomer immediately revealed a large number of conspicuous chemical shift changes upon removal of the His₆-tag in several regions of the protein, most importantly in the vicinity of the active site. This suggests that the His₆-tag may interfere with DERA activity. Indeed, a comparison of DERA_{EC} monomer enzyme activity for the C-terminally His₆-tagged and the tagless variants determined in 50 mM HEPES, pH 6.8, revealed a 15-fold increase in DERA_{EC} monomer kinetic rate constant, k_{cat} , after removal of the C-terminal His₆-tag (Table 1). This clearly demonstrates that only tagless DERA can be considered highly active and we recommend using this form for all future biochemical and structural studies involving DERA_{EC}. Unfortunately the extent of the chemical shift changes also means that the previous backbone-only chemical shift assignments of ^{His}DERA_{EC} monomer (Dick et al. 2016a) (BMRB accession number 25904) are inadequate for analysing NMR spectra of tagless DERA_{EC} monomer. This forced us to start over and record a complete set of backbone and side-chain assignment experiments in order to obtain the NMR resonance assignments of the DERA_{EC} monomer in its biochemically and structurally relevant high activity form.

Sequence-specific assignments for 96% of all the backbone atoms, in 50 mM HEPES buffer, have been achieved. In total, 246 out of 251 possible non-proline backbone amide groups, 94.2% of all the C' atoms (244 out of 259), 99.2% of all the C α (257 out of 259) and 98.9% of all the H α atoms (274 out of 277) have been successfully assigned. Additionally, we have successfully assigned the sidechain NH₂ groups of all the asparagine and glutamine residues. Backbone amides for D22, V206, S238, S239 and L240 could not be assigned under the used NMR conditions because they are line-broadened beyond detection, suggesting conformational exchange on an intermediate timescale in these regions. According to the crystal structure of DERA_{EC} (Heine et al. 2001), residues 206, 238, 239 and 240 are located at the interface directly coordinating the phosphate group of dR5P whereas residue 22 is located in close proximity to the region interacting with the aldehyde group of dR5P. Besides the backbone assignments in 50 mM HEPES buffer, we have successfully assigned 81.7% (558 out of 683) sidechain carbons and 86% (797

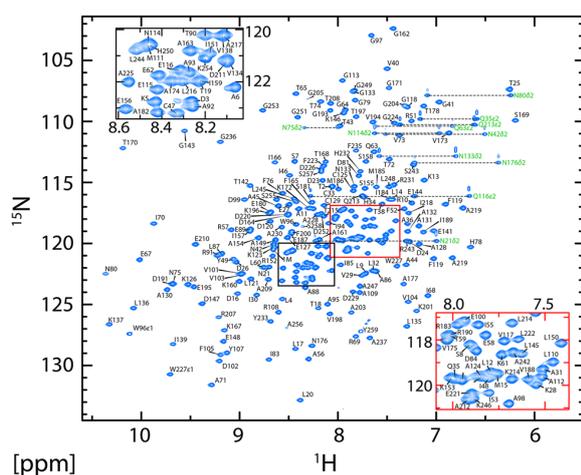


Fig. 1 2D ¹H-¹⁵N HSQC spectrum of DERA_{EC} monomer in 50 mM HEPES at pH 6.8, 0.03% NaN₃ and 90/10% (v/v) H₂O/D₂O mixture recorded at 800 MHz and 25 °C. Horizontal dashed lines represent NH₂ sidechain resonances. Assignment information for the most crowded regions of the spectrum is given in the insets

Table 1 Michaelis–Menten kinetics of aldol cleavage by DERA_{EC} monomer with and without a C-terminal His₆-tag

Enzyme	k_{cat} [s ⁻¹]	K_M [mM]
DERA _{EC} monomer	29.3 ± 0.8	0.344 ± 0.024
^{His} DERA _{EC} monomer	1.94 ± 0.09	0.011 ± 0.003

Triplicate measurements were performed at 19 °C in 50 mM HEPES at pH 6.8 with different 2-deoxy-D-ribose-5-phosphate (dR5P) concentrations ranging from 0.02 to 1.5 mM

out of 927) of all the sidechain protons. These include 96.2% of all the C β atoms (232 out of 241) and 95.4% of all the H β atoms (372 out of 390). The unassigned sidechain atoms included serine γ protons, threonine γ 1 protons, lysine ζ , arginine ϵ and η , histidine NH δ and NH ϵ atoms, and ζ CH groups from tyrosine and phenylalanine residues. The chemical shift resonance assignments for tagless, highly active DERA_{EC} monomer in 50 mM HEPES buffer, pH 6.8, have been deposited with the Biological Magnetic Resonance Data Bank under the accession number 27048.

A quantitative analysis of the aforementioned backbone amide chemical shift changes between ^{His}DERA_{EC} monomer and tagless DERA_{EC} monomer is shown in Fig. 2. For a more accurate assessment of the effect of the His₆-tag itself, this comparison was carried out in the phosphate buffer used previously (Dick et al. 2016a).

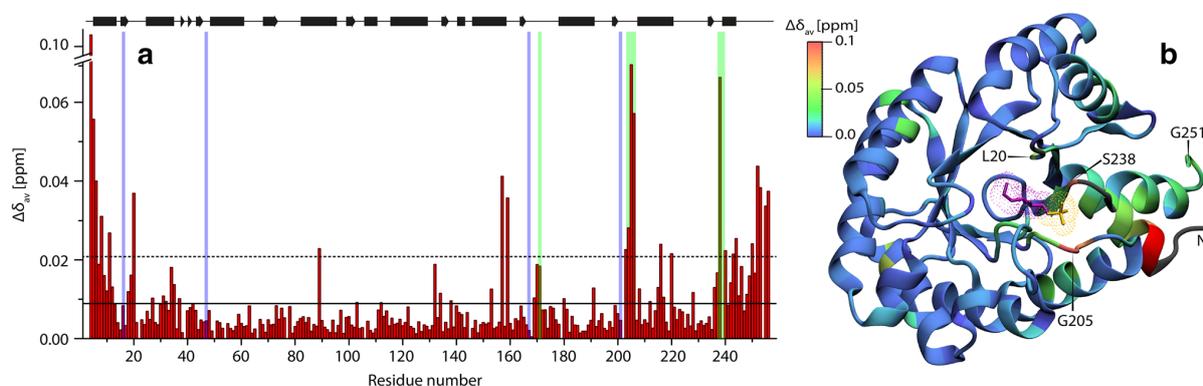


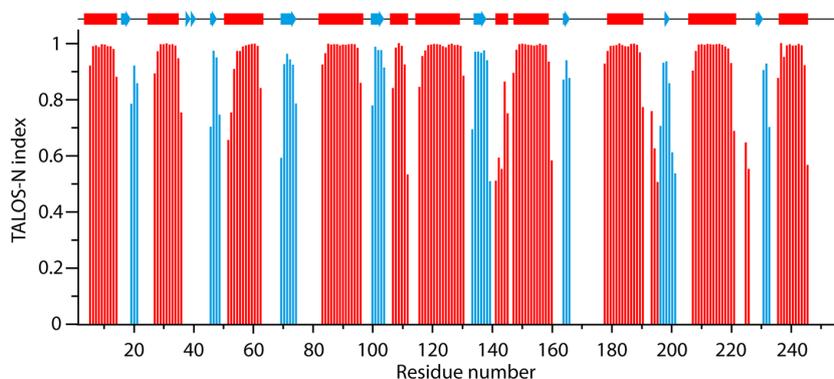
Fig. 2 Weighted average ¹H and ¹⁵N chemical shift differences, $\Delta\delta_{av}$, between DERA_{EC} monomer and ^{His}DERA_{EC} monomer. 2D ¹H-¹⁵N HSQC spectra for both the proteins were recorded under NMR conditions employed by (Dick et al. 2016a). Unlike in HEPES buffer, resonances for V206, S238 and L240 are severely line-broadened but still detectable above the noise threshold of the ¹H-¹⁵N HSQC spectrum in phosphate buffer. $\Delta\delta_{av}$ were calculated using the equation $\Delta\delta_{av} = \{(\Delta H_N^2 + \Delta N^2/25)/2\}^{1/2}$. In **a** the continuous black line indicates the mean $\Delta\delta_{av}$ value of 0.009 ppm and the dotted black line indicates mean plus one standard deviation at 0.02 ppm. Residues in

Sequence-specific backbone chemical shift assignments for the tagless DERA_{EC} monomer in 50 mM potassium phosphate buffer, pH 6.8, were verified using 3D HNCACB and 3D CBCA(CO)NH experiments. The chemical shift changes in the N-terminal region could be a consequence of the “Gly-Pro” overhang from the protease cleavage site. Significant chemical shift changes upon removal of the C-terminal His₆-tag also occur in the C-terminal region and in several additional regions of the protein including the residues involved in coordination of the phosphate group of dR5P (residues 204, 205, 206 and 238) and residues in the vicinity of the active site (residues 11 and 20).

Using the obtained ¹H^N, ¹⁵N, ¹³C^α, ¹H^α, ¹³C^γ, and ¹³C^β assignments (in 50 mM HEPES buffer, pH 6.8), a TALOS-N (Shen and Bax 2013) prediction of secondary structure elements was performed (Fig. 3). It is evident from the

the enzyme active site (D16, C47, K167, K210) are highlighted in blue and residues involved in phosphate coordination (G171, G204, G205, V206, S238, S239), as observed in the DERA_{EC} crystal structure (PDB: 1JCL) (Heine et al. 2001) are highlighted in green. A schematic representation of the secondary structure elements, observed in DERA_{EC} crystal structure, is shown at the top. In **b** the $\Delta\delta_{av}$ are represented on the crystal structure of DERA_{EC} using the colour scale shown in the figure. The open-form of dR5P is represented in magenta with the phosphate group highlighted in yellow

Fig. 3 Secondary structure of DERA_{EC} monomer derived from backbone chemical shifts. The red and blue bars represent α -helices and β -strands, respectively, and the height of the bars reflects the probability of the TALOS-N neural network secondary structure prediction. As a comparison, the secondary structure of DERA_{EC} monomer observed in the crystal structure (PDB id: 5EKY) is shown at the top



chemical shift data that DERA_{EC} monomer in solution shows the presence of eight beta-strands and eleven alpha-helices. The TALOS-N predictions are in accordance with the canonical TIM barrel fold of type I aldolases, in close agreement with the crystal structure of DERA_{EC} monomer (Dick et al. 2016a).

Acknowledgements The authors acknowledge access to the Jülich-Düsseldorf Biomolecular NMR Center that is jointly run by the Forschungszentrum Jülich and Heinrich-Heine-University Düsseldorf. The authors thank Dr Markus Dick for providing the gene encoding the C-terminal His₆-tagged DERA_{EC} monomer in the pET21a vector and Nicole Müller for cloning the gene encoding the DERA_{EC} monomer in pET15b-Kombi-P vector.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Barbas CF, Wang YF, Wong CH (1990) Deoxyribose-5-phosphate aldolase as a synthetic catalyst. *J Am Chem Soc* 112:2013–2014. doi:10.1021/ja00161a064
- Britton J, Meneghini LM, Raston CL, Weiss GA (2016) Accelerating enzymatic catalysis using vortex fluidics. *Angew Chem Int Ed* 55:11387–11391. doi:10.1002/anie.201604014
- Cao T-P, Kim J-S, Woo M-H, Choi JM, Jun Y, Lee KH, Lee SH (2016) Structural insight for substrate tolerance to 2-deoxyribose-5-phosphate aldolase from the pathogen *Streptococcus suis*. *J Microbiol* 54:311–321. doi:10.1007/s12275-016-6029-4
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277–293
- Dick M et al (2016a) Mechanism-based inhibition of an aldolase at high concentrations of its natural substrate acetaldehyde: structural insights and protective strategies. *Chem Sci* 7:4492–4502
- Dick M, Weiergräber OH, Classen T, Bisterfeld C, Bramski J, Gohlke H, Pietruszka J (2016b) Trading off stability against activity in extremophilic aldolases. *Sci Rep*. doi:10.1038/srep17908
- Findeisen M, Brand T, Berger S (2007) A ¹H-NMR thermometer suitable for cryoprobes. *Magn Reson Chem* 45:175–178
- Greenberg WA, Varvak A, Hanson SR, Wong K, Huang H, Chen P, Burk MJ (2004) Development of an efficient, scalable, aldolase-catalyzed process for enantioselective synthesis of statin intermediates. *Proc Natl Acad Sci USA* 101:5788–5793. doi:10.1073/pnas.0307563101
- Heine A, DeSantis G, Luz JG, Mitchell M, Wong CH, Wilson IA (2001) Observation of covalent intermediates in an enzyme mechanism at atomic resolution. *Science* 294:369–374. doi:10.1126/science.1063601
- Heine A, Luz JG, Wong CH, Wilson IA (2004) Analysis of the class I aldolase binding site architecture based on the crystal structure of 2-deoxyribose-5-phosphate aldolase at 0.99 Å resolution. *J Mol Biol* 343:1019–1034. doi:10.1016/j.jmb.2004.08.066
- Hoffee P, Rosen O, Horecker B (1965) The mechanism of action of Aldolases VI. Crystallization of deoxyribose 5-phosphate aldolase and the number of active sites. *J Biol Chem* 240:1512–1516
- Hoffee P, Snyder P, Sushak C, Jargiello P (1974) Deoxyribose-5-P aldolase: subunit structure and composition of active site lysine region. *Arch Biochem Biophys* 164:736–742
- Hyberts SG, Milbradt AG, Wagner AB, Arthanari H, Wagner G (2012) Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional poisson gap scheduling. *J Biomol NMR* 52:315–327. doi:10.1007/s10858-012-9611-z
- Hyberts SG, Robson SA, Wagner G (2013) Exploring signal-to-noise ratio and sensitivity in non-uniformly sampled multi-dimensional NMR spectra. *J Biomol NMR* 55:167–178. doi:10.1007/s10858-012-9698-2
- Machajewski TD, Wong CH (2000) The catalytic asymmetric aldol reaction. *Angew Chem Int Ed* 39:1352–1375
- Nicholas PC (1988) Effects of buffer concentrations, pH, triose phosphate isomerase and glycerol-3-phosphate dehydrogenase on the measurement of the activity of aldolase and glyceraldehyde-3-phosphate dehydrogenase. *Biochem Soc Trans* 16:752–753. doi:10.1042/bst0160752
- Rose IA, O'Connell EL (1969) Studies on the interaction of aldolase with substrate analogues. *J Biol Chem* 244:126–134
- Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution. *Prog Nucl Magn Reson Spectrosc* 34:93–158
- Shen Y, Bax A (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR* 56:227–241. doi:10.1007/s10858-013-9741-y
- Tonkin ML, Halavaty AS, Ramaswamy R, Ruan J, Igarashi M, Ngõ HM, Boulanger MJ (2015) Structural and functional divergence of the aldolase fold in *Toxoplasma gondii*. *J Mol Biol* 427:840–852
- Tozzi MG, Camici M, Mascia L, Sgarrella F, Ipata PL (2006) Pentose phosphates in nucleoside interconversion and catabolism. *FEBS J* 273:1089–1101
- Vranken WF et al (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59:687–696. doi:10.1002/prot.20449
- Windle CL, Müller M, Nelson A, Berry A (2014) Engineering aldolases as biocatalysts. *Curr Opin Chem Biol* 19:25–33
- Wishart DS et al (1995) ¹H, ¹³C and ¹⁵N chemical shift referencing in biomolecular NMR. *J Biomol NMR* 6:135–140

3.2 Conformational Sampling of the Intrinsically Disordered C-terminal Tail of DERA is Important for Enzyme Catalysis

Summary

The previously obtained set of backbone and side chain chemical shift assignments enabled us to perform structural and dynamic NMR studies on DERA_m with focus on the C-terminal tail, and in particular the C-terminal residue Y259. In this publication, we show that the C-terminal tail of DERA_m is flexible in solution and transiently samples a conformation allowing Y259 to enter the active site (closed state). Based on NOE distance restraints, the solution structure of the C-terminal tail in the closed state was derived. Transition between the closed states and open states may be necessary for substrate and/or product diffusion. The importance of Y259 in the catalytic reaction has been demonstrated in numerous biochemical studies, however the structural basis of its involvement remained elusive. Using time resolved hydrogen/deuterium (H/D) exchange experiments of propanal, in presence of DERA_m and DERA_m Y259F, we conclusively show that Y259 ^oOH group facilitates the C2 proton exchange, and thus may facilitate the imine-enamine transition in the multistep reaction of *ec*DERA.

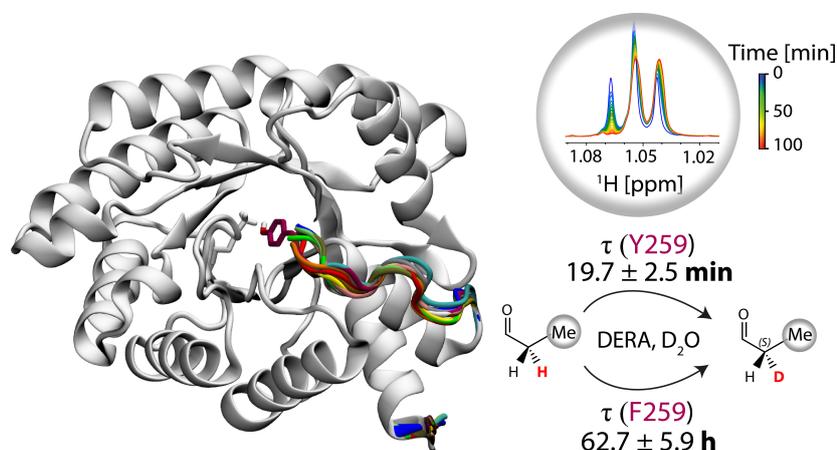


Figure 3.1: Graphical table of contents

Acknowledgments Reprint Permission

This section contains a complete reprint of the publication and supporting information. Reprinted with permission from *ACS Catalysis* **2018**, 8, 3971-3984, doi:10.1021/acscatal.7b04408. Copyright 2018 American Chemical Society.

Conformational Sampling of the Intrinsically Disordered C-Terminal Tail of DERA Is Important for Enzyme Catalysis

Marianne Schulte,^{†,‡} Dušan Petrović,^{‡,§,¶} Philipp Neudecker,^{†,‡} Rudolf Hartmann,[‡] Jörg Pietruszka,^{§,||} Sabine Willbold,[⊥] Dieter Willbold,^{†,‡,¶} and Vineet Panwalkar^{*,†,‡,¶}

[†]Institut für Physikalische Biologie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany

[‡]Institute of Complex Systems 6 (ICS-6): Structural Biochemistry, Forschungszentrum Jülich, 52425 Jülich, Germany

[§]Institute of Bioorganic Chemistry, Heinrich-Heine-Universität im Forschungszentrum Jülich, 52425 Jülich, Germany

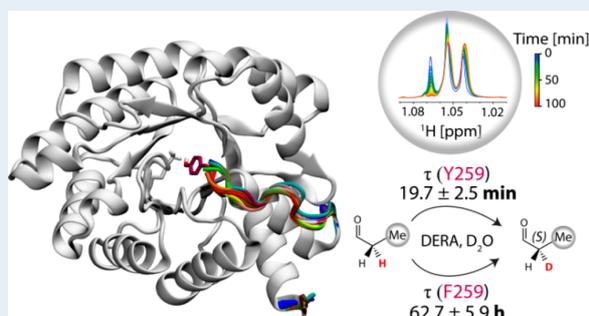
^{||}Institute of Bio- and Geosciences 1 (IBG-1): Biotechnology, Forschungszentrum Jülich, 52425 Jülich, Germany

[⊥]Central Institute of Engineering, Electronics and Analytics (ZEA-3), Forschungszentrum Jülich, 52425 Jülich, Germany

Supporting Information

ABSTRACT: 2-Deoxyribose-5-phosphate aldolase (DERA) catalyzes the reversible conversion of acetaldehyde and glyceraldehyde-3-phosphate into deoxyribose-5-phosphate. DERA is used as a biocatalyst for the synthesis of drugs such as statins and is a promising pharmaceutical target due to its involvement in nucleotide catabolism. Despite previous biochemical studies suggesting the catalytic importance of the C-terminal tyrosine residue found in several bacterial DERAs, the structural and functional basis of its participation in catalysis remains elusive because the electron density for the last eight to nine residues (i.e., the C-terminal tail) is absent in all available crystal structures. Using a combination of NMR spectroscopy and molecular dynamics simulations, we conclusively show that the rarely studied C-terminal tail of *E. coli* DERA (*ecDERA*) is intrinsically disordered and exists in equilibrium between open and catalytically relevant closed states, where the C-terminal tyrosine (Y259) enters the active site. Nuclear Overhauser effect distance restraints, obtained due to the presence of a substantial closed state population, were used to derive the solution-state structure of the *ecDERA* closed state. Real-time NMR hydrogen/deuterium exchange experiments reveal that Y259 is required for efficiency of the proton abstraction step of the catalytic reaction. Phosphate titration experiments show that, in addition to the phosphate-binding residues located near the active site, as observed in the available crystal structures, *ecDERA* contains previously unknown auxiliary phosphate-binding residues on the C-terminal tail which could facilitate in orienting Y259 in an optimal position for catalysis. Thus, we present significant insights into the structural and mechanistic importance of the *ecDERA* C-terminal tail and illustrate the role of conformational sampling in enzyme catalysis.

KEYWORDS: aldolase, TIM-barrel fold, Hamiltonian replica exchange, NMR H/D exchange, NOE



INTRODUCTION

The fundamental principle of enzyme catalysis is the ability of an enzyme to decrease the transition-state energy (or destabilize the ground state) and thereby accelerate the chemical reactions which would otherwise not occur on biologically relevant time scales.^{1–3} Enzymes are intrinsically dynamic, and the emerging consensus is that conformational sampling plays a crucial role in enzyme catalysis.^{4–13} However, describing the direct link between protein motions and enzymatic function remains challenging. Under native conditions, enzymes sample various conformations separated by energy barriers that determine the rates of exchange between these substates.^{10,14,15} It has been shown that some of the substates sampled by enzymes in the absence of a substrate resemble the substrate-bound state,^{6,16–19} thus demonstrating

the importance of conformational sampling in enzyme catalysis.^{9,20}

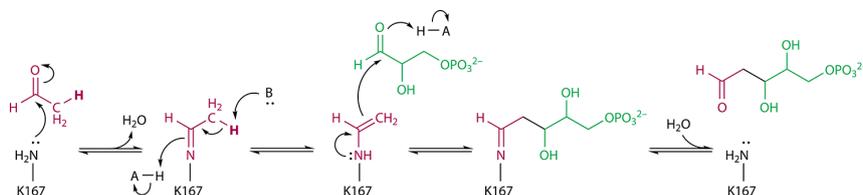
2-Deoxyribose-5-phosphate aldolase (DERA, E.C. 4.1.2.4) is a class I aldolase with a triosephosphate isomerase (TIM) barrel fold and a highly conserved sequence in bacteria. In vivo, DERA catalyzes the reversible conversion of acetaldehyde and glyceraldehyde-3-phosphate (G3P) to generate deoxyribose-5-phosphate (dR5P)^{21,22} and provides key intermediates essential for different metabolic pathways, e.g., glycolysis, the pentose-phosphate pathway, and nucleotide catabolism.²³ Therefore, DERA is considered as a promising pharmaceutical target against human bacterial pathogens.^{24,25} In addition to this,

Received: December 21, 2017

Revised: March 24, 2018

Published: March 27, 2018

Scheme 1. Mechanism of the DERA-Catalyzed Aldol Reaction of Acetaldehyde (Magenta) and G3P (Green) Proceeding via a Schiff-Base Intermediate with K167, Giving dRSP^a



^aIn addition, D102, K201, and a water molecule were proposed to act as an acid (A)/base (B) in a proton relay system responsible for abstracting a C2 proton (highlighted in boldface).³³

DERA is unique in catalyzing the cross-aldol condensation between two aldehydes and is utilized as an environmentally friendly biocatalyst for the synthesis of drugs such as statins.^{26–28} However, the absence of a phosphate group from the substrate severely curtails the catalytic efficiency of DERA,²⁹ thereby restricting its use as an efficient biocatalyst to phosphorylated aldehydes. Additionally, mutations of residues interacting with the dRSP phosphate group have also been shown to drastically decrease the activity of *ecDERA*.³⁰ Several efforts have been made to re-engineer DERAs for accelerated enzyme catalysis³¹ and acceptance of nonphosphorylated substrates.³²

On the basis of *Escherichia coli* DERA (*ecDERA*) crystal structures with carbinolamine and Schiff base intermediates, as well as site-directed mutagenesis, Heine et al.³³ have proposed a reaction mechanism of the reversible conversion of acetaldehyde and G3P to generate dRSP, which proceeds through a Schiff base formation with the active site K167 residue (Scheme 1). After the formation of an imine, a proton relay system composed of D102, K201, and a conserved water molecule facilitates the abstraction of a C2 proton (i.e., a proton attached to carbon adjacent to the aldehyde group), allowing the formation of an enamine. The enamine further performs a nucleophilic attack onto the carbonyl carbon of the acceptor aldehyde (G3P) to yield dRSP. The crystal structures also reveal that, within the *ecDERA* active site, the most important substrate coordinating residues are T18, D102, K167, and T170, whereas residues G205 and S238 directly coordinate the dRSP phosphate group, forming the major part of a phosphate-binding site.³³

Despite several biochemical studies suggesting that the C-terminal tyrosine (i.e., Y259) is crucial for efficient catalysis of dRSP breakdown,^{33,34} the absence of electron density for the last eight C-terminal residues (residues D252–Y259 in *ecDERA*; i.e., the C-terminal tail) in all available crystal structures of DERAs containing a C-terminal tyrosine (including *ecDERA*)^{24,32,33,35,36} has precluded acquiring insights into the role of the tyrosine as well as the C-terminal tail. As observed in *ecDERA*, the C-terminal tyrosine residue of class I rabbit muscle fructose-1,6-disphosphate aldolase (FBPA) is important for the catalysis of dihydroxyacetone phosphate (DHAP) breakdown. Unlike *ecDERA*, however, crystal structures of FBPA reveal that, upon binding DHAP, the phosphate dianion recruits the disordered C-terminal region to the active site.³⁷

To investigate the structural and functional basis of Y259 in the *ecDERA* catalytic mechanism, we have carried out a combined molecular dynamics (MD) and NMR spectroscopy study on its monomeric variant, DERAm (carrying K58E and Y96W mutations).^{32,36} Since the high molecular weight (56

kDa) of the dimeric wild-type *ecDERA* makes acquiring atomic-resolution NMR structural and dynamic information challenging, we chose DERAm, which has been shown to have a nearly identical tertiary structure and exhibits enzymatic activity similar to that of the dimeric *ecDERA*.³² In this study, we show that the C-terminal tail is flexible in solution and provide the first direct structural evidence that, in the absence of the substrate, it samples several substates including a catalytically relevant conformation where Y259 is located inside the active site (i.e., the closed state). We present the solution-state structure of the catalytically relevant closed state, derived using NOE distance restraints, in the substrate-free form of DERAm. We further reveal that, in the closed state conformation, the Y259 ^oOH group is responsible for efficient substrate C2 proton exchange and, therefore, could enable a fast transition from the imine to enamine intermediates. Chemical shift perturbations with potassium phosphate reveal that *ecDERA* contains two distinct sets of phosphate-binding residues: the main phosphate-binding residues near the active site as observed from the reaction-intermediate bound crystal structures³³ and previously unknown auxiliary phosphate-binding residues on the C-terminal tail, which could aid in stabilizing the catalytically active conformation of Y259. Thus, our data provide novel insights into DERA function and highlight the role of conformational sampling in the catalysis of dRSP breakdown.

EXPERIMENTAL SECTION

Protein Expression and Purification. Cloning of the DERAm construct was performed as described previously.³⁸ Y259F, S257D/S258D, and the C-terminal tail deletion mutant of DERAm, D252stop (where the new C-terminal residue is G251), were generated with the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, Düsseldorf-Ratingen, Germany) according to the manufacturer's instructions using the primers given in Table S1. The recombinant protein production and purification were carried out as described previously.³⁸

Activity Assay. Enzyme activity was determined at 20 °C in a retro-aldol reaction with 0.02–5.0 mM dRSP in 50 mM HEPES buffer (pH 6.8) using a coupled enzyme assay as described previously.^{32,39} In this assay, the G3P generated in the DERA-catalyzed retro-aldol reaction is isomerized to DHAP by triosephosphate isomerase. DHAP is further reduced to glycerol-3-phosphate by glycerol-3-phosphate dehydrogenase through the oxidation of NADH (Sigma-Aldrich Chemie GmbH, Munich, Germany). The decrease in NADH concentration was monitored at 340 nm.

NMR Experiments. [$U\text{-}^{13}\text{C},^{15}\text{N}$] and [$U\text{-}^{15}\text{N}$] DERAm and its mutants were prepared in 50 mM HEPES buffer (pH 6.8), unless stated otherwise, 0.03% (w/v) NaN_3 , and 1 mM 2,2,3,3-d(4)-3-(tetramethylsilyl)propionic acid sodium salt (TMSP, Thermo Fisher Scientific, Karlsruhe, Germany) in a 90%/10% (v/v) $\text{H}_2\text{O}/\text{D}_2\text{O}$ mixture. The protein concentration was 800 μM . The NMR spectra were recorded on spectrometers operating at proton frequencies of 600, 800, and 900 MHz, equipped with cryogenically cooled z-gradient probes, and at 25 $^\circ\text{C}$. The sample temperature was calibrated using a perdeuterated methanol sample as described by Findeisen et al.⁴⁰ ^1H , ^{13}C , and ^{15}N backbone and side chain resonance assignments for DERAm were published previously (BMRB entry: 27048).³⁸ The backbone assignments for DERAm mutants were confirmed using standard multidimensional heteronuclear NMR experiments.⁴¹ Spectra were processed using the NMRPipe package⁴² and analyzed with CcpNMR Analysis.⁴³

$\{^1\text{H}\}\text{-}^{15}\text{N}$ Heteronuclear NOE. $\{^1\text{H}\}\text{-}^{15}\text{N}$ NOE values for DERAm in 50 mM HEPES (pH 6.8) at 25 $^\circ\text{C}$ were derived from pairs of interleaved spectra recorded with (NOE) and without (reference) proton saturation during the recycle delay. A recycle delay of 14 s was used at 14.1 T. The $\{^1\text{H}\}\text{-}^{15}\text{N}$ NOE values were calculated from peak intensity ratios obtained from the NOE and reference spectra, with uncertainties estimated from background noise of the spectra.

Solvent Paramagnetic Relaxation Enhancement (sPRE). 2D $^1\text{H}\text{-}^{15}\text{N}$ HSQC spectra of DERAm and the Y259F mutant (300 μM each) in 50 mM HEPES (pH 6.8) at 25 $^\circ\text{C}$ were recorded in the presence of 0.5 and 1.0 mM [$\text{Gd}(\text{DTPA-BMA})$]⁴⁴ and in the absence of the paramagnetic agent. The HSQC spectra were recorded with $1024^* \times 140^*$ complex data in the ^1H and ^{15}N dimensions, respectively, with acquisition times of 106.5 ms (t_{HN}) and 67.7 ms (t_{N}). The recycle delay was 5 s, and 16 scans were collected over 6.5 h per experiment. The experiment was performed in duplicate. The ratio of peak intensities in the presence and absence of the paramagnetic agent ($I_{\text{para}}/I_{\text{dia}}$) was used to qualitatively assess the solvent accessibility for DERAm and the Y259F mutant.

NMR Structure Calculation. All NOESY-HSQC experiments were recorded at 800 MHz and 25 $^\circ\text{C}$. Distance restraints were obtained from ^{13}C - and ^{15}N -edited NOESY spectra recorded with mixing times of 100–110 ms.

The NOE distance restraint derived structural ensemble of DERAm closed state conformation was calculated starting from the crystal structure (PDB ID: SEKY).³² Residues missing from the crystal structure (i.e., N-terminal M1–T2 and C-terminal D252–Y259) were added with Swiss-PdbViewer 4.1.0,⁴⁵ and hydrogen atoms were added with NIH version 1.2.1 of XPLOR 3.851 (XPLOR-NIH 1.2.1).^{46,47} Structure preparation was followed by 1000 steps of Powell minimization⁴⁸ based on the bond geometry parameter file paralhgdg.pro modified as described previously⁴⁹ with harmonic restraints to keep the heavy atoms as close to the crystal structure as possible (atomic RMSDs of 0.28 \AA for the backbone and 0.37 \AA for all heavy atoms) while reducing the deviation from the ideal covalent geometry of the force field used for the subsequent structure calculation. The resulting heavy-atom positions of residues D3–H250 were kept fixed, except for the side chains of L245, K246, and H250 and the carbonyl group of H250. The closed conformation of the C-terminal region was determined from 48 unambiguous medium- to long-range distance restraints derived from the three-dimensional NOESY spectra in an iterative

procedure. The NOEs were manually classified into medium (<3.8 \AA) and long (>5.3 \AA) distance restraints on the basis of their corresponding cross peak volumes, i.e., medium and weak. Intra-residual, sequential, and possibly also medium-range NOE cross-peaks in the highly mobile C-terminal region are expected to contain significant contributions from conformations other than the closed state conformation and were therefore not included in the structure calculation. These experimental restraints served as an input for the calculation of 120 structures using restrained molecular dynamics according to a three-stage simulated annealing protocol⁵⁰ using floating assignment of prochiral groups⁵¹ with XPLOR-NIH 1.2.1,^{46,47} as described previously.⁵² The Gaussian conformational database potential⁵³ with a cutoff of 10.0 standard deviations⁵⁴ was included in the target function in order to improve the stereochemical properties. The 26 structures showing the lowest energy values (excluding conformational database potential) and fewest distance restraint violations were selected for further characterization using XPLOR-NIH 1.2.1^{46,47} and PROCHECK-NMR 3.4.⁵⁵

Chemical Shift Perturbation Analysis. Potassium phosphate (KP_i) at concentrations ranging from 0 to 150 mM was titrated against 800 μM DERAm and the Y259F, D252stop, and S257D/S258D mutants in 50 mM HEPES buffer (pH 6.8) to obtain the affinity toward inorganic phosphate. 2D $^1\text{H}\text{-}^{15}\text{N}$ HSQC spectra for each titration point were recorded at 25 $^\circ\text{C}$ and 600 MHz (DERAm and the S257D/S258D and D252stop mutants) and 800 MHz (Y259F mutant). The data matrix of the 2D $^1\text{H}\text{-}^{15}\text{N}$ HSQC at 600 MHz consisted of $160^* \times 1024^*$ complex data points with acquisition times of 77.5 ms (t_{N}) and 107 ms (t_{HN}) and $160^* \times 832^*$ data points and 59 ms (t_{N}) and 64.5 ms (t_{HN}) at 800 MHz. Sixteen scans per titration point were collected with a recycle delay of 1.3 s. The total measuring time was 2 h per 2D $^1\text{H}\text{-}^{15}\text{N}$ HSQC experiment. The equilibrium dissociation constant, K_{D} , was obtained from changes in the weighted average chemical shift differences $\Delta\delta_{\text{av}} = [(\Delta\delta_{\text{HN}}^2 + \Delta\delta_{\text{N}}^2/25)/2]^{1/2}$ assuming a two-state model⁵⁶

$$\Delta\delta_{\text{obs}} = \frac{1}{2}\Delta\delta_{\text{max}} \left[1 + X + \frac{K_{\text{D}}}{[\text{P}_0]} - \sqrt{\left(1 + X + \frac{K_{\text{D}}}{[\text{P}_0]} \right)^2 - 4X} \right] \quad (1)$$

where $[\text{P}_0]$ is the total protein concentration and X is the molar ratio of ligand to protein. In total, chemical shift changes for 21 residues were used to report the average K_{D} values for DERAm and the Y259F mutant and 16 residues for the D252stop mutant, whereas 23 residues could be used for the S257D/S258D mutant to describe the respective affinities toward P_i .

1D ^1H Hydrogen/Deuterium Exchange Experiments. A 200 mM solution of propanal (Sigma-Aldrich Chemie GmbH, Munich, Germany) was incubated in 50 mM potassium phosphate in degassed D_2O (pD 6.8) with 0.5 mM TMSP. In aqueous solution propanal exists in equilibrium with its hydrate, propane-1,1-diol. The chemical shift assignments of the 1D ^1H spectrum of propanal were confirmed using a 2D $^1\text{H}\text{-}^1\text{H}$ COSY spectrum. A reference sample without DERA was recorded as time point $t = 0$. After addition of 5 μM DERAm to the reaction mixture, 1D ^1H spectra were recorded

with a relaxation delay of 20 s, 16384 data points, a spectral width of 9579 Hz, and four scans resulting in 128 s per experiment. A series of 1D spectra was recorded for 3 h. In another experiment, upon the addition of 5 μM Y259F mutant, 1D ^1H spectra were recorded using the same acquisition parameters with 32 scans and resulting in 13 min 57 s per 1D experiment, over a period of 4 days. Peak heights were extracted using NMRDraw.⁴² To quantify the rate of hydrogen/deuterium (H/D) exchange of the C2 proton, the collapse of the triplet representing the methyl group of propanal at 1.06 ppm ($t, {}^3J_{\text{H-H}} = 7.3$ Hz) and propane-1,1-diol at 0.91 ppm ($t, {}^3J_{\text{H-H}} = 7.5$) to a doublet (with coupling constants of $d, {}^3J_{\text{H-H}} = 7.4$ Hz for propanal and $d, {}^3J_{\text{H-H}} = 7.6$ Hz for propane-1,1-diol) was monitored and the ratio of the upfield peak and the downfield peak of each triplicate was plotted as a function of time. The intensity ratios could be fit to the monoexponential equation

$$y(t) = y_0 + Ae^{R_0 t} \quad (2)$$

with $R_0 = 1/\tau$ where y is the intensity ratio, y_0 is the intensity ratio of the reference (without DERA) representing $t = 0$, R_0 is the rate and τ is the time constant describing the H/D exchange kinetics. The fitting was performed using OriginPro 8.5G (OriginLab Corporation, Friedrichsdorf, Germany).

1D ^{31}P NMR Experiments. 1D ^{31}P NMR spectra of 50 mM KPi (pH 6.8) with 800 μM DERAm and its mutants were recorded on a Bruker Avance III HD with a Prodigy cryoprobe at 600 MHz and 25 $^\circ\text{C}$ with 16384* complex data points and an acquisition time of 672.8 ms and 86161 scans with an experimental time of 27.5 h per experiment.

MD Simulations. System Setup. The crystal structure of DERAm (PDB ID: 5EKY),³² resolved at 1.1 \AA resolution, was used as the starting structure for MD simulations. As this structure contains several unassigned residues due to the absence of electron density (i.e., N-terminal M1–T2 and C-terminal D252–Y259), we used the MODELER 9.14 tool⁵⁷ to prepare the full-length protein structure. A model where the C-terminus is positioned laterally to the TIM barrel was selected (K167–Y259 distance of ~ 26 \AA).

For MD simulations, GROMACS 5.1 suite^{58,59} was used. The protein was represented with the Amber 99SB*-ILLN force field^{60–62} in combination with the explicit TIP3P water model.⁶³ For titratable residues, the protonation states were assigned to a pH of 6.8, on the basis of the PROPKA 3 estimate.⁶⁴ The protein was centered in a cubic box, at least 15 \AA away from any edge. A larger than usual box size was used so that the flexible C-terminus does not come in contact with the periodic image even in the extended state. The simulation box was filled with ~ 22000 water molecules, and the charge was neutralized with Na^+ ions. The steepest descent algorithm was used to minimize the system (maximal force of 500 $\text{kJ mol}^{-1} \text{nm}^{-1}$) before equilibrations. In NVT equilibration, the system was heated to 298 K (v-rescale thermostat⁶⁵) and equilibrated for 200 ps. At this stage, positional restraints were applied to all protein atoms (force constant of 1000 $\text{kJ mol}^{-1} \text{nm}^{-2}$). During 2.5 ns NPT equilibration, the restraints were gradually reduced from 1000 to 5 $\text{kJ mol}^{-1} \text{nm}^{-2}$, while the pressure was kept at 1 bar (Berendsen barostat⁶⁶).

The system was modeled under periodic boundary conditions, and the particle mesh Ewald method⁶⁷ was used to treat electrostatic interactions. The short-range nonbonded interactions were calculated under the cutoff of 10 \AA . All bonds

were constrained using the LINCS algorithm.⁶⁸ An integration step of 2.0 fs was used.

HREX-MD Simulation. The Hamiltonian replica exchange (HREX)-MD was performed in GROMACS patched with the Plumed 2.3 plugin.^{69,70} Twelve replicas were simulated, where the Hamiltonian affecting the C-terminus and several loops over the active site were scaled (i.e., residues 19–25, 75–82, 168–178, 202–208, 249–259). The Hamiltonian scaling factors were 1.00, 0.95, 0.91, 0.87, 0.83, 0.79, 0.76, 0.72, 0.69, 0.66, 0.63, and 0.60, which corresponds to the temperature range of 298–497 K. The exchanges between replicas were attempted every 4 ps, leading to an exchange acceptance rate of $\sim 30\%$. Production HREX-MD simulation was carried out with the NPT ensemble (v-rescale thermostat and the Parrinello–Rahman barostat⁷¹). Each replica was simulated for 220 ns (accumulated sampling time of 2.64 μs), with coordinates of the system collected every 5 ps. The replica with the unperturbed Hamiltonian (i.e., scaling factor of 1.00) was used for the analysis.

MD Simulations. A set of 120 frames, where Y259 was observed in the active site (distance between K167 ^5N and Y259 $^{\text{O}} < 6$ \AA) was selected from HREX-MD. From each frame, an unrestrained MD simulation was set up. The water box was reduced to at least 10 \AA away from any edge, while preserving the cubic shape. The system was minimized and equilibrated as initially described. Production MD simulations were 50 ns long, providing a total of 6 μs of sampling, and coordinates of the system were saved every 20 ps.

Structural Analysis. The analysis of MD trajectories was performed using GROMACS tools and the MDTraj library.⁷² Hydrogen bonds were identified according to the Baker–Hubbard criteria,⁷³ i.e., where the H...acceptor distance was shorter than 2.5 \AA and the donor–H...acceptor angle was greater than 120° , in at least 30% of frames. For the identification of the hydrophobic contacts between residues, all side chain C atoms not directly bound to a heteroatom were considered. Given a residue pair, a hydrophobic contact was defined if the shortest distance between the two residues was lower than 4.5 \AA , in at least 30% of the frames.

RESULTS

Efficient Catalysis by ecDERA Requires Y259 $^{\text{O}}\text{H}$. The ecDERA reaction mechanism, proposed by Heine et al.,³³ describes the reversible conversion of acetaldehyde and G3P to dRSP, through a Schiff-base intermediate with the catalytic K167. Several studies have indicated that the side chain hydroxyl group of the C-terminal tyrosine is crucial for DERA catalysis^{33,34} and the Y259F mutation results in a reduction in the catalytic activity of 2 orders of magnitude.³³ Similarly, the DERAm variant, used in our study, shows an ~ 100 -fold decrease in the kinetic rate constant due to the Y259F mutation (k_{cat} , Table 1), further highlighting the importance of this residue in the catalytic mechanism.

C-Terminal Tail of DERAm Is Flexible in Solution and Transiently Occupies the Active Site. In all the available

Table 1. Catalytic Activities of DERAm and Its Y259F Mutant Using dRSP as a Substrate

DERA variant	k_{cat} (s^{-1})	K_{M} (mM)	$k_{\text{cat}}/K_{\text{M}}$ ($\text{s}^{-1} \text{M}^{-1}$)
DERAm	19.0 ± 1.0	0.31 ± 0.01	$(6.1 \pm 0.1) \times 10^4$
DERAm Y259F	0.20 ± 0.05	0.16 ± 0.01	$(1.3 \pm 0.3) \times 10^3$

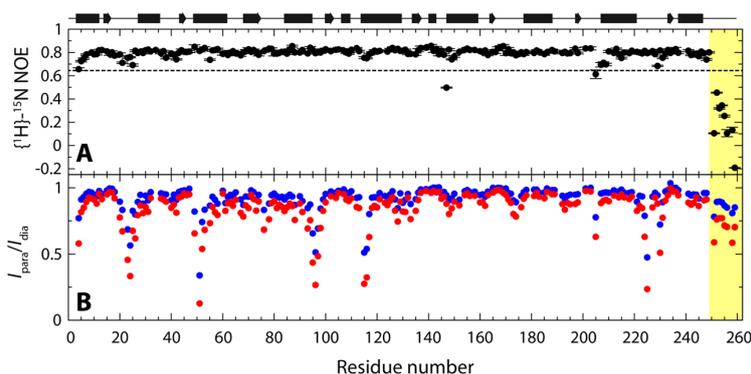


Figure 1. Flexibility of DERAm in solution. (A) Steady-state heteronuclear $\{^1\text{H}\}\text{-}^{15}\text{N}$ NOE values recorded at 25 °C and 14.1 T. The dashed black line is drawn at an NOE value of 0.65. (B) Intensity ratio ($I_{\text{para}}/I_{\text{dia}}$) of the backbone amide resonances of DERAm at two concentrations of the paramagnetic agent [Gd(DTPA-DMA)]: 0.5 mM (blue) and 1.0 mM (red). A schematic representation of the secondary structure elements is shown at the top. Residues in the C-terminal tail are highlighted in yellow.

crystal structures of DERAs containing a C-terminal tyrosine,^{24,32,33,35,36,74} the electron density for the C-terminal tail is poorly defined, which prevents understanding the structural and functional basis of the participation of Y259 in catalysis. To characterize the dynamics of the C-terminal tail in solution, we acquired $\{^1\text{H}\}\text{-}^{15}\text{N}$ heteronuclear NOE (hetNOE) values for DERAm. The hetNOE values, obtained for 204 out of total 251 nonproline residues, suggest that the TIM-barrel fold of DERAm (residues 2–250) is rigid on the picosecond to nanosecond time scale with only G205 and D147 showing hetNOE values lower than 0.65 (Figure 1A). The C-terminal tail shows hetNOE values consistently lower than 0.65, indicating high flexibility in solution, which explains the lack of well-defined electron density in the crystal structures.

Many recent experimental and computational studies have shown that, under native conditions, enzymes have an intrinsic ability to sample open and closed forms in the absence of a substrate.^{9,20,75,76} Upon substrate binding one of these conformations gets stabilized through local rearrangements.²⁰ For the possible participation of DERAm Y259 in the enzymatic reaction, it is plausible that the C-terminal tail would need to sample conformations which allow Y259 to enter the active site (closed state). However, for the substrate binding and product release, the enzyme should preferentially be in a conformation that does not block the active site (i.e., open state). A highly flexible C-terminal tail of DERA leads potentially to many open states which would have higher solvent accessibility in comparison to the closed conformation. Paramagnetic relaxation enhancements, induced using soluble paramagnetic compounds (sPREs), provide a straightforward approach toward structural and dynamic analysis of biomolecules using solvent accessibilities.^{77–79} sPREs have been used to detect transiently populated conformers and changes in the solvent accessibilities of intrinsically disordered regions of proteins.^{80,81}

In our study, we used sPREs of the backbone amide group, with [Gd(DTPA-DMA)] as a paramagnetic agent, to assess the solvent accessibility of DERAm. Intensity ratios ($I_{\text{para}}/I_{\text{dia}}$) of resonances in 2D $^1\text{H}\text{-}^{15}\text{N}$ HSQC spectra recorded with (0.5 mM and 1.0 mM [Gd(DTPA-DMA)]) and without the paramagnetic solvent indicate that several loop regions (N21–E26, A93–A98, E115–V117, G224, A225, A230, R231) and the N-terminus of helix 3 (residues Y49–I53) are the most surface exposed regions of the protein (Figure 1B). In

contrast, the flexible C-terminal tail shows significantly higher $I_{\text{para}}/I_{\text{dia}}$ values in comparison with the aforementioned surface-exposed residues. These observations indicate that the C-terminal tail may sample conformations with low solvent-accessible surface area: e.g., a closed state.

To investigate the possibility of the C-terminal tail adopting a closed state, thereby allowing Y259 to enter the active site, we recorded 2D $^1\text{H}\text{-}^{15}\text{N}$ HSQC spectra of DERAm Y259F mutant. A quantitative analysis of the backbone amide chemical shift differences between DERAm and the Y259F mutant revealed that most of the significant chemical shift changes (i.e., $\Delta\delta_{\text{av}} > 0.015$ ppm) occurred to residues located within the DERAm active site (i.e., T18, L20, T168, T170, G171, A203, G204, G205, and R207) (Figure 2). These residues have been shown to be crucial for DERA–substrate interaction.³³ The observed perturbations demonstrate that the flexible C-terminal tail of DERAm can sample the closed conformation, thereby bringing Y259 into the active site, in the absence of the substrate. Additionally, the $I_{\text{para}}/I_{\text{dia}}$ ratios between DERAm and the Y259F mutant in the presence of 1 mM [Gd(DTPA-DMA)] are nearly identical with each other (Figure S2). This indicates that the absence of the $^{\text{OH}}$ group of Y259 does not result in large scale conformational changes and the aforementioned chemical shift perturbations arise primarily due to the absence of the Y259 side chain hydroxyl group.

MD Simulations Suggest That the Closed State Is Stabilized by a Weak-Interaction Network. The inability to resolve the conformation of the C-terminal tail of *ec*DERA from X-ray crystallography^{32,33,35} was at the origin of an idea that the tail is intrinsically disordered and samples many states.^{33,35} However, the early kinetic investigation pointed out the crucial role of the C-terminus for catalysis by DERA (Table 1). Although the C-terminus is long enough to extend and protrude into the active site,³³ no structure of such a conformation is available. To understand the closed conformation of DERA, where Y259 is in the active site, we explored the conformational ensemble of DERAm using enhanced sampling HREX-MD simulations. We recently demonstrated how HREX-MD can be used to study conformational ensembles of enzyme active sites¹⁰ and here present the sampling on a larger scale. To that extent, we enhanced the sampling of the C-terminal tail together with the loops on the catalytic face. While the secondary structure elements of the TIM (α/β)₈ barrel fold remain stable during 220 ns of HREX-

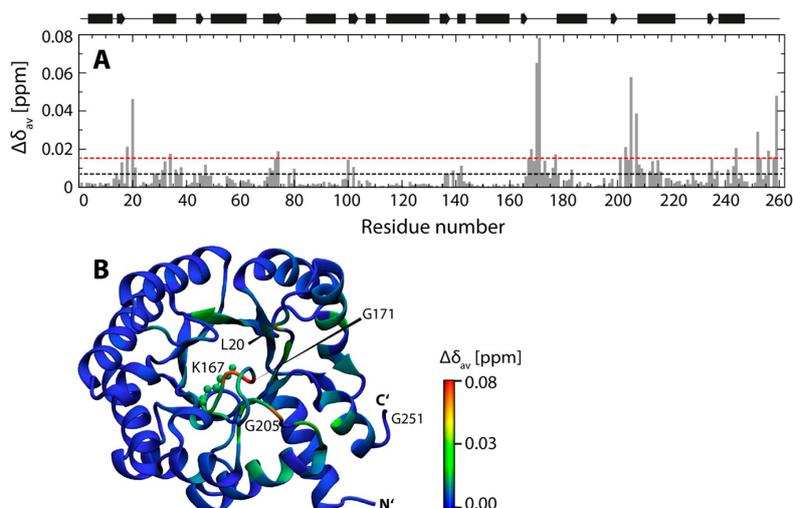


Figure 2. Weighted ^1H – ^{15}N chemical shift differences ($\Delta\delta_{\text{av}}$) between DERAm and the Y259F mutant. (A) The black dotted line represents the average $\Delta\delta_{\text{av}}$ value, whereas the red dotted line represents the average $\Delta\delta_{\text{av}}$ plus one standard deviation (SD). A schematic representation of the secondary structure elements is shown at the top. (B) The $\Delta\delta_{\text{av}}$ values are mapped onto the ecDERA crystal structure (PDB ID: 1JCL) using the coloring scheme shown. The catalytic K167 side chain is shown in a ball and stick representation.

MD (RMSD value of 2 Å from the initial structure), the complete protein backbone indicates much more dynamics (RMSD of up to 6 Å, Figure S3) which comes from the highly flexible C-terminal tail.

To analyze if the C-terminus protrudes into the active site, we measured the distance between the catalytic K167 and C-terminal Y259 residue. Figure S4 shows that the C-terminal tail of DERAm explores a wide range of conformations. While there are multiple open states which present the majority of the complete conformational ensemble, a low closed-state population was identified where the K167–Y259 distance was shorter than 6 Å (Figure S4A). In the open states, the C-terminal tail samples a wide range of conformations, as indicated by structures representing maxima on the distance distribution plot (Figure S4B). The high relative free energy of the closed state, in comparison to the open states, leads to a low occurrence of this state in the HREX-MD simulations. To further sample the closed state, we performed an additional 6 μs of MD (120 simulations starting from different closed state frames identified in the HREX-MD, each 50 ns long). Figure S5A shows that a much higher number of the closed state conformations were sampled, which allowed us to investigate the interactions that stabilize this state.

We used the complete closed state ensemble, identified with MD (around 80000 structures), to investigate the noncovalent interactions in the closed state, which we roughly divide into polar (Figure 3A) and hydrophobic (Figure 3B). However, as the C-terminal tail was found to open in each of the 120 closed state simulations (Figure S5B), these stabilizing interactions are only transient.

The hydrophobic contacts are defined here based on the proximity of nonpolar C atoms in any residue pair, i.e., below the cutoff of 4.5 Å. The distance distributions for hydrophobic contacts are shown in Figure S6. These interactions, stabilizing the closed conformation of the C-terminal tail of DERAm, could be split into two groups: those stabilizing N-terminal half of the tail and those stabilizing Y259 in the active site. The former group involves a particularly complex network of

contacts through the aliphatic side chain of L248, K5, L9, L12, and V40, which seems to tether the C-terminal helix 10 (S239–L248 in PDB ID: 5EKY) with the rest of the protein. L245, located on the helix 10, forms contacts with residues H250 and D252 on the tail (Figure S6L,M). D252 also forms contacts with A242, and A256 is further in contact with L241 and A242, both located on the helix 10 (Figure 3b2). Y259 forms an extensive network of hydrophobic contacts with aliphatic side chains of residues located in the active site: i.e., T18, L20, and K172 (Figure 3a1,b1).

Polar interactions were further split into salt bridges and hydrogen bonds (Figure S7). Only one salt bridge was identified in the closed state, between K172, located on the loop over the active site, and the Y259 C-terminal COO^- group (Figure 3a1 and Figure S7A). Furthermore, several hydrogen bonds were identified between the C-terminal tail and the protein core (i.e., G249–K246, H250–L245, and S255–Q35, Figure S7B–D) and also within the tail itself (A256–G253, S257–K254, and S258–S255, Figure S7E–G). However, the last three hydrogen bonds observed in the C-terminal tail (i.e., residues 253–258) are rarely present simultaneously, appearing together only in 5% of the total closed state ensemble, whereas at least one of the three is present in more than 70% of the sampled conformations.

NOE-Derived NMR Structure of the Closed State. The classical MD simulations indicate that the closed state of the DERAm C-terminal tail gets stabilized through a range of hydrophobic interactions with residues in the C-terminal helix 10 (Figure 3A,b1,b2). To verify the closed state structural representation and confirm the MD-observed interactions that stabilize the closed state in solution, we recorded a 2D ^1H – ^{15}N HSQC spectrum of the D252stop mutant. A quantitative analysis of the backbone amide chemical shift differences between DERAm and the D252stop mutant revealed that most of the significant chemical shift changes (i.e., $\Delta\delta_{\text{av}} > 0.038$ ppm) occurred not only to residues located around the active site but also to residues located in the helix 10 (A242, S243, L244, L245, and K246) (Figure S8). These chemical shift

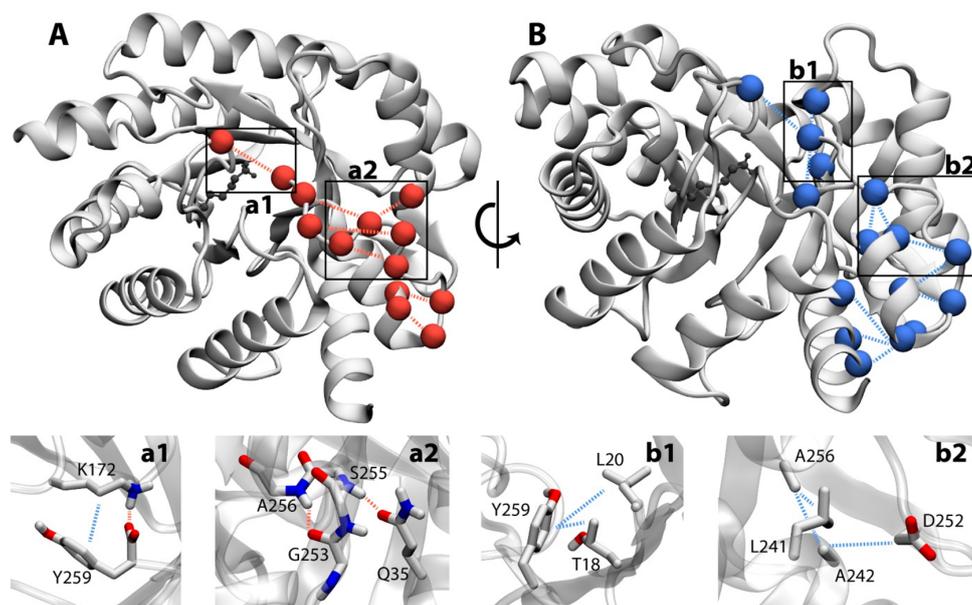


Figure 3. Transient noncovalent interactions stabilize the closed state conformation. (A) Polar and (B) hydrophobic interactions stabilize the C-terminal tail in the closed state conformation. The side chain of the catalytic K167 is highlighted in dark gray. (a1) The hydrophobic interaction between the K172 side chain and the aromatic ring of Y259 as well as the salt bridge between the K172 5 N-group and the C-terminal COO^- group. (a2) Hydrogen bonds between A256 and G253 backbone as well as S255 N to Q35 $^{\epsilon}$ O. (b1) The C-terminal tyrosine Y259 is further stabilized through hydrophobic interactions with T18 and L20. (b2) Residues A256 and D252 have hydrophobic interactions with residues L241 and A242 located at the C-terminal helix 10.

changes coincide with the hydrophobic contacts observed from the MD simulations. Therefore, we recorded NOESY spectra in an attempt to obtain distance restraints between the C-terminal tail and rest of the protein.

Analyses of the NOESY spectra revealed the presence of an unambiguous NOE network between residues in the C-terminal tail and the rest of the protein: in particular, the active site (Figure 4). Although the hetNOE values and TALOS-N secondary structure prediction, using NMR chemical shifts,³⁸ show conclusively that the C-terminal tail is disordered in solution (Figure 1A), the presence of NOEs indicates that the closed conformations of the C-terminal tail could nevertheless be populated in the conformational equilibrium for a substantial fraction of the time. Owing to the steep r^{-6} dependence of the NOE intensity on the interproton distance, r , long-range contacts are expected to give rise to detectable NOE cross-peaks even if these contacts are only transiently populated for a fraction of the NOESY mixing time used.⁸²

We used a total of 48 unambiguous NOE distance restraints (Table S2), between the residues in the C-terminal tail and the rest of the protein (i.e., helix 10 and active site), for the calculation of an ensemble of 26 structures representing the closed state of DERAm (Figure 4A), which is provided as a PDB file in the Supporting Information. Each of these 26 structures individually satisfies the full set of 48 distance restraints without seriously violating the stereochemical quality. None of the C-terminal tail residues lie in the disallowed region of the Ramachandran plot, and only S257 and S258 are in the generously allowed region.

The contribution of the distance restraints to the target function was $0.43 \pm 0.29 \text{ kcal mol}^{-1}$, the root-mean-square deviation from the distance restraints was $0.013 \pm 0.004 \text{ \AA}$, and

no distance restraint was violated by more than 0.12 \AA . The transient presence of Y259 in the active site is conclusively supported by an extensive network of NOEs between the Y259 side chain and residues in the active site (Figure 4A inset). A comparison of inter-residue cross-peak intensities of NOEs between Y259 and other active site residues (Figure 4B–D) indicates that the ϵ atoms of Y259 protrude more deeply in the active site than the δ groups. The distance between the side chain amino group of the active site K167 and the $^{\eta}$ OH group of Y259 in the NMR ensemble was $4.3 \pm 0.5 \text{ \AA}$ ($n = 26$), in support of the observations made by the MD simulations. The population of the closed state represented by the calculated NMR structure was estimated to be $\sim 13\%$ from intensities of NOEs between L20 $\text{H}^{\delta 1*}$ and Y259 $\text{H}^{\epsilon*}$. To estimate the population, we assumed that in open states the contribution to the NOE intensity is zero due to larger interproton distances. The interproton distances were averaged with r^{-6} weighting, and intra-residue NOE between L20 $\text{H}^{\delta 1*}$ and L20 H^{α} was used as an internal distance reference.

It is important to note that although the NMR ensemble satisfies all 48 experimentally derived distance restraints and exhibits structural features very similar to those observed in the MD simulations, this closed state represents only a small—albeit significant—fraction of the full conformational ensemble sampled by the intrinsically disordered C-terminal tail. Moreover, our MD simulations point out that not all stabilizing interactions are simultaneously required to maintain the DERAm closed state: i.e., to keep Y259 inside the active site. This is also illustrated by fluctuations of the distance between the side chain amino group of the active site K167 and the hydroxyl group of Y259 in the apoenzyme (Figure S5). Therefore, due to the flexible nature of the C-terminal tail, we cannot rule out the existence of a more diverse closed state

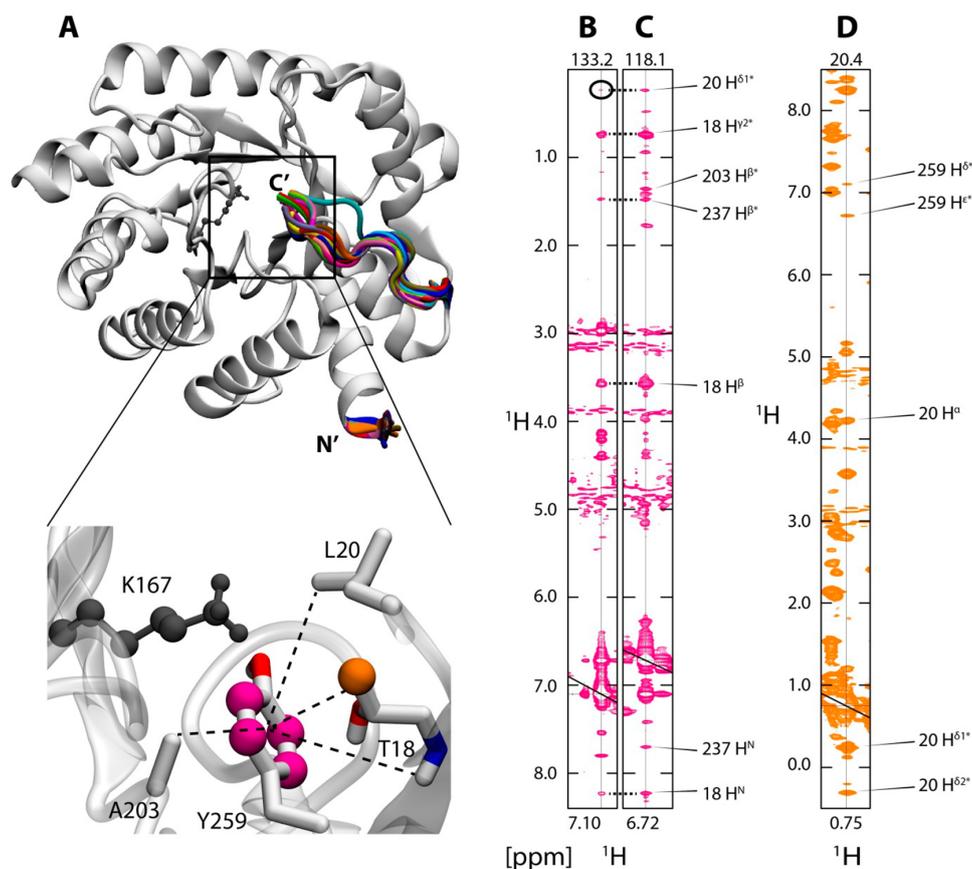


Figure 4. NOE distance restraint derived structural ensemble of DERAm C-terminal tail closed state. (A) A set of 48 NOEs satisfies the closed state ensemble. The inset shows that the presence of Y259 inside the active site is supported by NOEs (dashed lines) to nearby residues. Strips from ^{13}C -edited NOESY spectra showing inter-residue NOE interactions for 259 $\text{H}^{\delta*}$ (B), 259 $\text{H}^{\epsilon*}$ (C), and 18 $\text{H}^{\gamma*}$ (D) protons. Weak NOE between 259 $\text{H}^{\delta*}$ and 20 $\text{H}^{\delta*}$ is highlighted by a black circle in (B). ^{13}C chemical shifts for 259 $\text{C}^{\delta*}$, 259 $\text{C}^{\epsilon*}$, and 18 $\text{C}^{\gamma*}$ are displayed at the top of the NOESY strips in (B)–(D), respectively.

conformational ensemble in comparison to that calculated using our limited unambiguous NOE data.

Y259 Facilitates C2 Proton Exchange in Propanal. The importance of the Y259 $^{\text{OH}}$ group in the *ec*DERA catalytic reaction is highlighted by an ~ 100 -fold reduction in the k_{cat} value reported herein (Table 1) and in previous studies.^{33,34} To understand the role of the Y259 $^{\text{OH}}$ group, we monitored the real-time H/D exchange of the C2 proton of propanal in the presence of DERAm and the Y259F mutant using 1D ^1H NMR in D_2O and KP_i buffer (Figure 5A). Propanal was chosen over acetaldehyde (natural substrate) because the latter can react in a sequential manner to form side products which covalently attach to and inactivate *ec*DERA.³² For monitoring of the C2 deprotonation, the transformation of the methyl group of both propanal and its hydrate, propane-1,1-diol, from a triplet into a doublet was monitored over time. The addition of DERAm leads to an immediate decrease in the intensity of the downfield resonance of the methyl group triplet and a simultaneous increase in the upfield resonance (Figure 5B), signifying exchange of a C2 proton with a deuterium. Fitting the change in the intensity ratio of the upfield and downfield peaks to a monoexponential equation (eq 2) yielded a time constant (τ) of 19.7 ± 2.5 min for DERAm-catalyzed C2 deprotonation (Figure 5C). However, the change in the aforementioned

intensity ratio in the presence of the Y259F mutant was observed to be significantly slower, with the fits yielding a time constant of 62.7 ± 5.9 h (Figure 5D), ~ 190 -fold higher than τ_{DERAm} . This drastic increase in τ indicates that the C2 proton abstraction is significantly more efficient in the presence of the Y259 side chain hydroxyl group than in its absence.

C-Terminal Tail Contains Auxiliary Phosphate-Binding Residues. The catalytic efficiency of *ec*DERA is severely reduced for nonphosphorylated substrates. The k_{cat} value is $68 \pm 1 \text{ s}^{-1}$ with dRSP as the acceptor, whereas it decreases by a factor of ~ 600 to $0.11 \pm 0.01 \text{ s}^{-1}$ when deoxyribose (dR) is used as the acceptor aldehyde.²⁹ A strong increase in K_{M} from 0.64 mM (dRSP) to 57.7 mM (dR) suggests that the phosphate group is vital for the enzyme–substrate interaction. Additionally, under our reaction conditions, we could not detect any catalytic activity for DERAm toward dR.

Considering that the phosphate group is crucial for enzyme–substrate interaction, we investigated the interaction between DERAm and phosphate using 2D ^1H – ^{15}N HSQC spectra recorded at different phosphate concentrations ranging from 0 mM to 100 mM KP_i (Figure 6A). As expected, the main phosphate-binding residues in the vicinity of the active site identified by X-ray crystallography (i.e., T18, L20, T170, G171, G204, G205, and A237)³³ show significant chemical shift

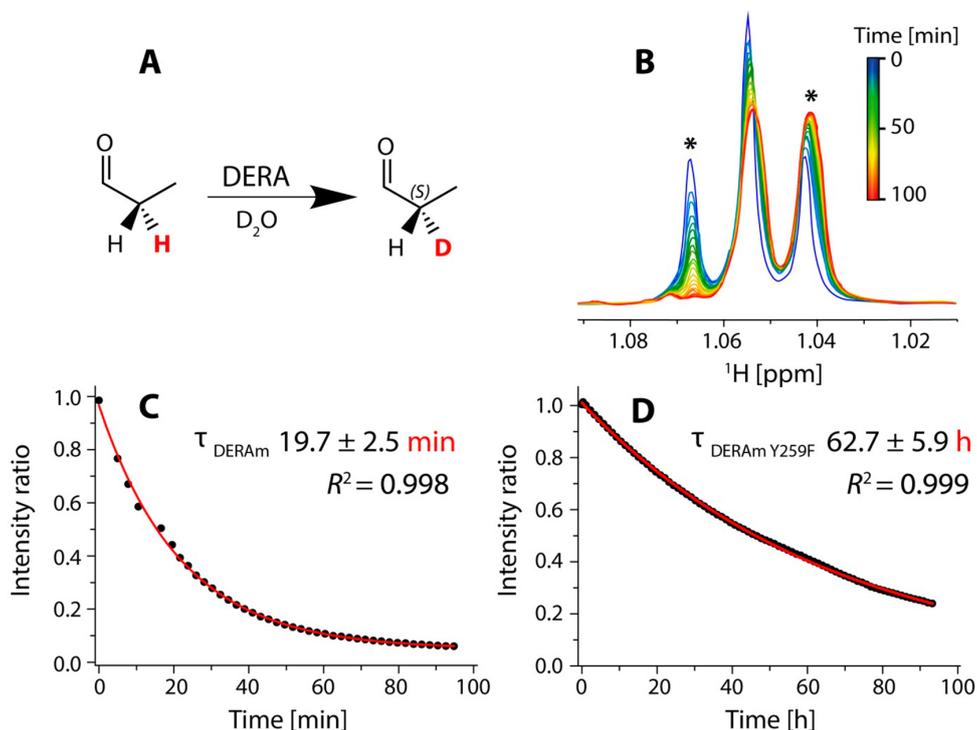


Figure 5. Substitution of the C2-proton of propanal with deuterium in the presence of DERA and the Y259F mutant studied with 1D proton NMR spectroscopy highlights the role of Y259. (A) Schematic of the DERA-catalyzed H/D exchange reaction. The proton undergoing exchange is highlighted in red. (B) Transformation of the propanal methyl group triplet into a doublet over time in the presence of 5 μM DERA. The ratio of the peaks at 1.068 and 1.043 ppm (marked with asterisks in (B)) was fit to a single-exponential equation in the presence of 5 μM DERA (C) and 5 μM Y259F mutant (D), respectively. The errors in (C) and (D) were estimated from triplicate measurements.

changes upon titration of KPi . Residues undergoing fast to intermediate exchange on the NMR time scale were used to extract K_D values (eq 1). The chemical shift changes for 16 residues in DERA (Table S3) were used to calculate an average K_D value of 18.8 ± 3.6 mM for the main phosphate-binding residues. Additionally, several residues on the C-terminal tail showed chemical shift changes upon interaction with phosphate. The average K_D value for these auxiliary phosphate-binding residues was 19.0 ± 0.7 mM ($n = 5$, Table S3), nearly identical with that for the main set of residues. The ^{31}P NMR spectrum, recorded on DERA variant in 50 mM KPi buffer (pH 6.8), shows the presence of one additional phosphate resonance in addition to the buffer resonance. This observation indicates that both binding sites may simultaneously coordinate the same molecule of inorganic phosphate (Figure S9).

To identify the key C-terminal tail auxiliary phosphate-binding residues, we carried out ^1H - ^{15}N HSQC titration of KPi with ^{15}N -labeled DERA mutants: Y259F and S257D/S258D. The affinity toward Pi for the main and auxiliary binding residues in the Y259F mutant were 17.9 ± 2.4 mM ($n = 16$) and 19.3 ± 0.4 mM ($n = 5$), respectively. The near-identical affinities for both sets of binding residues, as well as those to the corresponding binding sites in DERA, indicate that the Y259 $^{\circ}\text{OH}$ group does not play any role in phosphate coordination. This is supported by the presence of a single protein-bound phosphate resonance in the ^{31}P NMR spectrum, which has a chemical shift identical with that of the bound

phosphate resonance in the ^{31}P spectrum of DERA variant (Figure S9).

For the S257D/S258D mutant, 2D ^1H - ^{15}N HSQC spectra were recorded with KPi concentrations ranging from 0 to 150 mM, in anticipation of the decreased phosphate affinity due to the presence of negatively charged aspartic acid residues. Residues Y259 and A256, flanking the mutation site, show a decrease in Pi affinity by more than 200%, whereas residues D258, G253, L241, A237, and T208 show an affinity decrease of more than 100% in comparison to their respective affinities in DERA and the Y259F mutant (Table S3). The remaining residues used for K_D determination located around the active site (i.e., D16, T18, L20, T170, N176, A177, A203, G204, G205, and R207) show a decrease in affinity between 47 and 87% (Table S3). Additionally, a ^{31}P NMR spectrum recorded on the S257D/S258D mutant shows a single protein-bound phosphate resonance which has a distinct chemical shift in comparison with the bound phosphate resonance observed in ^{31}P NMR spectra of DERA and Y259F mutant (Figure S9). This observation indicates that S257 and S258 might be the key phosphate interacting residues on the C-terminal tail and demonstrates that the perturbation of the phosphate affinity of the auxiliary phosphate-binding residues has a direct effect on the affinity of the main phosphate-binding residues.

DISCUSSION

In this report, we show that mutation of the C-terminal Y259 into a phenylalanine leads to a drastic drop in the catalytic activity of the monomer variant of *ecDERA* (i.e., DERA)

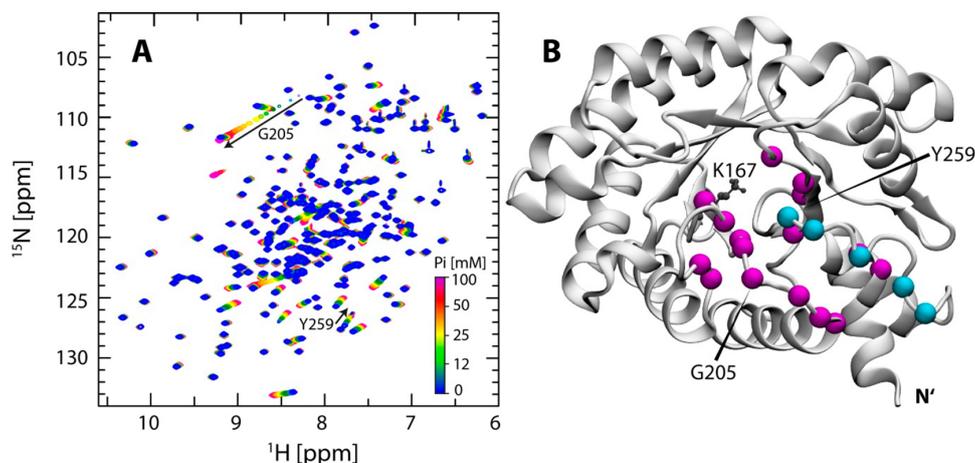


Figure 6. Phosphate-binding sites in DERAm. (A) Overlay of 2D ^1H – ^{15}N HSQC spectra of DERAm recorded in the presence of KPi , ranging from 0 mM to 100 mM. The main and auxiliary phosphate-binding residues show significant changes in backbone amide chemical shift. In addition to chemical shift changes for various residues, extensive changes in line widths are observed. (B) Residues showing significant chemical shift changes upon addition of KPi are highlighted (magenta, main; cyan, auxiliary) on a model of the DERAm closed state. The catalytic K167 side chain is shown in a ball and stick representation.

(Table 1), similar to the effect observed in wild-type *ecDERA*.³³ Such an observation indicates that the $^{\text{O}}\text{H}$ group of Y259 plays a crucial role in the function of *ecDERA*. Despite the presence of significant structural and biochemical work on *ecDERA*,^{29,30,32,33,35,36} precise structural information on how its intrinsically disordered C-terminal tail facilitates the catalytic process has been largely unaddressed. Using NMR chemical shift perturbations, we provide the first evidence demonstrating that the intrinsically disordered C-terminal tail samples conformations which allow Y259 to enter the active site (i.e., closed state), in the absence of the substrate (Figure 2). Such presampling of functionally relevant conformations in the absence of substrates has been observed for several enzymes, highlighting the importance of conformational flexibility for the catalytic process. For example, loops within the structural core of enzymes such as cyclophilin A,⁶ adenylate kinase,^{18,83} and RNase A¹⁶ have the ability to sample catalytically relevant “closed” state conformations in the absence of a substrate, and the frequency of the sampling of such conformations correlates with the catalytic turnover rates. We also attempted to quantify the exchange rate between the open and closed states of the DERAm C-terminal tail. However, even at 5 °C, the exchange process was too fast to obtain reliable quantitative data from NMR CPMG relaxation dispersion experiments (i.e., $k_{\text{ex}} > 10000 \text{ s}^{-1}$). Since DERAm-catalyzed dR5P breakdown proceeds with a k_{cat} value of 19 s^{-1} , fast sampling of the open and closed states, in the substrate-free form, is unlikely to be the rate-limiting step of DERAm catalysis. Experiments to probe the effect of the natural substrate (dR5P) on the C-terminal tail conformational sampling were hampered by the instability of the system. Upon addition of dR5P, chemical shift as well as line width changes in a series of 2D [^1H – ^{15}N] HSQC spectra were observed over time (Figure S10). These time-dependent changes could be caused by acetaldehyde molecules, formed during the reversible aldol reaction, which can react in a sequential manner to form side products³² which may covalently attach to the enzyme.

Despite the conformational plasticity of the C-terminal tail, the population of the closed state was sufficiently large to give

rise to NOEs between the C-terminal tail and DERAm active site as well as helix 10 (Figure 4B–D). Using the unambiguous NOE distance restraints, we were able to derive a structural ensemble of the catalytically relevant closed state. The NMR ensemble concurs with a closed state ensemble generated through the MD simulations. Medium- to weak-intensity NOEs were observed between residues showing hydrophobic interactions in the MD simulations: i.e., Y259–L20, A256–A242, A246–L241, and others (Table S2). From a structural point of view, it appears that the closed state is not ideal for the substrate entry and/or product exit, as Y259 blocks the entrance to the active site pocket (Figure S11). Therefore, a transition between open and closed states may be necessary for substrate and product diffusion, as observed for rabbit muscle FBPA.^{37,84} Additionally, deletion of the C-terminal tail (D252stop mutant) leads to a significant decrease in DERAm activity in comparison to full-length DERAm. Interestingly, the activity of D252stop mutant ($k_{\text{cat}} = 0.48 \text{ s}^{-1}$ and $K_{\text{M}} = 0.12 \text{ mM}$) was moderately higher (2.4-fold increase in k_{cat}) than that of the Y259F mutant. This increase could be attributed to efficient substrate entry and/or product release in the absence of the C-terminal tail (i.e., the open conformation) and indicates that Y259 could be the only C-terminal tail residue participating in the catalytic process.

In spite of the abundance of biochemical data stressing the importance of Y259 in DERA catalysis, the role of this crucial residue remained inconclusive. Studies with *Salmonella typhimurium* DERA, which is 96.5% sequence identical with *ecDERA*, have shown that Y259 does not play a role in facilitating active site Schiff base formation.³⁴ In FBPA, the C-terminal tyrosine residue is proposed to be a general base in the reaction with DHAP, enabling the transformation of the imine to enamine intermediate.^{37,85} However, Heine et al.³³ used a combination of X-ray crystallography and 1D ^1H NMR to rule out the role of Y259 as a general base in the catalytic mechanism of *ecDERA*. Instead, it was proposed that a structurally conserved water molecule participates in a proton relay, together with D102 and K201, responsible for the C2 proton abstraction from the substrate.

Our study shows that the C2 proton of propanal is abstracted by DERAm and the Y259F mutant, similarly to the observations by Heine et al.³³ However, analysis of the real-time propanal H/D exchange performed herein, using a 14-fold lower concentration of the enzyme (relative to Heine et al.), reveals an ~ 190 -fold increase in the time constant for the Y259F mutant catalyzed C2 proton abstraction (Figure 5), making it the rate-limiting step for this mutant. A similar effect was observed for FBPA after hydrolysis of the carboxy-terminal peptide bond through carboxypeptidase treatment.⁸⁵ Therefore, participation of Y259 in the *ec*DERA catalytic mechanism, together with residues D102 and K201, cannot be ruled out. Two putative mechanisms could explain the role that the Y259 $^{\text{OH}}$ group plays in lowering the activation energy ($\Delta\Delta G^{\ddagger}_{\text{Y259F-DERAm}}$ value of ~ 3 kcal mol $^{-1}$, obtained from k_{cat} in Table 1 using the Eyring equation) for the proton-abstraction step. (1) Y259 $^{\text{OH}}$ acts as a general base to directly abstract the C2 proton, as suggested previously for other class I aldolases.^{37,86} In its absence, an alternative mechanism such as the participation of a conserved water molecule could drive the substrate proton abstraction, albeit inefficiently. (2) Y259 $^{\text{OH}}$ participates in the proton-relay system, originally proposed for *ec*DERA,³³ by maintaining a desirable electrostatic environment for fast proton abstraction. The absence of the Y259 $^{\text{OH}}$ group (e.g., in the Y259F and D252stop mutants) perturbs this conducive environment, leading to a much slower proton abstraction.

Overlay of the NMR structure of the DERAm closed state with the crystal structure of *ec*DERA bound to the carbinolamine intermediate (PDB: 1JCL) reveals the presence of a steric clash between Y259 and the substrate (Figure 7). Therefore, after the substrate enters the active site, a reorientation of the side chains of Y259 and other active site residues might be necessary.

Using a combination of chemical shift perturbation analysis and ^{31}P NMR with different DERAm mutants, we have revealed that the C-terminal tail coordinates the active site phosphate moiety through previously unknown auxiliary phosphate-binding residues (Figure 6 and Figure S9) involving amino acids 256–259. For FBPA, a water-mediated coordina-

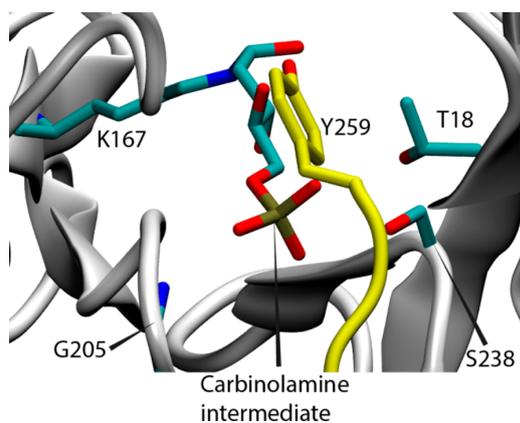


Figure 7. Y259 side chain clashed with the carbinolamine intermediate. An overlay of the NMR DERAm closed state structure (yellow) with the crystal structure of *ec*DERA in complex with the carbinolamine reaction intermediate (PDB: 1JCL) reveals the presence of a steric clash between Y259 and the carbinolamine intermediate.

tion of the C-terminal Y363 with inorganic phosphate and DHAP phosphate was proposed to generate a reactive phenolate ion which carries out stereospecific proton abstraction to generate the enamine intermediate.³⁷ However, for DERAm we observed identical phosphate-binding affinities regardless of the C-terminal residue: i.e., Y259 or F259. Furthermore, the deprotonation of Y259 through a phosphate coordinating water molecule, to generate a phenolate ion, can also be ruled out for DERAm due to the near-identical chemical shifts of the Y259 epsilon carbon (C^{ϵ}) in 50 mM HEPES and 50 mM KPi buffers (Y259 $^{\epsilon}\text{C}$, 118.00 ppm in HEPES buffer;³⁸ and Y259 $^{\epsilon}\text{C}$, 118.04 ppm in KPi buffer). Therefore, the auxiliary phosphate-binding residues might have an alternative function.

The phosphate-binding region of *ec*DERA active site shows the presence of A237, S238, and S239, forming a noncanonical phosphate-binding site.³⁰ Interestingly, the auxiliary phosphate-binding residues form a noncanonical motif (residues A256, S257, and S258) as well. Mutation of S257 and S258 into aspartic acids decreased P_i affinity not only at the C-terminal tail but also in the active site. This demonstrates that the residues in the two noncanonical phosphate-binding regions coordinate the same phosphate molecule located at the active site. Since deletion of the C-terminal tail results in a slight increase (1.9-fold, Table S3) in phosphate affinity in the vicinity of the active site, the auxiliary phosphate-binding residues are not required for the phosphate interaction of the main phosphate-binding residues. Therefore, we hypothesize that these auxiliary residues allow the C-terminal tail to dock onto the phosphate group of the substrate, thereby stabilizing the closed state conformation. Such docking may allow reorientation Y259 in an optimal position, avoiding a steric clash with the substrate and enabling participation in the C2 proton exchange step of the DERA catalytic reaction.

CONCLUSIONS

We have uncovered novel insights into the role of the rarely studied intrinsically disordered C-terminal tail on a structural level and provided a more detailed picture of the *ec*DERA catalytic reaction mechanism. Although a comparison between the static crystal structures of *ec*DERA in the apo- and substrate-bound states has provided insights into understanding the DERA–substrate interactions,^{32,33,35} our data demonstrate that conformational sampling of the C-terminal tail is a key feature for efficient catalysis. The dynamic aspects related to enzyme catalysis are becoming more evident with the development of both experimental (e.g., NMR⁸⁷ and multi-temperature and X-ray-free electron laser crystallography¹⁴) and computational (e.g., MD⁸⁸ and Monte Carlo simulations⁸⁹) methods. The TIM-barrel fold is one of the most common among enzymes, and many recent attempts to design de novo enzymes showed an incredible evolvability of the TIM-barrel scaffolds.^{90–92} The importance of loop dynamics within the TIM-barrel fold family of enzymes, for the catalytic activity, has been highlighted previously.^{93–97} However, to our knowledge, the results provided herein represent the first structural evidence in the TIM-barrel fold family where an intrinsically disordered C-terminal tail, located outside the fold, not only samples catalytically relevant conformations in the absence of a substrate but also participates in the catalytic reaction. Therefore, our current contribution not only provides significant insights related to DERAs specifically but also

deepens the understanding of the involvement of the intrinsic motions in catalysis of the TIM-barrel fold enzymes.

Furthermore, DERAs found in several human pathogens, e.g. *Salmonella typhimurium*, *Klebsiella pneumoniae*, *Vibrio cholerae*, and others (Figure S12), show a high sequence identity (81% for *V. cholerae* and >94% identity for the others presented) with *ecDERA*, unlike human DERA with only 38% identity.⁹⁸ All of these pathogens (except *V. cholerae*) have an identical C-terminal tail sequence. Since DERA plays a crucial role in microbial survival by participating in nucleotide catabolism and providing key metabolic intermediates such as acetaldehyde and G3P, it is considered to be a promising drug target against human pathogens.^{24,25} Considering the high sequence identity between the aforementioned bacterial DERAs, their structural and dynamic properties should be identical with those of *ecDERA*. The insights uncovered herein, together with the closed state structure, should therefore aid in not only the development of better DERA variants for efficient biocatalysis but also small-molecule inhibitors targeting DERAs from human bacterial pathogens.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscatal.7b04408.

Tables S1–S3 and Figures S1–S12 as described in the text (PDF)

File of the DERAm closed state NMR ensemble (PDB)

■ AUTHOR INFORMATION

Corresponding Author

*V.P.: tel,++49 2461 61 9449; e-mail, v.panwalkar@fz-juelich.de.

ORCID

Dušan Petrović: 0000-0002-1834-7358

Jörg Pietruszka: 0000-0002-9819-889X

Dieter Willbold: 0000-0002-0065-7366

Vineet Panwalkar: 0000-0003-1621-1606

Present Address

#D.P.: Department of Cell and Molecular Biology, Uppsala University, BMC Box 596, S-751 24 Uppsala, Sweden.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Professors Tobias Madl, Lynn Kamerlin, and Gunnar Schröder for fruitful discussions and critical reading of the manuscript. The authors acknowledge access to the Jülich-Düsseldorf Biomolecular NMR Center that is jointly run by the Forschungszentrum Jülich and Heinrich-Heine-Universität Düsseldorf and the computing time granted by JARA-HPC Vergabegremium and VSR commission on the supercomputer JURECA (project JICS69 to D.P. and Prof. Birgit Strodel) at Forschungszentrum Jülich. We thank Dr. Thomas Classen for providing the plasmid encoding DERAm. M.S. acknowledges the Ministry of Innovation, Science, and Research of the German federal state of North Rhine-Westphalia and the Heinrich-Heine-Universität Düsseldorf (scholarship within the CLIB-Graduate Cluster Industrial Biotechnology) for financial support. D.P. was financially supported by the Jürgen Manchot Foundation.

■ REFERENCES

- Andrews, L. D.; Fenn, T. D.; Herschlag, D. Ground state destabilization by anionic nucleophiles contributes to the activity of phosphoryl transfer enzymes. *PLoS Biol.* **2013**, *11*, e1001599.
- Zhang, Y.; Schramm, V. L. Ground-state destabilization in orotate phosphoribosyltransferases by binding isotope effects. *Biochemistry* **2011**, *50*, 4813–4818.
- Griswold, W. R.; Castro, J. N.; Fisher, A. J.; Toney, M. D. Ground-state electronic destabilization via hyperconjugation in aspartate aminotransferase. *J. Am. Chem. Soc.* **2012**, *134*, 8436–8438.
- Hammes-Schiffer, S. Catalytic efficiency of enzymes: A theoretical analysis. *Biochemistry* **2013**, *52*, 2012–2020.
- Benkovic, S. J.; Hammes-Schiffer, S. A perspective on enzyme catalysis. *Science* **2003**, *301*, 1196–1202.
- Eisenmesser, E. Z.; Millet, O.; Labeikovsky, W.; Korzhnev, D. M.; Wolf-Watz, M.; Bosco, D. A.; Skalicky, J. J.; Kay, L. E.; Kern, D. Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **2005**, *438*, 117–121.
- Campbell, E.; Kaltenbach, M.; Correy, G. J.; Carr, P. D.; Porebski, B. T.; Livingstone, E. K.; Afriat-Jurnou, L.; Buckle, A. M.; Weik, M.; Hoffelder, F.; Tokuriki, N.; Jackson, C. J. The role of protein dynamics in the evolution of new enzyme function. *Nat. Chem. Biol.* **2016**, *12*, 944–950.
- Luk, L. Y.; Loveridge, E. J.; Allemann, R. K. Protein motions and dynamic effects in enzyme catalysis. *Phys. Chem. Chem. Phys.* **2015**, *17*, 30817–30827.
- Boehr, D. D.; Nussinov, R.; Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **2009**, *5*, 789–796.
- Petrović, D.; Frank, D.; Kamerlin, S. C. L.; Hoffmann, K.; Strodel, B. Shuffling active site substrate populations affects catalytic activity: The case of glucose oxidase. *ACS Catal.* **2017**, *7*, 6188–6197.
- Pabis, A.; Risso, V. A.; Sanchez-Ruiz, J. M.; Kamerlin, S. C. L. Cooperativity and flexibility in enzyme evolution. *Curr. Opin. Struct. Biol.* **2018**, *48*, 83–92.
- Wolfenden, R. Massive thermal acceleration of the emergence of primordial chemistry, the incidence of spontaneous mutation, and the evolution of enzymes. *J. Biol. Chem.* **2014**, *289*, 30198–30204.
- Schramm, V. L. Transition states and transition state analogue interactions with enzymes. *Acc. Chem. Res.* **2015**, *48*, 1032–1039.
- Keedy, D. A.; Kenner, L. R.; Warkentin, M.; Woldeyes, R. A.; Hopkins, J. B.; Thompson, M. C.; Brewster, A. S.; Van Benschoten, A. H.; Baxter, E. L.; Uevirojnangkoorn, M.; McPhillips, S. E.; Song, J.; Alonso-Mori, R.; Holton, J. M.; Weis, W. I.; Brunger, A. T.; Soltis, S. M.; Lemke, H.; Gonzalez, A.; Sauter, N. K.; Cohen, A. E.; van den Bedem, H.; Thorne, R. E.; Fraser, J. S. Mapping the conformational landscape of a dynamic enzyme by multitemperature and XFEL crystallography. *eLife* **2015**, *4*, e07574.
- Otten, R.; Liu, L.; Kenner, L. R.; Clarkson, M. W.; Mavor, D.; Tawfik, D. S.; Kern, D.; Fraser, J. S. Rescue of conformational dynamics in enzyme catalysis by directed evolution. *Nat. Commun.* **2018**, *9*, 1314.
- Beach, H.; Cole, R.; Gill, M. L.; Loria, J. P. Conservation of μ s enzyme motions in the apo- and substrate-mimicked state. *J. Am. Chem. Soc.* **2005**, *127*, 9167–9176.
- Boehr, D. D.; McElheny, D.; Dyson, H. J.; Wright, P. E. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **2006**, *313*, 1638–1642.
- Henzler-Wildman, K. A.; Thai, V.; Lei, M.; Ott, M.; Wolf-Watz, M.; Fenn, T.; Pozharski, E.; Wilson, M. A.; Petsko, G. A.; Karplus, M.; Hubner, C. G.; Kern, D. Intrinsic motions along an enzymatic reaction trajectory. *Nature* **2007**, *450*, 838–844.
- Jackson, C. J.; Foo, J. L.; Tokuriki, N.; Afriat, L.; Carr, P. D.; Kim, H. K.; Schenk, G.; Tawfik, D. S.; Ollis, D. L. Conformational sampling, catalysis, and evolution of the bacterial phosphotriesterase. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 21631–21636.
- Bahar, I.; Chennubhotla, C.; Tobi, D. Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Curr. Opin. Struct. Biol.* **2007**, *17*, 633–640.

- (21) Machajewski, T. D.; Wong, C. H. The catalytic asymmetric aldol reaction. *Angew. Chem., Int. Ed.* **2000**, *39*, 1352–1375.
- (22) Hoffee, P. Deoxyribose-5-phosphate aldolase from *Salmonella typhimurium*. *Methods Enzymol.* **1975**, *42*, 276–279.
- (23) Tozzi, M. G.; Camici, M.; Mascia, L.; Sgarrella, F.; Ipata, P. L. Pentose phosphates in nucleoside interconversion and catabolism. *FEBS J.* **2006**, *273*, 1089–1101.
- (24) Cao, T. P.; Kim, J. S.; Woo, M. H.; Choi, J. M.; Jun, Y.; Lee, K. H.; Lee, S. H. Structural insight for substrate tolerance to 2-deoxyribose-5-phosphate aldolase from the pathogen *Streptococcus suis*. *J. Microbiol.* **2016**, *54*, 311–321.
- (25) Tonkin, M. L.; Halavaty, A. S.; Ramaswamy, R.; Ruan, J.; Igarashi, M.; Ngo, H. M.; Boulanger, M. J. Structural and functional divergence of the aldolase fold in *Toxoplasma gondii*. *J. Mol. Biol.* **2015**, *427*, 840–852.
- (26) Ošljaj, M.; Cluzeau, J.; Orkić, D.; Kopitar, G.; Mrak, P.; Časar, Z. A highly productive, whole-cell DERA chemoenzymatic process for production of key lactonized side-chain intermediates in statin synthesis. *PLoS One* **2013**, *8*, e62250.
- (27) Jennewein, S.; Schürmann, M.; Wolberg, M.; Hilker, I.; Luiten, R.; Wubbolts, M.; Mink, D. Directed evolution of an industrial biocatalyst: 2-deoxy-D-ribose-5-phosphate aldolase. *Biotechnol. J.* **2006**, *1*, 537–548.
- (28) Greenberg, W. A.; Varvak, A.; Hanson, S. R.; Wong, K.; Huang, H.; Chen, P.; Burk, M. J. Development of an efficient, scalable, aldolase-catalyzed process for enantioselective synthesis of statin intermediates. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 5788–5793.
- (29) DeSantis, G.; Liu, J.; Clark, D. P.; Heine, A.; Wilson, I. A.; Wong, C. H. Structure-based mutagenesis approaches toward expanding the substrate specificity of D-2-deoxyribose-5-phosphate aldolase. *Bioorg. Med. Chem.* **2003**, *11*, 43–52.
- (30) Ma, H.; Szeler, K.; Kamerlin, S. C.; Widersten, M. Linking coupled motions and entropic effects to the catalytic activity of 2-deoxyribose-5-phosphate aldolase (DERA). *Chem. Sci.* **2016**, *7*, 1415–1421.
- (31) Britton, J.; Meneghini, L. M.; Raston, C. L.; Weiss, G. A. Accelerating enzyme catalysis using vortex fluidics. *Angew. Chem., Int. Ed.* **2016**, *55*, 11387–11391.
- (32) Dick, M.; Hartmann, R.; Weiergräber, O. H.; Bisterfeld, C.; Classen, T.; Schwarten, M.; Neudecker, P.; Willbold, D.; Pietruszka, J. Mechanism-based inhibition of an aldolase at high concentrations of its natural substrate acetaldehyde: structural insights and protective strategies. *Chem. Sci.* **2016**, *7*, 4492–4502.
- (33) Heine, A.; DeSantis, G.; Luz, J. G.; Mitchell, M.; Wong, C.-H.; Wilson, I. A. Observation of covalent intermediates in an enzyme mechanism at atomic resolution. *Science* **2001**, *294*, 369–374.
- (34) Hoffee, P.; Snyder, P.; Sushak, C.; Jargiello, P. Deoxyribose-5-P aldolase: Subunit structure and composition of active site lysine region. *Arch. Biochem. Biophys.* **1974**, *164*, 736–742.
- (35) Heine, A.; Luz, J. G.; Wong, C. H.; Wilson, I. A. Analysis of the class I aldolase binding site architecture based on the crystal structure of 2-deoxyribose-5-phosphate aldolase at 0.99 Å resolution. *J. Mol. Biol.* **2004**, *343*, 1019–1034.
- (36) Dick, M.; Weiergräber, O. H.; Classen, T.; Bisterfeld, C.; Bramski, J.; Gohlke, H.; Pietruszka, J. Trading off stability against activity in extremophilic aldolases. *Sci. Rep.* **2016**, *6*, 17908.
- (37) St-Jean, M.; Sygusch, J. Stereospecific proton transfer by a mobile catalyst in mammalian fructose-1,6-bisphosphate aldolase. *J. Biol. Chem.* **2007**, *282*, 31028–31037.
- (38) Schulte, M.; Stoldt, M.; Neudecker, P.; Pietruszka, J.; Willbold, D.; Panwalkar, V. ¹H, ¹³C, and ¹⁵N backbone and sidechain resonance assignments of a monomeric variant of *E. coli* deoxyribose-5-phosphate aldolase. *Biomol. NMR Assignments* **2017**, *11*, 197–201.
- (39) Nicholas, P. C. Effects of buffer concentrations, pH, triose phosphate isomerase and glycerol-3-phosphate dehydrogenase on the measurement of the activity of aldolase and glyceraldehyde-3-phosphate dehydrogenase. *Biochem. Soc. Trans.* **1988**, *16*, 752–753.
- (40) Findeisen, M.; Brand, T.; Berger, S. A ¹H-NMR thermometer suitable for cryoprobes. *Magn. Reson. Chem.* **2007**, *45*, 175–178.
- (41) Sattler, M.; Schleucher, J.; Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 93–158.
- (42) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **1995**, *6*, 277–293.
- (43) Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins: Struct., Funct., Genet.* **2005**, *59*, 687–696.
- (44) Pintacuda, G.; Otting, G. Identification of protein surfaces by NMR measurements with a paramagnetic Gd (III) chelate. *J. Am. Chem. Soc.* **2002**, *124*, 372–373.
- (45) Guex, N.; Peitsch, M. C. SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. *Electrophoresis* **1997**, *18*, 2714–2723.
- (46) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Marius Clore, G. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **2003**, *160*, 65–73.
- (47) Brünger, A. T. *X-PLOR: version 3.1: A system for X-ray crystallography and NMR*; Yale University Press: New Haven, CT, USA, 1992.
- (48) Powell, M. J. D. Restart procedures for the conjugate gradient method. *Math. Program.* **1977**, *12*, 241–254.
- (49) Neudecker, P.; Nerkamp, J.; Eisenmann, A.; Nourse, A.; Lauber, T.; Schweimer, K.; Lehmann, K.; Schwarzing, S.; Ferreira, F.; Rosch, P. Solution structure, dynamics, and hydrodynamics of the calcium-bound cross-reactive birch pollen allergen Bet v 4 reveal a canonical monomeric two EF-hand assembly with a regulatory function. *J. Mol. Biol.* **2004**, *336*, 1141–1157.
- (50) Kharrat, A.; Macias, M. J.; Gibson, T. J.; Nilges, M.; Pastore, A. Structure of the dsRNA binding domain of *E. coli* RNase III. *EMBO J.* **1995**, *14*, 3572–3584.
- (51) Holak, T. A.; Nilges, M.; Oschkinat, H. Improved strategies for the determination of protein structures from NMR data: The solution structure of acyl carrier protein. *FEBS Lett.* **1989**, *242*, 218–224.
- (52) Neudecker, P.; Schweimer, K.; Nerkamp, J.; Scheurer, S.; Vieths, S.; Sticht, H.; Rosch, P. Allergic cross-reactivity made visible: Solution structure of the major cherry allergen Pru av 1. *J. Biol. Chem.* **2001**, *276*, 22756–22763.
- (53) Kuszewski, J.; Clore, G. M. Sources of and solutions to problems in the refinement of protein NMR structures against torsion angle potentials of mean force. *J. Magn. Reson.* **2000**, *146*, 249–254.
- (54) Neudecker, P.; Sticht, H.; Rosch, P. Improving the efficiency of the Gaussian conformational database potential for the refinement of protein and nucleic acid structures. *J. Biomol. NMR* **2001**, *21*, 373.
- (55) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.
- (56) Grzesiek, S.; Stahl, S. J.; Wingfield, P. T.; Bax, A. The CD4 determinant for downregulation by HIV-1 Nef directly binds to Nef. Mapping of the Nef binding surface by NMR. *Biochemistry* **1996**, *35*, 10256–10261.
- (57) Šali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779–815.
- (58) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
- (59) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854.
- (60) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 1950–1958.

- (61) Best, R. B.; Hummer, G. Optimised molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J. Phys. Chem. B* **2009**, *113*, 9004–9015.
- (62) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 712–725.
- (63) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (64) Olsson, M. H. M.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.
- (65) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (66) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (67) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N.log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (68) Hess, B. P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.
- (69) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961–1972.
- (70) Bussi, G. Hamiltonian replica exchange in GROMACS: A flexible implementation. *Mol. Phys.* **2014**, *112*, 379–384.
- (71) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (72) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernandez, C. X.; Schwantes, C. R.; Wang, L. P.; Lane, T. J.; Pande, V. S. MDTraj: A modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* **2015**, *109*, 1528–1532.
- (73) Baker, E. N.; Hubbard, R. E. Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* **1984**, *44*, 97–179.
- (74) Jiao, X.-C.; Pan, J.; Xu, G.-C.; Kong, X.-D.; Chen, Q.; Zhang, Z.-J.; Xu, J.-H. Efficient synthesis of a statin precursor in high space-time yield by a new aldehyde-tolerant aldolase identified from *Lactobacillus brevis*. *Catal. Sci. Technol.* **2015**, *5*, 4048.
- (75) Wurm, J. P.; Holdermann, I.; Overbeck, J. H.; Mayer, P. H.; Sprangers, R. Changes in conformational equilibria regulate the activity of the Dcp2 decapping enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 6034–6039.
- (76) Papaleo, E.; Saladino, G.; Lambrugh, M.; Lindorff-Larsen, K.; Gervasio, F. L.; Nussinov, R. The role of protein loops and linkers in conformational dynamics and allostery. *Chem. Rev.* **2016**, *116*, 6391–6423.
- (77) Hocking, H. G.; Zangger, K.; Madl, T. Studying the structure and dynamics of biomolecules by using soluble paramagnetic probes. *ChemPhysChem* **2013**, *14*, 3082–3094.
- (78) Madl, T.; Güttler, T.; Görlich, D.; Sattler, M. Structural analysis of large protein complexes using solvent paramagnetic relaxation enhancements. *Angew. Chem., Int. Ed.* **2011**, *50*, 3993–3997.
- (79) Hartmüller, C.; Günther, J. C.; Wolter, A. C.; Wöhnert, J.; Sattler, M.; Madl, T. RNA structure refinement using NMR solvent accessibility data. *Sci. Rep.* **2017**, *7*, 5393.
- (80) Emmanouilidis, L.; Schütz, U.; Tripsianes, K.; Madl, T.; Radke, J.; Rucktäschel, R.; Wilmanns, M.; Schliebs, W.; Erdmann, R.; Sattler, M. Allosteric modulation of peroxisomal membrane protein recognition by farnesylation of the peroxisomal import receptor PEX19. *Nat. Commun.* **2017**, *8*, 14635.
- (81) Sun, Y.; Friedman, J. I.; Stivers, J. T. Cosolute paramagnetic relaxation enhancements detect transient conformations of human uracil DNA glycosylase (hUNG). *Biochemistry* **2011**, *50*, 10724–10731.
- (82) Dyson, H. J.; Wright, P. E. Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* **2004**, *104*, 3607–3622.
- (83) Wolf-Watz, M.; Thai, V.; Henzler-Wildman, K.; Hadjipavlou, G.; Eisenmesser, E. Z.; Kern, D. Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nat. Struct. Mol. Biol.* **2004**, *11*, 945–949.
- (84) Rago, F.; Saltzberg, D.; Allen, K. N.; Tolan, D. R. Enzyme substrate specificity conferred by distinct conformational pathways. *J. Am. Chem. Soc.* **2015**, *137*, 13876–13886.
- (85) Rose, I. A.; O'Connell, E. L.; Mehler, A. H. Mechanism of the aldolase reaction. *J. Biol. Chem.* **1965**, *240*, 1758–1765.
- (86) Zhu, X.; Tanaka, F.; Lerner, R. A.; Barbas, C. F., III; Wilson, I. A. Direct observation of an enamine intermediate in amine catalysis. *J. Am. Chem. Soc.* **2009**, *131*, 18206–18207.
- (87) Palmer, A. G., III Enzyme dynamics from NMR spectroscopy. *Acc. Chem. Res.* **2015**, *48*, 457–465.
- (88) Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. Computational enzyme design. *Angew. Chem., Int. Ed.* **2013**, *52*, 5700–5725.
- (89) Mandell, D. J.; Kortemme, T. Backbone flexibility in computational protein design. *Curr. Opin. Biotechnol.* **2009**, *20*, 420–428.
- (90) Sterner, R.; Höcker, B. Catalytic versatility, stability, and evolution of the ($\beta\alpha$) 8-barrel enzyme fold. *Chem. Rev.* **2005**, *105*, 4038–4055.
- (91) Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. Kemp elimination catalysts by computational enzyme design. *Nature* **2008**, *453*, 190–195.
- (92) Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; Clair, J. L. S.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* **2010**, *329*, 309–313.
- (93) Massi, F.; Wang, C.; Palmer, A. G. Solution NMR and computer simulation studies of active site loop motion in triosephosphate isomerase. *Biochemistry* **2006**, *45*, 10787–10794.
- (94) Williams, J. C.; McDermott, A. E. Dynamics of the flexible loop of triose-phosphate isomerase: The loop motion is not ligand gated. *Biochemistry* **1995**, *34*, 8309–8319.
- (95) Guallar, V.; Jacobson, M.; McDermott, A.; Friesner, R. A. Computational modeling of the catalytic reaction in triosephosphate isomerase. *J. Mol. Biol.* **2004**, *337*, 227–239.
- (96) Katebi, A. R.; Jernigan, R. L. The critical role of the loops of triosephosphate isomerase for its oligomerization, dynamics, and functionality. *Protein Sci.* **2014**, *23*, 213–228.
- (97) Joseph, D.; Petsko, G.; Karplus, M. Anatomy of a conformational change: Hinged “lid” motion of the triosephosphate isomerase loop. *Science* **1990**, *249*, 1425–1428.
- (98) Salleron, L.; Magistrelli, G.; Mary, C.; Fischer, N.; Bairoch, A.; Lane, L. DERA is the human deoxyribose phosphate aldolase and is involved in stress response. *Biochim. Biophys. Acta, Mol. Cell Res.* **2014**, *1843*, 2913.

Supporting Information

Conformational Sampling of the Intrinsically Disordered C-terminal Tail of DERA is Important for Enzyme Catalysis

Marianne Schulte,^{†,‡} Dušan Petrović,^{‡,#} Philipp Neudecker,^{†,‡} Rudolf Hartmann,[‡] Jörg Pietruszka,^{‡,¶} Sabine Willbold,[§] Dieter Willbold,^{†,‡} Vineet Panwalkar^{†,‡,*}

[†] Institut für Physikalische Biologie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany

[‡] Institute of Complex Systems 6 (ICS-6): Structural Biochemistry, Forschungszentrum Jülich, 52425 Jülich, Germany

[‡] Institute of Bioorganic Chemistry, Heinrich-Heine-Universität im Forschungszentrum Jülich, 52425 Jülich, Germany

[¶] Institute of Bio- and Geosciences 1 (IBG-1): Biotechnology, Forschungszentrum Jülich, 52425 Jülich, Germany

[§] Central Institute of Engineering, Electronics and Analytics (ZEA-3), Forschungszentrum Jülich, 52425 Jülich, Germany

Present address

[#] D.P.: Department of Cell and Molecular Biology, Uppsala University, BMC Box 596, S-751 24 Uppsala, Sweden

*Corresponding author: v.panwalkar@fz-juelich.de

Table S1: Primers sequences used to generate DERAm mutants with the mutation site underlined

Construct	Primers 5' – 3' sequence
Y259F_fwd	GAGCGCCAGCAGCT <u>TT</u> CTAACTCGAGGATG
Y259F_rws	CATCCTCGAGTTAGAA <u>G</u> CTGCTGGCGCTC
D252stop_fwd	CTGAAAGCGCTGGGTCACGGCT <u>AA</u> GGTAAGAGCGCCAGCAGCTAC
D252stop_rws	GTAGCTGCTGGCGCTCTTACCT <u>AG</u> CCGTGACCCAGCGCTTTCAG
S257D/S258D_fwd	GCGACGGTAAGAGCGCC <u>GACG</u> ACTACTAACTCGAGGATG
S257D/S258D_rws	CATCCTCGAGTTAGTAG <u>TCG</u> TCGGCGCTCTTACCGTCGC

Table S2: NOE distance restraints used for calculation of DERAm closed state conformation

NOE		
Atom A	Atom B	Intensity
Y259 H ^{δ*}	T18 H ^N	Weak
Y259 H ^{δ*}	T18 H ^β	Weak
Y259 H ^{δ*}	T18 H ^{γ2*}	Weak
Y259 H ^{δ*}	L20 H ^{δ1*}	Weak
Y259 H ^{δ*}	A237 H ^{β*}	Weak
Y259 H ^{ε*}	T18 H ^N	Weak
Y259 H ^{ε*}	T18 H ^β	Weak
Y259 H ^{ε*}	T18 H ^{γ2*}	Weak
Y259 H ^{ε*}	L20 H ^{δ1*}	Weak
Y259 H ^{ε*}	L20 H ^{δ2*}	Weak
Y259 H ^{ε*}	A203 H ^α	Weak
Y259 H ^{ε*}	A203 H ^{β*}	Weak
Y259 H ^{ε*}	A237 H ^N	Weak
Y259 H ^{ε*}	A237 H ^{β*}	Weak
A256 H ^α	L240 H ^α	Weak
A256 H ^α	A242 H ^{β*}	Medium
G253 H ^N	A242 H ^{β*}	Medium
G253 H ^N	K246 H ^{γ*}	Weak
G253 H ^N	K246 H ^{ε*}	Weak
G253 H ^{α1}	A242 H ^{β*}	Medium
G253 H ^{α1}	S243 H ^N	Weak
G253 H ^{α1}	K246 H ^N	Weak
G253 H ^{α1}	K246 H ^{β1}	Weak
G253 H ^{α1}	K246 H ^{β2}	Weak
G253 H ^{α2}	A242 H ^{β*}	Medium
G253 H ^{α2}	S243 H ^N	Weak
G253 H ^{α2}	K246 H ^N	Weak
G253 H ^{α2}	K246 H ^{β1}	Weak
G253 H ^{α2}	K246 H ^{β2}	Weak
D252 H ^α	K246 H ^{γ*}	Medium
G251 H ^N	L245 H ^{δ2*}	Weak
G251 H ^N	K246 H ^{δ*}	Weak
G251 H ^{α*}	L245 H ^{δ2*}	Medium
H250 H ^{δ2}	L248 H ^N	Weak
H250 H ^{δ2}	L248 H ^{β1}	Medium
H250 H ^{δ2}	L248 H ^{δ*}	Medium
K246 H ^{ε*}	A242 H ^{β*}	Weak
K246 H ^{ε*}	G253 H ^N	Weak
L245 H ^N	H250 H ^{β2}	Weak
A242 H ^{β*}	G253 H ^N	Weak
A242 H ^{β*}	G253 H ^{α1}	Weak
A242 H ^{β*}	K254 H ^α	Weak
A242 H ^{β*}	S255 H ^α	Weak
A242 H ^{β*}	S255 H ^{β2}	Weak
A237 H ^α	Y259 H ^{δ*}	Medium
A237 H ^α	Y259 H ^{ε*}	Weak
G171 H ^{α1}	Y259 H ^{δ*}	Medium
G171 H ^{α2}	Y259 H ^{ε*}	Medium

Table S3: Residue-specific K_D values, towards inorganic phosphate, for DERAm variants derived from NMR

Residue number	K_D [mM]			
	DERAm	Y259F	S257D/S258D	D252stop
16	16.5 ± 0.7 ^a	17.9 ± 0.8	26.8 ± 1.2	9.5 ± 0.5
18	16.1 ± 0.8	15.4 ± 0.7	29.4 ± 1.3	8.5 ± 0.4
20	18.7 ± 0.9	19.4 ± 0.9	35.2 ± 1.7	10.2 ± 0.5
170	9.9 ± 0.5	11.1 ± 0.5	14.1 ± 0.6	6.3 ± 0.3
171	23.7 ± 1.3	19.8 ± 1.0	50.6 ± 2.8	12.9 ± 0.7
176	15.8 ± 0.8	16.4 ± 0.7	23.4 ± 1.1	8.9 ± 0.4
177	17.4 ± 0.9	17.3 ± 0.8	29.0 ± 1.5	9.7 ± 0.5
203	19.6 ± 1.0	20.4 ± 1.0	38.7 ± 2.1	11.4 ± 0.6
204	23.0 ± 1.3	17.7 ± 0.8	37.9 ± 2.0	9.8 ± 0.6
205	18.2 ± 0.9	17.7 ± 0.8	31.1 ± 1.6	9.7 ± 0.5
206 ^b	-	-	36.5 ± 2.7	-
207	18.5 ± 0.9	18.2 ± 0.9	29.6 ± 1.5	10.1 ± 0.5
208	20.4 ± 1.1	20.2 ± 1.0	42.8 ± 2.2	10.9 ± 0.5
209	25.6 ± 1.5	17.0 ± 0.7	28.3 ± 1.4	9.4 ± 0.5
237	21.1 ± 1.2	21.5 ± 1.1	49.9 ± 2.8	12.3 ± 0.6
241	20.6 ± 1.1	19.2 ± 1.0	41.5 ± 2.4	11.8 ± 1.0
242	16.4 ± 1.4	17.0 ± 1.0	30.1 ± 1.6	8.1 ± 0.4
252	18.9 ± 1.1	19.6 ± 0.9	37.1 ± 1.9	-
253	18.4 ± 1.0	18.5 ± 0.9	40.0 ± 1.9	-
256	18.5 ± 0.9	19.4 ± 0.9	58.7 ± 3.1	-
257 ^b	-	-	45.0 ± 2.3	-
258	20.3 ± 1.3	19.5 ± 0.9	52.5 ± 3.0	-
259	18.9 ± 1.0	19.5 ± 0.9	61.7 ± 3.6	-
Average	18.9 ± 3.2 ^c	18.2 ± 2.2	37.8 ± 11.3	10.0 ± 1.6

^a Fit error.

^b K_D values could not be obtained for DERAm and the Y259F mutant due to resonance overlap.

^c Standard deviation

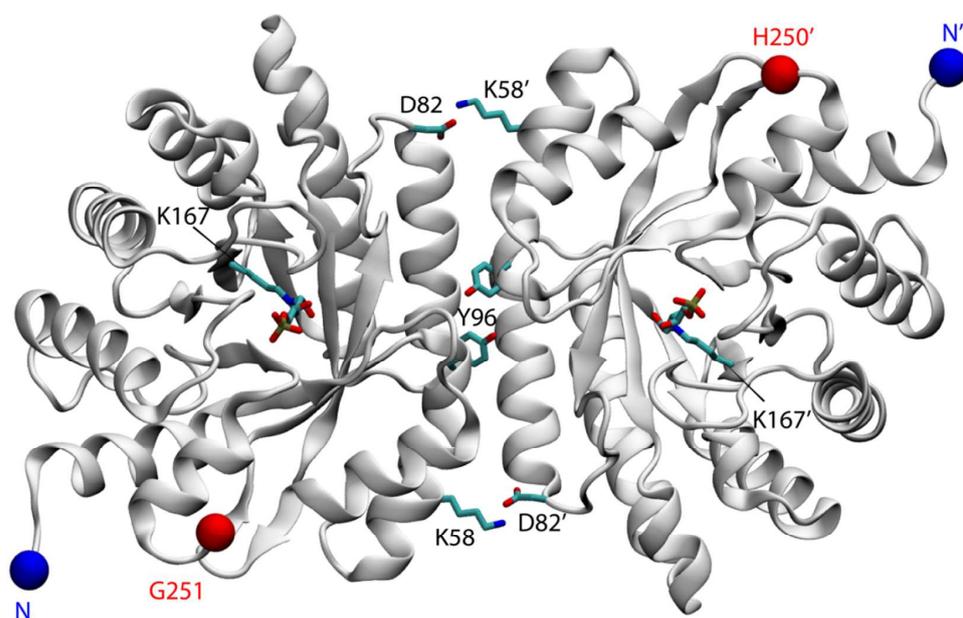


Figure S1: The crystal structure of the wild-type dimeric *ecDERA* with the carbinolamine intermediate.¹ Residues K58 and Y96, located at the dimer interface, were mutated (K58E and Y96W) to generate the monomeric variant DERAm. The side chains of K58, Y96 and the Schiff-base forming K167 are labelled.

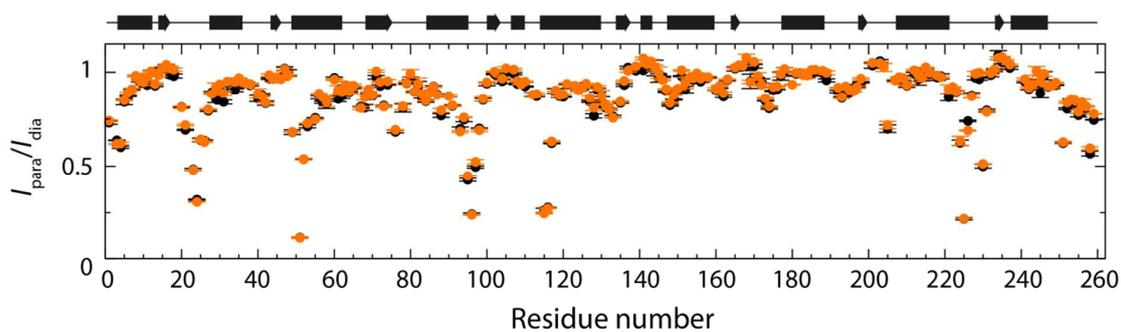


Figure S2: Comparison of solvent accessibility for DERAm and the Y259F mutant using sPRE agent. The intensity ratio of the backbone amide resonances of DERAm (black) and the Y259F mutant (orange) in the presence and absence of 1 mM soluble paramagnetic agent: [Gd(DTPA-DMA)]. The errors were obtained through duplicate measurements. A schematic representation of the secondary structure elements is shown at the top.

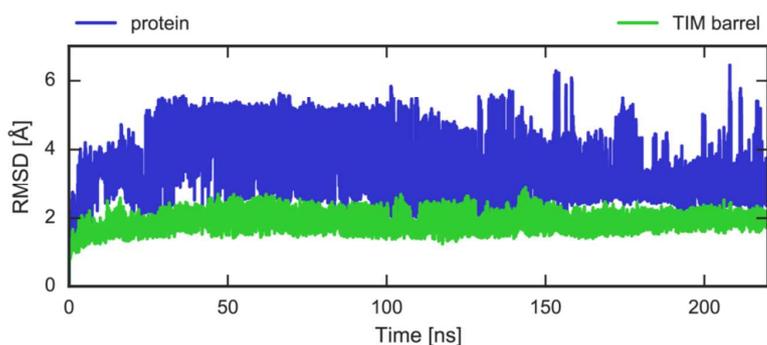


Figure S3: DERA backbone RMSD from the HREX-MD simulation. While the TIM (α/β)₈ barrel is stable during the entire simulation (green), the whole protein appears very flexible (blue) due to the C-terminal flexibility.

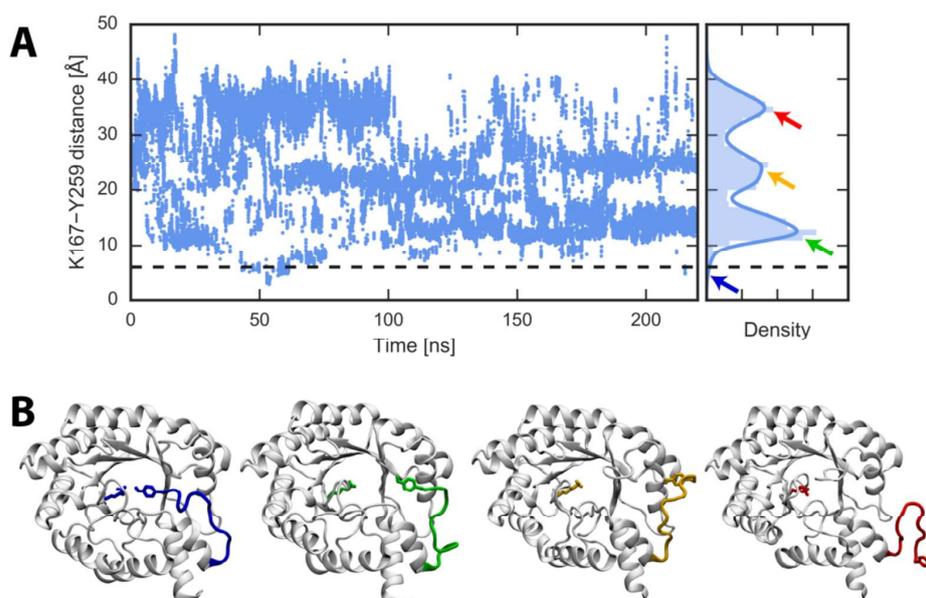


Figure S4: HREX-MD sampling of the C-terminal tail conformational ensemble. (A) The conformational sampling as a function of the distance between the active site (represented by the K167 ¹⁵N-atom) and C-terminus (represented by the Y259 ¹⁸O-atom). Dashed lines show the 6 Å cutoff, below which we assume that C-terminus is in the closed state. (B) Four representative structures corresponding to the closed state (blue) and several open states indicated by the arrows in (A). The catalytic K167 and the C-terminal tail are highlighted in (B).

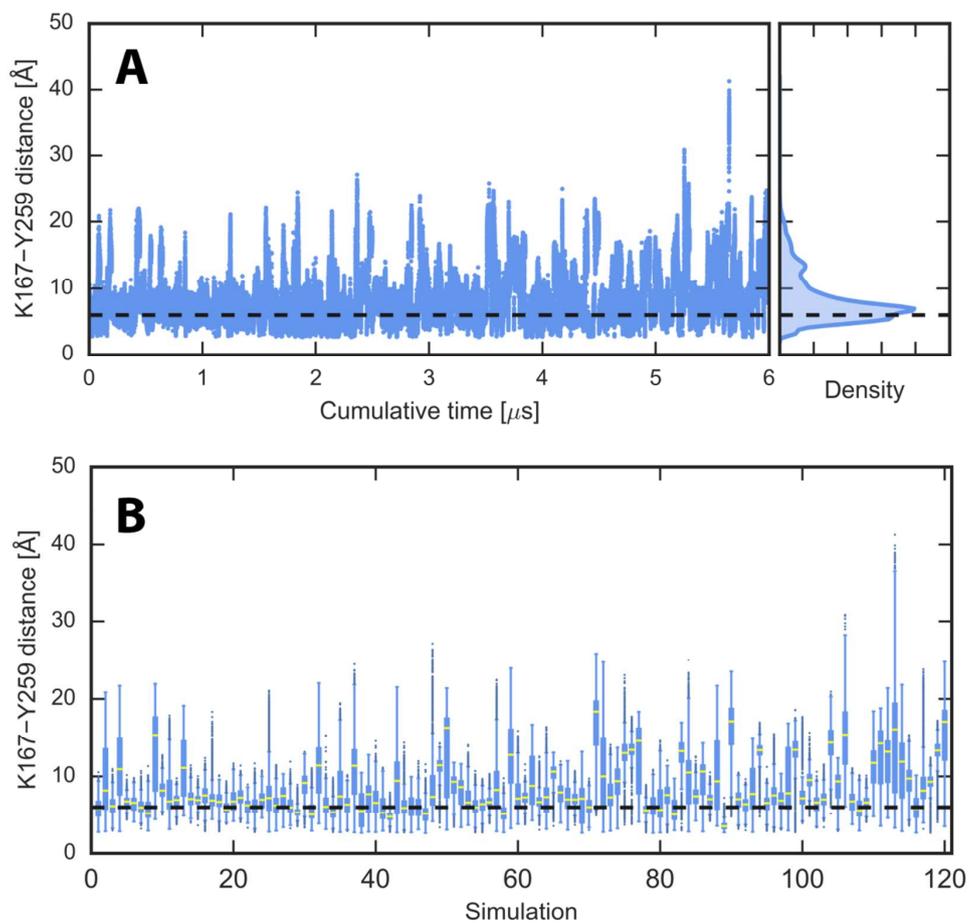


Figure S5: MD sampling of the C-terminal tail conformational ensemble starting from the closed state. (A) Distance distribution from all 120 MD simulations and (B) from each simulation separately. The dashed lines indicate the 6 Å cutoff. Panel (B) shows the K167-Y259F distance distribution as a box plot. In every simulation conformations are sampled in which the K167-Y259 distance is above 6 Å, demonstrating that within every 50 ns long simulation Y259 exits the active site pocket. Such behavior further illustrates the transient nature of the closed state.

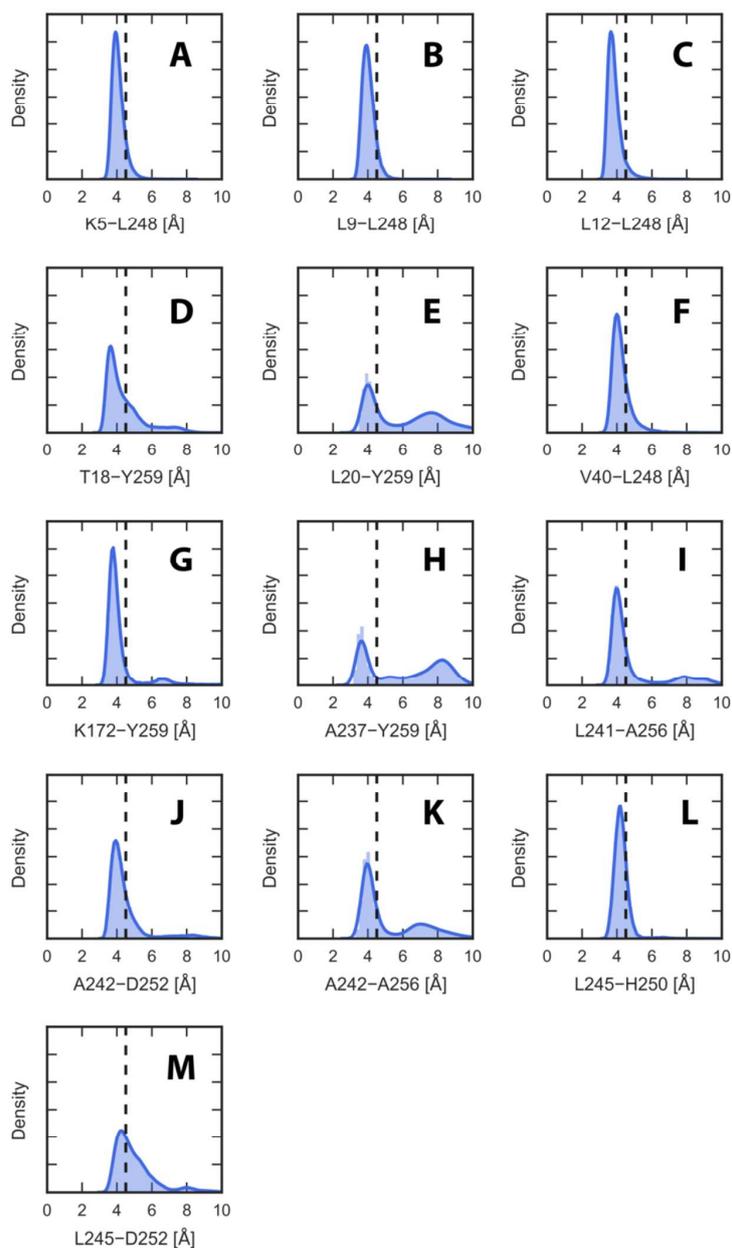


Figure S6: Hydrophobic contacts in the closed ensemble (K167-Y259 distance < 6 Å) from combined 6 μ s DERAm MD simulations. The minimal distance between the residue pairs (any hydrophobic C-atom contacts) is shown as density distribution, and the cutoff of 4.5 Å is shown as dashed line. (A) K5-L248, (B) L9-L248, (C) L12-L248, (D) T18-Y259, (E) L20-Y259, (F) V40-L248, (G) K172-Y259, (H) A237-Y259, (I) L241-A256, (J) A242-D252, (K) A242-A256, (L) L245-H250, (M) L245-D252.

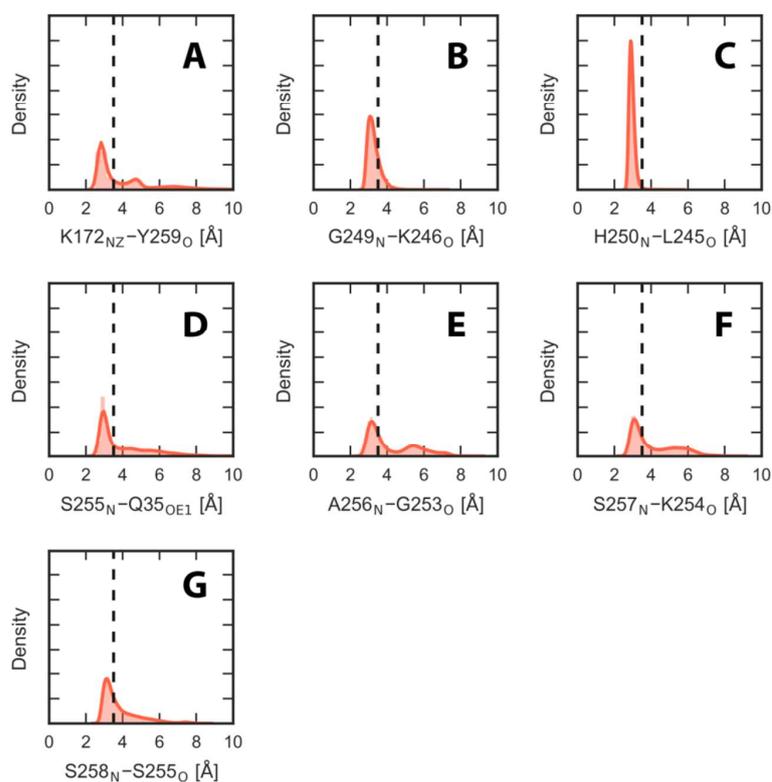


Figure S7: Polar interactions (salt bridges and H-bonds) in the closed ensemble (K167-Y259 distance < 6 Å) from combined 6 μ s DERAM MD simulations. Distance distribution between the donor and acceptor is shown as density distribution, and the cutoff of 3.5 Å is shown as the dashed line. Salt bridge (A) K172-Y259, and H-bonds (B) G249-K246, (C) H250-L245, (D) S255-Q35, (E) A256-G253, (F) S257-K254, (G) S258-S255.

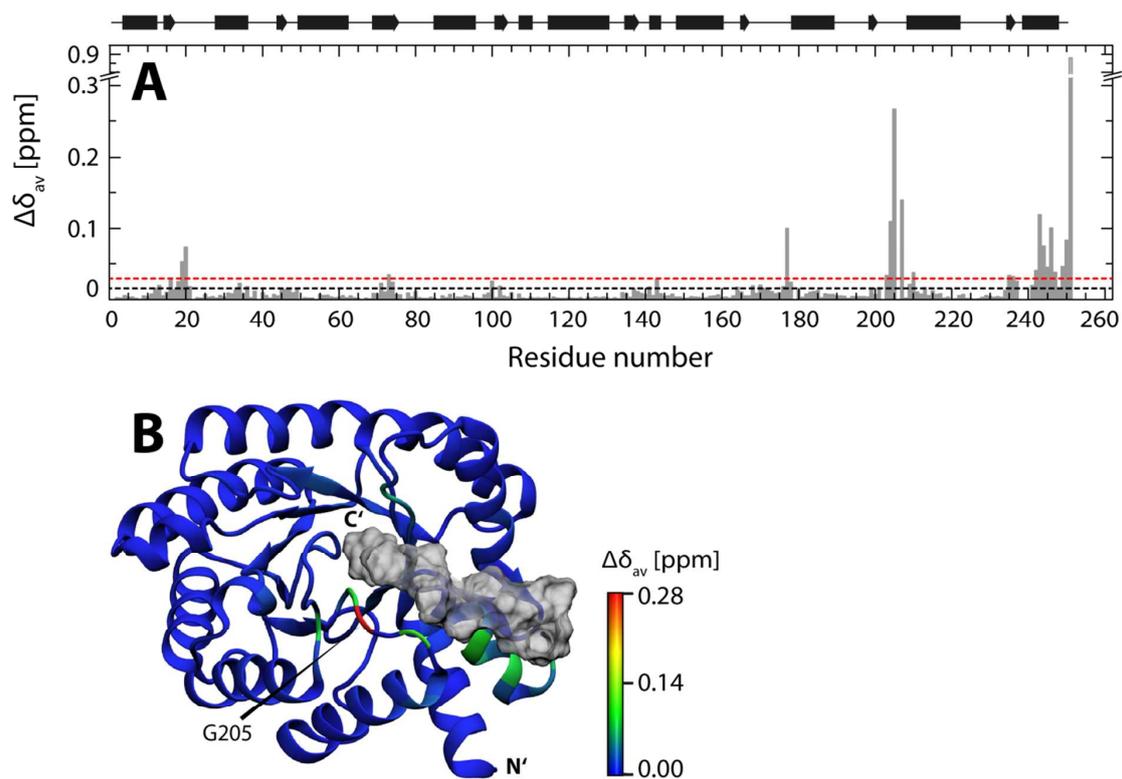


Figure S8: Changes in backbone amide chemical shifts of DERAμ due to deletion of the C-terminal tail. (A) Weighted ^1H - ^{15}N chemical shift differences between DERAμ and the D252stop mutant. The black dotted line represents the average $\Delta\delta_{av}$ value whereas the red dotted line represents the average $\Delta\delta_{av}$ plus one S.D. (B) The $\Delta\delta_{av}$ are mapped onto the DERAμ structure using the coloring scheme shown. The C-terminal tail is represented in grey. A schematic representation of the secondary structure elements is shown at the top.

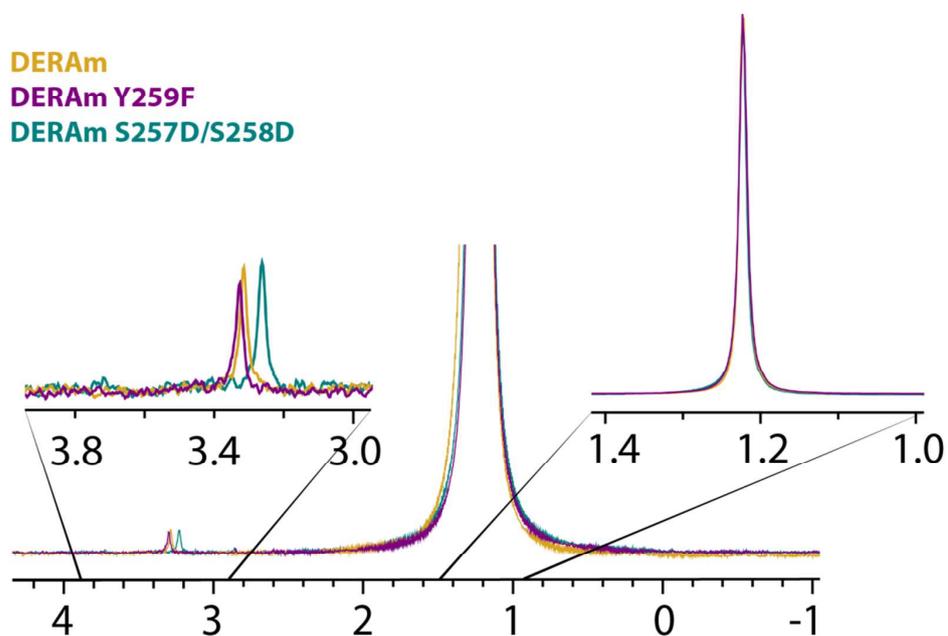


Figure S9: ³¹P NMR spectra of DERAm variants. Overlay of 1D ³¹P NMR spectra of three DERAm variants recorded in 50 mM KPi (pH 6.8) at 25 °C and 600 MHz. The spectra were referenced to the buffer phosphate. A single phosphate resonance at ~3.4 ppm other than the buffer resonance (1.23 ppm) was observed which represents the protein-bound phosphate.

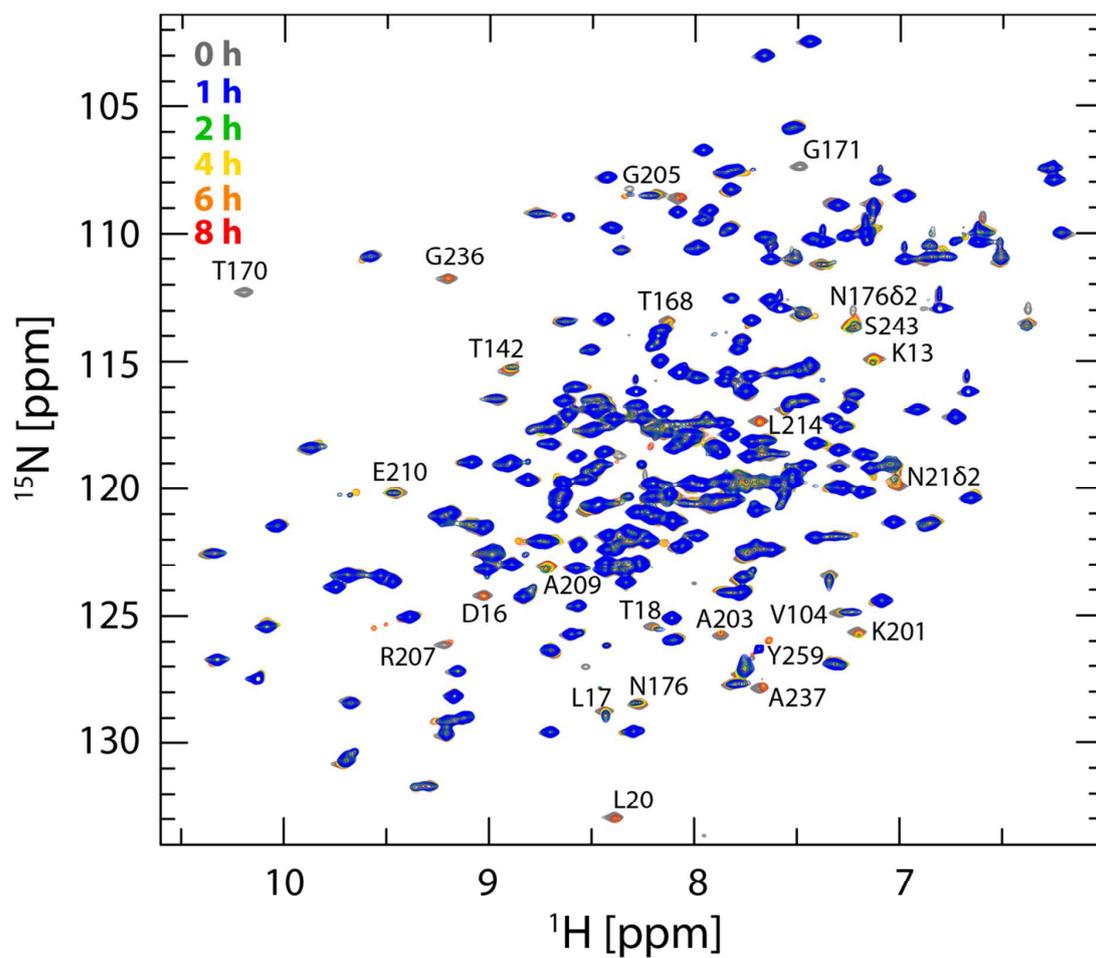


Figure S10: Overlay of a 2D [^1H - ^{15}N] HSQC spectrum of DERAm before (gray) and after the addition of 1.7 mM dR5P over time (blue to red). Spectra were recorded over an incubation time of 8 h at 25 °C and 600 MHz. Resonances undergoing chemical shift and linewidth changes over time are labeled, suggesting the formation of side-products² that may interact with the protein.

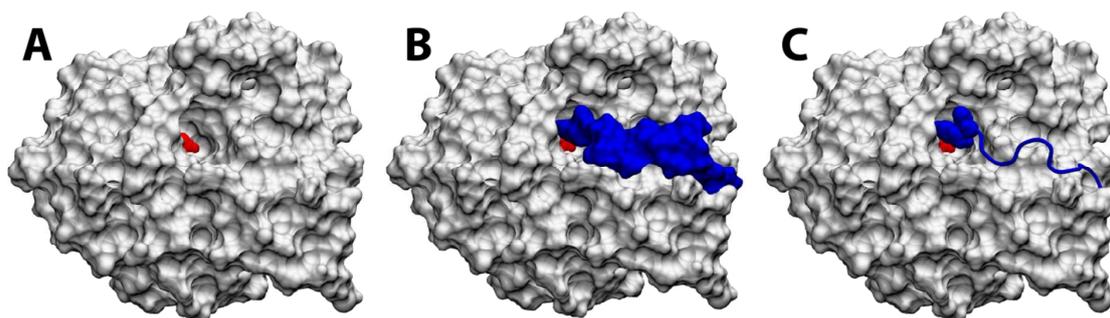


Figure S11: Surface representation of DERA with the catalytic K167 highlighted in red. (A) The crystal structure (PDB: 5EKY) without the C-terminal tail. (B) In the lowest energy NMR structure of the closed state, the C-terminal tail (rendered as blue surface) mostly covers the active site. (C) The C-terminal tail rendered as cartoon indicates that Y259, highlighted in VDW representation, is responsible for closing the active site pocket in the closed conformation.

A

<i>E. coli</i> (P0A6L0)	1	MTDLKASSLRALKLMDLTTLNDDDT	DEKVI	ALCHOAKTPVGNTAAI	CI	YPRFI	PI	ARKTL
<i>Shigella flexneri</i> (Q83P02)	1	MTDLKASSLRALKLMDLTTLNDDDT	DEKVI	ALCHOAKTPVGNTAAI	CI	YPRFI	PI	ARKTL
<i>Klebsiella pneumoniae</i> (Q7WT44)	1	MTDLKASSLRALKLMDLTTLNDDDT	NEKVI	ALCHOAKTPVGNTAAI	CI	YPRFI	PI	ARKTL
<i>Salmonella typhimurium</i> (Q8ZJV8)	1	MTDLKASSLRALKLMDLTTLNDDDT	NEKVI	ALCHOAKTPVGNTAAI	CI	YPRFI	PI	ARKTL
<i>Citrobacter koseri</i> (A0A078LIC5)	1	MTDLKASSLRALKLMDLTTLNDDDT	NEKVI	ALCHOAKTPVGNTAAI	CI	YPRFI	PI	ARKTL
<i>Cronobacter sakazakii</i> (A7MGB0)	1	MTDLKASSLRALKLMDLTTLNDDDT	NEKVI	ALCHOAKTPVGNTAAI	CI	YPRFI	PI	ARKTL
<i>Vibrio cholerae</i> (A0A0E4CB48)	1	MSELKVAALRALKLMDLTTLNDDT	DAKVI	QLCHAKSPVGNTAAI	CI	YPRFI	PI	ARKTL
<i>E. coli</i> (P0A6L0)	61	KEQGTPEIRI	ATVTNFPHGNDI	DI	ALAEETRAAI	AYGADEVVVFPYRAL	MAGNE	QVGF
<i>Shigella flexneri</i> (Q83P02)	61	KEQGTPEIRI	ATVTNFPHGNDI	EI	ALAEETRAAI	AYGADEVVVFPYRAL	MAGNE	QVGF
<i>Klebsiella pneumoniae</i> (Q7WT44)	61	NAQGTPEIRI	ATVTNFPHGNDI	DI	ALAEETRAAI	AYGADEVVVFPYRAL	AGNE	QVGF
<i>Salmonella typhimurium</i> (Q8ZJV8)	61	KEQGTPEIRI	ATVTNFPHGNDI	DI	ALAEETRAAI	AYGADEVVVFPYRAL	AGNE	QVGF
<i>Citrobacter koseri</i> (A0A078LIC5)	61	KEQGTPEIRI	ATVTNFPHGNDI	DI	ALAEETRAAI	AYGADEVVVFPYRAL	AGNE	QVGF
<i>Cronobacter sakazakii</i> (A7MGB0)	61	NEQGTPEIRI	ATVTNFPHGNDI	DI	ALAEETRAAI	AYGADEVVVFPYRAL	MAGNE	QVGF
<i>Vibrio cholerae</i> (A0A0E4CB48)	61	REQGTPEIRI	ATVTNFPHGNDI	EI	AVAEETKAAV	AYGADEVVVFPYRAL	AGNE	QVGF
<i>E. coli</i> (P0A6L0)	121	LVKACK	EACAAANVLLKVI	I	ETGELK	DEALI	RKASEI	SI
<i>Shigella flexneri</i> (Q83P02)	121	LVKACK	EACAAANVLLKVI	I	ETGELK	DEALI	RKASEI	SI
<i>Klebsiella pneumoniae</i> (Q7WT44)	121	LVKACK	EACAAANVLLKVI	I	ETGELK	EEALI	RKASEI	SI
<i>Salmonella typhimurium</i> (Q8ZJV8)	121	LVKACK	DACAAANVLLKVI	I	ETGELK	EEALI	RKASEI	SI
<i>Citrobacter koseri</i> (A0A078LIC5)	121	LVKACK	DACAAANVLLKVI	I	ETGELK	EEALI	RKASEI	SI
<i>Cronobacter sakazakii</i> (A7MGB0)	121	LVKACK	DACAAANVLLKVI	I	ETGELK	EEALI	RKASEI	SI
<i>Vibrio cholerae</i> (A0A0E4CB48)	121	LVK	ACKACGDKVLLKVI	I	ETGELK	EEALI	RKAS	QI
<i>E. coli</i> (P0A6L0)	181	SARI	MMEVIRDMGV	KT	VGFKPAGGVRT	AEDAQKYLAI	ADELFGADWAD	ARHYRF
<i>Shigella flexneri</i> (Q83P02)	181	SARI	MMEVIRDMGV	KT	VGFKPAGGVRT	AEDAQKYLAI	ADELFGADWAD	ARHYRF
<i>Klebsiella pneumoniae</i> (Q7WT44)	181	SARI	MLEVIRDMGV	KT	VGFKPAGGVRT	AEDAQFLAI	ADELFGADWADS	RHYRF
<i>Salmonella typhimurium</i> (Q8ZJV8)	181	SARI	MMEVIRDMGV	KT	VGFKPAGGVRT	AEDAQFLAI	ADELFGADWADS	RHYRF
<i>Citrobacter koseri</i> (A0A078LIC5)	181	SARI	MMEVIRDMGV	KT	VGFKPAGGVRT	AEDAQFLAI	ADELFGADWADS	RHYRF
<i>Cronobacter sakazakii</i> (A7MGB0)	181	SARI	MLEVIRDMGV	KT	VGFKPAGGVRT	AEDAQKYLAVADEL	FGADWAD	ARHYRF
<i>Vibrio cholerae</i> (A0A0E4CB48)	180	YARM	MLEVIRDMGV	KT	VGFKPAGGVRT	AEDAQYLA	MADEL	FGDWADS
<i>F. coli</i> (P0A6L0)	241	LASLLKAL	GHDGKSA	SSY				
<i>Shigella flexneri</i> (Q83P02)	241	LASLLKAL	GHDGKSA	SSY				
<i>Klebsiella pneumoniae</i> (Q7WT44)	241	LASLLKAL	GHDGKSA	SSY				
<i>Salmonella typhimurium</i> (Q8ZJV8)	241	LASLLKAL	GHDGKSA	SSY				
<i>Citrobacter koseri</i> (A0A078LIC5)	241	LASLLKAL	GHDGKSA	SSY				
<i>Cronobacter sakazakii</i> (A7MGB0)	241	LASLLKAL	GHDGKSA	SSY				
<i>Vibrio cholerae</i> (A0A0E4CB48)	240	L	NLLNTLEVT	D	KADP	AAY		

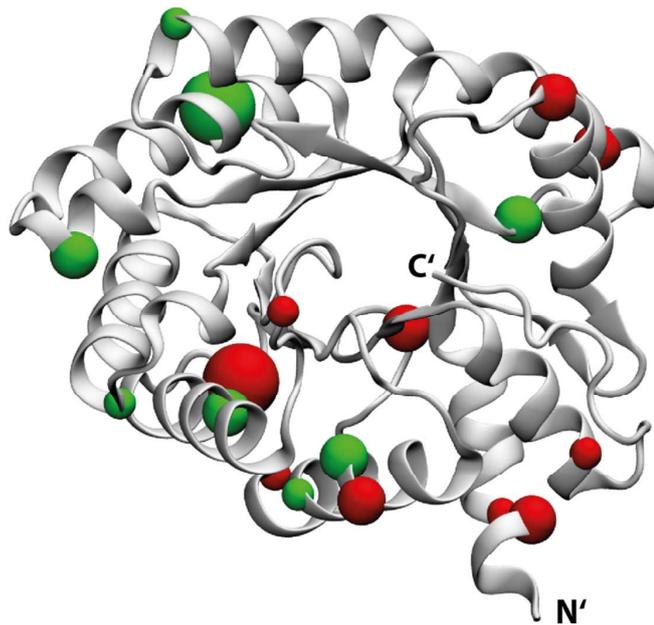
B

Figure S12: Comparison of amino acids sequences of *ecDERA* with DERA enzymes from human bacterial pathogens. (A) Sequence alignment of *ecDERA* with DERA enzymes from other human pathogens, created using T-coffee³: *S. flexneri* (99.2 % identity), *K. pneumoniae* (94.6 % identity), *S. typhimurium* (96.5 % identity), *C. koseri* (95.8 % identity), *C. sakazakii* (93.8 % identity) and *V. cholerae* (80.8 % identity). The UniProt accession numbers are shown in the bracket. Partially conserved residues are highlighted in green, non-conserved residues are highlighted in red. (B) The positions of the partially conserved (green) or non-conserved (red) residues are shown spheres, on the *ecDERA* structure. The size of the spheres corresponds to the number of sequences shown in panel (A) deviating from a particular residue (minimum two) at a given position.

REFERENCES

1. Heine, A.; DeSantis, G.; Luz, J. G.; Mitchell, M.; Wong, C.-H.; Wilson, I. A. Observation of covalent intermediates in and enzyme mechanism at atomic resolution. *Science* **2001**, *294*, 369-374.
2. Dick, M.; Hartmann, R.; Weiergräber, O. H.; Bisterfeld, C.; Classen, T.; Schwarten, M.; Neudecker, P.; Willbold, D.; Pietruszka, J. Mechanism based inhibition of an aldolase at high concentrations of its natural substrate acetaldehyde: structural insights and protective strategies. *Chem. Sci.* **2016**, *7*, 4492-4502.
3. Notredame, C.; Higgins, D. G.; Heringa, J. T-coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **2000**, *302*, 205-217.

3.3 The Nedd4-1 WW Domain Recognizes the PY Motif Peptide through Coupled Folding and Binding Equilibria

Summary

The focus of this work is to understand the role of conformational sampling in molecular recognition of the high affinity third domain (WW3*) from hNedd4-1 to the PY motif of the α -hENaC peptide. Comparison of the NOE-derived structure of WW3* in the apo state (pdb: 5AHT) with the previously solved structure of WW3* in complex with the α -hENaC peptide [137] reveals only minor changes in side chain conformations, which were further investigated by MD simulations. Closer inspection of the dynamics, however, reveals that WW3* in the apo state exhibits pronounced chemical exchange on the ms timescale which is quenched upon peptide binding. Multi-temperature CPMG relaxation dispersion experiments revealed that WW3* in the apo state exists in an equilibrium between the natively folded, binding-competent conformation and an unfolded state, which is populated $\sim 20\%$ at 37 °C. These results highlight that conformational sampling of WW3* is crucial for peptide recognition, as the α -hENaC peptide selectively binds the natively, folded binding-competent state.

Acknowledgments Reprint Permission

This section contains a complete reprint of the publication and supporting information. Reprinted with permission from *Biochemistry* **2016**, 55(4), 659-674, doi:[10.1021/acs.biochem.5b01028](https://doi.org/10.1021/acs.biochem.5b01028). Copyright 2016 American Chemical Society.

The Nedd4–1 WW Domain Recognizes the PY Motif Peptide through Coupled Folding and Binding Equilibria

Vineet Panwalkar,^{†,||} Philipp Neudecker,^{||,⊥} Michael Schmitz,[†] Justin Lecher,^{||,⊥} Marianne Schulte,^{||,⊥} Karima Medini,^{‡,§} Matthias Stoldt,^{||,⊥} Margaret A. Brimble,^{‡,‡,§} Dieter Willbold,^{||,⊥} and Andrew J. Dingley^{*,||}

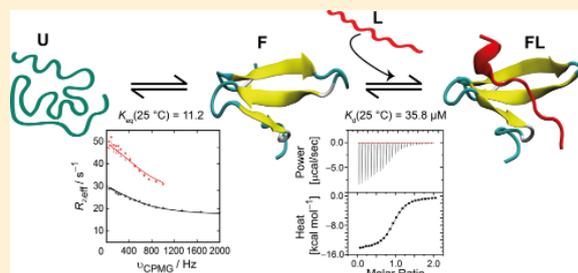
[†]School of Chemical Sciences, [‡]School of Biological Sciences, and [§]The Maurice Wilkins Centre for Molecular Biodiscovery, The University of Auckland, Auckland 1142, New Zealand

^{||}ICS-6 (Strukturbiochemie), Forschungszentrum Jülich, 52425 Jülich, Germany

[⊥]Institut für Physikalische Biologie, Heinrich-Heine-Universität, 40225 Düsseldorf, Germany

Supporting Information

ABSTRACT: The four WW domains of human Nedd4–1 (neuronal precursor cell expressed developmentally downregulated gene 4–1) interact with the PPxY (PY) motifs of the human epithelial Na⁺ channel (hENaC) subunits, with the third WW domain (WW3*) showing the highest affinity. We have shown previously that the α -hENaC PY motif binding interface of WW3* undergoes conformational exchange on the millisecond time scale, indicating that conformational sampling plays a role in peptide recognition. To further understand this role, the structure and dynamics of hNedd4–1 WW3* were investigated. The nuclear Overhauser effect-derived structure of apo-WW3* resembles the domain in complex with the α -hENaC peptide, although particular side chain conformations change upon peptide binding, which was further investigated by molecular dynamics simulations. Model-free analysis of the ¹⁵N nuclear magnetic resonance spin relaxation data showed that the apo and peptide-bound states of WW3* have similar backbone picosecond to nanosecond time scale dynamics. However, apo-WW3* exhibits pronounced chemical exchange on the millisecond time scale that is quenched upon peptide binding. ¹H_N and ¹⁵N Carr–Purcell–Meiboom–Gill (CPMG) relaxation dispersion experiments at various temperatures revealed that apo-WW3* exists in an equilibrium between the natively folded peptide binding-competent state and a random coil-like denatured state. The thermodynamics of the folding equilibrium was determined by fitting a thermal denaturation profile monitored by circular dichroism spectroscopy in combination with the CPMG data, leading to the conclusion that the unfolded state is populated to ~20% at 37 °C. These results show that the binding of the hNedd4–1 WW3* domain to α -hENaC is coupled to the folding equilibrium.



Ubiquitination of cellular proteins is an abundant posttranslational modification process in eukaryotes that was originally considered solely as a marker for proteasome degradation;¹ however, ubiquitination is now known to regulate various cellular activities.^{2,3} This diversity in function arises because of the large variety of E3 ubiquitin ligases that catalyze ubiquitin chain formation, and the modular architecture of these ligases mediates substrate specificity and ligase activity. The human Nedd4 (neuronal precursor cell expressed developmentally downregulated gene 4) family of HECT-type E3 ubiquitin ligases contains nine members. The prototypic member of the family, Nedd4–1, was originally discovered as a developmentally downregulated gene in mouse brain.⁴ The human epithelial Na⁺ channel (hENaC) was the first recognized substrate of Nedd4–1.^{5,6} Subsequent results show that human Nedd4–1 (hNedd4–1) is both a positive and negative regulator of the activity of signaling receptors,⁷ including the insulin-like growth factor-1 receptor.⁸ hNedd4–

1 is overexpressed in various cancers and is thus a potential anticancer drug target, possibly through the design of peptide therapeutics.⁹

The domain architecture of Nedd4 proteins includes an N-terminal C2 domain that binds Ca²⁺ and phospholipids, a central region composed of up to four WW domains that are responsible for the recognition of substrates and a C-terminal HECT ubiquitin ligase domain.¹⁰ Nedd4 proteins form an autoinhibited conformation via an interdomain interaction between the C2 and HECT domains.^{11,12} The catalytic activity of Nedd4 proteins is activated upon the C2 domain binding Ca²⁺, which then leads to recruitment of Nedd4 to the plasma membrane.^{13,14} Additionally, the autoinhibited state of Nedd4–1 is released by adapter proteins via interaction with the WW

Received: September 17, 2015

Revised: December 15, 2015

Published: December 18, 2015

domains of Nedd4–1.¹⁵ Thus, the WW domains play a role in both regulating activity and substrate recognition by Nedd4 family members.

WW domains are small interaction modules of 38–40 residues that have two conserved tryptophans and an invariant proline.^{16–19} Structures of WW domains in complex with their cognate peptides show that the first tryptophan and the invariant proline are important for the stability of the domain, whereas the second tryptophan is essential for peptide recognition^{20–25} and is one of two conserved aromatic residues that form the hydrophobic “XP pocket” that binds a conserved proline in proline-rich motifs recognized by WW domains. WW domains adopt a common three-stranded twisted antiparallel β -sheet fold but fall into five different classes because they recognize different proline-rich motifs. Nedd4 WW domains bind PY motif (L/PPxY) peptides and are members of class I WW domains.²⁶

The WW domains of hNedd4–1 bind the α -hENaC PY motif with different affinities.^{27,28} The strongest binding was reported for the third WW domain (WW3*) of hNedd4–1 with a dissociation constant (K_d) of $\sim 5 \mu\text{M}$; WW1 showed no affinity, and WW2 and WW4 showed binding that was ~ 25 -fold weaker than that of WW3*.²⁸ Moreover, a construct consisting of WW domains 2, 3, and 4 of Nedd4–1 did not bind the α -hENaC peptide with higher affinity, showing that the WW3* domain is sufficient for high-affinity substrate recognition.²⁸ Examination of a sequence alignment among the four WW domains of hNedd4–1 shows that the majority of the WW3* residues directly interacting with the α -hENaC peptide either are sequence identical or have a conserved substitution in the other three WW domains (Figure 1). This high degree of

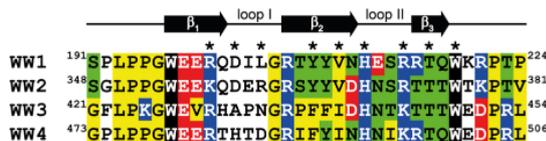


Figure 1. Alignment of the four human Nedd4–1 WW domains. Sequence homology between domains is highlighted for nonpolar (yellow), polar (green), acidic (red), and basic (blue) amino acids, and the conserved tryptophans are highlighted in black. The secondary structure is shown above the sequence alignment, and asterisks mark those residues of the WW3* domain that are involved in peptide interaction. Sequence numbering is according to NCBI sequence entry NM_006154.3, and the sequence alignment was performed with CLUSTALW.⁹⁷

sequence homology extends to other species, including rat, mouse, *Drosophila*, and *Xenopus* Nedd4 isoforms.²⁷ The majority of the sequence diversity between the WW domains of hNedd4–1 resides in the β_1 – β_2 loop, and biophysical studies combined with mutagenesis of rat Nedd4 WW4 have shown that residues of this loop play a role in peptide binding affinity,²⁹ in agreement with the general notion that sequence divergence in the loops encodes specificity for WW domain–peptide complexes.²² Nonetheless, studies of the WW domain from the human peptidyl-prolyl isomerase (Pin1) showed that besides chemical recognition, the flexibility of the PIN1 WW β_1 – β_2 loop also augments peptide affinity.³⁰ Thus, it appears that both chemical recognition and dynamics play a role in defining the peptide affinity of WW domains, and understanding these determinants for hNedd4–1 should aid in the

development of approaches to modulating Nedd4 interaction networks associated with human disease.

We have reported a structure–dynamics study of the Nedd4–1 WW3*– α -hENaC peptide complex. In this study, residues located at the domain–peptide interface were found to undergo conformational exchange on the microsecond to millisecond time scale, suggesting that conformational sampling might direct peptide recognition.³¹ To further understand the importance of motion in peptide recognition, we have investigated the structure and dynamics of hNedd4–1 WW3*. The NOE-derived structure of the apo-WW3* domain is very similar to that of the domain in complex with the α -hENaC peptide, whereas the dynamics of the domain show dramatic changes upon peptide recognition. Here, microsecond to millisecond motions are observed for many residues in the apo state, and these motions are quenched upon peptide binding. Moreover, CD melt and relaxation dispersion analysis showed that in the apo state an unfolded form of the WW domain exists, and at 37 °C, $\sim 20\%$ of the protein population adopts this state. The mechanism of peptide binding by the WW domain is thus reminiscent of coupled folding and binding equilibria.

EXPERIMENTAL PROCEDURES

Sample Preparation. The WW3* domain of hNedd4–1 (G416–A458) was expressed and purified as described previously.²⁸ A peptide representing the extended PY motif in the α -hENaC subunit (⁶³⁸/TAPPPAYATLG⁶⁴⁸’, where a prime is used to distinguish it from the WW domain sequences) was synthetically produced.³¹ For structure determination, a 1.8 mM sample containing the [^U-¹³C,¹⁵N]hNedd4–1 WW3* domain was prepared in 20 mM sodium phosphate buffer (pH 6.5), 50 mM NaCl, 0.1% (w/v) Na₂S₂O₃, and 1 mM DSS in a 93%/7% (v/v) H₂O/D₂O mixture. The samples used for relaxation experiments were a 1.5 mM sample containing [^U-¹³N]hNedd4–1 WW3* and a 1.5 mM sample containing [^U-¹⁵N]hNedd4–1 WW3* with 4.5 mM α -hENaC peptide. Both samples were prepared in the buffer used for structure calculations.

Isothermal Titration Calorimetry. For ITC measurements, the protein samples were dialyzed extensively against 20 mM sodium phosphate and 50 mM NaCl (pH 6.5). Experiments were performed with a MicroCal VP-ITC instrument (GE Healthcare Europe GmbH, Freiburg, Germany). Twenty-five aliquots of 1.52 μL of 8 mM α -hENaC peptide were injected into 200 μL of 0.8 mM hNedd4–1 WW3* domain, with a mixing speed of 300 rpm. The temperature was regulated at 25 °C. The experiment was conducted in triplicate. Following subtraction of the corresponding heats of dilution, the binding isotherms were fitted by nonlinear least-squares optimization using 100 Levenberg–Marquardt iterations to obtain the binding stoichiometry (n), the apparent enthalpy of interaction (ΔH_A^{app}), and the apparent equilibrium dissociation constant (K_d^{app}). Data processing was performed with the Origin ITC analysis software supplied by GE Healthcare.

Chemical Shift Perturbation Analysis. The α -hENaC peptide at concentrations ranging from 0 to 300 μM was titrated against 10 μM hNedd4–1 WW3* domain in 20 mM sodium phosphate and 50 mM NaCl (pH 6.5). Two-dimensional (2D) ¹H–¹⁵N HSQC spectra for each titration point were recorded at 25 °C and 800 MHz. The data matrix of the 2D ¹H–¹⁵N HSQC consisted of 128* \times 960* data points

(where n^* refers to complex points) with acquisition times of 49 ms (t_N) and 75 ms (t_{HN}). A total of 64–80 scans per complex t_N increment were collected. The recycle delay was 1.25 s. The total measuring time was 3–3.8 h. The equilibrium dissociation constant, K_d^{app} , was obtained from changes in the weighted average chemical shift differences³² $\Delta\delta_{av} = [(\Delta\delta_{HN}^2 + \Delta\delta_N^2/25)/2]^{1/2}$ assuming a one-site binding model:

$$\Delta_{obs} = 0.5\Delta_{maximum} \left[1 + X + \frac{K_d^{app}}{[P_0]} - \sqrt{\left(1 + X + \frac{K_d^{app}}{[P_0]} \right)^2 - 4X} \right] \quad (1)$$

where $[P_0]$ is the total protein concentration and X is the molar ratio of the ligand to protein.

Resonances for residues L423, W427, V429, H431, G435, R436, D441, N443, D451, L454, and K455 were used to report the average K_d^{app} for the hNedd4–1 WW3*– α -hENaC complex. These residues showed linear chemical shift changes upon titration and could be fit by the one-site binding model (eq 1), which neglects any complications from the coupling of the binding equilibrium to the folding equilibrium.

NMR Measurements for Structure Determination.

NMR spectra were recorded at 25 °C on NMR spectrometers equipped with cryogenically cooled z-gradient probes operating at ¹H frequencies of 600 and 900 MHz. Backbone and aliphatic and aromatic side chain ¹H, ¹⁵N, and ¹³C resonance assignments for the hNedd4–1 WW3* domain were obtained from multidimensional heteronuclear NMR experiments (Table S1).^{33,34} Proton chemical shifts were referenced to DSS, whereas the ¹⁵N and ¹³C chemical shifts were indirectly referenced according to the ratios given by Wishart et al.³⁵ Data sets were processed using NMRPipe³⁶ and analyzed using CcpNMR Analysis.³⁷

Distance restraints were obtained from ¹⁵N- and ¹³C-edited NOESY spectra with mixing times between 150 and 180 ms. Backbone dihedral restraints for the ϕ and ψ angles were derived from backbone chemical shifts using TALOS+.³⁸ ³J scalar couplings providing information about the χ_1 angles were obtained from quantitative ³J_{NH β} -HNHB³⁹ and ³J_{H α H β} -HAHB-(CACO)NH⁴⁰ experiments.

Structure Calculations. NOE cross peak assignments of the acquired ¹⁵N- and ¹³C-edited NOESY spectra were obtained by an iterative procedure using a combination of manual and automatic steps. The tolerances for automatic assignments by Aria version 2.3.1⁴¹ were 0.03–0.06 and 0.05–0.09 ppm for the ¹H direct and indirect dimensions, respectively, and 0.5 ppm for the heteronuclear dimensions. Structures were calculated by a combination of Aria version 2.3.1⁴¹ and CNS version 1.21⁴² (including the Aria patchset) using the PARALLHDG force field with a log-harmonic potential^{43,44} and automatic restraint weighting. All molecular dynamics parameters were used in the default configuration. The 15 lowest-energy structures of the 100 calculated were further refined in a 9 Å explicit water shell as a final step in the Aria procedure. The stereochemical quality of the refined models was assessed via PROCHECK-NMR⁴⁵ and MolProbity.⁴⁶ VMD software was used for the superposition and visualization of structures.⁴⁷

NMR Relaxation Experiments. Backbone dynamics for the hNedd4–1 WW3* domain and the WW3* domain in complex with the α -hENaC peptide were derived from ¹⁵N relaxation experiments⁴⁸ at static magnetic field strengths of 14.1, 18.8, and 21.2 T and a temperature of 25 °C (Tables S2 and S3). R_1 experiments were conducted using inversion recovery times between 30.2 and 1811.7 ms, with a recycle delay of 3.0 s. The $R_{1\rho}$ experiments were conducted according to the method of Korzhnev et al.⁴⁹ with spin-lock periods between 5 and 150 ms, and ¹⁵N spin-lock field strengths of 1.87 kHz (21.2 T), 1.84 kHz (18.8 T), and 1.72 kHz (14.1 T). A recycle delay of 3.0 s was used. R_1 and $R_{1\rho}$ values were determined by nonlinear least-squares fitting of the intensities quantified in Analysis by two-parameter monoexponential equations using CurveFit (A. G. Palmer, III, Columbia University, New York, NY). R_2 values were determined from the measured R_1 and $R_{1\rho}$ rates.⁵⁰ ¹H–¹⁵N heteronuclear NOE values were measured from pairs of interleaved spectra recorded with (NOE) and without (control) proton saturation during the recycle delay. A recycle delay of 12 s was used at 14.1 T, whereas a delay of 15 s was used at the higher fields. Reported errors for rates were standard deviations from Monte Carlo simulations in CurveFit. The heteronuclear steady state ¹H–¹⁵N NOE values were calculated from peak intensity ratios obtained from the NOE and control spectra, with uncertainties estimated from the background noise of the spectra.

¹⁵N single-quantum^{51,52} and ¹H_N single-quantum⁵³ CPMG relaxation dispersion (RD) experiments were conducted at 14.1 and 21.2 T. ¹⁵N single-quantum CPMG data sets for apo-hNedd4–1 WW3* were acquired at 5, 10, 15, 20, and 25 °C, whereas data sets for the hNedd4–1 WW3* domain– α -hENaC complex were recorded at 5, 15, and 25 °C. ¹H_N single-quantum CPMG experiments were conducted at 5 °C. Data were processed with NMRPipe and quantified with Analysis. In each ¹⁵N CPMG experiment, 14 or 21 different CPMG frequencies $\nu_{CPMG} = 1/2\delta$, where δ is the time between consecutive refocusing pulses, ranging up to 2000 (1000) Hz at 14.1 (21.2) T were sampled during a constant-time relaxation interval (T_{CPMG}) of 48 ms. Because of the large ¹⁵N sweep width at 21.2 T, the experiments were conducted at two separate ¹⁵N offsets (118 and 108 ppm) to reduce offset effects. In each ¹H_N CPMG experiment, 15 different CPMG frequencies ranging up to 2000 Hz at 14.1 and 21.2 T were sampled during a T_{CPMG} of 30 ms. Global exchange parameters (exchange rates and equilibrium populations) and residue specific values (¹⁵N chemical shift differences and intrinsic relaxation rates $R_{2,0}$) were extracted by a nonlinear least-squares fitting procedure whereby experimental dispersion profiles [$R_2(\nu_{CPMG}) = R_{2,0} + R_{ex}(\nu_{CPMG})$] were fit to those calculated by deriving the evolution of magnetization during the CPMG interval by solving the Bloch–McConnell equations numerically for a two-site exchange model. This fitting routine was performed as described previously,^{54–56} and errors of the fitted parameters were calculated from the covariance matrix.⁵⁷ Absolute signs of ¹⁵N chemical shift differences between exchanging states were obtained from differences in ¹⁵N peak positions in 2D ¹H–¹⁵N HMQC and 2D ¹H–¹⁵N HSQC spectra recorded at 21.2 T and/or between two 2D ¹H–¹⁵N HSQC spectra obtained at 14.1 and 21.2 T.⁵⁸ Absolute signs of the ¹H_N chemical shift differences were determined from the difference in ¹H_N resonance positions in 2D ¹H–¹⁵N HSQC spectra recorded at 5 °C, and 14.1 and 21.2 T.⁵⁹

Model-Free Analysis. Model-free analysis was performed using the method for the combined optimization of the global diffusion tensor and local model-free parameters implemented in the software relax (version 2.0),^{60–62} as described in more detail previously.³¹

Circular Dichroism Spectropolarimetry. Far-UV circular dichroism (CD) spectra were recorded using a temperature-controlled JASCO J-1100 spectropolarimeter (Jasco, Easton, MA). Apo-hNedd4–1 WW3*, the hNedd4–1 WW3*– α -hENaC peptide complex, and the α -hENaC peptide samples in 20 mM sodium phosphate and 50 mM NaCl (pH 6.5) were equilibrated for 30 min at the set temperature. Spectra were recorded at 0, 25, 50, and 75 °C. CD spectra were recorded over the wavelength range of 200–260 nm in 0.2 nm intervals with a scan speed of 2 nm min⁻¹. The path length of the quartz cuvette was 2.0 mm. The WW3* sample concentration was 55 μ M (275 μ M peptide). For each measurement, five scans were averaged and a spectrum of the buffer alone was subtracted. For the complex data, both the buffer signal and the signal from 83% of a 275 μ M peptide signal were subtracted. The subtraction of 83% of the free peptide CD signal derives from calculation of the concentration of the peptide–domain complex using a standard heteronuclear bimolecular binding scheme, where [peptide] = 275 μ M, [WW3*] = 55 μ M, and the ITC-derived K_d value is 44 μ M (see Results). Approximately 10% of the apo WW3* domain is also present in the sample of the complex. This was not corrected for in the CD spectrum of the complex because of the uncertainty associated with the K_d value. CD spectra were normalized to yield mean residue ellipticity.

Equilibrium thermal denaturation CD profiles of the apo-WW3* domain were acquired by monitoring the ellipticity at 230 nm as a function of temperature employing a constant scan rate of 0.5 deg min⁻¹ over the range of 0–98 °C. Signal averaging at each temperature step was 4 s, and the WW3* concentration was 55 μ M. A buffer only signal was subtracted. The thermal denaturation curve and fractions of the unfolded domain obtained from relaxation dispersion analysis at various temperatures were fitted to a standard equation describing a two-state transition (eq 2⁶³) using a previously reported excess heat capacity for folding ($\Delta C_{p,F} = C_{p,F} - C_{p,U}$) value of -1745 J K⁻¹ mol⁻¹ for the WW domain of YAP65.⁶⁴ This value is in close agreement with the value derived using empirical relationships.

$$\varepsilon = \frac{\varepsilon_U^0 - \varepsilon_F^0 + (m_U - m_F)T}{1 + \exp\left\{\frac{1}{R}\left[\Delta C_{p,F} \ln\left(\frac{T}{T_m}\right) - \left(\frac{1}{T} - \frac{1}{T_m}\right)(\Delta H_F - T_m \Delta C_{p,F})\right]\right\}} + \varepsilon_F^0 + m_F T \quad (2)$$

where T is the temperature, ε is the ellipticity at 230 nm (millidegrees), R is the universal gas constant, T_m is the thermal midpoint of unfolding, ΔH_F is the molar enthalpy of folding at T_m , ε_U^0 and ε_F^0 are the ellipticities of the unfolded and folded domain at 0 K, respectively, and m_U and m_F are the temperature slopes of the ellipticities of the unfolded and folded domain, respectively.

Equilibrium unfolding transitions were normalized to the fraction of unfolded domain (F_U):

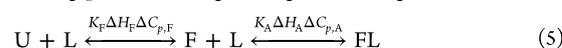
$$F_U = \frac{1}{1 + \exp\left\{\frac{1}{R}\left[\Delta C_{p,F} \ln\left(\frac{T}{T_m}\right) - \left(\frac{1}{T} - \frac{1}{T_m}\right)(\Delta H_F - T_m \Delta C_{p,F})\right]\right\}} \quad (3)$$

The simultaneous fits of the CD thermal denaturation curve (eq 2) and fractions of the unfolded domain from CPMG analysis (eq 3) used the leastsq optimizer module from Python scipy package version 0.13.3 to minimize the combined χ^2 function:

$$\chi^2 = \sum \frac{[\varepsilon - \varepsilon(T)]^2}{\sigma_\varepsilon^2} + f_{\text{CPMG}} \times \sum \frac{[F_U - F_U(T)]^2}{\sigma_{F_U}^2} \quad (4)$$

Scaling factor f_{CPMG} used in weighting the relaxation dispersion-derived fraction unfolded data relative to the CD data was established by minimizing the resulting χ^2 totals for the fraction unfolded data points as a function of scaling factor f_{CPMG} . Minimization was considered converged when the χ^2 total did not improve by more than 5%.

A coupled folding–binding model was used to extract the K_d and ΔH_A values of binding to the active state using a combination of the ITC and CD data, as described previously.⁶⁵ The common thermodynamic equilibrium model describing protein folding and ligand binding is



In this model, the folded domain (F) is in equilibrium with the peptide-bound state (FL) and the partially or fully unfolded form (U) is thought to be binding-incompetent, so that the population of the “encounter complex” (UL) is not significant. Indeed, the chemical exchange from domain unfolding and/or folding is quenched in the complex, and we do not observe any additional resonances in 2D ¹H–¹⁵N HSQC spectra of the complex that would suggest any significant population of UL. The thermodynamic parameters ΔH_F , ΔH_A , $\Delta C_{p,F}$, and $\Delta C_{p,A}$ are changes in enthalpy and heat capacity for the folding and binding steps. K_F and $K_A = 1/K_d$ are the folding and ligand association equilibrium constants, respectively. The ITC measurements contain contributions from both folding and binding reactions. Thus, the thermodynamic parameters are related to the measured binding parameters ΔH_A^{app} and K_d^{app} as

$$K_A^{\text{app}}(T) = (1 - F_U)K_A(T) \Leftrightarrow K_d = (1 - F_U)K_d^{\text{app}} \quad (6)$$

$$\Delta H_A^{\text{app}}(T) = \Delta H_A(T) + F_U \Delta H_F \quad (7)$$

where

$$\Delta H_F(T) = \Delta H_F(T_m) + \Delta C_{p,F}(T - T_m) \quad (8)$$

Thus, ΔH_A and K_d values were extracted using eqs 6–8, the CD melt analysis of the apo-WW domain, and the measured ΔH_A^{app} and K_d^{app} from ITC at 25 °C.

Molecular Dynamics Simulations. GROMACS version 4.6.5⁶⁶ was used to perform MD simulations on apo-hNedd4–1 WW3* and the hNedd4–1 WW3*– α -hENaC complex. The starting structures used for the simulations were the same lowest-energy structures from the NMR ensembles used for model-free analysis [Protein Data Bank (PDB) entries 2M3O for the hNedd4–1 WW3*– α -hENaC complex and SAHT for the apo-hNedd4–1 WW3* structure]. Prior to MD simulations, the structures were solvated with the TIP4P⁶⁷ water model and energy minimized via steepest descent minimization. The solvated apo-hNedd4–1 WW3* system consisted of 6387 water molecules in a 197.1 nm³ cubic box, whereas the hNedd4–1 WW3*– α -hENaC complex contained 6983 water molecules in a 222.2 nm³ cubic box. Six Na⁺ and seven Cl⁻ ions were added to balance the net charge and maintain the salt concentration of the simulated systems at 50 mM. A 1 ns

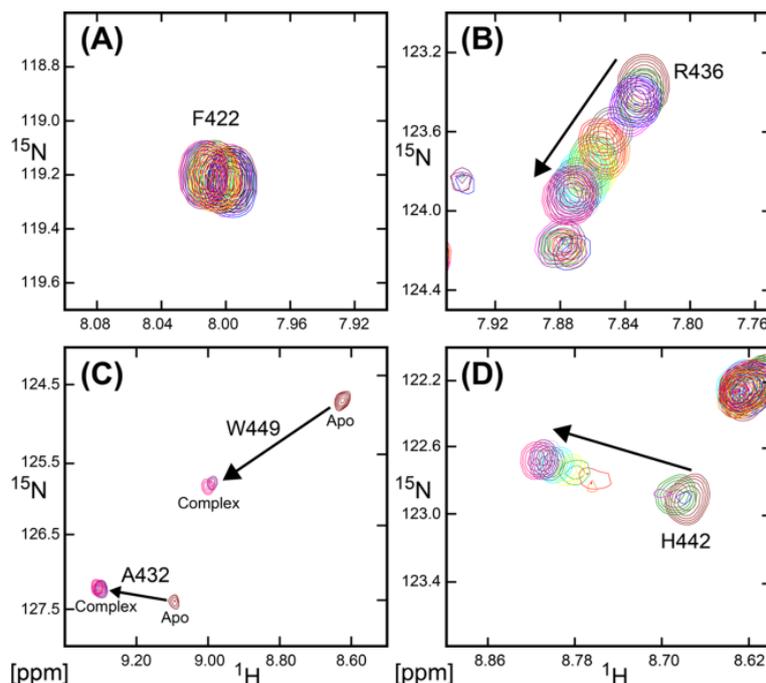


Figure 2. Different chemical exchange processes observed during titration of the α -hENaC peptide against the WW3* domain (10 μ M). The peptide was titrated to 30-fold excess (0–300 μ M) with 15 titration points. (A) The resonance in the 15 overlaid 2D ^1H – ^{15}N HSQC representing F422 shows small chemical shift change during the titration. (B) The resonance for R436 shows a progressive shift during the titration, indicative of fast exchange. Resonances representing residues located at the peptide binding interface, W449, A432, and H442, show characteristics of intermediate-to-slow exchange as the peptide is titrated (C and D). Resonances for W449 and A432 were observed only at the lowest (apo) and highest (complex) peptide concentrations.

equilibration was conducted in the *NVT* ensemble with a final temperature of 25 $^{\circ}\text{C}$, followed by a 1 ns equilibration using the *NPT* ensemble with the Parrinello–Rahman barostat at 1 atm. To calculate long-range electrostatic interactions, the particle mesh Ewald method^{68,69} was used with a 1 nm cutoff. The AMBER99SBnmr1-ILDN force field⁷⁰ with periodic boundary conditions was employed to conduct the simulations, and snapshots were taken every 2 ps. Production runs of 100 ns were conducted for the apo and complex systems using the leapfrog integrator and a time step of 2 fs. Bond lengths were constrained using the LINCS algorithm.

Backbone chemical shifts were calculated for each frame of the MD trajectory in both systems (4 ps spacing) using SPARTA+.⁷¹ SPARTA+ was also used to compute the backbone chemical shifts of the lowest-energy NMR structures (PDB entries 2M3O and 5AHT) and the X-ray crystal structures (PDB entries 4N7F and 4N7H). The water molecules within the crystal structure were removed; the structure was solvated as described above, and the shifts were computed after a 200-step minimization had been performed with the AMBER99SBnmr-ILDN force field.

NMR Data and Coordinates. NMR resonance assignments of the hNedd4–1 WW3* domain have been deposited in the BioMagResBank (accession code 25349). The structural coordinates have been deposited in the PDB (5AHT).

RESULTS

Binding of the PPXY Motif by the hNedd4–1 WW3* Domain Leads to Changes in Chemical Exchange Processes.

The binding affinity of hNedd4–1 WW3* for

the α -hENaC peptide was determined by isothermal titration calorimetry (ITC) (Figure S1 and Table S4). Fitting of the binding isotherms (run in triplicate) derived from the titrations gave an average apparent K_d (K_d^{app}) value of (44.53 ± 1.84) μM at 25 $^{\circ}\text{C}$ with a favorable apparent association enthalpy term (ΔH_A^{app}) of (-58.20 ± 0.29) kJ mol^{-1} that offsets the unfavorable entropy term of $-106 \text{ J K}^{-1} \text{ mol}^{-1}$. The energy terms are in close agreement with previous ITC results for the WW domain from the human Yes-associated protein binding its cognate PY ligand at pH 6.0 and 25 $^{\circ}\text{C}$.⁷² The K_d^{app} is ~ 10 -fold higher than the reported value using surface plasmon resonance (SPR).²⁸ This discrepancy in K_d^{app} is most likely because of the different buffer conditions, as previously shown for the rat Nedd4 WW domains where a change in pH from 7 to 6 resulted in a slight decrease of the binding affinity presumably due to changes in the ionization state of histidine rings involved in binding.²⁴ In the hNedd4–1 WW3* domain, the ionization state of H442, a key residue involved in recognition of Y644', would differ between the previous SPR study (pH 8.0) and this study (pH 6.5), thereby potentially affecting peptide affinity.

To further characterize the binding affinity, a chemical shift perturbation titration was performed using the assigned 2D ^1H – ^{15}N HSQC spectra of apo and peptide-bound forms of the hNedd4–1 WW3* domain (Figure S2). Only residues in the third β -strand ($\beta 3$) (T446, T447, and T448) and F438 (XP pocket) showed weighted chemical shift differences ($\Delta\delta_{\text{av}}$) larger than one standard deviation (SD) above the mean $\Delta\delta_{\text{av}}$ (Figure S3), suggesting that peptide binding does not lead to changes in the overall fold of the domain. Nonetheless, the $\Delta\delta_{\text{av}}$ values for residues in $\beta 3$ suggest that this strand undergoes the

Table 1. Assignments and Structural Statistics for the 15 Lowest-Energy Structures of the Nedd4-1 WW3* Domain

assignment statistics	
no. of residues (418–458)	41
molecular mass (Da)	4805
completeness of resonance assignments (%)	
protein, backbone ^a (193/200)	96.5
protein, all atoms ^b (302/319)	95.4
structural statistics	
NOE distance restraints	
assigned	
intraresidue ($i = j$)	1328
interresidue	416
sequential ($ i - j = 1$)	390
medium-range ($1 < i - j < 5$)	266
long-range ($ i - j \geq 5$)	256
ambiguous	264
root-mean-square deviation value from experimental distance restraints (Å) (1592)	0.120 ± 0.015
root-mean-square deviation value from TALOS+ torsion angle restraints (deg) (60)	0.980 ± 0.264
deviations from idealized covalent geometry	
bonds (Å)	0.0035 ± 0.0002
angles (deg)	0.45 ± 0.017
impropers (deg)	1.10 ± 0.07
coordinate precision (Å) ^e	
core region (423–457)	
backbone	0.71 ± 0.19
all heavy atoms	0.99 ± 0.16
RPF scores ^d	
recall	0.79
precision	0.88
F measure	0.83
DP score	0.71
Ramachandran statistics (%) ^e	
residues in most favored regions	85.3 ± 4.1
residues in additionally allowed regions	13.9 ± 3.8
residues in generously allowed regions	0.0 ± 0.0
residues in disallowed regions ^f	0.8 ± 0.0

^aObtained from the ¹H^N, ¹³C^α, ¹⁵N, ¹³C', and ¹H^β resonances. ^bRoutinely assigned ¹H, ¹⁵N, and ¹³C resonances are taken into account, excluding the N-terminal and Lys amino groups, guanidino groups of Arg, side chain hydroxyl protons of Ser, Thr, and Tyr, carboxyl resonances of Asp and Glu, and quaternary aromatic carbons. ^c¹H belonging to the same methyl group and Phe, Tyr ¹H^δ, ¹H^ε are counted as one signal. ^dCoordinate precision is the average root-mean-square deviation of all the structures from the average structure coordinates. ^eCalculated RPF scores according to Huang et al.⁹⁶ ^fCalculated with PROCHECK-NMR.⁴⁵ ^gResidues in the disallowed region are outside the core region of the domain (L423–K457).

largest conformational transition upon peptide binding. Interestingly, the observed narrowing of resonance line widths upon peptide interaction indicates that the Nedd4–1 WW3*– α -hENaC peptide complex adopts a stable WW fold that does not undergo chemical exchange on the microsecond to millisecond time scale (Figure S2) and that peptide binding quenches chemical exchange.

The ¹H–¹⁵N HSQC titration of the α -hENaC peptide with the ¹⁵N-labeled WW3* domain showed that exchange processes between free and bound forms ranged from slow-to-intermediate to fast on the NMR time scale. In general, resonances that show the smallest chemical shift change are in fast exchange, whereas those that show the largest chemical shift changes are in slower exchange. For resonances representing residues at the N-terminus (M418–K425) and C-terminus (I456–A458) and residues within the core of the domain that are not involved in ligand recognition (i.e., R436, D441, D451, and V429), the chemical shift changes observed were small and in the fast exchange regime (Figure 2A,B and Figure S3). Resonances for peptide-interacting residues and residues adjacent to these interacting residues (R430, H431,

A432, N434, F438, I440, H442, T447, W449, and K445) underwent intermediate-to-slow exchange during the peptide titration experiment (Figure 2C,D). Residues along the β 3 strand (T444, T446, and T448) were also observed to undergo slow-to-intermediate exchange processes on the NMR time scale. Residues undergoing fast-to-intermediate exchange were used to extract K_d^{pp} values (eq 1). Chemical shift changes for L423, W427, V429, H431, G435, R436, D441, N443, D451, L454, and K455 were used for calculating the K_d^{pp} of the domain–peptide interaction (Figure S4). The average K_d^{pp} value for the hNedd4–1 WW3*– α -hENaC peptide complex interaction was calculated to be (30.41 ± 2.32) μ M ($n = 11$) and is in good agreement with the ITC analysis.

The Backbone Structure of the WW3* Domain from hNedd4–1 Is Similar in the Apo and Peptide-Bound States, but the Side Chain Conformations Are Different. Standard triple-resonance experiments (Table S1) were used to obtain backbone and side chain chemical shift assignments of the hNedd4–1 WW3* domain. Near complete assignments were obtained for the backbone (96.5%) and side chain (95.4%) resonances for the WW3* domain (Table 1).

Structural restraints for the WW3* domain included backbone torsion angle restraints from TALOS+;³⁸ 1592 NOEs and side chain χ_1 angles for residues N434, D441, H442, D451, and R453 were restrained to one of the staggered conformations (60° , 180° , and -60°) $\pm 60^\circ$ as determined by combined $^3J_{\alpha\beta}$ and $^3J_{N\beta}$ coupling analysis (Figure S5). 3J coupling data to define the staggered rotamers of the $C\alpha-C\beta$ bond of other residues gave either intermediate coupling values indicative of dynamic averaging between rotameric states (i.e., E419, Q420, F422, L423, K425, R430, H431, L454, and K455) or insufficient coupling data (e.g., severe line broadening) to define unambiguously the rotamer state of residues W427, R436, F438, N443, K445, and W449. In comparison, severe line broadening was not as prevalent in the peptide-bound state with coupling values indicative of distinct rotamer states being adopted for residues W427, R430, N434, F438, D441, H442, N443, K445, W449, D451, and R453 (Figure S5).

Superposition of the 15 lowest-energy models of the Nedd4-1 WW3* domain is shown in Figure 3A. The backbone atom coordinate root-mean-square deviation (rmsd) was (0.71 ± 0.19) Å for the backbone heavy atoms

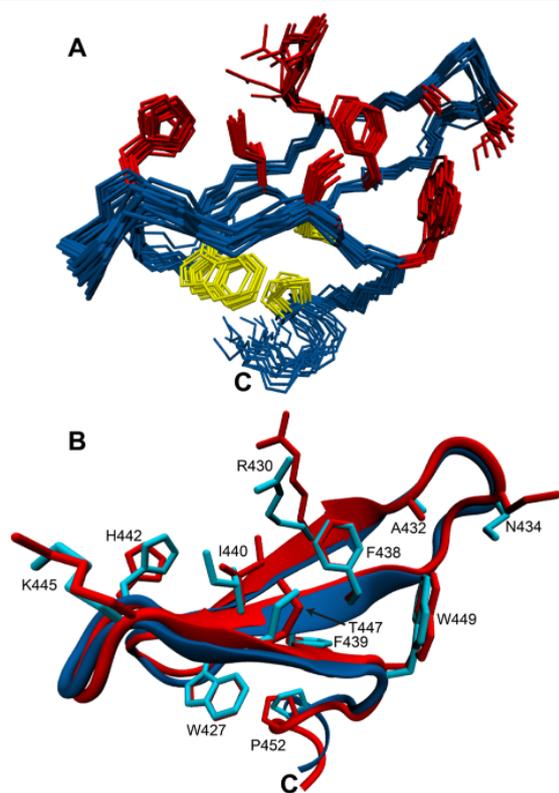


Figure 3. Solution structure of the Nedd4-1 WW3* domain. (A) Superposition of the 15 lowest-energy structures (423–457). The backbone of the WW domain is colored blue. There are three antiparallel β -strands between W427 and H431 (β_1), R436 and D441 (β_2), and T446 and T448 (β_3). Side chains of residues that are part of the conserved hydrophobic core and involved in peptide binding are colored yellow and red, respectively. (B) Superposition of the lowest-energy structures of the WW3* domain in the apo (blue) and peptide-bound (red) states. The side chains of residues involved in peptide binding and the hydrophobic core are labeled. (C) Close-up view of the peptide binding site showing the interaction between the peptide backbone (yellow) and the protein side chains (red).

in the ordered (core) region (423–457) and (0.99 ± 0.16) Å for all heavy atoms over the same region. Structural statistics are listed in Table 1. The structure is similar to determined structures of apo WW domains (Figure S6A),^{20,73} including the Nedd4 WW domains from *Drosophila* (backbone rmsd of 1.19 Å),²⁹ rat (0.84 Å),²⁴ and human (0.79 Å)³¹ in complex with their cognate peptides (Figure S6B), and to the recently determined Nedd4-1 apo-WW3* crystal structure (0.65 Å) (Figure S6C).⁷⁴ As shown in Figure S6C, the backbone conformation and side chain orientations of residues involved in peptide recognition and domain stability are very similar between the solution and crystal structures, with only R430 and I440 showing rotamer state disparity (see description below).

The peptide binding epitope of the Nedd4-1 WW3* domain can be divided into three regions: the “XP” groove (W449, F438, T447, and A432), a hydrophobic pocket (F438, I440, and R430), and the tyrosine binding pocket (H442, K445, and I440) (Figure 3). Despite the small difference in the backbone rmsd between the peptide-bound and apo states of 0.79 Å, particular side chains of residues involved in peptide recognition adopt different conformations between the two structural states (Figure 3B).

The side chain of R430 is displaced by ~ 2.3 Å (as measured using the C_γ atom) upon peptide binding to facilitate interaction with L647', and this is in accord with the displacement observed for R430 in the Nedd4-1 WW3*–ARRDC3 complex.⁷⁴ The side chain χ_1 angle of I440 changes from *trans* in the apo form to *gauche*⁺ in the peptide-bound state (Figure 3B). In contrast, in the crystal structure of the apo-WW3* domain (Figure S6C), the side chain χ_1 angle of I440 changes from $\pm 60^\circ$ (there are two structures in the asymmetric unit of the apo state) to *gauche*⁺ in the peptide-bound state. Thus, the structural data of the apo-WW3* domain suggest that the side chain of I440 adopts all three χ_1 rotamer states, but in the complex, the side chain adopts predominantly the *gauche*⁺ rotamer. The χ_2 angle of I440 adopts the commonly found *trans* state in both apo and peptide-bound states, which is supported by the relationship between the $C^{\delta 1}$ chemical shift of 16.4 ppm and rotamer state.⁷⁵ As previously reported,⁷⁴ these structural changes create the hydrophobic binding pocket for L647', which comprises the $\gamma 1$ methylene and $\delta 1$ methyl groups of I440 and the aliphatic side chain of R430. For Y644' binding, peptide interaction leads to the displacement of residues H442 and K445 by ~ 1.5 Å to facilitate Y644' binding, and the back face of this site is formed by the $\gamma 2$ methyl group of I440.

The side chain orientations of F438 and W449 (XP pocket) show only minor conformational differences between the apo and peptide-bound states (Figure 3B). In the structure of the apo-WW3* domain, the side chain of R436 (side chain not shown in Figure 3) packs across the underside of the indole ring of W449, as supported by NOEs between the $H^\beta/H^\gamma/H^\delta$ and H^ϵ atoms of R436 and the $H^{\eta 2}$ and $H^{\epsilon 3}$ atoms of W449, respectively. Similar conformations are present in the crystal structures of the apo state (PDB entry 4N7F) and the WW3*–ARRDC3 peptide complex (PDB entry 4N7H), and the solution structure of the *Drosophila* WW3*–Comm peptide complex (PDB entry 2EZ5). This stacking arrangement restricts the W449 side chain conformation and thus restrains opening of the XP pocket.

To probe the conformational state of the XP pocket, 100 ns MD simulations were performed. In support of the structural data, no fluctuations in the χ_1 torsion angle in either the

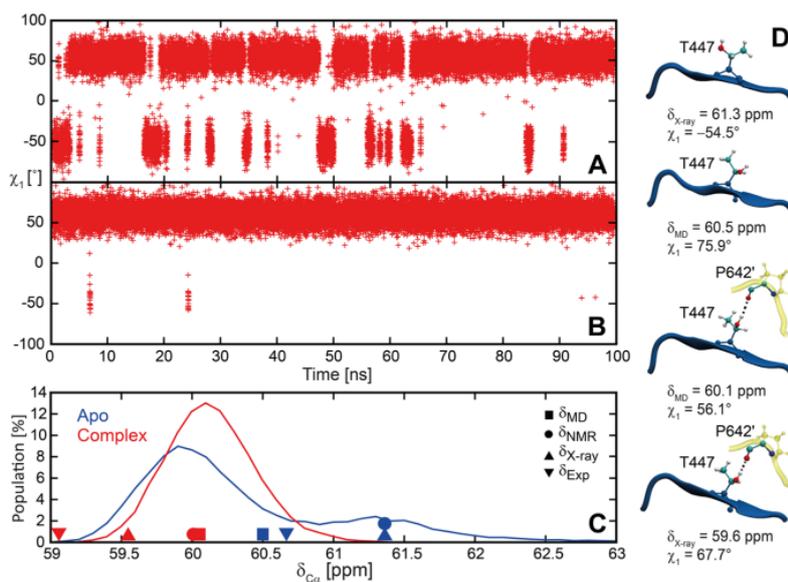


Figure 4. χ_1 torsion angle values and chemical shift predictions of T447 $^{13}\text{C}\alpha$ atoms in 100 ns MD simulations of the WW3* domain in the apo and peptide-bound states. Values of the χ_1 torsion angle of T447 in the apo-WW3* domain (A) and the WW3*–peptide complex (B) for snapshots saved every 4 ns of the MD trajectories. (C) The normalized distribution of the SPARTA+–predicted shifts of T447 $^{13}\text{C}\alpha$ from the trajectories of the apo (blue) and complex (red) states. Chemical shifts were predicted for snapshots saved every 4 ns. The average values of the SPARTA+ predictions over the entire MD trajectory (square) and obtained from the lowest-energy NMR (circle) and X-ray (triangle) structures for the apo (blue) and WW3*–peptide complex (red) states are shown. The experimentally determined values (inverted triangles) are shown for the apo (blue) and complex (red) states. (D) Conformation of T447 observed in the X-ray structures of the hNedd4–1 WW3* domain in the apo and peptide-bound states. Representative structural snapshots of T447 at the average SPARTA+ $^{13}\text{C}\alpha$ chemical shift values for the apo and peptide-bound states. The dotted line shows the presence of the T447 $\text{OH}^{\text{T1}}\cdots\text{O}^{\text{H}}\text{C P642}'$ H-bond.

peptide-free or -bound state were observed for F438 and W449 over the MD simulation period (Figure S7). Furthermore, measured φ and ψ torsion angles for each 4 ps snapshot taken (data not shown) showed that backbone structural fluctuations were small and the backbone conformation was similar between the apo and peptide-bound states.

In Figure 4A, the χ_1 of the T447 side chain, which forms the front face of the XP pocket (Figure 3), fluctuates between *gauche*⁺ and *gauche*[−] throughout the course of the 100 ns simulation for the apo state. In contrast, in the peptide-bound state (Figure 4B), the χ_1 of T447 adopts solely the *gauche*[−] rotamer during the simulation because the formation of the H-bond between the side chain OH^{T1} group of T447 and the backbone carbonyl group of P642' stabilizes this particular rotamer (Figure 4D). Figure 4C shows the normalized distribution of the $^{13}\text{C}\alpha$ chemical shift predictions of T447 determined by SPARTA+ for each 4 ps snapshot taken, together with representative conformations from the MD trajectory and the X-ray structure (Figure 4D). The distribution of the T447 $^{13}\text{C}\alpha$ shift indicates that the presence of a bimodal equilibrium distribution of different χ_1 rotamers produces a prediction of the experimental chemical shift more accurate than that of the rotamer observed in either the NMR or crystal structure for the apo-WW3* domain. Thus, the MD simulations show that the side chain of T447 in the apo state fluctuates between nonligand binding-competent states and ligand binding-competent states, and upon peptide binding, the side chain rotamer fluctuations of T447 are quenched and the side chain adopts solely ligand binding-competent states.

^{15}N Dynamics of the WW3* Domain Reveals That Chemical Exchange Processes Are Quenched in the

Peptide-Bound State. In a previous study,³¹ we reported a dynamics analysis of the Nedd4–1 WW3*– α -hENaC peptide complex with the observation of R_{ex} contributions for a number of residues. However, closer inspection of the data revealed that the sample contained inadvertently a minor population of the apo state. Nonetheless, the original set of data provided impetus for this study, showing that microsecond to millisecond dynamics present in the WW domain is likely to be crucial in peptide recognition. Thus, relaxation data of the complex have also been recorded. Nuclear ^{15}N spin relaxation data acquired at 14.1, 18.8, and 21.2 T (Figure S8) were used to quantify picosecond to nanosecond dynamics of the WW3* domain using the model-free formalism. In contrast to the similar R_1 values between the two states (Figure S8), the overall average R_2 rates for residues of apo-WW3* are noticeably higher than the R_2 rates derived from the peptide-bound WW3* domain, indicative of R_{ex} contributions to the R_2 rates for residues of the domain in the apo state. These high R_2 values for the apo WW3* domain could arise from a self-association equilibrium, as observed for an SH2 domain⁴⁸ and the FBP28 WW domain;⁷⁶ however, average R_2 values for samples at 250 and 1300 μM , after excluding residues that undergo fast internal motion and/or conformational exchange,⁵⁰ were (7.40 ± 0.10) and (7.59 ± 0.11) s^{-1} ($n = 18$), respectively, and are not significantly different. The average $\{^1\text{H}\}$ – ^{15}N heteronuclear NOE values for the core region (423–457) of the Nedd4–1 WW3*– α -hENaC complex were 0.70 ± 0.03 (14.1 T), 0.75 ± 0.03 (18.8 T), and 0.74 ± 0.03 (21.2 T), whereas lower values of 0.60 ± 0.02 (14.1 T), 0.70 ± 0.03 (18.8 T), and 0.72 ± 0.03 (21.2 T) were observed for the apo-WW3* domain.

The rotational diffusion tensor of the apo-WW3* domain is best represented by an ellipsoid described by the isotropic component for diffusion $D_{\text{iso}} = (3.87 \pm 0.03) \times 10^7 \text{ s}^{-1}$, the anisotropy of diffusion $D_a = (4.59 \pm 0.15) \times 10^7 \text{ s}^{-1}$, the rhombicity $D_r = 0.24 \pm 0.02$, and the three Euler angles $\alpha = (132.5 \pm 1.6)^\circ$, $\beta = (78.8 \pm 1.4)^\circ$, and $\gamma = (95.8 \pm 1.4)^\circ$. The global rotational correlation time (τ_m) of the domain was calculated to be 4.30 ns, which is in reasonable agreement with the HYDRONMR-determined value of 4.8 ns using the three-dimensional structure and an atomic element radius of 3.1 Å.⁷⁷ The diffusion tensor of the Nedd4-1 WW3*- α -hENaC peptide complex is best represented by a prolate spheroid described by $D_{\text{iso}} = (3.97 \pm 0.02) \times 10^7 \text{ s}^{-1}$, $D_a = (5.43 \pm 0.99) \times 10^6 \text{ s}^{-1}$, $D_{\parallel}/D_{\perp} = 1.14 \pm 0.03$, the polar angle $\theta = (102.2 \pm 6.9)^\circ$, and azimuthal angle $\varphi = (98.3 \pm 8.9)^\circ$. The τ_m was determined to be 4.20 ns, suggesting that peptide binding causes a slight compaction of the WW3* domain. The data for 34 of a possible 37 native non-proline residues of the apo-WW3* domain construct could be measured and fitted satisfactorily to one of the 10 models (Table S5): 1 (m2; S^2 and τ_e), 16 (m4; S^2 , τ_e , and R_{ex}), 4 (m5; S^2 , S_f^2 , and τ_e), 4 (m6; S^2 , τ_e , S_f^2 , and τ_s), 8 (m7; S^2 , S_f^2 , τ_e , and R_{ex}), and 1 (m8; S^2 , τ_e , S_f^2 , τ_s , and R_{ex}). For the WW3* domain in complex with the α -hENaC peptide, 36 residues could be fitted (Table S6): 3 (m1; S^2), 17 (m2), 1 (m3; S^2 and R_{ex}), 3 (m4), 6 (m5), and 6 (m6). The majority of residues (58%) in the complex were satisfactorily fit to motional models with one or two parameters (m1–m3). In contrast, only one residue was assigned to these models in the apo state with 33 spins assigned to models that contained three or more parameters (m4–m8), with 13 spins described by models m6 to m8. This result shows that in the apo state backbone motions occur over a wide time scale and the motional characteristics of the WW3* residues are more multifaceted in the apo state than in the complex state. An overview of the relaxation analysis is presented in Figure 5.

The N- and C-termini of the WW3* domain in both states are highly mobile with S^2 values of ≤ 0.50 , whereas the overall average S^2 values for residues excluding the two termini (i.e., residues 423–455) are 0.79 ± 0.04 (apo) and 0.84 ± 0.02 (complex), indicating that peptide binding restricts the amplitude of motions across the WW3* domain. Residues in the β -strands show picosecond to nanosecond motions with amplitudes lower than the averages with S^2 values of 0.80 ± 0.02 (apo) and 0.85 ± 0.02 (complex) for β_1 (W427–H431), 0.82 ± 0.03 (apo) and 0.88 ± 0.02 (complex) for β_2 (R436–D441), and 0.79 ± 0.02 (apo) and 0.89 ± 0.02 (complex) for β_3 (T446–T448). While residues in all three β -strands show lower-amplitude motions (i.e., greater rigidity) upon peptide recognition, it is strand β_3 that undergoes the largest reduction in mobility (i.e., $\Delta S^2 = 0.10$), indicating that peptide binding affects the fast time scale motions of this strand the most. Experimental and MD simulations have shown that strand β_3 forms last (or dissociates first from the core fold) during the folding (unfolding) of the FBP28 (formin binding protein 28)^{78,79} and YAP (yes kinase-associated protein)⁸⁰ WW domains. The greater flexibility of strand β_3 observed herein might reflect the lower-energy barrier for this strand to sample other non-native conformers. Such motion might promote peptide recognition because T447 and W449 are located within and at the end of strand β_3 .

In a dynamics study of the Pin1 WW domain binding of the cognate peptide, Cdc25 gave rise to a noticeable increase in the S^2 values for residues in loop I (average $\Delta S^2 = 0.19$) but not

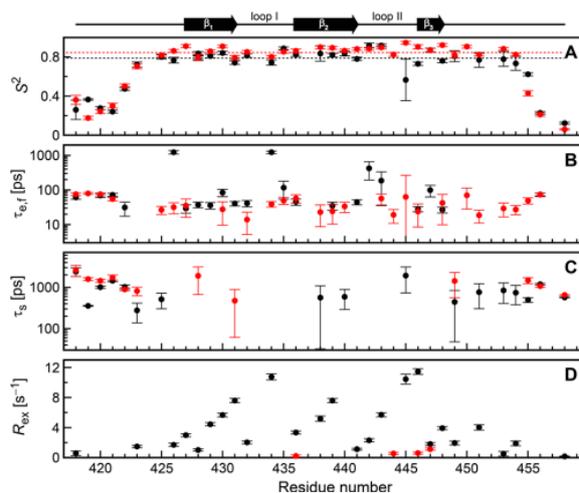


Figure 5. Summary of the model-free parameters of residues of the hNedd4-1 WW3* domain (black) and hNedd4-1 WW3*- α -hENaC complex (red) obtained using the extended model-free formula with the model selection approach. The global rotation of the WW3* domain in both states was determined to be asymmetric. (A) Lipari-Szabo order parameters, S^2 ($=S_f^2 \times S_c^2$). The black dashed lines represent the average S^2 values [0.79 ± 0.04 (apo) and 0.84 ± 0.02 (complex)] over the ordered region (L423–K455) of the WW3* domain. (B) Effective fast internal correlation time ($\tau_e = \tau_f$), (C) slow effective internal correlation time ($\tau_e = \tau_s$), and (D) apparent chemical exchange contribution, R_{ex} , to relaxation rate R_2 . A schematic representation of the secondary structure elements is shown at the top.

strand β_3 (average $\Delta S^2 = 0.03$), indicative of the enhanced rigidity of loop I upon peptide binding.³⁰ Moreover, this study showed that reducing the length of loop I from six to five residues reduced the flexibility for subnanosecond motions of the backbone NH groups concomitant with reduced peptide affinity, suggesting that the mobility of this loop augments peptide affinity. Such dramatic changes in loop I dynamics for the hNedd4-1 WW3* domain were not observed (Figure 5A), because loop I is a more structured four-residue type I G1 bulge turn. In accord with loop I dynamic behavior of the Pin1 WW domain, the greater subnanosecond motions of strand β_3 of apo-WW3* might function in peptide affinity.

The change in free entropy, i.e., conformational entropy (ΔS_{conf}),⁸¹ and excluding other entropic contributions such as solvation that are not examined herein, resulting from changes in the S^2 values was estimated to be an upper bound value of $T\Delta S_{\text{conf}} = -17.53 \text{ kJ mol}^{-1}$ at 25 °C (excluding outlier K445), which is comparable to the $T\Delta S_{\text{A}}^{\text{pp}}$ measured by ITC for this binding reaction (Table S4). Thus, these small restrictions in backbone motion on the picosecond to nanosecond time scale upon the WW3* domain binding its cognate peptide might account for part of the unfavorable association entropy term derived from the ITC fits. Additional restriction in side chain motion upon peptide binding is also likely to contribute to changes in entropy, as observed for T447 in the MD simulations (Figure 4). Moreover, relaxation parameters obtained for the W422 and W449 indole NH^{e1} groups were also fit by relax.⁶⁰ For W422, the S^2 values in the apo and complex state were indistinguishable with values of 0.675 ± 0.016 (apo) and 0.678 ± 0.012 (complex). In contrast, peptide recognition by W449 leads to the NH^{e1} group S^2 value increasing from 0.640 ± 0.012 (apo) to 0.717 ± 0.012

(complex), which translates into a change in $T\Delta S_{\text{conf}}$ of -0.67 kJ mol^{-1} at 25 $^{\circ}\text{C}$, showing that peptide binding restricts picosecond to nanosecond time scale motions of this side chain.

Twenty-five residues of the WW3* domain have contributions to the effective ^{15}N R_2 values due to chemical or conformational exchange on the microsecond to millisecond time scale, R_{ex} , between 0.2 and 11.5 s^{-1} (Figure 5D). The largest R_{ex} contributions determined by model-free analysis were observed for residues N434 ((10.8 ± 0.4) s^{-1}), K445 ((10.5 ± 0.7) s^{-1}), and T446 ((11.5 ± 0.4) s^{-1}), which are located in loop I, loop II, and strand β_3 , respectively, indicative of the widespread R_{ex} observed for residues of the apo-WW3* domain. Upon peptide binding, the R_{ex} contributions of the hNedd4–1 WW3* residues are quenched with only residues R436, T444, T446, and T447 showing small R_{ex} contributions [0.2 – 1.1 s^{-1} (Figure 5D)]. This quenching of R_{ex} suggests that maybe only a subensemble of Nedd4–1 WW3* conformations is ligand binding-competent.

CPMG Relaxation Dispersion Analysis of the Apo and Complex Forms of the WW3* Domain Reflects a Coupled Folding–Binding Equilibrium. To investigate the microsecond to millisecond motions in more detail, ^{15}N and $^1\text{H}_\text{N}$ single-quantum CPMG relaxation dispersion analysis was conducted at 600 and 900 MHz. Thirty residues exhibited relaxation dispersions for the WW3* domain in the apo state over the temperature range of 5 – 25 $^{\circ}\text{C}$, whereas relaxation dispersions were essentially absent in the complex state (Figures S9 and S10), supporting the notion described above that peptide binding quenches chemical exchange. Initially, relaxation dispersion profiles for the 30 residues were fit simultaneously by a two-site exchange model using only the 5 $^{\circ}\text{C}$ data where exchange is the slowest. Subsequently, data sets at 10 and 15 $^{\circ}\text{C}$ were added successively to the fitting routine. Although accurate data fitting was increasingly difficult at 20 and 25 $^{\circ}\text{C}$ because the exchange rate was in the fast exchange regime, negligible changes, i.e., within derived errors, were observed in the fitted parameters at the lower temperatures when these two data sets were included in the fit routine. Thus, relaxation dispersion profiles for the 30 residues at the five temperatures were fit simultaneously by a two-site exchange model (Table 2). The results of the fitting showed that residues undergo a conformational exchange process with a k_{ex} of (3639 ± 68) s^{-1} at 5 $^{\circ}\text{C}$ increasing to (13230 ± 372) s^{-1} at 20 $^{\circ}\text{C}$. The population of the excited state increased from $(2.24 \pm 0.04)\%$ at 5 $^{\circ}\text{C}$ to $(6.89 \pm 0.42)\%$ at 20 $^{\circ}\text{C}$. There is an

Table 2. Exchange Parameters Extracted from Fitting a Two-Site Model of Chemical Exchange to CPMG Relaxation Dispersion Data Recorded on the Apo-hNedd4–1 WW3* Domain at Multiple Temperatures^a

T ($^{\circ}\text{C}$)	$k_{\text{ex,FU}}$ (s^{-1})	P_{U} (%)
5	3638 ± 68	2.24 ± 0.04
10	5282 ± 142	2.67 ± 0.12
15	6381 ± 203	2.78 ± 0.15
20	13230 ± 372	6.89 ± 0.42
25	11917 ± 993	4.67 ± 0.73

^aAll temperature-dependent data were fit simultaneously assuming an $\text{F} \leftrightarrow \text{U}$ exchange model, where F is the folded state and U is the unfolded (excited) state. Exchange parameters at 20 and 25 $^{\circ}\text{C}$ have higher uncertainties because the exchange process is in the fast exchange regime.

excellent correlation between the ^{15}N and $^1\text{H}_\text{N}$ chemical shifts of the excited state reconstructed from the CPMG fit and random coil chemical shifts⁸² with low rmsd values of 1.17 ppm for ^{15}N , which is similar to the unfolding equilibrium of SH3 domain mutants,⁵⁴ and 0.37 ppm for $^1\text{H}_\text{N}$ (Figure 6). Clearly, the apo-WW3* domain exists in equilibrium between the native state, which must be the binding-competent state given the relatively minor tertiary structure changes upon binding, and a random coil-like unfolded state, and binding of the peptide shifts the equilibrium to predominantly the ligand binding-competent state. Such a linking of equilibria between unfolded–folded states and peptide-free and peptide-bound states is reminiscent of a coupled folding–binding mechanism.

The hNedd4–1 WW3* Domain Unfolds in a Single Broad Transition with Low Thermal Stability. The CD spectrum of the hNedd4–1 WW3* domain (Figure 7A) shows the characteristic profile observed for WW domains.⁸³ The positive CD band with a maximum at ~ 230 nm is due to local interactions involving aromatic side chains. In the WW3* structure, this arises from the packing of W427 into the hydrophobic core that also consists of L423, F439, and P452 (Figure 3A). The low-wavelength region of the CD spectrum is characteristic of β -II proteins, which have a β -sheet/random coil ratio lower than those of typical β -sheet proteins that have CD spectra with a characteristic negative band at ~ 218 nm.⁸⁴ Moreover, there will be a CD contribution from the unfolded WW3* domain, as determined from the CPMG analysis ($\sim 7\%$). The addition of the α -hENaC peptide to WW3* results in clear changes in the CD spectrum. In particular, the stronger intensity of the CD band at 230 nm is due to the peptide folding to form a poly(Pro)II helix and stacking of the P641' side chain between the side chains of W449 and F438 (i.e., XP groove).⁸⁴ Additionally, the negative band shifts from ~ 208 to ~ 218 nm, indicative that upon peptide binding there is negligible unfolded domain present and thus the majority of the WW domain adopts a well-folded, antiparallel β -sheet structure.

Figure 7B shows the thermal denaturation profile of the apo-WW3* domain monitored by CD spectroscopy at 230 nm. The unfolding profile shows a cooperative thermal transition, which is very broad as expected for a domain this size. The width of the transition makes it more difficult to fit the linear pre- and posttransition baselines. To overcome this difficulty and obtain a more accurate picture of the thermodynamics of denaturation, we used our knowledge of the unfolded state population at low temperatures from the CPMG analysis and simultaneously fitted the CD thermal denaturation profile and fractions of the unfolded domain obtained from the relaxation dispersion analysis at various temperatures (Table 2) using a two-state equilibrium model. The solid line in Figure 7B represents the best fit to the data. The fit shows that the data can be described by a two-state model with a temperature midpoint (T_{m}) = (49.7 ± 0.3) $^{\circ}\text{C}$, an enthalpy of folding (ΔH_{F}) = $-(103.2 \pm 0.5)$ kJ mol^{-1} , an entropy of folding (ΔS_{F}) = (-0.32 ± 0.00) $\text{kJ mol}^{-1} \text{K}^{-1}$, and a global (χ^2 values from both CD and relaxation dispersion data fitting) reduced χ^2 value of 2.9 with an f_{CPMG} value of 1.4 (see Experimental Procedures). The thermodynamic values are similar to those reported for other WW domains,⁸⁵ including the WW4 domain from Nedd4.⁸⁶ The normalized fraction unfolded profile is presented in Figure 7C, along with the relaxation dispersion-derived fraction unfolded data at five temperatures. The unfolded state is populated to $\sim 20\%$ at physiological temperature (37 $^{\circ}\text{C}$), which is significantly higher than those of the more thermally stable

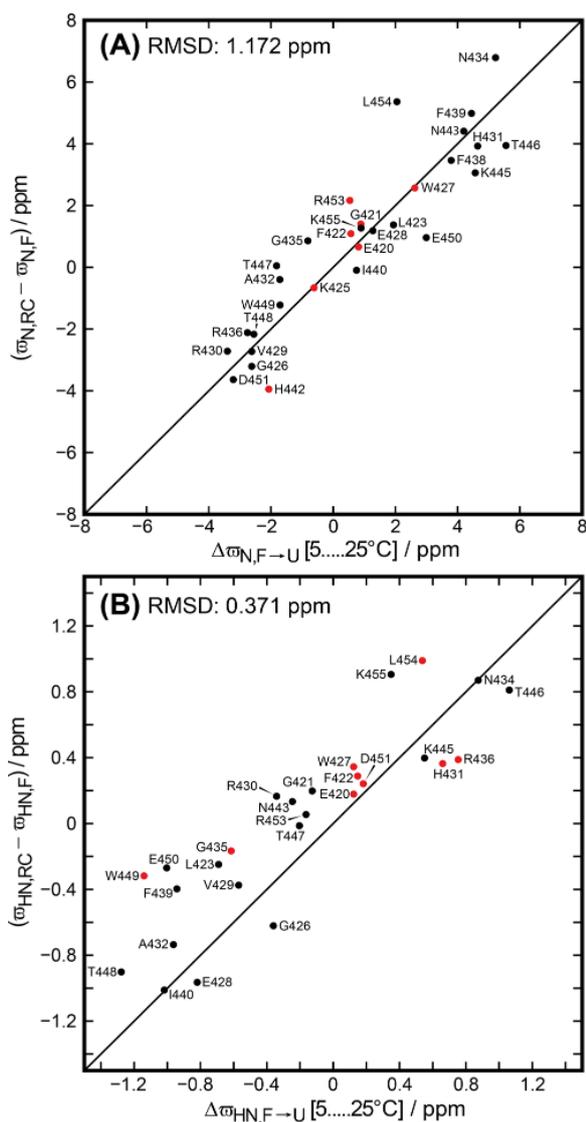


Figure 6. Comparison of chemical shift differences between excited and folded states of the hNedd4-1 WW3* domain extracted from simultaneous fits of CPMG data at multiple temperatures using a two-site model of $F \leftrightarrow U$ exchange ($\Delta\varpi_{F \rightarrow U} = \varpi_U - \varpi_F$), with differences between random coil, ϖ_{RC} , and folded, ϖ_F , (A) ^{15}N and (B) $^1\text{H}_\text{N}$ shift values. The excellent correlation between the ^{15}N and $^1\text{H}_\text{N}$ chemical shifts extracted for the excited state with random coil values identifies this state as unfolded. Data points are labeled with assignment information. Because CPMG experiments yield only the magnitude, $|\Delta\varpi_{F \rightarrow U}|$, the absolute signs of the chemical shift differences between the exchanging states were obtained by comparing 2D ^1H - ^{15}N HSQC and 2D ^1H - ^{15}N HMQC spectra at two different fields.^{58,59} Signs of data points colored red could not be determined because of resonance overlap or insufficient signal to noise and were inferred from the random coil chemical shifts, ϖ_{RC} .

Pin1 and FBP WW domains, where the denatured state is barely populated at $\leq 37^\circ\text{C}$,^{85,87} but similar to the simulated unfolded profile of the Nedd4 WW4 domain studied at pH 6.0.⁸⁶

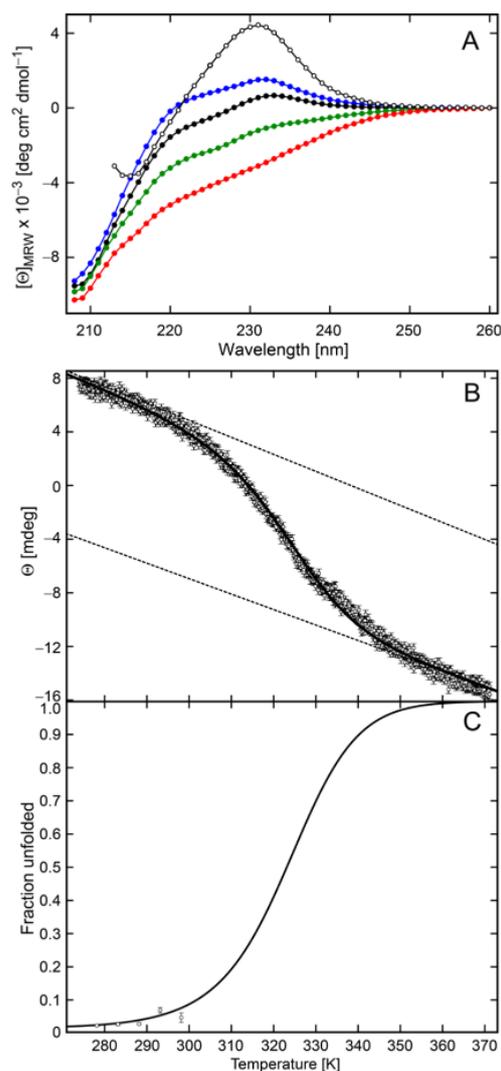


Figure 7. CD thermal denaturation analysis of the hNedd4 WW3* domain. (A) Temperature dependency of the far-UV CD spectra of hNedd4-1 WW3* at 0 (blue), 25 (black), 50 (green), and 75 °C (red). The free peptide corrected hNedd4-1 WW3*- α -hENaC complex (1:5 ratio) spectrum is also shown (\circ). Data collection was terminated at 213 nm (complex) and 208 nm (apo) because of the high tension voltage (dynode) registered at lower wavelengths. (B) CD thermal denaturation profiles recorded at 230 nm of apo-hNedd4-1 WW3* (\circ). The dotted lines represent the baselines obtained from the fitting routine for the native and denatured states of the apo thermal denaturation profile. The solid line through experimental data corresponds to the best fit of the data by the two-state equilibrium model. (C) Normalized equilibrium thermal denaturation curves of the WW3* domain in the apo state. Data points between 5 and 25 °C are the fractions of the unfolded domain obtained from relaxation dispersion analysis (Table 2).

The thermal denaturation result supports the chemical shift changes observed in the 2D ^1H - ^{15}N HSQC spectra of the apo-hNedd4-1 WW3* domain between 5 and 37 °C (Figure S11). Here, the rmsds between random coil backbone ^{15}N chemical shifts⁸² and the chemical shifts of the resonances at 5, 25, and 37 °C were determined to be 3.3, 2.9, and 2.8 ppm, indicating

that as the temperature increases the population of the unfolded domain contributing to the NMR chemical shift values increases. Nonetheless, these rmsd values are noticeably larger than the value of ~ 1 ppm reported for urea-denatured apomyoglobin⁸⁸ and unfolded mutant SH3 domains,⁵⁴ demonstrating that the folded apo-hNedd4–1 WW3* domain is the predominant population at ≤ 37 °C and the chemical shift changes are primarily due to an increasing percentage of the unfolded state population.

Because the hNedd4–1 WW3* domain is in coupled folding–binding equilibria, the thermodynamics of folding from the CD melt analysis and the apparent thermodynamics of binding from ITC were used to reconstruct ΔH_A and K_d values describing binding of the ligand to the native state using eqs (6–8). On the basis of the unfolded state population of $F_U = 7.5\%$ at 25 °C from the combined CD/CPMG fit (Figure 7), the ΔH_A and K_d values determined were -67.63 kJ mol⁻¹ and 41 μ M at 25 °C, respectively.

DISCUSSION

In this report, the NOE-derived structure of the apo-WW3* domain from Nedd4–1 was found to adopt a fold similar to that of the previously determined structure of the α -hENaC peptide-bound WW3* domain³¹ and in agreement with the crystal structure.⁷⁴ Backbone amide chemical shift perturbations of the WW3* domain upon α -hENaC peptide interaction revealed that resonances from the $\beta 3$ strand undergo the largest chemical shift changes (Figure S3). Moreover, strand $\beta 3$ undergoes the greatest rigidification on the picosecond to nanosecond time scale upon peptide interaction, as determined by S^2 analysis (Figure 5). These observations are congruent with folding studies that showed that the third strand is the first to dissociate,^{78–80,89} suggesting that lower-energy barriers allow this strand to sample other conformers that might facilitate domain folding and peptide recognition.

¹H_N and ¹⁵N CPMG relaxation dispersion analysis showed that the apo-WW3* domain exists in an equilibrium between the ligand binding-competent folded state (ground state) and an unfolded state (Figure 6). This unfolded state was quenched upon peptide binding, leading to negligible R_{ex} contributions to R_2 values (Figure 5 and Figures S9–S10). Thus, the binding of the hNedd4–1 WW3* domain to the peptide is coupled to the folding equilibrium. The derived k_{ex} values from the relaxation dispersion analysis concur with rates of folding of other WW domains, which have been reported to be 10000 s⁻¹ (lifetime of ~ 100 μ s).^{64,78,85} Combination of CPMG relaxation dispersion and CD thermal unfolding analysis revealed that the unfolded state of this WW domain is populated to $\sim 20\%$ at 37 °C. This population is considerably higher than those from previous studies of Pin1 and FBP28 WW domains,^{85,87} where the unfolded state is barely populated at ≤ 37 °C, and explains why this coupled folding and binding equilibrium has not been observed in these two WW domains. This higher population of unfolded species reflects the lower thermal stability of the hNedd4–1 WW3* domain under the conditions of this study compared with those of the Pin1 ($T_m = 59$ °C) and FBP28 ($T_m = 64$ °C) WW domains. The exact biological significance of this low thermal stability and relatively high population of the unfolded hNedd4–1 WW3* domain at physiological temperature are unclear, but as observed for intrinsically disordered proteins,⁹⁰ this feature of WW3* appears to be thermodynamically advantageous with respect to complex formation. In a study of the YAP WW domain,⁸³ mutation of the tryptophan

crucial in maintaining domain integrity shifted the folding equilibrium to predominantly the unfolded state, but addition of the cognate ligand results in PY-ligand induced folding without a dramatic change in the K_d^{pp} compared with that of the more stable wild-type domain. This is an extreme case but highlights the fact that the energy barrier between the unfolded and folded macrostates can be low, which is also observed for the hNedd4 WW4 domain,⁸⁶ and that enthalpy–entropy compensation facilitates ligand binding by coupling the peptide binding event to WW domain folding.⁸³

The observed structural changes in side chain orientations upon ligand binding⁷⁴ show that peptide binding influences the rotamer state adopted by particular side chains. Determination of the side chain rotamer populations by comprehensive ³J analysis⁴⁰ or residual dipolar couplings⁹¹ of the χ_1 angle should provide insights into whether changes in side chain rotamer states play a role in peptide recognition. Interestingly, the side chain rotamer for the peptide binding residue T447 positioned on the front face of the XP pocket of apo-WW3* follows a bimodal distribution in the MD simulations (Figure 4), with the major state resembling the side chain rotamer of the peptide-bound state and the minor state the apo conformation present in the experimentally determined structures (PDB entries SAHT and 4N7F). Thus, T447 rotation between two distinct rotamer states for the apo state suggests a conformational selection approach for peptide binding. Two other key residues, W449 and F438, of the XP pocket showed no jumps between distinct rotameric states in the MD simulations (Figure S7). However, the S^2 value of the side chain indole group of W449 increased upon peptide interaction, and the MD simulations of W449 also showed that the ¹³C α chemical shift distribution narrowed in the simulation of the peptide-bound WW3* domain without a major shift in the δ_{Ca} maximum, suggesting that motions sampled by this residue are further restricted when the peptide is bound (Figure S7F). Future efforts directed at studying side chain dynamics of the WW3* domain should reveal how side chain motion facilitates peptide recognition and the formation of a stable hNedd4–1 WW3*– α -hENaC complex.

Coupled folding–binding mechanisms for protein–protein interactions have been presented in detail in intrinsically disordered proteins (IDPs) and proteins with intrinsically disordered regions (IDRs).⁹⁰ IDPs and IDRs have been found to interact with their target binding partner by either an induced fit or conformational selection binding mechanism;⁹² depending on the flux,⁹³ i.e., rate constants and concentrations of all species, a combination of the two mechanisms is plausible. Although it is conceivable that flux through both pathways exists for the coupled folding–binding mechanism of the WW3* domain, we have no evidence of an encounter complex (UL) between the peptide (L) and the unfolded state (U) being populated significantly, because no additional resonances in the 2D ¹H–¹⁵N HSQC spectra of the complex indicative of a population of UL were observed. Thus, even though our data for coupled folding and binding equilibria are consistent with a conformational selection mechanism, we cannot rule out the possibility that folding and peptide binding also follow an induced fit mechanism without an in-depth kinetic study, which is beyond the scope of this work.

The lower stability and folding energy barrier of the Nedd4–1 WW3* domain might be crucial for self-regulation of Nedd4, where WW domain flexibility may facilitate the self-regulatory interaction between the C2 and HECT domains.^{11,12} Whether

the thermodynamic properties of the isolated WW3* domain are retained in the presence of the neighboring domains or whether there is coordination between the four WW domains remains unresolved and represents a worthy future study. Nonetheless, the isolated WW3* domain from the hNedd4-1 domain has been shown to be sufficient to bind and trigger ubiquitination of hENaC and assist in the viral budding process of Ebola and Marburg viruses.⁹⁴ Because the results presented herein provide a detailed picture of the stabilizing interactions and dynamics of the binding interface, such information could be used to investigate the interactions of previously identified Nedd4-1 specific substrates via homology modeling of the complexes with the Nedd4-1 WW3* structure. This modeling might aid in the design of mutagenesis experiments for functional assays and in the development of WW constructs that manipulate protein interaction networks. Such efforts should aid in the design of therapeutics that target hNedd4-1 dysfunction.⁹⁵

■ ASSOCIATED CONTENT

■ Supporting Information

Supporting materials may be accessed free-of-charge online at: The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.biochem.5b01028](https://doi.org/10.1021/acs.biochem.5b01028).

NMR experiments used for structure determination (Table S1), NMR relaxation experiment parameters (Tables S2 and S3), ITC results (Table S4), and *relax* results for apo-WW3* (Table S5) and the WW3* domain in complex with its cognate peptide (Table S6), ITC results (Figure S1), 2D ¹H-¹⁵N HSQC spectra (Figure S2), chemical shift perturbation analysis (Figure S3), titration curves with NMR data (Figure S4), ³J_{αβ} and ³J_{Nβ} couplings (Figure S5), superposition of WW structures (Figure S6), MD simulation analysis of F438 and W449 (Figure S7), backbone ¹⁵N NMR relaxation data (Figure S8), CPMG relaxation dispersion analysis of G426 (Figure S9) and N443 (Figure S10), and 2D ¹H-¹⁵N HSQC temperature series (Figure S11) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*ICS-6 (Strukturbiochemie), Forschungszentrum Jülich, Wilhelm-Johnen-Str., 52425 Jülich, Germany. Phone: ++49 2461 61 9487. Fax: ++49 2461 61 2023. E-mail: a.dingley@fz-juelich.de.

Funding

The authors thank the Maurice Wilkins Centre for Molecular Biodiscovery for funding peptide synthesis.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge access to the Jülich-Düsseldorf Biomolecular NMR Center that is jointly run by the Forschungszentrum Jülich and Heinrich Heine University Düsseldorf. We thank Dr. Tet Verne Lee (School of Biological Sciences, The University of Auckland) for providing the WW3* domain bacterial overexpression construct.

■ ABBREVIATIONS

CPMG, Carr-Purcell-Meiboom-Gill; FBP28, formin binding protein 28; hENaC, human epithelial Na⁺ channel; HSQC, heteronuclear single-quantum coherence; IDPs, intrinsically disordered proteins; IDRs, intrinsically disordered regions; ITC, isothermal titration calorimetry; hNedd4-1, neuronal precursor cell expressed developmentally downregulated gene 4-1; Pin1, peptidyl-prolyl isomerase; rmsd, root-mean-square deviation; YAP, yes kinase-associated protein.

■ REFERENCES

- (1) Hershko, A., Ciechanover, A., and Rose, I. A. (1979) Resolution of the ATP-dependent proteolytic system from reticulocytes: a component that interacts with ATP. *Proc. Natl. Acad. Sci. U. S. A.* 76, 3107-3110.
- (2) Clague, M. J., Liu, H., and Urbe, S. (2012) Governance of endocytic trafficking and signaling by reversible ubiquitylation. *Dev. Cell* 23, 457-467.
- (3) Teixeira, L. K., and Reed, S. I. (2013) Ubiquitin ligases and cell cycle control. *Annu. Rev. Biochem.* 82, 387-414.
- (4) Kumar, S., Tomooka, Y., and Noda, M. (1992) Identification of a set of genes with developmentally down-regulated expression in the mouse brain. *Biochem. Biophys. Res. Commun.* 185, 1155-1161.
- (5) Staub, O., Dho, S., Henry, P., Correa, J., Ishikawa, T., McGlade, J., and Rotin, D. (1996) WW domains of Nedd4 bind to the proline-rich PY motifs in the epithelial Na⁺ channel deleted in Liddle's syndrome. *EMBO J.* 15, 2371-2380.
- (6) Staub, O., Yeager, H., Plant, P. J., Kim, H., Ernst, S. A., and Rotin, D. (1997) Immunolocalization of the ubiquitin-protein ligase Nedd4 in tissues expressing the epithelial Na⁺ channel (ENaC). *Am. J. Physiol.* 272, C1871-C1880.
- (7) Yang, B., and Kumar, S. (2010) Nedd4 and Nedd4-2: closely related ubiquitin-protein ligases with distinct physiological functions. *Cell Death Differ.* 17, 68-77.
- (8) Cao, X. R., Lill, N. L., Boase, N., Shi, P. P., Croucher, D. R., Shan, H., Qu, J., Sweezer, E. M., Place, T., Kirby, P. A., Daly, R. J., Kumar, S., and Yang, B. (2008) Nedd4 controls animal growth by regulating IGF-1 signaling. *Sci. Signaling* 1, ra5.
- (9) Mund, T., Lewis, M. J., Maslen, S., and Pelham, H. R. (2014) Peptide and small molecule inhibitors of HECT-type ubiquitin ligases. *Proc. Natl. Acad. Sci. U. S. A.* 111, 16736-16741.
- (10) Rotin, D., Staub, O., and Haguenaer-Tsapis, R. (2000) Ubiquitination and endocytosis of plasma membrane proteins: role of Nedd4/Rsp5p family of ubiquitin-protein ligases. *J. Membr. Biol.* 176, 1-17.
- (11) Mari, S., Ruetalo, N., Maspero, E., Stoffregen, M. C., Pasqualato, S., Polo, S., and Wiesner, S. (2014) Structural and functional framework for the autoinhibition of nedd4-family ubiquitin ligases. *Structure* 22, 1639-1649.
- (12) Wiesner, S., Ogunjimi, A. A., Wang, H. R., Rotin, D., Sicheri, F., Wrana, J. L., and Forman-Kay, J. D. (2007) Autoinhibition of the HECT-type ubiquitin ligase Smurf2 through its C2 domain. *Cell* 130, 651-662.
- (13) Wang, J., Peng, Q., Lin, Q., Childress, C., Carey, D., and Yang, W. (2010) Calcium activates Nedd4 E3 ubiquitin ligases by releasing the C2 domain-mediated auto-inhibition. *J. Biol. Chem.* 285, 12279-12288.
- (14) Escobedo, A., Gomes, T., Aragon, E., Martin-Malpartida, P., Ruiz, L., and Macias, M. J. (2014) Structural basis of the activation and degradation mechanisms of the E3 ubiquitin ligase Nedd4L. *Structure* 22, 1446-1457.
- (15) Mund, T., and Pelham, H. R. (2009) Control of the activity of WW-HECT domain E3 ubiquitin ligases by NDFIP proteins. *EMBO Rep.* 10, S01-S07.
- (16) Bork, P., and Sudol, M. (1994) The WW domain: a signalling site in dystrophin? *Trends Biochem. Sci.* 19, 531-533.

- (17) Chen, H. I., and Sudol, M. (1995) The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules. *Proc. Natl. Acad. Sci. U. S. A.* 92, 7819–7823.
- (18) Hofmann, K., and Bucher, P. (1995) The rsp5-domain is shared by proteins of diverse functions. *FEBS Lett.* 358, 153–157.
- (19) Sudol, M., Bork, P., Einbond, A., Kastury, K., Druck, T., Negrini, M., Huebner, K., and Lehman, D. (1995) Characterization of the mammalian YAP (Yes-associated protein) gene and its role in defining a novel protein module, the WW domain. *J. Biol. Chem.* 270, 14733–14741.
- (20) Macias, M. J., Gervais, V., Civera, C., and Oschkinat, H. (2000) Structural analysis of WW domains and design of a WW prototype. *Nat. Struct. Biol.* 7, 375–379.
- (21) Macias, M. J., Hyvönen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M., and Oschkinat, H. (1996) Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. *Nature* 382, 646–649.
- (22) Zarrinpar, A., and Lim, W. A. (2000) Converging on proline: the mechanism of WW domain peptide recognition. *Nat. Struct. Biol.* 7, 611–613.
- (23) Huang, X., Poy, F., Zhang, R., Joachimiak, A., Sudol, M., and Eck, M. J. (2000) Structure of a WW domain containing fragment of dystrophin in complex with beta-dystroglycan. *Nat. Struct. Biol.* 7, 634–638.
- (24) Kanelis, V., Rotin, D., and Forman-Kay, J. D. (2001) Solution structure of a Nedd4 WW domain-ENaC peptide complex. *Nat. Struct. Biol.* 8, 407–412.
- (25) Verdecia, M. A., Bowman, M. E., Lu, K. P., Hunter, T., and Noel, J. P. (2000) Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nat. Struct. Biol.* 7, 639–643.
- (26) Kay, B. K., Williamson, M. P., and Sudol, M. (2000) The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J.* 14, 231–241.
- (27) Henry, P. C., Kanelis, V., O'Brien, M. C., Kim, B., Gautschi, I., Forman-Kay, J., Schild, L., and Rotin, D. (2003) Affinity and specificity of interactions between Nedd4 isoforms and the epithelial Na⁺ channel. *J. Biol. Chem.* 278, 20019–20028.
- (28) Lott, J. S., Coddington-Lawson, S. J., Teesdale-Spittle, P. H., and McDonald, F. J. (2002) A single WW domain is the predominant mediator of the interaction between the human ubiquitin-protein ligase Nedd4 and the human epithelial sodium channel. *Biochem. J.* 361, 481–488.
- (29) Kanelis, V., Bruce, M. C., Skrynnikov, N. R., Rotin, D., and Forman-Kay, J. D. (2006) Structural Determinants for High-Affinity Binding in a Nedd4 WW3* Domain-Comm PY Motif Complex. *Structure* 14, 543–553.
- (30) Peng, T., Zintsmaster, J. S., Namanja, A. T., and Peng, J. W. (2007) Sequence-specific dynamics modulate recognition specificity in WW domains. *Nat. Struct. Mol. Biol.* 14, 325–331.
- (31) Bobby, R., Medini, K., Neudecker, P., Lee, T. V., Brimble, M. A., McDonald, F. J., Lott, J. S., and Dingley, A. J. (2013) Structure and dynamics of human Nedd4-1 WW3 in complex with the alphaENaC PY motif. *Biochim. Biophys. Acta, Proteins Proteomics* 1834, 1632–1641.
- (32) Grzesiek, S., Stahl, S. J., Wingfield, P. T., and Bax, A. (1996) The CD4 determinant for downregulation by HIV-1 Nef directly binds to Nef. Mapping of the Nef binding surface by NMR. *Biochemistry* 35, 10256–10261.
- (33) Dingley, A. J., Lorenzen, I., and Grotzinger, J. (2008) NMR analysis of viral protein structures. *Methods Mol. Biol.* 451, 441–462.
- (34) Sattler, M., Schleucher, J., and Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* 34, 93–158.
- (35) Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) ¹H, ¹³C and ¹⁵N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* 6, 135–140.
- (36) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6, 277–293.
- (37) Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins: Struct., Funct., Genet.* 59, 687–696.
- (38) Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* 44, 213–223.
- (39) Dux, P., Whitehead, B., Boelens, R., Kaptein, R., and Vuister, G. W. (1997) Measurement of ¹⁵N-¹H coupling constants in uniformly ¹⁵N-labeled proteins: Application to the photoactive yellow protein. *J. Biomol. NMR* 10, 301–306.
- (40) Vajpai, N., Gentner, M., Huang, J. R., Blackledge, M., and Grzesiek, S. (2010) Side-chain χ_1 conformations in urea-denatured ubiquitin and protein G from ³J coupling constants and residual dipolar couplings. *J. Am. Chem. Soc.* 132, 3196–3203.
- (41) Rieping, W., Habeck, M., Bardiaux, B., Bernard, A., Malliavin, T. E., and Nilges, M. (2007) ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* 23, 381–382.
- (42) Brunger, A. T. (2007) Version 1.2 of the Crystallography and NMR system. *Nat. Protoc.* 2, 2728–2733.
- (43) Bernard, A., Vranken, W. F., Bardiaux, B., Nilges, M., and Malliavin, T. E. (2011) Bayesian estimation of NMR restraint potential and weight: a validation on a representative set of protein structures. *Proteins: Struct., Funct., Genet.* 79, 1525–1537.
- (44) Nilges, M., Bernard, A., Bardiaux, B., Malliavin, T. E., Habeck, M., and Rieping, W. (2008) Accurate NMR structures through minimization of an extended hybrid energy. *Structure* 16, 1305–1312.
- (45) Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8, 477–486.
- (46) Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B. r., Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35, W375–383.
- (47) Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graphics* 14, 33–38.
- (48) Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ¹⁵N NMR relaxation. *Biochemistry* 33, 5984–6003.
- (49) Korzhnev, D. M., Skrynnikov, N. R., Millet, O., Torchia, D. A., and Kay, L. E. (2002) An NMR experiment for the accurate measurement of heteronuclear spin-lock relaxation rates. *J. Am. Chem. Soc.* 124, 10743–10753.
- (50) Tjandra, N., Wingfield, P., Stahl, S., and Bax, A. (1996) Anisotropic rotational diffusion of perdeuterated HIV protease from ¹⁵N NMR relaxation measurements at two magnetic fields. *J. Biomol. NMR* 8, 273–284.
- (51) Loria, J. P., Rance, M., and Palmer, A. G. (1999) A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR Spectroscopy. *J. Am. Chem. Soc.* 121, 2331–2332.
- (52) Tollinger, M., Skrynnikov, N. R., Mulder, F. A., Forman-Kay, J. D., and Kay, L. E. (2001) Slow dynamics in folded and unfolded states of an SH3 domain. *J. Am. Chem. Soc.* 123, 11341–11352.
- (53) Ishima, R., and Torchia, D. A. (2003) Extending the range of amide proton relaxation dispersion experiments in proteins using a constant-time relaxation-compensated CPMG approach. *J. Biomol. NMR* 25, 243–248.
- (54) Neudecker, P., Zarrine-Afsar, A., Choy, W. Y., Muhandiram, D. R., Davidson, A. R., and Kay, L. E. (2006) Identification of a collapsed intermediate with non-native long-range interactions on the folding

pathway of a pair of Fyn SH3 domain mutants by NMR relaxation dispersion spectroscopy. *J. Mol. Biol.* 363, 958–976.

(55) Korzhnev, D. M., Neudecker, P., Mittermaier, A., Orekhov, V. Y., and Kay, L. E. (2005) 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, 15602–15611.

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

(57) Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007) *Numerical Recipes in C*, Cambridge University Press, Cambridge, U.K.

(58) Skrynnikov, N. R., Dahlquist, F. W., and Kay, L. E. (2002) Reconstructing NMR spectra of “invisible” excited protein states using HSQC and HMQC experiments. *J. Am. Chem. Soc.* 124, 12352–12360.

(59) Bouvignies, G., Korzhnev, D. M., Neudecker, P., Hansen, D. F., Cordes, M. H., and Kay, L. E. (2010) A simple method for measuring signs of $^1\text{H}^{\text{N}}$ chemical shift differences between ground and excited protein states. *J. Biomol. NMR* 47, 135–141.

(60) d’Auvergne, E. J., and Gooley, P. R. (2006) Model-free model elimination: a new step in the model-free dynamic analysis of NMR relaxation data. *J. Biomol. NMR* 35, 117–135.

(61) d’Auvergne, E. J., and Gooley, P. R. (2008) Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. *J. Biomol. NMR* 40, 107–119.

(62) d’Auvergne, E. J., and Gooley, P. R. (2008) Optimisation of NMR dynamic models II. A new methodology for the dual optimization of the model-free parameters and the Brownian rotational diffusion tensor. *J. Biomol. NMR* 40, 121–133.

(63) Kemmer, G., and Keller, S. (2010) Nonlinear least-squares data fitting in Excel spreadsheets. *Nat. Protoc.* 5, 267–281.

(64) Ferguson, N., Johnson, C. M., Macias, M., Oschkinat, H., and Fersht, A. (2001) Ultrafast folding of WW domains without structured aromatic clusters in the denatured state. *Proc. Natl. Acad. Sci. U. S. A.* 98, 13002–13007.

(65) Cliff, M. J., Williams, M. A., Brooke-Smith, J., Barford, D., and Ladbury, J. E. (2005) Molecular recognition via coupled folding and binding in a TPR domain. *J. Mol. Biol.* 346, 717–732.

(66) Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B., and Lindahl, E. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29, 845–854.

(67) Abascal, J. L. F., and Vega, C. (2005) A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* 123, 234505.

(68) Darden, T., York, D., and Pedersen, L. G. (1993) Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092.

(69) Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.* 103, 8577–8593.

(70) Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Genet.* 78, 1950–1958.

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

(72) Koepf, E. K., Petrassi, H. M., Sudol, M., and Kelly, J. W. (1999) WW: An isolated three-stranded antiparallel beta-sheet domain that unfolds and refolds reversibly; evidence for a structured hydrophobic cluster in urea and GdnHCl and a disordered thermal unfolded state. *Protein Sci.* 8, 841–853.

(73) Kato, Y., Hino, Y., Nagata, K., and Tanokura, M. (2006) Solution structure and binding specificity of FBP11/HYPA WW domain as Group-II/III. *Proteins: Struct., Funct., Genet.* 63, 227–234.

(74) Qi, S., O’Hayre, M., Gutkind, J. S., and Hurley, J. H. (2014) Structural and biochemical basis for ubiquitin ligase recruitment by arrestin-related domain-containing protein-3 (ARRDC3). *J. Biol. Chem.* 289, 4743–4752.

(75) Hansen, D. F., Neudecker, P., and Kay, L. E. (2010) Determination of isoleucine side-chain conformations in ground and excited states of proteins from chemical shifts. *J. Am. Chem. Soc.* 132, 7589–7591.

(76) Ferguson, N., Berriman, J., Petrovich, M., Sharpe, T. D., Finch, J. T., and Fersht, A. R. (2003) Rapid amyloid fiber formation from the fast-folding WW domain FBP28. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9814–9819.

(77) Garcia de la Torre, J., Huertas, M. L., and Carrasco, B. (2000) HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J. Magn. Reson.* 147, 138–146.

(78) Davis, C. M., and Dyer, R. B. (2013) Dynamics of an ultrafast folding subdomain in the context of a larger protein fold. *J. Am. Chem. Soc.* 135, 19260–19267.

(79) Petrovich, M., Jonsson, A. L., Ferguson, N., Daggett, V., and Fersht, A. R. (2006) Phi-analysis at the experimental limits: mechanism of beta-hairpin formation. *J. Mol. Biol.* 360, 865–881.

(80) Ibragimova, G. T., and Wade, R. C. (1999) Stability of the beta-sheet of the WW domain: A molecular dynamics simulation study. *Biophys. J.* 77, 2191–2198.

(81) Yang, D., and Kay, L. E. (1996) Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding. *J. Mol. Biol.* 263, 369–382.

(82) Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) ^1H , ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J. Biomol. NMR* 5, 67–81.

(83) Koepf, E. K., Petrassi, H. M., Ratnaswamy, G., Huff, M. E., Sudol, M., and Kelly, J. W. (1999) Characterization of the structure and function of W → F WW domain variants: identification of a natively unfolded protein that folds upon ligand binding. *Biochemistry* 38, 14338–14351.

(84) Sreerama, N., and Woody, R. W. (2003) Structural composition of betaI- and betaII-proteins. *Protein Sci.* 12, 384–388.

(85) Jäger, M., Nguyen, H., Crane, J. C., Kelly, J. W., and Gruebele, M. (2001) The folding mechanism of a beta-sheet: the WW domain. *J. Mol. Biol.* 311, 373–393.

(86) Cobos, E. S., Iglesias-Bexiga, M., Ruiz-Sanz, J., Mateo, P. L., Luque, I., and Martinez, J. C. (2009) Thermodynamic characterization of the folding equilibrium of the human Nedd4-WW4 domain: at the frontiers of cooperative folding. *Biochemistry* 48, 8712–8720.

(87) Nguyen, H., Jäger, M., Moretto, A., Gruebele, M., and Kelly, J. W. (2003) Tuning the free-energy landscape of a WW domain by temperature, mutation, and truncation. *Proc. Natl. Acad. Sci. U. S. A.* 100, 3948–3953.

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

(89) Karanicolas, J., and Brooks, C. L. r. (2003) The structural basis for biphasic kinetics in the folding of the WW domain from a form-binding protein: lessons for protein design? *Proc. Natl. Acad. Sci. U. S. A.* 100, 3954–3959.

(90) Wright, P. E., and Dyson, H. J. (2014) Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* 16, 18–29.

(91) Li, F., Grishaev, A., Ying, J., and Bax, A. (2015) Side chain conformational distributions of a small protein derived from model-free analysis of a large set of residual dipolar couplings. *J. Am. Chem. Soc.* 137, 14798–14811.

- (92) Wright, P. E., and Dyson, H. J. (2009) Linking folding and binding. *Curr. Opin. Struct. Biol.* 19, 31–38.
- (93) Hammes, G. G., Chang, Y. C., and Oas, T. G. (2009) Conformational selection or induced fit: a flux description of reaction mechanism. *Proc. Natl. Acad. Sci. U. S. A.* 106, 13737–13741.
- (94) Timmins, J., Schoehn, G., Ricard-Blum, S., Scianimanico, S., Vernet, T., Ruigrok, R. W., and Weissenhorn, W. (2003) Ebola virus matrix protein VP40 interaction with human cellular factors Tsg101 and Nedd4. *J. Mol. Biol.* 326, 493–502.
- (95) Bedford, L., Lowe, J., Dick, L. R., Mayer, R. J., and Brownell, J. E. (2011) Ubiquitin-like protein conjugation and the ubiquitin-proteasome system as drug targets. *Nat. Rev. Drug Discovery* 10, 29–46.
- (96) Huang, Y. J., Powers, R., and Montelione, G. T. (2005) Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. *J. Am. Chem. Soc.* 127, 1665–1674.
- (97) Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948.

Supporting Information

The Nedd4-1 WW domain recognizes the PY motif peptide through coupled folding and binding equilibria

Vineet Panwalkar,^{†,‡} Philipp Neudecker,^{‡,§} Michael Schmitz,[†] Justin Lecher,^{‡,§} Marianne Schulte,^{‡,§} Karima Medini,^{1,¶} Matthias Stoldt,^{‡,§} Margaret A. Brimble,^{†,1,¶} Dieter Willbold,^{‡,§} and Andrew J. Dingley^{‡,}*

[†]School of Chemical Sciences, ¹School of Biological Sciences and [¶]The Maurice Wilkins Centre for Molecular Biodiscovery, The University of Auckland, Auckland 1142, New Zealand

[‡]ICS-6 (Strukturbiochemie), Forschungszentrum Jülich, 52425 Jülich, Germany

[§]Institut für Physikalische Biologie, Heinrich-Heine-Universität, 40225 Düsseldorf, Germany

Table S1. NMR data acquired for structure determination of the apo-hNedd4-1 WW3* domain

Experiment	Time (h)	Acquisition times (ms)	Data matrices
2D ^1H - ^{15}N HSQC	0.5	78.0 (t_{HN}) \times 67.9 (t_{N})	700* (H) \times 128* (N) ^a
2D ^1H - ^{13}C CT HSQC	1.5	122.3 (t_{H}) \times 23.1 (t_{C})	1024* (H) \times 256* (C)
3D HNCO	15.0	106.9 (t_{HN}) \times 17.0 (t_{N}) \times 20.4 (t_{C})	1024* (H) \times 32* (N) \times 40* (C)
3D HNCA	33.5	122.3 (t_{HN}) \times 22.3 (t_{N}) \times 14.2 (t_{C})	1024* (H) \times 42* (N) \times 64* (C)
3D CBCA(CO)NH	18.5	106.9 (t_{HN}) \times 17.0 (t_{N}) \times 6.4 (t_{C})	1024* (H) \times 32* (N) \times 48* (C)
3D H(CCO)NH	38.0	122.3 (t_{HN}) \times 12.7 (t_{N}) \times 14.2 (t_{C})	1024* (H) \times 24* (N) \times 64* (C)
3D CC(CO)NH	66.5	122.3 (t_{HN}) \times 22.3 (t_{N}) \times 6.1 (t_{C})	1024* (H) \times 42* (N) \times 64* (C)
3D ^1H - ^{15}N TOCSY-HSQC	24.5	136.5 (t_{HN}) \times 10.3 (t_{N}) \times 6.7 (t_{H})	1024* (H) \times 20* (N) \times 50* (H)
3D HCCH-TOCSY	90.0	131.3 (t_{H}) \times 3.4 (t_{C}) \times 25.7 (t_{H})	512* (H) \times 38* (C) \times 100* (H)
2D (HB)CB(CGCD)HD	1.0	53.4 (t_{H}) \times 4.3 (t_{C})	750* (H) \times 32* (C)
2D (HB)CB(CGCDCE)HE	1.0	53.4 (t_{H}) \times 4.3 (t_{C})	750* (H) \times 32* (C)
3D ^1H - ^{15}N NOESY-HSQC ^b	77.5	76.2 (t_{HN}) \times 18.7 (t_{N}) \times 11.3 (t_{H})	1024* (H) \times 46* (N) \times 128* (H)
3D ^1H - ^{13}C NOESY-HSQC	163.0	60.8 (t_{H}) \times 10.9 (t_{C}) \times 14.1 (t_{H})	768* (H) \times 94* (C) \times 73* (H)
3D ^1H - ^{13}C NOESY-HSQC (aromatic region)	43.0	65.9 (t_{H}) \times 19.8 (t_{C}) \times 9.6 (t_{H})	832* (H) \times 36* (C) \times 50* (H)

^a n^* refers to complex points. ^bNOESY experiments were run at 900 MHz, whereas the other experiments were recorded at 600 MHz.

Table S2. Summary of the NMR relaxation experiments recorded for the apo-hNedd4-1**WW3* domain**

Experiment	Time (h)	Acquisition times (ms)	Data matrices	Magnetic field (T)
2D ¹⁵ N R _{1ρ} relaxation	17	64.0 (t _{HN}) × 52.6 (t _N)	615* (H) × 96* (N) ^a	14.1
2D ¹⁵ N R ₁ relaxation	22	64.0 (t _{HN}) × 52.6 (t _N)	614* (H) × 96* (N)	14.1
2D { ¹ H}- ¹⁵ N heteronuclear NOE	36	65.8 (t _{HN}) × 52.6 (t _N)	633* (H) × 96* (N)	14.1
2D ¹⁵ N SQ CPMG dispersion	29	106.9 (t _{HN}) × 58.2 (t _N)	1024* (H) × 92* (N)	14.1
2D ¹ H- ¹⁵ N HSQC	10	106.9 (t _{HN}) × 140.5 (t _N)	1024* (H) × 256* (N)	14.1
2D ¹ H- ¹⁵ N HMQC	10	106.9 (t _{HN}) × 140.5 (t _N)	1024* (H) × 256* (N)	14.1
2D ¹ H SQ CPMG dispersion	44	106.9 (t _{HN}) × 52.6 (t _N)	1024* (H) × 96* (N)	14.1
2D ¹⁵ N R _{1ρ} relaxation	21	64.6 (t _{HN}) × 40.0 (t _N)	832* (H) × 104* (N)	18.8
2D ¹⁵ N R ₁ relaxation	24	64.6 (t _{HN}) × 40.0 (t _N)	832* (H) × 104* (N)	18.8
2D { ¹ H}- ¹⁵ N heteronuclear NOE	39	64.6 (t _{HN}) × 40.0 (t _N)	832* (H) × 104* (N)	18.8
2D ¹⁵ N R _{1ρ} relaxation	17	70.5 (t _{HN}) × 35.1 (t _N)	1024* (H) × 96* (N)	21.2
2D ¹⁵ N R ₁ relaxation	22	70.5 (t _{HN}) × 35.1 (t _N)	1024* (H) × 96* (N)	21.2
2D { ¹ H}- ¹⁵ N heteronuclear NOE	36	70.5 (t _{HN}) × 35.1 (t _N)	1024* (H) × 96* (N)	21.2
2D ¹⁵ N SQ CPMG dispersion	29	70.5 (t _{HN}) × 35.1 (t _N)	1024* (H) × 96* (N)	21.2
2D ¹ H- ¹⁵ N HSQC	13	71.3 (t _{HN}) × 93.6 (t _N)	1024* (H) × 256* (N)	21.2
2D ¹ H- ¹⁵ N HMQC	13	71.3 (t _{HN}) × 93.6 (t _N)	1024* (H) × 256* (N)	21.2
2D ¹ H SQ CPMG dispersion	41	71.3 (t _{HN}) × 35.1 (t _N)	1024* (H) × 96* (N)	21.2

^an* refers to complex points.

Table S3. Summary of the NMR relaxation experiments recorded for the hNedd4-1 WW3*– α -hENaC complex

Experiment	Time (h)	Acquisition times (ms)	Data matrices	Magnetic field (T)
2D ^{15}N R_2 relaxation	20	136.4 (t_{HN}) \times 56.6 (t_{N})	1024* (H) \times 110* (N)	14.1
2D ^{15}N R_1 relaxation	22	136.4 (t_{HN}) \times 51.4 (t_{N})	1024* (H) \times 100* (N)	14.1
2D $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE	32	136.4 (t_{HN}) \times 65.8 (t_{N})	1024* (H) \times 128* (N)	14.1
2D ^{15}N SQ CPMG dispersion	24	106.9 (t_{HN}) \times 58.2 (t_{N})	1024* (H) \times 92* (N)	14.1
2D ^{15}N $R_{1\rho}$ relaxation	21	64.6 (t_{HN}) \times 39.5 (t_{N})	832* (H) \times 96* (N)	18.8
2D ^{15}N R_1 relaxation	24	64.6 (t_{HN}) \times 39.5 (t_{N})	832* (H) \times 96* (N)	18.8
2D $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE	39	64.6 (t_{HN}) \times 39.5 (t_{N})	832* (H) \times 96* (N)	18.8
2D ^{15}N $R_{1\rho}$ relaxation	18	70.5 (t_{HN}) \times 32.9 (t_{N})	1024* (H) \times 90* (N)	21.2
2D ^{15}N R_1 relaxation	22	70.5 (t_{HN}) \times 32.9 (t_{N})	1024* (H) \times 90* (N)	21.2
2D $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE	36	70.5 (t_{HN}) \times 32.9 (t_{N})	1024* (H) \times 90* (N)	21.2
2D ^{15}N SQ CPMG dispersion	29	70.5 (t_{HN}) \times 35.1 (t_{N})	1024* (H) \times 96* (N)	21.2

Table S4. Thermodynamic parameters of association derived from fitting the ITC measurements of the hNedd4-1 WW3*– α -hENaC complex (see Figure S1)

n^a	0.88 ± 0.02
K_d^{app}	$(44.53 \pm 1.84) \mu\text{M}$
ΔH_A^{app}	$(-58.20 \pm 0.29) \text{kJ mol}^{-1}$
$T\Delta S_A^{\text{appb}}$	$(-31.67 \pm 0.04) \text{kJ mol}^{-1}$
ΔG_A^{app}	$(-26.57 \pm 0.29) \text{kJ mol}^{-1}$

^a n = stoichiometry

^bTemperature (T) = 25 °C.

Table S5. Results from model-free analysis of the apo-hNedd4 WW3* domain

Residue	Model	S^2	ΔS^2	S_f^2	ΔS_f^2	$\tau_e < 100$ ps or τ_f [ps]	$\Delta \tau_f$ [ps]	$\tau_e > 100$ ps or τ_s [ps]	$\Delta \tau_s$ [ps]	R_{ex} [s^{-1}]	ΔR_{ex} [s^{-1}]
418	8	0.260	0.099	0.652	0.039	61.60	7.35	2409	572	0.58	0.31
419	5	0.365	0.007	0.801	0.014	–	–	355	9	–	–
420	6	0.276	0.015	0.731	0.014	69.67	8.87	1008	63	–	–
421	6	0.241	0.014	0.719	0.014	72.53	7.50	1437	68	–	–
422	6	0.475	0.02	0.811	0.023	31.17	13.74	1024	119	–	–
423	7	0.722	0.027	0.838	0.03	–	–	277	139	1.48	0.16
425	5	0.800	0.021	0.882	0.024	–	–	513	209	–	–
426	4	0.767	0.029	–	–	–	–	1234	138	1.71	0.21
427	4	0.835	0.019	–	–	29.32	8.00	–	–	2.98	0.20
428	4	0.790	0.019	–	–	37.68	6.99	–	–	1.01	0.14
429	4	0.809	0.018	–	–	35.72	7.83	–	–	4.45	0.20
430	4	0.841	0.018	–	–	84.38	20.54	–	–	5.67	0.25
431	4	0.741	0.018	–	–	40.58	6.03	–	–	7.61	0.29
432	4	0.813	0.019	–	–	41.13	8.76	–	–	2.03	0.18
434	4	0.742	0.027	–	–	–	–	1230	107	10.75	0.41
435	2	0.888	0.014	–	–	–	–	117	63	–	–
436	4	0.823	0.019	–	–	45.32	9.37	–	–	3.34	0.20
438	7	0.834	0.077	0.919	0.035	–	–	561	528	5.19	0.36
439	4	0.818	0.019	–	–	35.11	9.10	–	–	7.60	0.28
440	5	0.834	0.025	0.91	0.022	–	–	591	301	–	–
441	4	0.780	0.017	–	–	44.71	7.49	–	–	1.12	0.13
442	4	0.916	0.025	–	–	–	–	426	233	2.31	0.20
443	4	0.917	0.016	–	–	–	–	185	150	5.71	0.28
445	7	0.566	0.212	0.818	0.054	–	–	1920	1186	10.46	0.66
446	4	0.730	0.018	–	–	28.20	4.77	–	–	11.47	0.377
447	4	0.871	0.019	–	–	99.59	36.54	–	–	1.81	0.20
448	4	0.759	0.018	–	–	26.86	5.22	–	–	3.93	0.21
449	7	0.807	0.056	0.884	0.04	–	–	445	397	1.95	0.25
451	7	0.769	0.075	0.873	0.037	–	–	763	459	4.03	0.35
453	7	0.778	0.075	0.901	0.033	–	–	841	434	0.51	0.31
454	7	0.734	0.071	0.866	0.031	–	–	745	370	1.89	0.32
455	5	0.622	0.016	0.857	0.015	–	–	495	65	–	–

456	6	0.227	0.014	0.665	0.012	74.37	6.59	1187	63	-	-
458	7	0.123	0.012	0.666	0.015	-	-	571	11	0.16	0.05

Table S6. Results from model-free analysis of the hNedd4 WW3*– α -hENaC peptide complex

Residue	Model	S^2	ΔS^2	S^2_f	ΔS^2_f	$\tau_c < 100$ ps or τ_f [ps]	$\Delta \tau_f$ [ps]	$\tau_c > 100$ ps or τ_s [ps]	$\Delta \tau_s$ [ps]	R_{ex} [s^{-1}]	ΔR_{ex} [s^{-1}]
418	6	0.361	0.046	0.684	0.019	75.86	7.21	2622	793	–	–
419	6	0.176	0.018	0.547	0.009	79.86	5.01	1588	127	–	–
420	6	0.242	0.021	0.668	0.014	75.12	7.04	1442	119	–	–
421	6	0.302	0.025	0.677	0.019	54.81	6.52	1746	233	–	–
422	5	0.502	0.022	0.836	0.015	–	–	892	74	–	–
423	5	0.703	0.026	0.851	0.014	–	–	817	194	–	–
425	2	0.817	0.013	–	–	26.55	7.04	–	–	–	–
426	2	0.861	0.015	–	–	31.66	11.20	–	–	–	–
427	2	0.91	0.015	–	–	35.58	19.38	–	–	–	–
428	5	0.794	0.033	0.856	0.028	–	–	1895	1226	–	–
429	1	0.858	0.015	–	–	–	–	–	–	–	–
430	2	0.907	0.014	–	–	27.56	17.92	–	–	–	–
431	5	0.793	0.022	0.839	0.05	–	–	474	413	–	–
432	2	0.85	0.015	–	–	14.05	8.87	–	–	–	–
434	2	0.801	0.013	–	–	39.11	7.14	–	–	–	–
435	2	0.853	0.015	–	–	49.67	11.61	–	–	–	–
436	4	0.861	0.019	–	–	57.59	15.27	–	–	0.23	0.14
438	2	0.898	0.016	–	–	22.85	14.18	–	–	–	–
439	2	0.892	0.015	–	–	24.20	13.94	–	–	–	–
440	2	0.864	0.015	–	–	33.40	11.22	–	–	–	–
441	1	0.881	0.014	–	–	–	–	–	–	–	–
442	1	0.882	0.015	–	–	–	–	–	–	–	–
443	2	0.896	0.016	–	–	56.49	19.48	–	–	–	–
444	4	0.823	0.019	–	–	19.07	8.22	–	–	0.56	0.14
445	2	0.944	0.015	–	–	62.80	203.00	–	–	–	–
446	4	0.903	0.02	–	–	24.02	15.52	–	–	0.61	0.15
447	3	0.867	0.02	–	–	–	–	–	–	1.13	0.16
448	2	0.92	0.015	–	–	42.26	32.35	–	–	–	–
449	5	0.822	0.035	0.913	0.02	–	–	1425	863	–	–
450	2	0.904	0.015	–	–	70.98	42.82	–	–	–	–
451	2	0.823	0.014	–	–	18.64	7.54	–	–	–	–
453	2	0.882	0.015	–	–	29.48	12.76	–	–	–	–
454	2	0.824	0.014	–	–	27.50	8.24	–	–	–	–

455	6	0.428	0.026	0.764	0.024	49.28	10.03	1471	242	-	-
456	6	0.211	0.018	0.712	0.015	73.41	9.11	1072	75	-	-
458	5	0.06	0.007	0.646	0.014	-	-	654	7	-	-

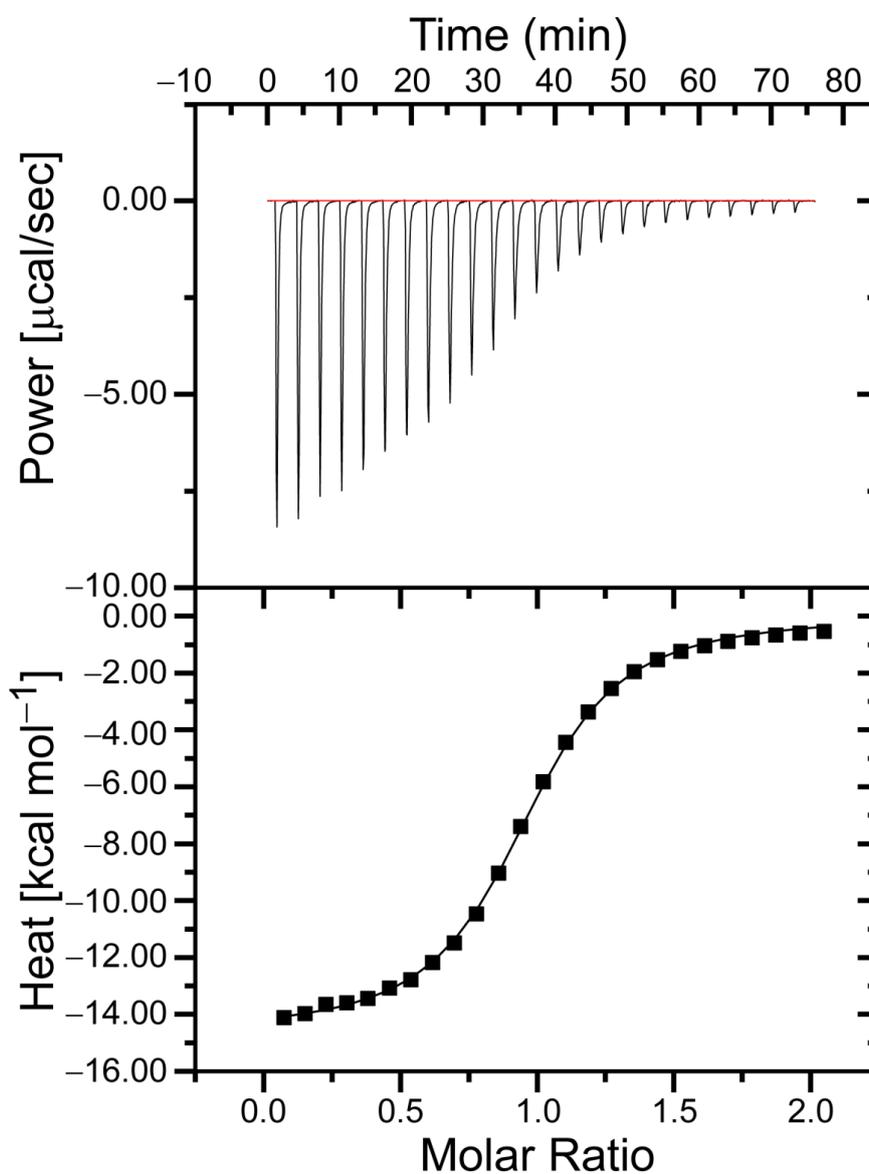


Figure S1. ITC data for the hNedd4-1 WW3* domain binding to the α -hENaC peptide. *A*, Baseline-subtracted raw ITC data for injections of the α -hENaC peptide into a solution of the WW3* domain at 25 °C. The raw data is expressed as a change in thermal power with respect to time over the period of the titration. *B*, Change in the molar heat as a function of the molar ratio of the α -hENaC peptide to the WW3* domain of Nedd4-1. The solid line shows the fit of the raw data to a single site binding model.

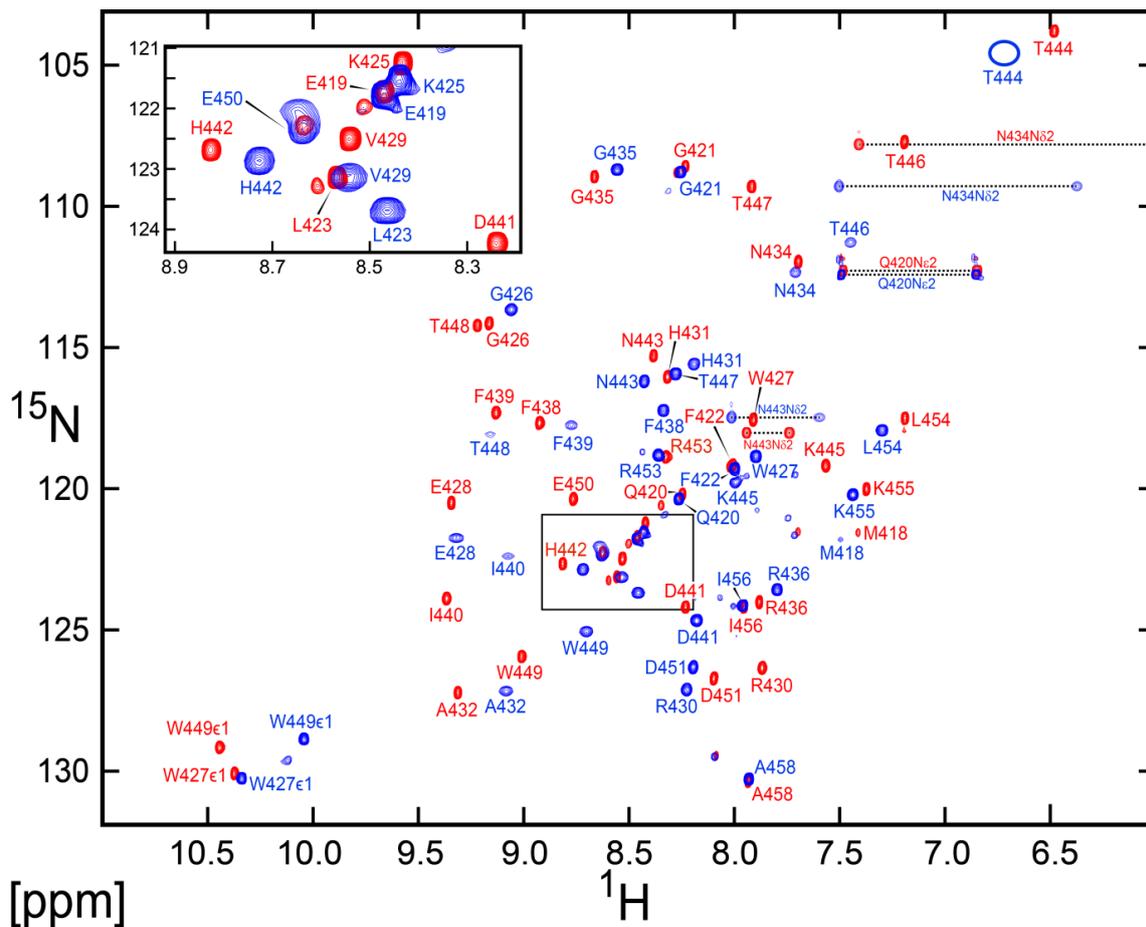


Figure S2. 2D ^1H - ^{15}N HSQC spectra of the WW3* domain from hNedd4-1. Overlay of 2D ^1H - ^{15}N HSQC spectra of apo (blue) and peptide-bound hNedd4-1 WW3* (red) recorded at 25 °C. Assignment information is provided. In the apo state, the resonance for residue T444 was only observable at a lower contour plotting threshold. The dotted lines connect side chain amide protons corresponding to Q420, N434 and N443.

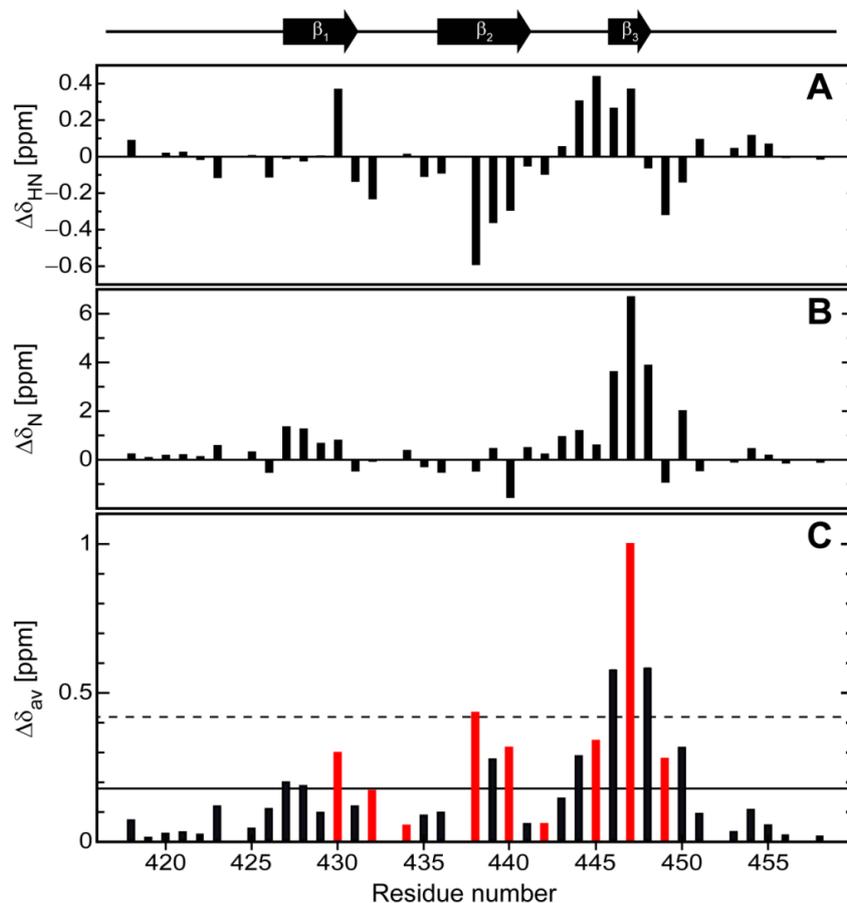


Figure S3. Changes in $^1\text{H}_\text{N}$ and ^{15}N chemical shifts between the apo-Nedd4-1 WW3* domain and the WW3* domain in complex with the α -hENaC peptide. (A) Changes in $^1\text{H}_\text{N}$ chemical shifts ($\Delta\delta_{\text{HN}} = \delta_{\text{HN}}(\text{apo}) - \delta_{\text{HN}}(\text{complex})$) between apo and peptide bound states. (B) Changes in ^{15}N chemical shifts ($\Delta\delta_{\text{N}} = \delta_{\text{N}}(\text{apo}) - \delta_{\text{N}}(\text{complex})$) between apo and peptide bound states. (C) Weighted average ^1H and ^{15}N chemical shift changes, $\Delta\delta_{\text{av}}$ ($\Delta\delta_{\text{av}} = [(\Delta\delta_{\text{HN}}^2 + \Delta\delta_{\text{N}}^2/25)/2]^{1/2}$), for residues of the hNedd4-1 WW3* domain upon binding the α -hENaC peptide at a molar ratio of protein to peptide of 1:30. The line represents the mean $\Delta\delta_{\text{av}}$ of 0.18 ppm for all residues, whereas the dotted line represents the mean plus one S.D. = 0.42 ppm. The bars in red indicate data for residues directly involved in the interaction with the α -hENaC peptide (i.e., R430, A432, N434, F438, I440, H442, K445, T447 and W449). A schematic representation of the secondary structure elements is shown at the top.

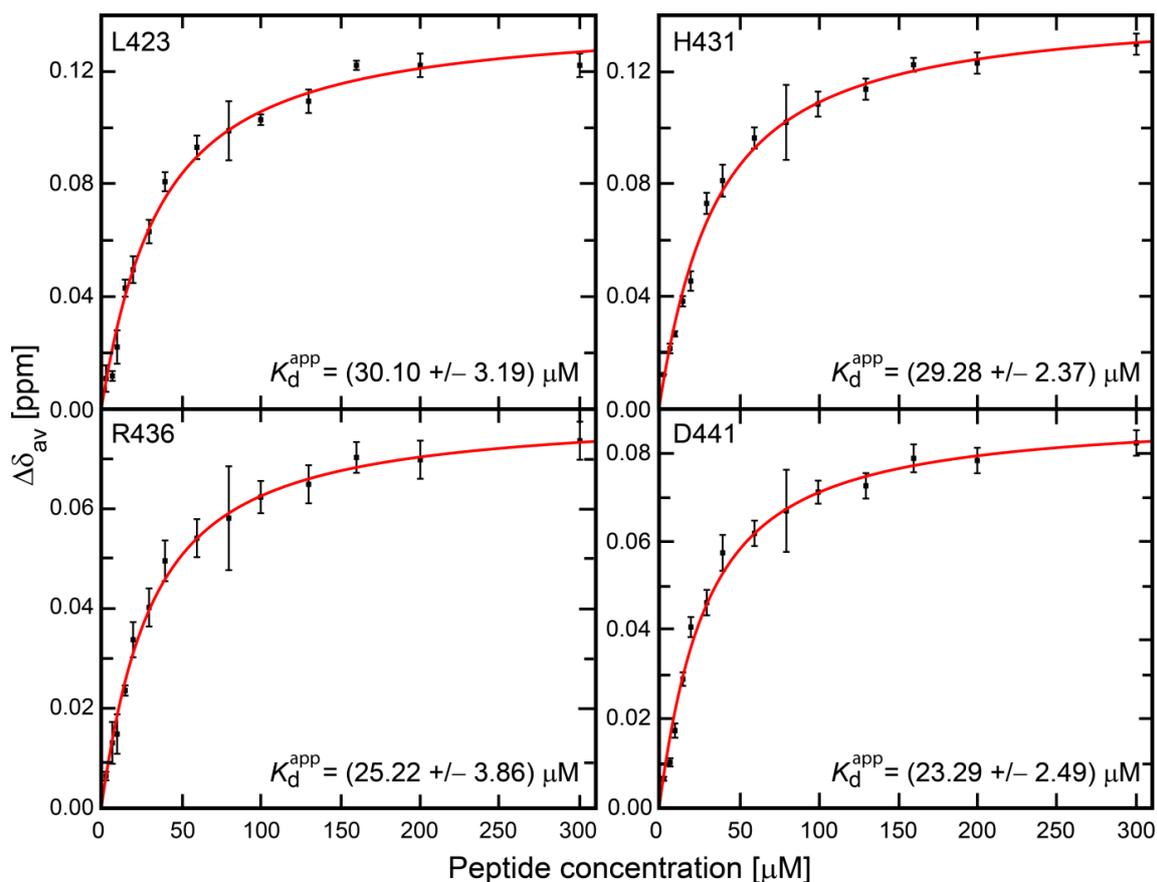


Figure S4. Selected titration curves showing the $\Delta\delta_{av}$ changes derived from 2D ^1H - ^{15}N HSQC spectra as a function of peptide concentration for the titration of ^{15}N -labeled hNedd4-1 WW3* with the α -hENaC peptide. The WW3* domain concentration was 10 μM and fifteen peptide concentrations were used: 0, 3, 7, 10, 15, 20, 30, 40, 60, 80, 100, 130, 160, 200 and 300 μM . Values of $\Delta\delta_{av}$ are plotted as symbols and the continuous red lines show the curves fitted individually to a one-site binding model (Eq. 1). K_d^{app} values for the individual fits are given in each panel. Errors are derived from two separate titrations.

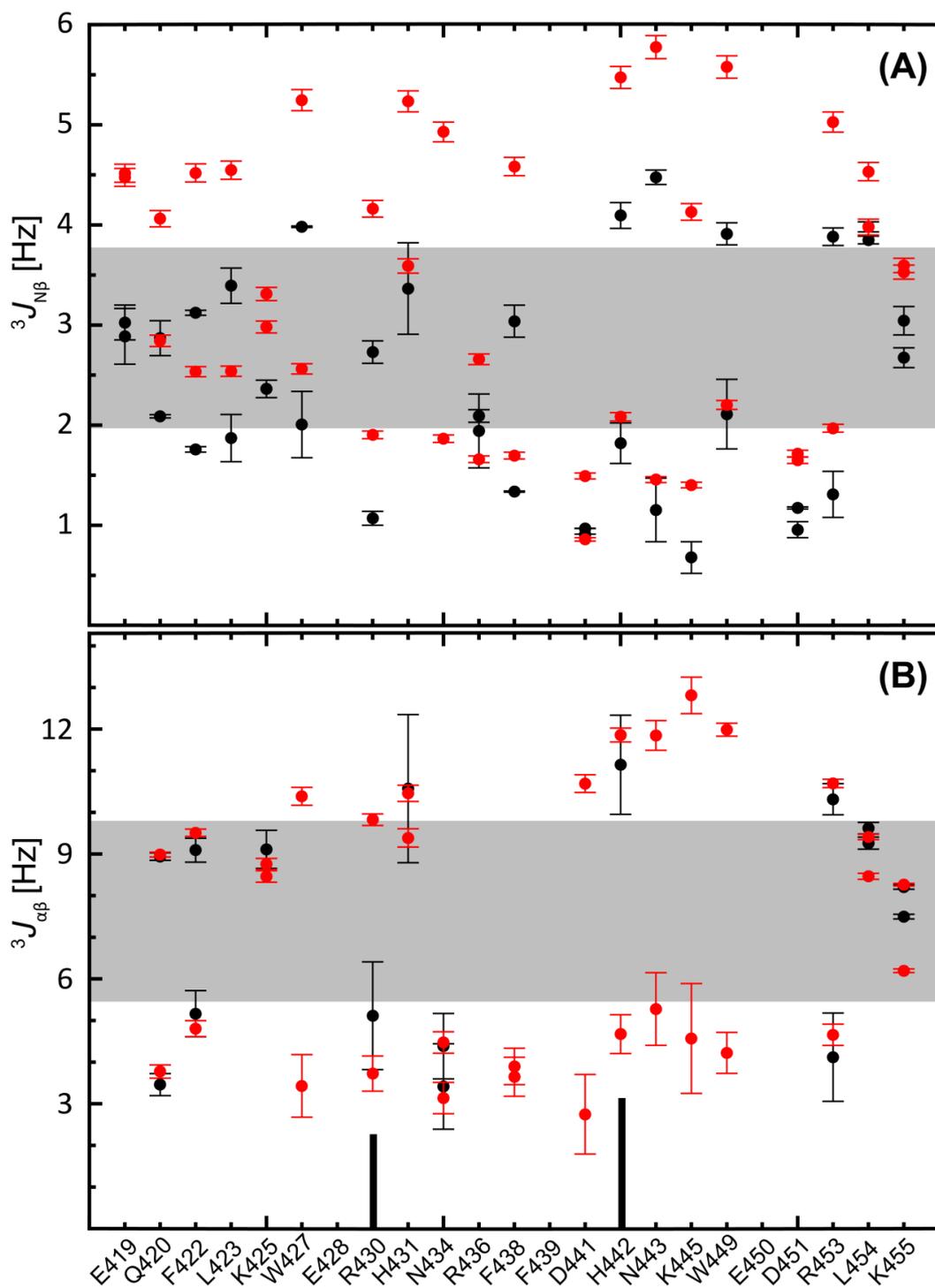


Figure S5. $^3J_{N\beta}$ (A) and $^3J_{\alpha\beta}$ (B) couplings for the apo-hNedd4-1 WW3* domain (black symbols) and the WW3* domain in complex with the α -hENaC peptide (red symbols).

Trans, *gauche*⁺ and *gauche*⁻ rotamer state boundaries were determined using the χ_1 -related

consensus Karplus coefficients inferred from self-consistent J couplings analysis¹ using an amplitude of χ_1 angle fluctuations of $\pm 20^\circ$ for each rotamer state. Thus, ${}^3J_{N\beta}$ coupling values of $2.0 \leq {}^3J_{N\beta} \leq 3.7$ Hz were considered to indicate dynamic averaging between χ_1 rotameric states, and ${}^3J_{\alpha\beta}$ couplings values of $5.5 \leq {}^3J_{\alpha\beta} \leq 9.8$ Hz were considered to indicate dynamic averaging between χ_1 rotameric states. These regions are shaded grey in the figure. Upper limits for ${}^3J_{\alpha\beta}$ (shown as bars) were derived for residues where one of the two cross peaks were missing using the ratio of the height of the lowest plotted contour line relative to the amplitude of the diagonal peak.² Distinct rotamer states that were used in structure calculations of the apo-hNedd4-1 WW3* domain were for residues N434, D441, H442, D451 and R453. For other residues, insufficient coupling data (e.g., due to severe line broadening) was available to define unambiguously the rotamer state of residues W427, R436, F438, N443, K445 and W449, or the coupling values were indicative of rotamer averaging (i.e., E419, Q420, F422, L423, K425, R430, H431, L454 and K455). For residues K425, H431 and K445, single coupling values were determined for the apo-state because the chemical shifts of the H β protons for these residues were indistinguishable (e.g., chemically equivalent), whereas in the complex resonances for each H β were observed. Data for E428, F439 and E450 were not obtained because the chemical shifts of the H β protons for these three residues were indistinguishable in both the apo and peptide bound states.

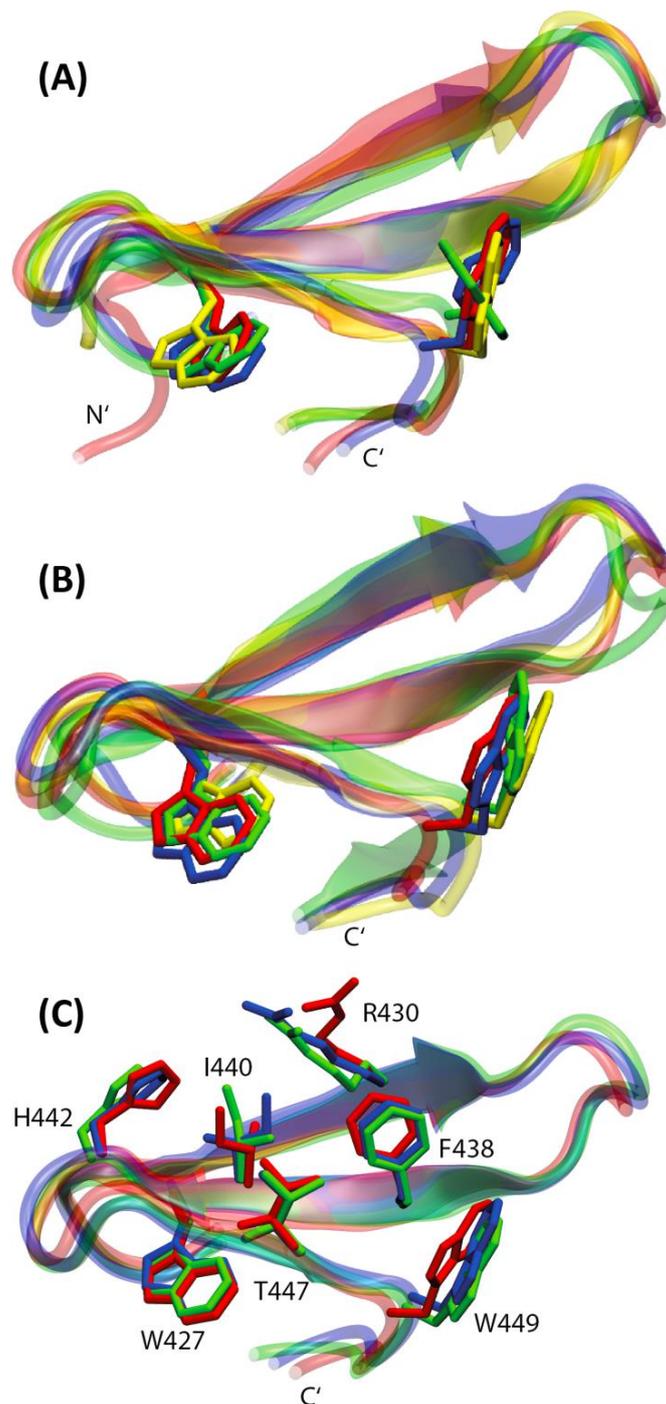


Figure S6. Comparison of the solution structure of the human Nedd4-1 WW3* domain with other WW domains. (A) Superposition of the human Nedd4-1 WW3* apo domain (red) with other apo-WW domains: human FBP11 (yellow, PDB ID: 1ZR7),³ mouse FBP28 WW domain (blue, 1E0L)⁴ and a *de novo* prototype WW domain (green, 1E0M).⁴ The two

conserved tryptophan residues are shown. (B) Superposition of the human Nedd4-1 WW3* apo domain (red), human Nedd4-1 WW3* domain in complex with the human α -ENaC peptide (blue, 2M3O),⁵ *Drosophila* Nedd4 WW3 domain in complex with the Comm (Commissureless) peptide (green, 2EZ5)⁶ and rat Nedd4 WW3 domain in complex with the rat ENaC β P2 peptide (yellow, 1I5H).⁷ The two conserved tryptophan residues are shown. (C) Superposition of the human Nedd4-1 WW3* apo domain structures solved by NMR (averaged structure; red) and X-ray crystallography⁸ (PDB ID: 4N7F; blue and green – there are two molecules in the asymmetric unit). The side chains of residues involved in peptide binding and the hydrophobic core are shown. Two key peptide binding residues, R430 and I440, show different side chain rotamers in the NMR and crystal structures. The side chain χ_1 angle of R430 adopts a *gauche*⁺ rotamer in subunit A of the crystal structure and a *gauche*⁻ rotamer in subunit B, whereas in the solution structure the rotamer is *gauche*⁺. The side chain χ_1 angle of I440 adopts a *gauche*⁺ side chain rotamer in subunit A and a *gauche*⁻ rotamer in subunit B, whereas in the solution structure the rotamer state is *trans*.

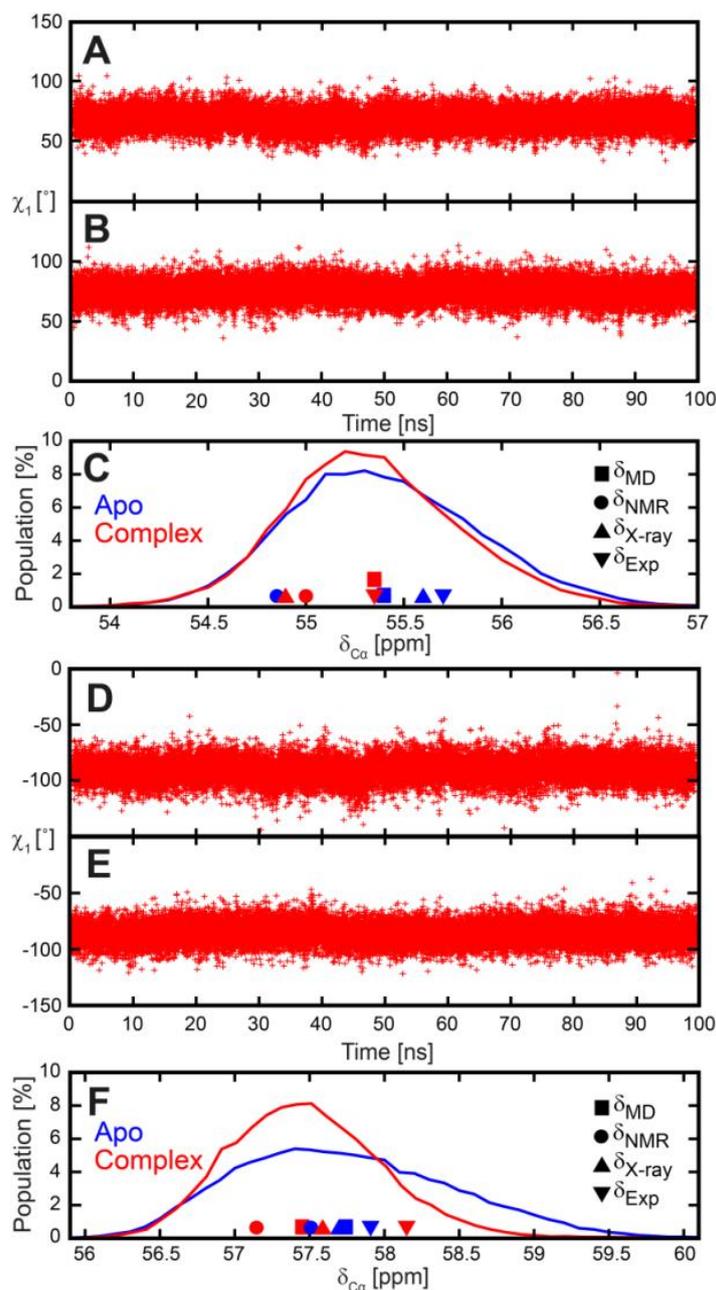


Figure S7. Chemical shift predictions of F438 and W449 $^{13}\text{C}\alpha$ in 100 ns MD simulations of WW3* in the apo and peptide bound states. The value of the χ_1 torsion angle of F438 (A,B) and W449 (D,E) in the apo (A, D) and complex (B, E) states for snapshots saved every 4 ns of the MD trajectories. Normalized distributions of the SPARTA+ predicted $\text{C}\alpha$ chemical shifts of F438 (C) and W449 (F) from the 100 ns MD trajectories of the apo state (blue line)

and peptide bound state (red line). Chemical shifts were predicted for snapshots saved each 4 ns. The average value of the SPARTA+ predictions over the entire MD trajectory are shown (apo, blue square; WW3*-peptide complex, red square). The SPARTA+ predictions obtained from the lowest energy NMR (circle) and X-ray (triangle) structures in the apo (blue) and peptide bound (red) states are shown. The experimentally determined values are shown for the apo (blue inverted triangle) and complex (red inverted triangle).

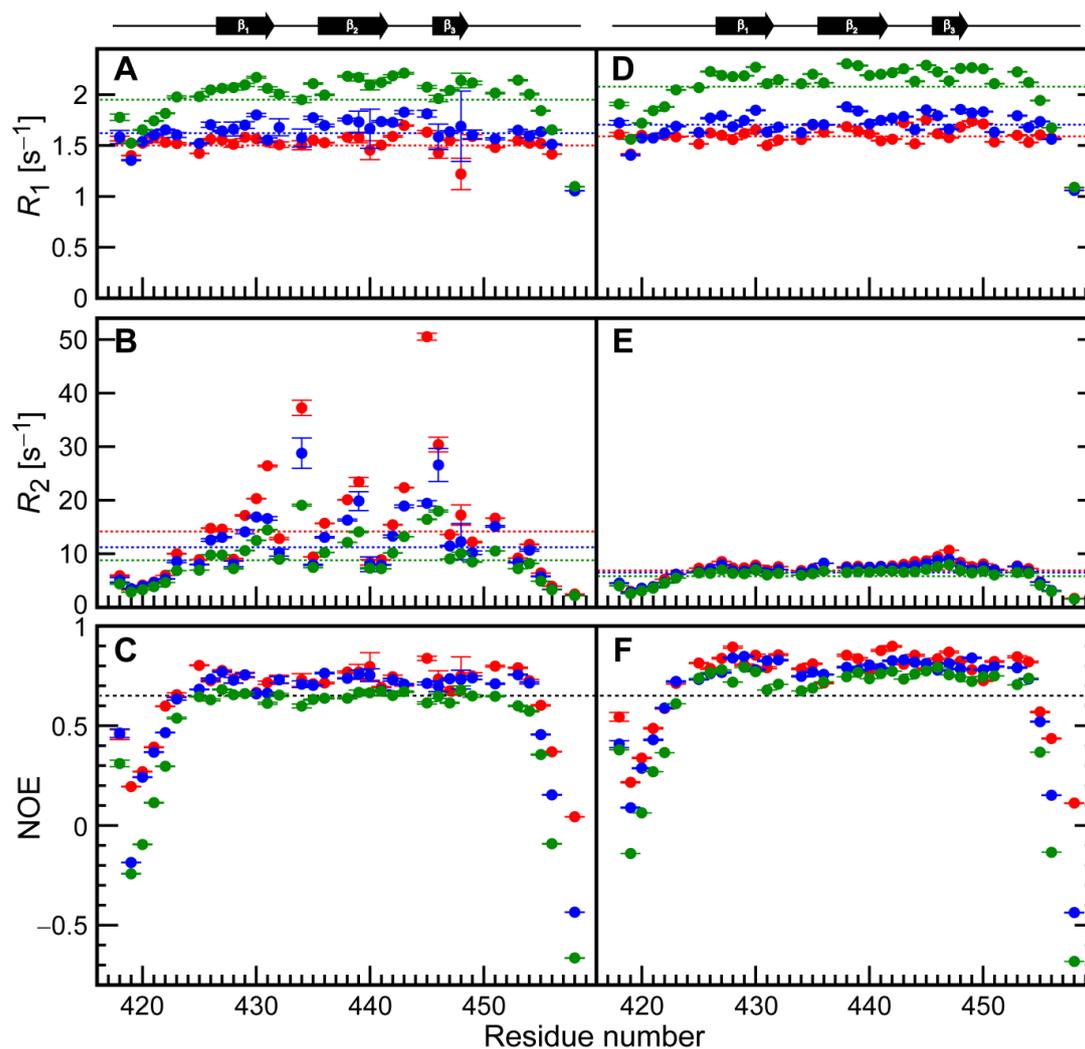


Figure S8. Backbone amide ^{15}N relaxation data of hNedd4-1 WW3* and the hNedd4-1 WW3*- α -hENaC peptide complex. Longitudinal R_1 relaxation rates (A, D), transverse R_2 relaxation rates (B, E) and steady-state heteronuclear $\{^1\text{H}\}$ - ^{15}N NOE (C, F) recorded at 25 °C and magnetic field strengths of 14.1 T (green), 18.8 T (blue) and 21.1 T (red). Data for the apo-state is presented in the three panels on the left (A, B, C) and data of the WW3* domain in complex with the α -hENaC peptide is presented in the three panels on the right (D, E, F). The dashed lines in (A, D) and (B, E) represent the average rates, whereas in (C, F), the black dashed line is drawn at an NOE value of 0.65. A schematic representation of the secondary structure is presented at the top.

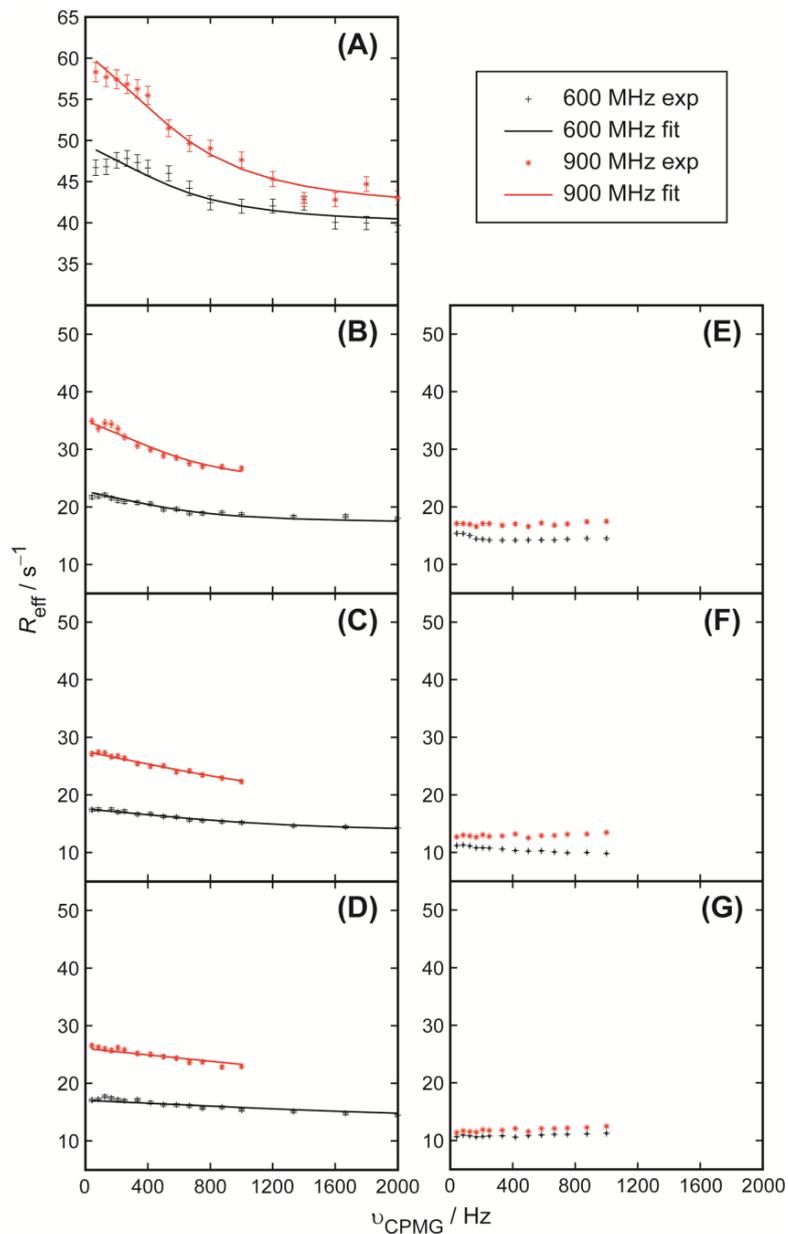


Figure S9. Effective $^1\text{H}_\text{N}$ and ^{15}N single-quantum (SQ) relaxation rates R_eff as a function of CPMG frequency ν_CPMG for G426 of the hNedd4 WW3* domain in the free (A, B, C and D) and peptide bound (E, F and G) states. (A) $^1\text{H}_\text{N}$ SQ R_eff at 5 °C; (B and E) ^{15}N SQ R_eff at 5 °C; (C and F) ^{15}N SQ R_eff at 15 °C; (D and G) ^{15}N SQ R_eff at 25 °C. Continuous lines are the relaxation dispersion profiles predicted using parameters generated from fitting a two-state exchange model to the $^1\text{H}_\text{N}$ and ^{15}N relaxation dispersion data.

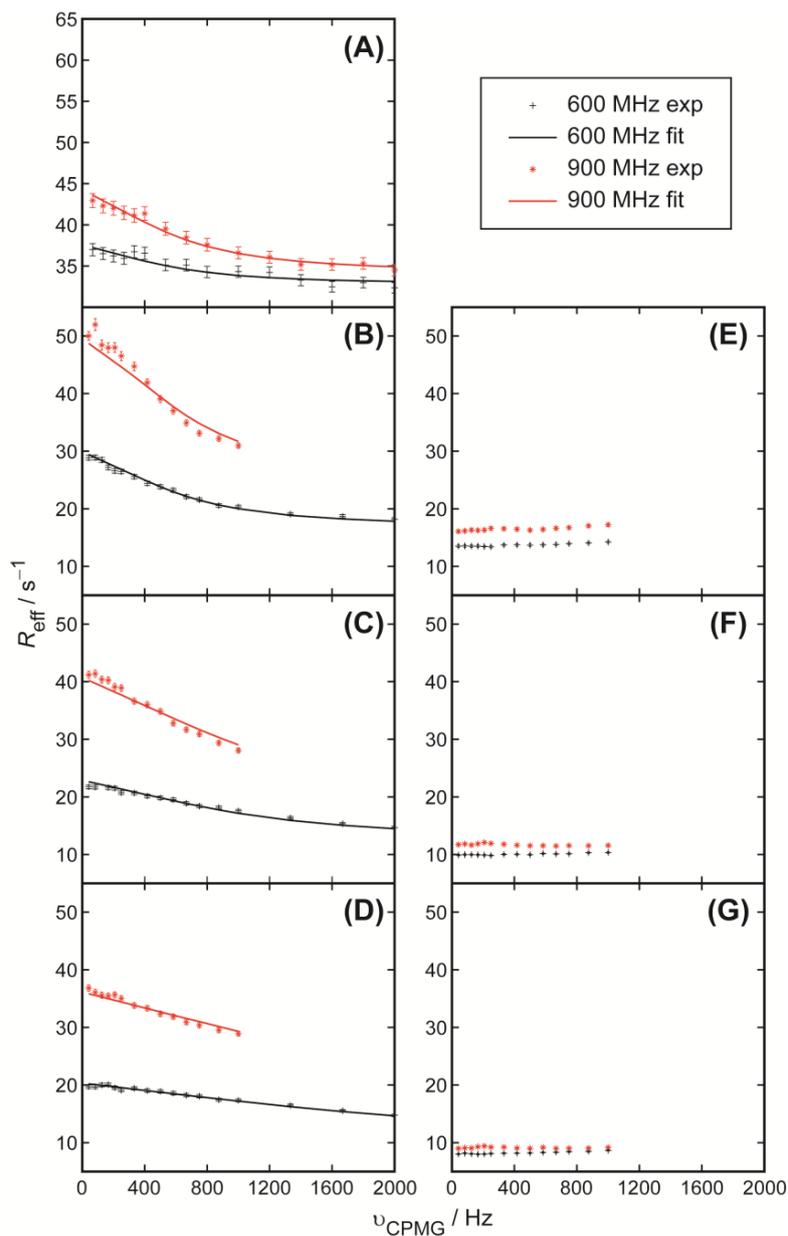


Figure S10. Effective $^1\text{H}_\text{N}$ and ^{15}N single-quantum (SQ) relaxation rates R_eff as a function of CPMG frequency ν_CPMG for N443 of the hNedd4 WW3* domain in the free (A, B, C and D) and peptide bound (E, F and G) states. (A) $^1\text{H}_\text{N}$ SQ R_eff at 5 °C; (B and E) ^{15}N SQ R_eff at 5 °C; (C and F) ^{15}N SQ R_eff at 15 °C; (D and G) ^{15}N SQ R_eff at 25 °C. Continuous lines are the relaxation dispersion profiles predicted using parameters generated from fitting a two-state exchange model to the $^1\text{H}_\text{N}$ and ^{15}N relaxation dispersion data.

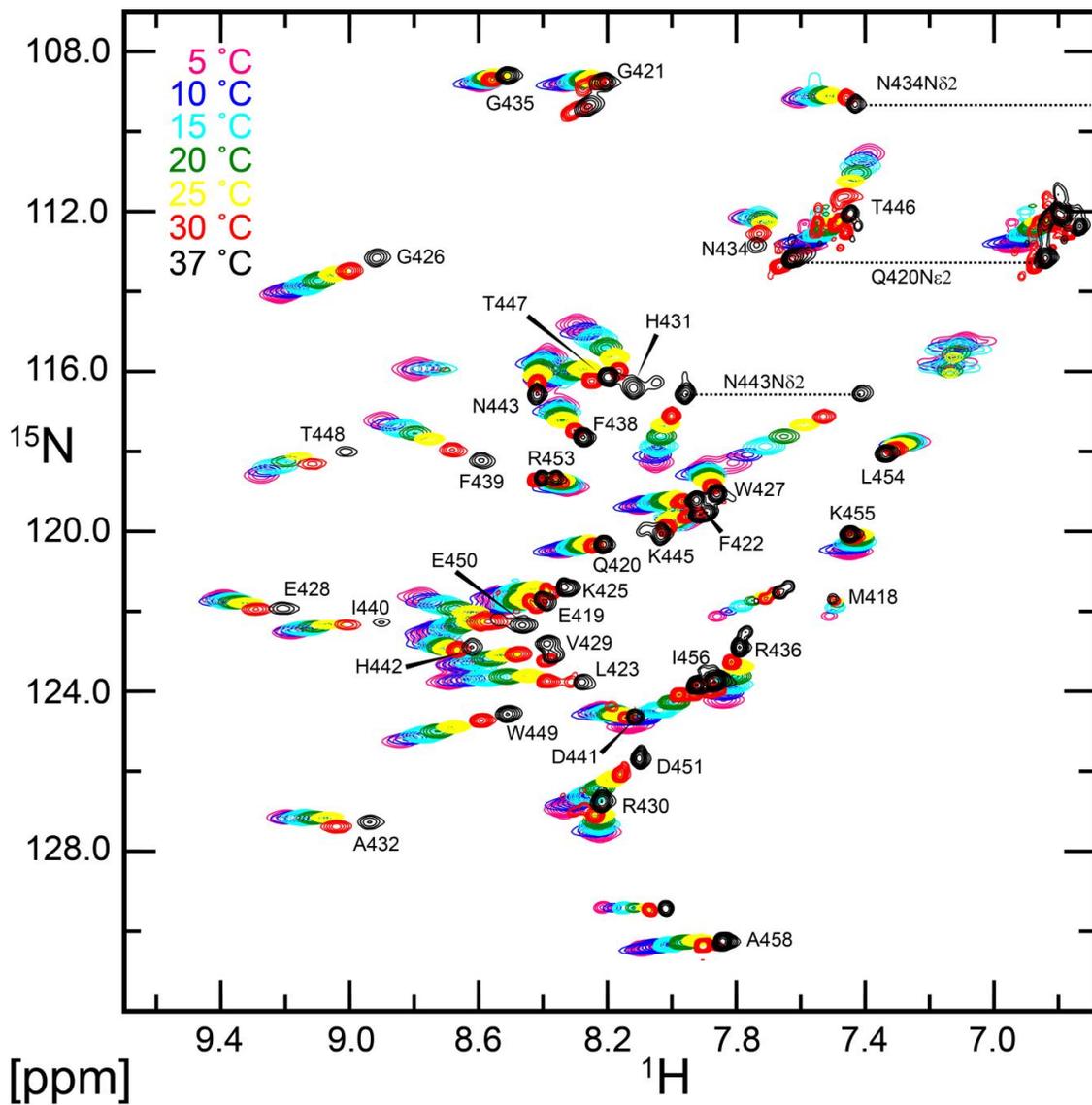


Figure S11. 2D ^1H - ^{15}N HSQC spectra of the apo-Nedd4-1 WW3* domain at 5, 10, 15, 20, 25, 30 and 37 °C. Assignment information is provided for backbone and side chain amide groups.

REFERENCES

- (1) Perez, C., Lohr, F., Ruterjans, H., and Schmidt, J. M. (2001) Self-consistent Karplus parametrization of 3J couplings depending on the polypeptide side-chain torsion χ_1 . *J. Am. Chem. Soc.* *123*, 7081–7093.
- (2) Dingley, A. J., Masse, J. E., Peterson, R. D., Barfield, M., Feigon, J., and Grzesiek, S. (1999) Internucleotide scalar couplings across hydrogen bonds in Watson-Crick and Hoogsteen base pairs of a DNA triplex. *J. Am. Chem. Soc.* *121*, 6019–6027.
- (3) Kato, Y., Hino, Y., Nagata, K., and Tanokura, M. (2006) Solution structure and binding specificity of FBP11/HYPA WW domain as Group-II/III. *Proteins* *63*, 227–234.
- (4) Macias, M. J., Gervais, V., Civera, C., and Oschkinat, H. (2000) Structural analysis of WW domains and design of a WW prototype. *Nat. Struct. Biol.* *7*, 375–379.
- (5) Bobby, R., Medini, K., Neudecker, P., Lee, T. V., Brimble, M. A., McDonald, F. J., Lott, J. S., and Dingley, A. J. (2013) Structure and dynamics of human Nedd4-1 WW3 in complex with the alphaENaC PY motif. *Biochim. Biophys. Acta* *1834*, 1632–1641.
- (6) Kanelis, V., Bruce, M. C., Skrynnikov, N. R., Rotin, D., and Forman-Kay, J. D. (2006) Structural Determinants for High-Affinity Binding in a Nedd4 WW3* Domain-Comm PY Motif Complex. *Structure* *14*, 543–553.
- (7) Kanelis, V., Rotin, D., and Forman-Kay, J. D. (2001) Solution structure of a Nedd4 WW domain-ENaC peptide complex. *Nat. Struct. Biol.* *8*, 407–412.
- (8) Qi, S., O'Hayre, M., Gutkind, J. S., and Hurley, J. H. (2014) Structural and biochemical basis for ubiquitin ligase recruitment by arrestin-related domain-containing protein-3 (ARRDC3). *J. Biol. Chem.* *289*, 4743–4752.

3.4 Data describing the solution structure of the WW3* domain from human Nedd4-1

Summary

This publication supplements the previous study, providing a more detailed description of the structure determination of apo-WW3* from hNedd4-1 by solution-state NMR. This data was complemented with MD simulations, illustrating conformational fluctuations not observed in the time- and molecule-averaged NMR observables. MD simulations suggest, for example, that χ_1 of residue T447, located in close proximity to the XP-pocket, fluctuates between a *gauche*⁺ and a *gauche*⁻ conformer state, whereas in complex with the α -hENaC peptide, it solely adopts the *gauche*⁻ rotamer.

Reprint

This section contains a complete reprint of the publication *Data in Brief* **2016**, 8, 605-612, doi:[10.1016/j.dib.2016.06.024](https://doi.org/10.1016/j.dib.2016.06.024). This article is available under the terms of the Creative Commons Attribution License (CC BY) which can be found here: <https://creativecommons.org/licenses/by/4.0/>.



ELSEVIER

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Data describing the solution structure of the WW3* domain from human Nedd4-1

Vineet Panwalkar^{a,*}, Marianne Schulte^{a,b}, Justin Lecher^{a,b},
Matthias Stoldt^{a,b}, Dieter Willbold^{a,b}, Andrew J Dingley^{a,*}^a ICS-6 Strukturbiochemie, Forschungszentrum Jülich, 52425 Jülich, Germany^b Institut für Physikalische Biologie, Heinrich-Heine-Universität, 40225 Düsseldorf, Germany

ARTICLE INFO

Article history:

Received 19 February 2016

Received in revised form

26 May 2016

Accepted 15 June 2016

Available online 22 June 2016

Keywords:

Chemical shift

Neuronal precursor cell expressed devel-

opmentally down-regulated gene 4-1

NMR

NOE distance restraints

WW domain

ABSTRACT

The third WW domain (WW3*) of human Nedd4-1 (Neuronal precursor cell expressed developmentally down-regulated gene 4-1) interacts with the poly-proline (PY) motifs of the human epithelial Na⁺ channel (hENaC) subunits at micromolar affinity. This data supplements the article (Panwalkar et al., 2015) [1]. We describe the NMR experiments used to solve the solution structure of the WW3* domain. We also present NOE network data for defining the rotameric state of side chains of peptide binding residues, and complement this data with χ_1 dihedral angles derived from 3J couplings and molecular dynamics simulations data.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Specifications Table

Subject area	Biochemistry, structural biology
More specific subject area	Nuclear magnetic resonance (NMR) spectroscopy
Type of data	Tables, figures

* Corresponding authors.

E-mail addresses: v.panwalkar@fz-juelich.de (V. Panwalkar), a.dingley@fz-juelich.de (A. Dingley).<http://dx.doi.org/10.1016/j.dib.2016.06.024>2352-3409/© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

How data was acquired	Heteronuclear multidimensional solution-state NMR spectroscopy and MD simulations from experimental structure.
Data format	Processed, analyzed
Experimental factors	The NMR experiments were performed on samples containing 1.5–1.8 mM WW3* domain (¹³ C, ¹⁵ N-labeled) from human Nedd4-1 in 20 mM sodium phosphate buffer (pH 6.5), 50 mM NaCl, 0.1% (w/v) NaN ₃ and 1 mM DSS in a 93%/7% (v/v) H ₂ O/D ₂ O mixture.
Experimental features	All NMR spectra were acquired at 25 °C on Bruker BioSpin Avance III HD 600 and Varian INOVA 900 spectrometers and data were processed using NMRPipe.
Data source location	ICS-6 (Strukturbiochemie), Forschungszentrum Jülich, Jülich, Germany
Data accessibility	Data are within this article and have been deposited in the RCSB Protein Data Bank (http://www.rcsb.org) under the accession number PDB: 5AHT and in the BioMagResBank (accession code: 25349).

Value of the data

- The NOE network defines clearly the side chain orientations of particular ligand-binding residues;
 - MD simulations provide atomistic descriptions of conformational fluctuations within the WW3* domain that are not observed in the NMR-derived structure of the domain;
 - This data set serves as a reference for future studies involving WW domains.
-

1. Data

We have collected 1592 NOE distance restraints from three-dimensional ¹⁵N-edited and ¹³C-edited NOESY spectra, which were processed using NMRPipe [2] and analyzed using CcpNMR Analysis [3]. The NOE dataset consists of 390 sequential, 416 intra-residue, 266 medium-range and 256 long-range NOE distance restraints. In addition, 60 dihedral angle restraints and five sidechain χ_1 angle restraints determined from combined $^3J_{\alpha\beta}$ and $^3J_{N\beta}$ couplings were used for structure calculation. The NOEs were picked manually and assigned in a semi-automated manner using the Aria 2.3.1 [4] software package. The structure calculation was carried out by a combination of Aria 2.3.1 and CNS version 1.21 [5] using the PARALLHDG force field. The protocol employed by Aria for calculation of the solution structure of the WW3* domain is provided as supplementary material. The experiments performed to acquire chemical shift assignments, 3J couplings and NOE distance restraints are summarized in Table 1. The 3J couplings and the subsequently determined rotameric state for the WW3* domain are given in Table 2.

We provide, as examples, the NOE networks for two key peptide binding residues I440 and T447 (Figs. 1 and 2), side chain rotamers of which differ between NMR and the crystal structures [6]. MD simulations data of χ_1 rotameric states of six key peptide binding residues (R430, F438, I440, H442, T447 and W449) over 100 ns in the apo and hENaC peptide bound state of the WW3* domain is provided (Fig. 3).

Table 1
NMR experiments used for structure determination.

Experiment	Sweep width (ppm)	Data matrices	Chemical shift offset (ppm)	Number of scans	Recycle delay (s)	Time (h)
Backbone assignments^a						
3D HNCO	16 (H) × 32 (N) × 13 (C)	1024* (H) × 32* (N) × 40* (C)	4.7 (H) × 117.1 (N) × 175.2 (C)	8	1.1	15
3D HNCA	12.5 (H) × 29 (N) × 28 (C)	1024* (H) × 42* (N) × 64* (C)	4.7 (H) × 117.1 (N) × 56.8 (C)	8	1.1	28
3D CBCA(CO)NH	16 (H) × 32 (N) × 50 (C)	1024* (H) × 32* (N) × 48* (C)	4.7 (H) × 117.1 (N) × 50 (C)	8	1.1	38
Side chain assignments						
3D H(CCO)NH	14 (H) × 32 (N) × 7.5 (H)	1024* (H) × 24* (N) × 64* (C)	4.7 (H) × 117.1 (N) × 3.0 (H)	16	1.1	38
3D CC(CO)NH	14 (H) × 32 (N) × 70 (C)	1024* (H) × 42* (N) × 64* (C)	4.7 (H) × 117.1 (N) × 42 (C)	16	1.1	57
3D ¹⁵ N-edited TOCSY	12.5 (H) × 32 (N) × 12.5 (H)	1024* (H) × 20* (N) × 50* (H)	4.7 (H) × 117.1 (N) × 4.7 (H)	16	1.1	24
3D HCCH-TOCSY	6.5 (H) × 74 (C) × 6.5 (H)	512* (H) × 38* (C) × 100* (H)	3.2 (H) × 45.2 (C) × 1.5 (H)	16	1.1	90
2D (HB)CB(CGCD)HD	15 (H) × 33 (C)	750* (H) × 32* (C)	4.7 (H) × 35 (C)	32	1.5	1
2D (HB)CB(CGCDDE)HE	15 (H) × 33 (C)	750* (H) × 32* (C)	4.7 (H) × 35 (C)	32	1.5	1
Distance restraints						
3D ¹⁵ N-edited NOESY	15 (H) × 27 (N) × 12.5 (H)	1024* (H) × 46* (N) × 128* (H)	4.7 (H) × 119 (N) × 4.7 (H)	8	1.2	80
3D ¹³ C-edited NOESY	14 (H) × 38 (C) × 6 (H)	768* (H) × 94* (C) × 73* (H)	4.7 (H) × 29 (C) × 2.8 (H)	16	1.1	161
3D ¹³ C-edited NOESY(aromatic region)	14 (H) × 23 (C) × 6 (H)	832* (H) × 36* (C) × 50* (H)	4.7 (H) × 123.4 (C) × 7.3 (H)	16	1.1	43
Dihedral restraints						
3D HNHB	12.5 (H) × 32 (N) × 12.5 (H)	1024* (H) × 21* (N) × 64* (H)	4.7 (H) × 117.1 (N) × 4.7 (H)	16	1.2	35
3D HAHBCACONH	12.5 (H) × 32 (N) × 12.5 (H)	1024* (H) × 10* (N) × 61* (H)	4.7 (H) × 117.1 (N) × 2.7 (H)	128	1.2	134

^a NMR backbone and side chain spectra as well as ²J data were recorded at 600 MHz, whereas distance restraint experiments were recorded at 900 MHz.

Table 2³J couplings and the subsequently derived side chain rotamer used in structure determination of the WW3* domain.

Residue	³ J coupling (Hz)		χ_1 angle
	³ J N β	³ J $\alpha\beta$	
N434	2.15 ± 0.89, 3.64 ± 0.50	3.42 ± 1.02, 4.38 ± 0.79	<i>gauche-trans</i>
D441	0.58 ± 0.19, 0.95 ± 0.12	N.D., N.D.	<i>gauche+</i>
H442	4.07 ± 0.09, 1.73 ± 0.22	3.06 ^a , 11.14 ± 1.19	<i>trans</i>
D451	1.15 ± 0.11, 0.85 ± 0.15	N.D., N.D.	<i>gauche+</i>
R453	1.43 ± 0.09, 0.85 ± 0.15	4.12 ± 1.06, 10.31 ± 0.37	<i>gauche+</i>

N.D. Not determined

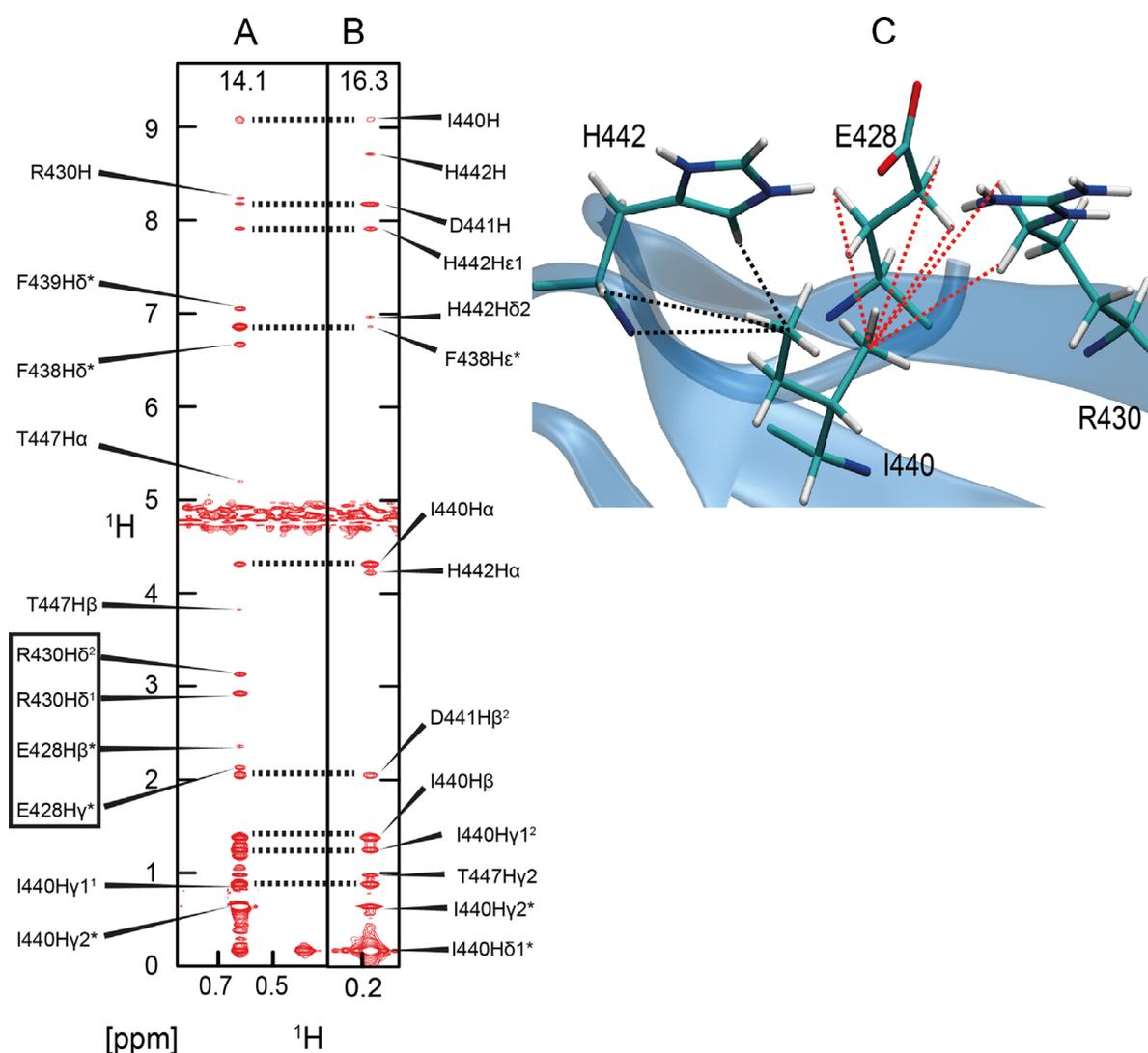
^a upper limit value for the ³J coupling.

Fig. 1. Strips from a ¹³C-edited NOESY spectrum for the δ_1 methyl protons (A) and the γ_2 methyl protons (B) of the residue I440 of the WW3* domain are shown. The ¹³C chemical shifts are shown at the top of each strip. The NOE network that gives rise to the *trans* rotamer for I440 is mapped onto the structure (C). The γ_2 methyl protons show NOEs to the β and γ protons of E428 as well as the δ protons of R430 (red dashed lines in Fig. 1C). The δ_1 methyl protons of I440 do not show NOEs to E428 and R430 but show NOEs to the amide proton and the α proton of H442 (black dashed lines in Fig. 1C). This NOE pattern defines the side chain conformation of I440.

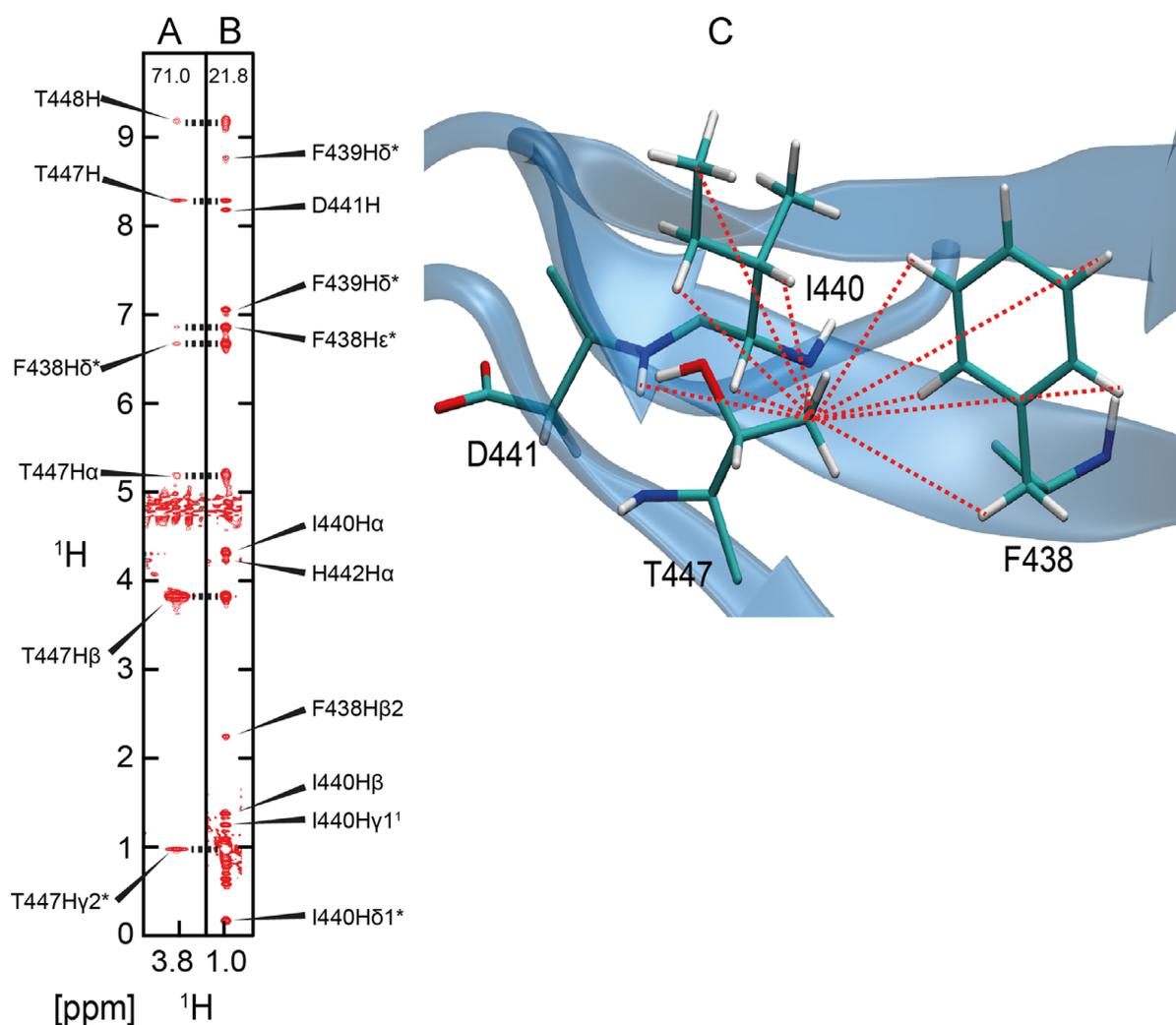


Fig. 2. Strips from a ^{13}C -edited NOESY spectrum for the β proton (A) and the γ_2 methyl protons (B) of the residue T337 of the WW3* domain are shown. The NOE network that gives rise to a *gauche+* rotamer is mapped onto the structure (C). This NOE pattern defines the side chain conformation of T447.

2. Experimental design, materials and methods

2.1. Protein expression, purification and NMR sample preparation

The WW3* domain (41 residues, 4.8 kDa) from neuronal precursor cell expressed developmentally down-regulated gene 4-1 (Nedd4-1) was overexpressed in *E. coli* BL21 (DE3)pLysS cells, as described previously [7,8]. Protein purification was performed as described previously [1,7,8].

2.2. NMR spectroscopy

Standard heteronuclear multidimensional NMR experiments [9] were performed on samples containing 1.5–1.8 mM WW3* domain (^{13}C , ^{15}N -labeled) from human Nedd4-1 in 20 mM sodium phosphate buffer (pH 6.5), 50 mM NaCl, 0.1% (w/v) NaN_3 and 1 mM DSS in a 93%/7% (v/v) $\text{H}_2\text{O}/\text{D}_2\text{O}$

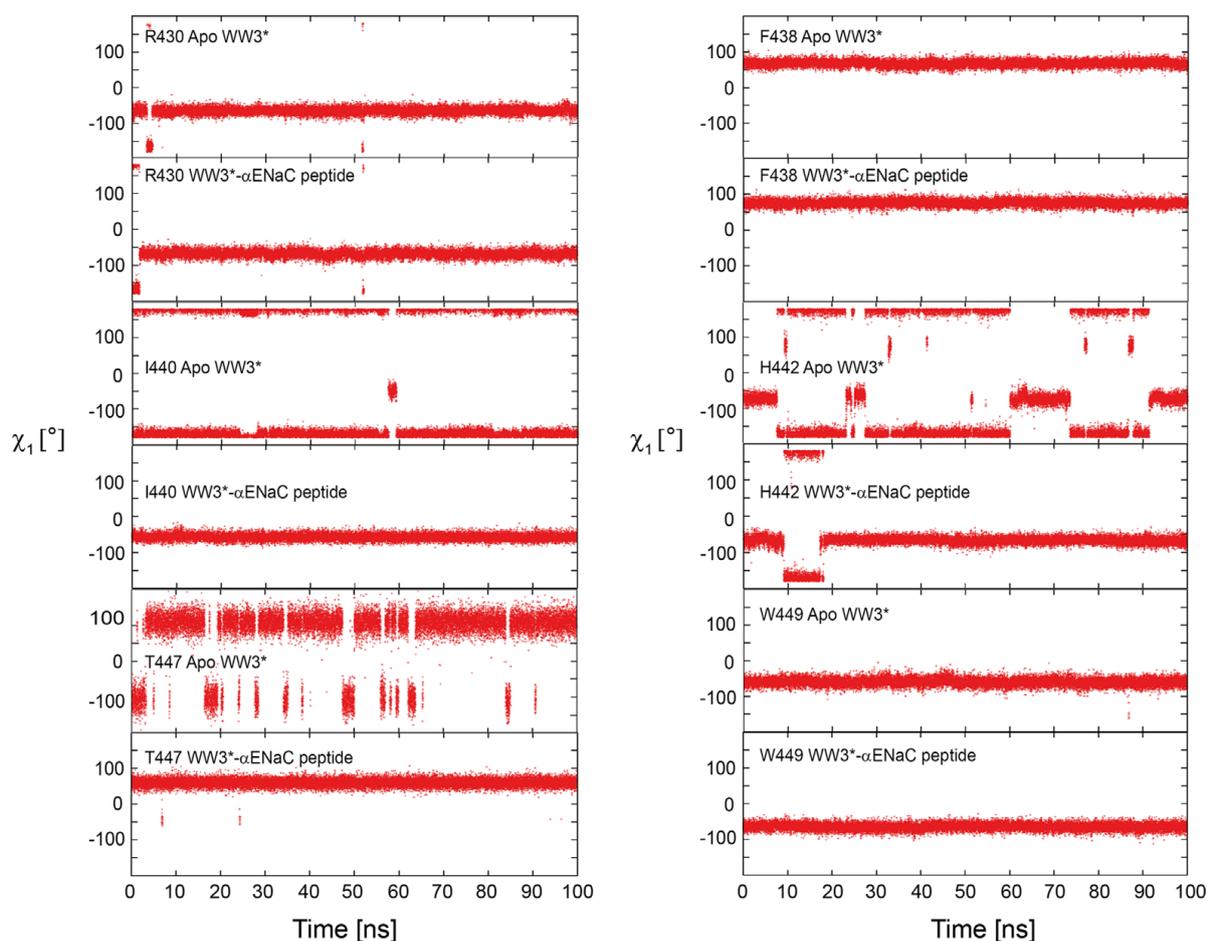


Fig. 3. Plots of side chain rotameric states for key peptide binding residues (R430, F438, I440, H442, T447 and W449) observed over 100 ns MD simulations of the apo- and hENaC peptide bound forms of the WW3* domain are shown.

mixture. NMR spectra were recorded at 25 °C on NMR spectrometers equipped with cryogenically cooled z-gradient probes operating at ^1H frequencies of 600 and 900 MHz. ^1H , ^{15}N and ^{13}C chemical shift assignments of the WW3* domain were obtained using experiments in Table 1. An example of a backbone sequential walk using three-dimensional (3D) HNCA and CBCA(CO)NH spectra between residues F438 and H442 is presented in Fig. 4. Near complete backbone (193/200 or 96.5%) and side chain assignments (302/319 or 94.5%) were obtained. To derive NOE distance restraints for structure calculation, ^{15}N -edited and ^{13}C -edited NOESY spectra were recorded using mixing times between 150 and 180 ms. Backbone dihedral angles were obtained from TALOS+ [10] using a combination of backbone ($^1\text{H}_\text{N}$, $^1\text{H}_\alpha$, $^{13}\text{C}_\alpha$, $^{13}\text{C}'$ and ^{15}N) and $^{13}\text{C}_\beta$ chemical shifts. Sidechain χ_1 dihedral angles were obtained from a combination of $^3J_{\alpha\beta}$ and $^3J_{\text{N}\beta}$ couplings derived from 3D HNHB [11] and 3D HAHB (CACO)NH [12] experiments (Table 2).

2.3. MD simulations

MD simulations were performed using parameters described in [1].

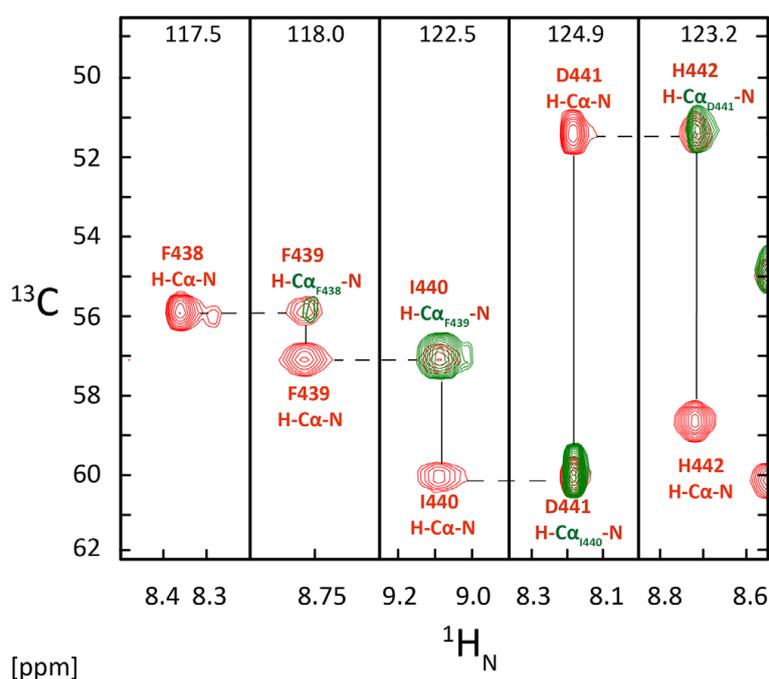


Fig. 4. Strips from 3D HNCA (red) and 3D CBCA(CO)NH (green) spectra illustrating the backbone sequential walk from F438 to H442 of the WW3* domain. The ^{15}N chemical shift is shown at the top of each strip.

Acknowledgments

Access to the Jülich-Düsseldorf Biomolecular NMR Center that is jointly run by the Forschungszentrum Jülich and Heinrich-Heine-Universität Düsseldorf is gratefully acknowledged.

Transparency document. Supporting information

Transparency document associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2016.06.024>.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2016.06.024>.

References

- [1] V. Panwalkar, P. Neudecker, M. Schmitz, J. Lecher, M. Schulte, K. Medini, M. Stoldt, M.A. Brimble, D. Willbold, A.J. Dingley, The Nedd4-1 WW domain recognizes the PY motif peptide through a coupled folding-binding equilibrium, *Biochemistry* (2015).
- [2] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR* 6 (1995) 277–293.
- [3] W.F. Vranken, W. Boucher, T.J. Stevens, R.H. Fogh, A. Pajon, M. Llinas, E.L. Ulrich, J.L. Markley, J. Ionides, E.D. Laue, The CCPN data model for NMR spectroscopy: development of a software pipeline, *Proteins Struct. Funct. Bioinform.* 59 (2005) 687–696.
- [4] W. Rieping, M. Habeck, B. Bardiaux, A. Bernard, T.E. Malliavin, M. Nilges, ARIA2: automated NOE assignment and data integration in NMR structure calculation, *Bioinformatics* 23 (2007) 381–382.
- [5] A.T. Brunger, Version 1.2 of the crystallography and NMR system, *Nat. Protoc.* 2 (2007) 2728–2733.
- [6] S. Qi, M. O’Hayre, J.S. Gutkind, J.H. Hurley, Structural and biochemical basis for ubiquitin ligase recruitment by arrestin-related domain-containing protein-3 (ARRDC3), *J. Biol. Chem.* 289 (2014) 4743–4752.

- [7] J.S. Lott, S.J. Coddington-Lawson, P.H. Teesdale-Spittle, F.J. McDonald, A single WW domain is the predominant mediator of the interaction between the human ubiquitin-protein ligase Nedd4 and the human epithelial sodium channel, *Biochem. J.* 361 (2002) 481–488.
- [8] R. Bobby, K. Medini, P. Neudecker, T.V. Lee, M.A. Brimble, F.J. McDonald, J.S. Lott, A.J. Dingley, Structure and dynamics of human Nedd4-1 WW3 in complex with the α ENaC PY motif, *Biochim. Biophys. Acta BBA Proteins Proteom.* 1834 (2013) 1632–1641.
- [9] A.J. Dingley, I. Lorenzen, J. Grötzinger, NMR analysis of viral protein structures, *Plant Virology Protocols: From Viral Sequence to Protein Function*, 2008, pp. 441–462.
- [10] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts, *J. Biomol. NMR* 44 (2009) 213–223.
- [11] P. Düx, B. Whitehead, R. Boelens, R. Kaptein, G.W. Vuister, Measurement of ^{15}N - ^1H coupling constants in uniformly ^{15}N -labeled proteins: application to the photoactive yellow protein, *J. Biomol. NMR* 10 (1997) 301–306.
- [12] N. Vajpai, M. Gentner, J.-r Huang, M. Blackledge, S. Grzesiek, Side-chain χ_1 conformations in urea-denatured ubiquitin and protein G from ^3J coupling constants and residual dipolar couplings, *J. Am. Chem. Soc.* 132 (2010) 3196–3203.

3.5 Proline Restricts Loop I Conformation of the High Affinity WW Domain from Human Nedd4-1 to a Ligand Binding-Competent Type I β -Turn

Summary

The four WW domains from hNedd4-1, despite having high sequence similarity, show differing peptide binding affinities towards the PY motif of the α -hENaC peptide. Sequence alignment of the four WW domains from hNedd4-1 reveals that the highest sequence diversity lies in loop I, suggesting its role in defining the WW domain peptide affinities. Loop I of the high affinity domain WW3* adopts a type I β -turn and exhibits the statistically preferred proline at the $i+1$ position found in β -turns. The stability and dynamics of loop I were characterized using a combination of MD simulation and experimental techniques. Mutation of the proline in $i+1$ position to a threonine (WW3* P433T), the equivalent residue in WW4, shifts the loop I conformation from a predominantly type I β -turn to an open, seven-residue Ω loop. Transitions between these two states in both, wild-type WW3* and WW3* P433T, were further investigated using cluster analysis. MD simulations and hetNOE values suggest that the WW3* P433T- α -hENaC peptide interaction locks the loop I into a type I β -turn. Hence, presence of proline at the $i+1$ position, results in a stable, peptide-binding competent, type I β -turn, which may affect peptide binding affinity.

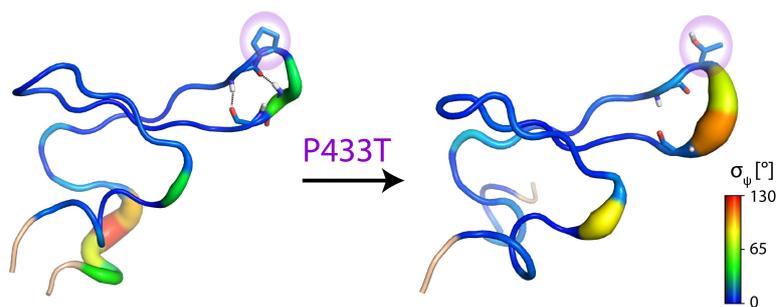


Figure 3.2: Graphical table of contents

Acknowledgments Reprint Permission

This section contains a complete reprint of the publication and supporting information. Reprinted with permission from *The Journal of Physical Chemistry B* **2018**, 122 (15), 4219-4230, doi:[10.1021/acs.jpcc.7b11637](https://doi.org/10.1021/acs.jpcc.7b11637). Copyright 2018 American Chemical Society.

Proline Restricts Loop I Conformation of the High Affinity WW Domain from Human Nedd4-1 to a Ligand Binding-Competent Type I β -Turn

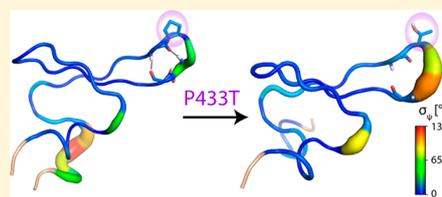
Marianne Schulte,^{†,§} Vineet Panwalkar,^{†,§} Stefan Freischem,^{†,§} Dieter Willbold,^{†,§} and Andrew J. Dingley^{*,†,§}

[†]ICS-6 (Strukturbiochemie), Forschungszentrum Jülich, 52425 Jülich, Germany

[§]Institut für Physikalische Biologie, Heinrich-Heine-Universität, 40225 Düsseldorf, Germany

Supporting Information

ABSTRACT: Sequence alignment of the four WW domains from human Nedd4-1 (neuronal precursor cell expressed developmentally down-regulated gene 4-1) reveals that the highest sequence diversity exists in loop I. Three residues in this type I β -turn interact with the PPxY motif of the human epithelial Na⁺ channel (hENaC) subunits, indicating that peptide affinity is defined by the loop I sequence. The third WW domain (WW3*) has the highest ligand affinity and unlike the other three hNedd4-1 WW domains or other WW domains studied contains the highly statistically preferred proline at the ($i + 1$) position found in β -turns. In this report, molecular dynamics simulations and experimental data were combined to characterize loop I stability and dynamics. Exchange of the proline to the equivalent residue in WW4 (Thr) results in the presence of a predominantly open seven residue Ω loop rather than the type I β -turn conformation for the wild-type apo-WW3*. In the presence of the ligand, the structure of the mutated loop I is locked into a type I β -turn. Thus, proline in loop I ensures a stable peptide binding-competent β -turn conformation, indicating that amino acid sequence modulates local flexibility to tune binding preferences and stability of dynamic interaction motifs.



INTRODUCTION

WW domains are small interaction modules that are ~40 residues in length with two conserved tryptophans and an invariant proline. These domains are present in eukaryotic signaling and structural proteins and function by mediating protein–protein interactions in cell signaling networks.^{1,2} WW domains adopt a three-stranded twisted antiparallel β -sheet fold with two intervening loops.³ Structures of WW domains in complex with peptides show that the first tryptophan and the invariant proline are required for domain stability, whereas the second tryptophan is essential for peptide recognition and is one of two conserved aromatic residues that form the hydrophobic “XP pocket” that binds a conserved proline in proline-rich sequences recognized by WW domains.^{2–8}

Depending on species and isoform, the E3 ubiquitin ligase Nedd4 (neuronal precursor cell expressed developmentally downregulated gene 4), which targets membrane proteins for endocytosis,⁹ contains 3–4 WW domains sandwiched between a C2 domain and a C-terminal HECT ubiquitin ligase domain. Human Nedd4 isoform 1 (hNedd4-1) contains four WW domains that bind the alpha human epithelial Na⁺ channel (α -hENaC) (L/P)PxY motif (PY motif) with differing affinity¹⁰ with the third domain (WW3*, where the asterisk designates a WW domain with high affinity for targets) showing the highest affinity. Sequence alignment of the four hNedd4-1 WW domains shows that of the 10 residues of WW3* involved in direct interaction with the α -hENaC peptide⁴ seven are

sequence identical or have a conserved substitution in the other three WW domains (Figure 1A). This high degree of sequence similarity and differing affinities of WW domains from Nedd4 extends to other species, including rat, mouse, *Drosophila*, and *Xenopus*.^{6,11,12}

The highest sequence diversity among the WW domains of hNedd4-1 exists in loop I, which forms a type I β -turn for the WW3* domain (Figure 1).^{4,13,14} Three residues in this β -turn of WW3* interact with the α -hENaC peptide,⁴ indicating that peptide affinity of the four WW domains is defined primarily by the sequence of loop I. Mutagenesis studies of Nedd4 have demonstrated that exchanging residues in loop I of the WW domains affects peptide binding affinity.^{11,12} Here, exchanging Ala and Pro of the loop I sequence (A432–R436) from hNedd4 WW3* with the corresponding His and Thr in WW4 decreased binding to the ENaC peptide by ~3-fold.¹¹ In the subsequent study of *Drosophila* Nedd4,¹² it was shown that the APN sequence in loop I is required for defining high affinity binding of Nedd4 WW3* domains. Ala at this position in loop I was considered to be sterically more suited because it creates a larger XP groove that accommodates tighter ligand binding, whereas the Pro restricts the dynamics of loop I in the apo-state, which may yield a smaller entropic cost upon ligand

Received: November 27, 2017

Revised: March 29, 2018

Published: March 29, 2018

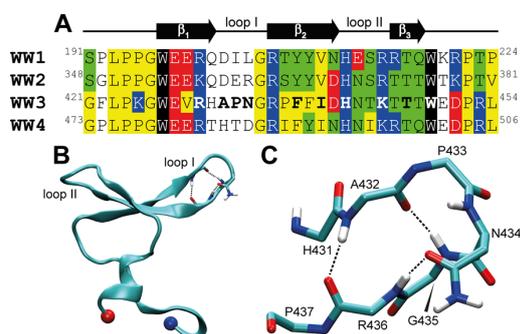


Figure 1. (A) Alignment of the four human Nedd4-1 WW domains. Sequence homology between domains is highlighted for nonpolar (yellow), polar (green), acidic (red), and basic (blue) amino acids, and the conserved tryptophans are highlighted in black. The secondary structure is shown above the sequence alignment, and bold lettering indicates those residues of the WW3* domain that are involved in peptide interaction (R430, A432, P433, N434, F438, I440, H442, K445, T447, and W449). Sequence numbering is according to NCBI sequence entry NM_006154.3, and the sequence alignment was performed with CLUSTALW.³⁰ This figure is a modified version from Panwalkar et al.¹³ (B) Energy minimized starting structure of the apo-state wild-type hNedd4-1 WW3* structure. H-bonds that provide loop I stability are shown. Blue and red spheres refer to the N- and C-termini, respectively. (C) Stick representation of the backbone and N434 side chain of the loop I substructure (A432–R436) in wild-type hNedd4-1 WW3*. Backbone G435–A432 and A432–R436 H-bonds, and the H-bond between the side chain carbonyl oxygen of the carboxamide group of N434 and the backbone amide group of R436 are indicated by black dashed lines.

binding and domain folding.^{11,12} Moreover, the Pro residue in WW3* provides contacts with the target peptide, whereas in the structure of the rat Nedd4-1 WW4–ENaC peptide complex the corresponding Thr residue does not interact with the peptide.^{11,12,14}

WW domains have been used extensively as models to study the kinetics and thermodynamics of β -sheet folding.^{15–25} Kinetic studies show that WW domains fold in less than 100 μ s,^{13,16,19,20,26} and this folding rate is sufficiently fast to make molecular dynamics simulations of WW folding feasible.^{27,28} For the folding of WW domains from human peptidyl-prolyl isomerase NIMA-interacting 1 (hPin1) protein, Formin-binding protein 28 (FBP28) and human yes-associated protein (hYAP), formation of loop I is the rate-limiting step.^{16,20,21} Mutational and kinetic analysis of the Pin1 WW domain showed that reducing the length of the 6-residue loop I to the frequently found 5-residue type I G-bulge β -turn of WW domains increased the stability and rate of folding of the domain but compromised ligand binding affinity.²¹ This observation highlights that the length of loop I of the Pin1 WW domain is optimized for function at the expense of folding energetics. Moreover, shortening loop I of the Pin1 WW domain simultaneously reduces flexibility and peptide binding affinity,²⁹ indicating that the intrinsic disorder and flexibility of this loop has also been selected for function.

These studies describing the role of loop I in WW domain folding and stability have focused on the human Pin1 WW domain (loop I sequence: SRSSGR), the FBP28 WW domain (TADGK) and the human YAP65 WW domain (TSSGQ). However, none of these WW domains contain a Pro at the ($i + 1$) position of type I β -turns, which has the highest statistical

occurrence at this position in type I β -turns.^{31,32} Consequently, characterizing the role of loop I from the WW3* domain of Nedd4-1, which contains a Pro at position ($i + 1$) in loop I, should augment results describing the function of loop I in WW domain dynamics, stability, folding and peptide recognition. In this report, the role of P433 in loop I of hNedd4-1 WW3* to conformation, flexibility and overall domain stability was examined by molecular dynamics (MD) simulations and complemented with experimental data. The results showed that mutation of P433 causes loop I to adopt a dynamic seven residue open loop conformation rather than the stable type I 3:5 β -turn. Thus, proline in loop I ensures a stable ligand binding-competent β -turn conformation, which likely affects the apparent peptide ligand affinity and may explain the lower affinity of the other hNedd4-1 WW domains toward the α -hENaC peptide.

EXPERIMENTAL SECTION

MD Simulations. MD simulations of the apo and the α -hENaC peptide-bound forms of the wild-type and P433T mutant of the hNedd4-1 WW3* domain were performed using the AMBER99SBnmr1-ILDN force field³³ with GROMACS version 4.6.5.³⁴ As starting structures for wild-type simulations, the NMR-derived lowest-energy structures of the apo-state (PDB ID: SAHT) and peptide-bound state (PDB ID: 2M3O) were used. The P433T mutation was introduced by using the tool “Rotamers” provided in Chimera (version 1.10.2)³⁵ and the Dunbrack rotamer library.³⁶ Prior to MD simulations, the structures were placed into cubic boxes with all protein atoms at a distance ≥ 1 nm from the box edges and solvated with the TIP4P³⁷ water model. Sodium and chloride ions were added as counterions to neutralize the total charge of the system and mimic a salt concentration that matched the experimental conditions followed by an energy minimization via the steepest descent method. The numbers of ions, protein, and solvent atoms, along with further details on the MD simulations are provided in Table S1.

One nanosecond equilibrations were conducted for each system in the NVT ensemble (constant number of particles N , constant volume V , and temperature T) with a final temperature of 25 °C, followed by a 1 ns equilibration using the NPT ensemble (constant pressure P) with the Parrinello–Rahman barostat³⁸ at 1 atm. The backbone root-mean-square-deviation (RMSD) of loop I between the wild-type and P433T mutant starting structures after energy minimization and equilibration was 0.29 Å for the apo-state and 0.36 Å in the peptide bound state.

Production runs were performed in the isothermal–isobaric NPT ensemble at 25 °C, using the velocity-rescaling thermostat³⁹ with a coupling constant of 0.1 s. Pressure was maintained at 1 atm using the Parrinello–Rahman barostat³⁸ with a time constant for pressure coupling of 2 ps. For constraining bond lengths, the LINCS algorithm⁴⁰ was used allowing a 2 fs time step. Electrostatic interactions were treated with the Particle-Mesh-Ewald algorithm^{41,42} with a Fourier grid spacing of 1.6 Å and cubic interpolation. Moreover, the minimum distance between the protein and its periodic images was at all times larger than the cutoff distance for short-range interactions, ensuring that the protein did not interact with its images during the simulations (Table S1). Production runs were conducted for 100 ns using the leapfrog integrator.

Multiple shorter (50 ns each) simulations were performed using different starting structures taken from the cluster analysis

(see below) of the apo-WW3* domain ($n = 20$ for wild-type and P433T mutant) and the WW3*- α -hENaC peptide complex ($n = 10$ for wild-type and P433T mutant). Energy minimization and equilibration was performed as described above.

MD Trajectory Analysis. Conformations of loop I sampled during the 100 ns simulation at time intervals of 4 ps were grouped according to RMSD. This clustering analysis of loop I (H431–P437) was performed using the gmx cluster utility of GROMACS 5.1³⁴ with the GROMOS clustering algorithm⁴³ and a backbone RMSD cutoff of 0.5 Å. The GROMOS algorithm first calculates the RMSD between all structures using only the backbone atoms of loop I. All structures that have an RMSD lower than the cutoff (0.5 Å) are considered as neighbors. The structure with the largest number of neighbors is considered as the representative structure of the largest cluster (cluster 1) and is eliminated from the pool of clusters along with all its neighbors. This procedure is repeated until there are no structures left in the pool. With all structures assigned to a cluster, it is possible to illustrate the clusters sampled over the simulation time and to define clusters that are directly or indirectly connected. To display the connectivity between clusters or main conformations, respectively, the program Gephi version 0.9.1⁴⁴ was used to generate network figures. GROMACS 5.1 gmx hbond tool with default settings (i.e., $d_{\text{N}\cdots\text{O}} \leq 0.35$ nm, and $\angle\text{N-H}\cdots\text{O} = 180 \pm 30^\circ$) was used to calculate the presence of H-bonds. The gmx mindist utility was used to calculate the distance between H-bond donor and acceptor atoms ($d_{\text{H}\cdots\text{O}}$) in the 50 ns simulations.

Backbone chemical shifts (i.e., $C\alpha$, C' , N , and H_N), and ϕ and ψ torsion angles were calculated for each frame of the MD trajectories using SPARTA+.⁴⁵

Cloning, Protein Production, and Purification. Mutagenesis to generate the P433T mutation in loop I was performed using the QuikChange II site-directed mutagenesis kit (Agilent Technologies, Santa Clara, CA, U.S.A.) according to the manufacturer's instructions with the forward 5'-GAAGTCCGGCATGCAACAATGGGAGGCCTT-3' and reverse 5'-AAGGCCTCCCATTGTTGCATGCCGGA-CTTC-3' primers (mutation codon is underlined). Recombinant protein production and purification of wild-type and the P433T mutant hNedd4-1 WW3* domain was carried out as described previously.¹⁰

NMR Spectroscopy. [U -¹³C, ¹⁵N] and [U -¹⁵N] P433T WW3* mutant samples and the [U -¹⁵N] WW3* wild-type sample were prepared in 20 mM sodium phosphate buffer (pH 6.5), 50 mM NaCl, 0.1% (w/v) NaN₃, and 1 mM DSS in a 90%/10% H₂O/D₂O mixture. The concentration of the WW domain samples was 0.8 mM. Samples of the WW domain in complex with the unlabeled α -hENaC peptide were prepared at a molar peptide to domain ratio of 5:1. NMR spectra were recorded on a Bruker Avance III spectrometer (Bruker BioSpin GmbH, Rheinstetten, Germany) equipped with a cryogenically cooled z-gradient probe and operating at a ¹H frequency of 600 MHz. Backbone ¹H, ¹⁵N and ¹³C assignments of the hNedd4-1 P433T WW3* mutant were obtained from standard multi-dimensional heteronuclear NMR experiments at 5 °C.⁴⁶ Proton chemical shifts were referenced to DSS, whereas the ¹⁵N and ¹³C chemical shifts were referenced indirectly according to the ratios given by Wishart et al.⁴⁷ A series of 2D ¹H–¹⁵N HSQC spectra over a temperature range of 5–25 °C were recorded to map the 5 °C backbone amide chemical shifts to the 2D ¹H–¹⁵N HSQC spectrum at 25 °C. ¹H–¹⁵N heteronuclear NOE values for the hNedd4-1 WW3* domain and the P433T

variant in the apo- and peptide-bound states at 5 °C were derived from pairs of interleaved spectra recorded with (NOE) and without (control) proton saturation during the recycle delay, which was 14 s.⁴⁸ The ¹H–¹⁵N NOE values were calculated from peak intensity ratios obtained from the NOE and control spectra, with uncertainties estimated from background noise of the spectra. Data sets were processed using NMRPipe⁴⁹ and analyzed using CcpNMR Analysis.⁵⁰

RESULTS AND DISCUSSION

Loop I of the Wild-Type hNedd4-1 WW3* Domain Adopts a Type I β -turn. Structural data shows that the first loop of the wild-type hNedd4-1 WW3* domain exists as a type I 3:5 β -turn with a G1 beta-bulge consisting of residues A432 (i), P433 ($i + 1$), N434 ($i + 2$), G435 ($i + 3$), and R436 ($i + 4$).^{4,13,14} This β -turn is characterized by the canonical backbone–backbone hydrogen bonds (H-bonds) between the N–H of G435 and the C=O of A432, and the N–H of A432 and the C=O of R436 (Figure 1B,C). Proline has the highest statistical occurrence at the ($i + 1$) position for type I β -turns because the pyrrolidine ring restricts the ϕ torsion angle to the favored value of -60° .³² In the crystal structure of WW3* (PDB ID: 4N7F), P433 adopts the conventional ϕ and ψ torsion angles for type I turns of -60° and -25° , respectively. Additionally, Asn and Gly at the ($i + 2$) and ($i + 3$) positions, respectively, are also statistically favored in type I β -turns.^{32,51} The dihedral angles of N434 ($\phi = -86^\circ$, $\psi = 3^\circ$) and G435 ($\phi = 89^\circ$, $\psi = 2^\circ$) are very close to the ideal values of the ($i + 2$) position for a type I β -turn ($\phi = -90^\circ$, $\psi = 0^\circ$) and the typically found backbone torsion angles for the ($i + 3$) position of a left-handed helix, respectively, which facilitates the return of the backbone to run antiparallel to its original direction after completion of the turn. Asn at the ($i + 2$) has also been shown to be favored⁵¹ because the side chain carboxamide group forms an H-bond with the backbone amide group of R436 due to the side chain adopting the gauche⁻ rotamer state (PDB ID: 4N7F; Figure 1B,C). Thus, the first loop in the hNedd4-1 WW3* domain contains amino acids with high positional potentials to promote the formation of the stable type I β -turn observed in the solution and crystal structures.^{4,13,14} In support of loop I forming a stable β -turn, NMR backbone dynamics analysis of the apo-state of the WW3* domain at 25 °C¹³ showed that the amplitude of motions (i.e., square of the generalized order parameter (S^2)) for residues in loop I (average $S^2 = 0.82$, $n = 4$) are very similar to residues located in the stable β -sheet (average $S^2 = 0.80$, $n = 12$), indicating that loop I of hNedd4-1 WW3* adopts a relatively rigid substructure on the pico- to nanosecond (ps–ns) time scale. The importance of P433 in stabilizing the type I β -turn was examined by exchanging this amino acid to Thr; the corresponding amino acid present in loop I of hNedd4-1 WW4 (Figure 1A). Ile and Glu located at the ($i + 1$) position in Nedd4-1 WW1 and WW2, respectively, were not selected because the van der Waals volumes are larger (Glu, 109 Å³; Ile, 169 Å³) than Pro (90 Å³) and Thr (93 Å³). Moreover the choice of Glu would introduce a negative charge to the loop region that may facilitate the formation of a salt bridge or H-bond with the proximal H431 side chain, which would likely affect the β -turn structure and stability. Alanine was also considered, however, this amino acid has a significantly lower van der Waals volume (67 Å³) when compared with that of proline, and Thr and Ala have similar ($i + 1$) potentials in type I β -turns, that is, 1.07 versus 0.94.³² Thus, the P433T mutation

was considered a suitable choice for this study. 2D ^1H - ^{15}N HSQC spectra of the wild-type WW3* domain and the P433T mutant were recorded to investigate structural changes to the WW3* caused by the mutation (Figure S1). Overlay of the 2D ^1H - ^{15}N HSQC spectra showed that the mutation did not affect the overall fold of the WW domain with only small chemical shift differences for resonances arising from residues located in loop I and in close structural proximity of loop I (R430, H431, and W449).

Loop I of the hNedd4-1 WW3* Domain Samples The Largest Conformational Space with the P433T Mutation Leading to Greater Conformational Heterogeneity and Flexibility of Loop I. Variation of the backbone conformation of the wild-type hNedd4 WW3* domain and the P433T mutant in the apo- and peptide-bound states was determined by calculating the standard deviation (σ) of ϕ and ψ torsion angles for all structural snapshots taken during each simulation (Table S2). The majority of residues in the core β -sheet region (K425–E450) for both wild-type and P433T have ϕ and ψ torsion angle deviations (σ_ϕ , σ_ψ) that show limited conformational sampling as illustrated in Figure 2, which shows σ_ψ values

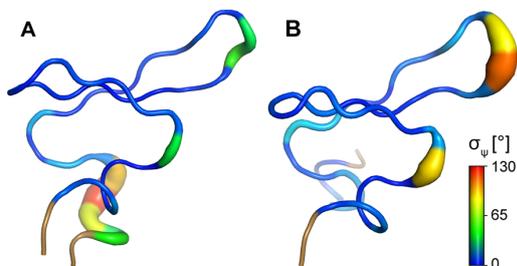


Figure 2. Standard deviations of the ψ torsion angle (σ_ψ) describing backbone conformational flexibility during the 100 ns simulations for (A) the apo-state wild-type WW3* domain and (B) the apo-state P433T mutant. The σ_ψ values were mapped onto the cluster 1 structure of each simulation (Figure 7). Structures are shown in tube representation and residues colored according to their σ_ψ value.

mapped onto the WW3* structure. Only residues N434 (50°) and G435 (62°) located in loop I, and E450 (51°) in the core β -sheet region of the apo-state of the wild-type WW3* domain have σ_ψ values well above the mean σ_ψ of 18° ($n = 26$) (Figure 2A), showing that loop I samples more conformational space than other residues in this core region. The large σ_ψ value for E450 is hypothesized to arise from transient side chain interactions with R436, which has a σ_ϕ value (Table S2) that is above the mean σ_ϕ value. A similar result is observed for the P433T mutant (Figure 2B); however, both residues N434 (79°) and G435 (108°) have larger σ_ψ values when compared with the mean σ_ψ of 24° ($n = 26$) for the core region and the values derived from the wild-type WW3* domain. The observed increase in σ_ψ values for N434 and G435 of the P433T mutant shows that the proline at the ($i + 2$) position restricts the conformational space that these two residues sample in the wild-type WW3* domain. In contrast to loop I, the backbone of residues in loop II displays a much narrower torsion angle distribution because of the extensive network of hydrogen bonds for this type I 4:6 β -turn affording a rigid substructure.^{14,52} In the presence of the α -hENaC peptide the σ_ψ and σ_ϕ values for residues in loop I decrease and are comparable between wild-type and P433T WW3* (Table S2,

Figure S2), showing that upon peptide binding loop I samples a more confined conformational space.

We have shown by NMR dynamics that the core β -sheet region of the wild-type WW3* domain at 25°C shows limited mobility on the ps–ns time scale.¹³ Thus, to characterize the effect of the P433T mutation to internal motions of the WW3* domain, the $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE experiment was recorded. However, severe line-broadening of many resonances in this HSQC-type experiment due to chemical exchange processes precluded collection of a satisfactory set of complementary data for the P433T mutant at 25°C , especially data for residues adjacent and part of loop I. We hypothesize, based on our previous observation of a fold-unfold equilibrium of the wild-type WW3* domain,¹³ that the P433T mutation has altered the kinetics (exchange rate) and thermodynamics (relative populations) of this equilibrium to give rise to the broad line widths at 25°C .

In a previous report, line shape analysis of 2D ^1H - ^{15}N HSQC spectra recorded on the wild-type WW3* domain at different temperatures showed that line widths were narrower at lower temperatures.⁵³ Thus, $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE experiments were recorded at 5°C in an effort to obtain data describing fast internal motions. Data recorded at this temperature provided a near complete set of $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE values for both proteins (Figure 3). Lower $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE values for residues in loop I were observed for the P433T mutant when compared with that of the wild-type WW3* domain (Figure 3A,B,E; $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE differences between wild-type and P433T for residues A432, N434, and G435 were 0.063 ± 0.001 , 0.103 ± 0.008 , and 0.043 ± 0.001 , respectively), indicating larger amplitudes of ps–ns time scale motions. The increase in the $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE values upon addition of the peptide shows that peptide binding quenches ps–ns time scale motion of this loop in both wild-type and P433T WW3* domains (Figure 3C,D,F), with larger differences generally observed for loop I amide groups of the mutant (Figure 3D). In summary, the MD simulations and $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE results show that mutation of P433 leads to greater conformational freedom and mobility (fast time scale motion) of loop I without significant perturbations to other regions of the domain stability or fast time scale motion.

Backbone ^{15}N Chemical Shifts Derived From the Simulated Data. Calculation of chemical shifts from structural ensembles derived from MD trajectories provides a link between simulation and NMR (experimental) data for validation purposes, and facilitates interpretation of experimental data such as dynamically averaged experimental chemical shifts.⁵⁴ Thus, chemical shifts of backbone nuclei for each frame of the MD trajectories were calculated with a focus on the ^{15}N chemical shifts because comparisons can be made with available experimental data at 25°C . Figure 4 presents the normalized distributions of SPARTA+ predicted backbone ^{15}N chemical shifts for A432, G435, and R436 (i.e., H-bond donors) from the simulations of wild-type and mutant WW3* domains in the apo and complex states. The experimental values and the average SPARTA+ predictions are also shown. The average ^{15}N predicted chemical shifts of A432 and G435 in both apo and complex states of the WW3* domain are similar to the experimental values, suggesting the simulations describe the systems reasonably well; however, the predicted average values for R436 differ from the experimental values noticeably. This difference could arise from the simulation period being too

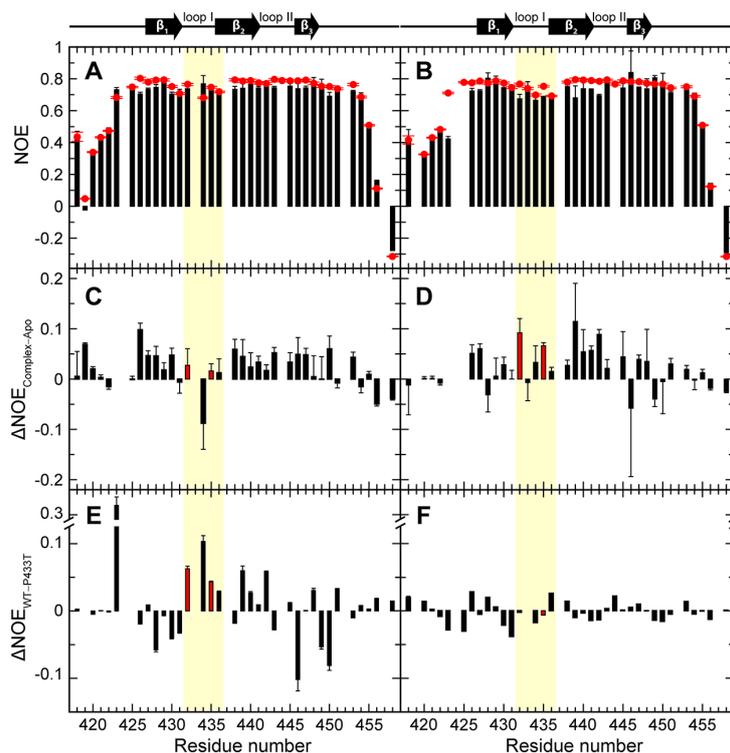


Figure 3. Backbone amide steady-state heteronuclear $\{^1\text{H}\}-^{15}\text{N}$ NOE values (hetNOE) of hNedd4-1 WW3* (A) and the P433T mutant (B) in the absence (black bars) and presence (red circles) of the α -hENaC peptide recorded at 5 °C and 14.1 T. (C,D) Changes in steady-state heteronuclear $\{^1\text{H}\}-^{15}\text{N}$ NOE values ($\Delta\text{NOE}_{\text{Complex-Apo}} = \text{hetNOE}_{\text{Complex}} - \text{hetNOE}_{\text{Apo}}$) between the complex and apo-state of the wild-type WW3* domain (C) and P433T mutant (D). Changes in steady-state heteronuclear $\{^1\text{H}\}-^{15}\text{N}$ NOE values ($\Delta\text{NOE}_{\text{WT-P433T}} = \text{hetNOE}_{\text{WT}} - \text{hetNOE}_{\text{P433T}}$) between the wild-type WW3* domain and the P433T mutant in the apo (E) and their respective α -hENaC peptide bound (F) states. The yellow shaded region highlights results for loop I and the red bars are for residues A432 and G435, which are backbone H-bond donor residues in loop I. A schematic representation of the secondary structure elements for wild-type WW3* is shown at the top of (A) and (B).

short to capture all conformational space and dynamics this residue samples. Nonetheless, the differences between experimental and predicted ^{15}N chemical shifts for R436 are in reasonable agreement (average difference = 2.31 ppm) with the standard deviation (2.45 ppm) reported in the SPARTA+ validation set of proteins.⁴⁵

The ^{15}N chemical shift distributions of A432 for both the P433T mutant and wild-type WW3* domain in the apo-state have similar broad profiles over a comparable chemical shift range. The σ_{ϕ} and σ_{ψ} values for A432 of the wild-type and P433T WW3* domains (Table S2) are modestly above the average values for the core region (and should be comparatively narrower than the distributions for G435 if torsion angles are the only factor influencing the chemical shifts; Table S2), suggesting that backbone conformational sampling by this residue does not fully explain the large chemical shift distribution observed. Thus, other structural factors also affect the A432 ^{15}N chemical shift; however, the slight narrowing of the chemical shift distributions in the complex state for wild-type and P433T WW3* domains does correlate with modest reductions in σ_{ϕ} and σ_{ψ} values.

The wider ^{15}N chemical shift distribution of G435 for the P433T WW3* domain compared with the results for the wild-type WW3* domain indicates that this residue samples a larger conformational space during the simulation of the P433T WW3* domain than that of the wild-type WW3* domain, which is in agreement with the larger backbone conformational

distribution observed for G435 in the mutant construct (Figure 2; Table S2). Moreover, as the ^{15}N chemical shift is influenced by the ψ angle of the preceding residue further substantiates the observed wider ^{15}N chemical shift distribution of G435 of the P433T mutant (Table S2; σ_{ψ} mutant for N434 = 79°, σ_{ψ} wild-type for N434 = 50°). In complex with the α -hENaC peptide the ^{15}N chemical shift distribution of G435 for both domains narrows and this correlates with the σ_{ϕ} and σ_{ψ} values, indicating that fluctuations in the backbone of G435 account predominantly for the observed chemical shift distributions.

The ^{15}N chemical shift distributions for R436 in both apo and complex states are similar. However, distributions of R436 ^{15}N shifts in both states for the wild-type WW3* domain and the P433T mutant in the complex state display apparent bimodal equilibrium distributions. Inspection of WW3* structures from the MD simulation of the wild-type WW3*–peptide complex reveals a correlation between the ^{15}N chemical shift and the χ_1 angle of R436. By binning chemical shifts based on the R436 χ_1 rotamer states (i.e., 60° (gauche⁻), -60° (gauche⁺), and 180° (trans)) we observe distinct chemical shift ranges, with average and σ of 118.7 ± 2.1, 120.2 ± 1.3, and 121.5 ± 1.5 ppm for the gauche⁻, gauche⁺, and trans rotamer states, respectively. The chemical shift averages and σ for each state overlap and thus mask the expected trimodal distribution in Figure 4. The results for the wild-type WW3* domain in the apo-state give even closer distributions with average and σ -values of 121.3 ± 1.5, 120.8 ± 1.2, and 122.2 ± 1.3 ppm for the

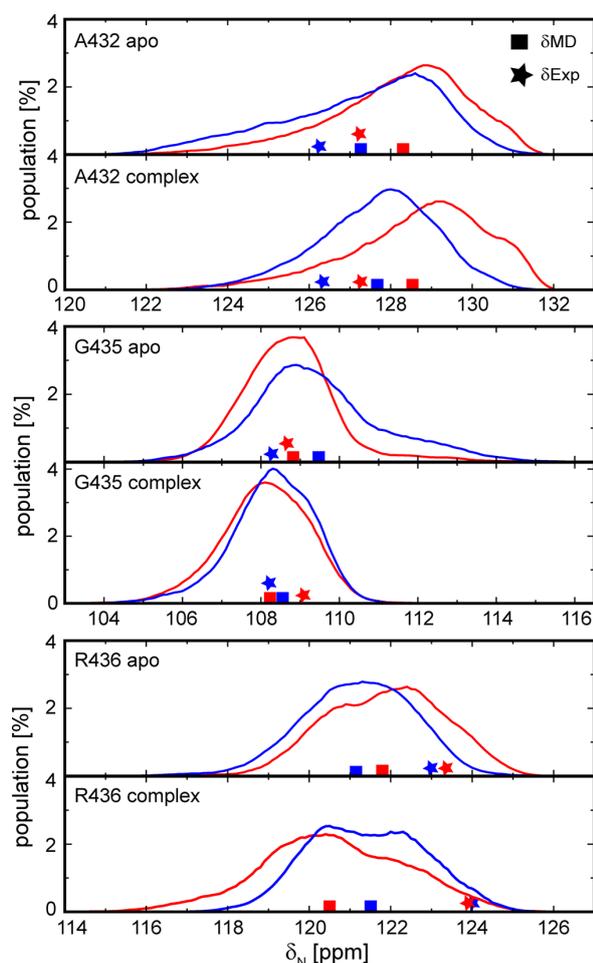


Figure 4. Normalized distributions of the SPARTA+ predicted ^{15}N chemical shift of A432, G435, and R436 from the MD trajectories of WW3* wild-type (red) and P433T mutant (blue) in both the apo and complex states. The average values of the SPARTA+ predictions over each MD trajectory are shown as squares (δ_{MD}). The experimentally measured values are shown as stars (δ_{Exp}).

gauche⁻, gauche⁺, and trans rotamer states, respectively. Differences in the ^{15}N chemical shift distributions for the rotamers of the P433T mutant are not as apparent, especially for the apo-state.

MD Simulations Reveal That Loop I of the WW3* Domain of hNedd4-1 Fluctuates Between a Four Residue β -Turn and a Seven Residue Ω Loop. The conformation of loop I of wild-type and P433T WW3* in the absence and presence of the α -hENaC peptide was examined in further detail. This was carried out by analysis of backbone torsion angles of residues R430 to P437 during the MD simulations (Figures S3–S6) and also visualized by Ramachandran plots of loop I residues (Figures S7–S11). The MD simulation of the apo-state of the wild-type WW3* domain showed that for the majority of the simulation ϕ and ψ torsion angles adopted values similar to those reported in the crystal structure for these residues (Figure S3). During the period between 30 and 47 ns ϕ and ψ torsion angles of N434 and G435 showed large deviations, whereas the ϕ and ψ torsion angles for the other residues displayed modest deviations, with

the values for P433 showing minimal changes over the entire simulation (Figure S3).

The effect of ϕ and ψ torsion angle changes to loop I stability was examined by focusing on residues G435 and A432 because the backbone amide groups of these residues are H-bond donors in the type I β -turn. Changes in the side chain χ_1 angle for residue N434 were also examined because this residue acts as an H-bond acceptor for the backbone to side chain R436–N434 H-bond (Figure 1C). Figure 5A,D shows changes in

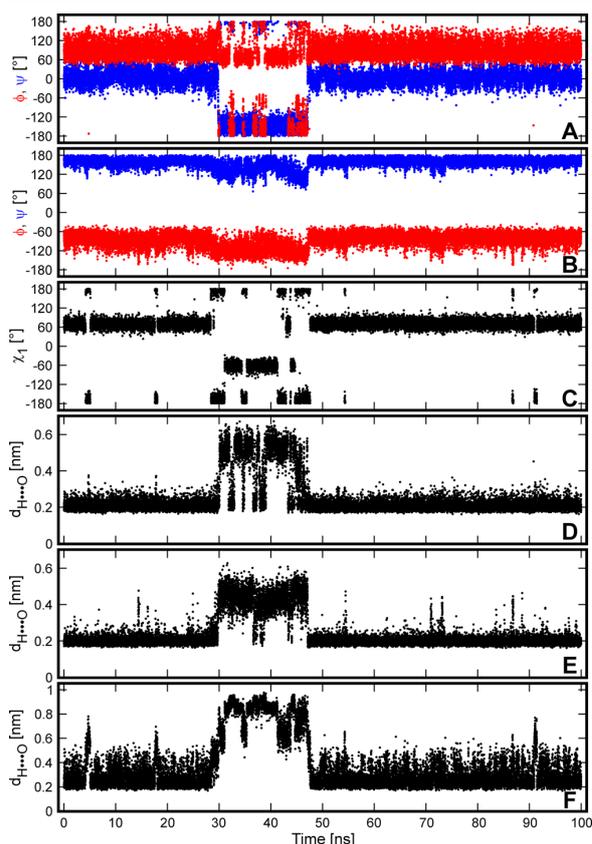


Figure 5. Torsion angles and H-bond distances of the apo-state wild-type WW3* domain during the 100 ns MD simulation with snapshots were taken every 4 ps. Changes in ϕ (red) and ψ (blue) torsion angles of (A) G435 and (B) A432. (C) Changes in the χ_1 torsion angle of N434. The backbone H \cdots O distance ($d_{\text{H}\cdots\text{O}}$) for the (D) G435–A432 and (E) A432–R436 H-bonds, and the $d_{\text{H}\cdots\text{O}}$ for the (F) R436 backbone to N434 side chain H-bond.

backbone torsion angles for G435 and changes in the G435–A432 H-bond length during the MD simulations. The average and standard deviation of ϕ and ψ torsion angles of G435 in wild-type apo-WW3* were $91 \pm 24^\circ$ and $6 \pm 22^\circ$, respectively, for 83% of the simulation, whereas between 30 and 47 ns the mean and standard deviation of the ψ torsion angle was $-131 \pm 67^\circ$, and the ϕ angle fluctuated between average values of $77 \pm 30^\circ$ (major population) of $-135 \pm 36^\circ$ (minor population). The change in ϕ and ψ torsion angles of G435 during this 17 ns period correlated with the disruption of the G435–A432 H-bond (Figures 5D and S12) with the H-bond reforming when the ϕ -angle adopted a minor population value (Figure S12). This disruption and reforming of the G435–A432 H-bond also

correlated with large fluctuation of the N434 torsion angles (Figure S3). A similar pattern was observed for the A432–R436 H-bond (Figure 5E) with the backbone ϕ and ψ torsion angles of A432 adopting values of $-83 \pm 17^\circ$ and $160 \pm 10^\circ$, respectively, during the majority of the simulation (Figure 5B). During the 30 to 47 ns period of the MD simulation these two angles adopted average and standard deviation values of $-114 \pm 20^\circ$ and $134 \pm 21^\circ$, respectively. However, in comparison with the G435–A432 H-bond results, the transient reforming of this A432–R436 H-bond during the 30–47 ns period showed no correlation with changes in the ϕ/ψ torsion angles of A432, presumably because the backbone torsion angles of A432 (and R436) are more restricted by the presence of the adjacent β -sheet. The χ_1 angle of N434 adopts the *gauche*⁻ rotamer (Figure 5C), whereas during the 30–47 ns period the rotamer state is either *gauche*⁺ or *trans*, and both these rotamer states cause disruption of the R436–N434 H-bond (Figure 5F). Closer inspection of the timeline of the disruption of the three H-bonds at ~ 30 ns reveals that the R436–N434 H-bond breaks 1.3 ns before the almost simultaneous rupture (the A432–R436 H-bond breaks ~ 350 ps before the G435–A432 H-bond) of the G435–A432 and A432–R436 H-bonds. The breaking of the three H-bonds during the MD simulation causes the conformation of loop I to change from a stable type I β -turn to a seven residue Ω loop,⁵⁵ which encompasses residues H431–P437 (Figure S12). Besides the absence of the H-bonds, the average (\pm standard deviation) C α -to-C α distance between H431 and R436 is larger than for the closed β -turn (5.6 ± 0.3 Å) with a value of 7.3 ± 0.6 Å. No disruption to H-bonding along the β -sheet is observed and the C α -to-C α distance between H431 and P437, which belong to the β -sheet structure, is similar between both loop structures due to the termini constricting. The reforming of the three H-bonds occurs within 500 ps at ~ 47 ns in the MD simulation with the R436–N434 H-bond the last to reform.

In contrast to the MD simulation results of the wild-type WW3* domain, loop I of the P433T mutant fluctuates throughout the simulation between the closed β -turn structure and the seven residue Ω loop structure with loop I adopting a seven residue loop conformation for $\sim 75\%$ of the simulation time. Fluctuations in the ψ and ϕ torsion angles of G435 during the 100 ns MD simulation caused concerted rupture and formation of the G435–A432 H-bond (Figure 6A,D). As observed for the wild-type WW3* domain, fluctuation of the ψ and ϕ torsion angles for A432 in P433T were not as pronounced as that observed for G435 (Figure 6B) and these fluctuations did not correlate strongly with changes in the A432–R436 H-bond distance (Figure 6E). The limited change in torsion angles for A432 is probably because of the adjacent β -sheet restricting backbone mobility. The χ_1 angle of N434 fluctuates continuously among the three rotamer states during the simulation with the R436–N434 H-bond calculated to be present for only $\sim 4\%$ of the simulation period (Figure 6C,F). Moreover, for both apo-state wild-type WW3* and the P433T mutant stable formation of the R436–N434 H-bond only occurred during the MD simulations when the other two H-bonds were not fluctuating between closed and open states (Figures 5 and 6), indicating that stable formation of the R436–N434 H-bond is reliant on a stable β -turn conformation.

Unlike the apo-state, the MD simulations of the peptide-bound wild-type and P433T mutant WW3* domains showed no large backbone torsion angle fluctuations for residues in loop I (Figures S4 and S6), indicating that binding of the

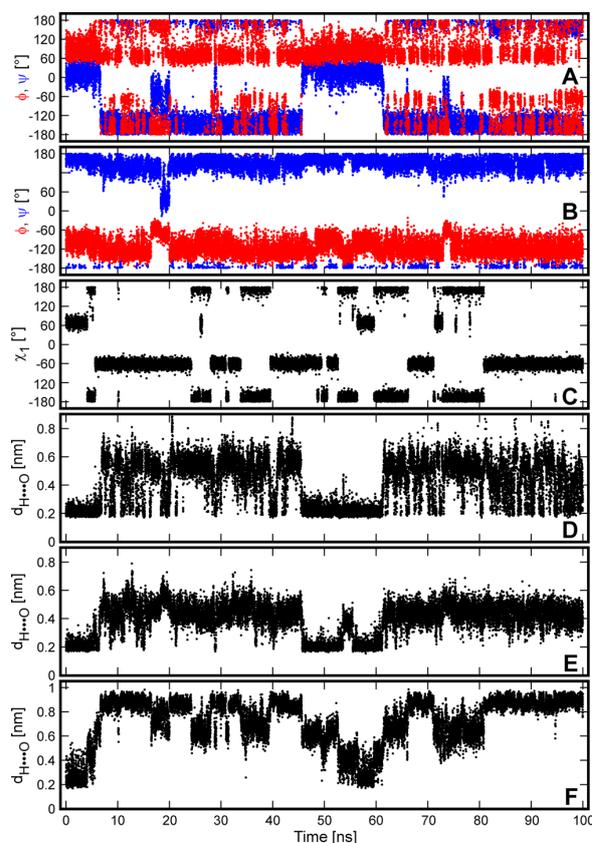


Figure 6. Torsion angles and H-bond distances of the apo-state P433T WW3* mutant during the 100 ns MD simulation with snapshots taken every 4 ps. Changes in ϕ (red) and ψ (blue) torsion angles of (A) G435 and (B) A432. (C) Changes in the χ_1 torsion angle of N434. The backbone H \cdots O distance ($d_{\text{H}\cdots\text{O}}$) for the (D) G435–A432 and (E) A432–R436 H-bonds, (F) and the $d_{\text{H}\cdots\text{O}}$ for the R436 backbone to N434 side chain H-bond.

peptide restricts loop I to the type I β -turn conformation. Consequently, H-bonds G435–A432 and A432–R436 remained essentially intact throughout the 100 ns simulation for both proteins, as indicated by the H-bond lengths remaining at ~ 0.2 nm (Figure S13A,B). In contrast, the R436–N434 H-bond was not stable throughout the simulation, and fluctuations between a formed and ruptured H-bond correlated with changes in the χ_1 angle for N434 (Figure S13C). Nonetheless, breaking of this H-bond did not lead to changes in the overall loop I structure (Figure S12D,E). This H-bond is also populated slightly more in the P433T (36%) than in the wild-type domain (30%). Surprisingly, for the wild-type WW3* domain in the apo-state (Figure 5F) this H-bond was present for 41% of the MD simulation, which initially seemed counterintuitive. However, analysis of the structures taken during the simulation reveals that W449 adopts a fixed state in the presence of the peptide because of the proline from the PY-motif peptide and the cation- π interaction with the side chain of R436, as reported in structural studies.^{13,14} This important cation- π interaction reduces the flexibility of the R436 side chain and leads to subtle changes in the backbone conformation that positions the amide group less favorably for formation of a stable H-bond with the side chain of N434.

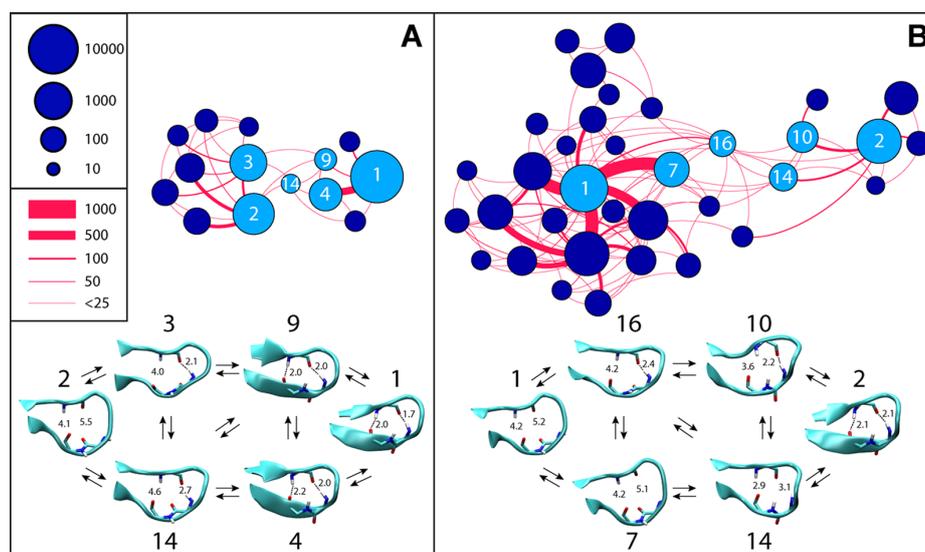


Figure 7. Cluster analysis of loop I of the hNedd4-1 WW3* domain wild-type (A) and P433T (B) in the apo-state. Clusters that are populated with more than 0.1% are illustrated by blue circles using a logarithmic scale. Transitions between the clusters are shown as lines and their width correlates to the frequency of the transition. Cluster sizes are ranked and indicated by numbers with cluster 1 being the largest cluster. Light blue circles indicate structures involved in the transition pathways between a type I β -turn and a seven residue Ω loop. Representative structures of these clusters are shown below the cluster network. Distances between the amide proton of G435 and the carbonyl oxygen of A432 and between the amide proton of A432 and the carbonyl oxygen of R436 are shown. H-bonds are indicated by a dotted line.

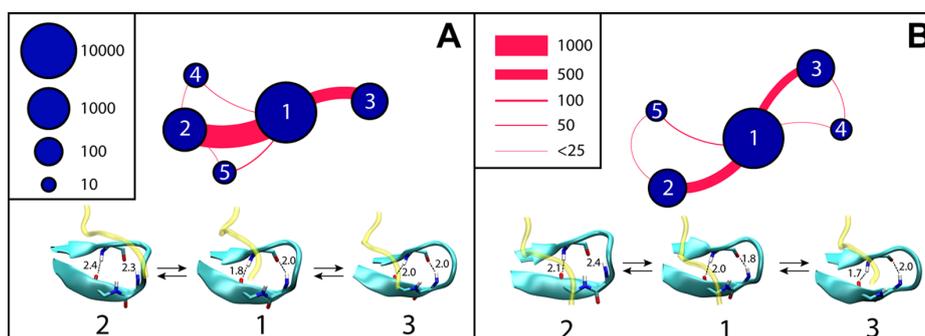


Figure 8. Cluster analysis of loop I of the hNedd4-1 WW3* domain wild type (A) and P433T (B) in the peptide-bound state. Clusters that are populated with more than 0.1% are illustrated by blue circles using a logarithmic scale. Transitions between the clusters are shown as lines and their width correlates to the frequency of the transition. Cluster sizes are ranked and indicated by numbers, with cluster 1 being the largest cluster. The representative structures of the three largest clusters comprising more than 99.5% of all sampled loop I structures are shown below the cluster network. H-bond distances between the donor proton of A432 and the carbonyl oxygen of R436 and between the amide proton of G435 and the carbonyl oxygen of A432 are shown. H-bonds are indicated by dotted lines. Residues 639–642 of the α -ENaC peptide are shown as a yellow, transparent cartoon representation.

We hypothesize that this conformational change to facilitate a tighter cation- π interaction in the complex state destabilizes the R436–N434 H-bond, thereby leading to a shift in the population distribution of the N434 χ_1 rotamer states to gauche⁺ and trans rotamers, which prevent formation of the R436–N434 H-bond because the side chain N434 carboxamide group faces away from the donor R436 amide.

Cluster Analysis Shows That the P433T Mutant Transitions Between More Conformational States Than the Wild-Type WW3* Domain. Cluster analysis of residues H431–P437 using the GROMOS algorithm with a backbone cutoff RMSD of 0.5 Å was performed to characterize the transition mechanism of loop I from a type I β -turn to a seven residue Ω loop. The R436–N434 H-bond was not examined in

this analysis because this H-bond was found to form only when loop I adopts the closed type I β -turn with the other two H-bonds present. In the wild-type apo-WW3* simulation, loop I fluctuates between the type I β -turn and a seven residue Ω loop with cluster 1 representing type I β -turn conformers and the most populated cluster observed during the simulation (80.2%; Table S3), whereas cluster 2 (9.7%) represents the majority of structures with a seven residue open loop I conformation. Figure 7A shows that these two clusters are not directly connected, but transition through intermediate states with multiple transition pathways observed from the cluster analysis. The representative structures of the clusters below the network figure show that clusters 1, 4, and 9 possess the H-bonds between G435–A432 and A432–R436 that give the type I β -

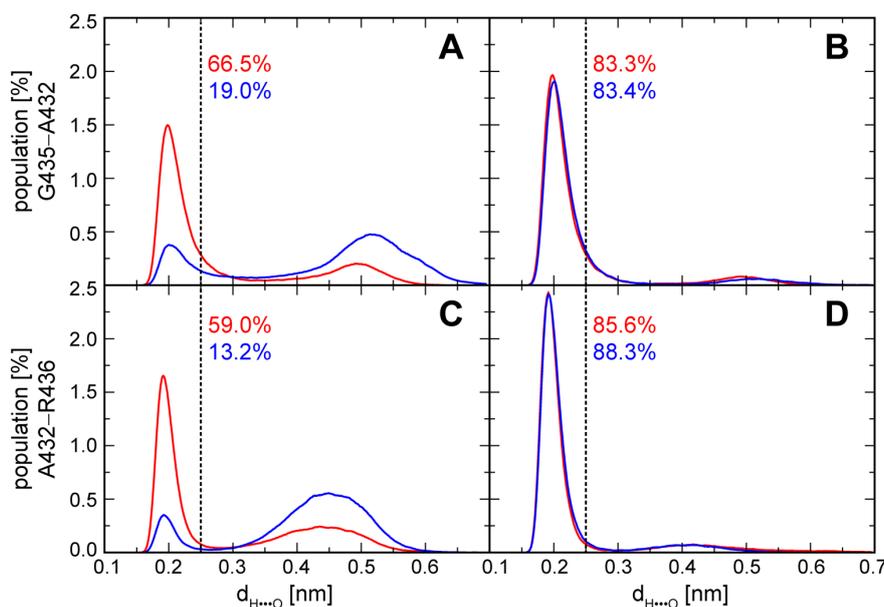


Figure 9. Distance distributions between the donor and acceptor atoms ($d_{\text{H}\dots\text{O}}$) for the two backbone H-bonds of loop I obtained from multiple 50 ns simulations of the hNedd4-1 wild-type and P433T WW3* domain (A,C) and in the corresponding complex with the α -hENaC peptide (B,D). In each panel, the red and the blue lines represent the $d_{\text{H}\dots\text{O}}$ distribution for the wild-type and P433T WW3* domain, respectively. The vertical black dotted line represents a $d_{\text{H}\dots\text{O}}$ distance of 0.25 nm. The H-bond populations (defined as being a $d_{\text{H}\dots\text{O}} < 0.25$ nm) are given in each panel.

turn. The H-bond between G435–A432 is only present in the representative structures of clusters 3 and 14. Thus, the transition pathway from the closed β -turn substructure to an open loop conformation involves the stepwise rupturing of the A432–R436 and G435–A432 H-bonds (Figure 5).

Cluster analysis of the apo-state P433T mutant shows significantly more transitions and a more complex transition network (Figure 7B). Clearly, the open state conformation (cluster 1) transitions between a much larger number of intermediates than the wild-type WW3* domain with many of the conformers not involved in the on-pathways between the closed type I β -turn (cluster 2, Figure 7B) and the largest cluster, the seven residue Ω loop (cluster 1). This analysis shows that the P433T mutation has reduced the energy barrier between structural states of the loop that enable this substructure at 25 °C to sample a much larger conformational space during the simulation, which is more kinetically restricted for the wild-type WW3* domain. Approximately 28% of all transition intermediates are intercluster transitions, whereas this is only 7% in the wild-type simulation of the apo-state. Cluster 1 comprises 30.3% of all sampled conformations for the P433T simulation, whereas cluster 2 is sampled in 17.4% of the simulation period. Clusters that are sampled during the transition from the type I β -turn to the seven residue Ω loop have only the H-bond between G435–A432 (clusters 10 and 16), showing that during particular transitions from a type I β -turn to a seven residue loop the H-bond between A432–R436 breaks before the one formed by G435–A432. Interestingly, in the $1 \rightarrow 7 \rightarrow 14 \rightarrow 2$ pathway the two intermediates 7 and 14 have neither H-bond formed but the structural transition from 7 to 14 juxtapositions the donor and acceptor groups to facilitate both H-bonds forming during the transition $14 \rightarrow 2$.

In both simulations of the complex state, only minor conformational changes to loop I are observed. The largest cluster in the wild-type WW3*– α -hENaC peptide complex

MD simulation is populated with over 92.8% of all sampled conformations and this value is 95.5% in the corresponding P433T WW3* mutant simulation (Figure 8; Table S4). The other clusters sampled during the complex simulations still exhibit both G435–A432 and A432–R436 H-bonds, showing that during the simulations loop I maintains a type I β -turn structure.

Absence of Proline in Loop I of WW3* Results in the Open State Conformation in Multiple MD Simulations.

Using the cluster analysis as a pool of different initial loop I conformations, we have conducted multiple short MD simulations to provide more reliable sampling of the conformational states and thus a more accurate estimate of the population of the type I β -turn and seven-residue Ω loop conformation. Twenty 50 ns MD simulations of both wild-type WW3* and the P433T mutant in the apo-state were performed (i.e., total of 1 μ s each). Starting structures were chosen based on the cluster populations (Table S5) and structures randomly distributed within a given cluster were selected (Table S6). The same approach was applied for the two constructs in the α -hENaC peptide bound state, except 10 MD simulations were run for both complexes (Tables S5 and S7). For the apo-state, these multiple simulations showed that the H-bond distance ($d_{\text{H}\dots\text{O}}$) defined as < 0.25 nm between G435–A432 and A432–R436 is sampled more often for the wild-type WW3* domain (66.5% and 59.0%, respectively) than for the P433T WW3* mutant (19.0% and 13.2%, respectively) (Figure 9A,C). MD simulations of the complex state showed that the G435–A432 and A432–R436 H-bonds in both wild-type and P433T WW3* domains were populated for more than 83% of the sampling time (Figure 9B,D). Interestingly, the MD simulations of the apo-state P433T mutant starting from a type I β -turn conformation (cluster 2, Table S6) showed that this protein spent a longer period in the open state (71%; Table S6) when compared with the corresponding results for the wild-type

WW3* (cluster 1; 31%). The results of the 60 MD simulations support the observations made from the 100 ns trajectories that the P433T mutant adopts a predominantly open seven residue Ω loop conformation for loop I, whereas complex formation with the α -hENaC peptide stabilizes the type I β -turn conformation for the wild-type and P433T WW3* domains.

CONCLUSIONS

In this report, MD simulations revealed that a single amino acid exchange, P433T, in loop I of the Nedd4-1 WW3* domain reduced the stability of the type I β -turn substructure, leading to this loop adopting a seven residue Ω loop during the majority of the 100 ns simulation. In both constructs, loop I is locked in the peptide binding-competent type I β -turn state in the presence of the α -hENaC peptide. These observations were supported through enhanced sampling (total of 3 μ s) conducted using multiple 50 ns MD simulations and different starting structures of each protein. The reduced sampling of conformational space of loop I in the peptide bound state arises because residues A432, P433, and N434 interact with the peptide.^{4,14} Although no structural data of the P433T mutant is available, it is conceivable that this amino acid in the P433T WW3* mutant also provides contacts with the peptide; although in the rat Nedd4-1 WW4- α ENaC peptide complex the equivalent Thr residue at this position in the loop does not interact with the peptide.¹¹ Consequently, this mutation may reduce the affinity of the P433T WW3* domain toward the α -hENaC peptide.

WW domains have been used extensively in thermodynamic and kinetic fold-unfold analysis,^{15,16,18-26,28,56-59} but unlike the Nedd4-1 WW3* domain the other WW domains studied do not contain a proline at the ($i + 1$) position. A number of these studies show that mutagenesis of the loop I substructure by either changing the length of the loop or exchanging amino acids modulates the thermodynamic stability of the domain. Thermal melting analysis of Nedd4-1 WW3* wild-type¹³ and P433T mutant showed that the mutation does reduce the thermal stability of the domain by 6 °C. The MD simulations illustrate that the P433T mutation results in the loop I adopting an open seven residue Ω loop conformation, lacking the three H-bonds present in Nedd4-1 WW3* wild-type, as the favored conformation (75% of the simulation; Figure 6) with minimal structural changes in other regions of the domain (Figure S1). The aforementioned 6 °C decrease in the thermal midpoint (T_m) and higher loop I flexibility (Figure 3D,E) for the WW3* domain due to the mutation could be attributed to loss of the three H-bonds from this loop (i.e., lower enthalpy of unfolding). Therefore, the favored seven residue Ω loop conformation, observed in the MD simulations, may represent the most populated state for loop I of the WW3* P433T mutant in solution. Currently, it is not possible to define accurately the enthalpy and entropy contributions that lead to this reduced T_m . Nonetheless, assuming negligible changes in entropy (although the change in loop size and structure will also reduce the entropy of unfolding) and taking the ΔH_{fold} of -103 kJ mol^{-1} reported for the wild-type¹³ this difference in T_m would correspond to an enthalpy change of at least 2 kJ mol^{-1} . Future efforts that combine experimental data with additional MD simulations of WW3* at different temperatures should provide an explanation for the reduced T_m .

The corresponding amino acid at the ($i + 1$) position for the other three WW domains of hNedd4-1 is not proline, suggesting that the loop I in these WW domains may not

adopt as stable a type I β -turn and could influence peptide affinity. Only the position-favored glycine at the ($i + 3$) position is sequence conserved in the four WW domains. Whether the absence of the proline in the other three Nedd4-1 WW domains influences the stability of this substructure requires further investigation; however, studies have shown that the hNedd4-1 WW4 domain displays higher thermal stability than WW3*,^{53,60} indicating that the absence of a proline in loop I of WW4 does not reduce the overall fold stability as observed for the P433T mutant. Thus, additional factors contribute to the thermal stability of the hNedd4-1 WW domains to match the stabilizing effect the proline provides to loop I.

In summary, the results provide evidence that P433 in the WW3* domain is a key residue for domain stability and mutation causes loop I to sample a much larger conformational space, though still increasing the entropic cost from restriction of conformational freedom by the threonine mutation. An understanding of the reduced thermal stability and the effect this has on the rate of folding remains unknown, as does the influence of this mutation on the affinity of the domain toward the α -hENaC peptide. By characterizing these aspects of the P433T mutant, we will further our understanding of WW domain stability and roles in biological processes.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcc.7b11637.

Tables presenting the parameters used in the MD simulations, standard deviations of the ϕ and ψ torsion angles adopted for wild-type and P433T WW3* during the MD simulation, cluster analysis of apo and complex states, and details of the individual 50 ns simulations of hNedd4-1 wild-type and P433T WW3*. Figures showing an overlay of 2D ^1H - ^{15}N HSQC spectra of hNedd4-1 wild-type and P433T WW3* recorded at 5 °C, changes in backbone conformation of the complex state, changes in ϕ and ψ torsion angles of residues R430-P437 during the MD simulations, Ramachandran plots of residues in loop I, representative structures taken from MD simulations showing various conformational states of loop I, and loop I H-bond distances and χ_1 angles of N434 during the MD simulations of the wild-type WW3* and P433T mutant in the peptide bound state (PDF)

AUTHOR INFORMATION

Corresponding Author

*Phone: ++49 2461 61 9487. Fax: ++49 2461 61 2023. E-mail: a.dingley@fz-juelich.de.

ORCID

Vineet Panwalkar: 0000-0003-1621-1606

Dieter Willbold: 0000-0002-0065-7366

Andrew J. Dingley: 0000-0002-6838-5803

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors acknowledge access to the Jülich-Düsseldorf Biomolecular NMR Center, which is jointly run by the Forschungszentrum Jülich and Heinrich-Heine-Universität

Düsseldorf. M. S. acknowledges the Ministry of Innovation, Science, and Research of the German federal state of North Rhine-Westphalia and the Heinrich-Heine-Universität Düsseldorf (scholarship within the CLIB-Graduate Cluster Industrial Biotechnology) for financial support. The authors gratefully acknowledge the computing time granted by the JARA-HPC Vergabegremium and VSR commission on the supercomputer JURECA at the Forschungszentrum Jülich. We thank Marcus Tusche for preparation of the P433T mutant construct. We thank Drs. Dušan Petrović, Martin Carballo Pacheco (Forschungszentrum Jülich), and Michael Schmitz (The University of Auckland) for useful discussions and Xue Wang (Forschungszentrum Jülich) for computational support.

REFERENCES

- (1) Ingham, R. J.; Gish, G.; Pawson, T. The Nedd4 Family of E3 Ubiquitin Ligases: Functional Diversity within a Common Modular Architecture. *Oncogene* **2004**, *23*, 1972–1984.
- (2) Verdecia, M. A.; Bowman, M. E.; Lu, K. P.; Hunter, T.; Noel, J. P. Structural Basis for Phosphoserine-Proline Recognition by Group IV WW Domains. *Nat. Struct. Biol.* **2000**, *7*, 639–643.
- (3) Macias, M. J.; Hyvönen, M.; Baraldi, E.; Schultz, J.; Sudol, M.; Saraste, M.; Oschkinat, H. Structure of the WW Domain of a Kinase-Associated Protein Complexed with a Proline-Rich Peptide. *Nature* **1996**, *382*, 646–649.
- (4) Bobby, R.; Medini, K.; Neudecker, P.; Lee, T. V.; Brimble, M. A.; McDonald, F. J.; Lott, J. S.; Dingley, A. J. Structure and Dynamics of Human Nedd4–1 WW3 in Complex with the α ENaC PY Motif. *Biochim. Biophys. Acta, Proteins Proteomics* **2013**, *1834*, 1632–1641.
- (5) Huang, X.; Poy, F.; Zhang, R.; Joachimiak, A.; Sudol, M.; Eck, M. J. Structure of a WW Domain Containing Fragment of Dystrophin in Complex with Beta-Dystroglycan. *Nat. Struct. Biol.* **2000**, *7*, 634–638.
- (6) Kanelis, V.; Rotin, D.; Forman-Kay, J. D. Solution Structure of a Nedd4 WW Domain-ENaC Peptide Complex. *Nat. Struct. Biol.* **2001**, *8*, 407–412.
- (7) Macias, M. J.; Gervais, V.; Civera, C.; Oschkinat, H. Structural Analysis of WW Domains and Design of a WW Prototype. *Nat. Struct. Biol.* **2000**, *7*, 375–379.
- (8) Wiesner, S.; Stier, G.; Sattler, M.; Macias, M. J. Solution Structure and Ligand Recognition of the WW Domain Pair of the Yeast Splicing Factor Prp40. *J. Mol. Biol.* **2002**, *324*, 807–822.
- (9) Rotin, D.; Staub, O.; Haguenaue-Tsapis, R. Ubiquitination and Endocytosis of Plasma Membrane Proteins: Role of Nedd4/Rsp5p Family of Ubiquitin-Protein Ligases. *J. Membr. Biol.* **2000**, *176*, 1–17.
- (10) Lott, J. S.; Coddington-Lawson, S. J.; Teesdale-Spittle, P. H.; McDonald, F. J. A Single WW Domain Is the Predominant Mediator of the Interaction between the Human Ubiquitin-Protein Ligase Nedd4 and the Human Epithelial Sodium Channel. *Biochem. J.* **2002**, *361*, 481–488.
- (11) Henry, P. C.; Kanelis, V.; O'Brien, M. C.; Kim, B.; Gautschi, I.; Forman-Kay, J.; Schild, L.; Rotin, D. Affinity and Specificity of Interactions between Nedd4 Isoforms and the Epithelial Na⁺ Channel. *J. Biol. Chem.* **2003**, *278*, 20019–20028.
- (12) Kanelis, V.; Bruce, M. C.; Skrynnikov, N. R.; Rotin, D.; Forman-Kay, J. D. Structural Determinants for High-Affinity Binding in a Nedd4 WW3* Domain-Comm PY Motif Complex. *Structure* **2006**, *14*, 543–553.
- (13) Panwalkar, V.; Neudecker, P.; Schmitz, M.; Lecher, J.; Schulte, M.; Medini, K.; Stoldt, M.; Brimble, M. A.; Willbold, D.; Dingley, A. J. The Nedd4–1 WW Domain Recognizes the PY Motif Peptide through Coupled Folding and Binding Equilibria. *Biochemistry* **2016**, *55*, 659–674.
- (14) Qi, S.; O'Hayre, M.; Gutkind, J. S.; Hurley, J. H. Structural and Biochemical Basis for Ubiquitin Ligase Recruitment by Arrestin-Related Domain-Containing Protein-3 (ARRDC3). *J. Biol. Chem.* **2014**, *289*, 4743–4752.
- (15) Dave, K.; Jager, M.; Nguyen, H.; Kelly, J. W.; Gruebele, M. High-Resolution Mapping of the Folding Transition State of a WW Domain. *J. Mol. Biol.* **2016**, *428*, 1617–1636.
- (16) Davis, C. M.; Dyer, R. B. Dynamics of an Ultrafast Folding Subdomain in the Context of a Larger Protein Fold. *J. Am. Chem. Soc.* **2013**, *135*, 19260–19267.
- (17) Deechongkit, S.; Nguyen, H.; Powers, E. T.; Dawson, P. E.; Gruebele, M.; Kelly, J. W. Context-Dependent Contributions of Backbone Hydrogen Bonding to Beta-Sheet Folding Energetics. *Nature* **2004**, *430*, 101–105.
- (18) Ferguson, N.; Berriman, J.; Petrovich, M.; Sharpe, T. D.; Finch, J. T.; Fersht, A. R. Rapid Amyloid Fiber Formation from the Fast-Folding WW Domain FBP28. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 9814–9819.
- (19) Ferguson, N.; Johnson, C. M.; Macias, M.; Oschkinat, H.; Fersht, A. Ultrafast Folding of WW Domains without Structured Aromatic Clusters in the Denatured State. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 13002–13007.
- (20) Jäger, M.; Nguyen, H.; Crane, J. C.; Kelly, J. W.; Gruebele, M. The Folding Mechanism of a Beta-Sheet: The WW Domain. *J. Mol. Biol.* **2001**, *311*, 373–393.
- (21) Jäger, M.; Zhang, Y.; Bieschke, J.; Nguyen, H.; Dendle, M.; Bowman, M. E.; Noel, J. P.; Gruebele, M.; Kelly, J. W. Structure-Function-Folding Relationship in a WW Domain. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 10648–10653.
- (22) Koepf, E. K.; Petrassi, H. M.; Sudol, M.; Kelly, J. W. WW: An Isolated Three-Stranded Antiparallel Beta-Sheet Domain That Unfolds and Refolds Reversibly; Evidence for a Structured Hydrophobic Cluster in Urea and GdnHCl and a Disordered Thermal Unfolded State. *Protein Sci.* **1999**, *8*, 841–853.
- (23) Petrovich, M.; Jonsson, A. L.; Ferguson, N.; Daggett, V.; Fersht, A. R. Phi-Analysis at the Experimental Limits: Mechanism of Beta-Hairpin Formation. *J. Mol. Biol.* **2006**, *360*, 865–881.
- (24) Sharpe, T.; Jonsson, A. L.; Rutherford, T. J.; Daggett, V.; Fersht, A. R. The Role of the Turn in Beta-Hairpin Formation During WW Domain Folding. *Protein Sci.* **2007**, *16*, 2233–2239.
- (25) Nguyen, H.; Jager, M.; Moretto, A.; Gruebele, M.; Kelly, J. W. Tuning the Free-Energy Landscape of a WW Domain by Temperature, Mutation, and Truncation. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 3948–3953.
- (26) Crane, J. C.; Koepf, E. K.; Kelly, J. W.; Gruebele, M. Mapping the Transition State of the WW Domain Beta-Sheet. *J. Mol. Biol.* **2000**, *298*, 283–292.
- (27) Piana, S.; Sarkar, K.; Lindorff-Larsen, K.; Guo, M.; Gruebele, M.; Shaw, D. E. Computational Design and Experimental Testing of the Fastest-Folding Beta-Sheet Protein. *J. Mol. Biol.* **2011**, *405*, 43–48.
- (28) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; et al. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330*, 341–346.
- (29) Peng, T.; Zintsmaster, J. S.; Namanja, A. T.; Peng, J. W. Sequence-Specific Dynamics Modulate Recognition Specificity in WW Domains. *Nat. Struct. Mol. Biol.* **2007**, *14*, 325–331.
- (30) Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948.
- (31) Guruprasad, K.; Rajkumar, S. Beta- and Gamma-Turns in Proteins Revisited: A New Set of Amino Acid Turn-Type Dependent Positional Preferences and Potentials. *J. Biosci.* **2000**, *25*, 143–156.
- (32) Hutchinson, E. G.; Thornton, J. M. A Revised Set of Potentials for Beta-Turn Formation in Proteins. *Protein Sci.* **1994**, *3*, 2207–2216.
- (33) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 1950–1958.
- (34) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; et al. Gromacs 4.5: A High-Throughput and Highly Parallel Open

Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845–854.

(35) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

(36) Dunbrack, R. L., Jr.; Cohen, F. E. Bayesian Statistical Analysis of Protein Side-Chain Rotamer Preferences. *Protein Sci.* **1997**, *6*, 1661–1681.

(37) Abascal, J. L. F.; Vega, C. A General Purpose Model for the Condensed Phases of Water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, 234505.

(38) Berendsen, H. J.; Postma, J. v.; van Gunsteren, W. F.; DiNola, A.; Haak, J. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(39) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(40) Hess, B. P-Lincs: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.

(41) Darden, T.; York, D.; Pedersen, L. G. Particle Mesh Ewald: An N log (N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(42) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(43) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. Peptide Folding: When Simulation Meets Experiment. *Angew. Chem., Int. Ed.* **1999**, *38*, 236–240.

(44) Bastian, M.; Heymann Jacomy, S. M. Gephi: An Open Source Software for Exploring and Manipulating Networks; *Proceedings of the Third International AAAI Conference on Weblogs and Social Media*, San Jose, CA, May 17–20, 2009; Association for the Advancement of Artificial Intelligence: Palo Alto, CA, 2009; pp 361–362.

(45) Shen, Y.; Bax, A. Sparta+: A Modest Improvement in Empirical NMR Chemical Shift Prediction by Means of an Artificial Neural Network. *J. Biomol. NMR* **2010**, *48*, 13–22.

(46) Dingley, A. J.; Lorenzen, L.; Grotzinger, J. NMR Analysis of Viral Protein Structures. *Methods Mol. Biol.* **2008**, *451*, 441–462.

(47) Wishart, D. S.; Bigam, C. G.; Yao, J.; Abildgaard, F.; Dyson, H. J.; Oldfield, E.; Markley, J. L.; Sykes, B. D. ¹H, ¹³C and ¹⁵N Chemical Shift Referencing in Biomolecular NMR. *J. Biomol. NMR* **1995**, *6*, 135–140.

(48) Farrow, N. A.; Muhandiram, R.; Singer, A. U.; Pascal, S. M.; Kay, C. M.; Gish, G.; Shoelson, S. E.; Pawson, T.; Forman-Kay, J. D.; Kay, L. E. Backbone Dynamics of a Free and Phosphopeptide-Complexed Src Homology 2 Domain Studied by ¹⁵N NMR Relaxation. *Biochemistry* **1994**, *33*, 5984–6003.

(49) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. Nmrpipe: A Multidimensional Spectral Processing System Based on Unix Pipes. *J. Biomol. NMR* **1995**, *6*, 277–293.

(50) Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline. *Proteins: Struct., Funct., Genet.* **2005**, *59*, 687–696.

(51) Lee, J.; Dubey, V. K.; Longo, L. M.; Blaber, M. A Logical or Redundancy within the Asx-Pro-Asx-Gly Type I Beta-Turn Motif. *J. Mol. Biol.* **2008**, *377*, 1251–1264.

(52) Anderson, J. M.; Jurban, B.; Huggins, K. N.; Shcherbakov, A. A.; Shu, L.; Kier, B.; Andersen, N. H. Nascent Hairpins in Proteins: Identifying Turn Loci and Quantitating Turn Contributions to Hairpin Stability. *Biochemistry* **2016**, *55*, 5537–5553.

(53) Panwalkar, V.; Neudecker, P.; Willbold, D.; Dingley, A. J. Multiple WW Domains of Nedd4–1 Undergo Conformational Exchange That Is Quenched Upon Peptide Binding. *FEBS Lett.* **2017**, *591*, 1573–1583.

(54) Robustelli, P.; Stafford, K. A.; Palmer, A. G., 3rd Interpreting Protein Structural Dynamics from NMR Chemical Shifts. *J. Am. Chem. Soc.* **2012**, *134*, 6365–6374.

(55) Leszczynski, J. F.; Rose, G. D. Loops in Globular Proteins: A Novel Category of Secondary Structure. *Science* **1986**, *234*, 849–855.

(56) Deechongkit, S.; Dawson, P. E.; Kelly, J. W. Toward Assessing the Position-Dependent Contributions of Backbone Hydrogen Bonding to Beta-Sheet Folding Thermodynamics Employing Amide-to-Ester Perturbations. *J. Am. Chem. Soc.* **2004**, *126*, 16762–16771.

(57) Karanicolas, J.; Brooks, C. L., 3rd Integrating Folding Kinetics and Protein Function: Biphasic Kinetics and Dual Binding Specificity in a WW Domain. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 3432–3437.

(58) Karanicolas, J.; Brooks, C. L., 3rd The Structural Basis for Biphasic Kinetics in the Folding of the WW Domain from a Formin-Binding Protein: Lessons for Protein Design? *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 3954–3959.

(59) Luo, Z.; Ding, J.; Zhou, Y. Temperature-Dependent Folding Pathways of Pin1 WW Domain: An All-Atom Molecular Dynamics Simulation of a Go Model. *Biophys. J.* **2007**, *93*, 2152–2161.

(60) Cobos, E. S.; Iglesias-Bexiga, M.; Ruiz-Sanz, J.; Mateo, P. L.; Luque, I.; Martinez, J. C. Thermodynamic Characterization of the Folding Equilibrium of the Human Nedd4-WW4 Domain: At the Frontiers of Cooperative Folding. *Biochemistry* **2009**, *48*, 8712–8720.

Supporting Information

Proline Restricts Loop I Conformation of the High Affinity WW Domain from Human Nedd4-1 to a Ligand Binding-Competent Type I β -Turn

*Marianne Schulte,^{‡,§} Vineet Panwalkar,^{‡,§} Stefan Freischem,^{‡,§} Dieter Willbold,^{‡,§} and Andrew
J. Dingley^{*,‡,§}*

[‡]ICS-6 (Strukturbiochemie), Forschungszentrum Jülich, 52425 Jülich, Germany

[§]Institut für Physikalische Biologie, Heinrich-Heine-Universität, 40225 Düsseldorf, Germany

Table S1. Parameters used in the MD simulations of the wild-type (WT) WW3* domain and the P433T mutant in both the apo and peptide bound states.

Property	apo WT	WT complex	apo P433T	P433T complex
# atoms	26280	28814	26244	22962
# water molecules	6387	6983	6278	5522
# Na ⁺ ions	20	20	20	16
# Cl ⁻ ions	21	21	21	17
Mean volume [nm ³]	198.6	217.8	198.3	173.7
min. protein distance to its periodic image [nm]	1.53	1.66	1.83	1.21

Table S2. Distribution (standard deviation) of the ϕ and ψ torsion angles for hNedd4-1 wild-type and P433T WW3* during the 100 ns MD simulations. Residues located in β -strands are shaded.

Residue	Wild-type				P433T mutant			
	apo		complex		apo		complex	
	σ_{ϕ}	σ_{ψ}	σ_{ϕ}	σ_{ψ}	σ_{ϕ}	σ_{ψ}	σ_{ϕ}	σ_{ψ}
418	22.8	60.1	28.5	56.5	11.7	11.5	23.0	73.4
419	24.1	48.9	28.4	82.5	12.1	12.6	22.5	55.2
420	24.3	81.7	26.9	87.6	15.2	18.2	21.9	69.3
421	91.8	133.8	95.2	118.8	44.8	22.5	80.4	54.8
422	59.2	98.5	33.1	17.4	19.9	19.6	20.2	16.1
423	24.9	20.1	27.4	13.8	27.3	16.2	24.4	13.1
424	10.3	23.4	9.3	9.5	10.4	14.4	9.7	14.7
425	10.8	12.2	10.3	10.9	16.8	13.1	18.5	40.2
426	15.5	23.7	12.8	17.4	14.6	27.3	38.2	25.2
427	25.2	11.7	18.4	9.8	27.8	12.1	24.5	10.3
428	12.2	12.5	10.6	9.9	15.6	14.1	11.6	10.1
429	16.0	9.3	12.0	7.5	16.7	9.8	12.1	7.7
430	9.7	14.3	8.6	9.6	10.8	16.4	9.2	9.3
431	20.8	14.8	16.2	11.6	27.2	15.0	14.9	11.1
432	21.8	16.2	18.5	10.8	25.2	22.4	16.8	9.8
433	9.0	12.9	8.9	13.5	23.0	17.3	15.5	16.1
434	22.1	49.6	16.7	21.0	26.3	78.8	16.5	18.7
435	54.4	62.2	23.9	24.7	94.9	107.7	22.9	22.6
436	26.3	13.8	29.0	13.3	27.4	12.4	27.7	8.9
437	10.6	12.0	9.8	10.2	10.7	12.4	9.5	9.3
438	12.7	11.0	12.0	9.8	13.9	11.7	11.5	9.2
439	14.2	10.0	11.1	9.3	16.0	10.2	10.6	9.0
440	11.4	8.7	10.6	9.3	13.4	10.1	10.2	8.7
441	18.4	13.5	12.5	10.8	19.1	15.6	16.5	11.8
442	12.2	13.5	11.3	13.6	12.4	13.8	12.0	14.0
443	19.2	16.0	13.8	12.7	19.7	15.3	15.1	13.8
444	18.7	10.9	15.5	9.2	18.1	11.2	16.1	10.9
445	7.9	15.6	7.8	9.7	8.0	14.9	8.4	13.9
446	21.7	16.1	11.7	9.2	20.9	13.2	21.6	14.3
447	15.8	13.7	10.1	9.2	19.3	12.5	13.1	11.0
448	12.4	16.0	9.2	8.1	12.0	12.3	9.9	8.0
449	15.4	17.9	10.0	10.6	19.7	21.1	9.6	10.3
450	23.1	51.2	15.6	12.7	23.1	85.8	14.5	11.6
451	16.6	8.1	12.7	7.8	37.8	8.6	12.5	6.3
452	9.8	10.0	10.8	12.4	12.5	10.7	9.4	11.0
453	9.2	11.6	24.2	18.0	9.3	11.9	12.4	12.7
454	18.2	21.9	19.5	20.8	18.9	19.3	17.8	27.5
455	24.5	15.9	27.7	29.6	23.5	16.7	29.3	70.3
456	22.7	17.5	22.2	10.7	22.1	13.3	25.2	15.1
457	-	-	-	-	-	-	-	-
458	-	-	-	-	-	-	-	-

Table S3. All clusters of loop I that were obtained using cluster analysis of the 100 ns simulation of apo-state hNedd4-1 wild-type and P433T WW3*.

Cluster ID	WT total # ^a	WT [%]	P433T total #	P433T [%]	Cluster ID	P433T total #	P433T [%]
1	20044 ^b	80.18	7583	30.33	44	10	0.04
2	2434	9.74	4357	17.43	45	10	0.04
3	1040	4.16	4188	16.75	46	9	0.04
4	475	1.90	1669	6.68	47	8	0.03
5	233	0.93	1236	4.94	48	8	0.03
6	152	0.61	751	3.00	49	7	0.03
7	141	0.56	738	2.95	50	7	0.03
8	77	0.31	730	2.92	51	7	0.03
9	66	0.26	528	2.11	52	7	0.03
10	55	0.22	357	1.43	53	6	0.02
11	50	0.2	284	1.14	54	6	0.02
12	43	0.17	269	1.08	55	6	0.02
13	36	0.14	267	1.07	56	6	0.02
14	36	0.14	200	0.80	57	5	0.02
15	19	0.08	148	0.59	58	5	0.02
16	17	0.07	144	0.58	59	5	0.02
17	15	0.06	136	0.54	60	5	0.02
18	12	0.05	117	0.47	61	4	0.02
19	10	0.04	100	0.40	62	4	0.02
20	9	0.04	99	0.40	63	4	0.02
21	7	0.03	96	0.38	64	4	0.02
22	5	0.02	86	0.34	65	4	0.02
23	4	0.02	73	0.29	66	4	0.02
24	4	0.02	59	0.24	67	3	0.01
25	4	0.02	56	0.22	68	3	0.01
26	3	0.01	55	0.22	69	3	0.01
27	2	0.01	54	0.22	70	3	0.01
28	2	0.00	49	0.20	71	2	0.01
29	2	0.00	47	0.19	72	2	0.01
30	1	0.00	40	0.16	73	2	0.01
31	1	0.00	39	0.16	74	2	0.01
32	1	0.00	38	0.15	75	2	0.01
33	1	0.00	36	0.14	76	2	0.01
34	–	–	35	0.14	77	1	0.00
35	–	–	30	0.12	78	1	0.00
36	–	–	22	0.09	79	1	0.00
37	–	–	21	0.08	80	1	0.00
38	–	–	19	0.08	81	1	0.00
39	–	–	17	0.07	82	1	0.00
40	–	–	15	0.06	83	1	0.00
41	–	–	15	0.06	84	1	0.00
42	–	–	12	0.05	85	1	0.00
43	–	–	10	0.04	86	1	0.00

^a The total number of structures and the percentage population are given.

^b Clusters shaded in gray were sampled more than 0.1% during the whole simulation time and are shown in Figure 7.

Table S4. All clusters of loop I that were obtained using cluster analysis of the 100 ns simulation of peptide-bound hNedd4-1 wild-type and P433T WW3*.

Cluster ID	WT total #^a	WT [%]	P433T total #	P433T [%]
1	23209 ^b	92.84	23862	95.49
2	1250	5.00	541	2.16
3	403	1.61	490	1.96
4	50	0.20	32	0.13
5	41	0.16	29	0.12
6	23	0.09	19	0.08
7	9	0.04	8	0.03
8	8	0.03	7	0.03
9	4	0.02	6	0.02
10	3	0.01	5	0.02
11	1	0.00	1	0.00
12	–	–	1	0.00

^a The total number of structures and the percentage population are given.

^b Clusters shaded in gray were sampled more than 0.1% during the total simulation time and are shown in Figure 8.

Table S5. Overview of starting structures of the 50 ns simulations. The starting structures were chosen based on the results of the cluster analysis.

Cluster ID	Population [%]	# Starting structures^a
WT apo (20 simulations)		
1	80.18	17
2	9.74	2
3	4.16	1
4	1.90	0
5	0.93	0
P433T apo (20 simulations)		
1	30.33	8
2	17.43	5
3	16.75	4
4	6.68	2
5	4.94	1
WT complex (10 simulations)		
1	92.84	9
2	5.00	1
3	1.61	0
4	0.20	0
5	0.16	0
P433T complex (10 simulations)		
1	95.49	10
2	2.16	0
3	1.96	0
4	0.13	0
5	0.12	0

^aThe total number of starting structures from a given cluster.

Table S6. Details of the individual 50 ns simulations of the apo-state hNedd4-1 wild-type ($n = 20$) and P433T WW3* ($n = 20$).

Cluster ID	Snapshot [ps] ^a	$d_{H...O}$ Start ^b [nm]		$d_{H...O}$ Mean ^c [nm]		Population of the closed state [%] ^d	Transitions ^e #
		G435-A432	A432-R436	G435-A432	A432-R436		
WT apo							
1	1000	0.198	0.202	0.328 ± 0.132	0.367 ± 0.112	24.65	3
	5000	0.198	0.234	0.213 ± 0.028	0.201 ± 0.025	96.78	0
	10000	0.204	0.205	0.208 ± 0.022	0.203 ± 0.028	96.66	0
	12500	0.197	0.188	0.316 ± 0.131	0.326 ± 0.120	42.21	2
	20000	0.215	0.176	0.312 ± 0.133	0.316 ± 0.137	54.12	1
	25000	0.198	0.235	0.349 ± 0.130	0.399 ± 0.104	14.88	1
	50000	0.189	0.196	0.255 ± 0.103	0.238 ± 0.087	80.00	1
	53000	0.282	0.210	0.221 ± 0.062	0.220 ± 0.066	88.42	2
	62000	0.215	0.190	0.348 ± 0.136	0.375 ± 0.127	29.41	1
	65000	0.219	0.191	0.259 ± 0.102	0.281 ± 0.132	69.09	2
	70000	0.195	0.198	0.232 ± 0.079	0.237 ± 0.093	82.81	1
	75000	0.175	0.200	0.214 ± 0.030	0.204 ± 0.033	94.70	0
	80000	0.232	0.192	0.207 ± 0.022	0.200 ± 0.025	97.80	0
	85000	0.180	0.173	0.208 ± 0.024	0.203 ± 0.030	95.73	0
	90000	0.197	0.213	0.322 ± 0.133	0.331 ± 0.130	44.84	1
	95000	0.252	0.201	0.213 ± 0.029	0.201 ± 0.027	97.02	0
100000	0.239	0.187	0.277 ± 0.113	0.290 ± 0.120	59.36	1	
2	30060	0.364	0.477	0.390 ± 0.127	0.420 ± 0.073	4.14	2
	47000	0.530	0.536	0.343 ± 0.131	0.399 ± 0.075	1.33	0
3	34648	0.207	0.398	0.361 ± 0.128	0.425 ± 0.072	3.60	0
P433T apo							
1	6900	0.571	0.481	0.216 ± 0.038	0.216 ± 0.055	85.51	1
	16000	0.504	0.452	0.488 ± 0.093	0.441 ± 0.058	0.07	0
	25000	0.444	0.448	0.457 ± 0.119	0.439 ± 0.063	0.54	0
	30000	0.550	0.407	0.472 ± 0.107	0.449 ± 0.060	0.32	0
	45000	0.600	0.449	0.542 ± 0.082	0.441 ± 0.050	0.15	0
	61700	0.504	0.485	0.455 ± 0.121	0.436 ± 0.068	0.93	0
	70000	0.486	0.417	0.468 ± 0.112	0.439 ± 0.061	0.36	0
91000	0.452	0.489	0.423 ± 0.130	0.426 ± 0.080	4.85	1	
2	500	0.215	0.187	0.355 ± 0.167	0.317 ± 0.134	53.11	1
	6000	0.223	0.176	0.352 ± 0.154	0.342 ± 0.123	37.66	6
	28900	0.227	0.172	0.463 ± 0.120	0.454 ± 0.079	3.82	1
	50000	0.188	0.203	0.403 ± 0.144	0.385 ± 0.116	22.40	5
	60000	0.240	0.194	0.396 ± 0.151	0.379 ± 0.127	28.53	1
3	8700	0.446	0.545	0.481 ± 0.114	0.449 ± 0.064	0.84	0
	30008	0.345	0.458	0.465 ± 0.115	0.432 ± 0.071	2.27	2
	70100	0.288	0.210	0.467 ± 0.113	0.444 ± 0.062	0.40	0
	97000	0.240	0.468	0.457 ± 0.134	0.426 ± 0.095	10.16	2
4	25200	0.634	0.490	0.473 ± 0.136	0.432 ± 0.073	4.00	2
	72220	0.614	0.462	0.459 ± 0.117	0.437 ± 0.073	2.74	2
5	36000	0.520	0.549	0.488 ± 0.098	0.466 ± 0.060	0.34	0

^a Snapshot taken at the given time point in the 100 ns simulation was used as the starting structure for the 50 ns simulation.

^b Distance between H-bond donor and acceptor atoms in the starting structure.

^c Mean distance and standard deviation between H-bond donor and acceptor atoms averaged over the 50 ns simulation.

^d Population (%) of the A432–R436 H-bond when $d_{\text{H}\cdots\text{O}} < 0.25$ nm.

^e Total number of transitions during the simulation is given. A transition is defined as the change in the A432–R436 H-bond $d_{\text{H}\cdots\text{O}}$ from < 0.25 nm to > 0.25 nm and vice-versa, for longer than 0.5 ns.

Table S7. Details of the individual 50 ns simulations of the peptide-bound hNedd4-1 wild-type ($n = 10$) and P433T WW3* ($n = 10$).

Cluster ID	Snapshot [ps] ^a	$d_{H...O}$ Start ^b [nm]		$d_{H...O}$ Mean ^c [nm]		Population of the closed state [%] ^d	Transitions ^e #
		G435-A432	A432-R436	G435-A432	A432-R436		
WT complex							
1	1000	0.228	0.206	0.215 ± 0.033	0.199 ± 0.023	97.79	0
	10000	0.168	0.182	0.207 ± 0.022	0.198 ± 0.016	99.05	0
	20000	0.208	0.204	0.212 ± 0.029	0.198 ± 0.021	98.19	0
	30000	0.216	0.176	0.295 ± 0.131	0.270 ± 0.104	63.95	1
	40000	0.207	0.188	0.380 ± 0.131	0.445 ± 0.123	10.30	1
	50000	0.223	0.195	0.212 ± 0.027	0.197 ± 0.017	98.82	0
	60000	0.214	0.172	0.207 ± 0.021	0.198 ± 0.018	98.70	0
	75000	0.202	0.231	0.215 ± 0.039	0.205 ± 0.040	94.58	2
	90000	0.191	0.203	0.209 ± 0.023	0.198 ± 0.017	98.71	0
2	53420	0.213	0.207	0.210 ± 0.026	0.203 ± 0.024	95.53	0
P433T complex							
1	1000	0.228	0.209	0.210 ± 0.022	0.200 ± 0.019	98.24	0
	10000	0.224	0.217	0.219 ± 0.031	0.203 ± 0.023	95.95	0
	20000	0.199	0.200	0.216 ± 0.030	0.199 ± 0.022	97.89	0
	30000	0.218	0.185	0.210 ± 0.022	0.199 ± 0.019	98.42	0
	40000	0.210	0.213	0.327 ± 0.161	0.291 ± 0.114	56.47	2
	50000	0.210	0.195	0.218 ± 0.032	0.200 ± 0.024	97.36	0
	60000	0.202	0.209	0.209 ± 0.022	0.199 ± 0.018	98.66	0
	70000	0.208	0.220	0.210 ± 0.022	0.201 ± 0.018	98.16	0
	80000	0.254	0.192	0.213 ± 0.027	0.202 ± 0.021	96.86	0
	90000	0.262	0.184	0.358 ± 0.150	0.310 ± 0.108	44.55	4

^a Snapshot taken at the given time point in the 100 ns simulation was used as the starting structure for the 50 ns simulation.

^b Distance between H-bond donor and acceptor atoms in the starting structure.

^c Mean distance and standard deviation between H-bond donor and acceptor atoms averaged over the 50 ns simulation.

^d Population (%) of the A432–R436 H-bond when $d_{H...O} < 0.25$ nm.

^e Total number of transitions during the simulation is given. A transition is defined as the change in the A432–R436 H-bond $d_{H...O}$ from < 0.25 nm to > 0.25 nm and vice-versa, for longer than 0.5 ns.

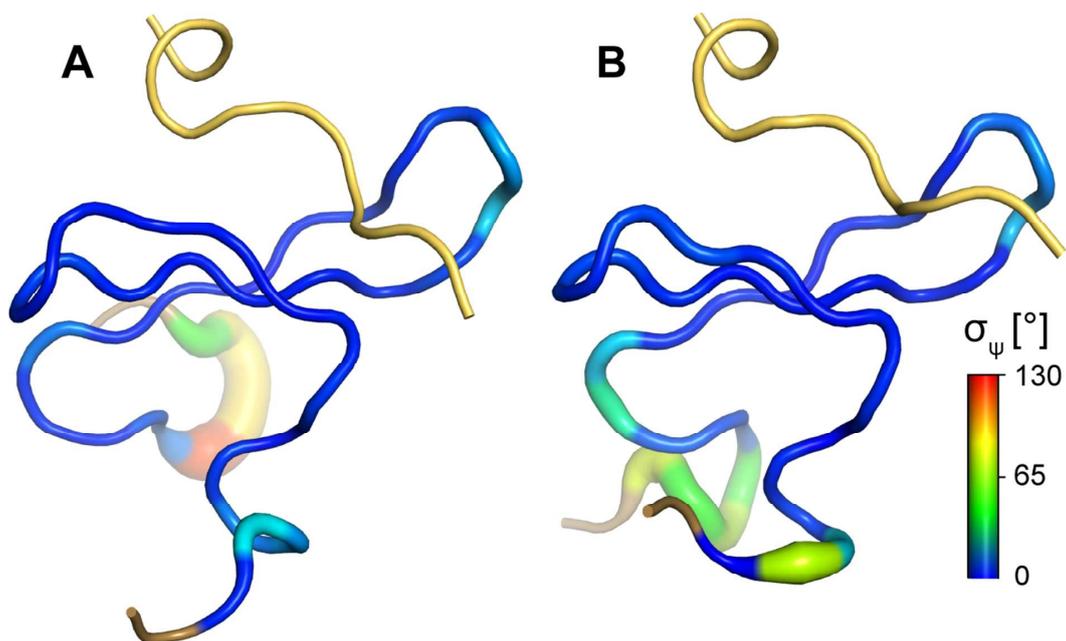


Figure S2. Changes in backbone conformation mapped to the WW domain structure in complex with the α -hENaC peptide for the hNedd4-1 wild-type WW3* domain (A) and the P433T mutant (B) during the 100 ns MD simulations. Backbone conformational distribution of the wild-type hNedd4 WW3* domain and the P433T mutant in the peptide-bound state was determined by calculating the σ_ψ for all structural snapshots taken during each simulation. Structures are shown in licorice representation and residues colored according to their σ_ψ value. The peptide is colored light brown.

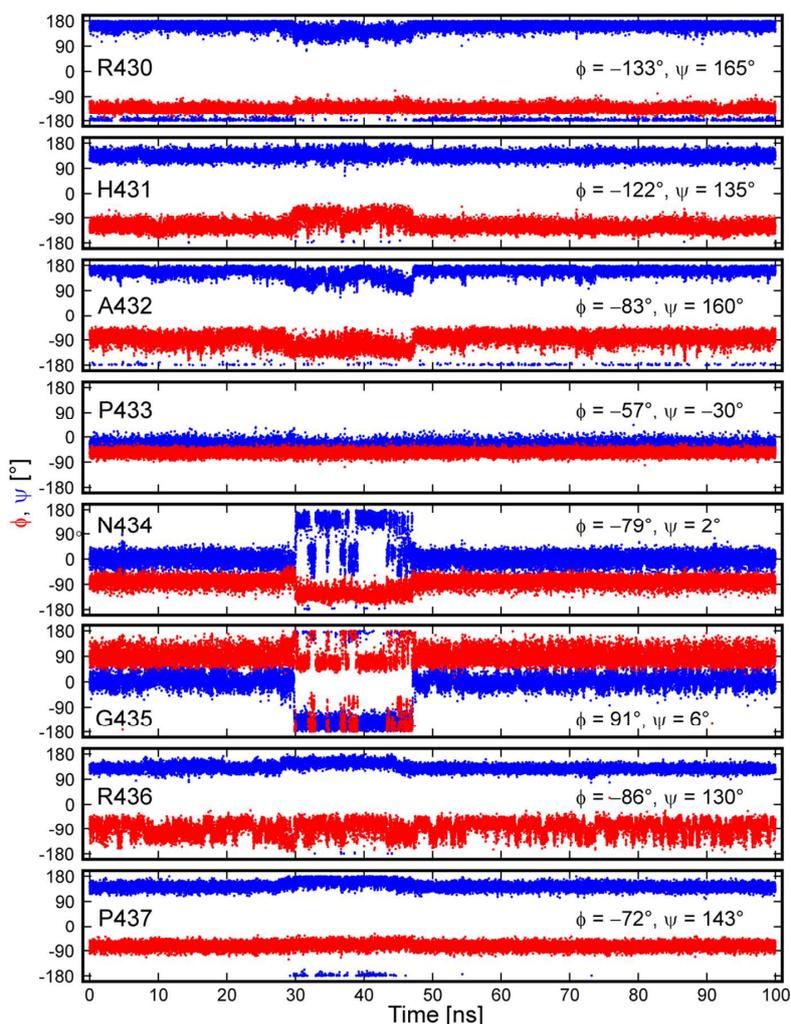


Figure S3. ϕ (red) and ψ (blue) torsion angles of R430–P437 during the 100 ns MD simulation of the wild-type WW3* domain in the apo-state. Snapshots of the structure were taken every 4 ps of the trajectory. Average ϕ and ψ torsion angles, excluding the simulation period 30–47 ns, are provided. The ϕ and ψ torsion angles for these residues taken from the crystal structure (PDB ID: 4N7F, molecule A) are -139° and 168° (R430), -126° and 135° (H431), -79° and 171° (A432), -60° and -25° (P433), -86° and 3° (N434), 89° and 2° (G435), -77° and 129° (R436), and -73° and 152° (P437).

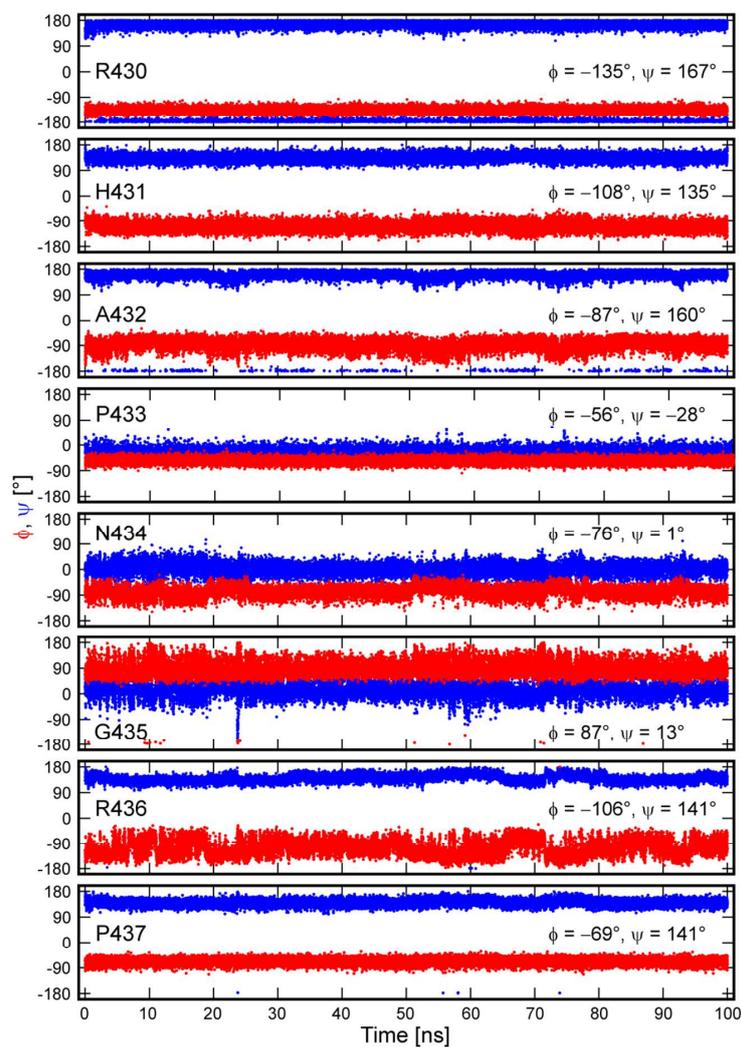


Figure S4. ϕ (red) and ψ (blue) torsion angles of R430–P437 during the 100 ns MD simulation of the wild-type WW3* domain in the complex state. Snapshots of the structure were taken every 4 ps of the trajectory. Average ϕ and ψ torsion angles are indicated. The ϕ and ψ torsion angles for these residues taken from the crystal structure (PDB ID: 4N7H) are -142° and 177° (R430), -111° and 134° (H431), -80° and 167° (A432), -51° and -38° (P433), -77° and 1° (N434), 95° and -8° (G435), -80° and 138° (R436), and -69° and 140° (P437).

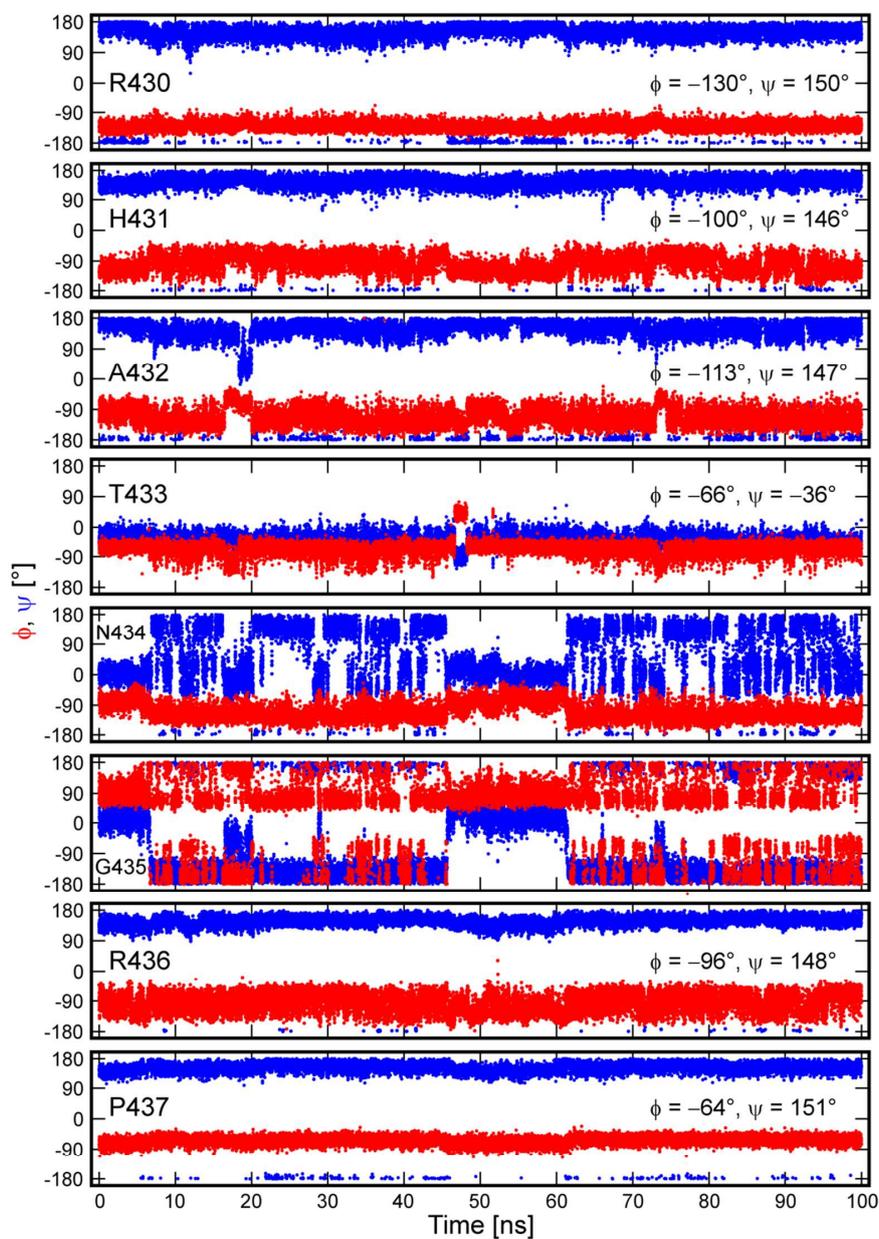


Figure S5. ϕ (red) and ψ (blue) torsion angles of R430–P437 during the 100 ns MD simulation of the P433T WW3* mutant in the apo-state. Snapshots of the structure were taken every 4 ps of the trajectory. Average ϕ and ψ torsion angles are indicated. The average ϕ and ψ torsion angles of N434 and G435 are not provided because these angles fluctuate substantially during the simulation with large standard deviations (see Table S1).

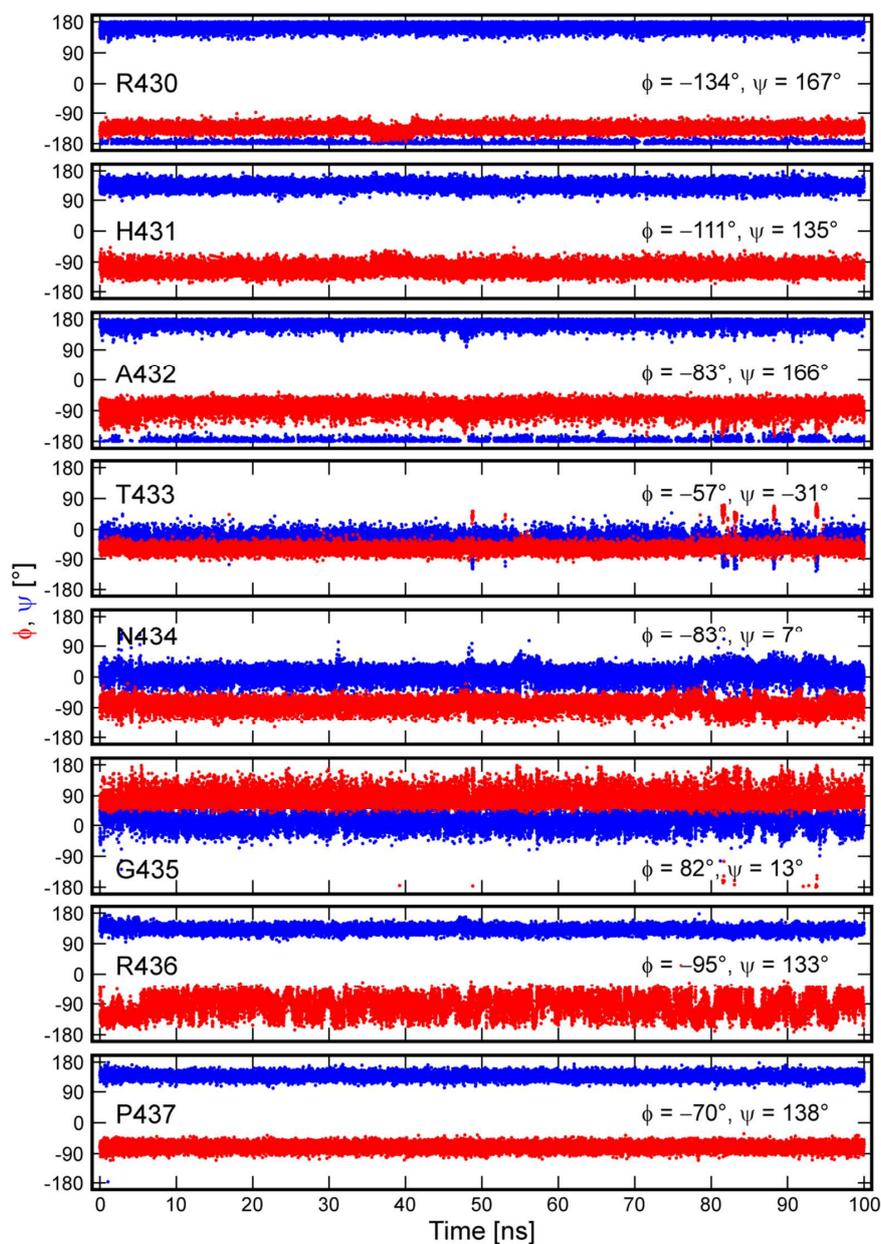


Figure S6. ϕ (red) and ψ (blue) torsion angles of R430–P437 during the 100 ns MD simulation of the P433T WW3* mutant in the complex state. Snapshots of the structure were taken every 4 ps of the trajectory. Average ϕ and ψ torsion angles are indicated.

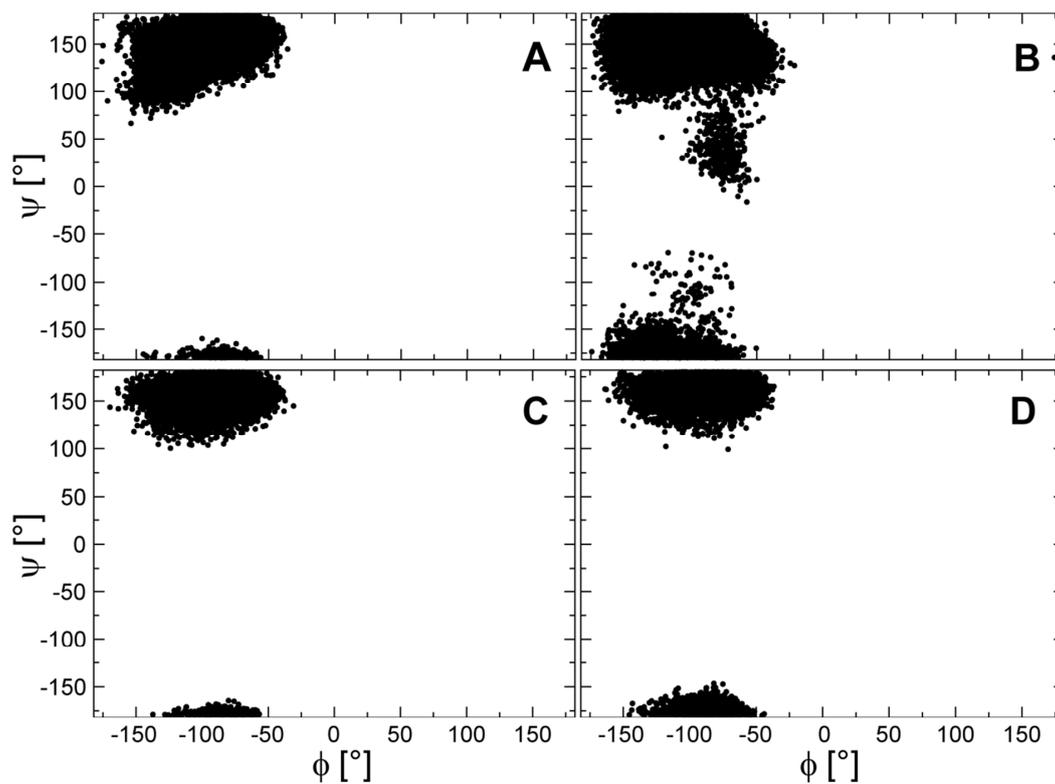


Figure S7. Ramachandran plots of A432 in wild-type WW3 (A,C) and the P433T mutant (B,D). The constructs in the apo-state (A,B) and in complex with the α -hENaC peptide (C,D).

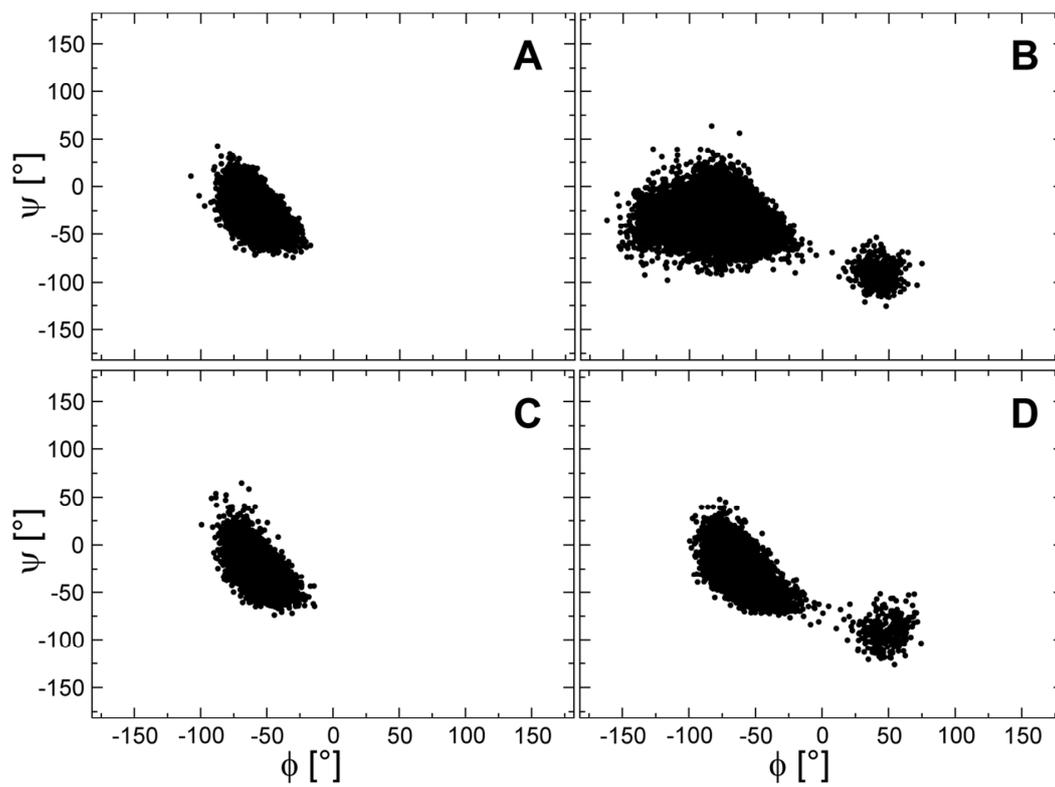


Figure S8. Ramachandran plots of P433 in wild-type WW3 (A,C) and T433 in the P433T mutant (B,D). The constructs in the apo-state (A,B) and in complex with the α -hENaC peptide (C,D).

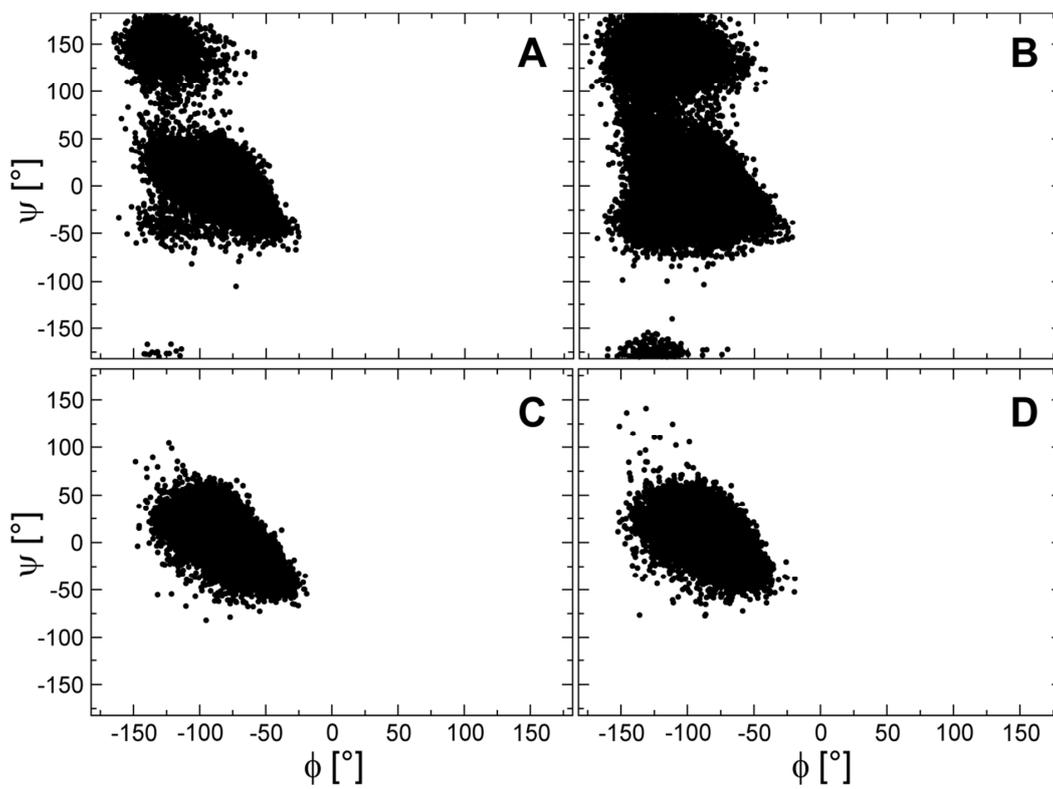


Figure S9. Ramachandran plots of N434 in wild-type WW3 (A,C) and the P433T mutant (B,D). The constructs in the apo-state (A,B) and in complex with the α -hENaC peptide (C,D).

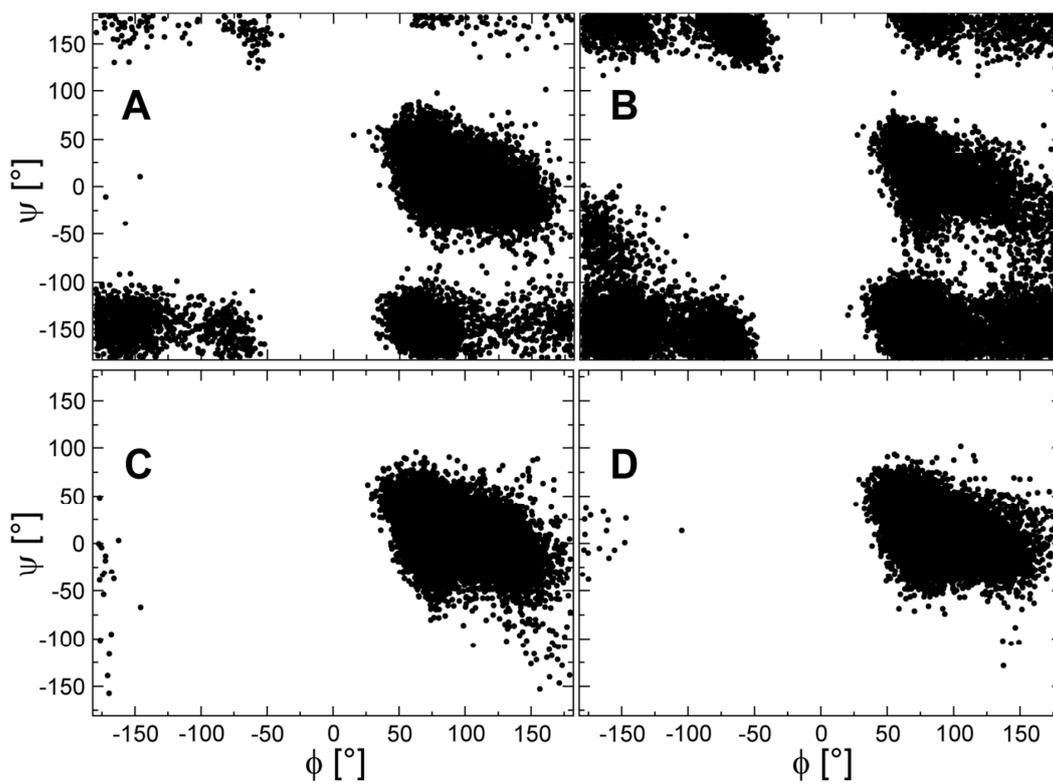


Figure S10. Ramachandran plots of G435 in wild-type WW3 (A,C) and the P433T mutant (B,D). The constructs in the apo-state (A,B) and in complex with the α -hENaC peptide (C,D).

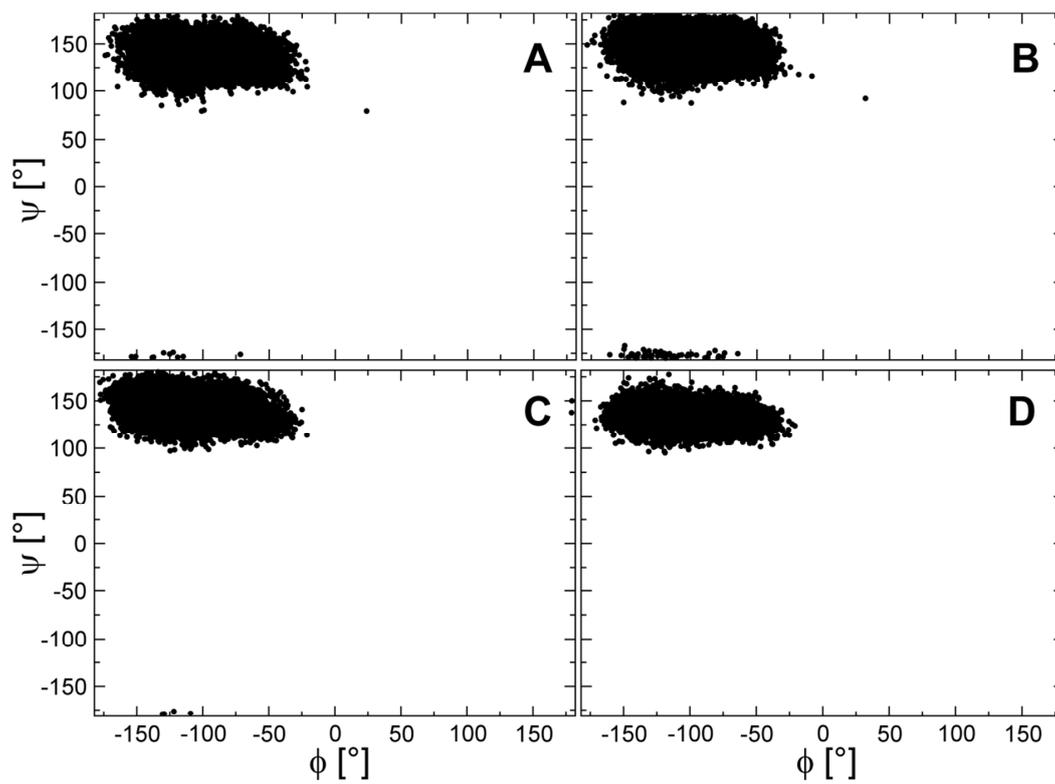


Figure S11. Ramachandran plots of R436 in wild-type WW3 (A,C) and the P433T mutant (B,D). The constructs in the apo-state (A,B) and in complex with the α -hENaC peptide (C,D).

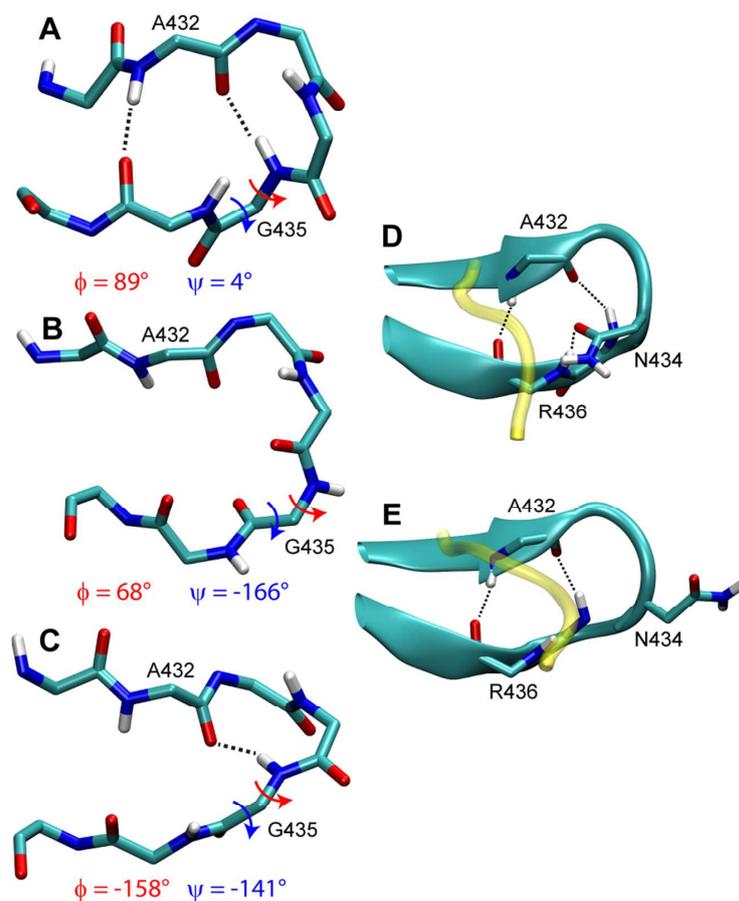


Figure S12. Representative structures taken from MD simulations of loop I of the wild-type hNedd4-1 WW3* apo-state showing the (A) closed β -turn state with both the G435–A432 and A432–R436 H-bonds present, (B) open state Ω loop (H431–P437) where the ψ and ϕ torsion angles adopt values that disrupt the H-bonds (i.e., representation of the major population during the 30–47 ns period of the simulation) and (C) the G435 ϕ torsion angle adopts a value (minor population) that enables formation of the G435–A432 H-bond. The $C\alpha$ to $C\alpha$ distances between H431 and R436 in (A), (B) and (C) are 5.6, 8.1 and 6.6 Å. (D,E) Loop I of the wild-type hNedd4-1 WW3* in complex with the peptide bound state showing that reorientation of the χ_1 angle of N434 disrupts the R436–N434 H-bond, but does not lead to disruption of the β -turn.

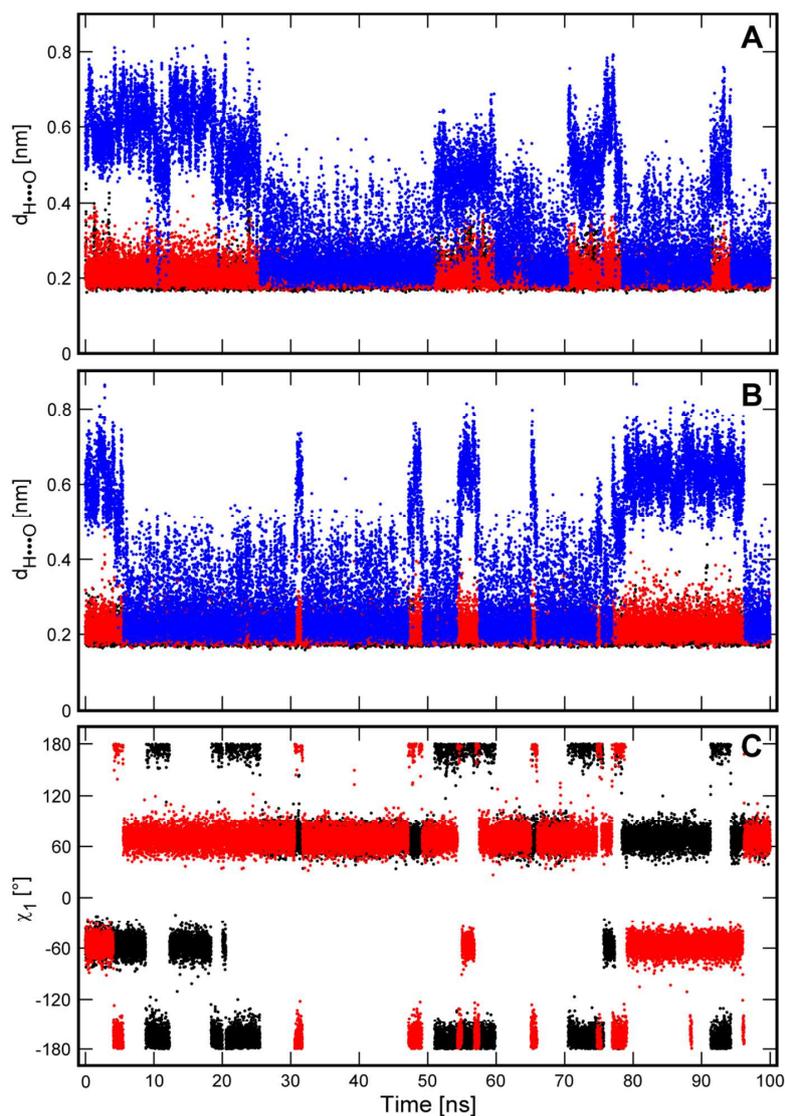


Figure S13. Loop I H-bond distances for (A) the wild-type hNedd4-1 WW3*– α -hENaC peptide complex and (B) the P433T hNedd4-1 WW3* mutant– α -hENaC peptide complex. G435–A432, A432–R436 and R436–N434 H-bond ($d_{\text{H}\cdots\text{O}}$) distances are shown in black, red and blue symbols, respectively. (C) Changes in the χ_1 torsion angle of N434 for the wild-type (black) and P433T mutant (red) in the α -hENaC peptide-bound state.

Chapter 4

Conclusions and Outlook

Protein structure and dynamics define their function. In this thesis, a combination of NMR spectroscopy and MD simulations were used, together, to highlight structural and dynamic determinants of *ec*DERA catalysis and WW domain-PY motif interaction.

The *ec*DERA catalyzes the reversible conversion of dR5P into acetaldehyde and G3P. Importance of the C-terminal Y259 in *ec*DERA catalysis of the breakdown of its natural substrate dR5P was highlighted by ~ 100 fold reduction in the catalytic turnover number upon Y259F mutation, using the monomeric variant DERAm. Hamiltonian Replica Exchange MD simulations suggested that the C-terminal tail may adopt conformations that allow Y259 to enter the active site in absence of the substrate. Presence of such a conformation in solution was confirmed using chemical shift perturbation analysis, comparing DERAm and DERAm Y259F mutant, and the observation of NOEs between Y259 and active-site residues. Additionally, transient, non-covalent interactions, stabilizing the C-terminal tail in the closed state, were identified. Based on NOE distance restraints, a solution-state structure of DERAm with the C-terminal tail in the catalytically-relevant, closed state was derived. Analysis of H/D exchange of the C2 proton of propanal, in presence of DERAm and the DERAm Y259F mutant, revealed that the exchange is significantly faster with Y259 than with F259 as C-terminal residue, highlighting the role of Y259 side chain in the proton abstraction step in the catalytic reaction. In order to probe if the exchange process between the closed state ensemble and the open states determines the catalytic rate constant, quantification of the exchange rate constant was attempted. However, the exchange process was too fast to obtain reliable quantitative data from NMR CPMG relaxation dispersion experiments, even at 5 °C. This indicates that the exchange between open and closed states is unlikely to be rate limiting. Experiments carried out in presence of the substrate (i.e. dR5P), to investigate the effect of the substrate on the C-terminal tail exchange process, were hampered by the instability of the system. For future investigation of the effect of the substrate on the motion of the C-terminal tail, use of an inactive variant

and/or an inhibitor may be a promising approach. The results presented in this work for DERAm provide a more complete picture of DERA reaction mechanism. Due to the high sequence identity of *ecDERA* with DERAs of several human pathogens, such as *Klebsiella pneumoniae* or *Salmonella typhimurium*, their structural and dynamic properties should be identical to *ecDERA*. Thus, the insights uncovered herein, may aid in the development of small molecule inhibitors targeting bacterial DERAs.

The hNedd4-1 mediates the transfer of ubiquitin from its cognate E2 ubiquitin-conjugating enzyme to the substrate. Substrate specificity and affinity is defined by the WW domains, with the third one (WW3*) showing highest peptide binding affinity to the α -hENaC peptide. Comparison of hNedd4-1 WW3* solution-state structures in the apo form and in complex with the α -hENaC peptide revealed only minor structural changes, such as side chain orientations. Analyses of the dynamic properties, however, revealed that apo-WW3* exhibits pronounced chemical exchange contributions on the ms timescale. WW3* in the apo state exists in an equilibrium between a natively folded binding-competent state and a random-coil like state. This chemical exchange is quenched in presence of the α -hENaC peptide, which locks the WW3* domain in the folded state. The largest sequence diversity amongst the four hNedd4-1 WW domains lies in loop I, indicating that different binding affinities may arise due to different loop I sequences. Unlike the other hNedd4-1 WW domains, WW3* exhibits a statistically preferred proline in the $i+1$ position of loop I. Mutation of this proline to a threonine, the corresponding residue in WW4, leads to a shift of the major loop I conformation from a type I β -turn to a Ω seven residue loop conformation, observed using MD simulations. Lower hetNOE values of loop I residues further indicate an increase in ps-ns motion. Thus, proline in the $i+1$ position results in a stable, peptide-binding competent, type I β -turn conformation. Moreover, a P to T mutation in loop I alters the overall protein stability, evident from a 6 °C reduction in melting temperature. Characterization of the effect of P433T mutation on the fold-unfold equilibrium of WW3* domain and its affinity to the α -hENaC peptide would significantly enhance the understanding of how this domain interacts with its cognate peptides. Such a knowledge of the molecular recognition processes involving WW domains may aid in gaining further insight in several human diseases that involve WW domains, e.g. Liddle's syndrome, Huntington's disease and cancer.

Appendix A

A.1 Material

A.1.1 Bacterial Strains

Bacterial strains used for DNA amplification and recombinant protein expression are listed in Table A.1. All bacterial strains are stored as 100 μL aliquots (XL10 Gold[®] as 20 μL aliquots) at -80 °C.

Table A.1: Overview of used bacterial strains

Strain	Genotype	Supplier
BL21(DE3)	$F^- ompT hsdS_B(r_B^-, m_B^-) gal dcm(DE3)$	Novagen, Darmstadt, Germany
Omnimax TM 2 T1 ^R	$F'\{proAB lacI^q lacZ\Delta M15 Tn10(Tet^R) \Delta(ccdAB)\} mcrA \Delta(mrr hsdRMS-mcrBC) \Phi 80(lacZ)\Delta M15 \Delta(lacZYA-argF)U169 endA1 recA1 supE44 thi-1 gyrA96 relA1 tonA panD$	Thermo Fisher Scientific, Karlsruhe, Germany
XL10 Gold [®]	$Tet^r \Delta(mcrA)183 \Delta(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F' proAB lacI^q Z\Delta M15 Tn10 (Tet^r)Amy Cam^r]$	Stratagene, La Jolla, USA

A.1.2 Chemicals and Biochemicals

Chemicals used for this research correspond to purity grade pro analysis. Standard chemicals were usually purchased from Merck KGaA (Darmstadt, Germany), Carl Roth GmbH (Karlsruhe, Germany), AppliChem (Darmstadt, Germany) and Sigma-Aldrich (St. Louis, USA). Table A.2 shows a selected list of chemicals and biochemicals.

Table A.2: Used chemicals and biochemicals

Chemicals and biochemicals	Supplier
β -Nicotinamide adenine dinucleotide, reduced disodium (NADH)	GERBU Biotechnik GmbH, Heidelberg, Germany
α -Glycerophosphate Dehydrogenase- Triosephosphate Isomerase from rabbit muscle	Sigma-Aldrich, St. Louis, USA
Pierce TM Unstained Protein MW Marker	Thermo Fisher Scientific, Karlsruhe, Germany
Propionaldehyde	Sigma-Aldrich, St. Louis, USA
Gd(DTPA-DMA)	Sigma-Aldrich, St. Louis, USA
D-Glyceraldehyde-3-phosphate solution	Sigma-Aldrich, St. Louis, USA
Deuterium oxide (D ₂ O) 99.90%	Sigma-Aldrich, St. Louis, USA
2-Deoxyribose-5-phosphate sodium salt	Sigma-Aldrich, St. Louis, USA
2,2,3,3-d(4)-3-(tetramethylsilyl)propionic acid sodium salt (TSP)	Thermo Fisher Scientific, Karlsruhe, Germany
[U- ¹⁵ N 99%] ammonium chloride	Cambridge Isotope Laboratories, Inc., Andover, USA
[U- ¹³ C 99%] D-glucose	Cambridge Isotope Laboratories, Inc., Andover, USA

A.1.3 Kits

Table A.3: Commercially obtained Kits

Kit	Supplier
innuPREP Plasmid Mini Kit	Analytik Jena, Jena, Germany
QIAGEN Plasmid Midi Kit	QIAGEN, Hilden, Germany
QuikChange II XL Site-directed Mutagenesis Kit	Agilent, Santa Clara, USA
NucleoSpin [®] Gel and PCR Clean-up	MACHEREY-NAGEL, Düren, Germany

A.2 Methods

A.2.1 Cloning of pET15b-kombi-P_deoC_K58EY96W

The gene encoding a monomeric variant of *E. coli* DERA, carrying K58E and Y96W mutations (DERAm), was PCR-amplified from the vector pET21a using the forward primer 5' – GGAGATGGGCCCATGACTGATC-3' and the reverse primer 5'-GTGGTGTCTCGAGTTAGTAGCTGC-3' containing *Bsp120I* and *XhoI* restriction sites, respectively. DNA digest of the vector (pET15b-Kombi-P) and the amplified gene was confirmed using agarose gel and the DNA was extracted using the NucleoSpin[®] kit (MACHEREY-NAGEL, Düren, Germany). Prior to ligation, the digested recipient vector was treated with shrimp alkaline phosphatase (SAP) (Thermo Fisher Scientific, Karlsruhe, Germany). 2 μ L ligation reaction was used for the transformation into *E. coli* OmniMAXTM 2 T1R competent cells. The resulting construct encodes DERAm with an N-terminal His₆-tag with a PreScission protease cleavage site (LEVLFG/GP). Point mutations were introduced using the QuikChange II XL site-directed mutagenesis kit (s. Table A.3) according to the manufacturer's instructions.

A.2.2 Transformation of Vector DNA into *E. coli*

Transformation refers to the process of the introduction of external DNA into chemically competent cells which is facilitated by chemical or physical methods. 100 μ L chemically competent cells (s. Table A.1), which are stored at -80 °C, were thawed on ice for 10 min. Approximately 100 ng of plasmid DNA was added and cautiously mixed with the cells. The solution was incubated on ice for 20 min. Subsequently, a 45 s heat shock at 42 °C was applied to the cells followed by incubation on ice for 2 min. 800 μ L LB medium was added and the cells were incubated for 60 min at 37 °C under moderate agitation and cultivated overnight at 37 °C on an LB agar plate containing the antibiotic ampicillin.

A.2.3 Protein Expression and Purification

For recombinant protein production of DERAm variant, the *E. coli* strain BL21(DE3), described in Table A.1, was used. After transformation, LB pre-cultures containing 100 μ g/ml ampicillin were inoculated with one colony and incubated overnight at 37 °C under moderate shaking at 120 rpm. Cells were centrifuged and resuspended in minimal medium (M9) containing isotopically enriched ¹⁵N ammonium chloride (and ¹³C glucose) as exclusive carbon and nitrogen source. The main M9 culture was inoculated to an OD₆₀₀ of ~0.2 and incubated at 37 °C under moderate shaking. After the culture had reached on optical density of 0.9 < OD₆₀₀ > 1.1, the expression was induced by adding isopropyl- β -D-1-thiogalactosid (IPTG) to a final concentration of 0.5 mM. Af-

ter 16 h of expression at 27 °C, cells were harvested by centrifugation (4000 rpm for 15 min at 10 °C). If cells were not processed directly, cell-pellets were stored at -20 °C. For protein purification, the cells pellets were thawed on ice and resuspended in lysis buffer (20 mM potassium phosphate (KP_i , pH 7.1)). Cells were lysed using a cell-disruptor (I&L Biosystems GmbH, Königswinter, Germany) in four rounds, and afterwards the lysate was centrifuged (15,000 rpm for 45 min at 4 °C) to remove the non-soluble components. The supernatant was applied twice onto an equilibrated gravity column packed with 5 ml nickel nitrilotriacetic acid (Ni-NTA) agarose. After washing the column-bound His₆-tagged protein with lysis buffer (2 x 8 ml) and washing buffer containing an additional 10 mM imidazole (2 x 4 ml) it was eluted with 250 mM imidazole (5 x 2 ml). Fractions containing DERAm were pooled and the N-terminal His₆-tag was enzymatically removed with glutathione-S-transferase (GST)-tagged PreScission protease (GE Healthcare, Freiburg, Germany), leaving two exogenous residues (GP) at the N-terminus. Protease digest was done overnight at 4 °C utilizing ~1 mg of PreScission protease and completeness was confirmed using SDS-PAGE (Figure A.2). The separation of the His₆-tag, further purification, and buffer exchange was accomplished using size-exclusion chromatography (SEC). The protein solution was applied in multiple rounds (5 ml / round) to a SEC-column (HighLoad 16/60 Superdex 75 prep-grade, GE Healthcare, Freiburg, Germany), pre-equilibrated with the NMR buffer. Fractions containing DERAm (or mutants), according to SDS-PAGE, elute around 69 ml corresponding to a protein size of 28 kDa, and were subsequently pooled and concentrated.

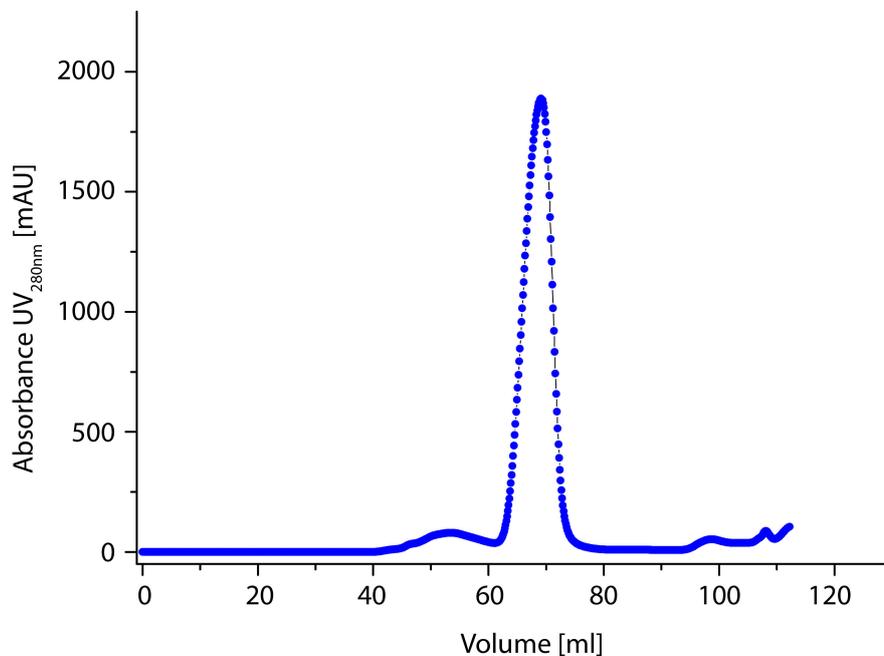


Figure A.1: Purification of DERA after digestion with PreScission protease was carried out by size-exclusion chromatography. The pure DERAm protein elutes at ~69 ml.

For protein concentration an Amicon chamber (Merck Milipore, Darmstadt, Germany) was used with a 3 kDa molecular weight cut-off filter. For long-term storage the protein samples were freeze-dried and stored at $-20\text{ }^{\circ}\text{C}$.

DERAm mutants described in this work (DERAm Y259F, DERAm S257DS258D and DERAm G252*) were expressed and purified according to the same protocol.

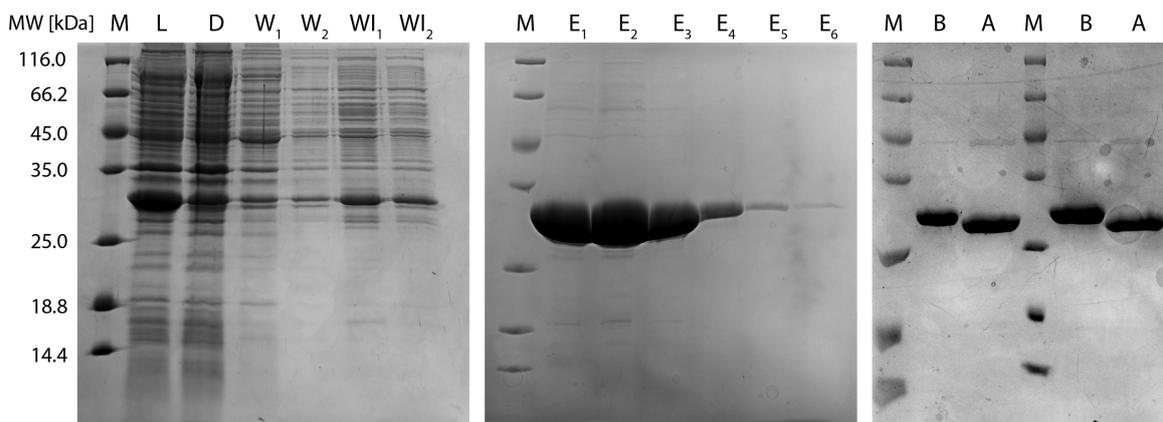


Figure A.2: Protein expression and purification of DERAm was confirmed by SDS-PAGE. As marker the PierceTM Unstained Protein MW Marker (M) was used. From left to right following samples are shown: lysate (L), flow-through (D), washing steps with lysis buffer (W_x), washing steps with 10 mM imidazole (WI_x), elutions (E_x) and PreScission digestion was confirmed by comparing the molecular weight before (B) and after digest (A) in duplicates.

A.2.4 NMR Spectroscopy

A.2.4.1 Sample Preparation

If not stated otherwise, NMR samples contained $800\text{ }\mu\text{L}$ [$\text{U-}^{13}\text{C}$, ^{15}N] or [$\text{U-}^{15}\text{N}$] protein in 50 mM HEPES (pH 6.8), with 0.03 % (w/v) NaN_3 in a 90/10 % (v/v) $\text{H}_2\text{O}/\text{D}_2\text{O}$ mixture. For referencing purposes, 2,2,3,3-d(4)-3-(tetramethylsilyl)propionic acid sodium salt (TSP, Thermo Fisher Scientific, Karlsruhe, Germany) to a final concentration of 1 mM was added.

A.2.4.2 NMR Spectrometers

Table A.4: NMR spectrometers

Spectrometer	Supplier
VNMRS with 21.1 T (900 MHz)	Agilent Technologies Inc., Santa Clara, USA
VNMRS with 18.8 T (800 MHz)	Agilent Technologies Inc., Santa Clara, USA
Bruker with 16.5 T (700 MHz)	Bruker, Billerica, USA
Bruker with 14.1 T (600 MHz)	Bruker, Billerica, USA
Bruker with 14.1 T (600 MHz)	Bruker, Billerica, USA

A.2.5 Databases

The RCSB Protein Data Bank (PDB; www.rcsb.org) was used for obtaining structural information about proteins, the Universal Protein Resource (UniProt) (www.uniprot.org) was used for protein sequence information. Sequence-specific NMR assignments were deposited in the BMRB database (www.bmrwisc.edu).

A.2.6 Software

Visualization of protein structures and generation of figures for this work was accomplished using the PyMOL Molecular Graphics System v. 1.8.6.0, (Schrödinger, LLC, New York, USA) and Visual Molecular Dynamics (VMD) v. 1.9.3 [140]. VMD was also used for visualizing molecular dynamics trajectories, which have been generated and analyzed using GROMACS v. 4.6 and later versions [141]. For protein sequence analysis (e.g. extinction coefficient, molecular weight or isoelectric point), protein and DNA sequence alignment, and translation of nucleotide sequences into protein sequences, tools from the ExpASy Bioinformatics Resource Portal were used [142].

In addition to PyMOL and VMD, the following programs were used for figure generation: Adobe Illustrator CS5 v. 15.0.0 (Adobe Systems, San José, USA), and Grace v. 5.1.25 (<http://plasma-gate.weizmann.ac.il/Grace/>). ChemDraw Professional v. 16.0 (PerkinElmer, Waltham, USA) was employed to draw the structural formula of chemical compounds and reaction mechanisms.

For setting up NMR experiments, the software TopSpin v. 3.2 (Bruker, Billerica, USA) and VnmrJ v. 4.2 (<http://openvnmrj.org/>) was used for Bruker and Varian

instruments (s. Table A.4), respectively. Data was processed and analyzed using the NMRPipe package v. 9.3 [143] and CcpNmr Analysis v. 2.4 [144]. Talos-N [26] was used to predict protein backbone torsion angles based on chemical shifts and SPARTA+ [68] predicted chemical shifts were used for analyzing and validating molecular dynamics simulations.

MATLAB 2015b (The MathWorks, Inc., Natick, USA) and OriginPro v. 8.5G (OriginLab Corporation, Friedrichsdorf, Germany) were used to analyze and fit data.

Presentations and Awards

Some parts of this work have been presented as a poster or talk at scientific conferences.

Molecular mechanism of *E. coli* 2-deoxyribose-5-phosphate aldolase catalysis

M. Schulte

Presented as a talk on the bio-N³MR Network Meeting, 09/2017, Jülich, Germany

Understanding the mechanistic role of the C-terminal tail of *E. coli* DERA in enzyme catalysis

M. Schulte, P. Neudecker, D. Petrović, R. Hartmann, J. Pietruszka, S. Willbold, D. Willbold, V. Panwalkar

Presented as a poster on the EMBO Practical Course NMR, 08/2017, Basel, Switzerland. **Awarded the poster prize.**

Understanding the mechanistic role of the C-terminal tail of *E. coli* DERA in enzyme catalysis

M. Schulte, P. Neudecker, D. Petrović, R. Hartmann, J. Pietruszka, S. Willbold, D. Willbold, V. Panwalkar

Presented as a poster on the EUROMAR, 07/2017, Warsaw, Poland. **Awarded the Suraj Manrao Student Poster Prize.**

A step towards understanding and optimizing enzyme catalysis

M. Schulte

Presented as a talk on the bio-N³MR Network Meeting, 09/2016, Düsseldorf, Germany.

Anti-aggregation approach to biotech proteins

M. Schulte, V. Panwalkar, P. Skozinski, C. Haist, A. Knapp, K.-E. Jäger, P. Neudecker, D. Willbold

Presented at the XXVIIth ICMRBS, 08/2016, Kyoto, Japan.

Bibliography

- [1] A. Heine, G. DeSantis, J. G. Luz, M. Mitchell, C.-H. Wong, and I. A. Wilson. Observation of covalent intermediates in an enzyme mechanism at atomic resolution. *Science*, 294(5541):369–374, 2001.
- [2] E. Luchinat and L. Banci. A unique tool for cellular structural biology: In-cell NMR. *Journal of Biological Chemistry*, 291(8):3776–3784, 2016.
- [3] D. S. Wishart, C. G. Bigam, J. Yao, F. Abildgaard, H. J. Dyson, E. Oldfield, J. L. Markley, and B. D. Sykes. ^1H , ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *Journal of Biomolecular NMR*, 6(2):135–140, 1995.
- [4] V. Tugarinov, W.-Y. Choy, V. Y. Orekhov, and L. E. Kay. Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proceedings of the National Academy of Sciences of the United States of America*, 102(3):622–627, 2005.
- [5] M. J. Minch. Orientational dependence of vicinal proton-proton NMR coupling constants: The Karplus relationship. *Concepts in Magnetic Resonance*, 6(1):41–56, 1994.
- [6] K. Wüthrich. *NMR of Proteins and Nucleic Acids*. A Wiley-interscience publication. Wiley, 1986.
- [7] L. E. Kay, M. Ikura, R. Tschudin, and A. Bax. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969)*, 89(3):496 – 514, 1990.
- [8] S. Grzesiek and A. Bax. Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *Journal of Magnetic Resonance (1969)*, 96(2):432–440, 1992.
- [9] D. R. Muhandiram and L. E. Kay. Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *Journal of Magnetic Resonance, Series B*, 103(3):203–216, 1994.
- [10] B. T. Farmer, R. A. Venters, L. D. Spicer, M. G. Wittekind, and L. Müller. A refocused and optimized HNCA: Increased sensitivity and resolution in large macromolecules. *Journal of Biomolecular NMR*, 2(2):195–202, 1992.

- [11] A. Bax and M. Ikura. An efficient 3D NMR technique for correlating the proton and ^{15}N backbone amide resonances with the α -carbon of the preceding residue in uniformly $^{15}\text{N}/^{13}\text{C}$ enriched proteins. *Journal of Biomolecular NMR*, 1(1):99–104, 1991.
- [12] R. T. Clubb, V. Thanabal, and G. Wagner. A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue $^1\text{H}^N$, ^{15}N , and $^{13}\text{C}'$ chemical shifts in ^{15}N - ^{13}C -labelled proteins. *Journal of Magnetic Resonance (1969)*, 97(1):213–217, 1992.
- [13] S. Grzesiek and A. Bax. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *Journal of the American Chemical Society*, 114(16):6291–6293, 1992.
- [14] S. Grzesiek, J. Anglister, H. Ren, and A. Bax. Carbon-13 line narrowing by deuterium decoupling in deuterium/carbon-13/nitrogen-15 enriched proteins. Application to triple resonance 4D J connectivity of sequential amides. *Journal of the American Chemical Society*, 115(10):4369–4370, 1993.
- [15] T. Yamazaki, W. Lee, M. Revington, D. L. Mattiello, F. W. Dahlquist, C. H. Arrowsmith, and L. E. Kay. An HNCA pulse scheme for the backbone assignment of ^{15}N , ^{13}C , ^2H -labeled proteins: Application to a 37-kDa Trp repressor-DNA complex. *Journal of the American Chemical Society*, 116(14):6464–6465, 1994.
- [16] M. Sattler and S. W. Fesik. Use of deuterium labeling in NMR: Overcoming a sizeable problem. *Structure*, 4(11):1245 – 1249, 1996.
- [17] D. Rovnyak, D. P. Frueh, M. Sastry, Z.-Y. J. Sun, A. S. Stern, J. C. Hoch, and G. Wagner. Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction. *Journal of Magnetic Resonance*, 170(1):15–21, 2004.
- [18] S. G. Hyberts, S. A. Robson, and G. Wagner. Exploring signal-to-noise ratio and sensitivity in non-uniformly sampled multi-dimensional NMR spectra. *Journal of Biomolecular NMR*, 55(2):167–178, 2013.
- [19] P. Schanda, H. Van Melckebeke, and B. Brutscher. Speeding up three-dimensional protein NMR experiments to a few minutes. *Journal of the American Chemical Society*, 128(28):9042–9043, 2006.
- [20] P. Schanda and B. Brutscher. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *Journal of the American Chemical Society*, 127(22):8014–8015, 2005.

- [21] R. K. Hite, P. Yuan, Z. Li, Y. Hsuing, T. Walz, and R. MacKinnon. Cryo-electron microscopy structure of the Slo2.2 Na⁺-activated K⁺ channel. *Nature*, 527(7577):198–203, 2015.
- [22] J. Du, W. Lü, S. Wu, Y. Cheng, and E. Gouaux. Glycine receptor mechanism elucidated by electron cryo-microscopy. *Nature*, 526(7572):224–229, 2015.
- [23] R. Zhang, G. M. Alushin, A. Brown, and E. Nogales. Mechanistic origin of microtubule dynamic instability and its modulation by EB proteins. *Cell*, 162(4):849–859, 2015.
- [24] Y. He, J. Fang, D. J. Taatjes, and E. Nogales. Structural visualization of key steps in human transcription initiation. *Nature*, 495(7442):481–486, 2013.
- [25] W. J. Metzler, M. Wittekind, V. Goldfarb, L. Mueller, and B. T. Farmer. Incorporation of ¹H/¹³C/¹⁵N-(Ile, Leu, Val) into a perdeuterated, ¹⁵N-labeled protein: Potential in structure determination of large proteins by NMR. *Journal of the American Chemical Society*, 118(28):6800–6801, 1996.
- [26] Y. Shen and A. Bax. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *Journal of Biomolecular NMR*, 56(3):227–241, 2013.
- [27] M.-S. Cheung, M. L. Maguire, T. J. Stevens, and R. W. Broadhurst. DANGLE: A bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *Journal of Magnetic Resonance*, 202(2):223 – 233, 2010.
- [28] S. Grzesiek, F. Cordier, V. Jaravine, and M. Barfield. Insights into biomolecular hydrogen bonds from hydrogen bond scalar couplings. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 45(3):275 – 300, 2004.
- [29] N. J. Baxter and M. P. Williamson. Temperature dependence of ¹H chemical shifts in proteins. *Journal of Biomolecular NMR*, 9(4):359–369, 1997.
- [30] P. Dossset, J.-C. Hus, D. Marion, and M. Blackledge. A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *Journal of Biomolecular NMR*, 20(3):223–231, 2001.
- [31] P. A. Kosen. Spin labeling of proteins. In *Nuclear Magnetic Resonance Part B Structure and Mechanism*, volume 177 of *Methods in Enzymology*, pages 86 – 121. Academic Press, 1989.
- [32] J. Iwahara, D. E. Anderson, E. C. Murphy, and G. M. Clore. EDTA-derivatized deoxythymidine as a tool for rapid determination of protein binding polarity to

- DNA by intermolecular paramagnetic relaxation enhancement. *Journal of the American Chemical Society*, 125(22):6634–6635, 2003.
- [33] G. Pintacuda and G. Otting. Identification of protein surfaces by NMR measurements with a paramagnetic Gd(III) chelate. *Journal of the American Chemical Society*, 124(3):372–373, 2002.
- [34] T. Madl, W. Bermel, and K. Zangger. Use of relaxation enhancements in a paramagnetic environment for the structure determination of proteins using NMR spectroscopy. *Angewandte Chemie International Edition*, 48(44):8259–8262, 2009.
- [35] T. Madl, T. Güttler, D. Görlich, and M. Sattler. Structural analysis of large protein complexes using solvent paramagnetic relaxation enhancements. *Angewandte Chemie International Edition*, 50(17):3993–3997, 2011.
- [36] M. Sakakura, S. Noba, P. A. Luchette, I. Shimada, and R. S. Prosser. An NMR method for the determination of protein-binding interfaces using dioxygen-induced spin-lattice relaxation enhancement. *Journal of the American Chemical Society*, 127(16):5826–5832, 2005.
- [37] J. P. Linge, M. Habeck, W. Rieping, and M. Nilges. ARIA: Automated NOE assignment and NMR structure calculation. *Bioinformatics*, 19:315–316, 2003.
- [38] P. Güntert and L. Buchner. Combined automated NOE assignment and structure calculation with CYANA. *Journal of Biomolecular NMR*, 62(4):453–471, 2015.
- [39] C. D. Schwieters, J. J. Kuszewski, N. Tjandra, and G. M. Clore. The Xplor-NIH NMR molecular structure determination package. *Journal of Magnetic Resonance*, 160(1):65 – 73, 2003.
- [40] Markus Zeeb and Jochen Balbach. Protein folding studied by real-time NMR spectroscopy. *Methods*, 34(1):65 – 74, 2004.
- [41] A. Mittermaier and L. E. Kay. New tools provide new insights in NMR studies of protein dynamics. *Science*, 312(5771):224–228, 2006.
- [42] M. Akke. NMR methods for characterizing microsecond to millisecond dynamics in recognition and catalysis. *Current Opinion in Structural Biology*, 12(5):642 – 647, 2002.
- [43] A. K. Mittermaier and L. E. Kay. Observing biological dynamics at atomic resolution using NMR. *Trends in Biochemical Sciences*, 34(12):601 – 611, 2009.
- [44] H. Y. Carr and E. M. Purcell. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys. Rev.*, 94:630–638, 1954.

- [45] S. Meiboom and D. Gill. Modified spin echo method for measuring nuclear relaxation times. *Review of Scientific Instruments*, 29(8):688–691, 1958.
- [46] Z. Luz and S. Meiboom. Nuclear magnetic resonance study of the protolysis of trimethylammonium ion in aqueous solution—Order of the reaction with respect to solvent. *The Journal of Chemical Physics*, 39(2):366–370, 1963.
- [47] J. P. Loria, M. Rance, and A. G. Palmer. A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy. *Journal of the American Chemical Society*, 121(10):2331–2332, 1999.
- [48] D. F. Hansen, P. Vallurupalli, P. Lundström, P. Neudecker, and L. E. Kay. Probing chemical shifts of invisible states of proteins with relaxation dispersion NMR spectroscopy: How well can we do? *Journal of the American Chemical Society*, 130(8):2667–2675, 2008.
- [49] P. Lundström, H. Lin, and L. E. Kay. Measuring $^{13}\text{C}^\beta$ chemical shifts of invisible excited states in proteins by relaxation dispersion NMR spectroscopy. *Journal of Biomolecular NMR*, 44(3):139–155, 2009.
- [50] A. L. Hansen and L. E. Kay. Quantifying millisecond time-scale exchange in proteins by CPMG relaxation dispersion NMR spectroscopy of side-chain carbonyl groups. *Journal of Biomolecular NMR*, 50(4):347–355, 2011.
- [51] F. A. A. Mulder, N. R. Skrynnikov, B. Hon, F. W. Dahlquist, and L. E. Kay. Measurement of slow (μs -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. *Journal of the American Chemical Society*, 123(5):967–975, 2001.
- [52] N. R. Skrynnikov, F. A. A. Mulder, B. Hon, F. W. Dahlquist, and 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. *Journal of the American Chemical Society*, 123(19):4556–4566, 2001.
- [53] A. J. Baldwin, T. L. Religa, D. F. Hansen, G. Bouvignies, and L. E. Kay. $^{13}\text{CHD}_2$ methyl group probes of millisecond time scale exchange in proteins by ^1H relaxation dispersion: An application to proteasome gating residue dynamics. *Journal of the American Chemical Society*, 132(32):10992–10995, 2010.

- [54] P. Neudecker, P. Robustelli, A. Cavalli, P. Walsh, P. Lundström, A. Zarrine-Afsar, S. Sharpe, M. Vendruscolo, and L. E. Kay. Structure of an intermediate state in protein folding and aggregation. *Science*, 336(6079):362–366, 2012.
- [55] D. D. Boehr, D. McElheny, H. J. Dyson, and P. E. Wright. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science*, 313(5793):1638–1642, 2006.
- [56] D. D. Boehr, R. Nussinov, and P. E. Wright. The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chemical Biology*, 5(11):789–796, 2009.
- [57] I. Bahar, C. Chennubhotla, and D. Tobi. Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Current Opinion in Structural Biology*, 17(6):633–640, 2007.
- [58] E. Z. Eisenmesser, O. Millet, W. Labeikovsky, D. M. Korzhnev, M. Wolf-Watz, D. A. Bosco, J. J. Skalicky, L. E. Kay, and D. Kern. Intrinsic dynamics of an enzyme underlies catalysis. *Nature*, 438(7064):117–121, 2005.
- [59] H. Beach, R. Cole, M. L. Gill, and J. P. Loria. Conservation of μ s-ms enzyme motions in the apo- and substrate-mimicked state. *Journal of the American Chemical Society*, 127(25):9167–9176, 2005.
- [60] K. A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M. A. Wilson, G. A. Petsko, M. Karplus, C. G. Hubner, and D. Kern. Intrinsic motions along an enzymatic reaction trajectory. *Nature*, 450(7171):838–844, 2007.
- [61] P. Vallurupalli, A. Sekhar, T. Yuwen, and L. E. Kay. Probing conformational dynamics in biomolecules via chemical exchange saturation transfer: A primer. *Journal of Biomolecular NMR*, 67(4):243–271, 2017.
- [62] A. G. Palmer, C. D. Kroenke, and J. P. Loria. [10] - Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. In L. T. James, V. Dötsch, and U. Schmitz, editors, *Nuclear Magnetic Resonance of Biological Macromolecules - Part B*, volume 339 of *Methods in Enzymology*, pages 204 – 238. Academic Press, 2001.
- [63] E. A. Cino, W.-Y. Choy, and M. Karttunen. Comparison of secondary structure formation using 10 different force fields in microsecond molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 8(8):2725–2740, 2012.

- [64] S. Rauscher, V. Gapsys, M. J. Gajda, M. Zweckstetter, B. L. de Groot, and H. Grubmüller. Structural ensembles of intrinsically disordered proteins depend strongly on force field: A comparison to experiment. *Journal of Chemical Theory and Computation*, 11(11):5513–5524, 2015.
- [65] J. Henriques, C. Cragnell, and M. Skepö. Molecular dynamics simulations of intrinsically disordered proteins: Force field evaluation and comparison with experiment. *Journal of Chemical Theory and Computation*, 11(7):3420–3431, 2015.
- [66] W. F. van Gunsteren, J. Dolenc, and A. E. Mark. Molecular simulation as an aid to experimentalists. *Current Opinion in Structural Biology*, 18(2):149 – 153, 2008.
- [67] A. K. Dehof, S. Loew, H.-P. Lenhof, and A. Hildebrandt. NightShift: NMR shift inference by general hybrid model training - a framework for NMR chemical shift prediction. *BMC Bioinformatics*, 14(1):98, 2013.
- [68] Y. Shen and A. Bax. SPARTA+: A modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *Journal of Biomolecular NMR*, 48(1):13–22, 2010.
- [69] B. Han, Y. Liu, S. W. Ginzinger, and D. S. Wishart. SHIFTX2: Significantly improved protein chemical shift prediction. *Journal of Biomolecular NMR*, 50(1):43–57, 2011.
- [70] C. C. F. Blake, D. F. Koenig, G. A. Mair, A. C. T. North, D. C. Phillips, and V. R. Sarma. Structure of hen egg-white lysozyme: A three-dimensional fourier synthesis at 2 Å resolution. *Nature*, 206:757–761, 1965.
- [71] M. P. Williamson, T. F. Havel, and K. Wüthrich. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ^1H nuclear magnetic resonance and distance geometry. *Journal of Molecular Biology*, 182(2):295 – 315, 1985.
- [72] E. Fischer. Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der deutschen chemischen Gesellschaft*, 27(3):2985–2993, 1894.
- [73] D. E. Koshland. Application of a theory of enzyme specificity to protein synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 44(2):98–104, 1958.
- [74] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes. The energy landscapes and motions of proteins. *Science*, 254(5038):1598–1603, 1991.

- [75] C. J. Tsai, S. Kumar, B. Ma, and R. Nussinov. Folding funnels, binding funnels, and protein function. *Protein Science : A Publication of the Protein Society*, 8(6):1181–1190, 1999.
- [76] S. Kumar, B. Ma, C. J. Tsai, N. Sinha, and R. Nussinov. Folding and binding cascades: Dynamic landscapes and population shifts. *Protein Science : A Publication of the Protein Society*, 9(1):10–19, 2000.
- [77] D. W. Miller and K. A. Dill. Ligand binding to proteins: The binding landscape model. *Protein Science : A Publication of the Protein Society*, 6(10):2166–2179, 1997.
- [78] T. R. Weikl and C. von Deuster. Selected-fit versus induced-fit protein binding: Kinetic differences and mutational analysis. *Proteins: Structure, Function, and Bioinformatics*, 75(1):104–110, 2009.
- [79] O. F. Lange, N.-A. Lakomek, C. Farès, G. F. Schröder, K. F. A. Walter, S. Becker, J. Meiler, H. Grubmüller, C. Griesinger, and B. L. de Groot. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science*, 320(5882):1471–1475, 2008.
- [80] D. Joseph, G. A. Petsko, and M. Karplus. Anatomy of a conformational change: Hinged “lid” motion of the triosephosphate isomerase loop. *Science*, 249(4975):1425–1428, 1990.
- [81] F. Massi, C. Wang, and A. G. Palmer. Solution NMR and computer simulation studies of active site loop motion in triosephosphate isomerase. *Biochemistry*, 45(36):10787–10794, 2006.
- [82] J. C. Williams and A. E. McDermott. Dynamics of the flexible loop of triosephosphate isomerase: The loop motion is not ligand gated. *Biochemistry*, 34(26):8309–8319, 1995.
- [83] D. A. Keedy, L. R. Kenner, M. Warkentin, R. A. Woldeyes, J. B. Hopkins, M. C. Thompson, A. S. Brewster, A. H. Van Benschoten, E. L. Baxter, M. Uervirojnangkoorn, S. E. McPhillips, J. Song, R. Alonso-Mori, J. M. Holton, W. I. Weis, A. T. Brunger, S. M. Soltis, H. Lemke, A. Gonzalez, N. K. Sauter, A. E. Cohen, H. van den Bedem, R. E. Thorne, and J. S. Fraser. Mapping the conformational landscape of a dynamic enzyme by multitemperature and XFEL crystallography. *eLife*, 4:e07574, 2015.
- [84] R. Sridharan, J. Zuber, S. M. Connelly, E. Mathew, and M. E. Dumont. Fluorescent approaches for understanding interactions of ligands with G protein coupled

- receptors. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1838(1, Part A):15 – 33, 2014.
- [85] C. A. Waudby, A. Ramos, L. D. Cabrita, and J. Christodoulou. Two-dimensional NMR lineshape analysis. *Scientific Reports*, 6:24826, 2016.
- [86] V. Panwalkar, P. Neudecker, M. Schmitz, J. Lecher, M. Schulte, K. Medini, M. Stoldt, M. A. Brimble, D. Willbold, and A. J. Dingley. The Nedd4–1 WW domain recognizes the PY motif peptide through coupled folding and binding equilibria. *Biochemistry*, 55(4):659–674, 2016.
- [87] M. J. Fedor and J. R. Williamson. The catalytic diversity of RNAs. *Nature Reviews Molecular Cell Biology*, 6(5):399–412, 2005.
- [88] U. K. Narta, S. S. Kanwar, and W. Azmi. Pharmacological and clinical evaluation of L-asparaginase in the treatment of leukemia. *Critical Reviews in Oncology/Hematology*, 61(3):208 – 221, 2007.
- [89] C. A. Masschelein, D. S. Ryder, and J.-P. Simon. Immobilized cell technology in beer production. *Critical Reviews in Biotechnology*, 14(2):155–177, 1994.
- [90] O. Colagrande, A. Silva, and M. D. Fumi. Recent applications of biotechnology in wine production. *Biotechnology Progress*, 10(1):2–18, 1994.
- [91] P. F. Fox and L. Stepaniak. Enzymes in cheese technology. *International Dairy Journal*, 3(4):509 – 530, 1993.
- [92] O. Kirk, T. V. Borchert, and C. C. Fuglsang. Industrial enzyme applications. *Current Opinion in Biotechnology*, 13(4):345 – 351, 2002.
- [93] K. Buchholz, V. Kasche, and U. T. Bornscheuer. *Biocatalysts and enzyme technology*. John Wiley & Sons, 2012.
- [94] E. C. Webb. *Enzyme nomenclature 1992. Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes*. Academic Press, 1992.
- [95] A. J. Barrett. Enzyme nomenclature. Recommendations 1992. *European Journal of Biochemistry*, 232(1-6):1–1, 1995.
- [96] M. Wolf-Watz, V. Thai, K. Henzler-Wildman, G. Hadjipavlou, E. Z. Eisenmesser, and D. Kern. Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nature Structural and Molecular Biology*, 11(10):945–949, 2004.

- [97] D. McElheny, J. R. Schnell, J. C. Lansing, H. J. Dyson, and P. E. Wright. Defining the role of active-site loop fluctuations in dihydrofolate reductase catalysis. *Proceedings of the National Academy of Sciences of the United States of America*, 102(14):5032–5037, 2005.
- [98] K. J. Laidler and M. C. King. Development of transition-state theory. *The Journal of Physical Chemistry*, 87(15):2657–2664, 1983.
- [99] R. Wolfenden, M. Snider, C. Ridgway, and B. Miller. The temperature dependence of enzyme rate enhancements. *Journal of the American Chemical Society*, 121(32):7419–7420, 1999.
- [100] J. G. Zalatan, T. D. Fenn, and D. Herschlag. Comparative enzymology in the alkaline phosphatase superfamily to determine the catalytic role of an active site metal ion. *Journal of Molecular Biology*, 384(5):1174–1189, 2008.
- [101] J. K. Lassila, J. G. Zalatan, and D. Herschlag. Biological phosphoryl-transfer reactions: Understanding mechanism and catalysis. *Annual Review of Biochemistry*, 80:669–702, 2011.
- [102] L. Michaelis and M. L. Menten. Die Kinetik der Invertinwirkung. *Biochemische Zeitschrift*, 49:333–369, 1913.
- [103] P. Valentin-Hansen, F. Boëtius, K. Hammer-Jespersen, and I. Svendsen. The primary structure of *Escherichia coli* K12 2-deoxyribose 5-phosphate aldolase. *European Journal of Biochemistry*, 125(3):561–566, 1982.
- [104] M. G. Tozzi, M. Camici, L. Mascia, F. Sgarrella, and P. L. Ippata. Pentose phosphates in nucleoside interconversion and catabolism. *FEBS Journal*, 273(6):1089–1101, 2006.
- [105] T. P. Cao, J. S. Kim, M. H. Woo, J. M. Choi, Y. Jun, K. H. Lee, and S. H. Lee. Structural insight for substrate tolerance to 2-deoxyribose-5-phosphate aldolase from the pathogen *Streptococcus suis*. *Journal of Microbiology*, 54(4):311–321, 2016.
- [106] M. L. Tonkin, A. S. Halavaty, R. Ramaswamy, J. Ruan, M. Igarashi, H. M. Ngo, and M. J. Boulanger. Structural and functional divergence of the aldolase fold in *Toxoplasma gondii*. *Journal of Molecular Biology*, 427(4):840–852, 2015.
- [107] A. Heine, J. G. Luz, C.-H. Wong, and I. A. Wilson. Analysis of the class I aldolase binding site architecture based on the crystal structure of 2-deoxyribose-5-phosphate aldolase at 0.99 Å resolution. *Journal of Molecular Biology*, 343(4):1019 – 1034, 2004.

- [108] M. Dick, R. Hartmann, O. H. Weiergraber, C. Bisterfeld, T. Classen, M. Schwarten, P. Neudecker, D. Willbold, and J. Pietruszka. Mechanism-based inhibition of an aldolase at high concentrations of its natural substrate acetaldehyde: Structural insights and protective strategies. *Chemical Science*, 7:4492–4502, 2016.
- [109] C. L. Windle, M. Müller, A. Nelson, and A. Berry. Engineering aldolases as biocatalysts. *Current Opinion in Chemical Biology*, 19(Supplement C):25 – 33, 2014.
- [110] H. J. M. Gijzen and C.-H. Wong. Unprecedented asymmetric aldol reactions with three aldehyde substrates catalyzed by 2-deoxyribose-5-phosphate aldolase. *Journal of the American Chemical Society*, 116(18):8422–8423, 1994.
- [111] W. A. Greenberg, A. Varvak, S. R. Hanson, K. Wong, H. Huang, P. Chen, and M. J. Burk. Development of an efficient, scalable, aldolase-catalyzed process for enantioselective synthesis of statin intermediates. *Proceedings of the National Academy of Sciences of the United States of America*, 101(16):5788–5793, 2004.
- [112] G. DeSantis, J. Liu, D. P. Clark, A. Heine, I. A. Wilson, and C.-H. Wong. Structure-Based mutagenesis approaches toward expanding the substrate specificity of D-2-Deoxyribose-5-phosphate aldolase. *Bioorganic & Medicinal Chemistry*, 11(1):43 – 52, 2003.
- [113] A. Hershko, A. Ciechanover, and I. A. Rose. Resolution of the ATP-dependent proteolytic system from reticulocytes: A component that interacts with ATP. *Proceedings of the National Academy of Sciences*, 76(7):3107–3110, 1979.
- [114] M. J. Clague, H. Liu, and S. Urbé. Governance of endocytic trafficking and signaling by reversible ubiquitylation. *Developmental Cell*, 23(3):457 – 467, 2012.
- [115] L. K. Teixeira and S. I. Reed. Ubiquitin ligases and cell cycle control. *Annual Review of Biochemistry*, 82(1):387–414, 2013.
- [116] M. Scheffner and S. Kumar. Mammalian HECT ubiquitin-protein ligases: Biological and pathophysiological aspects. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1843(1):61 – 74, 2014.
- [117] S. Kumar, Y. Tomooka, and M. Noda. Identification of a set of genes with developmentally down-regulated expression in the mouse brain. *Biochemical and Biophysical Research Communications*, 185(3):1155 – 1161, 1992.
- [118] O. Staub, S. Dho, P. Henry, J. Correa, T. Ishikawa, J. McGlade, and D. Rotin. WW domains of Nedd4 bind to the proline-rich PY motifs in the epithelial Na⁺

- channel deleted in Liddle's syndrome. *The EMBO Journal*, 15(10):2371–2380, 1996.
- [119] O. Staub, H. Yeager, P. J. Plant, H. Kim, S. A. Ernst, and D. Rotin. Immunolocalization of the ubiquitin-protein ligase Nedd4 in tissues expressing the epithelial Na⁺ channel (ENaC). *American Journal of Physiology - Cell Physiology*, 272(6):C1871–C1880, 1997.
- [120] B. Yang and S. Kumar. Nedd4 and Nedd4-2: Closely related ubiquitin-protein ligases with distinct physiological functions. *Cell death and Differentiation*, 17(1):68–77, 2010.
- [121] X. R. Cao, N. L. Lill, N. Boase, P. P. Shi, D. R. Croucher, H. Shan, J. Qu, E. M. Sweezer, T. Place, P. A. Kirby, R. J. Daly, S. Kumar, and B. Yang. Nedd4 controls animal growth by regulating IGF-1 signaling. *Science signaling*, 1(38):ra5–ra5, 2008.
- [122] C. Chen and L. E. Matesic. The Nedd4-like family of E3 ubiquitin ligases and cancer. *Cancer and Metastasis Reviews*, 26(3):587–604, 2007.
- [123] F. Bernassola, M. K., A. Ciechanover, and G. Melino. The HECT family of E3 ubiquitin ligases: Multiple players in cancer development. *Cancer Cell*, 14(1):10 – 21, 2008.
- [124] J. Wang, Q. Peng, Q. Lin, C. Childress, D. Carey, and W. Yang. Calcium activates Nedd4 E3 ubiquitin ligases by releasing the C2 domain-mediated auto-inhibition. *Journal of Biological Chemistry*, 285(16):12279–12288, 2010.
- [125] T. Mund and H. R. B. Pelham. Control of the activity of WW-HECT domain E3 ubiquitin ligases by NDFIP proteins. *EMBO reports*, 10(5):501–507, 2009.
- [126] P. Bork and M. Sudol. The WW domain: A signalling site in dystrophin? *Trends in Biochemical Sciences*, 19(12):531 – 533, 1994.
- [127] H. I. Chen and M. Sudol. The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules. *Proceedings of the National Academy of Sciences*, 92(17):7819–7823, 1995.
- [128] M. J. Macias, V. Gervais, C. Civera, and H. Oschkinat. Structural analysis of WW domains and design of a WW prototype. *Nature Structural Biology*, 7:375–379, 2000.
- [129] A. Zarrinpar and W. A. Lim. Converging on proline: The mechanism of WW domain peptide recognition. *Nature Structural Biology*, 7:611–613, 2000.

- [130] X. Huang, F. Poy, R. Zhang, A. Joachimiak, M. Sudol, and M. J. Eck. Structure of a WW domain containing fragment of dystrophin in complex with β -dystroglycan. *Nature Structural Biology*, 7:634–638, 2000.
- [131] V. Kanelis, D. Rotin, and J. D. Forman-Kay. Solution structure of a Nedd4 WW domain–ENaC peptide complex. *Nature Structural Biology*, 8:407–412, 2001.
- [132] M. A. Verdecia, M. E. Bowman, K. P. Lu, T. Hunter, and J. P. Noel. Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nature Structural Biology*, 7:639–643, 2000.
- [133] B. K. Kay, M. P. Williamson, and M. Sudol. The importance of being proline: The interaction of proline-rich motifs in signaling proteins with their cognate domains. *The FASEB Journal*, 14(2):231–241, 2000.
- [134] J. S. Lott, S. J. Coddington-Lawson, P. H. Teesdale-Spittle, and F. J. McDonald. A single WW domain is the predominant mediator of the interaction between the human ubiquitin-protein ligase Nedd4 and the human epithelial sodium channel. *Biochemical Journal*, 361(3):481–488, 2002.
- [135] P. C. Henry, V. Kanelis, M. C. O’Brien, B. Kim, I. Gautschi, J. Forman-Kay, L. Schild, and D. Rotin. Affinity and specificity of interactions between Nedd4 isoforms and the epithelial Na⁺ channel. *Journal of Biological Chemistry*, 278(22):20019–20028, 2003.
- [136] M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T. J. Gibson, and D. G. Higgins. Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21):2947–2948, 2007.
- [137] R. Bobby, K. Medini, P. Neudecker, T. V. Lee, M. A. Brimble, F. J. McDonald, J. S. Lott, and A. J. Dingley. Structure and dynamics of human Nedd4-1 WW3 in complex with the α ENaC PY motif. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1834(8):1632 – 1641, 2013.
- [138] V. Kanelis, M. C. Bruce, N. R. Skrynnikov, D. Rotin, and J. D. Forman-Kay. Structural determinants for high-affinity binding in a Nedd4 WW3* domain-Comm PY motif complex. *Structure*, 14(3):543 – 553, 2006.
- [139] T. Peng, J. S. Zintsmaster, A. T. Namanja, and J. W. Peng. Sequence-specific dynamics modulate recognition specificity in WW domains. *Nature Structural and Molecular Biology*, 14(4):325–332, 2007.

- [140] W. Humphrey, A. Dalke, and K. Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.
- [141] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845–854, 2013.
- [142] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel, and A. Bairoch. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, 31(13):3784–3788, 2003.
- [143] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, 6(3):277–293, 1995.
- [144] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, and E. D. Laue. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Structure, Function, and Bioinformatics*, 59(4):687–696, 2005.