Electronic Thesis and Dissertation Repository

11-27-2012 12:00 AM

Experimental and computational analysis of the structure and dynamics of intrinsically disordered proteins

Elio Anthony Cino
The University of Western Ontario

Supervisor
Dr. Wing-Yiu Choy
The University of Western Ontario

Graduate Program in Biochemistry
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy
© Elio Anthony Cino 2012

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Amino Acids, Peptides, and Proteins Commons

EXPERIMENTAL AND COMPUTATIONAL ANALYSIS OF THE STRUCTURE AND DYNAMICS OF INTRINSICALLY DISORDERED PROTEINS

(Spine title: Experimental and computational analysis of disordered proteins)

(Thesis format: Integrated Article)

by

Elio Anthony <u>Cino</u>

Graduate Program in Biochemistry

A thesis submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

The School of Graduate and Postdoctoral Studies

The University of Western Ontario

London, Ontario, Canada

THE UNIVERSITY OF WESTERN ONTARIO
School of Graduate and Postdoctoral Studies

CERTIFICATE OF EXAMINATION

Supervisor   Examiners

_____   _____
Dr. Wing-Yiu Choy   Dr. Lars Konermann

Co-supervisor

                                                    _____
                                                    Dr. Graeme Hunter

_____
Dr. Mikko Karttunen

                                                    _____
                                                    Dr. Joe Mymryk

Supervisory Committee

_____   _____
Dr. Gary Shaw   Dr. Valerie Booth

_____
Dr. Lars Konermann

The thesis by

Elio Anthony Cino

entitled:

Experimental and computational analysis of the structure and dynamics of intrinsically disordered proteins

is accepted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

_____   _____
            Date   Chair of the Thesis Examination Board

Abstract

Intrinsically disordered proteins (IDPs) are abundant in cells and have central roles in protein-protein interaction networks. Many are involved in cancer, aging and neurodegenerative diseases. The structure and dynamics of IDPs is intimately related to their interactions with binding partners. Because IDPs are inherently flexible and do not have a single conformation, conventional methods and conditions for determining structure and dynamics of globular proteins may not be directly applicable. Nuclear magnetic resonance (NMR) spectroscopy is one of the primary techniques characterizing the structures and dynamics of IDPs, but one cannot rely solely on NMR data. A primary aim of this work was to use Molecular Dynamics (MD) simulations in conjunction with NMR and other biophysical techniques to achieve a deeper understanding of the structure and dynamics of IDPs. To establish suitable parameters and force field choice for simulating IDPs, extensive MD simulations were performed and the results were compared to experimental data. Using computational and experimental techniques, the interactions between peptides from 9 disordered proteins with a common target were interrogated. The findings allowed us to determine key factors in modulating the affinities of the various interactions and highlighted the importance of Linear Motifs (LMs) in IDP target recognition and binding. IDP binding was also investigated from the perspective of the binding partner. The backbone resonances of the ~32 kDa target were assigned and the binding interface was mapped in the presence of a peptide from a disordered binding partner. Chemical shift changes distant from the interaction site indicated that IDP binding is a complex process, which should be studied from the perspectives of the partner and target. Because IDPs are highly sensitive to environmental conditions, the effects of molecular crowding on the dynamics of IDPs were also investigated. I found that crowding might have differential effects on the conformational propensities of distinct regions of some IDPs. This information will help to understand the behavior of IDPs in cellular environments and to determine suitable conditions for accurately studying them. This work has helped to improve the understanding of how IDP structure and dynamics relate to target binding.

Keywords

Co-authorship

All of the works presented in this thesis represent published, or to be published first author publications. In all works, one or both of my supervisors, Dr. Wing-Yiu Choy and Dr. Karttunen appear as authors. While I have been the primary planner, performer, analyzer, writer and editor of each work, they have assisted in all of these aspects as well. They have also provided financial resources. Dr. Choy was responsible for developing the LS3 model for chapter 6.

In chapter 3, Dr. Wong-Ekkabut contributed to the experimental design, data analysis and manuscript editing.

For chapter 4, second author, Ryan Killoran, assisted in planning, performing and analysis of the ITC and bioinformatics experiments.

Dr. Fan, Dr. Yang and Dr. Ryan McKay assisted with collection and analysis of the NMR data for chapter 5.

Anne Brickenden assisted with cloning and molecular biology for chapter 5.

Dedication

For Mom & Dad

Acknowledgements

This work would not have been possible without the help, guidance and support of my amazing supervisors, Dr. Wing-Yiu (James) Choy and Dr. Mikko Karttunen.

I would also like to acknowledge my advisory committee members, Dr. Gary Shaw and Dr. Lars Konermann for their valuable input and suggestions.

Thanks to all of the Choy lab members, past and present for your cooperation, thoughtful discussions and support. Special thanks to Anne Brickenden for teaching me everything I know about bench work.

Table of contents

List of tables

List of figures

List of abbreviations, symbols and nomenclature

Å	Angstrom
Amber	Assisted model building with energy refinement
BPTF	Nucleosome-remodelling factor
Cby	Chibby protein
CHARMM	Chemistry at HARvard molecular mechanics
CNS	Crystallography and NMR system
COSY	Correlation spectroscopy
C.V.	Circular variance
D_2O	Deuterium Oxide
Da	Dalton
DSS	2,2'-dimethyl-2-silapentane-5-sulfonate
DSSP	Dictionary of protein secondary structure
DTT	Dithiothreitol
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
FAC1	Fetal Alz-50 clone 1
GROMOS	GROningen MOlecular Simulation
HCl	Hydrochloric acid
HSQC	Heteronuclear Single Quantum Coherence
Hz	Hertz
IDP	Intrinsically disordered protein
IKKβ	Inhibitor of nuclear factor kappa-B kinase
IPTG	Isopropyl β-D thiogalactopyranoside
ITC	Isothermal titration calorimetry
K_a	Association constant
K_d	Dissociation constant
Keap1	Kelch-like ECH-associated protein 1
LM	Linear motif
M	Molar
MD	Molecular dynamics
MHz	Megahertz
MoRF	Molecular recognition feature
NaCl	Sodium chloride
$NaPO_4$	Sodium phosphate
NOE	Nuclear overhauser effect
NOESY	Nuclear overhauser spectroscopy
NMR	Nuclear magnetic resonance
Nrf2	Nuclear factor erythroid 2-related factor 2
OPLS	Optimized potentials for lipid simulations
PALB2	Partner and localizer of BRCA2
PGAM5	Serine/threonine protein phosphatase
ppm	Parts per million
PTMA	Prothymosin alpha
PDB	Protein data bank
PBC	Periodic boundary condition

PPI   Protein-protein interaction
PSE   Preformed structural element
R_1 Longitudinal relaxation
R_2 Transverse relaxation
RMSD   Root mean square deviation
SPC   Simple point charge
SSP   Secondary structure propensity
TC-1   Thyroid cancer protein 1
TEV   Tobacco etch virus
TOCSY   Total correlation spectroscopy
Tris   Tris(hydroxymethyl)aminomethane
TROSY   Transverse relaxation optimized spectroscopy
WTX   Wilms tumor gene on X chromosome
VMD   Visual molecular dynamics

Ala (A)   Alanine
Arg (R)   Arginine
Asn (N)   Asparagine
Asp (D)   Aspartic acid
Cys (C)   Cysteine
Gln (Q)   Glutamine
Glu (E)   Glutamic acid
Gly (G)   Glycine
His (H)   Histidine
Ile (I)   Isoleucine
Leu (L)   Leucine
Lys (K)   Lysine
Met (M)   Methionine
Phe (F)   Phenylalanine
Pro (P)   Proline
Ser (S)   Serine
Thr (T)   Threonine
Trp (W)   Tryptophan
Tyr (Y)   Tyrosine
Val (V)   Valine

1   Introduction

1.1   Intrinsically disordered proteins

Intrinsically disordered proteins (IDPs) are a biologically functional class of proteins that comprise ~30% of the eukaryotic proteome (1-4). The abundance of IDPs in organisms suggests that they are essential for numerous functions. It was once thought that a protein must adopt a defined three-dimensional structure to function properly; however, the discovery of biologically active disordered proteins illustrates that IDPs carry out their functions through different mechanisms than globular proteins. The defined conformation of a globular protein is often important for stabilizing a single interaction site, allowing it to partake in a specific and high affinity interaction. In contrast, the different possible conformations of an IDP often allow for specific, but generally lower affinity interactions with numerous different targets (5). These properties are well suited to their roles in signaling pathways, where reversible binding and the ability to interact with multiple partners is often required (5).
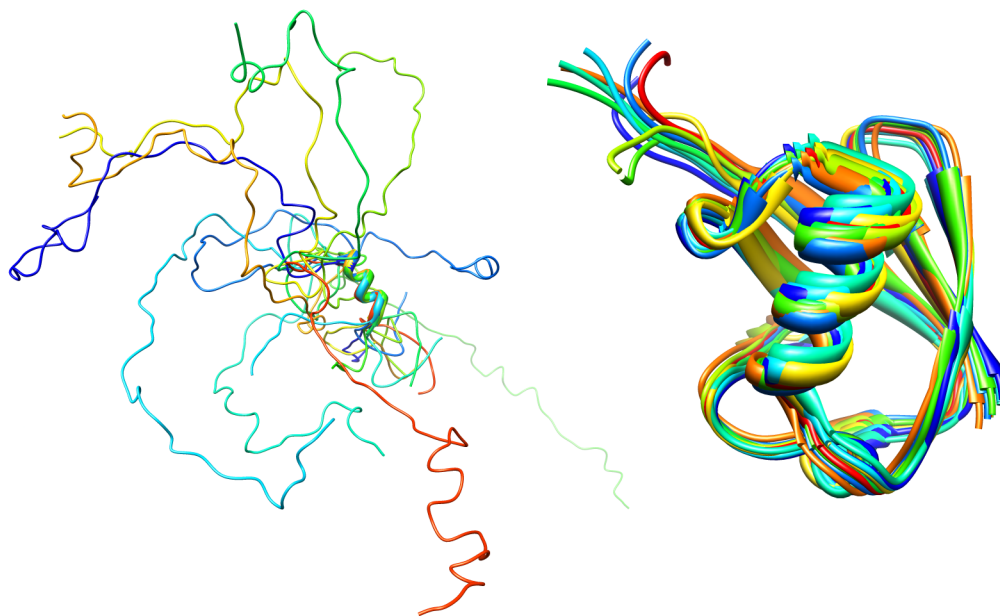
Figure 1.1 NMR structural ensembles of the intrinsically disordered Thylakoid soluble phosphoprotein TSP9 and globular protein, Ubiquitin.

Thylakoid soluble phosphoprotein TSP9 (left, PDB id: 2FFT) and Ubiquitin (right, PDB id: 1D3Z).

The structural differences between IDPs and globular proteins (Figure 1.1) illustrate that disorder is encoded in their sequences. IDPs have different amino acid compositions than globular proteins. They are enriched in charged, polar and the structure-breaking residues, glycine and proline (1, 6). Hydrophobic and aromatic content is also lower in IDPs (1, 6). As a result, IDPs usually lack hydrophobic cores, which stabilize globular proteins. Due to the absence of hydrophobic cores, IDPs are more dynamic compared to globular proteins. Ordered regions of globular proteins undergo relatively small fluctuations around their equilibrium backbone atom positions over time. In comparison, IDPs usually exhibit significant changes in their φ and ψ angles over time and may not have specific equilibrium values (6). Although IDPs lack stable tertiary structures, they may contain elements of secondary structure, which are often crucial for their functionality (eg. target binding).

The dynamic properties of IDPs are intimately related to the timescale of conformational exchange within the ensemble (Figure 1.1), which govern target recognition and how these proteins function. Different structures in the ensemble can participate in the interactions with distinct targets; therefore, the rate of exchange between conformers can have significant impact on the protein function (7, 8).

1.2   Target binding by IDPs

Protein-protein interaction (PPI) networks are essential for most biological functions. IDPs often act as hubs in PPI networks, where they regulate biochemical processes through low-affinity and high-specificity protein-protein interactions (2). Although IDPs are involved in crucial biological functions, such as signal transduction and transcription (5), there is relatively limited data, compared to globular proteins, describing how they interact with binding partners. Knowledge of the mechanisms that

IDPs employ to bind to their targets will aid in development of therapeutic approaches targeting these interactions (9).

Studies have shown that the unique structural properties of IDPs are important for their ability to act as hubs in PPI networks (10-12). Like proteins in the unfolded state, IDPs do not adopt completely random coil conformations (13-16). Many IDPs have considerable conformational propensities (17-19). Segments of IDPs that contain residual structure may act as molecular recognition features (MoRFs) for binding to their targets (17, 19). MoRFs are defined as 5-25 residue target binding regions, which may contain residual structure in their unbound states (17). It is possible for IDPs to contain multiple MoRFs along their sequence allowing them to interact with different binding partners. Also, binding to different targets can cause a MoRF to adopt completely different structures (17, 20, 21). For example, the same region of the intrinsically disordered N-terminal of p53 can adopt a helix or a sheet depending on its binding partners (21).

Some IDPs interact with targets through preformed structural elements (PSEs) that resemble the bound state (19, 22-24), while others may couple conformational changes with target binding (coupled folding and binding model). For IDPs that bind using PSEs, the binding region structure is already formed in the unbound state. In the coupled folding and binding model, the IDP undergoes a disorder-to-order transition upon binding to a target (2). Binding of IDPs may also be modulated by a combination of these two mechanisms (25).
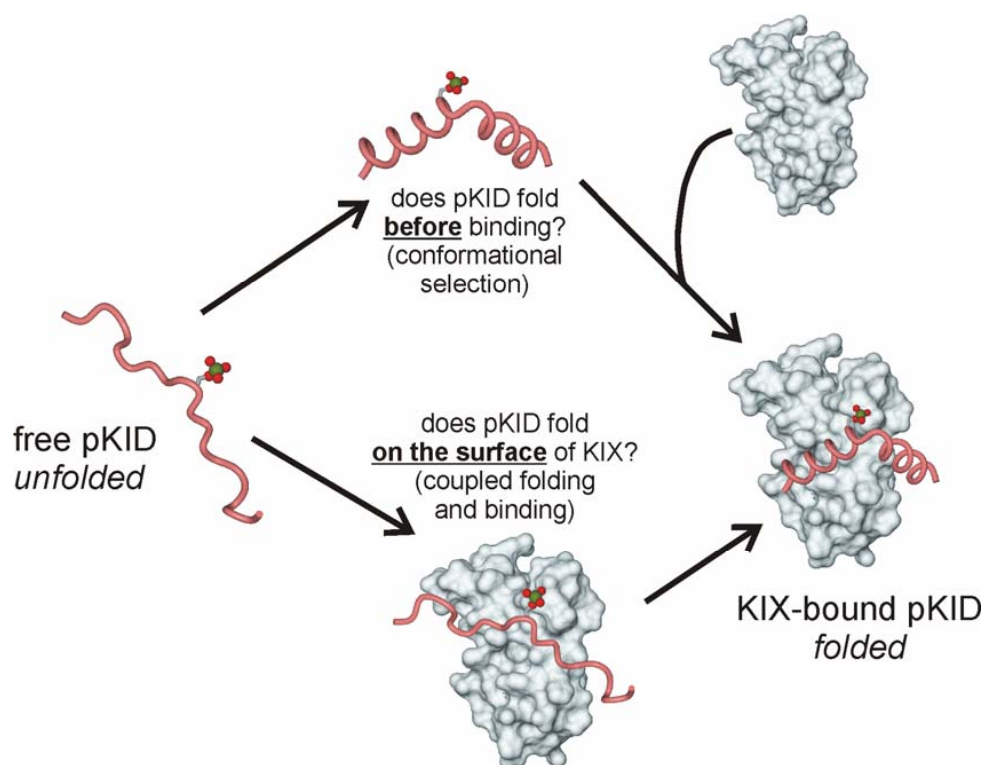
Figure 1.2 Mechanisms of target binding by IDPs.

Possible mechanism(s) for the interaction between the intrinsically disordered phosphorylated kinase inducible activation domain (pKID) and KIX domain of the CREB binding protein (26). Usage of this figure has been granted by the Nature Publishing Group (licence number 2993681394339).

1.3   IDPs and diseases

IDPs have been found to be associated with cancers, neurodegenerative diseases and aging (5, 27). Because their structural plasticity often allows them to interact with a large number of targets, IDPs are enriched in signaling networks (5). They have been shown to be involved in various activities, such as signal transduction, apoptosis, cell differentiation, and neuron function (28). In addition to missignalling diseases, IDPs are also involved in diseases related to protein misfolding (5). Mutations, exposure to toxins, aberrant posttranslational modifications and other factors can lead to misfolding of these

proteins, which can have serious consequences. The term 'misfolding' may seem counterintuitive because IDPs do not adopt stable folds. However, in several neurodegenerative disorders, the normal, disordered form of a protein may convert to a 'misfolded' conformation that is prone to aggregation. The build up of non-functional, ordered and highly stable amyloid fibrils in various tissues can result in specific pathological conditions depending on the protein (5). Because IDPs typically have low structural preferences compared to folded proteins, it is thought that they can transform into an aggregate-prone conformation more easily. Understanding the links between sequence, structure and target binding by IDPs is crucial for improving our understanding of their roles in disease and developing treatments and cures.

1.4   Techniques for characterizing IDPs

Nuclear magnetic resonance spectroscopy (NMR) and X-ray crystallography are commonly used methods to obtain atomic details of protein structure and dynamics. The dynamic nature of IDPs makes acquiring diffracting crystals of them in unbound states nearly impossible (29). Therefore, NMR is the primary technique for studying the structure and dynamics of IDPs (30, 31). NMR can yield a wealth of data, but there are limitations. Data collected by NMR are averaged over time and represent an ensemble average. For globular proteins, the protein core is stable and the ensemble average is usually a good representation of a true physical state. For highly flexible polymers, like IDPs, the ensemble average may not represent a realistic physical state. To comprehensively study the structure and dynamics of IDPs, and their relationship to target binding, NMR techniques were combined with molecular dynamics (MD) simulations and other biophysical techniques, such as isothermal calorimetry (ITC) for many of the projects in this thesis. A brief introduction to each of these techniques is provided below.

1.4.1   NMR spectroscopy

NMR is a technique used to determine the chemical environment of atoms (32-35). This information can be used to learn about the structure, dynamics, chemical environment, etc of the molecules that the atoms are contained in. The technique exploits the magnetic properties of specific nuclei in order to obtain this information. In a magnetic field (B_o), NMR active nuclei, atoms that have odd number of protons, neutrons, or both, and spin values of ½ (eg. ^1H, ^{15}N, ^{13}C) will precess in either spin-aligned or spin-opposed orientations parallel to the field (Figure 1.3). Nuclei in the spin-aligned state have a slightly lower energy and are slightly more populated than the spin opposed nuclei. By applying electromagnetic radiation, in the form of radio waves, these nuclei can be temporarily be excited to the higher energy, unaligned state. Data collected during the return to the lower energy state gives information about the local environment of the nuclei (chemical shift). This is the basis of all NMR experiments. Various types and patterns of electromagnetic pulses are combined to generate specific NMR experiments to obtain the desired information (eg. chemical environment, dynamics, etc) about the nuclei.

NMR can be used to determine protein structures and numerous other properties, such as dynamics. The process of determining a protein structure by NMR can be divided into two parts: assignment and restraint collection. Assignment refers to the determination of the chemical shift values of spin ½ nuclei. Next, restraints are gathered (distances, angles, orientations, etc) for the assigned atoms, which are used to computationally fold the polypeptide in such a way that the restraints are satisfied. The two steps are not always mutually exclusive. For example, through space distance restraints may be also used to assist with or verify the resonance assignment. For small proteins and peptides (~30 residues or less), homonuclear NMR experiments (eg. ^1H signals only) may be sufficient to assign the proton resonances (via ^1H-^1H COSY and TOCSY experiments) and determine the structure (via ^1H-^1H NOESY). This approach is used in chapter 4 to assign and collect distance restraints for several ~20-mer peptides. Due to spectral crowding and overlap in the ^1H dimension, multidimensional, heteronuclear experiments are required to assign and determine the structures of larger

proteins and peptides. In chapter 5, several heteronuclear experiments were used to assign backbone resonances for a ~32 kDa protein domain. In this case, the goal of assignment was not to determine the structure. Instead, we used the assigned ^{1}H-^{15}N HSQC spectrum to determine the residue-specific chemical shift changes upon addition of a peptide from a binding partner. This allowed us to map the binding interface (chemical shift mapping) onto a previously determined crystal structure. Furthermore, assignment of the ^{13}Cα and ^{13}Cβ resonances allowed for the determination of secondary structure content of the polypeptide. NMR can also be used to study protein dynamics. Relaxation measurements allow one to determine motions occurring on ps-ns timescales, which can help to identify structured and flexible regions of proteins. These types of experiments are employed in chapter 6 to measure changes in the dynamics of IDPs under crowded conditions. The experiments mentioned here are just a few examples of the uses of NMR.
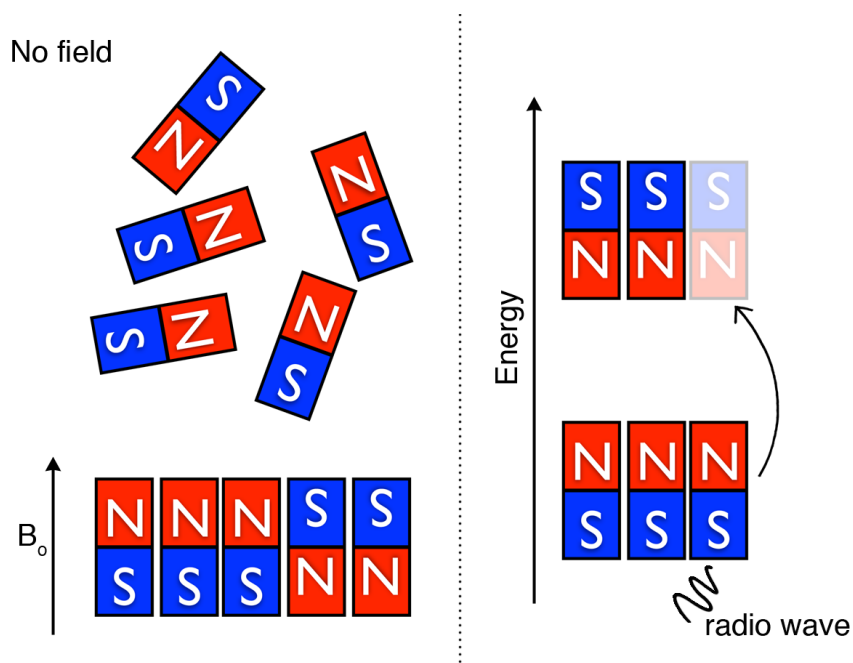
Figure 1.3 The basis of NMR experiments.

NMR active nuclei precess in either spin-aligned or spin-opposed orientations parallel to the field (B_o). Lower energy nuclei can be temporarily be excited to the

higher energy state. Data collected as the nuclei return to equilibrium gives information about the local environment of the nuclei.

1.4.2   MD simulations

MD is a simulation of the movement of particles, accomplished by solving Newton's equation of motion (2nd law) for a system of interacting particles:

$$\bar{F}_i = m_i \bar{a}_i = m_i \frac{d\bar{v}_i}{dt} = m_i \frac{d^2\bar{r}_i}{dt^2}$$

Where F_i is the force exerted on particle i, m_i is the mass of particle i, a_i is the acceleration of particle i, v_i is the velocity of particle i and r_i is the position of particle i at time t.

Using this equation, a trajectory (particle positions as a function of time) can be calculated by integration once the initial positions and velocities of the particles are known. The positions may come from a known structure (eg. Crystal or NMR protein structure) and the velocities are often taken from a Maxwellian distribution at the desired temperature. From these values, the forces on the particles (usually atoms) are calculated. The particles are allowed to move for a short period of time (the timestep), and integration gives their $t+1$ positions. Time is moved forward according to the timestep. This process is repeated as long as necessary. A number of different integration algorithms may be used, such as leap frog and velocity Verlet (36, 37).

The calculation of forces is the most time-consuming process in the generation of a trajectory. The forces may also be expressed as the negative gradient of the potential energy:

$$\bar{F}_i = -\nabla_i V$$

The potential energy function (V) is the sum of terms for bonded and non-bonded energies. These energies are determined as a function of the atom positions (Cartesian coordinates, r) in the system:

$$V(r) = E_{bonded} + E_{non-bonded}$$

The bonded interactions typically contain terms for stretching, angle-bending, dihedral angle and, usually, a term for improper dihedral angles, which functions to maintain the planarity of planar molecules (Figure 1.4):

$$E_{bonded} = E_{bond-stretch} + E_{angle-bend} + E_{dihedral} + E_{improper-dihedral}$$
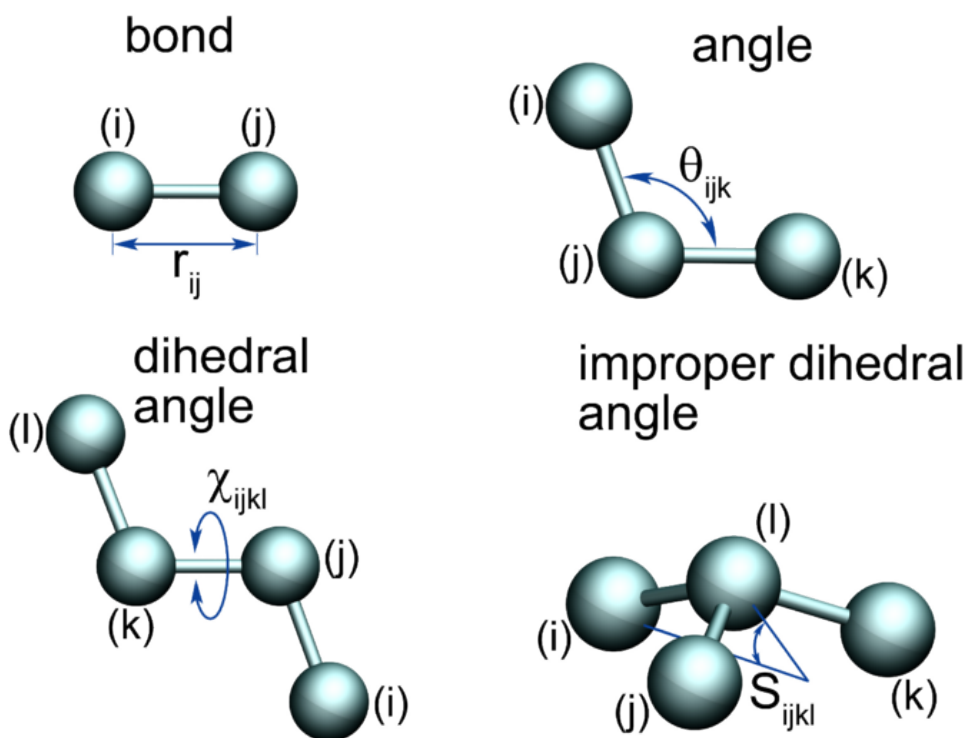
Figure 1.4 Bonded energy terms for MD simulations.

Figure usage granted under MesoBioNano (MBN) explorer academic license agreement (http://www.mbnexplorer.com/academic-license-agreement) (38).

The non-bonded energy consists of terms for Van der Waals and electrostatic interaction energies. Van der Waals forces are approximated using the Lennard-Jones potential:

$$V_{LJ} = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right]$$

Where ε is the well depth, σ is the distance where the inter-particle potential is zero and r is the distance between the particles. Figure 1.5 illustrates examples of Lennard-Jones potentials.
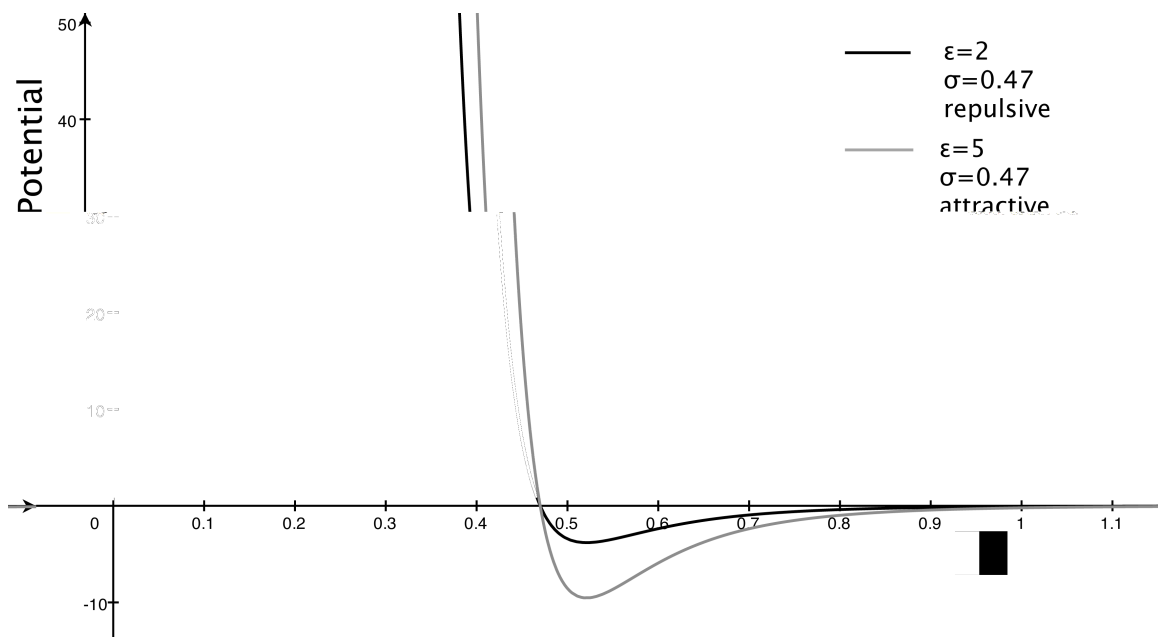
Figure 1.5 Examples of Lennard-Jones potentials.

r is the distance between particles in nanometers and the Lennard-Jones potential is given in kJ/mol. All Lennard-Jones potentials include repulsive and attractive components. Here, the grey curve has a greater attractive component compared to the black one.

The Coulomb interactions between two charged particles is given by:

$$V_C(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_o \varepsilon_r r_{ij}}$$

Where q is the charge on each particle, ε_r is the relative dielectric constant and r_{ij} is the separation distance.

The values associated with each of the energy terms are dictated by the force field being used for the simulation. Chapter 2 details some of the most commonly used force fields for biomolecular simulations.

When conducting MD simulations of biomolecules, such as proteins, it is usually desirable to mimic experimental, or laboratory conditions (eg. temperature, pressure). Direct integration of Newton's equation of motion will result in the system energy being conserved, which corresponds to an isolated system. In contrast, laboratory experiments are typically "open" systems. To reproduce experimental conditions it is usually desirable to run simulations under the NPT ensemble, where the number of particles, pressure and temperature of the system are kept constant. This is accomplished by coupling the system to temperature and/or pressure baths. Thermostat (39, 40) and barostat (41-44) algorithms can be employed to achieve the NPT ensemble.

One final concept that warrants introduction is that of periodic boundary conditions (PBCs). Typically, in a simulation, it is desirable to mimic a bulk system (e.g., protein in solution); however, because this would be too computationally expensive, a finite system with a small number of particles is preferred. Using PBCs, a finite system is surrounded by translated copies of itself (Figure 1.6). This eliminates boundary effects that would be present at the system edges if PBCs were not employed. In some cases, one may want to perform a simulation without the use of PBCs (e.g., a crystal).
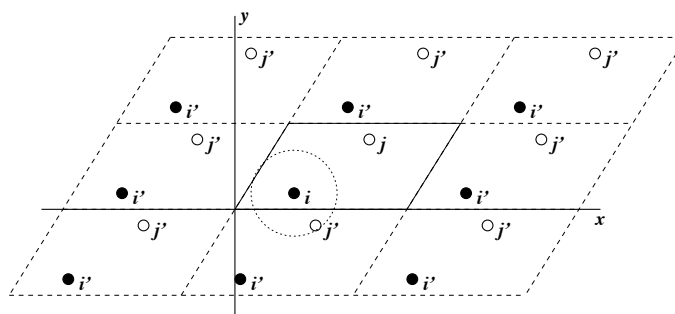
Figure 1.6 Periodic boundary conditions in 2D.

The system (center) is surrounded by translated copies of itself, eliminating edge effects and mimicking a bulk solvent environment (45). Figure usage granted under the GNU General Public License (http://www.gnu.org/copyleft/gpl.html).

MD simulations have become increasingly valuable tools in the field of biochemistry. Berni Alder invented the technique in the late 1950's and in 1959 performed the first MD simulation – of a group of interacting hard spheres under periodic boundary conditions (46). The first protein MD simulation was performed in 1975 of the bovine pancreatic trypsin inhibitor (BPTI) (47). Since then, advances in computing power have allowed for MD simulations to be used to study considerably more complex systems. They have become important predictive tools. Today, MD simulations are routinely used for NMR structure determination, refining X-ray crystal structures, protein-ligand docking, drug discovery/refinement, protein folding and countless other purposes. As computational power and software algorithms improve, the uses and applications of MD simulations will expand correspondingly.

1.4.3   ITC

ITC is a method to quantitatively measure the thermodynamic parameters of interactions in solution (48). The binding affinity (K_a), enthalpy changes (ΔH) and stoichiometry (n) of the interaction between two or more molecules are measured

directly. Determination of these parameters allows one to derive the Gibbs free energy and entropy changes (ΔG and ΔS, respectively) using the relationship:

$$\Delta G = -RT \ln K_a = \Delta H - T\Delta S$$

Where R is the gas constant and T is the temperature.

Briefly, in an ITC experiment, a binding target is loaded into the sample cell and subjected to a stepwise titration of precise volumes of ligand into the same cell (Figure 1.7). The reference cell contains water or buffer. A constant power is applied to the reference cell heater. During the titration, sensitive thermocouples measure the temperature differences between the reference and sample cells. In an exothermic reaction, heat is evolved in the sample cell and the power to the heater is decreased. In an endothermic reaction, heat is absorbed in the sample cell and the heater is activated. The heat input required to maintain the same temperatures in the sample and reference cells is measured throughout the experiment (Figure 1.8). Integration of the heat input with respect to time as a function of the molar ratio ((ligand)/(target)) gives the thermodynamic parameters of interest.
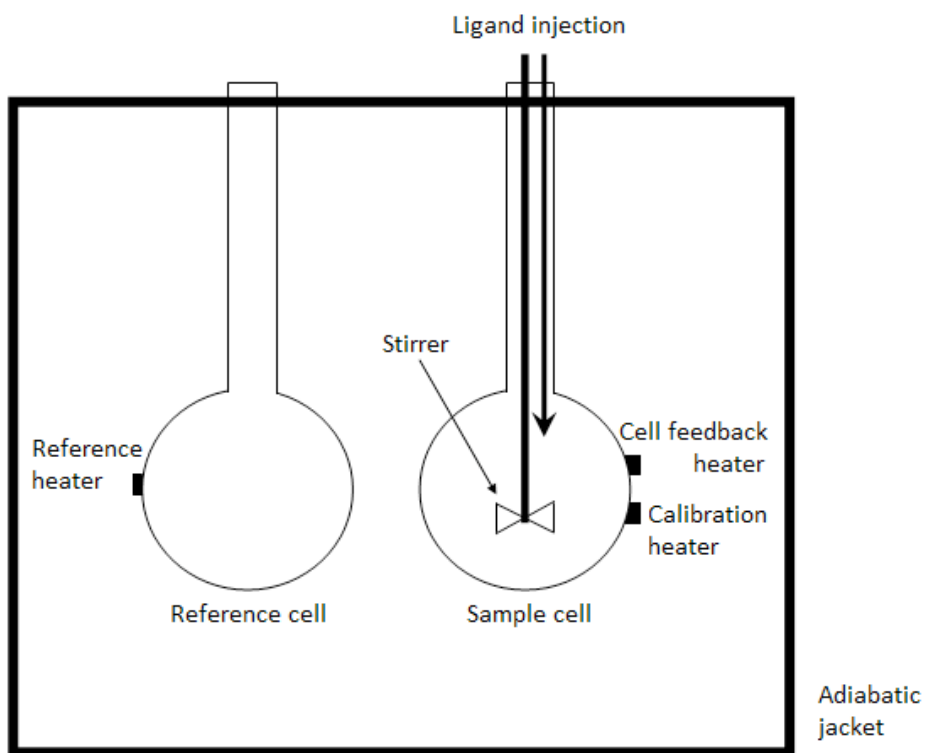
Figure 1.7 ITC instrument schematic.

Figure has been released to the public domain without conditions.
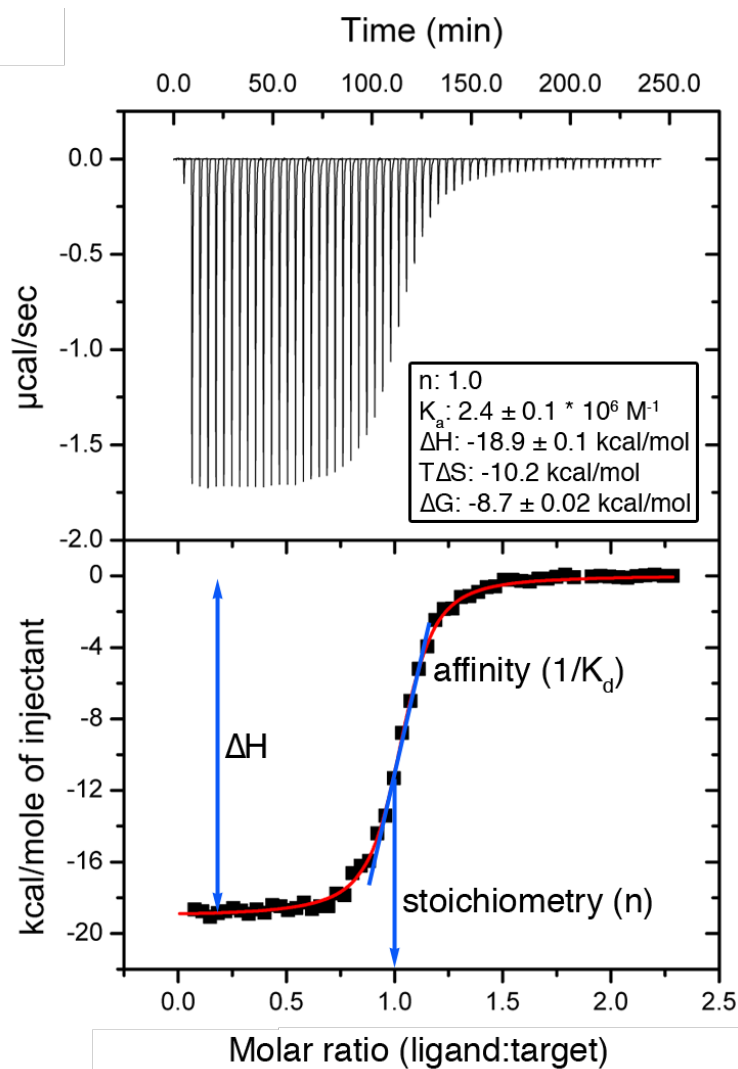
Figure 1.8 Typical ITC data

The top panel shows the raw ITC data and the bottom panel shows the binding isotherm.

1.5   Significance and aims

The primary aim of this work was to improve the understanding of how the structure and dynamics of IDPs are related to target binding. To accomplish this, experimental and computational techniques were used to investigate the molecular mechanisms by which several IDPs interact with a common binding partner. The

common partner is the Kelch domain of Kelch-like ECH-associated protein 1 (Keap1). This domain is a key component in regulating the cellular response to oxidative stress (49). It adopts a β-propeller conformation that comprises an interaction site for numerous proteins (50-55) (Figure 1.9, left). Perhaps the most important interaction that the Kelch domain is involved in is with the transcription factor Nuclear factor erythroid 2-related factor 2 (Nrf2). The intrinsically disordered Neh2 domain of Nrf2 can bind two Kelch domain monomers via distinct motifs with different binding affinities (54, 56). Under normal cellular conditions, association of both motifs with the Kelch domain leads to its degradation (54) (Figure 1.9, right). However, when cells are exposed to conditions of oxidative stress (eg. reactive oxygen species, toxins, etc), the lower affinity interaction is disrupted, and Nrf2 is able to promote transcription of cytoprotective genes (54). Because the overexpression of these genes may have clinical benefits, development of compounds that induce the Keap1-Nrf2 pathway are being actively developed for various purposes (see chapter 4 for details) (57). Moreover, misregulation of this pathway can result in cancers, premature aging and neurodegenerative diseases (58, 59). In recent years, numerous other proteins that bind to the same interface on the Kelch domain and share similar binding motif sequences with Nrf2 have been discovered, but not extensively studied (53, 60-66). Some of the proteins have been shown to promote cytoprotective gene expression by disrupting the low affinity Nrf2-Kelch domain interaction (52, 62, 65, 66), while others have different, or unknown roles. Several of the works presented in this thesis focus on the relationships between sequences and structures in controlling the affinities of the various protein-protein interactions. The findings should be useful for improving the general understanding of the mechanisms used by IDPs to interact with targets, deciphering the biological roles of these interactions, and development of therapeutic agents that target the Nrf2-Kelch domain interaction.

Top view of an Neh2 peptide bound to Kelch (PDB code: 1X2R).
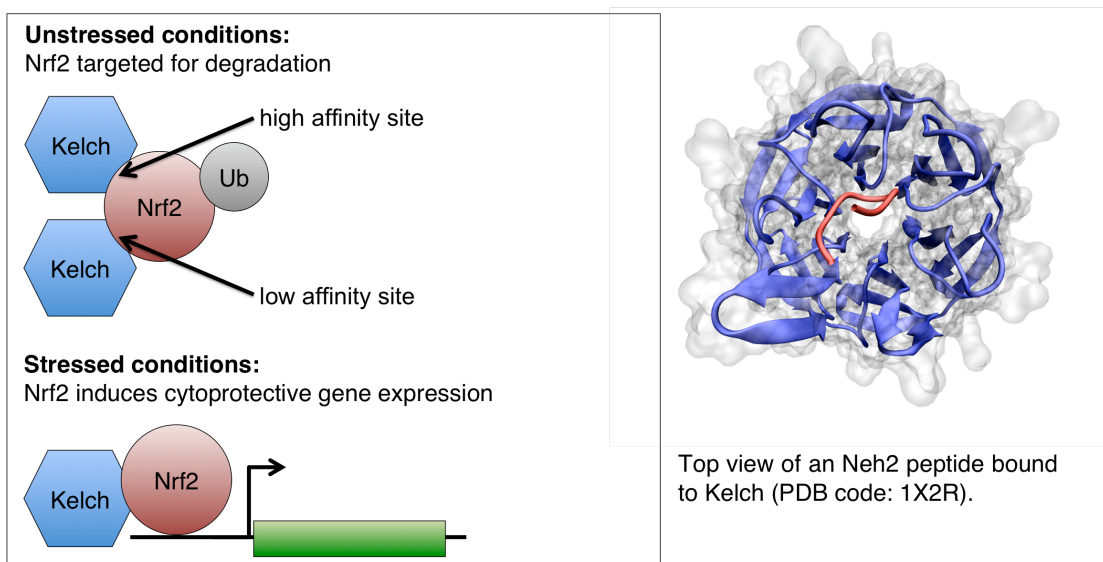
Figure 1.9 The structure of the Kelch domain of Keap1 and its interaction with the Neh2 domain of Nrf2.

Two-site binding is disrupted under conditions of oxidative stress preventing Nrf2 degradation and allowing promotion of cytoprotective gene expression. The structure corresponding to PDB id 1X2R was obtained from (67) and rendered with VMD (68).

1.6   Thesis outline

The following chapters (2-6) represent individual, first author publications, in their unmodified forms. A primary aim of this work was to use Molecular Dynamics (MD) simulations in conjunction with NMR and other biophysical techniques to achieve a deeper understanding of the structure and dynamics of IDPs. Although NMR is a powerful tool, the data represents an ensemble average, which can be limiting when studying a molecule rapidly undergoing conformational changes like an IDP. Because MD trajectories allow for the examination of discrete states and the interconversion between conformations, they can be used to complement experimental techniques. However, because IDPs have been recently discovered relative to their globular counterparts, parameters and protocols for accurately simulating them have not been widely established. Chapter 2 describes our efforts towards determining suitable

parameters and force field selection for performing MD simulations of IDPs. This was accomplished by performing extensive MD simulations of a peptide from the high affinity motif of Nrf2 with 10 different force fields and different parameters. The results were compared to experimental data. After determining suitable combinations of parameters for simulating IDPs, we applied the methodology to more diverse IDP systems. In chapters 3 and 4, we combine MD simulations with experimental techniques to dissect the mechanisms used by 9 different IDPs for binding to a common target, the Kelch domain of Keap1. Our findings provide valuable insights into the mechanisms used by IDPs for target binding and should also help to elucidate the biological roles of the various protein-protein interactions. In chapter 5, we report the backbone resonance assignments for the Kelch domain of Keap1 and map its binding interface with a peptide from the high affinity motif of Nrf2. This study was important for examining IDP binding from the perspective of a target, allowing us to more thoroughly understand IDP interactions. Chapter 6 focuses on determining how molecular crowding affects the dynamics of IDPs. Inside cells, the concentration of macromolecules can reach up to 400 g/L. In such crowded environments, proteins are expected to behave differently than *in vitro*. The dynamic properties of IDPs are intimately related to the timescale of conformational exchange within the ensemble, which govern target recognition and how these proteins function. Therefore, assessing how these properties are affected by environmental conditions is crucial for accurately studying this class of proteins. The major conclusions from this thesis and possible future directions are discussed in chapter 7.

1.7   References

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

2. Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* 293:321-31.

3. Dunker AK et al. (1998) Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pac Symp Biocomput* 473-84.

4. Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ (2000) Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform* 11:161-71.

5. Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys* 37:215-46.

6. Radivojac P et al. (2007) Intrinsic disorder and functional proteomics. *Biophys J* 92:1439-56.

7. Mittag T, Kay LE, Forman-Kay JD (2010) Protein dynamics and conformational disorder in molecular recognition. *J Mol Recognit* 23:105-16.

8. Smock RG, Gierasch LM (2009) Sending signals dynamically. *Science* 324:198-203.

9. Cheng Y et al. (2006) Rational drug design via intrinsically disordered protein. *Trends Biotechnol* 24:435-42.

10. Patil A, Nakamura H (2006) Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. *FEBS Lett* 580:2041-5.

11. Dosztányi Z, Chen J, Dunker AK, Simon I, Tompa P (2006) Disorder and sequence repeats in hub proteins and their implications for network evolution. *J Proteome Res* 5:2985-95.

12. Haynes C et al. (2006) Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput Biol* 2:e100.

13. Gall C, Xu H, Brickenden A, Ai X, Choy WY (2007) The intrinsically disordered TC-1 interacts with Chibby via regions with high helical propensity. *Protein Sci* 16:2510-8.

14. Choy WY, Forman-Kay JD (2001) Calculation of ensembles of structures representing the unfolded state of an SH3 domain. *J Mol Biol* 308:1011-32.

15. Morar AS, Olteanu A, Young GB, Pielak GJ (2001) Solvent-induced collapse of alpha-synuclein and acid-denatured cytochrome c. *Protein Sci* 10:2195-9.

16. Shortle D, Ackerman MS (2001) Persistence of native-like topology in a denatured protein in 8 M urea. *Science* 293:487-9.

17. Mohan A et al. (2006) Analysis of molecular recognition features (MoRFs). *J Mol Biol* 362:1043-59.

18. Sigalov AB, Zhuravleva AV, Orekhov VY (2007) Binding of intrinsically disordered proteins is not necessarily accompanied by a structural transition to a folded form. *Biochimie* 89:419-21.

19. Fuxreiter M, Simon I, Friedrich P, Tompa P (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J Mol Biol* 338:1015-26.

20. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J* 272:5129-48.

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

22. Wright PE, Dyson HJ (2009) Linking folding and binding. *Curr Opin Struct Biol* 19:31-8.

23. Tsai CJ, Ma B, Sham YY, Kumar S, Nussinov R (2001) Structured disorder and conformational selection. *Proteins* 44:418-27.

24. Kumar S, Ma B, Tsai CJ, Sinha N, Nussinov R (2000) Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci* 9:10-9.

25. Espinoza-Fonseca LM (2009) Reconciling binding mechanisms of intrinsically disordered proteins. *Biochem Biophys Res Commun* 382:479-82.

26. Sugase K, Dyson HJ, Wright PE (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 447:1021-5.

27. Iakoucheva LM, Brown CJ, Lawson JD, Obradović Z, Dunker AK (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol* 323:573-84.

28. Raychaudhuri S, Dey S, Bhattacharyya NP, Mukhopadhyay D (2009) The role of intrinsically unstructured proteins in neurodegenerative diseases. *PLoS One* 4:e5566.

29. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6:197-208.

30. Eliezer D (2009) Biophysical characterization of intrinsically disordered proteins. *Curr Opin Struct Biol* 19:23-30.

31. Mittag T, Forman-Kay JD (2007) Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol* 17:3-14.

32. Rule GS, and Hitchens TK (2005) *Fundamentals of protein NMR spectroscopy* (Springer).

33. Rabi II, Zacharias JR, Millman S, Kusch P (1992) Milestones in magnetic resonance: 'a new method of measuring nuclear magnetic moment' . 1938. *J Magn Reson Imaging* 2:131-3.

34. Bloch F, Hansen WW, Packard M (1946) The nuclear induction experiment. *Phys Rev* 70:474-485.

35. Bloch F, Hansen WW, Packard M (1946) Nuclear induction *Phys Rev* 70:460-474.

36. Swope WC, Andersen HC, Berens PH, Wilson KR (1982) A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J Chem Phys* 76:637.

37. Berendsen HJC, and van Gunsteren WF (1986) in *Molecular-Dynamics Simulation of Statistical-Mechanical Systems* (North-Holland, Amsterdam), p 43-65.

38. Solov'yov IA, Yakubovich AV, Nikolaev PV, Volkovets I, Solov'yov AV (2012) MesoBioNano explorer-A universal program for multiscale computer simulations of complex molecular structure and dynamics. *J Comput Chem* (in press).

*3*9. Berendsen HJC, Postma JPM, Van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3684-3690.

40. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126:014101.

41. Hoover WG (1985) Canonical dynamics: Equilibrium phase-space distributions. *Phys Rev A* 31:1695-1697.

42. Nose S, Klein ML (1983) Constant pressure molecular dynamics for molecular systems. *Mol Phys* 50:1055-1076.

43. Nosé S (1984) A molecular dynamics method for simulations in the canonical ensemble. *Mol Phys* 52:255-268.

44. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* 52:7182-7190.

45. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput* 4:435-447.

46. Alder BJ, Wainwright TE (1959) Studies in molecular dynamics. I. General method. *J Chem Phys* 31:459-466.

47. Levitt M, Warshel A (1975) Computer simulation of protein folding *Nature* 253:694.

48. Pierce MM, Raman CS, Nall BT (1999) Isothermal titration calorimetry of protein-protein interactions. *Methods* 19:213-221.

49. Itoh K et al. (1999) Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain. *Genes Dev* 13:76-86.

50. Li X, Zhang D, Hannink M, Beamer LJ (2004) Crystal structure of the Kelch domain of human Keap1. *J Biol Chem* 279:54750-8.

51. Tong KI, Kobayashi A, Katsuoka F, Yamamoto M (2006) Two-site substrate recognition model for the Keap1-Nrf2 system: a hinge and latch mechanism. *Biol Chem* 387:1311-20.

52. Lo SC, Li X, Henzl MT, Beamer LJ, Hannink M (2006) Structure of the Keap1:Nrf2 interface provides mechanistic insight into Nrf2 signaling. *EMBO J* 25:3605-17.

53. Lo SC, Hannink M (2006) PGAM5, a Bcl-XL-interacting protein, is a novel substrate for the redox-regulated Keap1-dependent ubiquitin ligase complex. *J Biol Chem* 281:37893-903.

54. Tong KI et al. (2006) Keap1 recruits Neh2 through binding to ETGE and DLG motifs: characterization of the two-site molecular recognition model. *Mol Cell Biol* 26:2887-900.

55. Padmanabhan B, Nakamura Y, Yokoyama S (2008) Structural analysis of the complex of Keap1 with a prothymosin alpha peptide. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 64:233-8.

56. Tong KI et al. (2007) Different electrostatic potentials define ETGE and DLG motifs as hinge and latch in oxidative stress response. *Mol Cell Biol* 27:7511-21.

57. Li Y, Paonessa JD, Zhang Y (2012) Mechanism of chemical activation of Nrf2. *PLoS One* 7:e35122.

58. Sporn MB, Liby KT (2012) NRF2 and cancer: the good, the bad and the importance of context. *Nat Rev Cancer* 12:564-71.

59. Zhang Y, Gordon GB (2004) A strategy for cancer prevention: stimulation of the Nrf2-ARE signaling pathway. *Mol Cancer Ther* 3:885-93.

60. Strachan GD et al. (2004) Fetal Alz-50 clone 1 interacts with the human orthologue of the Kelch-like Ech-associated protein. *Biochemistry* 43:12113-22.

61. Kim JE et al. (2010) Suppression of NF-kappaB signaling by KEAP1 regulation of IKKbeta activity through autophagic degradation and inhibition of phosphorylation. *Cell Signal* 22:1645-54.

62. Komatsu M et al. (2010) The selective autophagy substrate p62 activates the stress responsive transcription factor Nrf2 through inactivation of Keap1. *Nat Cell Biol* 12:213-23.

63. Niture SK, Jaiswal AK (2011) INrf2 (Keap1) targets Bcl-2 degradation and controls cellular apoptosis. *Cell Death Differ* 18:439-51.

64. Tian H et al. (2012) Keap1: One stone kills three birds Nrf2, IKKβ and Bcl-2/Bcl-xL. *Cancer Lett* 325:26-34.

65. Camp ND et al. (2012) Wilms Tumor Gene on X Chromosome (WTX) Inhibits Degradation of NRF2 Protein through Competitive Binding to KEAP1 Protein. *J Biol Chem* 287:6539-50.

66. Ma J et al. (2012) PALB2 interacts with KEAP1 to promote NRF2 nuclear accumulation and function. *Mol Cell Biol* 32:1506-17.

67. Padmanabhan B et al. (2006) Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer. *Mol Cell* 21:689-700.

68. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14:33-8, 27-8.

2 Comparison of secondary structure formation using 10 different force fields in microsecond molecular dynamics simulations

Elio A. Cino[†], Wing-Yiu Choy[†*] and Mikko Karttunen[‡*]

Department of Biochemistry[†], 1151 Richmond St. North, The University of Western Ontario, London, Ontario, Canada N6A 5C1, Department of Chemistry[‡], 200 University Avenue West, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

*Corresponding authors: mikko.karttunen@uwaterloo.ca, jchoy4@uwo.ca

2.1 Abstract

We have compared molecular dynamics (MD) simulations of a β-hairpin forming peptide derived from the protein Nrf2 with ten biomolecular force fields using trajectories of at least one microsecond long. The total simulation time was 37.2 microseconds. Previous studies have shown that different force fields, water models, simulation methods and parameters can affect simulation outcomes. The MD simulations were done in explicit solvent with a 16-mer Nrf2 β-hairpin forming peptide using Amber ff99SB-ILDN, Amber ff99SB*-ILDN, Amber ff99SB, Amber ff99SB*, Amber ff03, Amber ff03*, GROMOS96 43a1p, GROMOS96 53a6, CHARMM27 and OPLS-AA/L force fields. The effects of charge-groups, terminal capping and phosphorylation on the peptide folding were also examined. Despite using identical starting structures and simulation parameters, we observed clear differences amongst the various force fields and even between replicates using the same force field. Our simulations show that the uncapped peptide folds into a native-like β-hairpin structure at 310 K when Amber ff99SB-ILDN, Amber ff99SB*-ILDN, Amber ff99SB, Amber ff99SB*, Amber ff03, Amber ff03*, GROMOS96 43a1p or GROMOS96 53a6 were used. The CHARMM27 simulations were able to form native hairpins in some of the elevated temperature simulations, while the OPLS-AA/L simulations did not yield native hairpin structures at any temperatures tested. Simulations that used charge-groups or peptide capping groups were not largely different from their uncapped counterparts with single atom charge-groups. On the other hand, phosphorylation of the threonine residue located at the β-turn significantly affected the hairpin formation. To our knowledge, this is the first study comparing such a large set of force fields with respect to β-hairpin folding. Such a comprehensive comparison will offer useful guidance to others conducting similar types of simulations.

2.2 Introduction

Atomistic molecular dynamics (MD) simulations are a versatile tool for studying protein folding and function. They can provide detailed atomistic information, which may be difficult to obtain by experimental techniques. Increases in computational power have allowed for simulations to reach experimentally relevant time scales at the microsecond level: MD simulations have been used to study the folding of peptides and small proteins (1-9) and to model other biological systems. The current record for an atomistic simulation of protein conformational changes, as far as we know, is 1 ms reached by Shaw et al. (7) for the 58-residue protein BPTI.

One of the major challenges in protein folding simulations is choosing an appropriate force field, see e.g. Ref (10). This is due to possible biases different force fields have towards certain types of secondary structure (3, 11-14). Ideally, the force field should be fully validated with experimental data, but that is typically not possible as it would involve validation against different structures and other physical properties from a large number of independent and fully validated experiments – mission impossible since experiments have their own error sources due to, e.g., instrumentation. While a completely transferable force field does not exist, modifications of existing force fields have led to improvements in agreement with experimental data (15-22).

In this work, we compared 10 biomolecular force fields with respect to the folding of a peptide derived from Nuclear factor erythroid 2-related factor 2 (Nrf2). Nrf2 is an important transcription factor that regulates the expression of genes responsive to oxidative stress (23, 24). The protein consists of six highly homologous regions (Neh1-6 domains). Structural analysis showed that the N-terminal Neh2 domain is intrinsically disordered, a novel class of proteins that are extremely dynamic in nature (25-29). Under homeostatic conditions, this domain binds two Kelch units of a Keap1 dimer through two separate motifs: a high affinity 'ETGE' motif and a lower affinity 'DLG' motif (30). Crystallographic data has shown that the 'ETGE' motif and its surrounding residues (residues 75-83) form a β-hairpin structure upon binding to the Kelch domain of Keap1 (PDB ids: 2FLU and 1X2R) (30, 31). NMR-derived ^{1}H, ^{1}H NOEs suggest that residual structure spanning from residues 74-85, likely in the form of a β-hairpin, also exists in the

free-state of Neh2 (30, 31). Other experimental data has shown that a peptide containing residues 74-87 can compete with the full-length Nrf2 for binding Keap1 (31). Here, we chose to use a 16-mer human Nrf2 peptide with the sequence [72]AQLQLDEETGEFLPIQ[87] for our MD simulations. This peptide contains the 'ETGE' motif and should be of sufficient length to form the necessary interactions to stabilize the β-hairpin structure. It is noteworthy that the phosphorylation of Thr-80 has been shown to impair the binding to Keap1 (31). Since Neh2 is largely disordered and lacks a tertiary structure, this β-hairpin likely folds independently, making it a good target for folding simulations (30).

In addition to Nrf2, several other proteins that contain 'ETGE'-like motifs have been shown to interact with the Kelch domain of Keap1. These include PGAM5 (32), FAC1 (33), PTMA (34), p62 (35), WTX (36) and PALB2 (37). Some of these Keap1 interacting proteins have only been recently discovered, which suggests that this list of targets may still be growing. Structures of PTMA (Prothymosin alpha) and p62 peptides in complex with Keap1 indicate that their 'ETGE'-like motifs bind to the same region as the 'ETGE' motif of Nrf2 and form similar hairpin structures in their bound states (31, 34, 35). Interestingly, MD simulations from our previous work showed that the binding motifs of Nrf2 and PTMA have tendency to form hairpin structures that resembled the bound state conformation even in the absence of Keap1 (9). With the list of Keap1 binding partners seemingly expanding and MD simulations becoming an increasingly important and predictive tool, it is important to establish appropriate simulation protocols for these systems, including force field choice.

β-hairpins are a type of protein supersecondary structure consisting of two hydrogen bonded antiparallel β-strands connected by a turn. These structural elements are common in globular proteins because they reverse the direction of the protein backbone, allowing the formation of compact structures. β-hairpin motifs are sometimes involved in protein-protein interactions and it has been suggested that they can act as nucleation sites for protein folding (31, 38-40). In this study, we compared folding simulations of a β-hairpin peptide derived from the intrinsically disordered Neh2 domain. Starting from an unfolded structure, we performed extensive (1-2 μs each, totaling 37.2 μs) atomistic

molecular dynamics simulations using 10 different force fields (details in next section). We selected these force fields primarily because they are commonly used in biomolecular simulations, including those of β-hairpin folding (3, 9, 41, 42).

Force field selection is a key factor in the outcome of protein folding simulations. Although force field modifications have led to improved agreements between MD simulations and experimental data, continued testing and comparison with experimental data is required to further these advances. While studies comparing different force fields have been conducted previously, very few of them had included such a large set of force fields with respect to folding of secondary structure elements (3, 14, 19, 43-45). Small proteins and peptides with folding times on the microsecond timescale are excellent systems to test and compare force fields; such trajectories provide reasonable sampling of conformations and sufficient length to examine the stability of the force field.

In this work, we compare MD simulations of a β-hairpin forming peptide derived from the protein Nrf2, performed with ten force fields. We assess their agreements with experimental data. The effects of elevated temperatures, charge-groups (46, 47), peptide capping and phosphorylation of Thr-80 with respect to β-hairpin formation were also examined. Despite using identical starting structures and simulation parameters, we observed clear differences amongst the various force fields and even between replicate simulations using the same force field. Such a comprehensive comparison will offer useful guidance to others conducting similar types of simulations and for improving force fields.

2.3 Simulation methodology

The starting structure for our MD simulations was generated based upon the amino acid sequence of a 16-mer human Nrf2 peptide ([72]AQLQLDEETGEFLPIQ[87]). We used the Crystallography & NMR System (CNS) (48) to generate an extended structure, which subsequently underwent simulated annealing. To avoid any potential bias to the bound-state conformation, a structure from the annealing simulations that did not resemble the bound state structure (PDB id: 2FLU) was chosen as the starting structure

(31). The exact same starting structure was used for all simulations. For the phosphorylated peptide (pThr-80) simulations, a dianionic phosphate group (PO_4^{2-}) was modeled onto residue Thr-80 of the same structure using chimera (49).

Force fields

We compared the peptide folding using the following force fields: Amber ff99SB-ILDN (15, 19, 20), Amber ff99SB*-ILDN (15, 17, 19, 20), Amber ff99SB (15, 19), Amber ff99SB* (15, 17, 19), Amber ff03 (15, 16), Amber ff03* (15-17), GROMOS96 43a1p (50, 51), GROMOS96 53a6 (21, 22), CHARMM27 (version c32b1) with CMAP (18, 52, 53) and OPLS-AA/L force fields (54-56). The '*' designations on the Amber force fields indicate the presence of a modification to the backbone dihedral potentials to improve agreement with experimental data (17). The 'ILDN' designation indicates the presence of a modification to the side-chain torsion potentials of isoleucine, leucine, aspartate and asparagine to improve agreement with quantum-mechanical calculations (20). Combination of the 'ILDN' and ff99SB* modifications has been demonstrated recently (44, 57). For a recent summary of the evolution of the Amber ff99 and ff03 series of force fields see the results section of (44) The 'p' designation on the GROMOS96 43a1 force field indicates the inclusion of phosphorylated amino acid parameters to the otherwise unmodified 43a1 parameters (50). One major difference between the GROMOS force fields and the others used in this study is that the GROMOS force fields are united atom and do not explicitly have all hydrogen atoms. The 'AA' and '/L' designations on the OPLS force field indicate all-atom and the inclusion of updated dihedral parameters from the original distribution (56).

Simulations of the same peptide with residue Thr-80 phosphorylated (pThr-80) were conducted with several of the above force fields in which phosphothreonine parameters were available. These included Amber ff99SB-ILDN, GROMOS96 43a1p and CHARMM27. Phosphothreonine parameters from (58) and (50) were used for the Amber ff99SB-ILDN and GROMOS96 43a1p force field simulations, respectively.

Phosphothreonine parameters included in the CHARMM27 force field distribution were used (18, 53).

Simulation details

A. *General parameters*. Simulations were performed using GROMACS (GROningen MAchine for Chemical Simulations) version 4.5 (47) Although GROMACS was used in this work, we expect that our findings will be applicable to other simulation software that utilizes the same force fields (59). Cubic boxes of linear size 6 nm were used and periodic boundary conditions were applied in all directions. Sodium (Na$^+$) and chloride (Cl$^-$) ions were added to neutralize the system and bring the salt concentration to 0.1 M. Na$^+$ and Cl$^-$ parameters specific to each force field distribution were used (60). Protein and non-protein atoms were coupled to their own temperature baths, which were kept constant at 310 K using the Parrinello-Bussi algorithm (61). This approach has been shown to perform very well in biomolecular simulations (46). Pressure was maintained isotropically at 1 bar using the Parrinello-Rahman barostat (62). A 2-fs timestep was employed. Prior to the production runs, the energy of each system was minimized using the steepest descents algorithm. This was followed by 2 ps of position-restrained dynamics with all non-hydrogen atoms restrained with a 1000 kJ mol^{-1} force constant. Initial atom velocities were taken from a Maxwellian distribution at 310 K. All bond lengths were constrained using the LINCS algorithm (63). A 1.0 nm cut-off was used for Lennard-Jones interactions. Dispersion corrections for energy and pressure were applied. Electrostatic interactions were calculated using the Particle-Mesh Ewald (PME) method (64) with 0.12 nm grid-spacing and a 1.0 nm real-space cut-off. No reaction-field or cutoff methods were tested as they have previously been shown be inferior to PME (65, 66). System coordinates were written out at 4 ps intervals during the production runs.

B. *System-specific attributes*. The protonation states of all ionizable residues were chosen based on their most probable state at pH 7. Unless specified, simulations were conducted with the amino and carboxyl terminals of the peptide left uncapped (NH$_3^+$ and COO$^-$, respectively) for each force field. When studying peptides from the interior of a protein sequence, it is common to add capping groups to the ends. This neutralizes the

unphysical charges introduced by the free N- and C-termini, which can potentially disrupt the native structure. To study the effects of peptide capping, several simulations with the N- and C-terminus capped with acetyl (ACE) and NH_2 groups, respectively, were performed (Table 2.1). The starting structure was solvated in SPC (simple point charge), TIP3P or TIP4P (67, 68) water. The compatibility of these water models with ions has been examined in detail by (60). A three-point water model (SPC or TIP3P) was recommended by GROMACS for all of the force fields used in this study, with the exception of OPLS-AA/L, in which the four-point (TIP4P) water model was the recommended choice (Table 2.1). Simulations with TIP3P and TIP4P were conducted for this force field (Table 2.1). The non-phosphorylated peptide systems each contained 17 Na^+ and 13 Cl^- ions, while for the pThr-80 systems two extra Na^+ ions were added to neutralize the dianionic phosphate group. For each force field, a simulation was conducted without the use of charge-groups (single atom charge groups); GROMACS uses the concept of charge groups to speed up simulations, see section "Domain Decomposition" in (47) for details. It has recently been shown that in some situations charge-groups can lead to pronounced unphysical effects (46). To examine the effect of charge-groups, additional simulations were conducted with the GROMOS96 and OPLS-AA/L force fields employing the default charge-groups for these force fields. Simulations performed with charge-groups are denoted with brackets around the force field name in the results section. For simulations conducted with the CHARMM27 force field, CMAP correction was applied (18). A few of the simulations were duplicated in order to assess reproducibility (Table 2.1). These systems did not use charge-groups, were prepared in the same manner as stated above and were assigned different initial atom velocities than their originals. Duplicated simulations are denoted with bracketed sequential numbering beside the force field name in the results section. We also performed elevated temperature simulations at 330, 350 and 370K with the Amber ff99SB*-ILDN (15, 17, 19, 20), Amber ff03*, (15-17) GROMOS96 53a6, (21, 22) CHARMM27 with CMAP (18, 52, 53) and OPLS-AA/L force fields (54-56). Using the initial and final (after 1 μs) system configurations at 310K, we reassigned the atom velocities at each higher temperature and performed 0.2 μs MD simulations.

In total, 28 individual 1 μs simulations were conducted and two of these trajectories were extended to 2 μs (Amber ff99SB* and OPLS-AA/L). An additional 7.2 μs of simulations at elevated temperature were performed. The cumulative simulation time was 37.2 μs. The simulations are summarized in Table 2.1.

Table 2.1 Summary of the MD simulations

Force field	Water	Elevated temp[a]	Capped[b]	pThr-80[c]	Charge-groups[d]	Duplicate[e]
Amber ff99SB-ILDN	TIP3P (7038)					Y
	TIP3P (7030)		Y			Y
	TIP3P (7036)			Y		
Amber ff99SB*-ILDN	TIP3P (7038)	Y				
Amber ff99SB	TIP3P (7038)					Y
Amber ff99SB*[f]	TIP3P (7038)					Y
Amber ff03	TIP3P (7038)					
Amber ff03*	TIP3P (7038)	Y				
GROMOS96 43a1p	SPC (7035)				Y	
	SPC (7030)		Y			
	SPC (7033)			Y	Y	
	SPC (7027)		Y	Y		
GROMOS96 53a6	SPC (7035)	Y			Y	Y
	SPC (7033)	Y	Y			
CHARMM27	TIP3P (7038)	Y				Y
	TIP3P (7030)			Y		
OPLS-AA/L[f]	TIP3P (7037)	Y			Y	
	TIP4P (6969)					

[a] 'Y' indicates that elevated temperature simulations were performed at 330, 350 and 370 K from the initial and final (after 1μs) system configurations.

[b] 'Y' indicates that the N and C termini of the peptide was capped with with acetyl (ACE) and NH_2 groups, respectively. They were otherwise left uncapped (NH_3^+ and COO^-, respectively).

[c] 'Y' indicates that residue Thr-80 was phosphorylated.

[d] 'Y' indicates that two simulations were performed: one with default GROMACS charge-groups and one without charge-groups.

[e] 'Y' indicates that two simulations, each 1 μs, were performed. Duplicates were always performed without charge-groups and were identical to the first simulation except for their initial atom velocities.

[f] The trajectory was extended to 2 μs.

Simulation analysis

We used either the full 1 μs trajectories or the last 0.1 μs for analysis. By restricting some analyses to the last 0.1 μs, we allowed as much time as possible for the simulations to converge to a stable conformation. Hydrogen bonds were analyzed as follows: A hydrogen bond between a donor (D-H) and an acceptor (A) was considered to be formed when the DA distance was less than 3.2 Å and the angle between the DA vector and the D-H bond (AD-H angle) was less than 35°. These geometric criteria for defining hydrogen bonds are consistent with those used in prior studies (69, 70). Secondary structure content was assessed with the DSSP algorithm (71).

2.4 Results

We have compared the secondary structures and free- and bound-state contact formations in MD simulations of a β-hairpin forming peptide derived from the intrinsically disordered Neh2 domain of Nrf2 conducted with 10 different biomolecular force fields. The DSSP algorithm was used to monitor the evolution of secondary structures over the entire 1 μs trajectories. β-hairpin formation was identified by inspection of the cluster center structures and the C_α-C_α atom pair distances during the last 0.1 μs. Resemblance to the native state was gauged via the presence or absence of experimentally determined ^1H, ^1H NOEs (30). We have also compared hydrogen bonds,

RMSDs and backbone dihedral angles in the MD structures to the peptide in complex with its binding target, Keap1 (31). The effects of elevated temperatures, terminal capping, charge-groups and phosphorylation of Thr-80 on hairpin folding were also assessed.

Assessing secondary structure formation at 310K

To compare the MD trajectories obtained with different force fields, we first assessed the secondary structure content over the course of our simulations at 310K using the program DSSP (71). In this analysis, we deemed simulations that had residues from their β-turn regions (^{77}DEET^{80}) in "turn" conformations (yellow) flanked by residues in β-sheet conformations (red) simultaneously, to have formed β-hairpins (Figure 2.1). For the uncapped peptides, the Amber ff99SB-ILDN (1), Amber ff99SB*-ILDN, Amber ff99SB (2), Amber ff99SB* (2), Amber ff03, Amber ff03*, GROMOS96 43a1p and GROMOS96 53a6 (1 & 2) simulations, including those which used charge-groups, appeared to adopt β-hairpins at some points in their trajectories (Figure 2.1A). Cluster center structures from the last 0.1 μs of the simulations, with the potential β-turn region (^{77}DEET^{80}) colored in black, are shown in Figure 2.2. The result clearly illustrates the 16-mer peptide folds into β-hairpin conformations by using these force fields. Intriguingly, the folding times vary greatly between ~0.05 μs and > 0.9 μs in these simulations (Figure 2.1A). The CHARMM27 simulations did not form hairpins and DSSP plots showed that helical content was present in the peptide (Figures 2.1A & 2.2A). None of the OPLS-AA/L simulations met our criteria for hairpin formation. Instead, these simulations were enriched in "bend" conformations (green) (Figure 2.1A). While there was some transient turn and β-sheet content in the expected locations, there were no pronounced β-hairpin signatures (Figures 2.1A & 2.2A). Extending of the OPLS-AA/L trajectory (without charge groups and TIP3P water) to 2 μs still did not yield a native-like hairpin (data not shown).

A
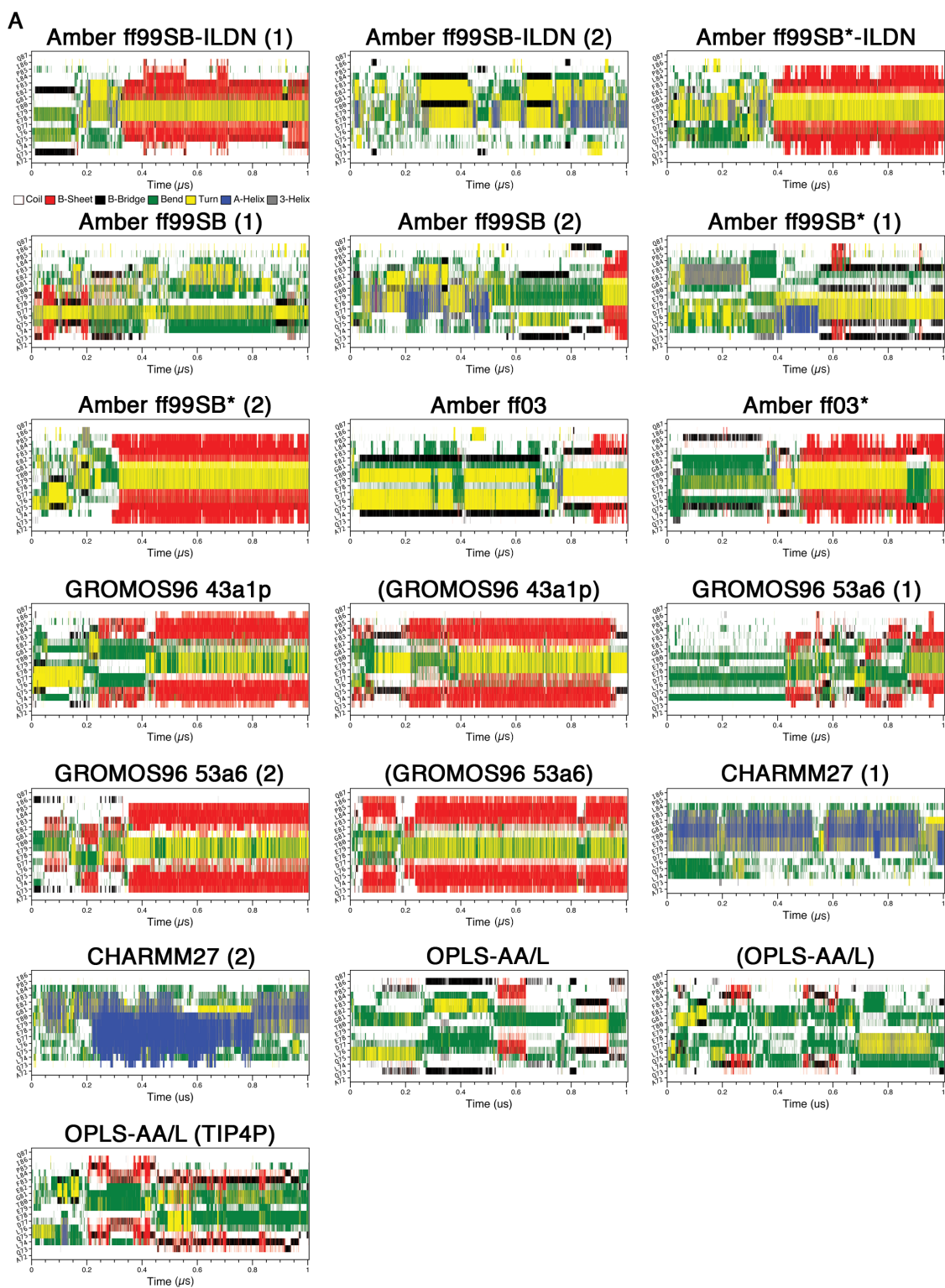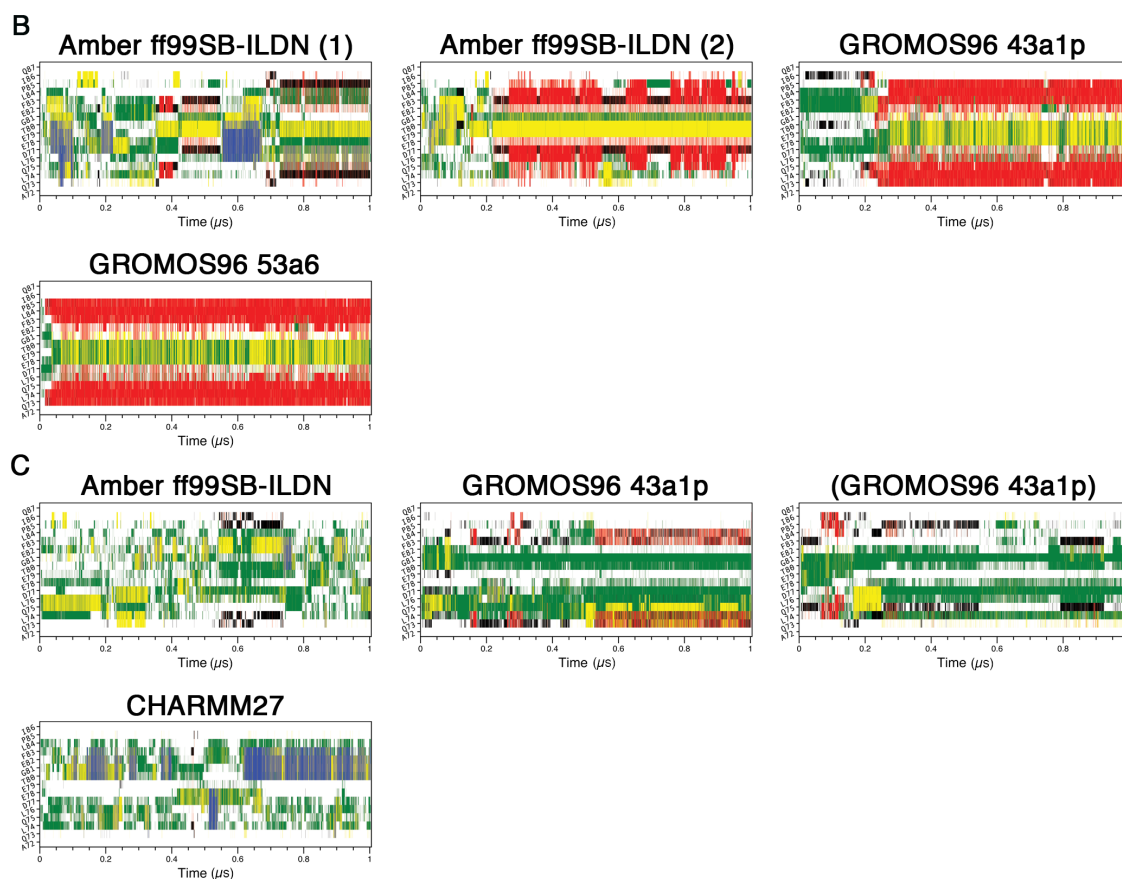
Figure 2.1 Secondary structure propensity analysis of the trajectories.

Secondary structure content was assessed with the DSSP algorithm (71): coil (white), β-sheet (red), β-bridge (black), bend (green), turn (yellow), α-helix (blue) and 3_{10} helix (grey). A) Uncapped peptide. B) Capped peptide. C) pThr-80 peptide.

A

Amber ff99SB-ILDN (1) Amber ff99SB-ILDN (2) Amber ff99SB*-ILDN

Amber ff99SB (1) Amber ff99SB (2) Amber ff99SB* (1)

Amber ff99SB* (2) Amber ff03 Amber ff03*

GROMOS96 43a1p (GROMOS96 43a1p) GROMOS96 53a6 (1)

GROMOS96 53a6 (2) (GROMOS96 53a6) CHARMM27 (1)

CHARMM27 (2) OPLS-AA/L (OPLS-AA/L)

OPLS-AA/L (TIP4P)
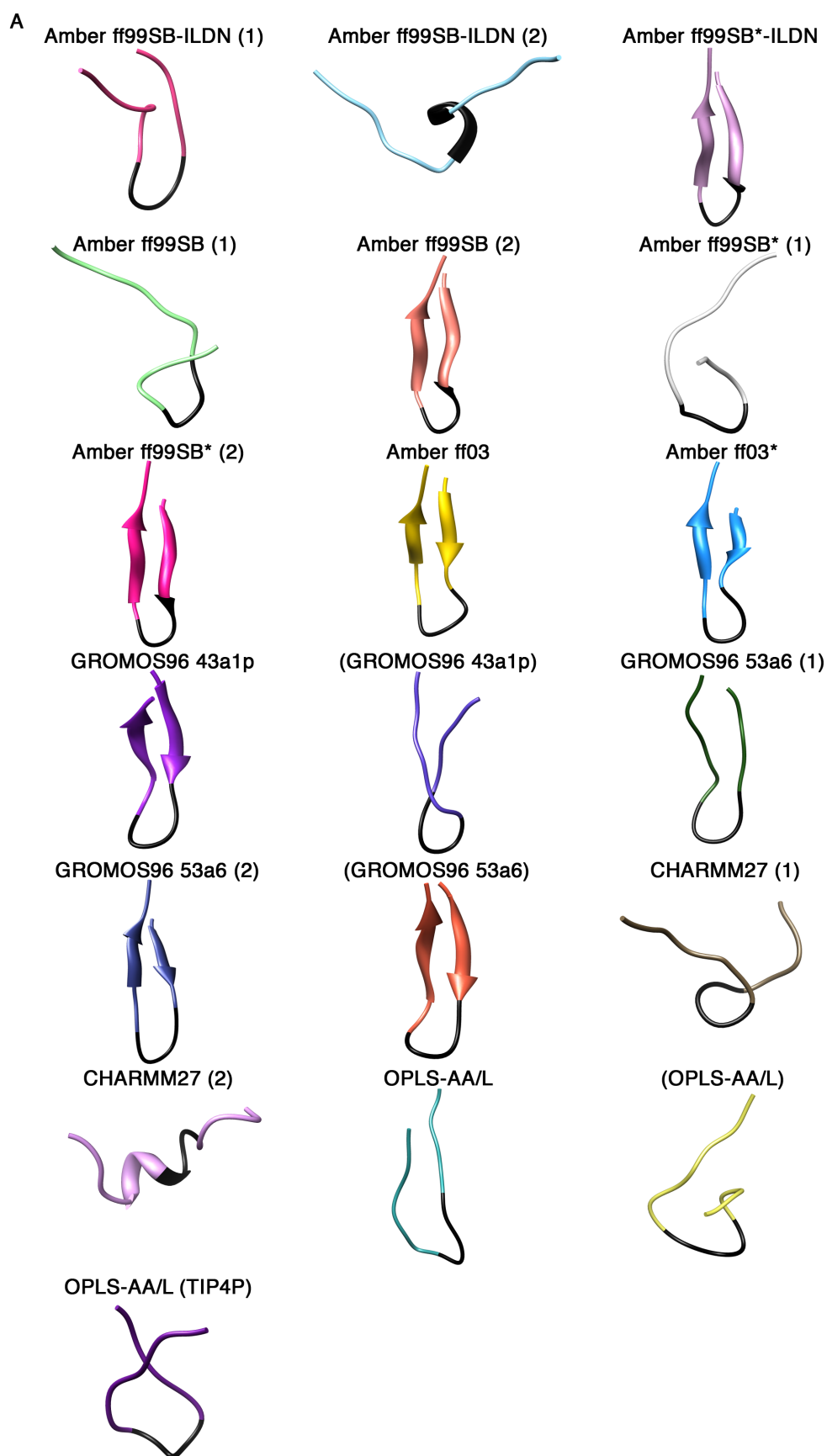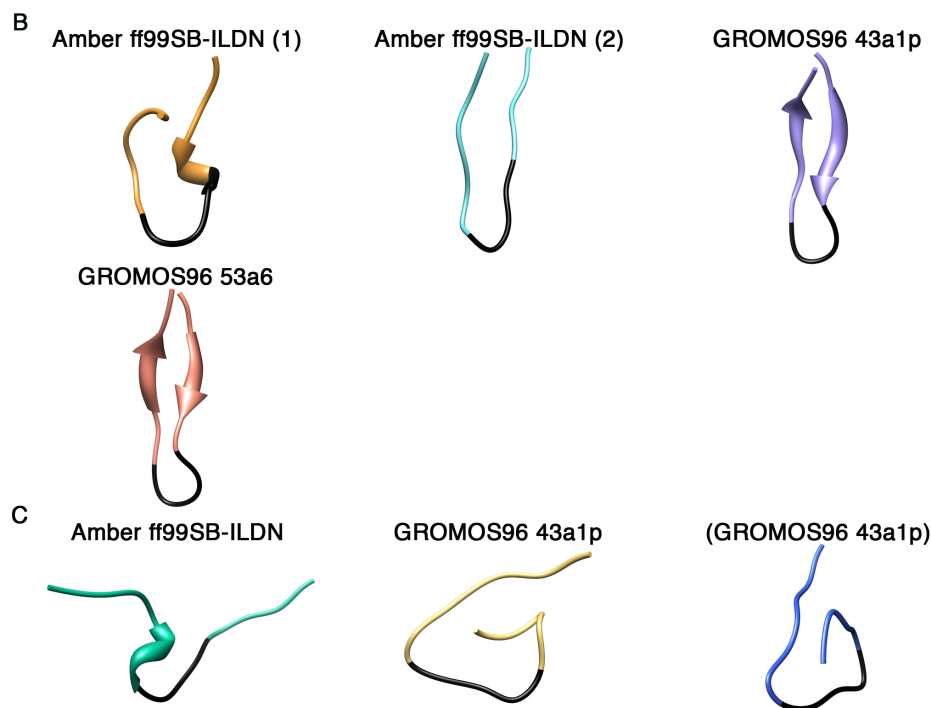
Figure 2.2 Cluster centroid structures from the last 0.1 μs of the simulations.

A single cluster represented all structures in each simulation and the center
structure was extracted. A) Uncapped peptide. B) Capped peptide. C) pThr-80
peptide.

There were clear differences between replicate runs using the same force fields.
Specifically, the Amber ff99SB-ILDN (2), Amber ff99SB (1) and Amber ff99SB* (1)
simulations did not converge upon β-hairpin conformations (Figures 2.1A & 2.2A). To
determine if a longer trajectory would lead to hairpin formation, we extended the Amber
ff99SB* (1) simulation to 2 μs. DSSP analysis, however, still was not indicative of a
hairpin structure (data not shown).

For the capped peptides, while the Amber ff99SB-ILDN (2), GROMOS96 43a1p
and GROMOS96 53a6 simulations yielded hairpin signatures throughout large parts of
the trajectories (Figure 2.1B), only the GROMOS96 force fields led to the formation of
well-defined β-hairpins in the last 0.1 μs of the simulations (Figure 2.2B). Again, there

were differences between the two Amber ff99SB-ILDN replicates. While the Amber ff99SB-ILDN (1) simulation did have turn and strand contents in the expected region, it appeared to be transient and not as pronounced as the Amber ff99SB-ILDN (2) hairpin signature (Figure 2.1B). Figure 2.2B shows that close to the end of the trajectory, Amber ff99SB-ILDN (1) structure adopted a short non-native helix before the turn region while hairpin structure that is slightly displaced from the expected location was observed in the Amber ff99SB-ILDN (2) trajectory.

The phosphorylation of Thr-80 located at the turn region appears to have significant effects on the peptide folding. β-hairpin formation was not observed in any of the pThr-80 simulations (Figures 2.1C & 2.2C). Interestingly, these simulations all displayed considerable bend content but failed to form a turn in the expected location (Figure 2.1C).

The averaged C_α-C_α atom pair distances (within 10 Å) were also plotted to identify β-hairpin formation in the simulations (Figure 2.3). In these plots, the β-turn (^{77}DEET^{80}) region, which the hairpin is approximately centered around, was indicated. For the uncapped peptides, the Amber ff99SB-ILDN (1), Amber ff99SB*-ILDN, Amber ff99SB (2), Amber ff99SB* (2), Amber ff03, Amber ff03*, GROMOS96 43a1p and GROMOS96 53a6 (1 & 2) simulations, including those which used charge-groups, appeared to form β-hairpins as evidenced by the cross-strand C_α-C_α contacts centered around the β-turn (Figure 3.3A). Like the DSSP plots, this analysis also revealed clear differences between the replicates of Amber ff99SB-ILDN, Amber ff99SB, and Amber ff99SB* simulations. While Amber ff99SB-ILDN (2) displayed no signature of β-hairpin structure, the hairpins formed in the Amber ff99SB (1) and Amber ff99SB* (1) simulations were found in different regions compared to the replicas (Figure 2.3A). The CHARMM27 simulations did not have cross-strand C_α-C_α contacts indicative of β-hairpin structures, but showed regions of compactness in the turn segment (Figure 2.3A). The OPLS-AA/L simulations without charge groups had some evident cross-strand contacts, but the β-turn was shifted from the expected location (Figure 2.3A); while the OPLS-AA/L simulation with default charge-groups did not appear to form a hairpin (Figure 2.3A). C_α-C_α contacts in the capped peptide simulations were indicative of hairpin structures. However, unlike the conformations observed with the GROMOS force

fields, the β-hairpin in the Amber ff99SB-ILDN simulations were shifted from the expected location (Figure 2.3B). None of the pThr-80 simulations had cross-strand contacts evident of β-hairpin structures (Figure 2.3C). It is worthwhile to note that in both GROMOS96 43a1p trajectories with Thr-80 phosphorylated there was evidence of close contacts between the positively charged N-terminus and the negatively charged phosphate group (Figure 2.3C). Since this may represent an unphysical interaction, we performed an additional simulation with a capped version of the peptide (Table 2.1). In this trajectory we still did not observe hairpin or turn formation and noticed similar close contacts between the N-terminal region and phosphate group (data not shown).

A
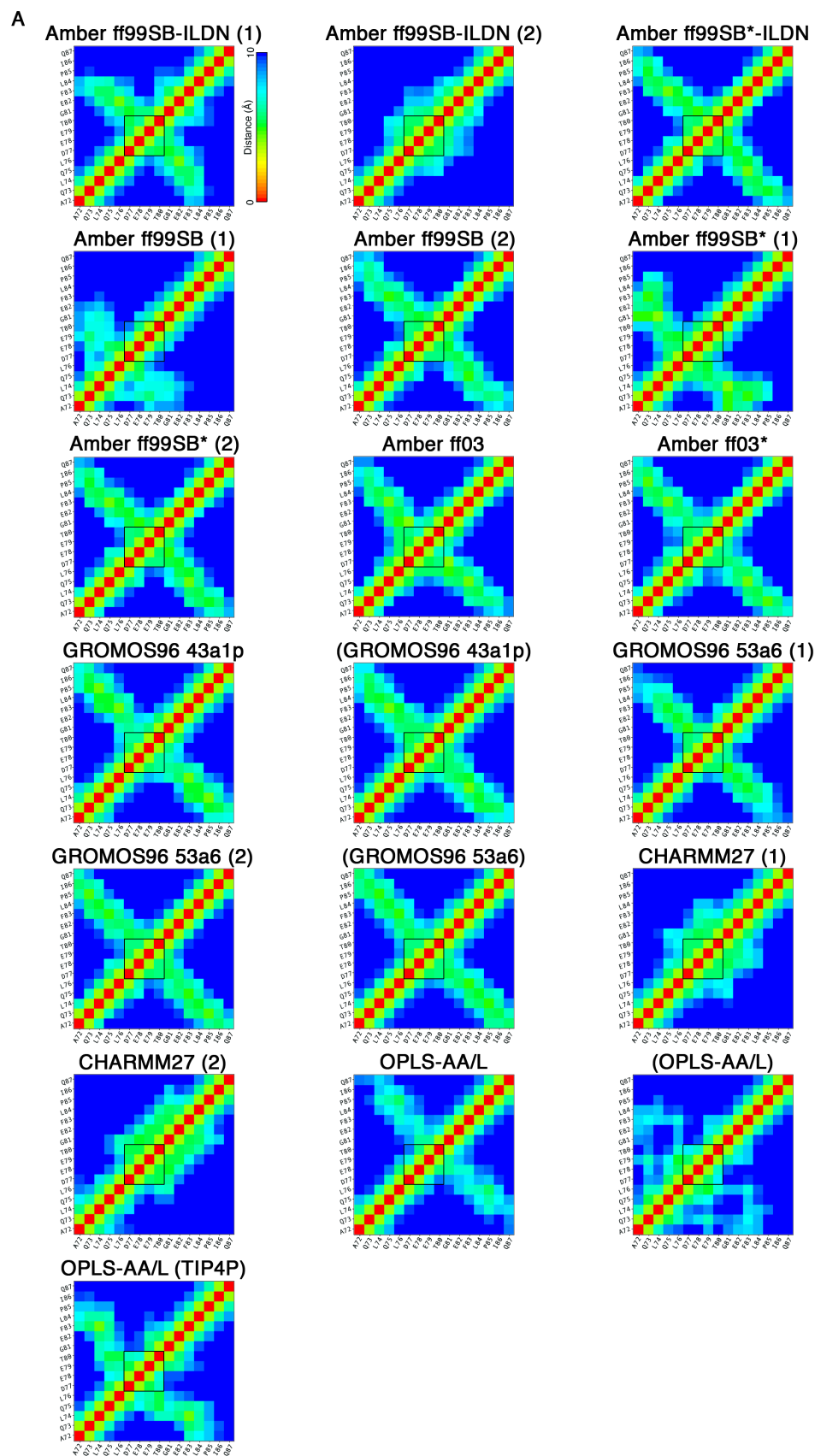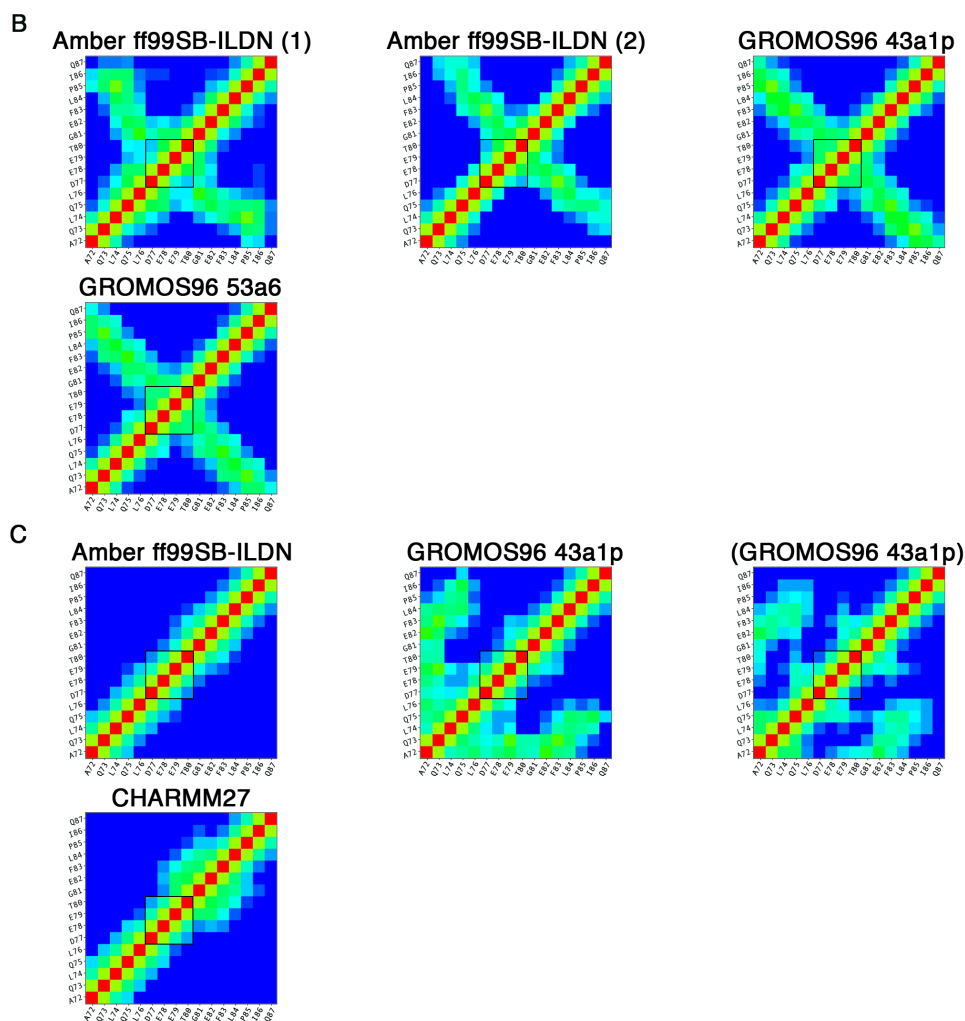
Figure 2.3 C_α-C_α atom pair distances.

Average C_α-C_α distances less than or equal to 10 Å during the last 0.1 μs of the MD simulations. Distances equal to or greater than 10 Å are colored blue. The black square indicates the β-turn ([77]DEET[80]) region. A) Uncapped peptide. B) Capped peptide. C) pThr-80 peptide.

Comparison to experimental data

Next, we compared the results of our simulations to experimental data. To begin, we assessed how many experimentally determined atomic contacts within the Nrf2 β-

hairpin were found in our simulations. Even though the free state structure of the 16-mer human Nrf2 peptide used in this study is not currently available, several atomic contacts within the β-hairpin region of mouse Nrf2 have been determined by NMR (30). The mouse Nrf2 contains the same β-hairpin sequence as the human version used in this study, except with a single conservative amino acid change of L74F (Figure 2.4A). Given that the human and mouse Nrf2 β-hairpin sequences are nearly identical, they are expected to adopt similar structures.

A

```
sp|Q16236|NF2L2_HUMAN    AQLQLDEETGEFLPIQ
sp|Q60795|NF2L2_MOUSE    AQFQLDEETGEFLPIQ
                         **:************
```

B

Structure labels: L84, 1, Q75, 2, F83, L76, E82, 3, D77, G81, T80, E78, E79

C

	Uncapped			Capped			pThr-80		
	1	2	3	1	2	3	1	2	3
Amber ff99SB-ILDN (1)	●	●	●	●	●	●	○	○	○
Amber ff99SB-ILDN (2)	○	○	●	○	●	○			
Amber ff99SB*-ILDN	●	●	●						
Amber ff99SB (1)	○	○	○						
Amber ff99SB (2)	●	●	●						
Amber ff99SB* (1)	○	○	○						
Amber ff99SB* (2)	●	●	●						
Amber ff03	●	●	●						
Amber ff03*	●	●	●						
GROMOS96 43a1p	●	●	●	●	●	●	●	●	○
(GROMOS96 43a1p)	●	●	●				●	●	○
GROMOS96 53a6 (1)	●	○	●	●	●	●			
GROMOS96 53a6 (2)	●	●	●						
(GROMOS96 53a6)	●	●	●						
CHARMM27 (1)	○	○	●				○	○	○
CHARMM27 (2)	○	○	●						
OPLS-AA/L	○	●	○						
(OPLS-AA/L)	●	○	○						
OPLS-AA/L (TIP4P)	●	○	○						

Figure 2.4 Nrf2 β-hairpin sequence alignment and native contacts.

A) Sequence alignments between the human and mouse Nrf2 β-hairpin segments were generated with ClustalW XXL (80). The Blosum scoring matrix (81) was used and gap penalties were set at their default values. Opening and end gap penalties were set to 10. Extending and separation gap penalties were set to 0.05. B) Three NMR-derived cross-strand ^{1}H, ^{1}H NOEs determined by (30) mapped onto a model Nrf2 β-hairpin backbone structure. C) Presence or absence of each native contact for each force field. Time-averaged distances < 6 Å during the last

0.1 μs of the simulations between hydrogen atom pairs matching those observed by Tong et al. (30) were considered to be native contacts.

We compared the NMR-derived cross-strand ^{1}H, ^{1}H NOEs determined by (30) to the corresponding time-averaged distances from our MD simulations. Time-averaged distances < 6 Å between hydrogen atom pairs matching those observed in (30) were considered to be native contacts. Because the united-atom GROMOS96 force fields used in this study do not explicitly represent every hydrogen atom, we restricted our analysis to backbone amide hydrogens, which were explicitly represented in all force fields. NOEs between adjacent residues and those involving F74, were excluded from the analysis. This reduced the number of experimentally determined native contacts used in this analysis to three (Q75 HN:L84 HN, L76 HN:L84 HN and D77 HN:E82 HN). They are depicted in Figure 2.4B.

The presence or absence of each of the three contacts is shown in Figure 2.4C. For the uncapped peptides, the Amber ff99SB-ILDN (1), Amber ff99SB*-ILDN, Amber ff99SB (2), Amber ff99SB* (2), Amber ff03, Amber ff03*, GROMOS96 43a1p and GROMOS96 53a6 simulations, including those which used charge-groups, had at least 2 of the 3 native contacts (Figure 2.4C). Once again, there were differences between the Amber replicas (Figure 2.4C). Notably, in the Amber ff99SB-ILDN (2), Amber ff99SB (1) and Amber ff99SB* (1) simulations, only one or none of the native contacts were present, while their replicas had all three (Figure 2.4C). The CHARMM27 and OPLS-AA/L simulations had only 1 out of the 3 native contacts (Figure 2.4C). The capped peptides were able to form all 3 native contacts, but differences between duplicates were also evident. The Amber ff99SB-ILDN (1) simulation had all 3 contacts while its duplicate had only 1 (Figure 2.4C). Native contacts were reduced in all pThr-80 simulations compared to their unphosphorylated counterparts (Figure 2.4C). Interestingly, 2 of the 3 native contacts were still present in the GROMOS96 43a1p pThr-80 trajectories (Figure 2.4C). In these simulations, while the two contacts in the β-sheet

region of the hairpin were present, the contact in the turn region was missing (Figure 2.4C).

In addition to NMR-derived contacts, backbone and side chain hydrogen bonds between Asp-77 and Thr-80 are present when Nrf2 is bound to Keap1 (PDB id: 2FLU) (31). We previously found that hydrogen bonds between these residues may also exist in the free state with high frequencies in simulations conducted with the GROMOS96 53a6 force field (9). Because hydrogen bonding between Asp-77 and Thr-80 may be correlated with β-hairpin formation, we calculated the frequencies of hydrogen bonding between these residues (Table 2.2). For the uncapped peptides, we observed high (> 0.68) frequencies of Asp-77 to Thr-80 hydrogen bonding in the Amber ff99SB-ILDN (1), Amber ff99SB*-ILDN, Amber ff99SB (2), Amber ff99SB* (2), Amber ff03*, GROMOS96 43a1p and GROMOS96 53a6 (1 & 2) simulations, including those which used charge-groups (Table 2.2). Like the aforementioned analyses, clear differences between some replicas were observed (Table 2.2). Specifically, the Amber ff99SB-ILDN (2), Amber ff99SB (1) and Amber ff99SB* (1) simulations had considerably less Asp-77 to Thr-80 hydrogen bonding compared to their duplicates (Table 2.2).

Table 2.2 Frequency of Asp-77 to Thr-80 hydrogen bonding[a]

Force field	Uncapped[b]	Capped[c]	pThr-80[d]
Amber ff99SB-ILDN (1)	0.95	0.00	0.00
Amber ff99SB-ILDN (2)	0.27	0.00	
Amber ff99SB*-ILDN	0.94		
Amber ff99SB (1)	0.00		
Amber ff99SB (2)	0.97		
Amber ff99SB* (1)	0.00		
Amber ff99SB* (2)	0.97		
Amber ff03	0.00		
Amber ff03*	0.86		
GROMOS96 43a1p	0.69	0.91	0.00
(GROMOS96 43a1p)[e]	0.94		0.00
GROMOS96 53a6 (1)	0.90	0.92	
GROMOS96 53a6 (2)	0.93		
(GROMOS96 53a6)	0.91		
CHARMM27	0.32		0.00
CHARMM36	0.22		
OPLS-AA/L	0.00		

(OPLS-AA/L)   0.00

OPLS-AA/L (TIP4P)   0.00

[a] Frequencies of 1 or more hydrogen bonds during the last 0.1 μs of the trajectories. Oxygen and nitrogen atoms were acceptors. Amine groups and the hydroxyl group of Thr-80 were donors. Intra-residue hydrogen bonds were excluded from the analysis. A hydrogen bond between a hydrogen donor (D-H) and a hydrogen acceptor (A) was judged to be formed when the DA distance (r) was less than 3.2 Å and the angle between the DA vector and the D-H bond (AD-H angle) was less than 35°.

[b] Values for the peptides with unmodified N and C termini (NH_3^+ and COO^-, respectively).

[c] Values for the peptides with capped N and C termini (ACE and NH_2, respectively).

[d] Values for the peptides with residue Thr-80 phosphorylated.

[e] Bracketed values indicate hydrogen bond frequencies for trajectories with default GROMACS charge groups.

It is noteworthy that in the Amber ff03 simulation, no hydrogen bonding between Asp-77 and Thr-80 was observed (Table 2.2). Because prior analysis showed that this trajectory formed a hairpin with 3 native contacts (Figure 2.4A), the complete lack of hydrogen bonding between these two residues was unexpected. Inspection of the trajectory showed that the side chains of Leu-76 and Asp-77 were on opposite sides of the hairpin compared to the other simulations (data not shown). Although Asp-77 was not in an appropriate orientation to form hydrogen bonds with Thr-80, we found that the frequency of forming 1 or more hydrogen bonds between Leu-76 and Thr-80 was 0.66 in this trajectory. Low frequencies of Asp-77 to Thr-80 hydrogen bonding was found in the CHARMM27 simulations (between 0.22 and 0.32) and was completely absent in the OPLS-AA/L simulations (Table 2.2).

For the capped peptides, no hydrogen bonding between Asp-77 and Thr-80 was observed in the Amber ff99SB-ILDN simulations, but was present in over 90% of the structures in the last 0.1 μs of the GROMOS96 trajectories (Table 2.2). While the capped Amber ff99SB-ILDN (1) simulation was found to have 3 native contacts (Figure 2.4C), cluster analysis indicated that there was a short non-native helix before its β-turn region (Figure 2.1B). Furthermore, the DSSP plot of this trajectory did not have a typical hairpin signature (Figure 2.2B). These factors likely contributed to the lack of Asp-77 to Thr-80 hydrogen bonding in this trajectory (Table 2.2). The capped Amber ff99SB-ILDN (2)

simulation had only 1 native contact (Figure 2.4C) and its β-turn was in a slightly displaced from the expected location (Figure 2.3B); factors that likely contributed to the lack of hydrogen bonding (Table 2.2). Hydrogen bonding between Asp-77 and Thr-80 was not observed in any of the pThr-80 simulations (Table 2.2).

We also compared the Nrf2 peptide structures from our simulations to that of the Keap1-bound state (PDB id: 2FLU) (31). This comparison is interesting because the ETGE motif of the disordered Nrf2 has been shown to have a tendency to form bound state-like structure even in the absence of the target (9, 30, 31). Since the Nrf2 β-hairpin does not adopt a well-defined in the free state (30), we restricted the RMSD calculations to backbone atoms only.

RMSDs were calculated separately for the β-turn, [77]DEET[80] and β-hairpin, [72]AQLQLDEETGEFL[84] regions. The RMSDs throughout the trajectories are plotted in Figure 2.5 and average RMSD values over the last 0.1 μs are shown in Figure 2.6 and summarized in Tables 2.3-2.5. For the uncapped peptides, the Amber ff99SB-ILDN (1), Amber ff99SB*-ILDN, Amber ff99SB (2), Amber ff99SB* (2), Amber ff03, Amber ff03*, GROMOS96 43a1p and GROMOS96 53a6 (1 & 2) simulations achieved average RMSDs < 1 and < 3 Å to the bound state β-turn and hairpin, respectively, including simulations which used charge-groups (Figure 2.6A). Again, there were some differences between replicas. The Amber ff99SB-ILDN (2) simulation had a β-turn region with an average RMSD < 1 Å, but when considering the full β-hairpin, the RMSD was larger than 4.8 Å (Figure 2.6A). Also, the Amber ff99SB (1) and Amber ff99SB* (1) simulations had substantially higher RMSDs compared to their duplicates (Figures 2.5A and 2.6A). The CHARMM simulations did not lead to bound state like β-hairpin (RMSDs > 4 Å), but had β-turn RMSDs below 1 Å (Figures 2.5A and 2.6A). The OPLS-AA/L simulations had both β-turn and hairpin RMSDs greater than 1 Å and 3 Å, respectively (Figures 2.5A and 2.6A), indicating significant deviations from the bound-state structure.

A
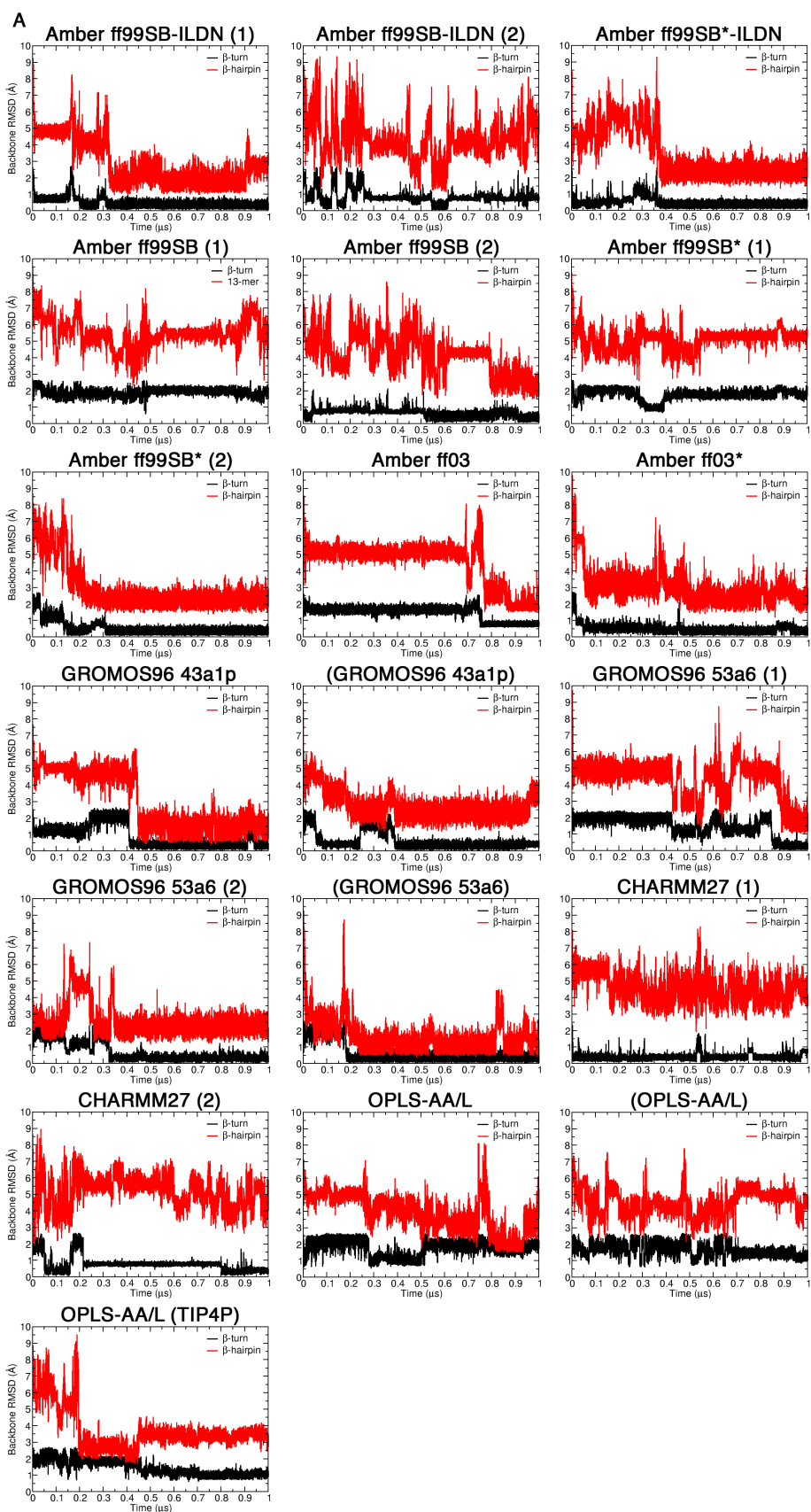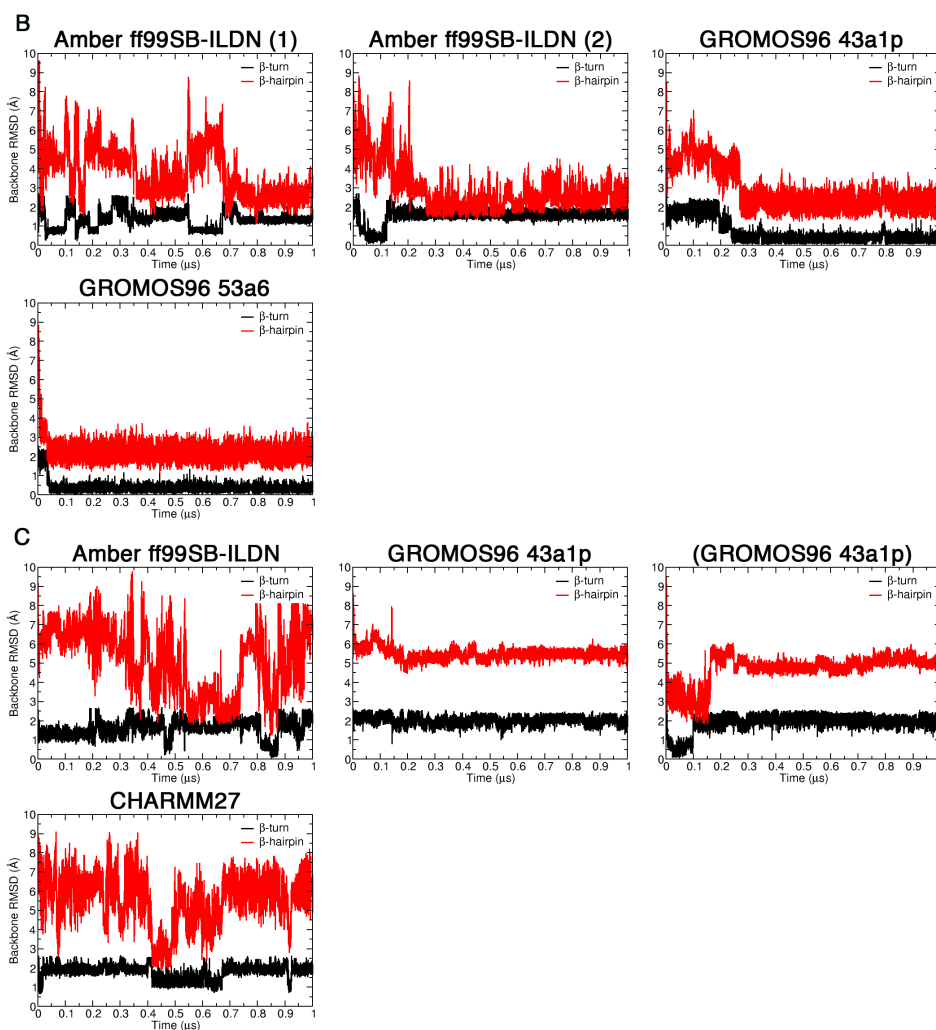
Figure 2.5 Backbone RMSDs between the bound state and MD structures throughout the trajectories.

RMSD values were calculated for the β-turn 4-mer, [77]DEET[80] (black) and β-hairpin 13-mer, [72]AQLQLDEETGEFL[84] (red) by least squares fitting the backbone atoms (N, C_α and C) from each frame to the corresponding atoms of bound state reference structure (PDB id: 2FLU) (31). A) Uncapped peptide. B) Capped peptide. C) pThr-80 peptide.

Figure 2.6 Average backbone RMSDs between the bound state and MD structures.

Average RMSD values were calculated over the last 0.1 μs of the simulations for the β-turn 4-mer, ^{77}DEET^{80} (black) and β-hairpin 13-mer, ^{72}AQLQLDEETGEFL^{84} (red) by least squares fitting the backbone atoms (N, C$_\alpha$ and C) from each frame to the corresponding atoms of bound state reference structure (PDB id: 2FLU) (31). A) Uncapped peptide. B) Capped peptide. C) pThr-80 peptide.

Table 2.3 Average backbone RMSDs between the bound state structure and MD structures of the uncapped peptides[a]

Force field	Backbone[b] (Å) ± sdev β-turn[c]	Backbone (Å) ± sdev β-hairpin[d]
Amber ff99SB-ILDN (1)	0.30 ± 0.10	2.74 ± 0.45
Amber ff99SB-ILDN (2)	0.77 ± 0.25	4.86 ± 1.14
Amber ff99SB*-ILDN	0.35 ± 0.08	2.41 ± 0.28
Amber ff99SB (1)	1.81 ± 0.19	6.10 ± 0.80
Amber ff99SB (2)	0.38 ± 0.13	2.53 ± 0.42
Amber ff99SB* (1)	1.86 ± 0.16	5.33 ± 0.14
Amber ff99SB* (2)	0.35 ± 0.09	2.32 ± 0.30
Amber ff03	0.77 ± 0.06	1.98 ± 0.24
Amber ff03*	0.46 ± 0.16	2.56 ± 0.39
GROMOS96 43a1p	0.50 ± 0.33	1.56 ± 0.43
(GROMOS96 43a1p)[e]	0.39 ± 0.09	2.88 ± 0.57
GROMOS96 53a6 (1)	0.25 ± 0.12	1.89 ± 0.41
GROMOS96 53a6 (2)	0.30 ± 0.12	2.32 ± 0.26
(GROMOS96 53a6)	0.32 ± 0.12	1.42 ± 0.56
CHARMM27	0.45 ± 0.21	4.45 ± 0.57
CHARMM36	0.39 ± 0.07	4.80 ± 0.71
OPLS-AA/L	1.84 ± 0.21	2.88 ± 0.87
(OPLS-AA/L)	1.33 ± 0.19	4.63 ± 0.57
OPLS-AA/L (TIP4P)	1.08 ± 0.11	3.48 ± 0.21

[a] Average RMSDs were calculated over the last 0.1 μs of the trajectories.
[b] Backbone atoms include N, C_α and C.
[c] β-turn ([77]DEET[80]).
[d] β-hairpin ([72]AQLQLDEETGEFL[84]).
[e] Average RMSDs for trajectory with default GROMACS charge groups.

Table 2.4 Average RMSDs between the bound state conformation and MD structures of the capped peptides[a]

Force field	Backbone[b] (Å) ± sdev β-turn[c]	Backbone (Å) ± sdev β-hairpin[d]
Amber ff99SB-ILDN (1)	1.39 ± 0.13	2.67 ± 0.34
Amber ff99SB-ILDN (2)	1.66 ± 0.10	2.65 ± 0.38
GROMOS96 43a1p	0.32 ± 0.14	2.18 ± 0.31
GROMOS96 53a6	0.29 ± 0.11	2.23 ± 0.31

[a] Average RMSDs were calculated over the last 0.1 μs of the trajectories.
[b] Backbone atoms include N, C_α and C.
[c] β-turn ([77]DEET[80]).
[d] β-hairpin ([72]AQLQLDEETGEFL[84]).

Table 2.5 Average RMSDs between the bound state conformation and MD structures of the pThr-80 peptides[a]

Force field	Backbone[b] (Å) ± sdev β-turn[c]	Backbone (Å) ± sdev β-hairpin[d]
Amber ff99SB-ILDN	1.85 ± 0.45	5.96 ± 1.19
GROMOS96 43a1p	2.01 ± 0.18	5.45 ± 0.17
(GROMOS96 43a1p)[e]	2.00 ± 0.18	5.04 ± 0.17
CHARMM27	1.89 ± 0.34	5.88 ± 0.93

[a] Average RMSDs were calculated over the last 0.1 μs of the trajectories.
[b] Backbone atoms include N, C_α and C.
[c] β-turn ([77]DEET[80]).
[d] β-hairpin ([72]AQLQLDEETGEFL[84]).
[e] Average RMSDs for trajectory with default GROMACS charge groups.

For the capped peptides, both Amber ff99SB-ILDN simulations had average RMSDs < 3 Å for the hairpin region, but their β-turns had RMSDs > 1 Å (Figures 2.5B and 2.6B). In comparison, both GROMOS96 force fields had RMSDs of < 1 and < 3 Å for the β-turn and hairpin, respectively (Figures 2.5B and 2.6B). These RMSDs were similar to their uncapped versions (Figures 2.5A and 2.6A). It is worthwhile to note that the capped GROMOS96 53a6 simulation converged to bound state like structure in < 0.05 μs, the fastest of all the simulations (Figure 2.5B). Among the simulations which had bound state like RMSDs, the amount of time it took to adopt these conformations varied between < 0.05 and ~ 0.9 μs, even for duplicates using the same force field (Figures 2.5A and B). However, once a bound state like structure was formed, it tended to remain stable. The β-turn and hairpin RMSDs were higher in all pThr-80 simulations compared to those of the unphosphorylated peptides (Figures 2.5C and 2.6C).

The convergence of the dihedral angles from the trajectories to those from the bound state structure was also assessed (PDB id: 2FLU) (31). The combined φ and ψ angles from the simulations and bound state structure are shown in Figure 2.7 and the average per-residue deviations are shown in Figure 2.8. For the uncapped peptides, the GROMOS96 43a1p with charge groups and 53a6 force field simulations had the lowest φ and ψ deviations from the bound state structure (Figure 2.8A). These simulations had combined φ and ψ deviations of < 7° and < 17° per residue from the bound state in their β-turn and hairpin regions, respectively (Figure 2.8A). The Amber ff99SB-ILDN (1) and

CHARMM simulations had combined ϕ and ψ deviations of ~ 10º in their β-turns, but deviated > 20º per residue when considering the entire hairpin (Figure 2.8A). For the capped peptides, both GROMOS96 force fields had slightly lower deviations compared to their uncapped counterparts and had considerably lower deviations than Amber ff99SB-ILDN (Figure 2.8B). The β-turn and hairpin deviations were higher in all pThr-80 simulations compared to those of the unphosphorylated peptides (Figure 2.8B).

A
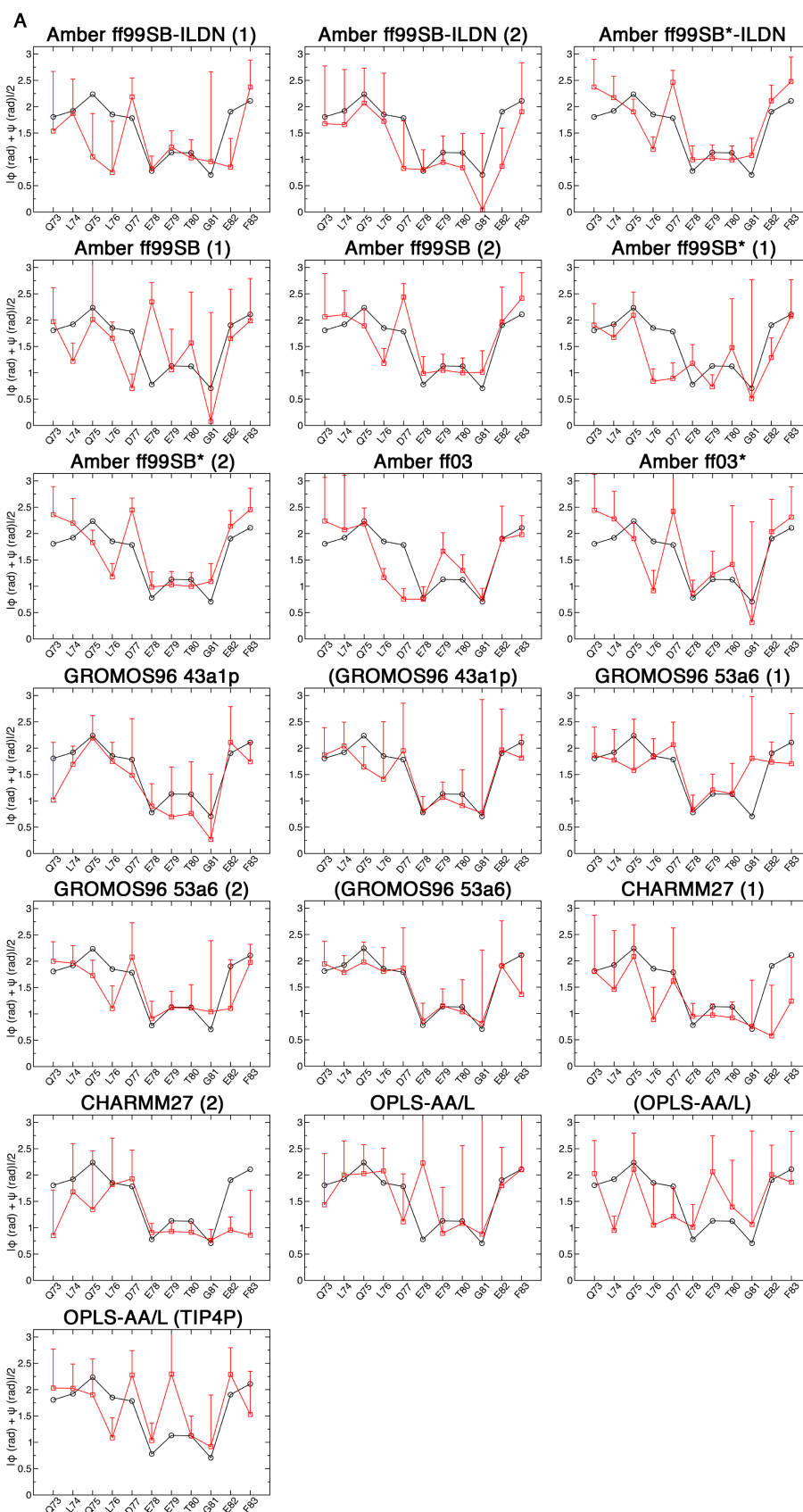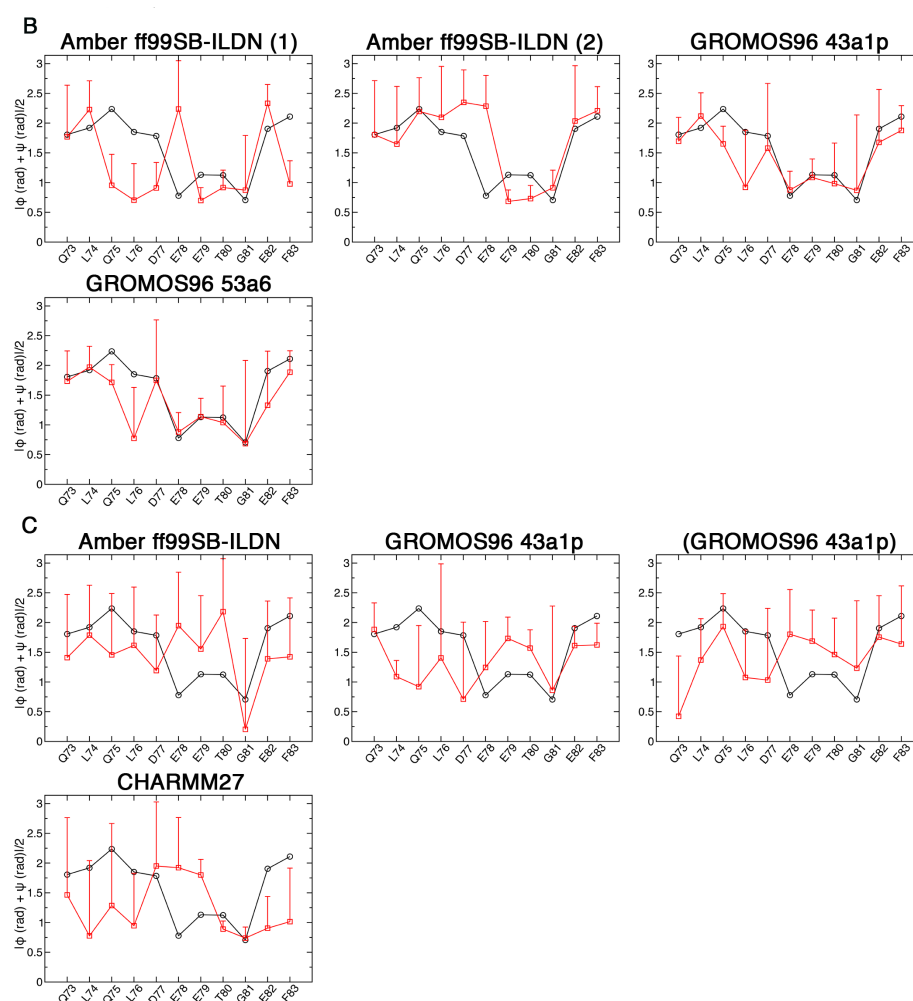
Figure 2.7 Comparison of the backbone dihedral angles from the bound state structure and MD simulations.

The φ and ψ angles for residues [73]QLQLDEETGEF[83] were converted to radians and the absolute values were summed and averaged. Black circles indicate the values from the bound state crystal structure (PDB id: 2FLU) (31). Red squares are the values over the last 0.1 μs of the simulations. A) Uncapped peptide. B) Capped peptide. C) pThr-80 peptide.
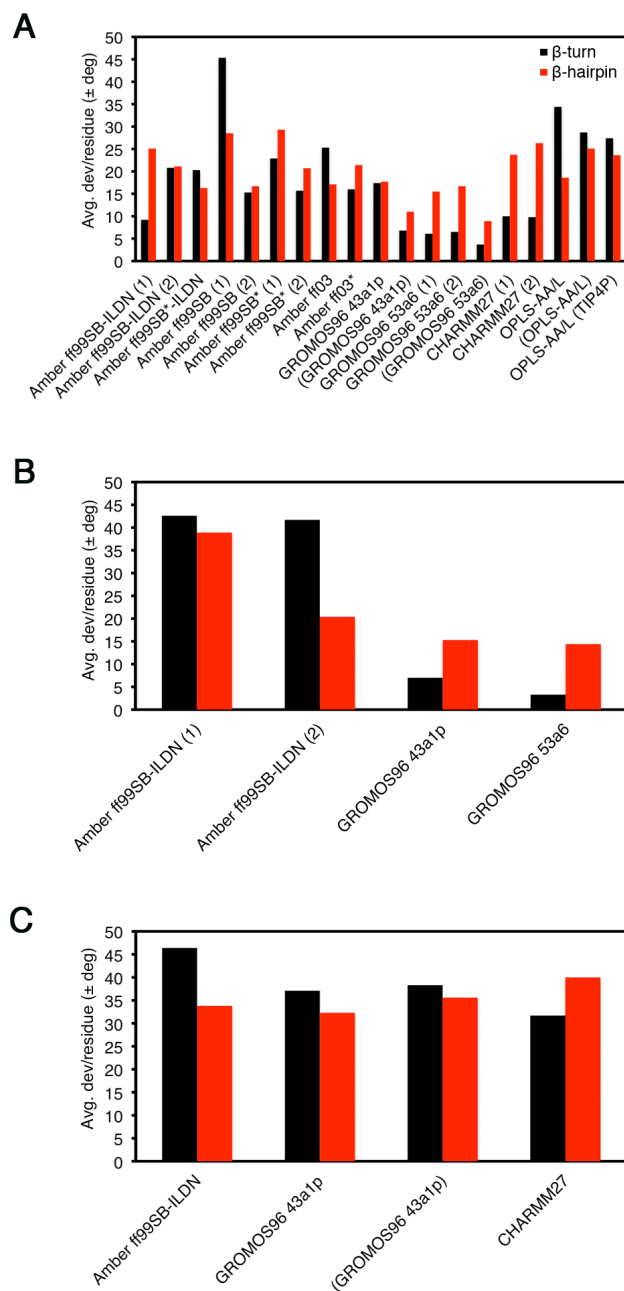
Figure 2.8 Average combined φ and ψ deviations per residue from the bound state crystal structure.

Black bars are for the β-turn 4-mer, [77]DEET[80] and red bars are for the β-hairpin 13-mer, [72]AQLQLDEETGEFL[84]. Data was analyzed over the last 0.1 μs of the simulations. A) Uncapped peptide. B) Capped peptide. C) pThr-80 peptide.

Secondary structure formation at elevated temperatures

Finally, we have performed MD simulations at elevated temperatures using a subset of force fields, Amber ff99SB*-ILDN and ff03*, GROMOS96 53a6, CHARMM27, and OPLS-AA/L, to identify secondary structure formation of the Neh2 peptide under these conditions. Using elevated temperatures provides an additional test to examine possible metastable states. The simulations were performed at 330, 350 and 370 K from both the initial and final (after 1 μs) system configurations at 310 K. Again, we used DSSP analysis to illustrate the evolution of secondary structures over the trajectories.

In the simulations starting from the initial (unfolded) system coordinates, hairpin formation, at the expected location, was observed in the Amber ff03*, GROMOS96 53a6 (2), and capped GROMOS96 53a6 simulations at 330 K (Figure 9). β-hairpin structures were also identified in the MD simulations using these force fields at 310 K as mentioned above (Figure 2.1). Hairpin conformation, which was not observed in CHARMM27 (1) at 310 K, was significantly populated in the trajectory at 330 K (Figure 2.9). At 350 K, the Amber ff03*, GROMOS96 53a6 (2) and capped GROMOS96 53a6 simulations still had hairpin signatures at some points in their trajectories, but β-hairpin structure was no longer observed in the CHARMM27 (1) simulation (Figure 2.9). On the other hand, a low population of hairpin conformation was observed in the 350 K OPLS-AA/L trajectory (Figure 2.9). Significant population of β-hairpin structure remained even at 370 K in GROMOS96 53a6 (2) and capped GROMOS96 53a6 simulations (Figure 2.9), while only transiently formed hairpin was observed in CHARMM27 (1). It is noteworthy that rapid hairpin folding and high thermal stability were observed in the capped GROMOS96 53a6 simulations at all elevated temperatures (Figure 2.9).
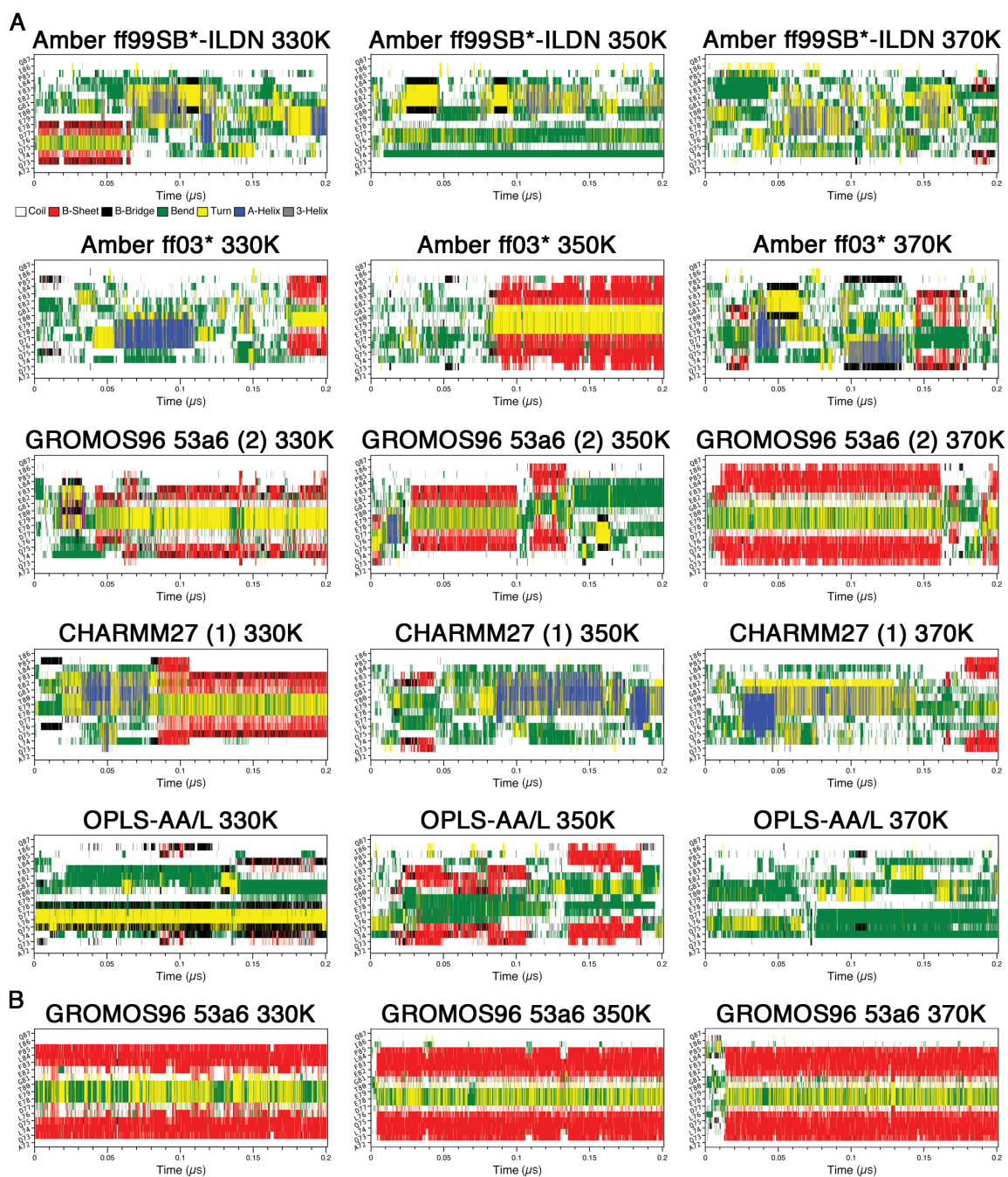
Figure 2.9 Secondary structure propensity analysis of the elevated temperature simulations from the initial system configurations.

Secondary structure content was assessed with the DSSP algorithm[71]: coil (white), β-sheet (red), β-bridge (black), bend (green), turn (yellow), α-helix (blue) and 3_{10} helix (grey). A) Uncapped peptide. B) Capped peptide.

In the simulations starting from the final system coordinates, the Amber ff99SB*-ILDN, Amber ff03*, GROMOS96 53a6 (2) and capped GROMOS96 53a6 trajectories, all of which formed hairpins at 310 K, maintained hairpin signatures at 330 K over 0.2 μs (Figure 10). On the other hand, the CHARMM27 (1) and OPLS-AA/L trajectories at 330 K were heavily biased by α-helical and bend conformations, respectively, similar to what were observed at 310 K (Figure 2.10). When the temperature was increased to 350 K, the hairpin signature in the Amber ff99SB*-ILDN trajectory disappeared shortly after ~ 0.1 μs, however the Amber ff03*, GROMOS96 53a6 (2) and capped GROMOS96 53a6 trajectories maintained their hairpins over the whole 0.2 μs period (Figure 2.10). The CHARMM27 (1) simulation at 350 K lost its helical properties after about 0.15 μs and appeared to form a hairpin shortly after (Figure 2.10). At 370 K, both the Amber ff99SB*-ILDN and Amber ff03* trajectories lost their hairpin signatures after ~ 0.1 μs, but the GROMOS96 53a6 (2) and capped GROMOS96 53a6 simulations still remained in hairpin conformations throughout almost the whole trajectories (Figure 2.10). On the other hand, the CHARMM27 (1) simulation at 370 K lost its helical property almost immediately and a turn conformation was present in the expected location, but a distinct hairpin signature was not observed (Figure 2.10). The OPLS-AA/L trajectories did not have any clear hairpin signatures at any of the temperatures (Figure 2.10). Once again, high thermal stability was observed in the GROMOS96 53a6 simulations (Figure 2.10).
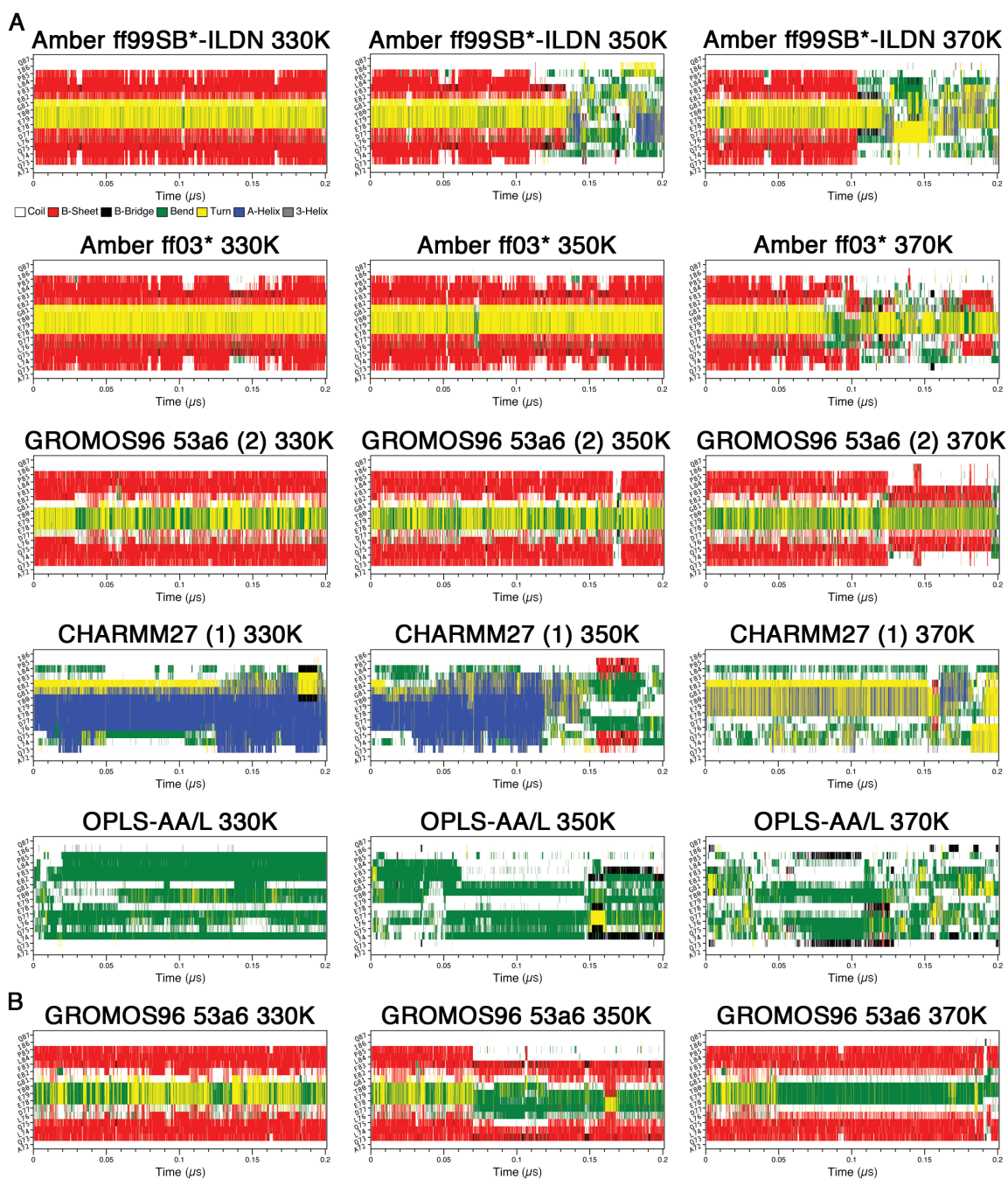
Figure 2.10 Secondary structure propensity analysis of the elevated temperature simulations from the final (after 1µs) system configurations.

Secondary structure content was assessed with the DSSP algorithm[71]: coil (white), β-sheet (red), β-bridge (black), bend (green), turn (yellow), α-helix (blue) and 3_{10} helix (grey). A) Uncapped peptide. B) Capped peptide.

2.5 Discussion and conclusions

We have examined the folding of a 16-mer polypeptide with 10 commonly used biomolecular force fields. The peptide used in this study is derived from the Neh2 domain of Nrf2. Despite that Neh2 has been characterized as being intrinsically disordered, the region encoded by the sequence of this peptide has been shown to contain β-hairpin structure (30, 31, 72). Various criteria were used to assess β-hairpin formation of this peptide and compare the results to experimental data. Although the simulations all used the same, non-native, starting structure and were performed with identical parameters, clear differences were observed between different force fields used and even between replicate simulations with the same force field.

While no single type of analysis was sufficient to thoroughly assess and compare β-hairpin formation, the DSSP plots were useful for visualizing potential hairpin formation in this work. In addition, these analyses were also useful in identifying other types of secondary structures. For example, DSSP plots of the CHARMM27 simulations showed that the Nrf2 peptide did not fold into hairpins, but had tendencies to form short α-helices (Figure 2.2A). This finding was not completely unexpected because CHARMM force fields have been known to have a bias towards helical structures, even when simulating the folding of all β proteins (3, 11-13, 73). In addition to CHARMM27, the Amber ff03 force field has also been shown to overstabilize helical structures. Lindorff-Larsen et al. (44) observed that while both CHARMM27 and Amber ff03 could fold the α-helical villin headpiece, proper folding of the β-sheet WW domain could not be achieved even in simulations that were 10 times the experimentally determined folding time in length. On the other hand, they found that the "helix-coil-balanced" Amber ff03* and recently developed CHARMM22* variants could achieve proper folding of both villin and the WW domain (8, 57).

The DSSP plots for the OPLS-AA/L force field simulations also were not indicative of hairpins, but showed considerable amounts of 'bend' content. This aligns with the finding of Cao et al. (74) that this force field did not produce the expected β-

hairpin structure of the H1 peptide. Interestingly, simulations of the H1 peptide performed with GROMOS96 43a1 yielded a β-hairpin structure consistent with experimental data (74). It is difficult to determine why the OPLS-AA/L simulations did not form a native-like hairpin structure in our simulations. It is possible that, in general, longer trajectories may be needed for convergence due to the rugged energy landscape and different barriers in these systems (75, 76). The weak hairpin signature observed in the DSSP plot of the OPLS-AA/L trajectory at 350 K supports this notion. Alternatively, there may be incompatibilities between our peptide sequence and OPLS-AA/L, such as high amounts of exposed hydrophobic content (74).

The C_α-C_α contacts plots also illustrated β-hairpin formation and helped to identify non-native hairpins. For example, these plots showed that the β-turn in the capped peptide Amber ff99SB-ILDN (2) simulation was slightly displaced from its expected location (Figure 2.2B). This likely explains the lack of Asp-77 to Thr-80 hydrogen bonding in this simulation (Table 2.2). Together, our findings from the DSSP and C_α-C_α contact analysis, suggested that the uncapped Amber ff99SB-ILDN (1), Amber ff99SB*-ILDN, Amber ff99SB (2), Amber ff99SB* (2), Amber ff03, Amber ff03*, GROMOS96 43a1p, GROMOS96 53a6 (1 & 2) and capped GROMOS96 43a1p and 53a6 simulations formed native-like β-hairpins.

Interestingly, the simulations that formed β-hairpins, as judged by DSSP and C_α-C_α contact analysis, also exhibited experimentally determined native contacts present in the free state of Nrf2. Furthermore, we observed that the presence or absence of native contacts was correlated with the frequency of Asp-77 to Thr-80 hydrogen bonding to some extent. Interactions between these residues are thought to be important for the hairpin structure (9, 31). Most of the simulations of the uncapped and unphosphorylated peptides that had two or more native contacts also had high frequencies of Asp-77 to Thr-80 hydrogen bonding. On the other hand, when 1 or 0 native contacts were present, there was usually less hydrogen bonding. One exception was the Amber ff03 simulation, which had all 3 native contacts, but lacked hydrogen bonding between Asp-77 to Thr-80. Figure 2.5 shows that in this simulation, Leu-76 and Asp-77 had large backbone dihedral angle deviations from the bound state structure, which could possibly explain the lower

hydrogen bonding with Thr-80. It is possible that alternate hydrogen bonds between Leu-76 and Thr-80 may have compensated. The evident positive correlation between native contact formation and a high frequency of Asp-77 to Thr-80 hydrogen bonding in our simulations supports prior suggestions that these interactions are vital for the hairpin structure (9, 31).

We also found that the simulations that formed β-hairpins converged upon conformations that were similar to the structure of Nrf2 bound to Keap1 (PDB id: 2FLU) (31). It is common for disordered proteins, like Nrf2, to contain preformed structural elements in their binding regions (9, 30, 77-79). Indeed, NMR data and our prior MD simulations indicated that Nrf2 adopts a hairpin structure in the free-state, which is high resemblance to its Keap1 bound form (PDB id: 2FLU) (9, 31). Therefore, it was expected that simulations with 2-3 free state native contacts also had low RMSDs to the bound state structure. The GROMOS96 simulations clearly had the lowest β-turn and hairpin RMSDs of all the simulations. These simulations also had very low dihedral angle deviations from the bound state structure.

In general, the simulations that used charge-groups or peptide capping groups were not largely different from their uncapped counterparts with single atom charge-groups. When studying peptides from the interior of a protein sequence, it is common to add capping groups to the ends. This neutralizes the unphysical charges introduced by the free N- and C-termini, which can potentially disrupt the native structure. However, we did not find that the uncapped termini had a detrimental effect on hairpin folding in our current simulations. The GROMOS96 force field simulations employing default charge-groups or peptide capping groups were highly consistent, in all aspects, to their uncapped counterparts. On the other hand, in both capped Amber ff99SB-ILDN replicates, the peptide folded into structures that were moderately different from their uncapped counterparts. It was difficult to determine the cause of this behavior and it could simply be a convergence issue.

The finding that none of the simulations where Thr-80 was phosphorylated formed β-hairpins was not surprising. Experimental data has shown that phosphorylation

of this residue can severely impair binding of Nrf2 to Keap1, likely due to a disruption of β-turn formation (31). Our pThr-80 simulations were consistent with this proposition and also suggest that β-turn disruption strongly impairs hairpin formation.

The evident differences between duplicate simulations in this work highlight the importance of replica simulations when performing MD simulations of folding. Even though all duplicate simulations here used identical starting structures and parameters, the assignment of different initial atom velocities led the simulations to follow different pathways. As a result, duplicate simulations did not always converge upon folded structures even with microsecond long trajectories. In this work, we have conducted simulations at elevated temperatures using a subset of force fields in order to gain insights into the temperature-dependence and metastability of conformational sampling. The results show that with the GROMOS96 53a6 force field, the Neh2 peptide continued to fold into β-hairpin conformation and remained stable even at higher temperatures. This is quite different from what was observed for Amber ff99SB*-ILDN as the hairpin structure becomes less stable under this force field when the temperature increases. Although native-like conformation was not observed in the microsecond long CHARMM27 (1) simulation at 310 K, the peptide quickly folded into a β-hairpin structure at 330 K. Therefore, the lack of conformational convergence at lower temperature may simply be due to insufficient sampling time. However, further increase in temperature (i.e. 350 and 370 K) again led to the disappearance of β-hairpin structure in the CHARMM27 (1) simulations. The results here also show that although long simulation times are necessary, it is important to have alternative methods of sampling conformations, such as replica-exchange and related methods (82-84).

Finally, this and other recent comparative studies (44, 45, 85) show the importance of using different criteria for assessing the properties of different force fields. In addition to more reliable simulations, such studies provide invaluable information about the collective non-additive properties of amino acids that are helpful in interpreting experiments.

2.6 Supplemental information

Two videos are available on YouTube (too large for direct inclusion):

Video 1: http://www.youtube.com/watch?v=AtDOJnVNC18&feature=channel&list=UL. The first and last 10 ns of the Amber ff99SB-ILDN (1), Amber ff99SB*-ILDN, Amber ff99SB (2), Amber ff03, Amber ff03* GROMOS96 43a1p, GROMOS96 53a6 (2), CHARMM27 (2) and OPLS-AA/L trajectories (without terminal capping or charge groups). For clarity, water, ions and hydrogens are not shown and rotation and translation of the peptide has been removed. Secondary structures were colored as follows in VMD: yellow - β-sheet (arrows indicate chain direction), purple - alpha helix, blue - 3_{10} helix, white - coil.

Video 2: at http://www.youtube.com/watch?v=HtRbtQ12eOI&feature=channel&list=UL. 0-400 ns of the Amber ff99SB* (2) trajectory. For clarity, water, ions and hydrogens are not shown and rotation and translation of the peptide has been removed. Secondary structures were colored as follows in VMD: yellow - β-sheet (arrows indicate chain direction), purple - alpha helix, blue - 3_{10} helix, white - coil.

2.7 Acknowledgements

2.8 References

1. Lei H, Wu C, Liu H, Duan Y (2007) Folding free-energy landscape of villin headpiece subdomain from molecular dynamics simulations. *Proc Natl Acad Sci U S A* 104:4925-30.

2. Ensign DL, Pande VS (2009) The Fip35 WW domain folds with structural and mechanistic heterogeneity in molecular dynamics simulations. *Biophys J* 96:L53-5.

3. Best RB, Mittal J (2011) Free-energy landscape of the GB1 hairpin in all-atom explicit solvent simulations with different force fields: Similarities and differences. *Proteins* 79:1318-28.

4. Zhou R (2003) Trp-cage: folding free energy landscape in explicit water. *Proc Natl Acad Sci U S A* 100:13280-5.

5. Monticelli L, Sorin EJ, Tieleman DP, Pande VS, Colombo G (2008) Molecular simulation of multistate peptide dynamics: a comparison between microsecond timescale sampling and multiple shorter trajectories. *J Chem Theory Comput* 29:1740-52.

6. Voelz VA, Bowman GR, Beauchamp K, Pande VS (2010) Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1-39). *J Am Chem Soc* 132:1526-8.

7. Shaw DE et al. (2010) Atomic-level characterization of the structural dynamics of proteins. *Science* 330:341-6.

8. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) How fast-folding proteins fold. *Science* 334:517-20.

9. Cino EA, Wong-Ekkabut J, Karttunen M, Choy WY (2011) Microsecond molecular dynamics simulations of intrinsically disordered proteins involved in the oxidative stress response. *PLoS One* 6:e27371.

10. Ulmschneider MB, Ulmschneider JP (2008) Folding Peptides into Lipid Bilayer Membranes *J Chem Theory Comput* 4:1807-1809.

11. Best RB, Buchete NV, Hummer G (2008) Are current molecular dynamics force fields too helical? *Biophys J* 95:L07-9.

12. Freddolino PL, Park S, Roux B, Schulten K (2009) Force field bias in protein folding simulations. *Biophys J* 96:3772-80.

13. Mittal J, Best RB (2010) Tackling force-field bias in protein folding simulations: folding of Villin HP35 and Pin WW domains in explicit water. *Biophys J* 99:L26-8.

14. Patapati KK, Glykos NM (2011) Three force fields' views of the 3(10) helix. *Biophys J* 101:1766-71.

15. Sorin EJ, Pande VS (2005) Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophys J* 88:2472-93.

16. Duan Y et al. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem* 24:1999-2012.

17. Best RB, Hummer G (2009) Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J Phys Chem B* 113:9004-15.

18. Mackerell AD, Feig M, Brooks CL (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* 25:1400-15.

19. Hornak V et al. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65:712-25.

20. Lindorff-Larsen K et al. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78:1950-8.

21. Oostenbrink C, Soares TA, van der Vegt NF, van Gunsteren WF (2005) Validation of the 53A6 GROMOS force field. *Eur Biophys J* 34:273-84.

22. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6 *Journal of Computational Chemistry* 25:1656-1676.

23. Itoh K et al. (1997) An Nrf2/small Maf heterodimer mediates the induction of phase II detoxifying enzyme genes through antioxidant response elements *Biochem Biophys Res Commun* 236:313-322.

24. Itoh K et al. (1999) Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain. *Genes Dev* 13:76-86.

25. Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ (2000) Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform* 11:161-71.

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

27. Dunker AK, Obradovic Z (2001) The protein trinity--linking function and disorder. *Nat Biotechnol* 19:805-6.

28. Radivojac P et al. (2007) Intrinsic disorder and functional proteomics. *Biophys J* 92:1439-56.

29. Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* 293:321-31.

30. Tong KI et al. (2006) Keap1 recruits Neh2 through binding to ETGE and DLG motifs: characterization of the two-site molecular recognition model. *Mol Cell Biol* 26:2887-900.

31. Lo SC, Li X, Henzl MT, Beamer LJ, Hannink M (2006) Structure of the Keap1:Nrf2 interface provides mechanistic insight into Nrf2 signaling. *EMBO J* 25:3605-17.

32. Lo SC, Hannink M (2006) PGAM5, a Bcl-XL-interacting protein, is a novel substrate for the redox-regulated Keap1-dependent ubiquitin ligase complex. *J Biol Chem* 281:37893-903.

33. Strachan GD et al. (2004) Fetal Alz-50 clone 1 interacts with the human orthologue of the Kelch-like Ech-associated protein. *Biochemistry* 43:12113-22.

34. Padmanabhan B, Nakamura Y, Yokoyama S (2008) Structural analysis of the complex of Keap1 with a prothymosin alpha peptide. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 64:233-8.

35. Komatsu M et al. (2010) The selective autophagy substrate p62 activates the stress responsive transcription factor Nrf2 through inactivation of Keap1. *Nat Cell Biol* 12:213-23.

36. Camp ND et al. (2012) Wilms Tumor Gene on X Chromosome (WTX) Inhibits Degradation of NRF2 Protein through Competitive Binding to KEAP1 Protein. *J Biol Chem* 287:6539-50.

37. Ma J et al. (2012) PALB2 interacts with KEAP1 to promote NRF2 nuclear accumulation and function. *Mol Cell Biol* 32:1506-17.

38. Blanco FJ, Rivas G, Serrano L (1994) A short linear peptide that folds into a native stable beta-hairpin in aqueous solution. *Nat Struct Biol* 1:584-90.

39. Fasan R et al. (2006) Structure-activity studies in a family of beta-hairpin protein epitope mimetic inhibitors of the p53-HDM2 protein-protein interaction. *Chembiochem* 7:515-26.

40. Stefanovic M et al. (2007) An 11-amino acid beta-hairpin loop in the cytoplasmic domain of band 3 is responsible for ankyrin binding in mouse erythrocytes. *Proc Natl Acad Sci U S A* 104:13972-7.

41. Best RB, Mittal J (2011) Microscopic events in β-hairpin folding from alternative unfolded ensembles. *Proc Natl Acad Sci U S A* 108:11087-92.

42. van der Spoel D, Lindahl E (2003) Brute-Force Molecular Dynamics Simulations of Villin Headpiece: Comparison with NMR Parameters *J Phys Chem B* 107:11178-11187.

43. Guvench O, MacKerell AD (2008) Comparison of protein force fields for molecular dynamics simulations. *Methods Mol Biol* 443:63-88.

44. Lindorff-Larsen K et al. (2012) Systematic Validation of Protein Force Fields against Experimental Data. *PLoS One* 7:e32131.

45. Ciarkowski J et al. (2012) Ensemble fits of restrained peptides' conformational equilibria to NMR data. Dependence on force fields: AMBER/8 ff03 versus ECEPP/3. *J Mol Graph Model* 32:67-74.

46. Wong-ekkabut J, Miettinen MS, Dias C, Karttunen M (2010) Static charges cannot drive a continuous flow of water molecules through a carbon nanotube. *Nat Nanotechnol* 5:555-7.

47. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation *J Chem Theory Comput* 4:435-447.

48. Brunger AT (2007) Version 1.2 of the Crystallography and NMR system. *Nat Protoc* 2:2728-33.

49. Pettersen EF et al. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605-12.

50. Hansson T, Nordlund P, Aqvist J (1997) Energetics of nucleophile activation in a protein tyrosine phosphatase *J Mol Biol* 265:118-127.

51. Christen M et al. (2005) The GROMOS software for biomolecular simulation: GROMOS05. *J Comput Chem* 26:1719-51.

52. Bjelkmar P, Larsson P, Cuendet MA, Hess B, Lindahl E (2010) Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from

Correction Maps, Virtual Interaction Sites, and Water Models *J Chem Theory Comput* 6:459-466.

53. MacKerell AD et al. (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins *J Phys Chem B* 102:3586-3616.

54. Jorgensen WL, Tirado-Rives J (1988) The OPLS (optimized potentials for liquid simulations) potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin *J Am Chem Soc* 110:1657-1666.

55. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids *J Am Chem Soc* 118:11225-11236.

56. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides *The Journal of Physical Chemistry B* 105:6474-6487.

57. Piana S, Lindorff-Larsen K, Shaw DE (2011) How robust are protein folding simulations with respect to force field parameterization? *Biophys J* 100:L47-9.

58. Craft JW, Legge GB (2005) An AMBER/DYANA/MOLMOL phosphorylated amino acid library set and incorporation into NMR structure calculations. *J Biomol NMR* 33:15-24.

59. Sapay N, Tieleman DP (2011) Combination of the CHARMM27 force field with united-atom lipid force fields. *Journal of Computational Chemistry* 32:1400-10.

60. Patra M, Karttunen M (2004) Systematic comparison of force fields for microscopic simulations of NaCl in aqueous solutions: diffusion, free energy of hydration, and structural properties. *J Comput Chem* 25:678-89.

61. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126:014101.

62. Parrinello, Rahman (1981) Polymorphic transitions in single crystals: A new molecular dynamics method *J Appl Phys* 52:7182-7190.

63. Hess B, Bekker H, Berendsen HJC, Johannes JGEM (1997) LINCS: A linear constraint solver for molecular simulations *J Chem Theory Comput* 18:1463-1472.

64. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N · log (N) method for Ewald sums in large systems *J Chem Phys* 98:10089-10092.

65. Patra M et al. (2003) Molecular dynamics simulations of lipid bilayers: major artifacts due to truncating electrostatic interactions. *Biophys J* 84:3636-45.

66. Patra M et al. (2007) Long-range interactions and parallel scalability in molecular simulations *Computer physics communications* 176:14-22.

67. Berendsen HJC, Postma JPM, and Gusteren WF (1981) in *Intermolecular forces*, p 331-342.

68. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water *J Chem Phys* 79:926-935.

69. Baker EN, Hubbard RE (1984) Hydrogen bonding in globular proteins *Prog Biophys Mol Biol* 44:97-179.

70. Murzyn K, Zhao W, Karttunen M, Kurdziel M, Róg T (2006) Dynamics of water at membrane surfaces: Effect of headgroup structure. *Biointerphases* 1:98-105.

71. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577-637.

72. Tong KI, Kobayashi A, Katsuoka F, Yamamoto M (2006) Two-site substrate recognition model for the Keap1-Nrf2 system: a hinge and latch mechanism. *Biol Chem* 387:1311-20.

73. Freddolino PL, Liu F, Gruebele M, Schulten K (2008) Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. *Biophys J* 94:L75-7.

74. Cao Z, Liu L, Wang J (2011) Why the OPLS-AA force field cannot produce the β-hairpin structure of H1 peptide in solution when comparing with the GROMOS 43A1 force field? *J Biomol Struct Dyn* 29:527-39.

75. Chan HS, Shimizu S, Kaya H (2004) Cooperativity principles in protein folding. *Methods Enzymol* 380:350-79.

76. Dill KA, Chan HS (1997) From Levinthal to pathways to funnels. *Nat Struct Biol* 4:10-9.

77. Fuxreiter M, Simon I, Friedrich P, Tompa P (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J Mol Biol* 338:1015-26.

78. Gall C, Xu H, Brickenden A, Ai X, Choy WY (2007) The intrinsically disordered TC-1 interacts with Chibby via regions with high helical propensity. *Protein Sci* 16:2510-8.

79. Sivakolundu SG, Bashford D, Kriwacki RW (2005) Disordered p27Kip1 exhibits intrinsic structure resembling the Cdk2/cyclin A-bound conformation. *J Mol Biol* 353:1118-28.

80. Larkin MA et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947-8.

81. Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A* 89:10915-9.

82. Fukunishi H, Watanabe O, Takada S (2002) On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction *J Chem Phys* 116:9058-67.

83. Lee MS, Olson MA (2011) Comparison of two adaptive temperature-based replica exchange methods applied to a sharp phase transition of protein unfolding-folding. *J Chem Phys* 134:244111.

84. Chaudhury S, Olson MA, Tawa G, Wallqvist A, Lee MS (2012) Efficient conformational sampling in explicit solvent using a hybrid replica exchange molecular dynamics method *J Chem Theory Comput* 8:677-87.

85. Beauchamp KA, Lin Y-S, Das R, Pande VS (2012) Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements *J Chem Theory Comput* 8:1409-1414.

3   Microsecond Molecular Dynamics Simulations of Intrinsically Disordered Proteins Involved in the Oxidative Stress Response

Elio A. Cino[†], Jirasak Wong-ekkabut[‡+], Mikko Karttunen[‡#*], and Wing-Yiu Choy[†*]

Department of Biochemistry[†], The University of Western Ontario, London, Ontario, Canada N6A 5C1, Department of Applied Mathematics[‡], The University of Western Ontario, London, Ontario, Canada N6A 5B7, Faculty of Science, Kasetsart University, 50 Phahon Yothin Rd., Chatuchak Bangkok 10900, Thailand[+] Department of Chemistry[#], University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

*Corresponding authors: mikko.karttunen@uwaterloo.ca, jchoy4@uwo.ca

Link: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0027371

3.1 Abstract

Intrinsically disordered proteins (IDPs) are abundant in cells and have central roles in protein-protein interaction networks. Interactions between the IDP Prothymosin alpha (ProTα) and the Neh2 domain of Nuclear factor erythroid 2-related factor 2 (Nrf2), with a common binding partner, Kelch-like ECH-associated protein 1 (Keap1), are essential for regulating cellular response to oxidative stress. Misregulation of this pathway can lead to neurodegenerative diseases, premature aging and cancer. In order to understand the mechanisms these two disordered proteins employ to bind to Keap1, we performed extensive 0.5-1.0 microsecond atomistic molecular dynamics (MD) simulations and isothermal titration calorimetry experiments to investigate the structure/dynamics of free-state ProTα and Neh2 and their thermodynamics of bindings. The results show that in their free states, both ProTα and Neh2 have propensities to form bound-state-like β-turn structures but to different extents. We also found that, for both proteins, residues outside the Keap1-binding motifs may play important roles in stabilizing the bound-state-like structures. Based on our findings, we propose that the binding of disordered ProTα and Neh2 to Keap1 occurs synergistically via preformed structural elements (PSEs) and coupled folding and binding, with a heavy bias towards PSEs, particularly for Neh2. Our results provide insights into the molecular mechanisms Neh2 and ProTα bind to Keap1, information that is useful for developing therapeutics to enhance the oxidative stress response.

3.2 Introduction

IDPs are a class of proteins that are biologically functional despite lacking well-defined structures (1-5). They are abundant in nature: 25-30% of eukaryotic proteins are predicted to be at least partially disordered, while up to 70% of signaling proteins may contain intrinsically disordered regions (6, 7). Compared to globular proteins, the amino acid compositions of IDPs are usually biased towards charged, polar and structure-breaking residues, such as glycine and proline (3, 8, 9). As a result, in the absence of binding partners, these proteins generally lack structured hydrophobic cores and display high conformational flexibility (3, 5).

Despite their dynamic nature, IDPs seldom adopt completely random coil conformations (10-13). In fact, many IDPs are found to possess considerable conformational propensities along their sequences (14-20). These transiently structured regions frequently act as molecular recognition features for target binding (16-18, 21, 22). Interestingly, interactions with different partners can also cause a disordered region to adopt distinct conformations (2, 18, 21-23). For example, the same region of the intrinsically disordered C-terminus of p53 can adopt either a helix or a ß-strand structure depending on the target it interacts with (23). These unique structural properties empower many IDPs to act as hubs in protein-protein interaction networks through low-affinity but yet highly specific binding (4, 21, 24-26). Therefore, it is not a surprise that IDPs are frequently associated with human diseases, in particular cancer and neurodegenerative diseases (27-29).

Even though IDPs are involved in crucial biological functions, the mechanisms by which they interact with targets are not well understood. Recent studies have shown that some IDPs undergo large conformational changes upon target binding (4, 30-32), while others have preformed structural elements (PSEs) that resemble the bound state conformations in a significant population of conformers in the ensemble (16, 33-35). It is noteworthy that these two mechanisms are not always independent; in many cases, the binding of IDPs to their targets involves a combination of both (36). Knowledge of the detailed mechanisms that IDPs employ to bind to their targets is critical for understanding

how this class of proteins function. More importantly, it will also aid in the development of therapeutic agents targeting these types of interactions (37, 38).

While X-ray crystallography is commonly used to determine protein structures with atomic-level accuracy, the dynamic nature of IDPs makes acquiring diffracting crystals of these proteins in free states extremely challenging (2). Nuclear magnetic resonance (NMR) spectroscopy has become the primary technique for the structural characterization of this class of proteins (39, 40). Despite the fact that NMR can yield a wealth of data, there are limitations. For an IDP undergoing fast conformational exchange on the NMR timescale, collected data are averaged over the entire ensemble of conformations sampled by the protein. Therefore, unlike for folded proteins, it is inappropriate to determine a single conformation to represent the disordered state. To circumvent this problem, molecular dynamics (MD) simulations have been used to complement the experimental techniques in order to establish better models for describing the dynamic nature of interconverting disordered state ensembles and, more importantly, the mechanisms by which IDPs interact with targets. For instance, MD simulations have been performed on both the bound and apo phosphorylated forms of intrinsically disordered kinase-inducible domain (KID) to investigate the molecular mechanism by which pKID interacts with KIX in signal transduction (41). Wu *et al.* have combined NMR spectroscopy and MD simulations to identify the structural reorganization of alpha-synuclein at low pH (42).

The objective of this work is to understand the molecular mechanisms that the disordered ProTα and Neh2 domain of Nrf2 use to bind Keap1 in the oxidative stress response pathway. Exposure to toxic reactive electrophiles from the environment as well as those generated by our own metabolism can disrupt the cellular functions, resulting in neurodegenerative diseases, cancer and aging (43). Nrf2 is a key transcription factor for genes responsive to oxidative stress (44, 45). The protein consists of six highly homologous regions (Neh1-6 domains). The Neh2 domain, which is located at the N-terminus of Nrf2, plays a regulatory role by interacting with an ubiquitously expressed inhibitor, Keap1 (45). Under homeostatic conditions, the Neh2 domain of Nrf2 binds to the Kelch domains of the monomeric units of a Keap1 dimer via a high affinity ETGE

motif and a lower affinity DLG motif (with K_d values of ~8 nM and ~0.5 µM), respectively (46). When both motifs are bound to a Keap1 dimer, Neh2 is (poly) ubiquitinated and subsequently degraded by the proteosome (45-48). When the cells are under oxidative stress conditions, the interaction of Keap1 and Nrf2 is disrupted, leading to the upregulation of Nrf2-mediated gene expression.

Recent studies have shown that ProTα can compete with Nrf2 for binding to Keap1, resulting in the upregulation of Nrf2-targeted cytoprotective genes (49, 50). ProTα is ubiquitously expressed in a wide variety of human tissues and besides the regulatory role it plays in the expression of oxidative stress response genes, the protein has also been found to be involved in other cellular processes such as cell proliferation, chromatin remodeling, transcriptional regulation and apoptosis (51-53). The Keap1-binding motif of ProTα (-NEENGE-) shares a similar sequence with that of the Neh2 (-DEETGE-). Crystal structures of ProTα and Neh2 peptides bound to the Kelch domain of Keap1 further reveal that these two proteins bind to the same site on the Kelch domain and form similar β-turn conformations (46, 50) (Figure 3.1). The Kelch domain adopts a six-bladed ß-propeller structure with each blade composed of four anti-parallel ß-strands (46, 50). Both ProTα and Neh2 bind to the positively charged face of the ß-propeller where the inter-blade loops are located and the electrostatic interactions are crucial for the stability of the complexes (46, 50). Interestingly, despite the high sequence identity and structural similarity of the binding motifs, ProTα seems to have a lower binding affinity to Keap1 (see result below) compared to Neh2 (only the ETGE motif is considered) (46, 49).

A ProTα-Kelch   Neh2-Kelch

B

C Overlay

83
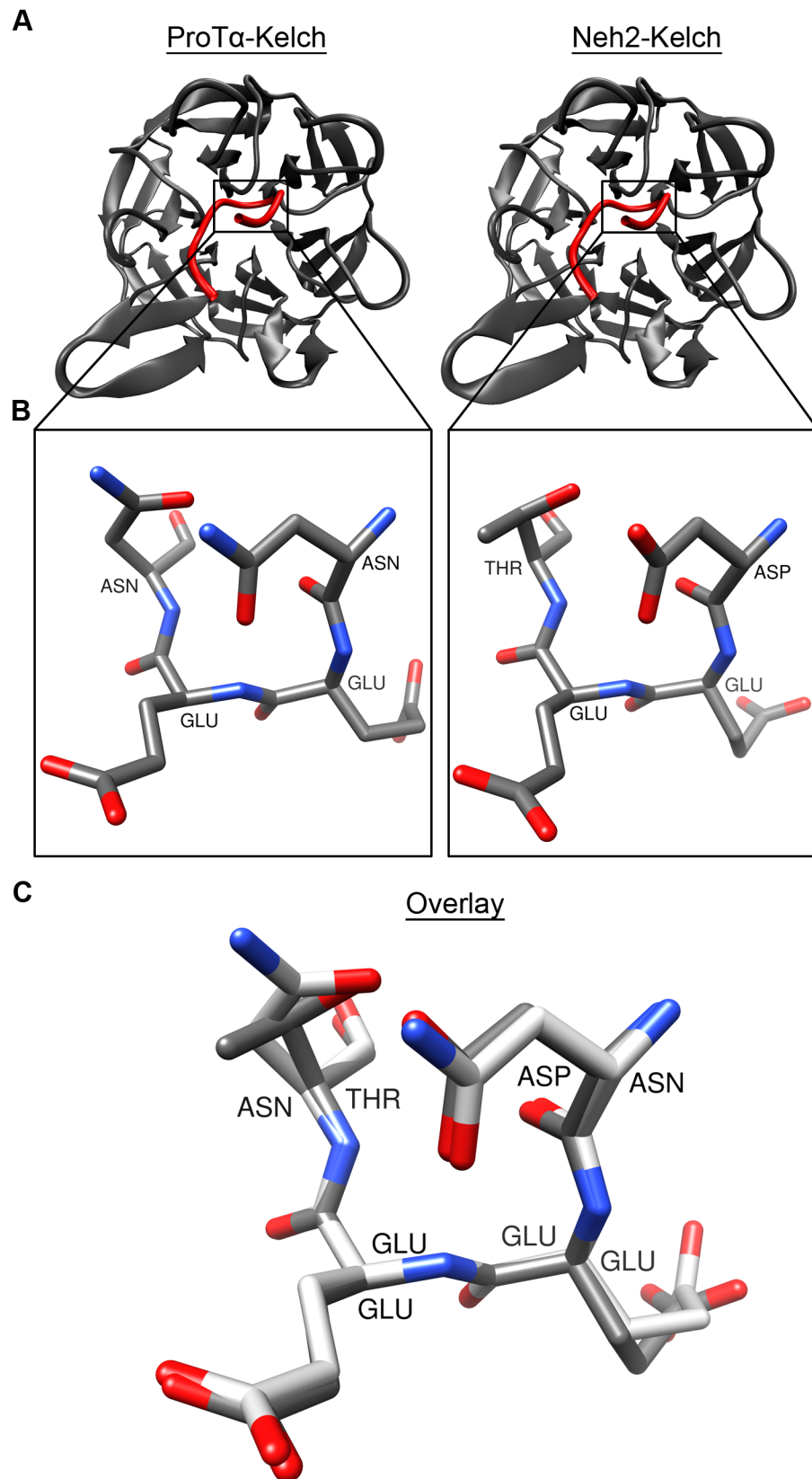
Figure 3.1 Crystal structures of ProTα and Neh2 peptides bound to the Kelch domain of Keap1.

A) Cartoon B-Spline representations of the ProTα-Keap1 and Neh2-Keap1 crystal structures (PDB ids: 2Z32 and 1X2R respectively (50, 55). Residues Asn-41 to Glu-48 of ProTα and Leu-76 to Leu-84 of Neh2 (red) are shown bound to the Kelch domain of Keap1 (grey). B) Licorice representations of the i to i+3 residues of the β-turns from the crystal structures (^{41}Asn-Glu-Glu-Asn^{44} and ^{77}Asp-Glu-Glu-Thr^{80}, of ProTα and Neh2 respectively). C) Overlay of the ProTα (white) and Neh2 (grey) β-turns.

Atomistic microsecond scale MD simulations were used to investigate the molecular mechanisms by which the intrinsically disordered ProTα and Neh2 interact with Keap1. In particular, we focused on whether their XEEXGE motifs bind to Kelch domain through coupled folding and binding, PSEs or a combination of both mechanisms. Our results show that in their free states, both the Keap1-binding motifs of ProTα and Neh2 display intrinsic propensities to form bound-state-like β-turns, and that the residues outside of the motifs may also contribute to the stability of the structural elements. We found that the Keap1-binding motif of Neh2 adopted a β-turn conformation that more closely resembled its bound-state structure than that of ProTα. Based on these results, we propose that binding occurs synergistically via a combination of PSEs and coupled folding and binding with a heavy bias towards PSEs, especially for Neh2. The better understanding of the binding mechanisms may provide insight into developing of therapeutics that can be used to promote cellular response to oxidative stress.

3.3 Materials and methods

Starting structures

The free state structure and dynamics of ProTα and Neh2 were investigated using

atomistic MD simulations. All starting structures were generated using the Crystallography & NMR System (CNS) software package (54). Briefly, extended structures were first generated based on the amino acid sequences of ProTα and Neh2. Each structure subsequently underwent a simulated annealing simulation using default CNS parameters from the anneal.inp script (without any restraints) (54). By using this procedure, we generated structures of peptides with identical sequences and lengths to those used to generate the crystal structures of mouse ProTα and Neh2 bound to Keap1 (PDB ids: 2Z32 and 1X2R respectively) (50, 55), the full-length mouse ProTα protein and a 32-mer mouse Neh2 peptide, as well as their human homologs. Table 1 summarizes the amino acid sequences used in the MD simulations and the lengths of the trajectories. Peptides with longer sequences (full-length ProTα protein and the 32-mer Neh2 peptide) were simulated to determine if residues outside of the Keap1 binding motif might be important for binding, while human sequences were simulated for cross-species comparison. To focus on the ETGE binding motif, the 32-mer Neh2 peptides instead of the full-length proteins were simulated in order to exclude the N-terminus DLG motif and the central helical region, which is not involved in Keap1 binding (46). To avoid biasing the sampling towards native-like conformations, conformers from the annealing simulations that did not resemble their bound-states were chosen as starting structures (Figure S1). The underlined residues in Table 1 comprise the Keap1-binding β-turns of ProTα and Neh2, determined from the crystal structures (50, 55), and are referred to as positions i through $i+3$ in this work (Figure 1).

MD simulations

All simulations were performed using GROMACS (GROningen MAchine for Chemical Simulations) version 4 (56), with the GROMOS96 53a6 united atom force-field parameter set (57, 58). This force field has been shown to be reliable in simulating proteins, including β-peptide folding (59). Protonation states of ionizable residues were chosen based on their most probable state at pH 7. The amino and carboxyl terminals of all systems were capped with NH_3^+ and COO^- groups respectively. The starting structures

were solvated in simple point charge (SPC) water (60), followed by the addition of sodium (Na^+) and chloride (Cl^-) ions to make the system charge neutral and bring the salt concentration to 0.1 M. The 16-mer ProTα and the 9-mer Neh2 systems (Table 3.1) contained between 9950 and 5926 water molecules and 43 to 26 molecules of salt, respectively. The full-length ProTα and the 32-mer Neh2 systems (Table 3.1) contained between 68146 and 16887 water molecules and 293 to 67 molecules of salt, respectively. The GROMOS parameterization of Na^+ and Cl^- was used, which has been shown to work well with SPC water (61). MD simulations were performed at constant temperature, pressure and number of particles (NPT ensemble). Protein and non-protein atoms were coupled to their own temperature baths, which were kept constant at 310 K using the weak coupling algorithm (62). Pressure was maintained isotropically at 1 bar using the Berendsen barostat (62). Prior to the production runs, the energy of each system was minimized using the steepest descents method. This was followed by 2 ps of position-restrained dynamics with all non-hydrogen atoms restrained with a 1000 kJ mol^{-1} force constant. The timestep was set to 2 fs. Initial atom velocities were taken from a Maxwellian distribution at 310 K. All bond lengths were constrained using the LINCS algorithm (63). Cut-off of 1.0 nm was used for Lennard-Jones interactions and the real part of the long-range electrostatic interactions, which were calculated using the Particle-Mesh Ewald (PME) method (64). 0.12 nm grid-spacing was used for PME. It is important to treat electrostatic interactions with accurate methods, such as PME, to avoid potential serious artifacts (65, 66). It has been shown that choosing simulation parameters, including thermostat and electrostatic treatment, is a subtle issue and that the choice of charge-groups may lead to unphysical effects (67). Baumketner *et al*. (68, 69) also reported that charge-group based truncation with reaction-field electrostatics may cause artificial repulsions between charged residues, identified as the microscopic reason behind artificial unfolding of protein in some simulations. Here, charge-groups were chosen to be small to avoid artifacts (67). Periodic boundary conditions were applied in all directions. This simulation protocol has been successfully applied in a number of our previous protein and membrane simulations (67, 70, 71). Simulations of the shorter peptide systems took ~1-2 weeks each using 32 cores, while the larger systems each took ~3-7 weeks using 64 cores. The cumulative simulation time for all of the trajectories was

~231, 000 CPU hours.

Table 3.1 Amino acid sequences of the simulated molecules and trajectory lengths.

System	Sequence	Simulation time (μs)
16-mer ProTα peptide (mouse)	[39]AQNEENGEQEADNEVD[54]	1.0
9-mer Neh2 peptide (mouse)	[76]LDEETGEFL[84]	1.0
Full-length ProTα (mouse)	[1]MSDAAVDTSSEITTKDLKEKKEVVEEAE NGRDAPANGNAQNEENGEQEADNEVDEE EEEGGEEEEEEEEGDGEEEDGDEDEEAEA PTGKRVAEDDEDDDVDTKKQKTEEDD[111]	0.5
32-mer Neh2 peptide (mouse)	[69]AFFAQFQLDEETGEFLPIQPAQHIQTDTS GSA[100]	0.5
Full-length ProTα (human isoform 2)	[1]MSDAAVDTSSEITTKDLKEKKEVVEEAE NGRDAPANGNANEENGEQEADNEVDEEE EEGGEEEEEEEEGDGEEEDGDEDEEAESA TGKRAAEDDEDDDVDTKKQKTDEDD[110]	0.5
32-mer Neh2 peptide (human isoform 1)	[69]AFFAQLQLDEETGEFLPIQPAQHIQSETS GSA[100]	0.5

Residues *i* through *i*+3 of the β-turn regions of the ProTα and Neh2 sequences, determined from the crystal structures (50, 55) are underlined.

Simulation analysis

To determine whether the binding motifs of ProTα and Neh2 have tendencies to adopt bound-state-like structures in their free states, coordinates from the MD trajectories were compared with the corresponding PDB crystal structures (PDB ids: 2Z32 and 1X2R respectively) (50, 55). Distance-based root-mean-square deviations (RMSD) were computed between structures at time *t* of the trajectory and the bound state reference determined from the crystal structure using the equation (56):

$$RMSD(t) = \left[\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \| r_{ij}(t) - r_{ij}(0) \|^2 \right]^{1/2}$$
(1)

where $r_{ij}(t)$ and $r_{ij}(0)$ are the distances between atoms i and j at time t of the trajectory and the same pair of atoms in the bound-state structure, respectively.

The C_α^i-C_α^{i+3} distances were calculated to determine if Keap1-binding β-turns of ProTα and Neh2 were formed during the simulations. To be defined as a β-turn, the C_α^i-C_α^{i+3} distance must be less than 7 Å (72). Residue specific dynamics of the β-turns were also probed by analyzing the circular variance (*C.V.*) of the φ and ψ dihedral angles over time. The *C.V.* is defined as (73):

$$CV = 1 - R/m$$
(2)

where m is the number of structures included in the analysis, and R is calculated using the following equation (73):

$$R = \sqrt{\frac{1}{2} \left[(\sum_{i=1}^{m} \cos \phi)^2 + (\sum_{i=1}^{m} \sin \phi)^2 + (\sum_{i=1}^{m} \cos \psi)^2 + (\sum_{i=1}^{m} \sin \psi)^2 \right]}$$
(3)

The value of *C.V.* ranges between 0 and 1. Lower values represent tighter clustering about the mean and higher values are indicative of greater φ and ψ variability.

Hydrogen bonds were analyzed as follows: A hydrogen bond between a donor (D-H) and an acceptor (A) was considered to be formed when the DA distance was less than 3.2 Å and the angle between the DA vector and the D-H bond (AD-H angle) was less than 35° (74, 75). Visualization of the structures was done using VMD (76) and Chimera (77).

Isothermal titration calorimetry (ITC) experiments

The Kelch domain (residues 324-612) of mouse Keap1 was expressed in *Escherichia coli* BL21 (DE3) grown in minimal M9 medium. The N-terminally His-tagged protein was purified by affinity chromatography using Ni Sepharose™ 6 Fast Flow beads (Amersham Biosciences). The tag was then cleaved by incubation with His-tagged tobacco etch virus (TEV) protease overnight at 25 ºC. The protein product was purified by passing the mixture through Ni Sepharose™ 6 Fast Flow beads (Amersham Biosciences).

ITC experiments were performed on a VP-ITC system (MicroCal) at 25 ºC. The Kelch domain was dialyzed against 50 mM phosphate buffer at pH 7, containing 100 mM NaCl and 1 mM DTT. Mouse ProTα (Ala-39 to Asp-54) and Neh2 (Leu-76 to Leu-84) peptides (GenScript) were also dissolved in the same buffer. All samples were filtered and degassed before the ITC experiments. Typically, 5 µL aliquots of 0.5 mM ProTα or Neh2 peptide were titrated to the sample cell containing 1.4 mL of 0.05 mM Kelch at 4-minute intervals. Heat changes after saturation were used to account for the heat of dilution. The binding stoichiometries (n), enthalpy changes (ΔH), binding constants (K$_a$), Gibbs free energy changes (ΔG) and entropy changes (ΔS) were calculated using the titration data.

3.4 Results and discussion

MD simulations were used to study the free-state structure and dynamics of ProTα and the Neh2 domain of Nrf2. The crystal structures revealed that the NEENGE and DEETGE motifs of ProTα and Neh2, respectively, bind to same site on the C-terminal Kelch domain of Keap1 (50, 55) (Figure 3.1). In particular, both the segments NEEN and DEET of ProTα and Neh2 occupied positions *i* through *i+3* of their respective β-turns and adopted highly similar structures in their bound states (Figure 3.1). We compared the structures of free-state ProTα and Neh2 peptides from the MD simulations to their corresponding bound-state conformations (50, 55) in order to determine whether ProTα and Neh2 interact with Keap1 via PSEs or coupled folding and binding mechanisms. MD simulations on the full-length ProTα protein and a 32-mer Neh2

peptide were also performed to determine if the residues outside the binding motifs play a role in binding. Finally, contributing factors to the β-turn propensities of ProTα and Neh2 were investigated through circular variance, C_α-C_α contact, and hydrogen-bond analyses.

Comparison of the free and bound-state structures

We first determined the average distance-based RMSD values (Eq. 1) between the free-state MD structures of ProTα and Neh2 peptides and their corresponding Keap1 bound-state conformations (Table 3.2). To focus on the turn structure that is crucial for the Keap1 binding, only the four residues that are involved in the β-turn formation (NEEN and DEET of ProTα and Neh2, respectively) were included in the following calculations. The all-atom RMSD values plotted over the trajectories reveal that the β-turn segment in the ProTα peptide sampled conformations with ~3 Å RMSD from the bound-state structure for the majority of the trajectory, and infrequently adopted lower RMSD (i.e. < 1.0 Å) bound-state like conformations (Figure 3.2A). In contrast, the 9-mer Neh2 peptide underwent conformational change between structures with ~1.0 Å and ~2.5 Å all-atom RMSD throughout the trajectory and adopted bound-state like conformations at multiple periods of time (Figure 3.2A; Video S3.1).

Table 3.2 Average distance-based RMSD values between the bound-state conformation and the MD structures.

System	C_α (Å) ± sdev	Backbone[a] (Å) ± sdev	All-atom (Å) ± sdev
16-mer ProTα peptide	1.17 ± 0.48	1.13 ± 0.40	2.47 ± 0.62
9-mer Neh2 peptide	1.02 ± 0.66	1.03 ± 0.59	1.73 ± 0.68
Full-length ProTα[b]	0.34 ± 0.12	0.44 ± 0.12	1.82 ± 0.25
32-mer Neh2 peptide[b]	0.18 ± 0.08	0.26 ± 0.07	0.85 ± 0.12

[a] Backbone atoms include N, C_α and C.
[b] The last 0.1 μs of the trajectory was used in the RMSD calculations.
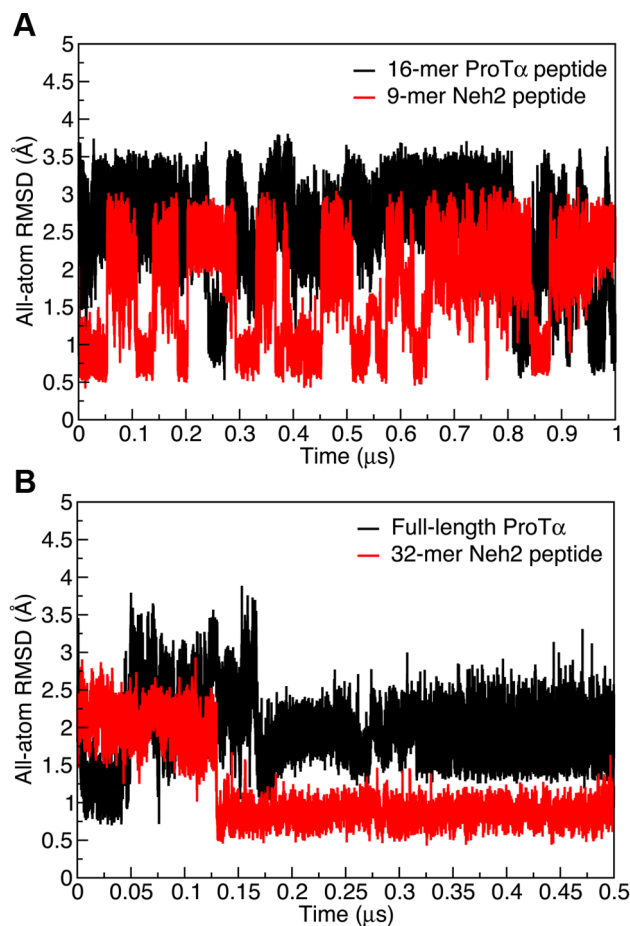
Figure 3.2 All-atom RMSD values between the MD and crystal structures.

The RMSD values were computed by subtracting the all-atom distance matrix at time t of the MD trajectories from the reference distance matrix determined from the crystal structures of the ProTα and Neh2 peptides bound to Keap1 (PDB ids: 2Z32 and 1X2R respectively) (50, 55). The distance matrices consisted of residues i through $i+3$ of the β-turn regions of the ProTα and Neh2 peptides determined from the crystal structures (50, 55).

Next, we determined if defined β-turns were formed by the free-state peptides. A good indicator of β-turn formation is that the distance between the C_α atoms of residues i and $i+3$ (C_α^i-C_α^{i+3}) is less than 7 Å (72). Based on this criterion, ~28% of the structures from the 16-mer ProTα peptide trajectory adopted a β-turn conformation in that particular

segment of the sequence, compared to ~53% of the structures from the 9-mer Neh2 trajectory (Figure 3.3A). The same data set was also plotted in terms of deviation from their corresponding $C_\alpha{}^i$-$C_\alpha{}^{i+3}$ values in the crystal structure (Figure 3.3B). The ProTα peptide had a single distribution of conformations, with an average $C_\alpha{}^i$-$C_\alpha{}^{i+3}$ deviation of ~2.2 Å from its bound state value (Figure 3.3B). In contrast, the $C_\alpha{}^i$-$C_\alpha{}^{i+3}$ distance deviations for the Neh2 peptide showed that significant populations of structures had deviations of <1.0 Å and >3.0 Å (Figure 3.3B). This finding was consistent with the RMSD data, which showed that the 9-mer Neh2 peptide transitioned between ~1 Å and ~2.5 Å all-atom RMSD throughout the trajectory (Figure 3.2A). Importantly, the RMSD data and $C_\alpha{}^i$-$C_\alpha{}^{i+3}$ distance distribution of the 9-mer Neh2 indicated that the free-state conformational ensemble of this peptide consists of both structures that closely resemble the bound-state β-turn conformation and ones that are comparably extended in that region (Video S1).
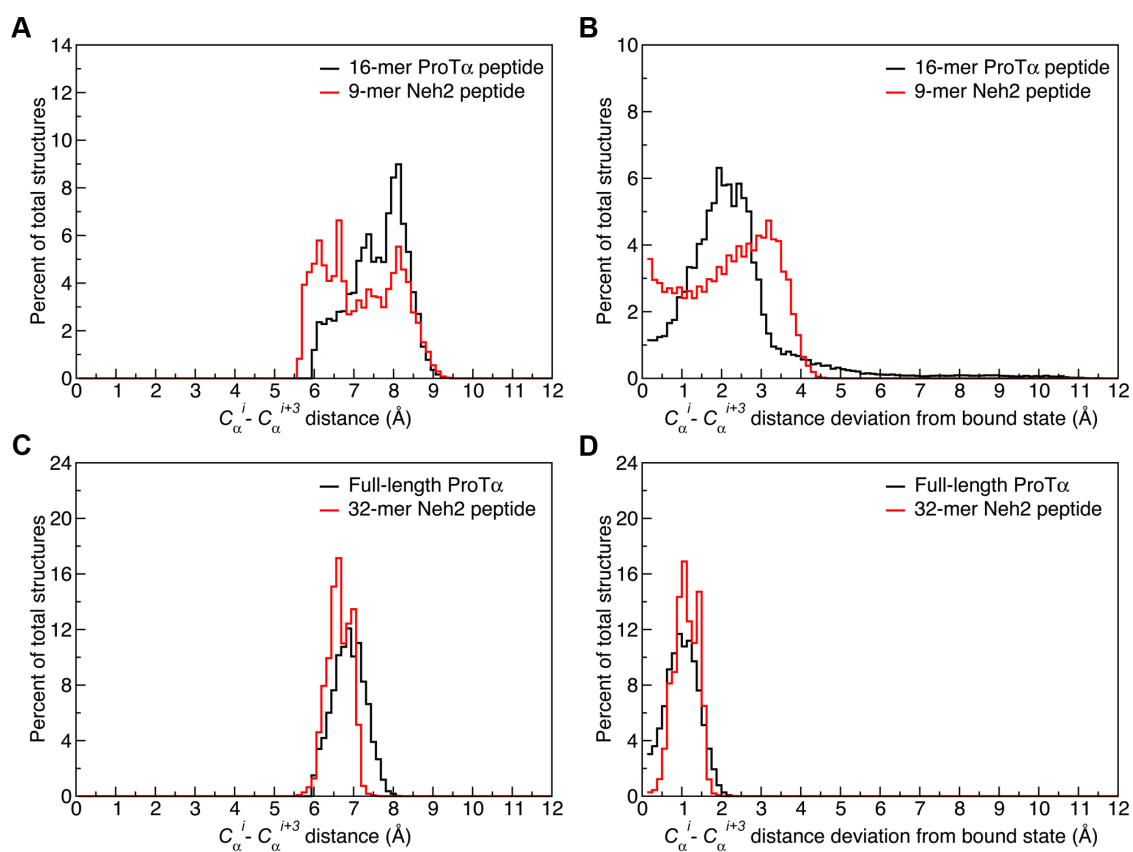
Figure 3.3 $C_\alpha{}^i$-$C_\alpha{}^{i+3}$ distances and their deviations from their crystal structure distances.

Panels B and D show the absolute deviations of $C_\alpha{}^i$-$C_\alpha{}^{i+3}$ distances from the corresponding distances in the crystal structures. Data were collected over the full 1.0 µs trajectories for the crystal structure peptides and the last 0.1 µs for the full-length ProTα and 32-mer Neh2. Deviations were calculated for $C_\alpha{}^i$-$C_\alpha{}^{i+3}$ pairs from the β-turns, determined from the crystal structures (50, 55), by subtraction of the *i* to *i*+3 distance at time *t* of the trajectory from the fixed distance of the corresponding atom pair from the crystal structures (PDB ids: 2Z32 and 1X2R) for ProTα and Neh2 respectively) (50, 55).

The above findings also indicate that during the 1-µs simulations, both the 16-mer ProTα and the 9-mer Neh2 peptides displayed intrinsic propensities of forming bound-state-like β-turn structures in the absence of Keap1. We realized that in the absence of Keap1, the peptides might not be long enough to form stable structures. To assess the contributions of residues outside the binding motifs in stabilizing the ß-turn conformation, MD simulations of the full-length ProTα protein and a 32-mer Neh2 peptide were performed. Structural resemblance to their Keap1-bound states was gauged by the same parameters as above.

Figure 3.2B shows the distance-based all-atom RMSD values between the MD structures and the corresponding bound-state crystal structures of full-length ProTα and the 32-mer Neh2 peptide over 0.5-µs trajectories. Like above, the analyses focused on the four residues that are involved in the β-turn formation. Interestingly, both the full-length ProTα protein and the 32-mer Neh2 peptide achieved lower and more stable all-atom RMSDs than their shorter counterparts (Figure 3.2). Specifically, the full-length ProTα converged to an all-atom RMSD of ~1 Å after a short period of simulation time despite having a starting structure with an RMSD ~2.6 Å (Figure S3.1). The values of RMSD fluctuated between ~0.75-3.75 Å in the first 0.18 µs and then stabilized at an all-atom RMSD around 2 Å for the remainder of the trajectory (Figure 3.2B). The 32-mer Neh2

peptide converged to an all-atom RMSD of less than 1 Å in about 0.13 µs and remained stable around that value for the rest of the trajectory (Figure 3.2B; Video S3.2). It is worth mentioning that the bound-state-like β-turn conformations formed by the full-length ProTα and the 32-mer Neh2 peptide closely resembled the ones adopted by their shorter peptide counterparts (Figure S3.2).

16-mer ProTα
All-atom RMSD = 3.17 Å

9-mer Neh2
All-atom RMSD = 1.78 Å

Full-length mouse ProTα
All-atom RMSD = 2.57 Å

32-mer mouse Neh2
All-atom RMSD = 2.77 Å

Full-length human isoform 2 ProTα
All-atom RMSD = 2.77 Å

32-mer human Neh2
All-atom RMSD = 2.57 Å
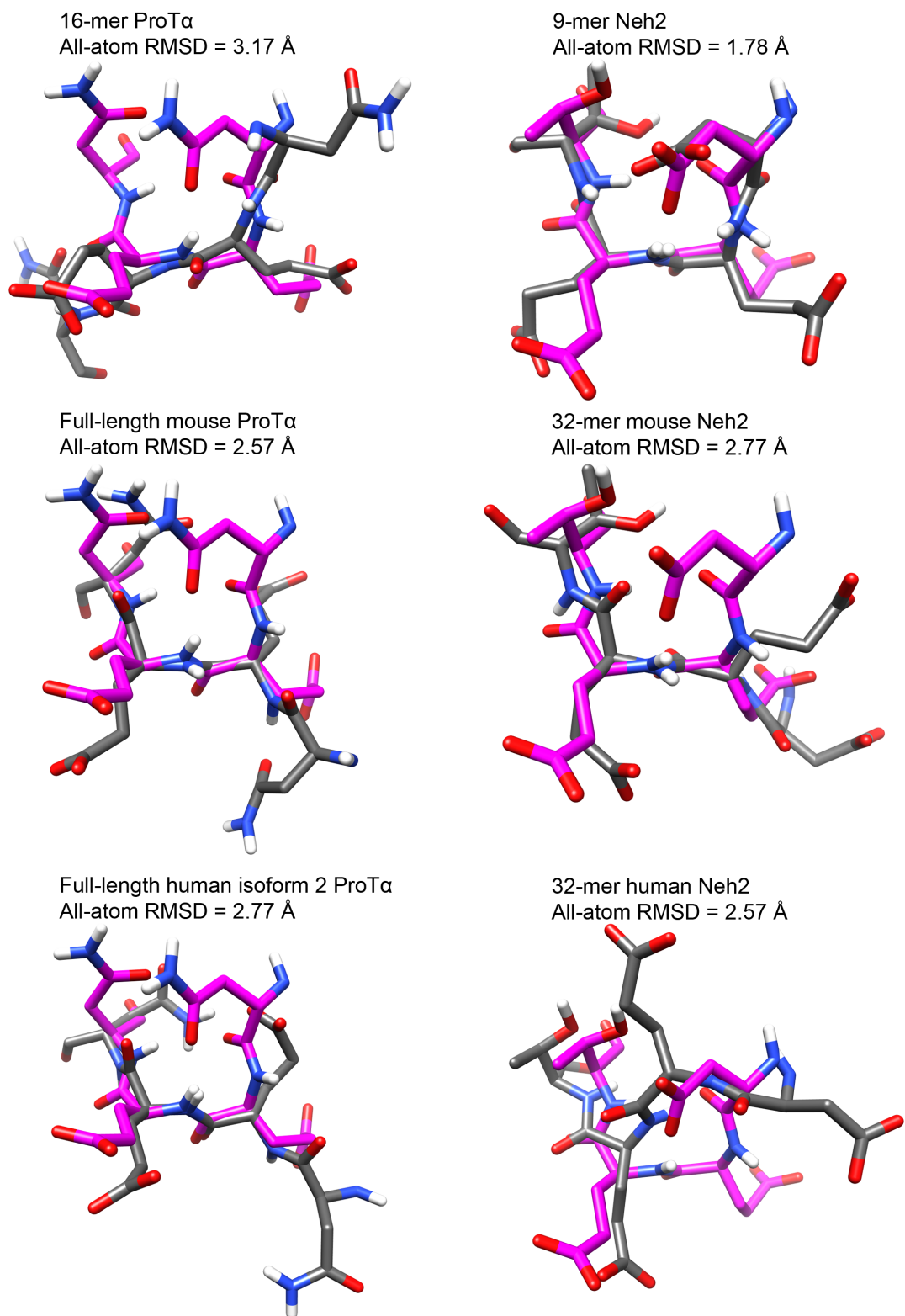
Figure S3.1 Overlays of the starting structure (grey) and crystal structure (pink) β-turns.

Residues *i* through *i*+3 of the β-turns from the starting structures, generated in CNS (54), were superimposed onto the corresponding residues from their bound state crystal structures. The RMSD values were computed by subtracting the all-atom distance matrix of the starting structures from the reference distance matrix determined from the crystal structures of the ProTα and Neh2 peptides bound to Keap1 (PDB ids: 2Z32 and 1X2R respectively) (50, 55). The distance matrices consisted of residues *i* through *i*+3 of the β-turn regions of the ProTα and Neh2 peptides determined from the crystal structures (50,55). The starting structures for human ProTα and Neh2 were compared to the mouse structures (PDB ids: 2Z32 and 1X2R) (50, 55) as their bound-state references. Hydrogen atoms were added for clarity.

ProTα
All-atom RMSD = 1.44 Å

Neh2
All-atom RMSD = 0.75 Å
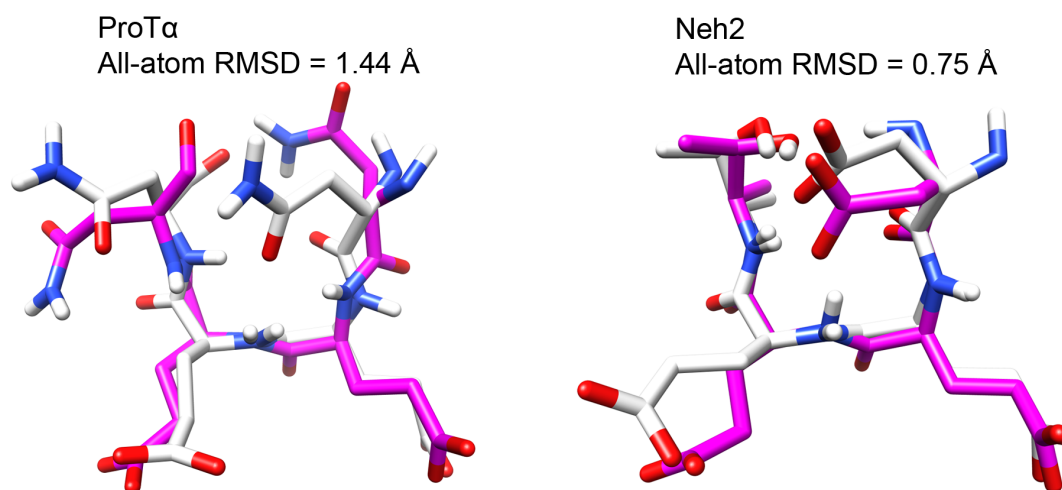
Figure S3.2 Overlays of the β-turn structures from the 16-mer ProTα and 9-mer Neh2 MD simulations (white) with those from the longer sequence simulations (pink).

The RMSD values were computed by subtracting the all-atom distance matrices. The distance matrices consisted of residues *i* through *i*+3 of the β-turn regions of the ProTα and Neh2 peptides determined from the crystal structures (50, 55). Centroid structures from the shorter peptide simulations with lowest RMSDs to the bound state (820-830 ns and 630-640 ns from the ProTα and Neh2

simulations, respectively) were superimposed onto the corresponding centroid structures from the last 100 ns of the longer sequence simulations.

The C_α^i-C_α^{i+3} distances were also calculated to appraise the formation of β-turn structure during the simulations. The results show that during the last 0.1 μs of the full-length ProTα trajectory, ~66% of the structures have the binding motif in β-turn conformations (C_α^i-C_α^{i+3} < 7 Å), compared to ~94% of the 32-mer Neh2 peptide structures (Figure 3.3C). It is noteworthy that both systems showed considerably smaller deviations from their bound-state C_α^i-C_α^{i+3} distances compared to their shorter counterparts (Figure 3.3D).

The superpositions of the cluster centroids of β-turn-forming residues from the MD simulations with their corresponding crystal structure atoms in Figure 3.4 further illustrate the structural similarities between the free and bound states for both ProTα and the Neh2 domain. The average distance-based RMSD values between the bound-state conformation and the MD structures were summarized in Table 3.2. Although both ProTα and Neh2 had average C_α and backbone RMSDs below 0.5 Å, the RMSDs and standard deviations increase considerably when all atoms were considered. It is clear that the side chains were not all in their bound state-like conformations. Figure S3 shows the distributions of side chain torsion angles in the NEEN and DEET motifs of ProTα and Neh2, respectively. The results suggest that although the backbones of these two proteins have strong propensity of forming ß-turn structure, the side chains within the turns are not restricted in torsion angle samplings. However, it is worthwhile to note that Thr-80 of Neh2 showed a clear preference for adopting a χ_1 angle that closely resembled its bound state value (Figure S3.3). This is discussed further in the following section.
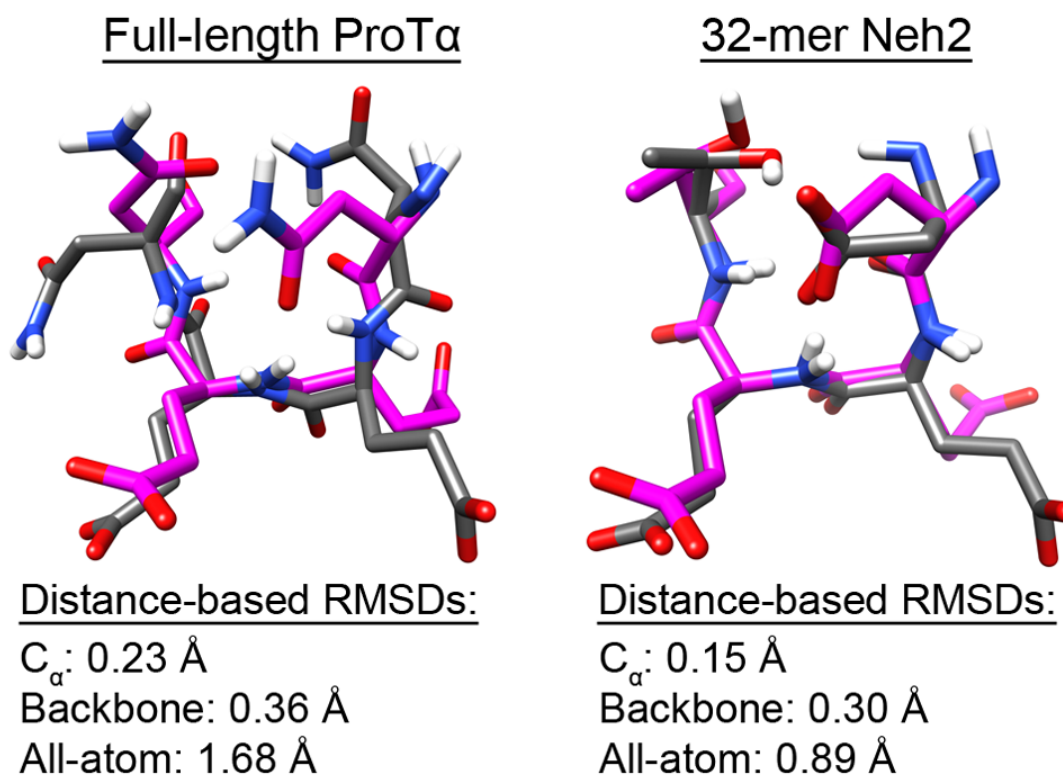
Figure 3.4 Overlay of the free and bound-state β-turns.

Residues i through $i+3$ of the β-turns from the full-length ProTα and the 32-mer Neh2 MD structures were superimposed onto the corresponding residues from their bound state crystal structures. Cluster centroids from the last 0.1 μs of the MD simulations (grey) were superimposed onto the corresponding C_α atoms from the crystal structures (pink) of ProTα and Neh2 bound to Keap1 (PDB ids: 2Z32 and 1X2R respectively) (50, 55). The single linkage clustering algorithm was used with a cutoff that included all structures from the last 0.1 μs. Hydrogens were added to the crystal structures for clarity. RMSD values were computed by subtracting the C_α, backbone or all-atom distance matrix of the centroid structures from the reference distance matrix determined from the crystal structures of the ProTα and Neh2 peptides bound to Keap1 (PDB ids: 2Z32 and 1X2R respectively) (50, 55).
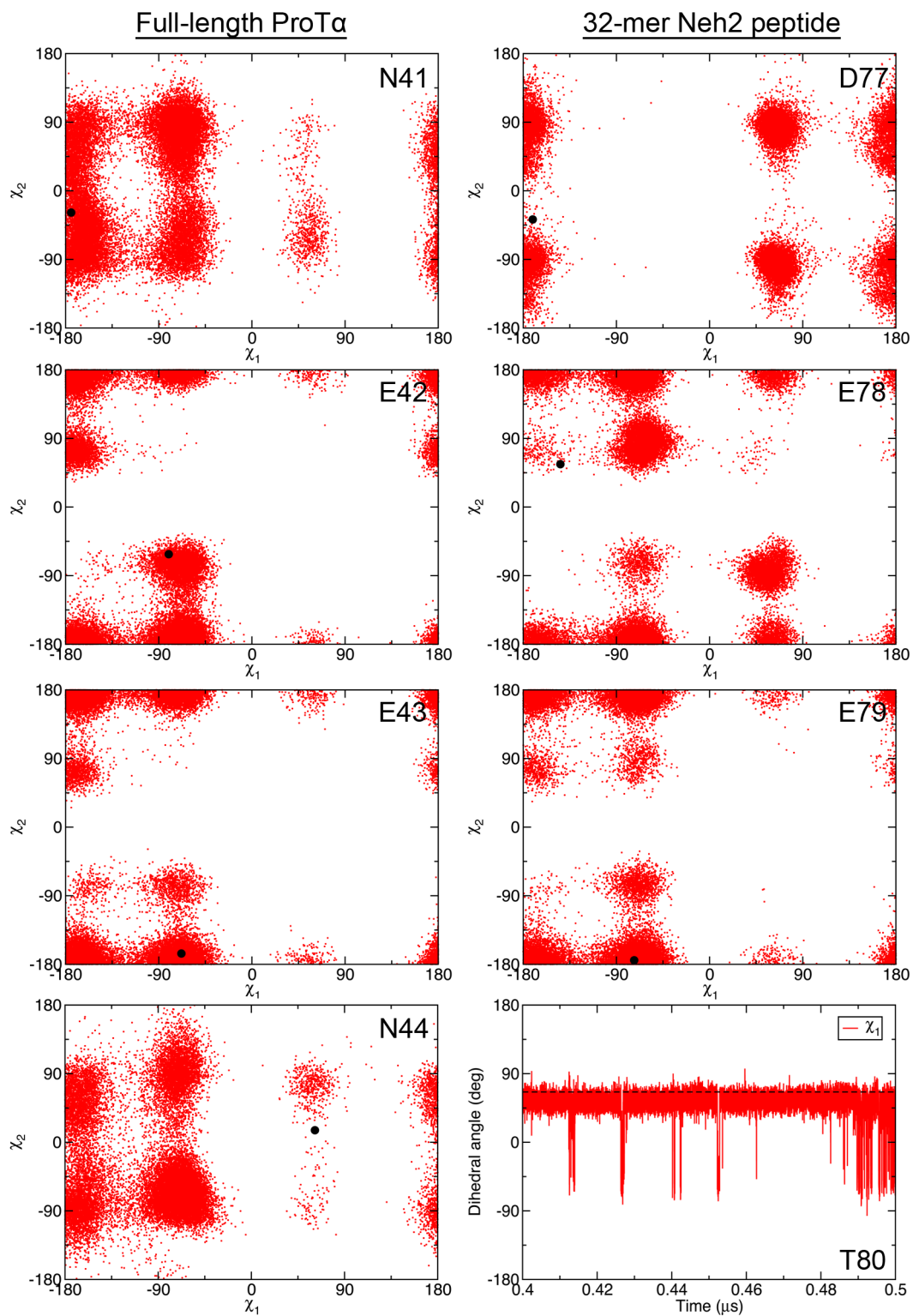
Figure S3.3 χ1 and χ2 angles from the MD and bound-state structures.

Plots of the sidechain χ1 and χ2 angles for residues i to i+3 of the β-turns are shown. Red dots indicate the angles from the last 0.1 μs of the full-length ProTα and 32-mer Neh2 trajectories. Black dots indicate the angles from the crystal structures (PDB ids: 2Z32 and 1X2R) for ProTα and Neh2 respectively) (50, 55).

Contributing factors to the β-turn propensities of ProTα and Neh2

To determine residue-specific convergences of amino acids in the torsion angle space, backbone dihedral angles of the Keap1-binding β-turns from the MD trajectories were compared to their corresponding bound-state values. Since ProTα and Neh2 peptides bind to the same site on the Kelch domain of Keap1 and adopt structurally similar β-turns (Figure 3.1), their bound-state φ and ψ angles are comparable as expected (Figure 3.5). MD simulations show that, in their free states, both ProTα and Neh2 had preferences of sampling dihedral angles around their bound-state values (Figure 3.5). Circular variance (*C.V.*) measurements were used to quantify the spread of φ and ψ angles over the last 0.1 μs of the trajectories. Both ProTα and Neh2 had similar *C.V.* (Eq. 2) values for residues *i* to *i+2*, while ProTα displayed a slightly lower circular variance for residue *i+3* compared to that of Neh2 (Figure 3.5). Snapshots over the last 0.1 μs of the trajectories illustrate that the β-turns of Neh2 and ProTα had limited backbone flexibilities (Figure 3.5).
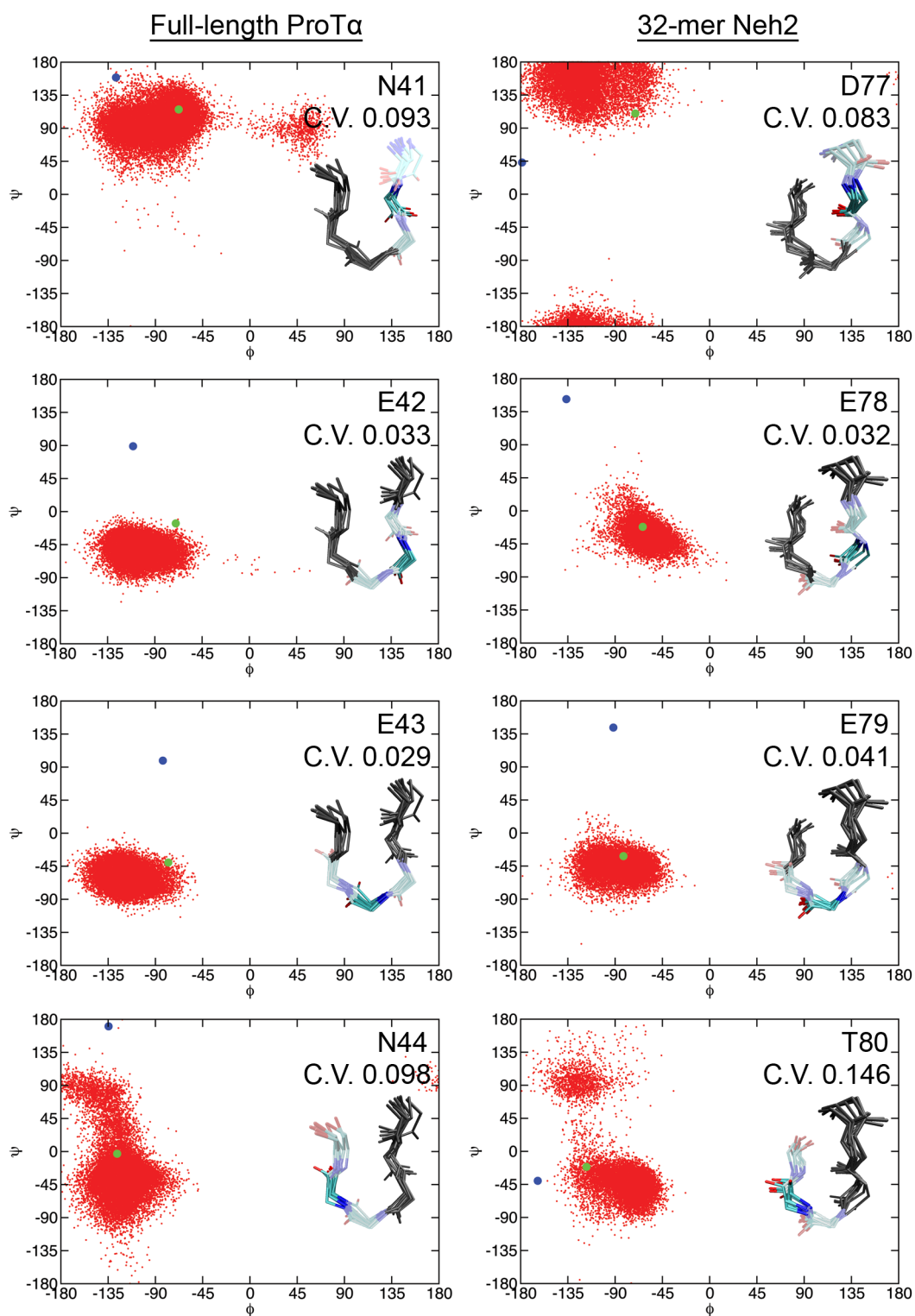
Figure 3.5 Ramachandran plots for residues *i* to *i*+3 of the ß-turns from the MD and crystal structures.

Red dots indicate the φ and ψ pair from the last 0.1 μs of the full-length ProTα and the 32-mer Neh2 trajectories. Blue circles indicate the angles of the starting structures. Green circles indicate the φ and ψ angle pair from the crystal structures (PDB ids: 2Z32 and 1X2R) (50, 55). Circular variance (*C.V.*) values and overlaid licorice representation snapshots from the last 0.1 μs of the simulations illustrate backbone mobility within the β-turns of ProTα and Neh2. Average circular variance values were calculated over the last 0.1 μs of the full-length ProTα and the 32-mer Neh2 peptide MD trajectories using the method described by MacArthur & Thornton (73).

Contacts between C_α-C_α atom pairs during the last 0.1 μs of the simulations were also examined. The contact plots and structures from the MD simulations show that the β-turns formed by ProTα and Neh2 at their Keap1-binding sites stretched out in both directions to form antiparallel β-sheets (Figure 3.6). This finding was in good agreement with previous NMR results, which suggest that residual structures may exist in regions surrounding the Keap1-binding motifs of disordered ProTα and Neh2 (46, 78). Interestingly, Neh2 has relatively higher ^{1}H-^{15}N heteronuclear NOE values in its Keap1-binding region, indicating a less dynamic free-state (46). Furthermore, chemical shift index values indicative of β-strand structure and the observance of ^{1}H, ^{1}H NOEs between the adjacent strands also evidence that residues on either side of the ETGE motif of Neh2 form a short β-sheet (46). Tong *et al.* suggested that interactions between the hydrophobic residues (Phe-74, Leu-76, Phe-83, and Pro-85) located on the β-strands may stabilize the antiparallel β-sheet structure (46). This proposal is supported by the ITC data showing that a long Neh2 segment containing the ETGE motif bound to the Kelch domain of Keap1 with higher affinity than the 9-mer peptide used in the current study ($K_d \approx 8$ nM vs $K_d \approx 182$ nM) (46). Similarly, Lo *et al.* (79) demonstrated that human Nrf2-derived 14-mer (LQLDEETGEFLPIQ) or 16-mer (AFFAQLQLDEETGEFL) peptides could compete with full-length Nrf2 for binding to Keap1 much better than a 10-mer peptide (LDEETGEFLP). Their ITC measurements showed that the human 16-mer Nrf2 peptide binds to the Kelch domain of Keap1 with $K_d \approx 20$ nM, an affinity similar to that of

the mouse homolog (79).
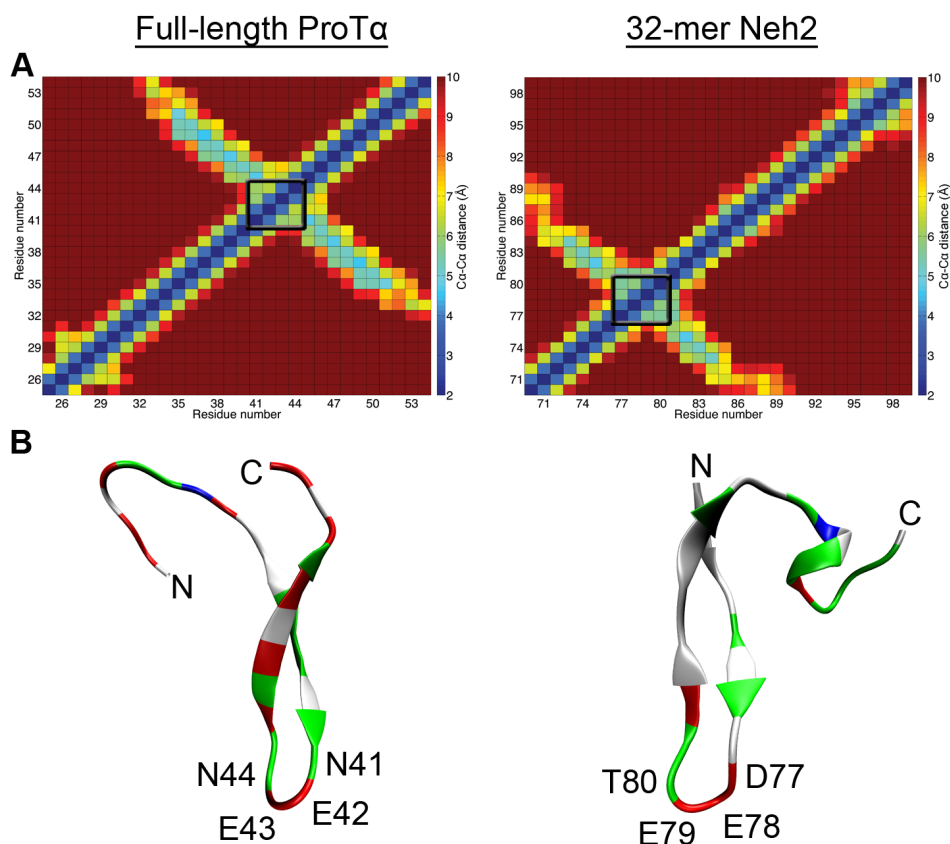
Figure 3.6 C_α-C_α contacts in the MD structures.

A) Average C_α-C_α distances over the last 0.1 μs of the full-length ProTα and 32-mer Neh2 MD trajectories. Distances equal to or greater than 10 Å are colored dark red and distances equal to or less than 2 Å are colored dark blue. The C_α^i-C_α^{i+3} atoms of the β-turns are indicated by the black boxes. B) Cartoon B-Spline representations colored by residue type of ther Keap1 binding regions of full-length ProTα and 32-mer Neh2 cluster centroids from the last 0.1 μs of the MD simulations. The single linkage clustering algorithm was used with a cutoff that included all structures from the last 0.1 μs. Residues comprising the XEEXGE Keap1-binding motifs are labeled. Directionality is indicated with the N and C labels.

In this work, we have measured the binding affinities of mouse 16-mer ProTα and 9-mer Neh2 peptides to the Kelch domain using ITC (Table 3.3; Figure S3.4). The large and negative entropy changes of 16-mer ProTα and 9-mer Neh2 peptides upon binding to Keap1 (Table 3.3) clearly reflect the significant reduction in conformational entropy of the peptides due to the disorder-to-order transition upon binding. Even though the 16-mer ProTα and the 9-mer Neh2 peptides have similar binding affinity to the Kelch domain, the former interacts more weakly with Keap1 compared to the Neh2 peptide with the same length (79). This observation is in good agreement with the lower propensity of the ß-turn formation in ProTα that is critical for the binding revealed by our MD simulations. It is noteworthy that unlike Neh2, ProTα lacks comparable hydrophobic content in the region surrounding the Keap1-binding motifs (Table 3.1). The deficiency in hydrophobic interactions may also account for the lower binding affinity between ProTα and Keap1.

Table 3.3 Thermodynamic parameters for the binding of ProTα and Neh2 peptides to the Kelch domain of Keap1.

Peptide	n^a	K_a^b $(10^6\ M^{-1})$	ΔH^b (kcal/mol)	$T\Delta S^b$ (kcal/mol)	ΔG^b (kcal/mol)
16-mer ProTα peptide	1.03	2.4 ± 0.1	-18.9 ± 0.1	-10.20	-8.70 ± 0.02
9-mer Neh2 peptide	1.02	3.7 ± 0.1	-19.0 ± 0.1	-10.05	-8.95 ± 0.02

[a] Binding stoichiometry

[b] K_a is the binding constant. ΔH, ΔS and ΔG are the change in enthalpy, entropy and Gibbs free energy upon binding (at temperature T=298 K), respectively.
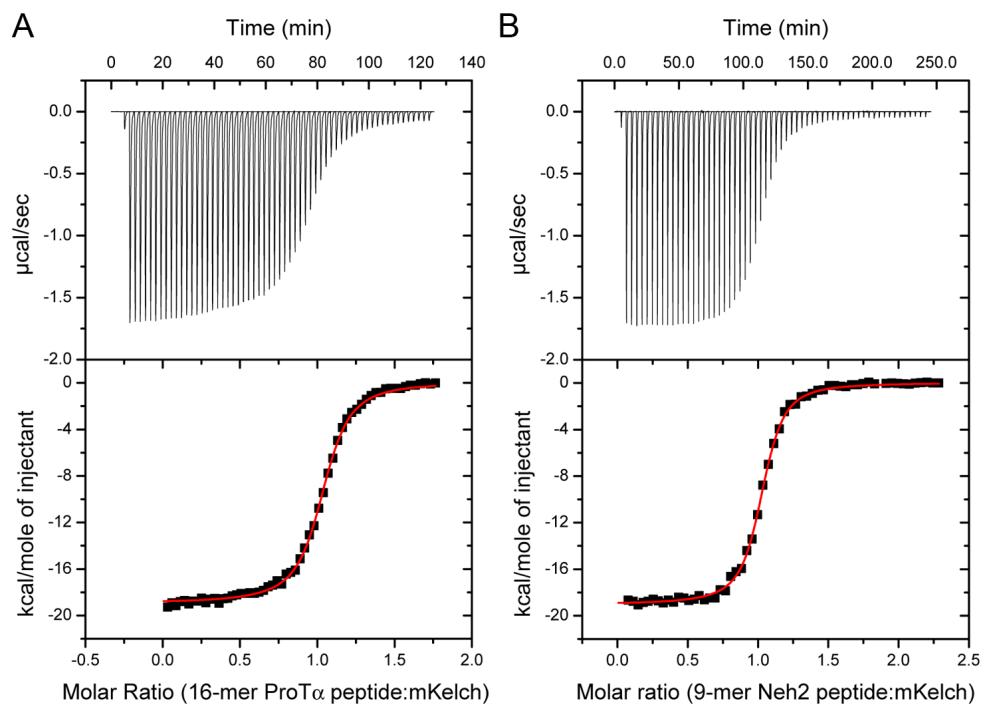
Figure S3.4 Isothermal titration calorimetry (ITC) measurements.

Panels A and B correspond to titrations of 16-mer ProTα and 9-mer Neh2 peptide to the mouse Kelch domain of Keap1, respectively. (Upper) The raw data of two ITC experiments each performed at 25° C. (Lower) The integrated heat changes, corrected for the heat of dilution, and the fitted curve assuming single-site binding.

Hydrogen bond analysis was conducted to help explain why the β-turns of ProTα and Neh2 converged to their bound state structures to different extents. Inspection of the MD structures from the last 0.1 μs of the simulations reveal that ProTα and Neh2 had different occurrence frequencies of hydrogen bonds within their Keap1-binding β-turns (Table 3.4). ProTα had at least one hydrogen bond present in only 14.3% of the structures, compared to a frequency of 98.6% for Neh2 (Table 3.4). The main differences arose from increased *i* to *i+3* and, to a lesser extent, *i* to *i+2* intra-turn hydrogen bonding in Neh2 compared to ProTα. For instance, hydrogen bonding between the side chains of Asp-77 and Thr-80 was observed in ~80% of the Neh2 structures, while the

corresponding side chain hydrogen bonding between Asn-41 and Asn-44 of ProTα was not observed in the MD trajectory (Table 3.4). The involvement of Thr-80 in intra-turn hydrogen bonds may explain why its χ_1 angle closely resembled its bound state value (Figure S3). Furthermore, hydrogen bonding between the side chain of Asn-41 and the main chain of Asn-44 was observed in only 6.7% of the ProTα conformations, while, in the Neh2 trajectory, over 77.5% of the conformations were found to have hydrogen bonding between the side chain of Asp-77 and the main chain of Thr-80 (Table 3.4). In addition, the side chain of Asp-77 and the main chain of Glu-79 in Neh2 also form hydrogen bond more frequently compared to the corresponding residues in ProTα (55% vs 0.4%) (Table 3.4). The result of this analysis suggested that the greater number and more frequent intra-turn hydrogen bonds formed by Neh2, particularly between the i and $i+3$ residues, may explain why it adopts more stable bound-state-like structure than ProTα. Interestingly, this finding qualitatively agrees with the difference in the residue-specific turn potentials for the β-turns of ProTα and Neh2. Using a table of overall turn potentials for each amino acid determined by Hutchinson & Thornton (80), the turn potentials for residues in the i to $i+3$ positions were summed. The NEEN and DEET sequences of Neh2 and ProTα had turn potentials of 4.87 and 5.03 respectively. The lower value for ProTα compared to Neh2 arose partly due to asparagine being slightly disfavored in position i compared to aspartic acid, but mainly because threonine was considerably more favored in position $i+3$ than asparagine.

Table 3.4 Frequencies of intra-turn hydrogen bond formations.

Atom involved	Full-length ProTα[a]	32-mer Neh2[a]
mc[b] i to mc i+2		0.196%
mc i to mc i+3		3.844%
mc i to sc i+3		27.808%
mc i+2 to sc i+3	0.284%	0.204%
sc[c] i to mc i+2	0.396%	55.368%
sc i to mc i+3	6.696%	77.524%
sc i+1 to mc i+2		0.428%
sc i to sc i+3		80.212%
sc i+2 to sc i+3	7.316%	0.572%
Intra-turn total	14.348%	98.644%

[a] Each frame from the last 0.1 μs of the mouse full-length ProTα and 32-mer Neh2 trajectories were used for the hydrogen bond calculations (25,000 structures). A hydrogen bond between a hydrogen donor (D-H) and a hydrogen acceptor (A) was judged to be formed when the DA distance (r) was less than 3.2 Å and the angle between the DA vector and the D-H bond (AD-H angle) was less than 35°. For clarity, only hydrogen bonds occur in more than 0.1% of the structures are listed and intra-residue hydrogen bonds are excluded.
[b] mc – main chain atoms
[c] sc – side chain atoms

As shown in Table 3.4, a large fraction of the intra-turn hydrogen bonds formed by Neh2 involve Thr-80. Studies reveal that mutating Thr-80 of Neh2 to alanine disrupts the interaction between these two proteins, making Nrf2 resistant to Keap1 mediated degradation. In contrast, a T80S mutant, which has the side chain hydroxyl group retained, behaved similarly to the wild type (79). Interestingly, the phosphorylation of Thr-80 has also been shown to severely decrease binding of Neh2 to Keap1 (79). The authors suggested that the negative charge introduced by the phosphorylation may disrupt the β-turn formation, preventing Neh2 from adopting a complementary structure to the binding site of Keap1 (79).

The attenuation of Keap1 binding when Thr-80 is mutated to alanine is likely due to the disruption of the β-turn structure. This idea is reinforced by our findings, which showed that the side chain of this residue is involved in the majority of intra-turn hydrogen bonds in the free state (Table 3.4). Moreover, residue-specific turn potential calculation also indicates that when the DEET sequence of Nrf2 is mutated to DEEA, the

turn potential falls below that of the ProTα sequence to 4.72. Therefore, Thr-80 may act as a function switch, allowing Nrf2 activity to be regulated efficiently by phosphorylation (79, 81).

Comparison of the mouse and human simulations

Finally, MD simulations were performed on the human homologs of full-length ProTα and the 32-mer Neh2 peptide (Table 3.1). The sequence alignments (Figure S3.5) indicate that there is a large percentage of sequence identity between the human and mouse versions of ProTα and Neh2. The human isoform 2 of ProTα used in this study contains 110 residues, which is shorter than the corresponding mouse sequence by one amino acid. The deletion site is located near the Keap1 binding region, immediately before the NEEN sequence. Besides the deletion, the human and mouse ProTα sequences are differ in only 5 other positions (Figure S3.5). For the 32-mer Neh2, there are three substitutions in the human sequence; one of them is located three residues upstream of the DEET ß-turn. The MD simulations of human ProTα and Neh2 therefore serve as pseudo duplicates of the mouse trajectories owing to the high sequence identities between the human and mouse versions of these two proteins. Moreover, the single-residue changes (deletion in ProTα and substitution in Neh2) close to the ß-turn sequences also allowed us to gauge the effects of mutations on the simulations.

```
Full-length    MSDAAVDTSSEITTKDLKEKKEVVEEAENGRDAPANGNAQNEENGEQEAD
ProTα          MSDAAVDTSSEITTKDLKEKKEVVEEAENGRDAPANGNAN-EENGEQEAD
(mouse)        **************************************: *********
Full-length    NEVDEEEEEGGEEEEEEEEGDGEEEDGDEDEEAEAPTGKRVAEDDEDDDV
ProTα          NEVDEEEEEGGEEEEEEEEGDGEEEDGDEDEEAESATGKRAAEDDEDDDV
(human         **********************************:.****.*********
isoform 2)     DTKKQKTEEDD
               DTKKQKTDEDD
               *******:***

32-mer         AFFAQFQLDEETGEFLPIQPAQHIQTDTSGSA
Neh2           AFFAQLQLDEETGEFLPIQPAQHIQSETSGSA
peptide        *****:****************::*****
(mouse)
32-mer
Neh2
peptide
(human
isoform 1)
```

Figure S3.5 Sequence alignments of the mouse and human full-length ProTα and 32-mer Neh2 constructs generated using ClustalW XXL.

The Blosum scoring matrix was used and gap penalties were set at their default values. Opening and end gap penalties were set to 10. Extending and separation gap penalties were set to 0.05.

The structure of a 16-mer human Neh2 peptide bound to human Keap1 (PDB id: 2FLU) (79) was compared to the structure of mouse Neh2-Keap1. Average distance-based RMSD calculations show that the residues comprising the β-turns in human and mouse Neh2 peptides adopt almost identical structures, with a backbone RMSD less than 0.1 Å in the bound-states (55, 79). For ProTα, the crystal structure of human ProTα-Keap1 was not currently available. Therefore, for consistency, in the following calculations, we continued to use the mouse structures (PDB ids: 2Z32 and 1X2R) (50, 55) as the bound-state references for the human MD data.

Due to the intrinsically disordered nature of ProTα and Neh2, the initial structures used for the simulations are not well-defined. To avoid the potential bias of conformational sampling, starting structures used in the MD simulations of human ProTα and Neh2 were different from that used for the mouse. Considering the residues

comprising the β-turns, the all-atom RMSDs between starting structures for the human and mouse sequences were 2.41 Å and 2.48 Å for the full-length ProTα proteins and the 32-mer Neh2 peptides, respectively.

Like the mouse versions, the β-turns of the human full-length ProTα and the 32-mer Neh2 peptide also converged to bound-state-like structures by the end of the trajectories (Figure S3.6). In the last 0.1 μs of the trajectories, both ProTα and Neh2 had C_α^i-C_α^{i+3} distance deviations around 1 Å from their mouse bound-state distances, with Neh2 having slightly closer C_α^i-C_α^{i+3} contacts (Figure S3.7). Interestingly, the 32-mer human Neh2 peptide adopted structures with about the same all-atom RMSDs to the bound-state conformation after a similar amount of simulation time compared to the mouse version (Figure S3.6 and Figure 3.2B). Meanwhile, the human ProTα was able to adopt structures with a lower all-atom RMSD to its bound state compared to its mouse counterpart (Figure S3.6 and Figure 3.2B).
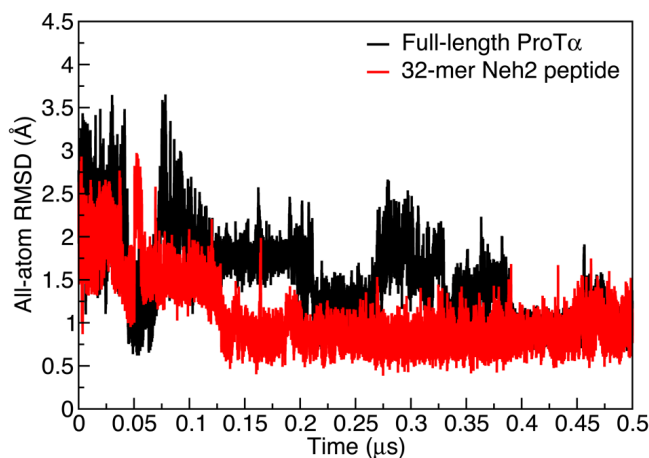
Figure S3.6 All-atom RMSD values between the MD and crystal structures.

The RMSD values were computed by subtracting the all-atom distance matrix at time *t* of the MD trajectories from the reference distance matrix determined from the crystal structures of the ProTα and Neh2 peptides bound to Keap1 (PDB ids: 2Z32 and 1X2R respectively) (50, 55). The distance matrices consisted of residues *i* through *i*+3 of the β-turn regions of the ProTα and Neh2 peptides determined from the crystal structures (50, 55).

Figure S3.7 C_α^i-C_α^{i+3} distances and their deviations from their crystal structure distances.

Panels A and B show the C_α^i-C_α^{i+3} distances and the deviations from the corresponding distances in the crystal structures respectively. Data was collected over the last 0.1 μs of the full-length human ProTα and human 32-mer Neh2 trajectories. Deviations were calculated for C_α^i-C_α^{i+3} pairs from the β-turns, determined from the mouse crystal structures (50, 55), by subtraction of the i to $i+3$ distance at time t of the trajectory from the fixed distance of the corresponding atom pair from the crystal structures (PDB ids: 2Z32 and 1X2R) for ProTα and Neh2 respectively) (50, 55).

The hydrogen bond analysis showed that, like the mouse homolog, human ProTα formed *i* to *i+3* hydrogen bonds less frequently compared to Neh2 (Table S3.1). For instance, hydrogen bonding between the side chains of Asn-40 and Asn-43 was observed in 24.2% of the ProTα structures compared to 63.7% for the corresponding Asp-77 and Thr-80 pair in Neh2 (Table S3.1). Furthermore, hydrogen bonding between the side chain of Asn-40 and the main chain of Asn-43 was observed in 61.3% of the ProTα conformations compared to 74.5% for the corresponding Asp-77 and Thr-80 pair in Neh2 (Table S3.1). The results from the human systems reinforce the notion that *i* to *i+3* hydrogen bonding between Asp-77 and Thr-80 of Neh2 might be more preferable than the corresponding Asn pair in ProTα.

Table S3.1 Frequencies of intra-turn hydrogen bond formations

Atoms involved	Full-length ProTα[a]	32-mer Neh2[a]
mc i to mc i+2	85.124%	0.16%
mc i to mc i+3		1.744%
mc i to sc i+3		15.884%
mc i+1 to sc i		1.708%
mc i+1 to sc i+2		0.56%
sc i to mc i+2	6.828%	49.164%
sc i to mc i+3	61.272%	74.504%
sc i+1 to mc i+2		0.328%
sc i+2 to mc i+3	0.112%	
sc i to sc i+3	24.172%	63.74%
sc i+2 to sc i+3	0.14%	0.536%
Intra-turn total	96.844%	95.248%

[a] Each frame from the last 0.1 μs of the human full-length ProTα and 32-mer Neh2 trajectories were used for the hydrogen bond calculations (25 000 structures). A hydrogen bond between a hydrogen donor (D-H) and a hydrogen acceptor (A) was judged to be formed when the DA distance (r) was less than 3.2 Å and the angle between the DA vector and the D-H bond (AD-H angle) was less than 35°. For clarity, only hydrogen bonds occur in more than 0.1% of the structures are listed and intra-residue hydrogen bonds are excluded.
[b] mc – main chain atoms
[c] sc – side chain atoms

Unlike the high similarities between the simulations of the mouse and the human Neh2, the intra-turn hydrogen bonding patterns of the human and mouse versions of

ProTα were less consistent (Tables 3.4 and S3.1). Higher occurrences of hydrogen bonding between the main chains of *i* and *i+2* residues, as well as between the side chain and main chain of *i* and *i+3* were found in human ProTα. We speculate that the discrepancies reflect lower simulation convergence due to the less restricted conformation sampling of free-state ProTα (78). However, the influence of starting structures and sequence differences cannot be ruled out. Further experimental studies are required to validate these findings.

3.5 Conclusion

In this work we have investigated how ProTα and Neh2 interact with a common binding partner, the Kelch domain of Keap1 using 0.5-1.0 μs MD simulations. Our main findings are that the XEEXGE Keap1 binding motifs of ProTα and Neh2 in their free states possess propensities to form bound-state-like structure to different extents. Neh2 was found to form a defined β-turn more frequently than ProTα and had lower RMSD to its bound state conformation. This may be attributed to a larger number of and more stable intra-turn hydrogen bonds. In particular, hydrogen bonding between Asp-77 and Thr-80 of Neh2 might be more preferable than the corresponding Asn pair in ProTα. However, we cannot rule out that other factors, such as the lack of comparable hydrophobic content surrounding the Keap1 binding region of ProTα. This may also contribute to the more dynamic nature of ProTα and its lower propensity for adopting bound-state-like conformations.

Addressing whether ProTα and Neh2 bind to Keap1 through PSEs, coupled folding and binding or a combination of both mechanisms was challenging. To conclude that binding occurs via PSEs, the free and bound state conformations would have to be highly similar or identical. The definition of being highly similar can be ambiguous, while restricting the definition to identical structures seems too stringent. In any protein-protein interaction there are likely to be a certain amount of structural changes upon binding. In this case, the backbone atoms of the β-turns overlay well with the crystal structure backbones, especially for Neh2. However, the side chain orientations of some

residues show considerable differences. It is clear that both mechanisms are at work to different extents. Because our data shows that the Keap1 binding regions of ProTα and Neh2 tend to form β-turns that have an obvious resemblance to their bound state conformations, we propose that binding occurs synergistically via a combination of PSEs and coupled folding and binding with a heavy bias towards PSEs, especially for Neh2.

3.6 Acknowledgements

3.7 References

1. Uversky VN (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* 11:739-56.

2. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6:197-208.

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

4. Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* 293:321-31.

5. Dunker AK et al. (1998) Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pac Symp Biocomput* 473-84.

6. Oldfield CJ et al. (2005) Comparing and combining predictors of mostly disordered proteins. *Biochemistry* 44:1989-2000.

7. Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ (2000) Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform* 11:161-71.

8. Vucetic S, Brown CJ, Dunker AK, Obradovic Z (2003) Flavors of protein disorder. *Proteins* 52:573-84.

9. Radivojac P et al. (2007) Intrinsic disorder and functional proteomics. *Biophys J* 92:1439-56.

10. Gall C, Xu H, Brickenden A, Ai X, Choy WY (2007) The intrinsically disordered TC-1 interacts with Chibby via regions with high helical propensity. *Protein Sci* 16:2510-8.

11. Choy WY, Forman-Kay JD (2001) Calculation of ensembles of structures representing the unfolded state of an SH3 domain. *J Mol Biol* 308:1011-32.

12. Morar AS, Olteanu A, Young GB, Pielak GJ (2001) Solvent-induced collapse of alpha-synuclein and acid-denatured cytochrome c. *Protein Sci* 10:2195-9.

13. Shortle D, Ackerman MS (2001) Persistence of native-like topology in a denatured protein in 8 M urea. *Science* 293:487-9.

14. Eliezer D, Kutluay E, Bussell R, Browne G (2001) Conformational properties of alpha-synuclein in its free and lipid-associated states. *J Mol Biol* 307:1061-73.

15. Demarest SJ et al. (2002) Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* 415:549-53.

16. Fuxreiter M, Simon I, Friedrich P, Tompa P (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J Mol Biol* 338:1015-26.

17. Sivakolundu SG, Bashford D, Kriwacki RW (2005) Disordered p27Kip1 exhibits intrinsic structure resembling the Cdk2/cyclin A-bound conformation. *J Mol Biol* 353:1118-28.

18. Mohan A et al. (2006) Analysis of molecular recognition features (MoRFs). *J Mol Biol* 362:1043-59.

19. Baker JM et al. (2007) CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. *Nat Struct Mol Biol* 14:738-45.

20. Mokhtarzada S, Yu C, Brickenden A, Choy WY (2011) Structural characterization of partially disordered human Chibby: insights into its function in the Wnt-signaling pathway. *Biochemistry* 50:715-26.

21. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J* 272:5129-48.

22. Vacic V et al. (2007) Characterization of molecular recognition features, MoRFs, and their binding partners. *J Proteome Res* 6:2351-66.

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

24. Patil A, Nakamura H (2006) Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. *FEBS Lett* 580:2041-5.

25. Dosztányi Z, Chen J, Dunker AK, Simon I, Tompa P (2006) Disorder and sequence repeats in hub proteins and their implications for network evolution. *J Proteome Res* 5:2985-95.

26. Haynes C et al. (2006) Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput Biol* 2:e100.

27. Iakoucheva LM, Brown CJ, Lawson JD, Obradović Z, Dunker AK (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol* 323:573-84.

28. Cheng Y, LeGall T, Oldfield CJ, Dunker AK, Uversky VN (2006) Abundance of intrinsic disorder in protein associated with cardiovascular disease. *Biochemistry* 45:10448-60.

29. Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys* 37:215-46.

30. Wang Y et al. (2011) Intrinsic disorder mediates the diverse regulatory functions of the Cdk inhibitor p21. *Nat Chem Biol* 7:214-21.

31. Dyson HJ, Wright PE (2002) Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* 12:54-60.

32. Sugase K, Dyson HJ, Wright PE (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 447:1021-5.

33. Wright PE, Dyson HJ (2009) Linking folding and binding. *Curr Opin Struct Biol* 19:31-8.

34. Tsai CJ, Ma B, Sham YY, Kumar S, Nussinov R (2001) Structured disorder and conformational selection. *Proteins* 44:418-27.

35. Kumar S, Ma B, Tsai CJ, Sinha N, Nussinov R (2000) Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci* 9:10-9.

36. Espinoza-Fonseca LM (2009) Reconciling binding mechanisms of intrinsically disordered proteins. *Biochem Biophys Res Commun* 382:479-82.

37. Salma P, Chintan C, Sriram S (2009) Intrinsically Unstructured Proteins: Potential Targets for Drug Discovery *Am J Infect Dis* 5:126-134.

38. Cheng Y et al. (2006) Rational drug design via intrinsically disordered protein. *Trends Biotechnol* 24:435-42.

39. Mittag T, Forman-Kay JD (2007) Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol* 17:3-14.

40. Eliezer D (2009) Biophysical characterization of intrinsically disordered proteins. *Curr Opin Struct Biol* 19:23-30.

41. Chen H-FF (2009) Molecular dynamics simulation of phosphorylated KID post-translational modification. *PLoS One* 4:e6516.

42. Wu KP, Weinstock DS, Narayanan C, Levy RM, Baum J (2009) Structural reorganization of alpha-synuclein at low pH observed by NMR and REMD simulations. *J Mol Biol* 391:784-96.

43. Veurink G, Fuller SJ, Atwood CS, Martins RN (2003) Genetics, lifestyle and the roles of amyloid beta and oxidative stress in Alzheimer's disease. *Ann Hum Biol* 30:639-67.

44. Itoh K et al. (1997) An Nrf2/small Maf heterodimer mediates the induction of phase II detoxifying enzyme genes through antioxidant response elements *Biochem Biophys Res Commun* 236:313-322.

45. Itoh K et al. (1999) Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain. *Genes Dev* 13:76-86.

46. Tong KI et al. (2006) Keap1 recruits Neh2 through binding to ETGE and DLG motifs: characterization of the two-site molecular recognition model. *Mol Cell Biol* 26:2887-900.

47. Itoh K et al. (2003) Keap1 regulates both cytoplasmic-nuclear shuttling and degradation of Nrf2 in response to electrophiles. *Genes Cells* 8:379-91.

48. Itoh K, Tong KI, Yamamoto M (2004) Molecular mechanism activating Nrf2-Keap1 pathway in regulation of adaptive response to electrophiles. *Free Radic Biol Med* 36:1208-13.

49. Karapetian RN et al. (2005) Nuclear oncoprotein prothymosin alpha is a partner of Keap1: implications for expression of oxidative stress-protecting genes. *Mol Cell Biol* 25:1089-99.

50. Padmanabhan P, Nakamura N, Yokoyama Y (2008) *Acta. Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* 64:233.

51. Haritos AA, Goodall GJ, Horecker BL (1984) Prothymosin alpha: isolation and properties of the major immunoreactive form of thymosin alpha 1 in rat thymus. *Proc Natl Acad Sci U S A* 81:1008-11.

52. Piñeiro A, Cordero OJ, Nogueira M (2000) Fifteen years of prothymosin alpha: contradictory past and new horizons. *Peptides* 21:1433-46.

53. Jiang X et al. (2003) Distinctive roles of PHAP proteins and prothymosin-alpha in a death regulatory pathway. *Science* 299:223-6.

54. Brunger AT (2007) Version 1.2 of the Crystallography and NMR system. *Nat Protoc* 2:2728-33.

55. Padmanabhan B et al. (2006) Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer. *Mol Cell* 21:689-700.

56. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation *J Chem Theory Comput* 4:435-447.

57. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6 *J Chem Theory Comput* 25:1656-1676.

58. Oostenbrink C, Soares TA, van der Vegt NF, van Gunsteren WF (2005) Validation of the 53A6 GROMOS force field. *Eur Biophys J* 34:273-84.

59. Huang W, Lin Z, van Gunsteren WF (2011) Validation of the GROMOS 54A7 Force Field with Respect to β-Peptide Folding *J Chem Theory Comput* 7:1237-1243.

60. Berendsen HJC, Postma JPM, and Gusteren WF (1981) in *Intermolecular forces*, p 331-342.

61. Patra M, Karttunen M (2004) Systematic comparison of force fields for microscopic simulations of NaCl in aqueous solutions: diffusion, free energy of hydration, and structural properties. *J Comput Chem* 25:678-89.

62. Berendsen HJC, Postma JPM, Van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath *J Chem Phys* 81:3684-3690.

63. Hess B, Bekker H, Berendsen HJC, Johannes JGEM (1997) LINCS: A linear constraint solver for molecular simulations *J Comput Chem* 18:1463-1472.

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

65. Patra M et al. (2003) Molecular dynamics simulations of lipid bilayers: major artifacts due to truncating electrostatic interactions. *Biophys J* 84:3636-45.

66. Karttunen M, Rottler J, Vattulainen I, Sagui C (2008) Electrostatics in Biomolecular Simulations: Where Are We Now and Where Are We Heading? *Current Topics in Membranes* 60:49-89.

67. Wong-ekkabut J, Miettinen MS, Dias C, Karttunen M (2010) Static charges cannot drive a continuous flow of water molecules through a carbon nanotube. *Nat Nanotechnol* 5:555-7.

68. Baumketner A (2009) Removing systematic errors in interionic potentials of mean force computed in molecular simulations using reaction-field-based electrostatics. *J Chem Phys* 130:104106.

69. Ni B, Baumketner A (2011) Effect of atom- and group-based truncations on biomolecules simulated with reaction-field electrostatics. *J Mol Model* 17:2889-93

70. Kaszuba K, Róg T, Bryl K, Vattulainen I, Karttunen M (2010) Molecular dynamics simulations reveal fundamental role of water as factor determining affinity of binding of beta-blocker nebivolol to beta(2)-adrenergic receptor. *J Phys Chem B* 114:8374-86.

71. Azzopardi PV et al. (2010) Roles of electrostatics and conformation in protein-crystal interactions. *PLoS One* 5:e9330.

72. Lewis PN, Momany FA, Scheraga HA (1973) Chain reversals in proteins. *Biochim Biophys Acta* 303:211-29.

73. MacArthur MW, Thornton JM (1993) Conformational analysis of protein structures derived from NMR data. *Proteins* 17:232-51.

74. Murzyn K, Zhao W, Karttunen M, Kurdziel M, Róg T (2006) Dynamics of water at membrane surfaces: Effect of headgroup structure. *Biointerphases* 1:98-105.

75. Baker EN, Hubbard RE (1984) Hydrogen bonding in globular proteins *Prog Biophys Mol Biol* 44:97-179.

76. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14:33-8, 27-8.

77. Pettersen EF et al. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605-12.

78. Yi S, Boys BL, Brickenden A, Konermann L, Choy WY (2007) Effects of zinc binding on the structure and dynamics of the intrinsically disordered protein prothymosin alpha: evidence for metalation as an entropic switch. *Biochemistry* 46:13120-30.

79. Lo SC, Hannink M (2006) PGAM5, a Bcl-XL-interacting protein, is a novel substrate for the redox-regulated Keap1-dependent ubiquitin ligase complex. *J Biol Chem* 281:37893-903.

80. Hutchinson EG, Thornton JM (1994) A revised set of potentials for beta-turn formation in proteins. *Protein Sci* 3:2207-16.

81. Bobrovnikova-Marjon E et al. (2010) PERK promotes cancer cell proliferation and tumor growth by limiting oxidative DNA damage. *Oncogene* 29:3881-95.

4   Binding of disordered proteins to a protein hub

Elio A. Cino[†], Ryan C. Killoran[†], Mikko Karttunen[‡*] and Wing-Yiu Choy[†*]

Department of Biochemistry[†], The University of Western Ontario, London, Ontario, Canada N6A 5C1, Department of Chemistry[‡], University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

*Corresponding author: jchoy4@uwo.ca, mikko.karttunen@uwaterloo.ca

4.2 Abstract

In the protein interactome, there are a small number of proteins with high connectivity. These proteins are commonly referred to as hubs, and are essential for interactome functionality and integrity. By combining experimental and computational approaches, we identified the Kelch domain of Keap1 as a crucial hub for binding intrinsically disordered proteins (IDPs) in oxidative stress response and apoptosis. Keap1 regulates transcriptional activity of nuclear factor erythroid 2-related factor 2 (NRF2), a master regulator of cytoprotective gene activation in the oxidative stress response pathway. Disorder predictions suggest that the majority of proteins that interact with the Kelch hub are intrinsically disordered in their binding regions. These targets share similar binding motifs, but have a wide range of binding affinities, spanning more than 2 orders of magnitude. Using nuclear magnetic resonance spectroscopy and molecular dynamic simulations, the major factors that govern the binding affinity and specificity of different disordered targets to the Kelch hub were determined. Good correlations ($r^2 > 0.75$) between the binding free-energy and protein dynamics and hydrophobicity were found. The results indicate that the binding affinity of different disordered targets to the Kelch hub is largely determined by the extent of preformed bound-state like conformation in their free-state structures. Based on the knowledge acquired, we have designed a high-binding affinity peptide that can specifically disrupt the Keap1-NRF2 interaction and has the potential for therapeutic applications. Overall, the work demonstrates a simple yet effective methodology for structurally characterizing IDP-protein interactions.

4.3 Introduction

Protein-protein interactions are the foundation upon which cells carry out their functions. Characterizing the interaction network is crucial for determining the functions of individual proteins, and identifying signaling pathways and their interconnections (1-6). Some proteins have much higher connectivity relative to the others. These proteins are referred to as hubs (1, 7-11). It is common for a protein to interact with a few targets but hubs, which may have tens to hundreds of network connections, are rare in comparison (6, 7, 12). Their ability to interact with numerous partners is thought to be essential for network functionality and stability. Because of the critical roles they play, disruption of hub interactions in the human protein interaction network are frequently associated with diseases (5, 13, 14).

The structural plasticity of IDPs is a crucial feature allowing them to interact with numerous different partners (15-19). Therefore, these proteins are frequently found to function as hubs. Interestingly, there are also examples of ordered hubs that interact almost exclusively with disordered partners (15, 18, 20-22). In contrast to protein-protein interactions between globular proteins, which usually involve larger interaction surfaces that are discontinuous in sequence, IDPs typically bind to ordered hubs using short (~6 residues) consecutive stretches of amino acid residues called linear motifs (LMs) (23-27). For example, the 14-3-3 family is one of the most studied hubs for IDPs. These are ~30 kDa dimeric proteins that adopt rigid structures capable of binding hundreds of ligands, which together may comprise 0.6% of the human proteome (18, 20, 22, 28-30). These interactions are involved in important cellular functions such as catalysis, regulation and localization (22, 29, 30). It is estimated that over 90% of the 14-3-3 binding partners are completely or partially disordered (20, 22). Binding is governed by phosphoserine and phosphothreonine containing motifs on the partners. Several distinct consensus motifs (modes), with varying binding affinities, have been discovered across the broad spectrum of 14-3-3 ligands (28, 29). In addition to phosphorylation of 14-3-3 binding motifs, the presence of structurally constrained anchor residues outside of the motifs are thought to play important roles in the stability and specificity of the interactions (31). Another established hub with a preference for binding disordered partners is the N-terminal β-

propeller domain (TD) domain of clathrin. Interactions between this hub and several disordered partners, including amphiphysins, AP180 and SNX9 have been demonstrated (32, 33). Using NMR, AP180 was found to be largely unstructured in free and clathrin bound states. Residues comprising the two clathrin binding motifs of AP180 had limited β-turn structures in their free and bound states and are more restricted in dynamics relative to their surrounding regions (33). Binding partners that remain dynamic in their free and bound states have been described to form 'fuzzy' complexes. Such complexes are thought to be important where specific, yet reversible binding is required (34-36).

Because LMs along an IDP sequence often represent potential target binding motifs, there have been considerable efforts towards identifying them (19, 23, 24, 37). The majority of work has been to detect motifs from sequences. Linear motifs often have distinct sequence characteristics compared to their surroundings, with the primary difference being an increased hydrophobic content (38), which may promote local structure formations. Such preformed structural elements often resemble their bound state conformations (39-42). A major hindrance in accurately detecting LMs with preformed structures or conformational propensity is a lack of free state structure data. Such information would provide unique insights into the relationships between sequence and structure, which could be used to verify and improve bioinformatics predictions of LMs. Another impeding factor in identifying, characterizing and comparing LMs is inconsistency in methods. For instance, to determine binding affinities and bound state structures, peptides containing the LM are frequently used. This is reasonable for IDPs because binding is typically a local event. However, because different binding partners for a given hub are typically discovered by different groups, there is often a lack of consistency in peptide lengths, alignments of LMs, organisms, buffers, methods, etc, which can be problematic if one wants to compare all binding partners simultaneously. While there are clear benefits of combining experimental data and predictive methods to predict and characterize LMs, variability in methods, experimental techniques can make it difficult to perform fair comparisons.

In this study, we have used bioinformatics tools for disorder prediction, isothermal titration calorimetry (ITC), nuclear magnetic resonance (NMR) spectroscopy, and molecular dynamic (MD) simulations to systematically characterize the interactions between Kelch-like ECH-associated protein 1 (Keap1), an ordered hub protein, and all of its known (to date) disordered partners that bind to the Kelch domain of the protein (Figure 4.1A). Keap1/NRF2 pathway is essential for regulating the cellular response to oxidative stress (43, 44). The Kelch domain is a ~32 kDa β-propeller located near the C-terminus of Keap1. To date, 9 different proteins have been shown to interact with it and most of them share similar 'GE'-containing motifs (Figure 4.1A). Scrutinizing the mechanisms they employ to interact with the Kelch domain is essential for understanding their functions. It will also provide insight into how the Kelch domain tasks as a hub for disordered proteins in the oxidative stress response, and possibly other biological processes. Our study illustrates that by using a systematic approach based on a combination of experimental and computational methods, one can obtain unique insights into factors regulating the affinities and specificities of interactions of disordered partners with ordered hub proteins. Our approach should be generally applicable to investigate the binding mechanisms of other systems involving disordered LMs.

A

NRF2 site 1	[71]FAQLQL	DEETGE	FLPIQPAQ[90]
PGAM5	[71]INVRKR	NVESGE	EELASKLD[90]
p62	[341]LSSKEV	DPSTGE	LQSLQMPE[360]
WTX	[280]SLEEPH	SPETGE	KVVAGEVN[299]
FAC1	[584]SAKAAD	DPENGE	RESHTPVS[603]
PALB2	[83]HIKTHL	DEETGE	KTSITLDV[102]
PTMA iso 1	[35]ANGNAE	NEENGE	QEADNEVD[54]
PTMA iso 2	[35]ANGNA-	NEENGE	QEADNEVD[53]
IKKβ	[28]NVIRWH	NQETGE	QIAIKQCR[47]

NRF2 site 2	[17]MDLIDIL	WRQD	IDLGVSREV[36]
BCL2	[181]LNRHLHT	WIQD	NGGWDAFVE[200]

B

Figure 4.1 Sequence analysis of the Kelch domain interacting proteins.

Sequence analysis of the Kelch domain interacting proteins. A) Manual sequence alignment and sequence logos of the site 1- and site 2-type regions of the Kelch domain binding proteins (91-93). B) Residue type fractions of the sequences. A,

F, I, L, M, P, V and W are hydrophobic; C, G, N, Q, S and T are polar; D and E are acidic; K, H and R are basic

4.4 Materials and methods

Expression and purification of the human Kelch domain

The Kelch domain of human Keap1 (residues 321-609), subcloned into the pET15b expression vector (from Dr. Mark Hannink, University of Missouri-Columbia), was expressed as an N-terminally His-tagged protein in *Escherichia coli* BL21 (DE3) and grown in minimal M9 medium. Protein expression was induced by addition of 0.5 mM IPTG at 18ºC for 24 h. The protein was purified from the crude cell lysate by affinity chromatography using Ni Sepharose™ 6 Fast Flow beads (Amersham Biosciences). The His-tag was then cleaved by incubation with human α-thrombin (Haematologic Technologies Inc.) overnight at 4ºC. The Kelch domain was purified from the cleavage mixture using a HiLoad Superdex-75 size-exclusion column (GE Healthcare) equilibrated with 50 mM sodium phosphate buffer, 100 mM NaCl, 1 mM DTT at pH 7.

Kelch domain interacting peptides

Peptides from the various site 1-type Kelch domain interacting proteins (Figure 4.1A) were synthesized (Tufts peptide synthesis). The sequences were manually aligned at their Kelch domain binding motifs, and extended on either side, up to 20 residues (Figure 4.1A). Because the sequences were from the interior regions of their respective proteins, acetyl and NH_2 groups were added on their N- and C-termini, respectively. IKKβ peptide was excluded from this study because its Kelch domain interacting region is located in a structured region of the protein (65). A WTX peptide with S286 phosphorylated was also synthesized. Phosphorylation of this residue has been demonstrated to occur *in vivo* (45). Two peptides were synthesized for PTMA, one for each isoform. The only difference between the two isoforms (in the full length proteins)

is the absence of E40 in isoform 2 (Figure 4.1A); therefore, the PTMA isoform 2 peptide was only 19 residues long. The peptides were dialyzed into 50 mM sodium phosphate buffer at pH 7 containing 100 mM NaCl and 1 mM DTT before use.

Isothermal titration calorimetry (ITC) experiments

ITC experiments were carried out on a VP-ITC instrument (MicroCal) at 25ºC. The protein and peptide samples were dialyzed into a buffer containing 50 mM sodium phosphate, 100 mM NaCl, 1 mM DTT at pH 7 and degassed before the experiments. ~40 μM Kelch was added to the 1.4 mL sample cell and subjected to stepwise titration with 5 μL aliquots of ~500 μM peptide. The equilibration period between each injection was 300 seconds. The association constant (K_a), molar binding stoichiometry (n) and the binding enthalpy (ΔH), entropy (ΔS) and Gibbs free energy (ΔG) were determined by fitting the binding isotherm to a single-binding-site model with Origin7 software (MicroCal). The heat changes after saturation were averaged and used to correct for the heats of dilution. All ITC experiments were performed in duplicate.

NMR spectroscopy

NMR experiments were performed on a 600 MHz Varian INOVA spectrometer (UWO Biomolecular NMR Facility). The peptide samples were prepared at a concentration of ~3 mM in a buffer containing 50 mM sodium phosphate, 100 mM NaCl, 1 mM DTT at pH 7.0. All samples contained 10% D_2O and trace 2,2-dimethyl-2-sila-pentane-5-sulfonic acid (DSS) for ^1H chemical shift referencing. The experiments were conducted at either 25 or 10ºC. Two-dimensional ^1H-^1H TOCSY and NOESY spectra were collected with spectral widths of 8000 Hz in both dimensions. WATERGATE solvent suppression was employed (82). TOCSY spectra were recorded with a 50 ms spin lock time. NOESY spectra were collected with a mixing time of 200 ms. The TOCSY and NOESY experiments were acquired using 4096 points with 256 increments. Sequential assignments for each peptide were obtained using the TOCSY and NOESY

spectra along with a natural abundance ^{1}H-^{15}N HSQC spectrum. All NMR data was processed with NMRPipe (83).

MD simulations

The amino acid sequences of all simulated peptides were the same as those used in the ITC experiments (Figure 4.1A and Table 4.1). We used the Crystallography & NMR System (CNS) (84) to generate an extended structure from each sequence. Simulated annealing was performed on each extended structure and resulting conformations that did not resemble the Neh2 domain site 1 region bound to the Kelch domain (PDB id: 2FLU) were used as starting structures (51). The N- and C-terminus of each structure was capped with acetyl (ACE) and NH$_2$ groups, respectively, using chimera (85). For the WTX peptide with S286 phosphorylated (WTX pS286), a dianionic phosphate group (PO$_4^{2-}$) was modeled onto S286 of the non-phosphorylated WTX peptide structure.

The MD simulations were performed using GROMACS (GROningen MAchine for Chemical Simulations) version 4.5 (86). All chemical species were represented by the GROMOS96 53a6 force field (87, 88), except in the WTX pS286 peptide simulation, where the GROMOS96 43a1p (89) force field was used. The starting structures were solvated in cubic boxes of linear size 6 nm with periodic boundary conditions applied in all directions. The SPC (simple point charge) water model was used (90). Protonation states of all ionizable residues were chosen based on their most probable state at pH 7. Histidine residues were protonated on ND1 only. Each system was neutralized and brought to an ionic strength of 0.1 M with sodium (Na$^+$) and chloride (Cl$^-$) ions. The simulations were performed using similar protocols as described in (39). System coordinates were written out at 4 ps intervals and each simulation was run for at least 1 μs, for a cumulative simulation time of 15 μs. The systems are summarized in Table 4.1.

Table 4.1 Summary of the MD simulations.

Protein	Sequence	Water & ions
NRF2 site 1	Ac-FAQLQLDEETGEFLPIQPAQ-NH$_2$	7007, 17 Na$^+$, 13 Cl$^-$
PGAM5	Ac-INVRKRNVESGEEELASKLD-NH$_2$	7019, 14 Na$^+$, 13 Cl$^-$
p62	Ac-LSSKEVDPSTGELQSLQMPE-NH$_2$	7012, 16 Na$^+$, 13 Cl$^-$
WTX	Ac-SLEEPHSPETGEKVVAGEVN-NH$_2$	7017, 17 Na$^+$, 13 Cl$^-$
WTX pS286	Ac-SLEEPHpSPETGEKVVAGEVN-NH$_2$	7010, 19 Na$^+$, 13 Cl$^-$
FAC1	Ac-SAKAADDPENGERESHTPVS-NH$_2$	7024, 16 Na$^+$, 13 Cl$^-$
PALB2	Ac-HIKTHLDEETGEKTSITLDV-NH$_2$	7019, 16 Na$^+$, 13 Cl$^-$
PTMA isoform 1	Ac-ANGNAENEENGEQEADNEVD-NH$_2$	7020, 21 Na$^+$, 13 Cl$^-$
PTMA isoform 2	Ac-ANGNANEENGEQEADNEVD-NH$_2$	7022, 20 Na$^+$, 13 Cl$^-$
IKKβ	Ac-NVIRWHNQETGEQIAIKQCR-NH$_2$	7005, 13 Na$^+$, 14 Cl$^-$
NRF2 site 2	Ac-MDLIDILWRQDIDLGVSREV-NH$_2$	7008, 16 Na$^+$, 13 Cl$^-$
BCL2	Ac-LNRHLHTWIQDNGGWDAFVE-NH$_2$	7001, 15 Na$^+$, 13 Cl$^-$
NRF2 site 1 E78P	Ac-FAQLQLDPETGEFLPIQPAQ-NH$_2$	7013, 16 Na$^+$, 13 Cl$^-$

4.5 Results and discussion

We have used a multidisciplinary approach involving bioinformatics, ITC, NMR and MD simulations to compare the sequences, binding parameters and free state structures of *all* Kelch domain interacting proteins identified to date. The study builds upon the findings from numerous investigations of Kelch domain interacting proteins (43, 45-59). While these protein-protein interactions have been identified, they have not all been thoroughly characterized and compared. Our study provides new insights into the relationship between sequence and structure in governing the binding affinity of the different targets to the hub protein Keap1. The findings will be useful in delineating the binding mechanisms of the different proteins – crucial information for understanding their biological roles. Importantly, our approach is generally applicable for investigations of the binding mechanisms of other systems involving disordered proteins.

Kelch domain interacting proteins are predominantly disordered

The Neh2 domain of NRF2 is responsible for mediating the interaction with the

Kelch domain of Keap1. Structural studies have shown that Neh2 is intrinsically disordered (57, 58). It has high and low affinity 'ETGE' and 'DLG' containing Kelch domain binding regions, respectively, which will be referred to as sites 1 and 2, respectively hereafter (57, 58). These sites are located on separate ends of the ~100 residue Neh2 domain, connected by a segment with high helical propensity (57). When both sites are bound to two separate Kelch domains, NRF2 is ubiquitinated, which targets it for proteasomal degradation (58). When only site 1 is bound, NRF2 avoids the degradation pathway and can promote expression of its target genes (58).

Several proteins have been shown to disrupt the low affinity site 2-Kelch domain interaction, allowing NRF2 to promote cytoprotective gene expression. These proteins include PGAM5, p62, WTX and PALB2 (45, 48, 50, 52). In addition, FAC1 (56), PTMA (54, 60), IKKβ (47, 49) and BCL2 (53) can also bind to the Kelch domain. PTMA is a highly disordered protein that functions as a vehicle for shuttling Keap1 into the nucleus (60-62). The purposes of the other protein-protein interactions are currently under intensive investigations (44, 63). Most of the Kelch domain interacting proteins contain sequences resembling the site 1 sequence of the Neh2 domain (Figure 4.1A) and will be referred to as site 1-type proteins hereafter. The one exception is BCL2, which contains a site 2-type sequence and will be referred to as a site 2-type protein (Figure 4.1A). Although not included in this study, because it does not appear to be directly linked to the oxidative stress response or apoptosis, myosin-VIIa has also been shown to interact with Kelch (64). It is worthwhile to mention that the region of human myosin-VIIa capable of binding Kelch includes a site 1 type sequence [1635]LDHDTGE[1641].

Structural information of only a few of the Kelch domain interacting proteins is currently available. Crystal structures of PTMA and p62 peptides in complex with the Kelch domain show that their site 1-type motifs bind to the same region as the NRF2 site 1 region and form similar β-hairpin structures in their bound states (48, 51, 54, 55). In its unbound state, the NRF2 site 1 region contains a short β-sheet structure, which is thought

to contribute to its higher binding affinity compared to the site 2 region (39, 40, 57). Our previous MD simulations showed that, in their free states, the NRF2 site 1 region and the site 1-type region of PTMA formed β-hairpin structures that resembled the bound state structure to different extents, with NRF2 forming a more defined hairpin that had a closer resemblance to its bound state structure compared to PTMA (40). Although the structure of a p62 peptide bound to the Kelch domain has been determined (48), its free state structure has not been examined and it is not known if it may also contain residual structure in its Kelch domain interacting region. Additionally, a homology structure of the IKKβ kinase domain, in the absence of the Kelch domain, illustrates that its site 1-type motif also adopts a β-hairpin conformation (65).

Based on the amino acids sequences of the various proteins (Figure 4.1A), we found that intrinsic disorder in the Kelch domain binding regions of the site 1-type binders may be a common attribute. NRF2 site 1 and PALB2 had PONDR-FIT (66) disorder tendencies of 0.50 or higher, while p62, WTX, FAC1 and PTMA had tendencies >0.70 (Figure S4.1). Furthermore, p62, WTX, FAC1 and PTMA are predicted to have long stretches of disorder around their binding regions (Figure S4.1). PGAM5 and IKKβ were the only site 1-type proteins with predicted disorder tendencies <0.5 (Figure S4.1). The binding region of PGAM5 (Figure 4.1A) had an average PONDR-FIT score of 0.36, but scored 0.59 when the metaPrDOS predictor was used (67). PONDR-FIT predicted that the N-terminal portion of PGAM5 is largely disordered, while the C-terminal is structured (Figure S4.1), suggesting that the Kelch domain binding region of PGAM5 is located on the border of a disordered region and structured domain. This notion is supported by the X-ray crystal structure of the C-terminal PGAM5 domain (PDB id: 3MXO), beginning at D90 (Figure 4.1A). The binding region of IKKβ is predicted to be in a structured part of the protein (Figure S4.1), which is consistent with the homology model (65) that illustrates that it is a well-folded protein (Figure 4.2). The site 2-type binders were predicted to have disorder tendencies of 0.40 for NRF2 and 0.07 for BCL2 (Figure S4.1). The predicted low disorder tendency of BCL2 was not surprising because its binding region is found in a well-folded part of the protein (68) (Figure 4.2). We also

used the MoRFpred server (69) to identify if the Kelch domain interacting regions of the various proteins are predicted to contain molecular recognition features (MoRFs). NRF2 site 1, PTMA and NRF2 site 2 were the only partners that were predicted to contain MoRFs at their binding sites (Figure S4.1). However, it is noteworthy that these two proteins were included in the training dataset of MoRFpred (69). Data presented throughout this work may be useful for further improving this predictive tool. Overall, the results suggest that while both intrinsically disordered and well-folded polypeptide segments are able to bind to the same interaction surface of the Kelch domain, the majority of the partners (identified to date) are disordered.
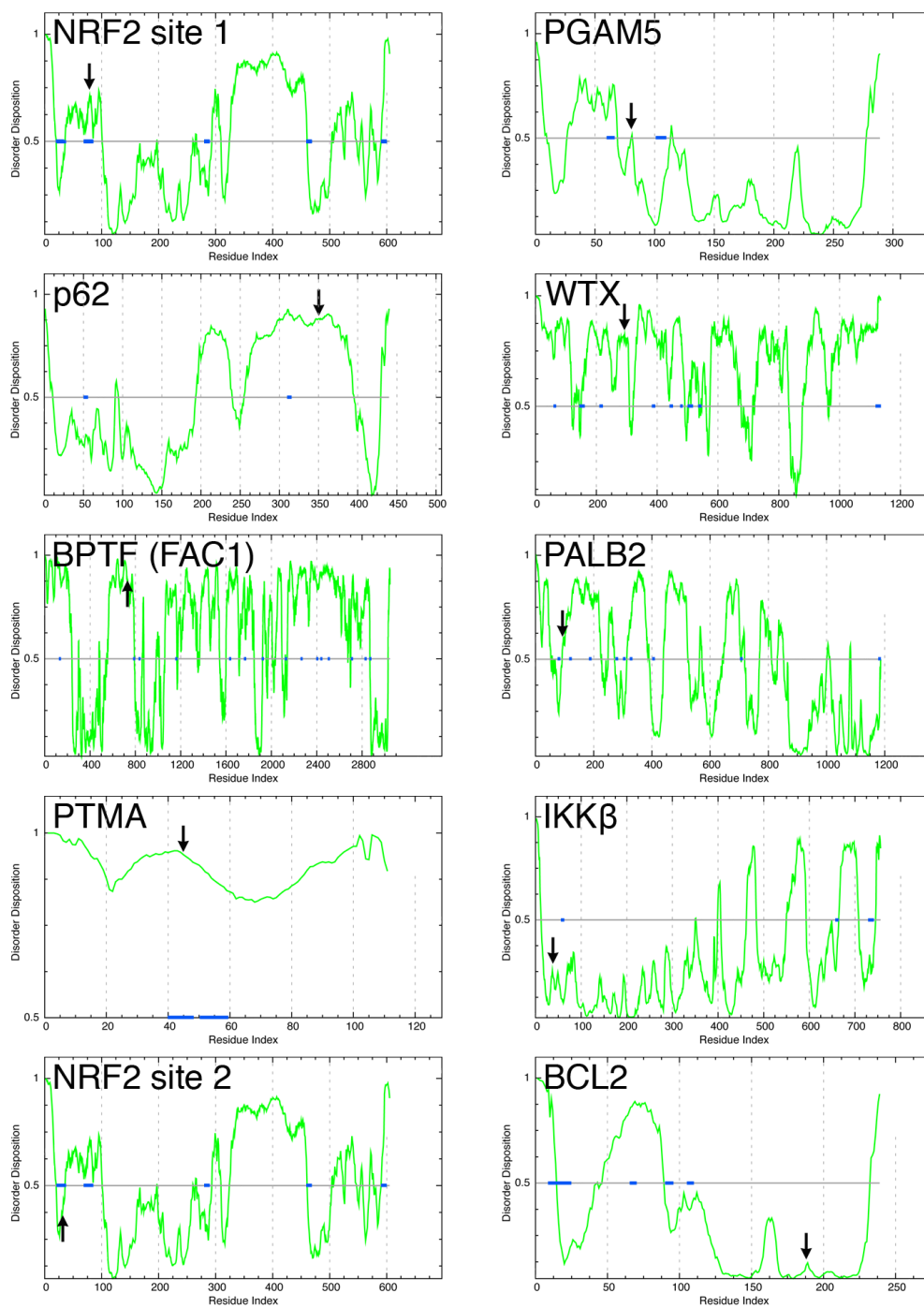
Figure S4.1 Disorder and MoRF predictions of the Kelch domain interacting proteins.

The disorder predictions were performed with PONDR-FIT (66). Arrows indicate the locations of the Kelch domain binding regions (Figure 4.1A). Blue lines

indicate predicted MoRFs (>3 residues) by the MoRFpred web server (69). The gene of FAC1 encodes part of a larger protein called bromodomain PHD finger transcription factor (BPTF). For the FAC1 plot, the prediction of the entire BPTF protein is shown. The residues 1-810 of FAC1 correspond to residues 140-939 of BPTF (97).

Sequence and structure comparison of the Kelch domain interacting proteins

To help assess factors that dictate the binding affinity and specificity to the Kelch domain, we performed an analysis of the amino acid sequences and available structures of the Kelch domain interacting proteins. The site 1-type proteins have sequence identity in a 6-residue stretch corresponding to the 'DEETGE' of NRF2 site 1 (Figure 4.1A). These residues comprise the Kelch domain binding interface (51) and will be referred to as positions *i* - *i*+5. G and E always occupy positions *i*+4 and *i*+5, respectively (Figure 4.1A). E is found at *i*+2 in all of the proteins, except p62, which contains an S at this position (Figure 4.1A). The other positions are more variable (Figure 4.1A). In order of highest to lowest occurrence frequencies, D/N or S is found at position *i*, E, P or V/Q at *i*+1 and T, N or S at *i*+3 (Figure 4.1A). Outside of this 6-residue stretch, there are no clear sequence similarities between the different site 1-type proteins (Figure 4.1A). Furthermore, there is no obvious sequence consensus between the site 1- and 2-type Kelch binding proteins (Figure 4.1A). The two site 2-type proteins have a short, 4-residue, 'WXQD' consensus region. BCL2 contains the sequence 'WIQD', which bears similarity to the 'WRQD' sequence of NRF2 site 2 (Figure 4.1A). Like the site 1-type proteins, these two proteins do not share apparent sequence consensus outside of this short motif (Figure 4.1A).

Next, we compared the residue type fractions of the sequences (Figure 4.1B). The hydrophobic content amongst the site 1-type binders varied considerably between 20-50% (Figure 4.1B). NRF2 site 1 had the highest fraction of hydrophobic residues, followed by p62/WTX, PGAM5/FAC1/IKKβ, PALB2 and PTMA (Figure 4.1B). Polar content in the site 1-type binders was more consistent, with a range of 25-40% (Figure

4.1B). PTMA, p62 and IKKβ had the highest fraction of polar content, followed by NRF2/WTX/FAC1/PALB2 and PGAM5 (Figure 4.1B). Acidic content fell into a range of 10-25% for all proteins, except PTMA, which had a considerably higher fraction of 40% (Figure 4.1B). Basic content varied between 0% in NRF2 and PTMA to 20% in PGAM5, PALB2 and IKKβ (Figure 4.1B). The site 2-type sequences of NRF2 and BCL2 had similar amounts of hydrophobic content between 40-45% (Figure 4.1B). BCL2 had considerably more polar and less acidic content compared to the NRF2 site 2 sequence (Figure 4.1B). Basic content was similar for both 20-mer sequences (Figure 4.1B).

We also compared the available structures of the Kelch domain binding proteins. Aside from IKKβ, for which a homology model structure exists (65), the free-state structures of the site 1 type Kelch domain binding proteins are otherwise not known. However, structures of NRF2 site 1 and 2, p62 and PTMA peptides bound to the Kelch domain have been determined (48, 54, 55, 59). In their bound states, the PTMA and p62 peptides both adopt β-hairpin structures with low rmsd (< 0.3 Å Cα rmsds for 8 atom pairs) to the NRF2 site 1 peptide bound to the Kelch domain (48, 54, 55) (Figure 4.2). The bound state structures of these peptides also reveal that they assume similar sidechain conformations (Figure 4.2). In its free state, the Kelch domain binding region of IKKβ (Figure 4.1A) also forms a β-hairpin that has considerable resemblance (~0.5 Å Cα rmsd for 8 atom pairs) to the bound state structure of the NRF2 site 1 peptide (Figure 4.2). Furthermore, the bound state structures of the NRF2 site 1, p62 and PTMA peptides and the free state structure of IKKβ have similar backbone and sidechain χ^1 dihedral angles in their binding interface residues (Table S1). When comparing the site 2-type binders, it was clear that in its free state, the 'WIQD' sequence of BCL2 (PDB id: 1G5M) (68) has considerable structural resemblance (<0.5 Å Cα rmsd for 4 atom pairs) to the 'WRQD' sequence of the NRF2 site 2 peptide bound to Kelch (PDB id: 2DYH) (59) (Figure 4.2). These residues appear to adopt a 'turn' conformation and share similar backbone, but not χ^1, dihedral angles (Figure 4.2 and Table S1). Intriguingly, although the site 1- and 2-type Kelch domain interacting proteins do not have obvious sequence similarities, the residues that are largely buried in the Kelch domain binding interface (EE in NRF2 site 1 and PTMA, PS in p62, QE in IKKβ and QD in NRF2 site 2 and BCL2)

have similar ϕ and ψ angles (Table S1).
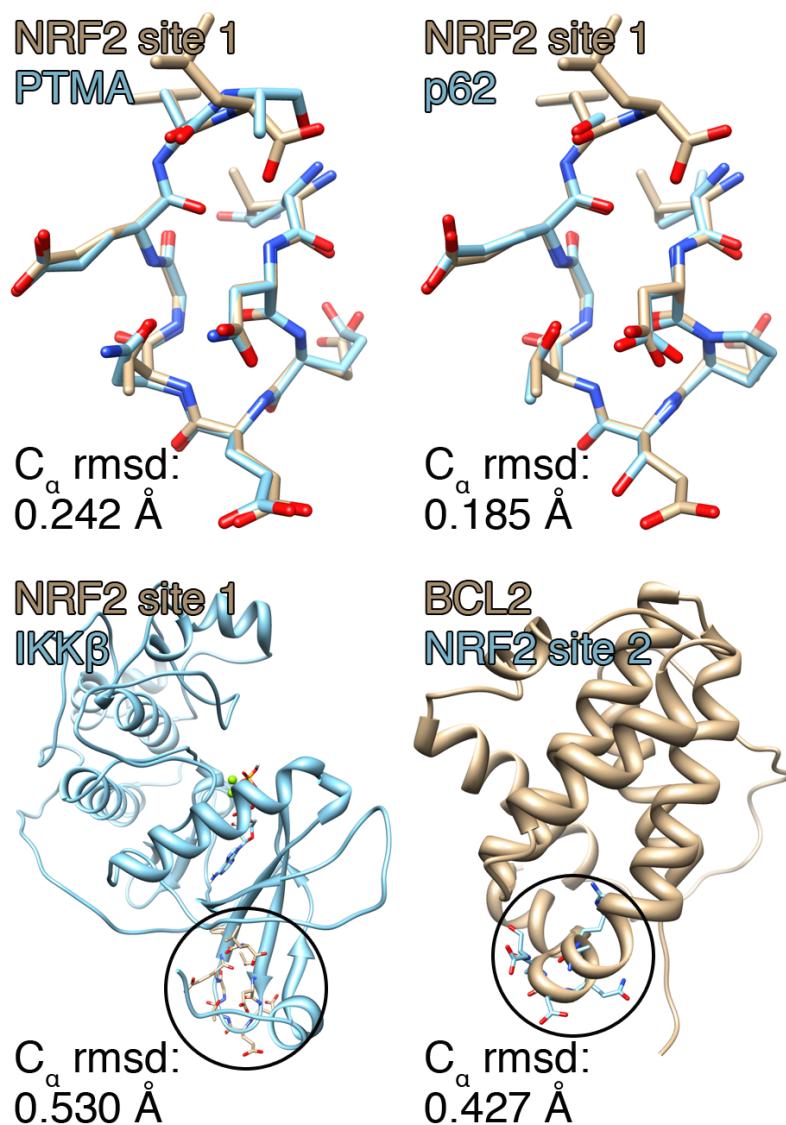
Figure 4.2 Structure comparisons of the Kelch domain binding proteins.

A) Superposition of NRF2 site 1 and PTMA peptides bound to the Kelch domain (PDB ids: 1X2R and 2Z32, respectively) (54, 55). The structures of the Kelch domain are not shown for clarity. B) Superposition of NRF2 site 1 and p62 peptides bound to the Kelch domain (not shown) (PDB ids: 1X2R and 3ADE, respectively) (48, 55). C) Superposition of an NRF2 site 1 peptide bound to the

Kelch domain (not shown) (PDB id: 1X2R) (55) and the free state IKKβ

homology model structure (PMDB id: 76858) (65). D) Superposition of an NRF2

site 2 peptide bound to the Kelch domain (not shown) (PDB id: 2DYH) (59) and

the free state BCL2 structure (PDB id: 1G5M) (68).

Table S1 Dihedral angles of the Kelch domain binding interface residues.

Protein	Residue	$\phi°$	$\psi°$	χ^{1o}
NRF2 site 1[a]	D77	-70.8	110.0	-170.8
	E78	-63.8	-21.8	-144.1
	E79	-82.0	-31.4	-72.8
	T80	-117.2	-21.2	66.0
	G81	85.3	-5.7	
	E82	-71.8	146.8	-67.0
p62[b]	D349	-87.3	109.3	-179.9
	P350	-47.7	-39.0	-26.0
	S351	-72.9	-31.1	-173.2
	T352	-121.8	-22.0	61.5
	G353	84.7	4.7	
	E354	-76.0	143.1	-80.5
PTMA[c]	N41	-67.8	115.6	-174.6
	E42	-70.1	-16.1	-80.4
	E43	-77.7	-39.9	-68.1
	N44	-126.4	-2.9	61.4
	G45	78.1	0.3	
	E46	-74.8	152.8	-51.2
IKKβ[d]	N34	-94.8	134.7	-168.9
	Q35	-91.5	-22.0	-83.2
	E36	-81.7	-36.7	-60.4
	T37	-102.5	-26.4	62.1
	G38	80.8	10.6	
	E39	-59.7	133.1	-166.9
NRF2 site 2[e]	W24	na[g]	na[g]	na[g]
	R25	-67.9	-16.5	58.5
	Q26	-77.6	-24.6	49.0
	D27	-76.5	-17.3	-88.4
BCL2[f]	W188	-63.7	-40.2	-179.6
	I189	-46.5	-59.0	-81.2
	Q190	-66.8	-30.2	-76.1
	D191	-64.5	-45.5	-150.4

[a] PDB id: 1X2R(55).
[b] PDB id: 3ADE(48).
[c] PDB id: 2Z32(54).

^d Protein model database (PMDB) id: 76858(65).
^e PDB id: 2DYH(59).
^f PDB id: 1G5M(68).
^g Atom positions needed to calculate the angle were not resolved in the structure.

Binding parameters of the Kelch domain interacting proteins

Using ITC, we determined the thermodynamic parameters of binding (Table 4.2) for peptides from the binding regions (Figure 4.1A) of all of the known (to date) site 1-type Kelch domain interacting proteins with high disorder tendencies. ITC thermograms and an additional set of binding parameters from duplicate experiments are available (Figure S4.2 and Table S4.2). Binding parameters have been previously reported for NRF2 (40, 57, 58), p62 (48) and PTMA isoform 2 (40). To our knowledge, parameters for PGAM5, WTX, WTX pS286, FAC1, PALB2 and PTMA isoform 1 have not been reported. Here, we determined the binding parameters for these proteins as well as the ones that have been previously reported. The reason for repeating the previously documented measurements was to insure that a systematic comparison with equal length peptides, same buffer conditions, protein/peptide concentrations, was performed.

Table 4.2 Thermodynamic parameters for the binding of the peptides to the human Kelch domain[a]

Protein	n[b]	$K_d{}^c$ (10^{-6} M)	ΔH^c (kcal/mol)	$T\Delta S^c$ (kcal/mol)	ΔG^c (kcal/mol)
NRF2 site 1	1.08	0.023 ± 0.002	-16.96 ± 0.046	-6.559	-10.40 ± 0.03
PGAM5	1.07	0.23 ± 0.010	-19.12 ± 0.056	-10.048	-9.07 ± 0.03
p62	0.97	1.3 ± 0.021	-18.38 ± 0.067	-10.345	-8.04 ± 0.03
WTX	1.04	0.25 ± 0.017	-18.04 ± 0.087	-9.034	-9.01 ± 0.04
WTX pS286	0.98	1.5 ± 0.140	-10.83 ± 0.143	-2.880	-7.95 ± 0.10
FAC1	0.99	1.1 ± 0.047	-15.11 ± 0.076	-6.976	-8.13 ± 0.04
PALB2	1.01	0.087 ± 0.007	-19.29 ± 0.109	-9.660	-9.63 ± 0.05
PTMA iso 1	1.07	11.6 ± 0.230	-14.74 ± 0.120	-8.020	-6.72 ± 0.05
PTMA iso 2	1.05	2.62 ± 0.052	-17.29 ± 0.064	-9.690	-7.60 ± 0.03

[a] The peptide sequences are shown in Figure 4.1A and Table 4.1.
[b] Binding stoichiometry.
[c] K_d is the dissociation constant. ΔH, ΔS and ΔG are the change in enthalpy, entropy and Gibbs free energy upon binding at T=298.15 K, respectively.
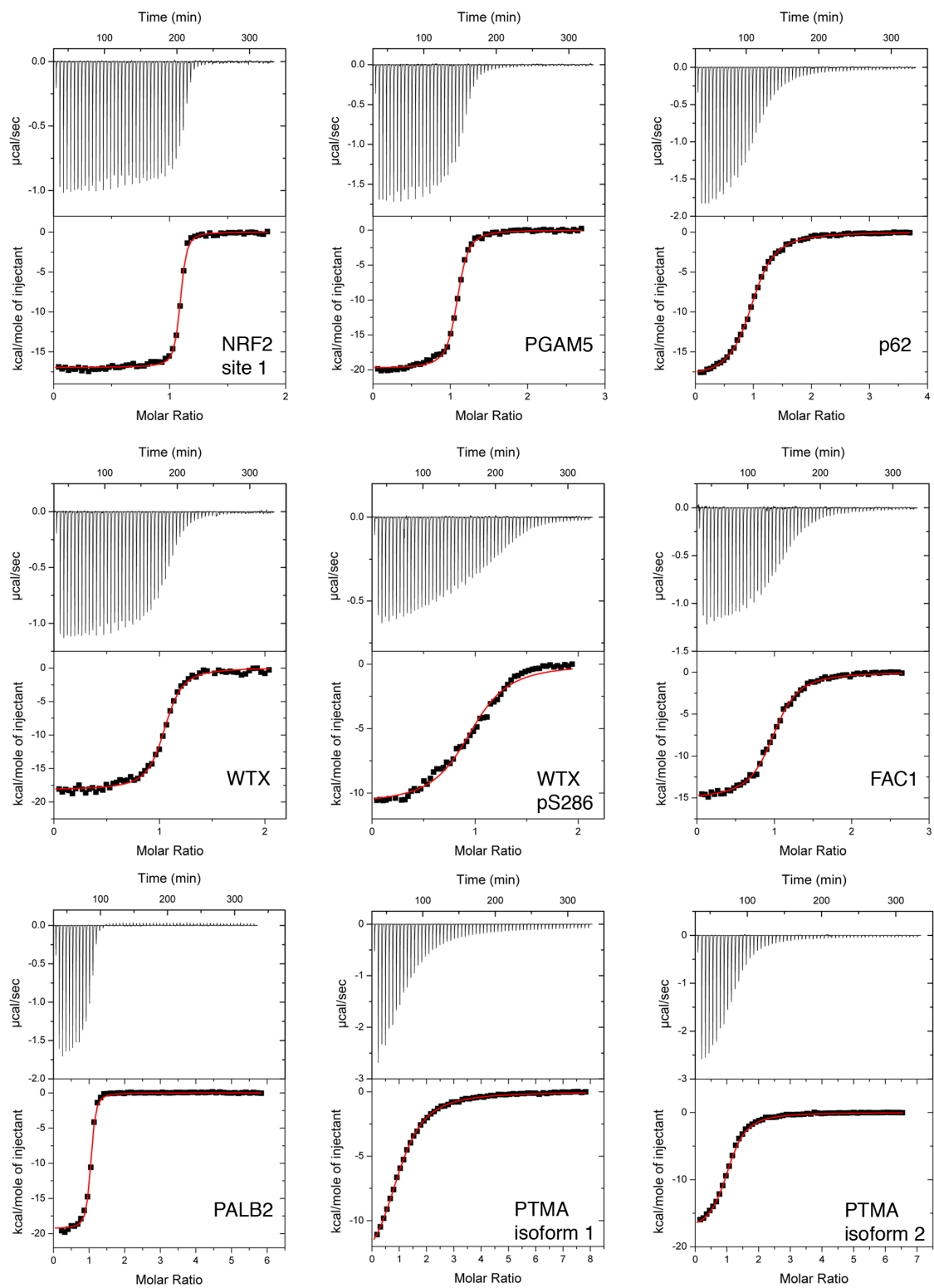
Figure S4.2 ITC thermograms.

Filtered and degassed aliquots of ~40 µM Kelch containing 50 mM sodium

phosphate, 100 mM NaCl, 1 mM DTT at pH 7 was added to the 1.4 mL sample cell and subjected to stepwise titration with 5 μL aliquots of ~500 μM peptide. The equilibration period between each injection was 300 seconds. The association constant (K_a), molar binding stoichiometry (n) and the binding enthalpy (ΔH), entropy (ΔS) and Gibbs free energy (ΔG) were determined by fitting the binding isotherm to a single-binding-site model with Origin7 software (MicroCal). The heat changes after saturation were averaged and used to correct for the heats of dilution. All ITC experiments were performed in duplicate. ITC experiments were carried out on a VP-ITC instrument (MicroCal) at 25°C. The protein and peptide samples were dialyzed into a buffer containing 50 mM sodium phosphate, 100 mM NaCl, 1 mM DTT at pH 7 and degassed before the experiments.

Table S4.2. Duplicate set of peptide to the human Kelch domain thermodynamic binding parameters[a]

Protein	n[b]	K_d[c] (10^{-6} M)	ΔH[c] (kcal/mol)	$T\Delta S$[c] (kcal/mol)	ΔG[c] (kcal/mol)
NRF2 site 1	1.08	0.024 ± 0.002	-17.56 ± 0.046	-7.185	-10.38 ± 0.03
PGAM5	1.08	0.18 ± 0.010	-19.78 ± 0.066	-10.584	-9.20 ± 0.03
p62	0.99	1.7 ± 0.038	-18.27 ± 0.088	-10.375	-7.90 ± 0.04
WTX	1.04	0.27 ± 0.020	-18.15 ± 0.094	-9.183	-8.97 ± 0.05
WTX pS286	0.99	1.6 ± 0.16	-10.39 ± 0.136	-2.480	-7.91 ± 0.10
FAC1	0.96	0.98 ± 0.069	-15.41 ± 0.124	-7.215	-8.20 ± 0.07
PALB2	0.96	0.13 ± 0.02	-19.30 ± 0.123	-9.869	-9.43 ± 0.06
PTMA iso 1	1.09	13.4 ± 0.32	-14.32 ± 0.153	-7.662	-6.66 ± 0.07
PTMA iso 2	1.04	2.93 ± 0.078	-17.53 ± 0.093	-10.077	-7.45 ± 0.04

[a] The peptide sequences are shown in Figure 4.1A and Table 4.1.
[b] Binding stoichiometry.
[c] K_d is the dissociation constant. ΔH, ΔS and ΔG are the change in enthalpy, entropy and Gibbs free energy upon binding at T=298.15 K, respectively.

A large variation in the binding affinity (K_d ranging from ~12 μM to 23 nM) was observed for different disordered Kelch domain interacting proteins. Out of all the peptides, the NRF2 site 1 peptide had the highest affinity for the Kelch domain (23 ± 2 nM), followed by PALB2 (87 ± 7 nM), PGAM5 (230 ± 10 nM), WTX (250 ± 17 nM), FAC1 ($1\,100 \pm 47$ nM), p62 ($1\,300 \pm 21$ nM), WTX pS286 ($1\,500 \pm 140$ nM), PTMA

isoform 2 (2 620 ± 52 nM) and PTMA isoform 1 (11 600 ± 230 nM). The K_d of the NRF2 site 1-Kelch domain interaction has been measured at 20 nM for a 16-mer peptide, which was consistent with our value of 23 ± 2 nM (51). It is interesting that PALB2, which contains the same 'LDEETGE' sequence as NRF2 site 1, was ~4-fold weaker. It has been demonstrated that PALB2 can compete with NRF2 for Kelch domain binding, and as a result, the authors proposed that PALB2 may have a similar, or higher affinity (52). Our ITC data showed that while the PALB2 peptide interacted with the Kelch domain slightly more favorably enthalpically than the NRF2 site 1 peptide, it lost considerable entropy upon binding. Our sequence analysis illustrates that, compared to NRF2 site 1, PALB2 has considerably less hydrophobic content in its binding region, which may allow for more conformational freedom. This could possibly explain why PALB2 loses considerable entropy upon binding, resulting in a weaker interaction. The affinity of the p62 peptide for the Kelch domain was 1300 ± 21 nM, which was on par with the reported value of 1851 ± 103 for a (mouse) fragment containing residues 168-391 (48). The high similarity between these measurements indicates that regions distant from the binding motif may not participate in the interaction. The relatively weak affinity of the p62-Kelch domain interaction may be attributed to the lacking of an E in the i+2 position of p62 (Figure 4.1A). The other binding partners all contain an E at this position and structures of NRF2 site 1 and PTMA peptides in complex with Kelch show that this residue forms favorable electrostatic interactions with a basic surface of Kelch (51, 54, 55). Although this suggests that the lower affinity of the p62 peptide would be primarily due to a less favorable enthalpic component of binding, p62 actually had a more favorable ΔH of binding compared to most of the other peptides, including NRF2. Because this could be due to number of possibilities, such as favorable intra-peptide interactions occurring upon binding, further experiments are necessary to determine the thermodynamic contributions of the E in position i+2. The unphosphorylated WTX peptide interacted comparably to PGAM5; however, upon phosphorylation of S286, binding was substantially decreased by ~6-fold (Table 4.2). The significant effect of WTX phosphorylation on its binding affinity parallels the effect of phosphorylation of residue T80 (position i+3) in NRF2, both of which occur *in vivo* (45, 51, 70). The ~5 fold lower affinity of PTMA isoform 1 compared to isoform 2 (Table 4.2) was intriguing. The

two isoforms have nearly identical sequences, with the major difference being a deletion of E at position *i*-1 in isoform 2 (Figure 4.1A). The contributing factor(s) for this difference are not clear, but we speculate that having amino acids with the same charge (E41 and E48) somewhat close to each other in a β-turn conformation is unfavorable. Similarly, the D589 and E597 pair in FAC1 would also be in close proximity, assuming a β-turn conformation is adopted.

There were some obvious trends in the ITC data, which led us to group the different proteins based on their binding affinities. These groupings may be helpful in deciphering the relationships between binding affinity and the biological functions of the various protein-protein interactions. Tier 1 consists of NRF2 only, which has a K_d of ~20 nM. NRF2 is known as the master regulator of the cellular oxidative stress response pathway (71-73). Tier 2 consists of PALB2, PGAM5 and WTX, which have K_d's in the ~100-200 nM range. The proteins have been shown to promote NRF2-mediated cytoprotective gene expression, by presumably, disrupting the low affinity site 2-Kelch domain interaction. Based on their affinities, these proteins should be able to easily disrupt this interaction, which has a K_d of ~1 µM (57). A third tier of proteins, consisting of FAC1, p62 and PTMA have dissociation constants > 1000 nM. PTMA contains a nuclear localization signal, and is thought to function as a vehicle for shuttling Keap1 into the nucleus. The transient nature of its shuttling role may explain its lower affinity. It should be noted that while the proteins discussed here interact with the Kelch domain of Keap1, the purpose of many of the interactions are not well established. The binding parameters reported here and hypothesis presented in several recent review articles may give insights into their possible roles (44, 63).

Free state structures of Kelch domain interacting peptides

NMR and MD simulations were used to assess the free state solution structures of the various Kelch domain interacting peptides. These experiments were crucial for assessing the relationships between peptide conformation in the free and bound states. We previously found evidence that, in their free states, the NRF2 site 1 region and the

site 1-type region of PTMA adopt β-hairpin conformations that resemble the bound state structures to different extents, with NRF2 forming a more defined hairpin that has a closer resemblance to its bound state structure compared to PTMA (40). The chemical shift assignments and assigned ^{1}H-^{15}N HSQC spectra for each peptide are available (Table S4.3 and Figure S4.3). The small ^{1}Hα chemical shifts deviations from their random coil values, and narrow range of ^{1}H peak dispersions (~1 ppm) suggest that the peptides are largely disordered in solution (Table S4.3 and Figure S4.3). The main purpose of assigning the ^{1}H resonances was to assess the solution structures of the various peptides using NOESY experiments. In Figure 4.3, lines mapped onto the peptide sequences, are used to illustrate any NOESY cross-peaks between protons >2 residues apart. Using this criterion, the NRF2 site 1 peptide had the largest number of cross-peaks of all the peptides (Figure 4.3). Several of the NRF2 site 1 peptide cross-peaks were between residues comprising the β-turn motif that forms the binding interface with the Kelch domain (Figure 4.3). The corresponding region in many of the other peptides also contained NOESY cross-peaks (Figure 4.3). In particular, cross-peaks between residues in the *i* and *i*+3 positions of the regions corresponding to the β-turn motif of the NRF2 site 1 peptide were usually observed (Figure 4.3). In addition to contacts within the β-turn, the NRF2 site 1 peptide also contained cross-peaks between hydrophobic residues on either side of the turn, suggestive of a hairpin structure (Figure 4.3). The p62 and WTX peptides also had detectable NOESY cross-peaks between residues outside of their binding motifs indicative of hairpin structures (Figure 4.3). Although NOESY cross-peaks between residues > 2 residues apart were not found for the FAC1, PALB2 and PTMA peptides, the presence of such contacts cannot be ruled out. Incomplete resonance assignments and overcrowding made analysis of these spectra challenging.

NRF2 site 1   Ac-FAQLQLDEETGEFLPIQPAQ-NH$_2$

PGAM5   Ac-INVRKRNVESGEEELASKLD-NH$_2$

p62   Ac-LSSKEVDPSTGELQSLQMPE-NH$_2$

WTX   Ac-SLEEPHSPETGEKVVAGEVN-NH$_2$

WTX pS286   Ac-SLEEPHpSPETGEKVVAGEVN-NH$_2$

Figure 4.3 Peptide NOESY connections.

The NMR experiments were performed at 25ºC with a 3 mM peptide sample in 50 mM sodium phosphate pH 7, 100 mM NaCl, 1 mM DTT and trace DSS for chemical shift referencing. Lines mapped onto the peptide sequences are used to illustrate any NOESY cross-peaks between protons > 2 residues apart.

Table S4.3. Chemical shifts assignments of the peptides.

NRF2 site 1 peptide, Ac-FAQLQLDEETGEFLPIQPAQ-NH$_2$

Residue	N$_H$	H$_N$	Hα	Hβ	Others
F	125.931	8.175	4.540	2.977, 3.119	Hδ: 7.261
A	125.300	8.269	4.272	1.328	
Q	119.427	8.217	4.298	1.968, 2.084	Hγ: 2.346, Hε: 6.867, 7.548
L	123.158	8.165	4.423	1.621	Hγ: 1.490, Hδ: 0.825
Q	121.318	8.596	4.449	1.977, 2.058	Hγ: 2.315, Hε: 6.850, 7.563
L	124.973	8.388	4.209		Hγ: 1.456, 1.495, Hδ: 0.661, 0.749
D	123.342	8.515	4.589	2.560, 2.846	
E	123.292	8.571	4.102	1.967, 2.056	
E	120.166	8.439	4.283	1.923, 2.090	Hγ: 2.384
T	111.394	8.022	4.383	4.275	Hγ: 1.211
G	111.218	8.426	3.810, 3.999		
E	119.788	7.931	4.258	1.786, 1.925	Hγ: 2.071, 2.150
F	120.922	8.426	4.667	2.960, 3.009	Hδ: 7.185
L	125.787	8.296	4.664	1.575	Hγ: 1.529, Hδ: 0.886
P			4.463	2.004, 2.258	Hγ: 1.863, Hδ: 3.616, 3.741
I	121.300	8.154	4.099	1.783	Hγ: 1.166, Hδ: 0.875
Q	125.711	8.450	4.275	1.984, 1.213	Hγ: 2.371
P			4.368	2.314, 2.021	Hγ: 1.908, Hδ: 3.676, 3.811
A	124.333	8.439	4.253	1.401	
Q	119.821	8.337	4.275	1.984, 2.123	Hγ: 2.371

PGAM5 peptide, Ac-INVRKRNVESGEEELASKLD-NH$_2$

Residue	N$_H$	H$_N$	Hα	Hβ	Others
I	125.561	8.145	4.121	1.826	Hγ: 1.461, 1.196, Hδ: 0.867
N	120.831	8.599	4.695	2.763, 2.873	Hδ: 6.901, 7.599
V	120.437	7.962	4.077	2.082	Hγ: 0.912
R	123.436	8.342	4.338	1.748, 1.824	Hγ: 1.624, Hδ: 3.008
K	122.556	8.393	4.342	1.743, 1.812	Hγ: 1.586, Hδ: 1.627
R	124.816	8.340	4.305	1.744, 1.825	Hγ: 1.626, Hδ: 3.182, Hε: 7.249
N	128.837	8.550	4.717	2.746, 2.831	Hδ: 6.928, 7.638
V	120.166	8.133	4.123	2.113	Hγ: 0.924

E	121.718	8.548	4.318	1.956, 2.087	Hγ: 2.282
S	116.273	8.328	4.442	3.869. 3.927	
G	111.198	8.484	3.971, 4.012		
E	120.511	8.263	4.283	1.928, 2.054	Hγ: 2.247
E	122.556	8.393	4.246	1.957, 2.033	Hγ: 2.233, 2.271
E	121.693	8.564	4.233	1.960, 2.020	Hγ: 2.275
L[a]	123.053	8.224	4.285	1.592, 1.665	Hδ: 0.904, 0.955
A	124.157	8.226	4.279	1.416	
S	114.277	8.133	4.397	3.862, 3.889	
K	123.147	8.198	4.286	1.782, 1.883	Hγ: 1.446, Hε: 3.266
L[a]					
D	121.151	8.213	4.546	2.654	

[a] L15 and L19 have similar H_N shifts. A set of ambiguous shifts for both residues is reported in the L15 row.

p62 peptide, Ac-LSSKEVDPSTGELQSLQMPE-NH$_2$

Residue	N_H	H_N	Hα	Hβ	Others
L	123.822	8.182	4.368	1.624	Hδ: 0.856, 0.898
S	116.678	8.286	4.419	3.852, 3.877	
S	115.561	8.572	4.458	3.910, 3.945	
K	127.719	8.272	4.334	1.746, 1.840	Hγ: 1.423, Hε: 3.155
E	121.822	8.335	4.322	1.918, 2.044	Hγ: 2.255
V	120.783	8.116	4.093	1.990	Hγ: 0.895
D	126.005	8.481	4.878	2.552, 2.864	
P			4.435	2.034, 2.323	Hδ: 3.685, 3.827
S	117.812	8.337	4.458	3.924, 3.990	
T	114.054	7.960	4.348	4.309	Hγ: 1.226
G	110.677	8.323	3.931, 3.995		
E	120.549	8.181	4.297	1.918, 2.041	Hγ: 2.220, 2.258
L	122.936	8.337	4.324	1.659	Hγ: 1.579, Hδ: 0.863
Q	123.326	8.391	4.294	1.964, 2.104	Hγ: 2.372
S	116.546	8.415	4.510	3.876, 3.911	
L	122.695	8.285	4.332	1.603, 1.655	Hδ: 0.882, 0.906
Q	120.946	8.269	4.312	1.963, 2.073	Hγ: 2.348
M	121.310	8.403	4.347		Hγ: 2.562, 2.635
P			4.412	1.926	Hδ: 3.687, 3.812
E	121.692	8.545	4.204	1.965, 2.031	Hγ: 2.292

WTX peptide, Ac-SLEEPHSPETGEKVVAGEVN-NH$_2$

Residue	N_H	H_N	Hα	Hβ	Others
S	121.319	8.303	4.426	3.832, 3.859	
L	124.232	8.428	4.398	1.627	Hδ: 0.868, 0.911
E	123.677	8.338	4.273	1.877, 2.016	Hγ: 2.267

E	121.578	8.298	4.253	1.889, 2.028	Hγ: 2.238
P			4.380	1.992, 2.257	Hγ: 1.838, Hδ: 3.680, 3.776
H	119.526	8.506	4.671	3.148, 3.162	Hδ: 7.165
S	119.102	8.430	4.291	3.832, 3.869	
P			4.433	2.026, 2.307	Hγ: 1.944, Hδ: 3.757
E	121.066	8.597	4.323	1.963, 2.084	Hγ: 2.299
T	114.070	8.167	4.353	4.293	Hγ: 1.212
G	111.053	8.377	3.988		
E	120.807	8.259	4.346	1.922, 2.026	Hγ: 2.224
K	122.843	8.403	4.342	1.779	Hγ: 1.412, Hδ: 1.742, Hε: 2.987
V	122.843	8.207	4.091	2.019	Hγ: 0.885, 0.921
V	125.563	8.337	4.104	2.027	Hγ: 0.922
A	128.821	8.455	4.331	1.392	
G	108.444	8.350	3.948		
E	na	8.109	4.286	1.793, 1.931	Hγ: 2.145, 2.233
V	121.436	8.286	4.088	2.080	Hγ: 0.931
N	122.584	8.531	4.762	2.747, 2.821	Hδ: 6.924, 7.602

WTX pS286 peptide, Ac-SLEEPHpSPETGEKVVAGEVN-NH$_2$

Residue	N$_H$	H$_N$	Hα	Hβ	Others
S	120.426	8.262	4.429	3.831, 3.866	
L	123.907	8.423	4.375	1.627	Hδ: 0.860, 0.917
E	123.341	8.319	4.267	1.916, 2.021	Hγ: 2.243
E	120.465	8.265	4.301	1.895, 2.026	Hγ: 2.249
P			4.535	2.013, 2.272	Hγ: 1.871, Hδ: 3.677, 3.839
H	118.983	8.501	4.688	3.194, 3.250	Hδ: 7.198, Hε: 8.644
pS	120.027	9.068	4.340	4.014, 4.080	
P			4.420	2.016, 2.292	Hγ: 1.899, Hδ: 3.656, 3.843
E	121.388	8.672	4.334	2.005, 2.086	Hγ: 2.321
T	114.342	8.228	4.349	4.270	Hγ: 1.206
G	110.618	8.410	3.983		
E	120.443	8.240	4.352	1.927	Hγ: 2.230
K	122.448	8.358	4.328	1.750, 1.799	Hγ: 1.432, Hε: 2.761
V	122.419	8.190	4.088	2.022	Hγ: 0.903, 0.939
V	125.197	8.320	4.086	2.084	Hγ: 0.931
A	128.434	8.449	4.315	1.396	
G	108.126	8.333	3.944		
E	na	8.073	4.243	1.792, 1.923	Hγ: 2.151
V	121.073	8.269	4.105	2.041	Hγ: 0.932, 0.914
N	122.276	8.515	4.688	2.761, 2.835	Hδ: 6.893, 7.575

FAC1 peptide, Ac-SAKAADDPENGERESHTPVS-NH$_2$

Residue	N$_H$	H$_N$	Hα	Hβ	Others
S[a]	na	8.289	4.425	3.841, 3.871	
A[b]	126.315	8.408	4.313	1.403	
K	121.013	8.232	4.291	1.737, 1.811	Hγ: 1.435, Hδ: 1.681
A					
A					
D	118.438	8.173	4.556	2.591, 2.641	
D	121.669	8.054	4.853	2.549, 2.747	
P			4.383	2.025, 2.301	Hγ: 1.961, Hδ: 3.799
E	119.725	8.525	4.227	1.957, 2.063	Hγ: 2.233, 2.278
N	118.715	8.199	4.712	2.796, 2.861	Hδ: 6.912, 7.675
G	109.298	8.316	3.955, 3.982		
E	120.685	8.252	4.296	1.954, 2.045	Hγ: 2.240, 2.268
R	123.285	8.316	4.298	1.781, 1.859	Hγ: 1.639, Hδ: 3.188, Hε: 7.178
E	121.568	8.408	4.269	1.935, 2.009	Hγ: 2.243
S					
H	118.816	8.095	4.664	3.429	
T	120.584	8.320	4.330	4.246	Hγ: 1.382
P			4.462	2.049, 2.321	Hγ: 1.914, Hδ: 3.719, 3.867
V	121.694	8.304	4.120	2.113	Hγ: 0.968
S					

[a] S1, S15 and S20 have similar H$_N$ and Hβ shifts. A set of ambiguous shifts for all three residues is reported in the S1 row.
[b] A2, A4 and A5 have similar N$_H$, H$_N$ and Hβ shifts. A set of ambiguous shifts for all three residues is reported in the A2 row.

PALB2 peptide, Ac-HIKTHLDEETGEKTSITLDV-NH$_2$

Residue	N$_H$	H$_N$	Hα	Hβ	Others
H					
I	125.362	8.281	4.105	1.812	Hγ: 1.190, Hδ: 0.899
K	122.602	8.463	4.346	1.783, 1.863	Hγ: 1.467, Hδ: 1.680
T					
H					
L					
D					
E					
E					
T					
G	111.026	8.373	3.975		

Residue	N_H	H_N	$H\alpha$	$H\beta$	Others
E	120.835	8.269	4.270	1.936, 2.055	Hγ: 2.263
K					
T					
S					
I					
T					
L					
D					
V	120.188	8.018	4.125	2.164	Hγ: 0.935

PTMA isoform 1 peptide, Ac-ANGNAENEENGEQEADNEVD-NH$_2$

Residue	N_H	H_N	$H\alpha$	$H\beta$	Others
A					
N					
G					
N					
A					
E					
N					
E					
E					
N					
G					
E					
Q					
E					
A					
D					
N					
E					
V	120.786	8.149	4.096	2.100	Hγ: 0.923
D					

PTMA isoform 2 peptide, Ac-ANGNANEENGEQEADNEVD-NH$_2$

Residue	N_H	H_N	$H\alpha$	$H\beta$	Others
A					
N					
G					
N					
A					
N					
E					
E					
N					
G					
E					
Q					
E					

A				
D				
N				
E				
V	120.727	8.144	4.073	2.092	Hγ: 0.928
D				

NRF2 site 1 E78P peptide, Ac-FAQLQLDPETGEFLPIQPAQ-NH$_2$

| Residue	N$_H$	H$_N$	Hα	Hβ	Others
F	126.184	8.286	4.541	2.954, 3.108	Hδ: 7.250
A	125.466	8.363	4.264	1.328	
Q	119.515	8.313	4.273	1.967, 2.095	Hγ: 2.358
L	123.128	8.247	4.499	1.614, 1.835	Hγ: 1.426, Hδ: 0.780
Q	121.109	8.820	4.506	1.981, 2.062	Hγ: 2.327
L	124.962	8.492	4.198	1.353	Hγ: 1.268, Hδ: 0.624, 0.545
D	125.926	8.769	4.850	2.533, 3.025	
P			4.381	2.019, 2.311	Hγ: 1.918, Hδ: 3.674, 3.830
E	117.769	8.444	4.273	2.117	Hγ: 2.226, 2.299
T	108.789	7.855	4.424	4.303	Hγ: 1.197
G	111.154	8.508	3.732. 4.047		
E	119.124	7.817	4.286	1.787, 1.986	Hγ: 2.116, 2.205
F	120.586	8.625	4.716	2.932, 2.971	Hδ: 7.159
L	125.899	8.508	4.699		Hγ: 1.560, 1.601, Hδ: 0.922
P			4.277	2.042, 2.309	Hδ: 3.889, 4.018
I	121.789	8.284	4.088	1.770	Hγ: 1.159, Hδ: 0.873
Q	126.173	8.599	4.607	1.920, 2.096	Hγ: 2.400
P			4.511	2.273	Hγ: 1.856, Hδ: 3.653, 3.798
A	124.693	8.570	4.266	1.401	
Q	120.176	8.469	4.277	1.987, 2.123	Hγ: 2.383
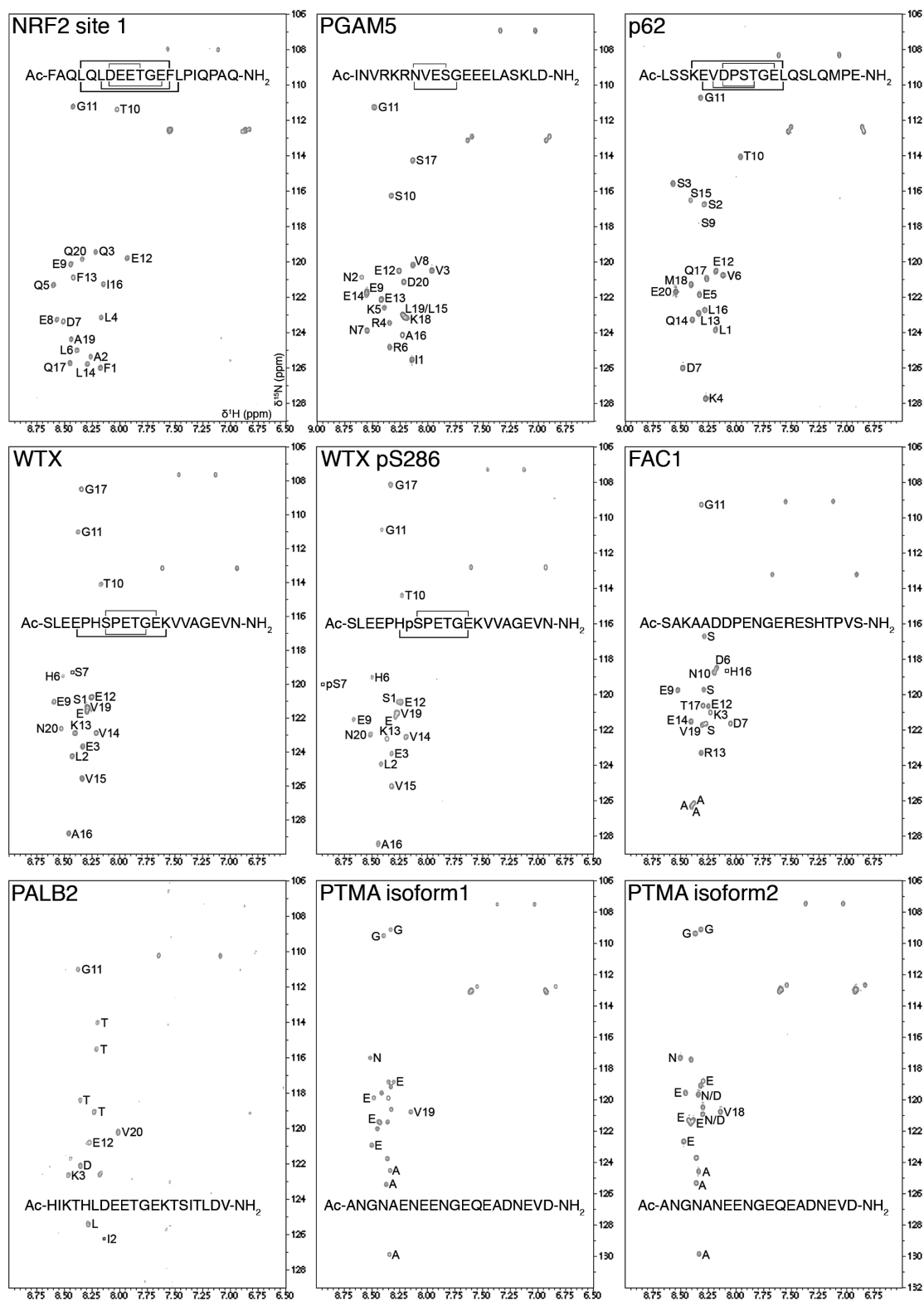
Figure S4.3 Assigned ^{1}H-^{15}N HSQC spectra of the peptides and NOESY connections.

The NMR experiments were performed at 25ºC with a 3 mM peptide sample in 50 mM sodium phosphate buffer at pH 7, 100 mM NaCl, 1 mM DTT and trace DSS for chemical shift referencing. Lines mapped onto the peptide sequences are used to illustrate any NOESY cross-peaks between protons > 2 residues apart.

Microsecond atomistic MD simulations were performed on the Kelch domain interaction peptides (Table 4.1) to investigate their free-state conformational propensities. To compare to the NMR result obtained, the $1/r^6$ averaged $C_\alpha^i - C_\alpha^{i+3}$ distances were calculated (Figure 4.4). The results clearly show that the regions of compactness identified in the NOESY experiments were also found in MD simulations of the peptide in the free states (The $1/r^6$ averaged distance matrices and final structures from the simulations are shown in Figure S4.4). Secondary structure assignments throughout the MD trajectories were also conducted (Figure S4.5). The NRF2 site 1, PGAM5, p62, FAC1, PALB2 and IKKβ peptides showed evidence of cross-strand contacts, which propagated away from their turn regions, and also had final structures with hairpin-like conformations (Figures 4.4 and S4.4). Several of these peptides had intra-turn contacts. Specifically, contacts between residues in position i, $i+2$ and $i+3$ were frequently observed (Figure S4.4). The NRF2 site 1 peptide also had contacts between hydrophobic residues on adjacent strands (Figure S4.4). The PGAM5 peptide showed evidence of cross-strand contacts from the N77 to E82 and E84 sidechains (Figures S4.4 and 4.1A). There was also indication of cross-strand contacts between the oppositely charged K344 and E352 sidechains of the p62 peptide (Figure S4.4 and 4.1A). The WTX peptide adopted a turn/bend conformation at the expected location, and had i to $i+2$ contacts, but did not have cross-strand contacts indicative of a hairpin (Figure S4.4). The WTX pS286 peptide had a turn at the expected location, and some evidence of cross-strand contacts, slightly offset from the turn location (Figure S4.4). In this peptide, obvious contacts between the pS286 sidechain phosphate group and a lysine residue on the opposite side of the turn were observed (Figure S4.4 and 4.1A). In our previous MD simulations, we found that phosphorylation of T80 in an NRF2 peptide severely inhibited formation of the expected β-hairpin structure (39). In the current MD simulations of WTX, it appeared

that S286 phosphorylation may actually enhance free state structure formation. This is supported by our ITC data (Table 4.2), which revealed that the WTX pS286 peptide had the smallest entropy change upon binding. However, this peptide interacted with the Kelch domain the least favorably enthalpically, which suggests that the peptide conformation induced by phosphorylation is not ideal for binding. Indeed, our MD simulations show that the β-turn adopted by this peptide is distorted due to interactions between the pS286 residue and a lysine on the opposite side of the turn (Figure S4.4). Both PTMA peptides did not show clear evidence of turn or hairpin formation in the expected location; however, the secondary structure analysis (Figure S4.5) showed these features occurred transiently throughout the trajectories. Despite the site 2-type peptide trajectories being extended to 2 μs, their 'WXQD' motifs rarely formed 'turn' structures that resembled the NRF2 site 2 bound state structure, or the BCL2 free state structure (data not shown). These trajectories were not analyzed further.
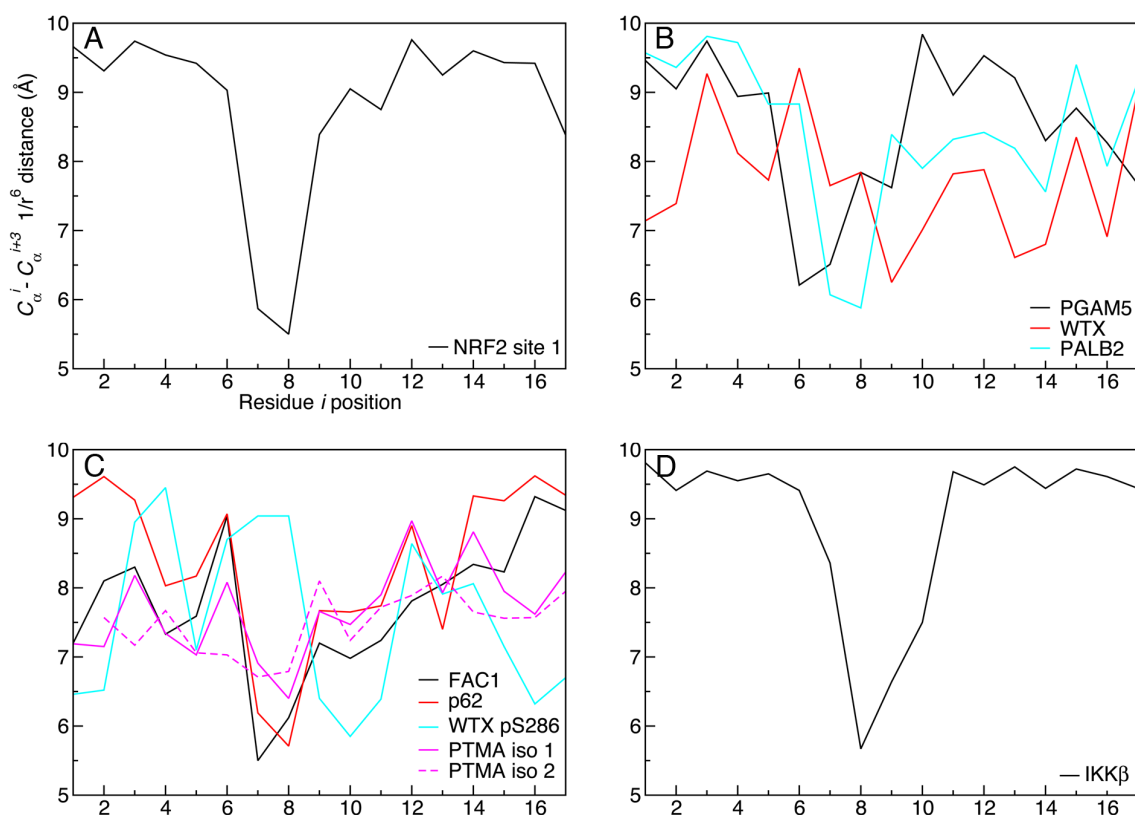

Figure 4.4 C_α^i-C_α^{i+3} $1/r^6$ **averaged distances from the MD simulations.**

The $1/r^6$ averaged distances were calculated using the g_rmsdist tool in GROMACS 4.5 (86) over the last 0.5 µs of the trajectories.
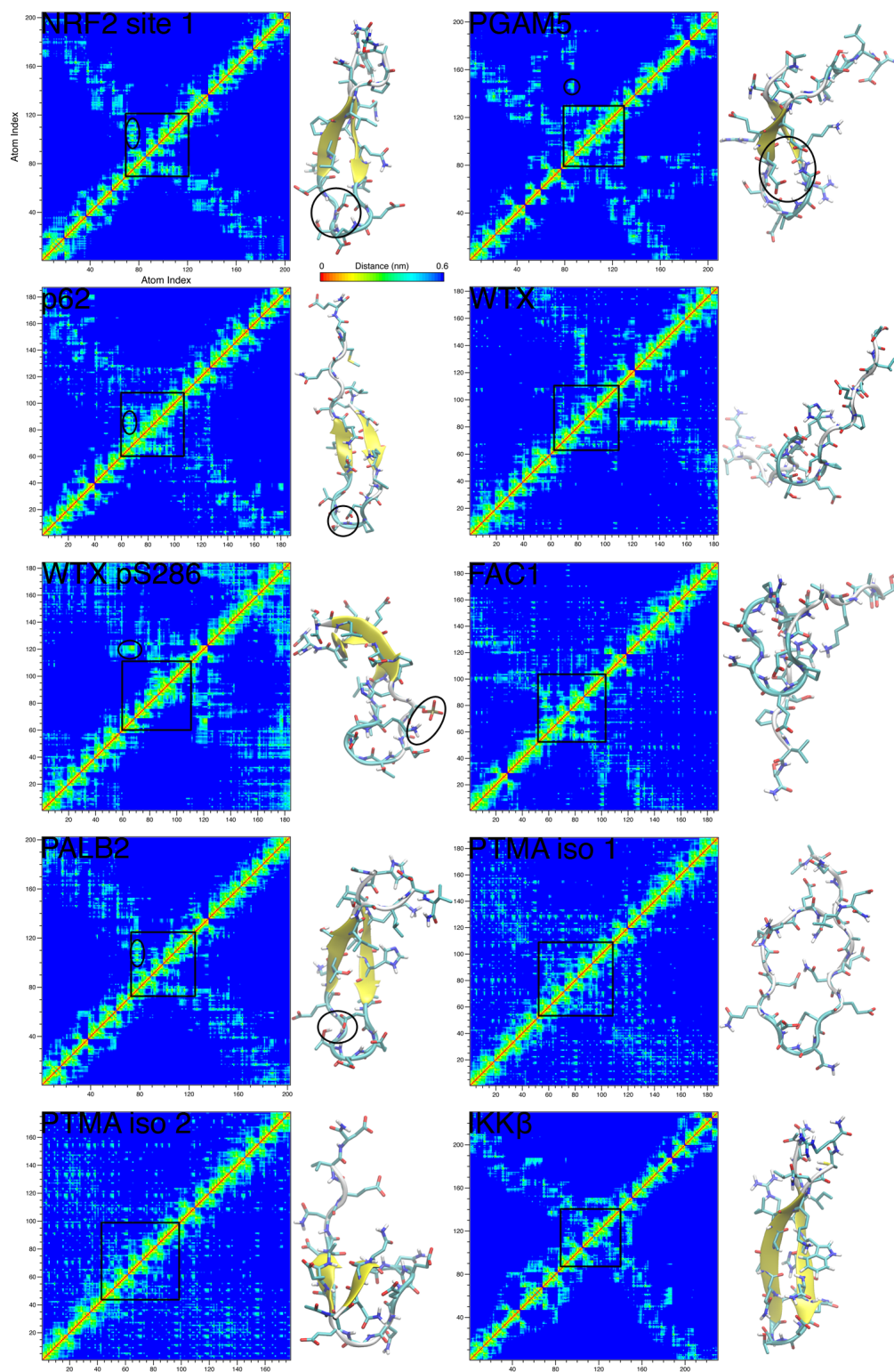
Figure S4.4 1/r⁶ averaged distances and final structures from the MD simulations.

The NMR averaged $1/r^6$ distances between all atoms pairs (atom index) were

calculated using the g_rmsdist tool in GROMACS 4.5(86) over the last 0.5 μs of the trajectories. Final (t=1.0 μs) structures of the peptides rendered with VMD (94). The potential turn forming regions (Figure 4.1A) are boxed and possible turn stabilizing contacts are indicated.
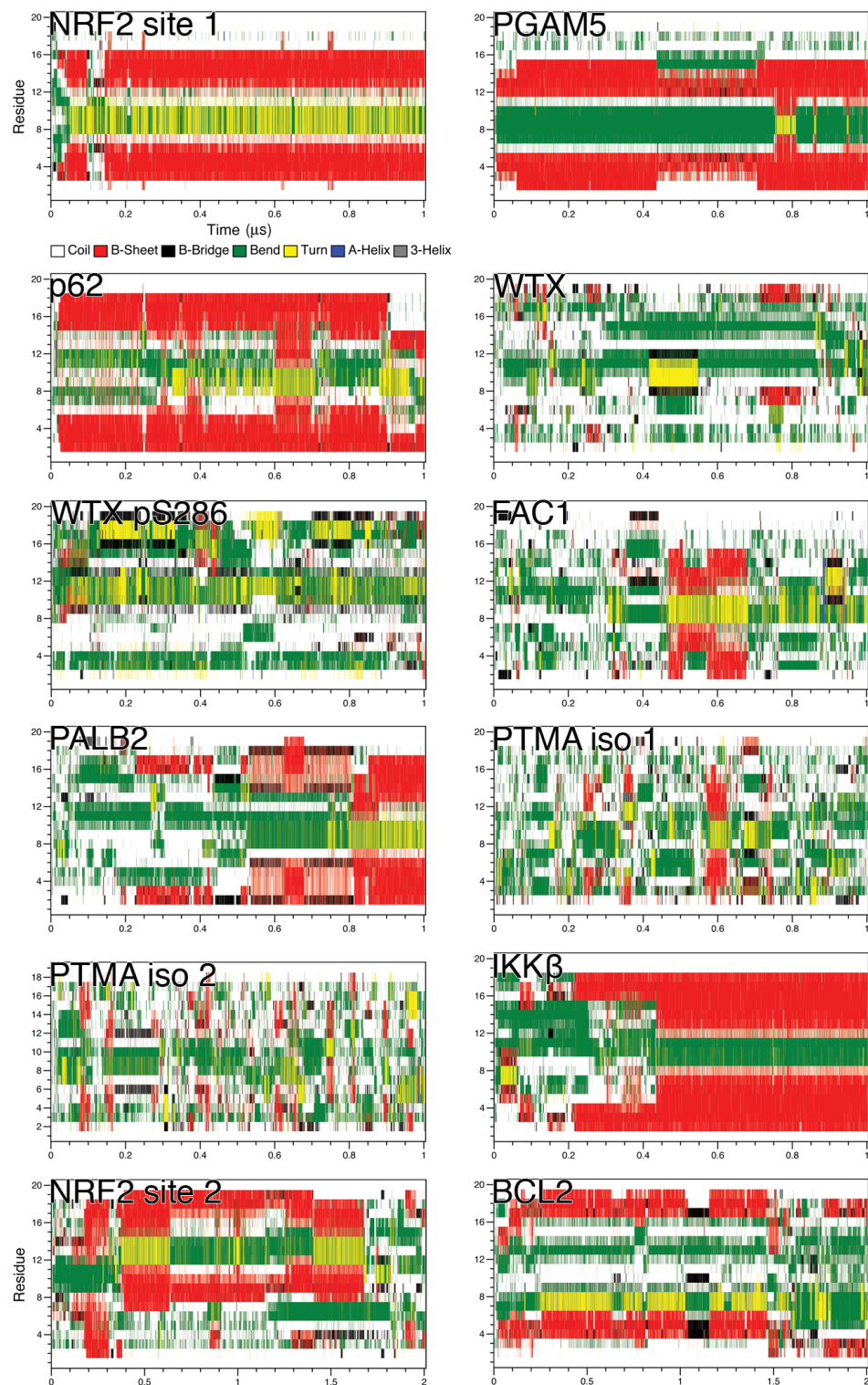
Figure S4.5 Evolution of secondary structures throughout the MD simulations.

The DSSP algorithm (98) was used for secondary structure assignments: coil

(white), β-sheet (red), β-bridge (black), bend (green), turn (yellow), α-helix (blue) and 3_{10} helix (grey).

The NMR and MD results suggest that a larger fraction of hydrophobic residues in the NRF2 peptide may act to stabilize the free state peptide conformation, lowering the entropy difference between the free and bound states, leading to higher binding affinity. While the peptides from the other Kelch domain interacting proteins lack comparable hydrophobic content, our NMR and MD data suggest that like NRF2, interactions between residues within their β-turns are likely to play important roles in stabilizing their free states. This idea is supported by various studies, which together clearly demonstrate that mutating residues within and surrounding the binding regions to amino acids with different characteristics usually decrease Kelch domain binding (45, 48, 50, 52). In addition to intra-turn contacts, we found that interactions between oppositely charged sidechains surrounding the turn might act to stabilize the free state conformations of the PGAM5, p62, WTX pS286 and possibly some of the other peptides (Figure S4.4). Overall, the analysis of the free states reveal that several, if not all, of the Kelch domain binding peptides analyzed in this study contain β-turn resembling LMs at their binding sites. These structures likely have some resemblance to their bound state conformations. It is anticipated that preformed structures are important features in regulating the binding affinities and other thermodynamic parameters of the different interactions.

A higher affinity Kelch domain interacting peptide

The findings from our analysis of the site 1-type proteins were used to create a peptide with a higher affinity for the Kelch domain than any of the natural peptides (identified to date). This peptide aided our interpretation and understanding of the thermodynamics of interaction with the Kelch domain and, possibly even more importantly, may be a potential therapeutic agent (74). Because the data presented here and in our previous work (40) points to a positive correlation between residual structure,

resembling the bound state conformation, in the free state and increased binding affinity, our peptide design aimed to further define the free state structure of the NRF2 site 1 peptide. By further restricting the bound state like conformation of this peptide, the entropic cost of binding may be reduced, potentially increasing binding affinity. Therefore, our attempt at designing a higher affinity peptide focused on reducing free state entropy.

NMR and MD data indicated that β-turn formation at the Kelch domain binding sites is a common feature of the various peptides, therefore, we aimed to increase the turn propensity of this region. The likely sites of turn formation in the site 1-type binders are the 'DEET'-like regions (Figure 4.1A) (40). For the site 2-type binders, the β-turn forming region is not as firmly established, however, the residues 'WRQD' appear to adopt a turn conformation in the bound state (59) (Figure 4.2). Using a table of turn potentials (75), we calculated the residue-specific and overall turn potentials of the 'DEET'-like regions of the various proteins (Table 4.3). The protein with the highest turn potential was p62, followed by FAC1, WTX, NRF2 site 1/PALB2, PTMA, IKKβ and PGAM5 (Table 4.3). The high turn potentials of p62, FAC1 and WTX result primarily from their sequences containing proline at position $i+1$ (Table 4.3). The site 2-type Kelch domain interacting proteins had lower turn propensities than the site 1 binders and NRF2 site 2 had a slightly lower turn potential than BCL2 (Table 4.3).

Table 4.3 Turn potentials of the Kelch domain binding proteins[a].

Protein	i	$i+1$	$i+2$	$i+3$	Total
NRF2 site 1	D(1.56)	E(1.35)	E(0.92)	T(1.20)	5.03
NRF2 site 1 E78P	D(1.56)	P(2.45)	E(0.92)	T(1.20)	6.13
PGAM5	N(1.54)	V(0.70)	E(0.92)	S(1.03)	4.19
p62	D(1.56)	P(2.45)	S(1.06)	T(1.20)	6.27
WTX	S(1.29)	P(2.45)	E(0.92)	T(1.20)	5.86
FAC1	D(1.56)	P(2.45)	E(0.92)	N(1.06)	5.99
PALB2	D(1.56)	E(1.35)	E(0.92)	T(1.20)	5.03
PTMA	N(1.54)	E(1.35)	E(0.92)	N(1.06)	4.87
IKKβ	N(1.54)	Q(0.94)	E(0.92)	T(1.20)	4.60
NRF2 site 2	W(0.62)	R(0.93)	Q(0.92)	D(0.99)	3.46
BCL2	W(0.62)	I(0.61)	Q(0.92)	D(0.99)	3.14

[a] Turn potentials are from (75).

Based upon the turn potential analysis, we made a single point mutation, E78P, to the natural, 20-mer NRF2 site 1 peptide (Figure 4.1A) and used ITC to assess its binding affinity to the Kelch domain. This mutation increases the turn potential to 6.13 (Table 4.3), possibly enriching the population of molecules with defined structure in solution (75). Our ITC measurements reveal that the E78P mutation indeed increases the binding affinity of the peptide (6.90 ± 1.07 nM) compared to the natural sequence by 3-4 fold (Tables 4.2 and 4.4). This increase in binding affinity arises primarily due to a decreased entropic cost of binding (Tables 4.2 and 4.4). The TΔS value is increased by ~1 kcal/mol, while the ΔH increased marginally (0.2 kcal/mol) compared to the natural peptide (Tables 4.2 and 4.4). The ITC thermogram for this measurement and an additional set of thermodynamic parameters from a duplicate experiment are available (Figure S4.6 and Table S4.4).

Table 4.4 Thermodynamic parameters for the binding of the E78P to the human Kelch domain[a]

Protein	n^{b}	K_d^{c} (10^{-9} M)	ΔH^{c} (kcal/mol)	$T\Delta S^{c}$ (kcal/mol)	ΔG^{c} (kcal/mol)
NRF2 site 1 E78P	0.994	6.90 ± 1.07	-16.76 ± 0.05	-5.635	-11.12 ± 0.03

[a] The peptide sequences are shown in Table 4.1.
[b] Binding stoichiometry.
[c] K_d is the dissociation constant. ΔH, ΔS and ΔG are the change in enthalpy, entropy and Gibbs free energy upon binding at T=298.15 K, respectively.
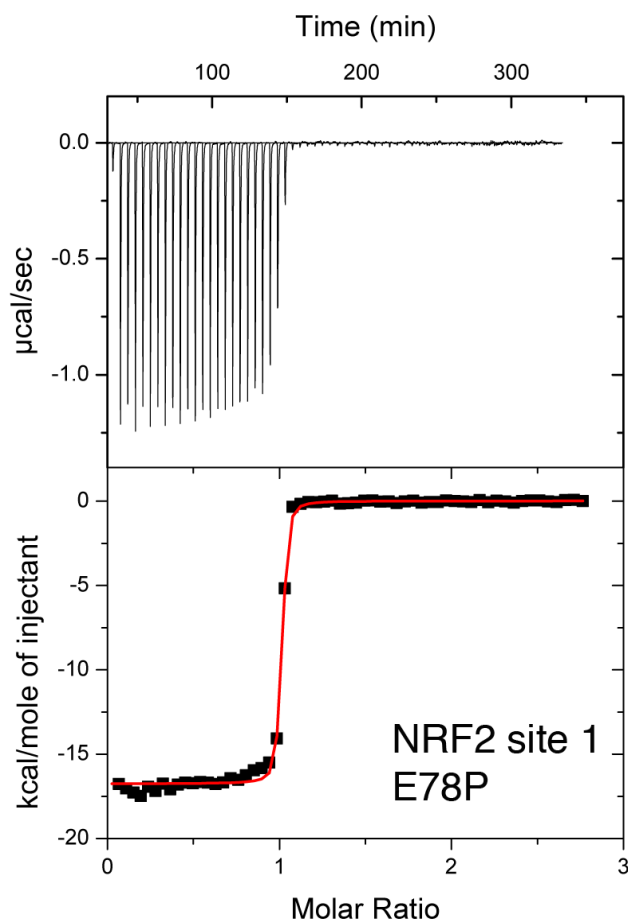
Figure S4.6 ITC thermogram for the NRF2 site 1 E78P peptide.

Filtered and degassed aliquots of ~40 μM Kelch containing 50 mM sodium phosphate, 100 mM NaCl, 1 mM DTT at pH 7 was added to the 1.4 mL sample cell and subjected to stepwise titration with 5 μL aliquots of ~500 μM peptide. The equilibration period between each injection was 300 seconds. The association constant (K_a), molar binding stoichiometry (n) and the binding enthalpy (ΔH), entropy (ΔS) and Gibbs free energy (ΔG) were determined by fitting the binding isotherm to a single-binding-site model with Origin7 software (MicroCal). The heat changes after saturation were averaged and used to correct for the heats of dilution. All ITC experiments were performed in duplicate. ITC experiments were carried out on a VP-ITC instrument (MicroCal) at 25ºC. The protein and peptide samples were dialyzed into a buffer containing 50 mM sodium phosphate, 100 mM NaCl, 1 mM DTT at pH 7 and degassed before the experiments.

Table S4.4: Duplicate set of NRF2 E78P peptide to the human Kelch domain thermodynamic binding parameters

Protein	n[b]	K_d[c] (10^{-9} M)	ΔH[c] (kcal/mol)	$T\Delta S$[c] (kcal/mol)	ΔG[c] (kcal/mol)
NRF2 site 1 E78P	1.04	5.95 ± 0.24	-16.65 ± 0.064	-5.367	- 11.28 ± 0.04

[a] The peptide sequences are shown in Table 4.1.
[b] Binding stoichiometry.
[c] K_d is the dissociation constant. ΔH, ΔS and ΔG are the change in enthalpy, entropy and Gibbs free energy upon binding at T=298.15 K, respectively.

NMR experiments and MD simulations were used to examine the free state structures of the E78P peptide. The chemical shifts assignments are available (Table S4.3). The assigned ^{1}H-^{15}N HSQC spectrum, NOESY contacts between protons >2 residues apart, 1/r^6 averaged distances, extracted from the MD simulation, and final MD structures are shown in Figure 4.5. The data confirms that the E78P peptide is able to adopt a hairpin structure with similar cross-strand contacts as the natural peptide (Figure 4.5). Overall, these findings provide new insights into the interplay between entropy and enthalpy in regulating interactions between IDPs and targets. Such information is useful for understanding how LMs can be regulated in IDPs in general, and provides guidance for making additional modifications the LM region of NRF2 in order to further increase its Kelch domain binding affinity.
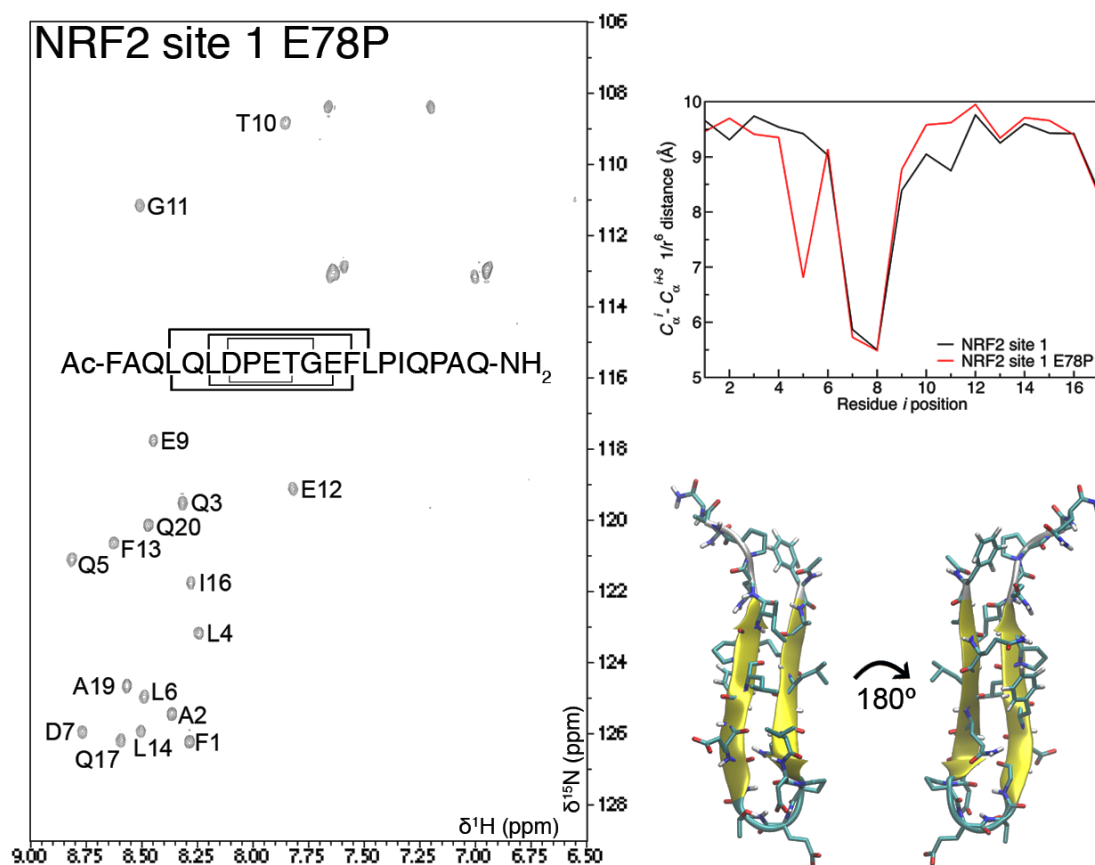
Figure 4.5 Assigned ^{1}H-^{15}N HSQC spectrum, NOESY connections, 1/r^{6} averaged distances and final structure from the MD simulations of the E78P peptide.

The NMR experiments were performed at 10ºC with a 3 mM peptide sample in 50 mM sodium phosphate pH 7, 100 mM NaCl, 1 mM DTT and trace DSS for chemical shift referencing. Lines mapped onto the peptide sequences are used to illustrate any NOESY cross-peaks between protons > 2 residues apart. The averaged 1/r^{6} distances were calculated using the g_rmsdist tool in GROMACS 4.5 (86) over the last 0.5 µs of the trajectory. The final (*t*=1.0 µs) structure of the peptide was rendered with VMD (94).

The development of higher affinity Kelch domain ligands, which can compete with NRF2, is an area of active research. Several NRF2 inducers (eg. bardoxolone methyl) are currently in development or clinical trials for the treatment of chronic kidney disease,

diabetes, cancer prevention, multiple sclerosis and oxidative tissue damage (76-80). However, to our knowledge, none of these compounds disrupt the NRF2-Kelch domain interaction directly. Instead, data suggests that these triterpenoids prevent NRF2 ubiquitination by reacting with specific thiol groups on the surface of Keap1 (76). Interestingly, these NRF2 inducers may actually promote interaction of NRF2 with Keap1 (76). While compounds such as bardoxolone methyl have shown positive clinical results, it is conceivable that such electrophilic compounds could covalently bind to cysteines of off-target proteins. Compounds, such as our E78P peptide, which directly and specifically bind to the NRF2 binding site on the Kelch domain through non-covalent interactions can be an alternative therapeutic agents. Head-to-tail cyclization or attachment of our E78P peptide to a cell-penetration peptide may further improve this peptide as a drug candidate (81).

Binding affinity correlations

Because all experiments and simulations were conducted with the same conditions for each peptide, we were able to assess correlations between the affinities of binding and various physical properties (Figure 4.6). We found that there was a good correlation ($r^2 = 0.77$) between Kyte Doolittle hydropathy index (higher values indicates more hydrophobic content in the sequence) and the free energy difference (ΔG) of binding (Figure 4.6A). The two main outliers in this correlation were the p62 and WTX pS286 peptides. For p62, this deviation is probably due to the lack of an E in position $i+2$. If an E was present at this position, the 6-mer motif of p62 gets converted to that of our NRF2 E78P peptide and would be expected to increase its binding affinity substantially to a level comparable the Tier 2 binders. The overestimated affinity of the WTX pS286 peptide was understandable considering the Kyte Doolittle scale does not include values for phosphorylated amino acids. Although phosphoserine was assigned the maximum negative value on the scale (-4.5), equivalent to arginine, a more negative value is probably appropriate.
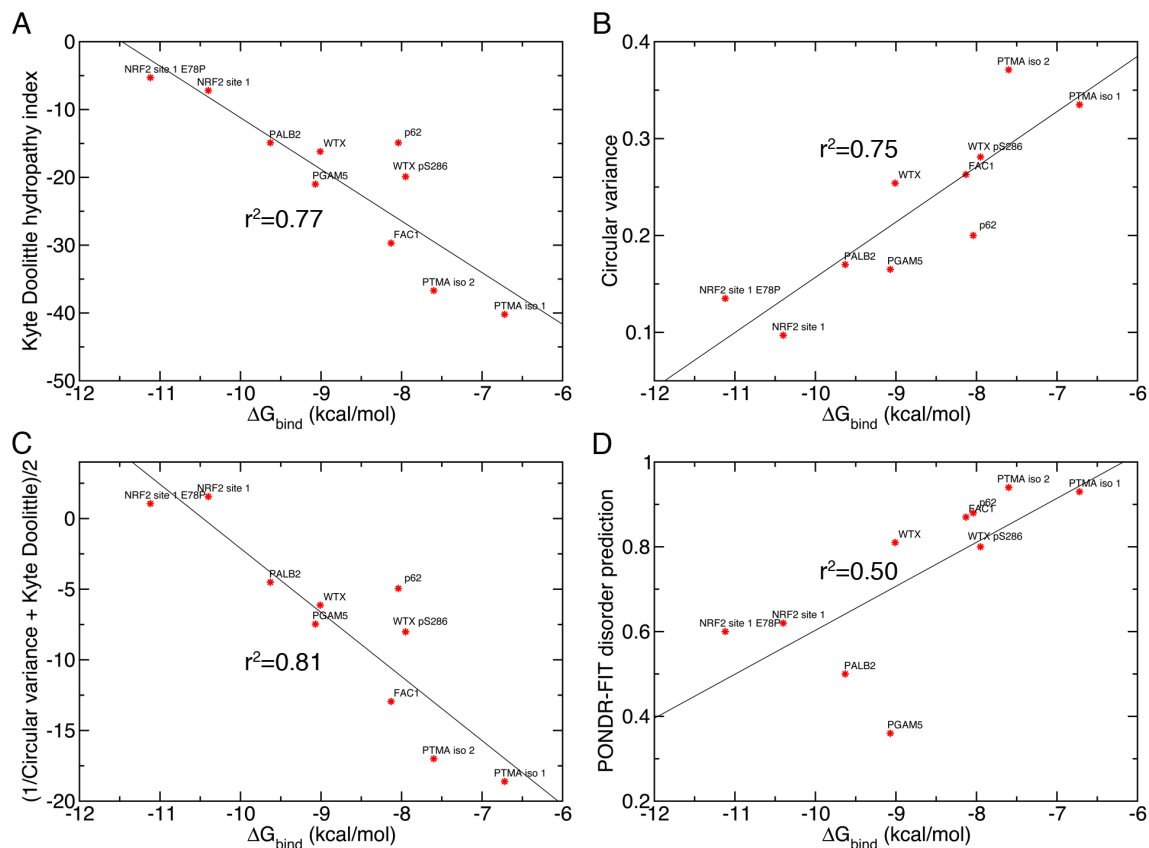
Figure 4.6 Correlations between ΔG's of binding and hydropathy, disorderness, and circular variances.

ΔG values were determined by ITC (Table 4.2) and plotted against Kyte Doolittle hydropathy indexes (A), average circular variances (B), combined circular variances and Kyte Doolittle indexes (C), and PONDR-FIT disorder predictions (D) of the peptides (Figure 4.1). Kyte Doolittle hydropathy index values were obtained from (95). For the WTX pS286 peptide, phosphoserine was assigned the maximum hydrophilic value on the scale (-4.5). The circular variances were calculated over the last 0.5 μs of the MD trajectories, using the method described by (96). Disorder predictions were performed on the full-length sequences and the average values for the segments in Figure 4.1A were plotted. For the WTX pS286 peptide, phosphoserine was changed to glutamic acid.

Circular variance values, a measure of backbone dynamics, extracted from the MD simulations, were also well correlated ($r^2 = 0.75$) with the ΔG of binding (Figure 4.6B). The trend clearly illustrates that lower free state dynamics corresponds to a higher binding affinity. By combining hydropathy indices with circular variances (Figure 4.6C), the correlation was moderately improved ($r^2 = 0.81$). The binding affinities were correlated to a lesser extent ($r^2 = 0.50$) to the PONDR-FIT disorder tendencies (Figure 4.6D). The major outlier here was PGAM5, with an average disorder probability of 0.36 for the 20-mer region of its sequence (Figure 4.6D). However, another disorder predictor, metaPrDOS (67) yielded a disorder probability of 0.59, which improves the correlation considerably $r^2 = 0.70$ (data not shown).

By establishing important parameters that are well correlated with the ΔG's of binding, it may be possible to identify novel LMs capable of binding Kelch and to assess their possible interaction strengths. One potential partner is the DNA replication licensing factor MCM3, which was found to have a physical interaction (by two hybrid methods) with Keap1 in Fruit fly (6). Although the function of this interaction is not known and experimental studies are required to confirm this interaction in humans, the human MCM3 sequence does harbor a site 1 type sequence, [379]TAAVTTDQETGERRLEAGAM[398]. This peptide has a hydropathy index of -13 and an MD simulation of this sequence did indeed form a hairpin conformation at the expected location and had an average C.V. value of 0.18. These diagnostics place it in the realm of the Tier 2 binders, but this needs to be experimentally verified.

4.6 Conclusions

Our findings suggest that intrinsic disorder coupled with a preformed β-turn resembling LM located at the binding site is a common feature among many of the Kelch domain interacting proteins. The LMs are differentially stabilized by intra-turn contacts and interactions between residues in close proximity to the binding region. The extent of turn stabilization is likely an important factor in modulating binding affinity. We found that the hydropathic indices and circular variance measurements of the peptides were well

correlated with the measured free energy difference of binding. These parameters may be useful for prediction the affinities of other possible Kelch domain interactions. Based on this knowledge, we have selectively mutated the turn region of the NRF2 site 1 peptide (E78P) to increase its binding affinity. The increase in affinity resulted from a lower entropic cost of binding, possibly due to a reduction in the conformational freedom of the free state. Importantly, this modified higher affinity peptide may have potential therapeutic applications. This work illustrates a simple, yet effective methodology for structurally characterizing LMs in the context of target binding. The results should also be useful for determining the biological roles of the various Kelch domain interactions and development of specific NRF2 inducers.

4.7 Acknowledgements

4.8 References

1. Itoh K et al. (1999) Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain. *Genes Dev* 13:76-86.

2. Stępkowski TM, Kruszewski MK (2011) Molecular cross-talk between the NRF2/KEAP1 signaling pathway, autophagy, and apoptosis. *Free Radic Biol Med* 50:1186-95.

3. Tong KI, Kobayashi A, Katsuoka F, Yamamoto M (2006) Two-site substrate recognition model for the Keap1-Nrf2 system: a hinge and latch mechanism. *Biol Chem* 387:1311-20.

4. Tong KI et al. (2006) Keap1 recruits Neh2 through binding to ETGE and DLG motifs: characterization of the two-site molecular recognition model. *Mol Cell Biol* 26:2887-900.

5. Li Y, Paonessa JD, Zhang Y (2012) Mechanism of chemical activation of Nrf2. *PLoS One* 7:e35122.

6. Padmanabhan B et al. (2006) Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer. *Mol Cell* 21:689-700.

7. Ma J et al. (2012) PALB2 interacts with KEAP1 to promote NRF2 nuclear accumulation and function. *Mol Cell Biol* 32:1506-17.

8. Lo SC, Hannink M (2006) PGAM5, a Bcl-XL-interacting protein, is a novel substrate for the redox-regulated Keap1-dependent ubiquitin ligase complex. *J Biol Chem* 281:37893-903.

9. Komatsu M et al. (2010) The selective autophagy substrate p62 activates the stress responsive transcription factor Nrf2 through inactivation of Keap1. *Nat Cell Biol* 12:213-23.

10. Camp ND et al. (2012) Wilms Tumor Gene on X Chromosome (WTX) Inhibits Degradation of NRF2 Protein through Competitive Binding to KEAP1 Protein. *J Biol Chem* 287:6539-50.

11. Strachan GD et al. (2004) Fetal Alz-50 clone 1 interacts with the human orthologue of the Kelch-like Ech-associated protein. *Biochemistry* 43:12113-22.

12. Padmanabhan B, Nakamura Y, Yokoyama S (2008) Structural analysis of the complex of Keap1 with a prothymosin alpha peptide. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 64:233-8.

13. Lee DF et al. (2009) KEAP1 E3 ligase-mediated downregulation of NF-kappaB signaling by targeting IKKbeta. *Mol Cell* 36:131-40.

14. Kim JE et al. (2010) Suppression of NF-kappaB signaling by KEAP1 regulation of IKKbeta activity through autophagic degradation and inhibition of phosphorylation. *Cell Signal* 22:1645-54.

15. Niture SK, Jaiswal AK (2011) INrf2 (Keap1) targets Bcl-2 degradation and controls cellular apoptosis. *Cell Death Differ* 18:439-51.

16. Tian H et al. (2012) Keap1: One stone kills three birds Nrf2, IKKβ and Bcl-2/Bcl-xL. *Cancer Lett* 325:26-34.

17. Lo SC, Li X, Henzl MT, Beamer LJ, Hannink M (2006) Structure of the Keap1:Nrf2 interface provides mechanistic insight into Nrf2 signaling. *EMBO J* 25:3605-17.

18. Kalia M, Kukol A (2011) Structure and dynamics of the kinase IKK-β--A key regulator of the NF-kappa B transcription factor. *J Struct Biol* 176:133-42.

19. Cino EA, Wong-Ekkabut J, Karttunen M, Choy WY (2011) Microsecond molecular dynamics simulations of intrinsically disordered proteins involved in the oxidative stress response. *PLoS One* 6:e27371.

20. Cino EA, Choy WY, Karttunen M (2012) Comparison of Secondary Structure Formation Using 10 Different Force Fields in Microsecond Molecular Dynamics Simulations. *J Chem Theory Comput* 8:2725-2740.

21. Piotto M, Saudek V, Sklenár V (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* 2:661-5.

22. Delaglio F et al. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277-93.

23. Brunger AT (2007) Version 1.2 of the Crystallography and NMR system. *Nat Protoc* 2:2728-33.

24. Pettersen EF et al. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605-12.

25. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation *J Chem Theory Comput* 4:435-447.

26. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6 *Journal of Computational Chemistry* 25:1656-1676.

27. Oostenbrink C, Soares TA, van der Vegt NF, van Gunsteren WF (2005) Validation of the 53A6 GROMOS force field. *Eur Biophys J* 34:273-84.

28. Hansson T, Nordlund P, Aqvist J (1997) Energetics of nucleophile activation in a protein tyrosine phosphatase *J Mol Biol* 265:118-127.

29. Berendsen HJC, Postma JPM, and Gusteren WF (1981) in *Intermolecular forces*, p 331-342.

30. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126:014101.

31. Parrinello, Rahman (1981) Polymorphic transitions in single crystals: A new molecular dynamics method *J Appl Phys* 52:7182-7190.

32. Hess B, Bekker H, Berendsen HJC, Johannes JGEM (1997) LINCS: A linear constraint solver for molecular simulations *Journal of Computational Chemistry* 18:1463-1472.

33. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N · log (N) method for Ewald sums in large systems *J Chem Phys* 98:10089-10092.

34. Gast K et al. (1995) Prothymosin alpha: a biologically active protein with random coil conformation. *Biochemistry* 34:13211-8.

35. Uversky VN et al. (1999) Natively unfolded human prothymosin alpha adopts partially folded collapsed conformation at acidic pH. *Biochemistry* 38:15009-16.

36. Yi S, Boys BL, Brickenden A, Konermann L, Choy WY (2007) Effects of zinc binding on the structure and dynamics of the intrinsically disordered protein prothymosin alpha: evidence for metalation as an entropic switch. *Biochemistry* 46:13120-30.

37. Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN (2010) PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta* 1804:996-1010.

38. Petros AM et al. (2001) Solution structure of the antiapoptotic protein bcl-2. *Proc Natl Acad Sci U S A* 98:3012-7.

39. Tong KI et al. (2007) Different electrostatic potentials define ETGE and DLG motifs as hinge and latch in oxidative stress response. *Mol Cell Biol* 27:7511-21.

40. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577-637.

41. Hutchinson EG, Thornton JM (1994) A revised set of potentials for beta-turn formation in proteins. *Protein Sci* 3:2207-16.

42. Altschul SF et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389-402.

43. Kerrien S et al. (2012) The IntAct molecular interaction database in 2012. *Nucleic Acids Res* 40:D841-6.

44. Itoh K et al. (1997) An Nrf2/small Maf heterodimer mediates the induction of phase II detoxifying enzyme genes through antioxidant response elements *Biochem Biophys Res Commun* 236:313-322.

45. Bobrovnikova-Marjon E et al. (2010) PERK promotes cancer cell proliferation and tumor growth by limiting oxidative DNA damage. *Oncogene* 29:3881-95.

46. Clements CM, McNally RS, Conti BJ, Mak TW, Ting JPY (2006) DJ-1, a cancer-and Parkinson's disease-associated protein, stabilizes the antioxidant transcriptional master regulator Nrf2 *Proceedings of the National Academy of Sciences* 103:15091-15096.

47. Kobayashi A et al. (2006) Oxidative and electrophilic stresses activate Nrf2 through inhibition of ubiquitination activity of Keap1. *Mol Cell Biol* 26:221-9.

48. Sykiotis GP, Bohmann D (2010) Stress-activated cap'n'collar transcription factors in aging and human disease. *Sci Signal* 3:re3.

49. Sporn MB et al. (2011) New synthetic triterpenoids: potent agents for prevention and treatment of tissue injury caused by inflammatory and oxidative stress. *J Nat Prod* 74:537-45.

50. Sporn MB, Liby KT (2012) NRF2 and cancer: the good, the bad and the importance of context. *Nat Rev Cancer* 12:564-71.

51. Zhang Y, Gordon GB (2004) A strategy for cancer prevention: stimulation of the Nrf2-ARE signaling pathway. *Mol Cancer Ther* 3:885-93.

52. Rojas-Rivera J, Ortiz A, Egido J (2012) Antioxidants in kidney diseases: the impact of bardoxolone methyl. *Int J Nephrol* 2012:321714.

53. Steel R, Cowan J, Payerne E, O'Connell MA, Searcey M (2012) Anti-inflammatory Effect of a Cell-Penetrating Peptide Targeting the Nrf2/Keap1 Interaction. *ACS Med Chem Lett* 3:407-410.

54. Schneider TD, Stephens RM (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* 18:6097-100.

55. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14:1188-90.

56. Larkin MA et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947-8.

57. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14:33-8, 27-8.

58. MacArthur MW, Thornton JM (1993) Conformational analysis of protein structures derived from NMR data. *Proteins* 17:232-51.

59. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157:105-32.

60. Giot L et al. (2003) A protein interaction map of Drosophila melanogaster. *Science* 302:1727-36.

5   **[1]H, ^{15}N and ^{13}C backbone resonance assignments of the Kelch domain of mouse Keap1**

Elio Cino[1], Jingsong Fan[2], Daiwen Yang[2], and Wing-Yiu Choy[1*]

[1]*Department of Biochemistry, The University of Western Ontario, London, Ontario, Canada N6A 5C1*

[2]*Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543*

*Corresponding author

E-mail: jchoy4@uwo.ca

Link: http://www.springerlink.com/content/l4813x05171g7n17/

5.1 Abstract

Keap1 is a multi-domain protein that functions as an inhibitor of the transcription factor Nrf2 in the cellular response to oxidative stress. Under normal conditions, Keap1 binds to Nrf2 via its C-terminal Kelch domain and the interaction ultimately leads to the ubiquitin-dependent degradation of Nrf2. It has been proposed that designing molecules to selectively disrupt the Keap1-Nrf2 interaction can be a potential therapeutic approach for enhancing the expression of cytoprotective genes. Here, we reported the ^{1}H, ^{13}C, and ^{15}N backbone chemical shift assignments of the Kelch domain of mouse Keap1. Further, unlabeled Nrf2 peptide containing the Kelch-binding motif was added to the ^{15}N-labeled Kelch sample. ^{1}H-^{15}N HSQC spectra of the protein in the absence and presence of an equimolar concentration of the Nrf2 peptide were presented. A significant number of resonance signals were shifted upon addition of the peptide, confirming the protein-peptide interaction. The results here will not just facilitate the further studies of the binding between Keap1 and Nrf2, it will also be valuable for probing interactions between the Kelch domain and small molecules, as well as a growing list of protein targets that have been identified recently.

Keywords

Oxidative stress response; Keap1; Kelch domain; Nrf2; protein-protein interaction

5.2 Biological context

Keap1 (Kelch-like ECH-associated Protein 1) is a 70-kDa protein that is rich in cysteine. It functions as the repressor of Nrf2 (nuclear factor E2-related factor 2), the key transcription factor that coordinates the cellular responses to oxidative stress. Keap1 is composed of three domains (1). The N-terminal BTB (Broad-Complex, Tramtrack and Bric-a-Brac) domain mediates the protein dimerization and is also responsible for the binding to the Cul3-dependent E3 ubiquitin ligase complex (1). The BTB domain is linked to the C-terminal Kelch domain by the intervening region (IVR), where cysteines that are pivotal for the Keap1 functions as chemical sensor are located (1). The C-terminal Kelch domain is responsible for target recognition. It directly interacts with the ETGE and DLG motifs present in the N-terminal Neh2 domain of Nrf2 (2-4). Under unstressed conditions, Keap1 binds to Nrf2 and targets this transcription factor for degradation (5).

The 32-kDa Kelch domain adopts a six-bladed β-propeller conformation and the structures in complex with peptides derived from Nrf2 that contain the binding motifs have been determined by X-ray crystallography (6-8). Due to the critical roles Keap1 and Nrf2 play in the oxidative stress response mechanism, disrupting their interaction has been proposed as a potential strategy to enhance the expression of cytoprotective genes. The backbone resonance assignments obtained here will facilitate the screening of inhibitors that can disrupt the Keap1-Nrf2 interaction. In addition to Nrf2, several novel targets of the Kelch domain of Keap1 have been identified recently. These include FAC1 (9), PGAM5 (10), WTX (11, 12), prothymosin α (13), p62 (14) and PALB2 (15). The assignments reported here will also serve as important starting points for studying the interaction of Keap1 with these proteins.

5.3 Methods and experiments

Sample preparation

The mouse Keap1 cDNA (GenBank# BC055732) was purchased from the American Type Culture Collection (ATCC) and the Kelch domain (residues 324-612) was cloned into the expression vector pDEST17 (Invitrogen). The TEV cleavage recognition sequence, 'ENLYFQG', was introduced between the histidine tag and insert. This vector was transformed into *E. coli* BL21(DE3). ^{15}N, ^{13}C and ^{2}H labeled protein was produced by growing *E. coli* in deuterated M9 media. Cells adapted in 70% D_2O were used to inoculate 1L of M9 prepared in 100% D_2O. The cells were then grown at 37ºC and overexpression was induced at an OD_{600} of 0.6 with 0.5 mM IPTG. After a 60-hour induction, the cells were centrifuged and the pellets were frozen. The N-terminally His-tagged protein was purified by affinity chromatography using Ni Sepharose™ 6 Fast Flow beads (Amersham Biosciences). The tag was cleaved by incubation with His-tagged tobacco etch virus (TEV) protease overnight at 25 ºC and the protein product was purified by passing the mixture through Ni Sepharose™ 6 Fast Flow beads (Amersham Biosciences). For NMR experiments, the purified protein was dialyzed against 50 mM sodium phosphate at pH 7 containing 50 mM NaCl and 1 mM DTT. The samples were then concentrated to ~0.3 mM. All samples contained 10% D_2O and 1 mM 2,2-dimethyl-2-sila-pentane-5-sulfonic acid (DSS) as ^{1}H and ^{13}C chemical shift references.

Sequential assignment experiments

NMR experiments for the backbone resonance assignment were conducted at 25ºC on Varian INOVA 800 MHz (NANUC) and Bruker Avance 800 MHz (Singapore) spectrometers equipped with cryogenic probes. Sequential assignments were obtained from ^{1}H-^{15}N TROSY-HSQC, HNCA, HN(CO)CA, HNCACB, HN(CO)CACB and ^{15}N-NOESY-HSQC spectra. The data was processed with NMRPipe (16) and analyzed using CARA (17).

Ligand binding experiment

To demonstrate the binding of the Kelch domain to a mouse Nrf2 peptide (Ac-^{72}AQFQLDEETGEFLP^{85}-NH$_2$; ordered from NEOpeptide), ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled Kelch domain (~ 200 μM) were collected on a Varian Inova 600 MHz spectrometer with cryogenic probe (UWO Biomolecular NMR Facility) at 25ºC in the absence and presence of equimolar concentration of the Nrf2 peptide.

5.4 Assignments and data deposition

The ^{1}H-^{15}N TROSY-HSQC of the Kelch domain of Keap1 (Figure 5.1a) had well dispersed peaks for a protein of this size. We were able to assign 91.3% of the ^{1}H^N and ^{15}N resonances of non-proline residues, 90.7% of all ^{13}Cα and 90% all ^{13}Cβ resonances. Repetitive sequences of the β-propeller structure (Figure 5.1b) made it difficult to obtain a higher percentage of resonance assignments. The ^{1}H, ^{15}N and ^{13}C α/β chemical shifts of the backbone resonances have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu), under BMRB accession number 18353.

a

b

```
     324                         340                              360                        380
GVGRLIYTAGGYFRQSLSYLEAYNPSNGSWLRLADLQVPRSGLAGCVVGGLLYAVGGR
NNSPDGNTDSSALDCYNPMTNQWSPCASMSVPRNRIGVGVIDGHIYAVGGSHGCIHHS
     440                         460                        480
SVERYEPERDEWHLVAPMLTRRIGVGVAVLNRLLYAVGGFDGTNRLNSAECYYPERNE
     500                  520                              540
WRMITPMNTIRSGAGVCVLHNCIYAAGGYDGQDQLNSVERYDVETETWTFVAPMRHHR
     560                  580                              600        612
SALGITVHQGKIYVLGGYDGHTFLDSVECYDPDSDTWSEVTRMTSGRSGVGVAVTMEP
```

Figure 5.1 ¹H-¹⁵N TROSY-HSQC spectrum and backbone resonance assignment of the ²H/¹³C/¹⁵N labeled Kelch domain of mouse Keap1.

a ^{1}H-^{15}N TROSY-HSQC spectrum and backbone resonance assignment of the ^{2}H/^{13}C/^{15}N labeled Kelch domain of mouse Keap1. The figure was generated using Sparky (21). **b** Protein sequence of the Kelch domain with unassigned resonances colored red and regions with high sequence identities boxed. The starting 'G' was a non-native residue from the TEV recognition sequence that remained after cleavage.

Figure 5.2 shows the residue-specific secondary structure propensity (SSP) scores determined using the program SSP on the basis of the assigned ^{13}C α/β chemical shifts (18). The result strongly indicates that the protein has an all-β fold (Figure 5.2). This is in good agreement with the crystal structure of the Kelch domain (PDB id: 1X2J), which shows that the protein adopts a six-bladed β-propeller conformation (7). The SSP scores we obtained here are also consistent with the DSSP analysis (19) and the secondary structure plot (20) of the crystal structure (7).
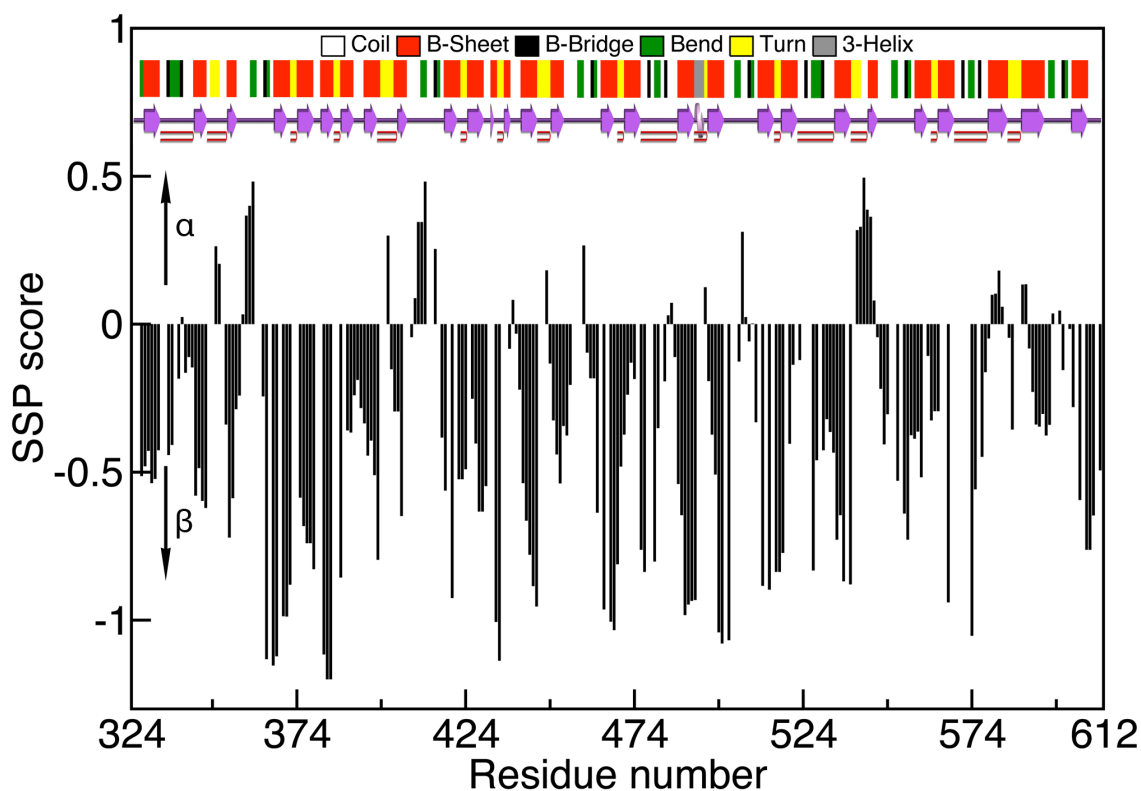
Figure 5.2 Secondary structure propensity (SSP) scores and DSSP analysis of the mouse Kelch domain of Keap1.

SSP scores were calculated based upon the [13]C α/β chemical shifts (18). The crystal structure of the mouse Kelch domain (PDB id: 1X2J) (7) was used for the DSSP analysis (19) and generation of the secondary structure cartoon diagram (20).

The overlay of the ^{1}H-^{15}N HSQC spectra of the Kelch domain in the absence and presence of an equimolar concentration of the 14-mer Nrf2 peptide is shown in Figure 5.3a. A significant number of resonance signals are shifted upon addition of the peptide (Figure 3a). To quantify the magnitude of peak shifts, combined chemical shift changes ($\Delta\omega = |\Delta^{15}N| + |\Delta^{1}H^{N}|$) were calculated, where $|\Delta^{15}N| + |\Delta^{1}H^{N}|$ are the absolute values of resonance frequency change (in Hz) in the ^{15}N and ^{1}H dimensions, respectively. Figure 5.3b shows that many of the traceable assigned residues with $\Delta\omega > 50$ Hz (in descending

order: G378, N504, G364, G477, G600, G570, G571, G417, G524, G379 and S431) are residues that comprise, or are in close proximity, to the binding interface as identified in the crystal structure (7).
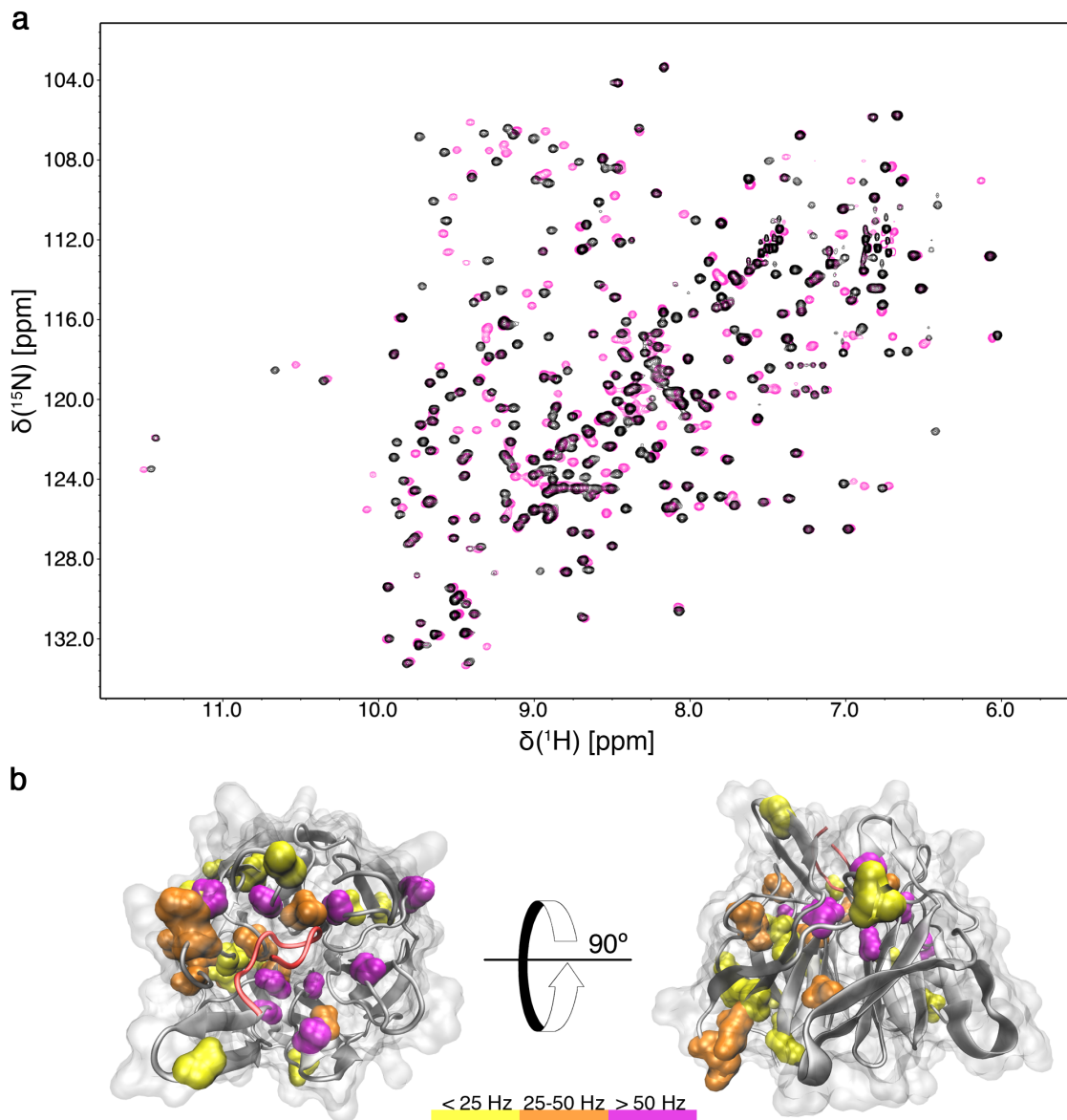

Figure 5.3 Overlay of ^{1}H-^{15}N HSQC spectra in the absence (black) and presence (pink) of an equimolar concentration of the Nrf2

a Overlay of ^{1}H-^{15}N HSQC spectra in the absence (black) and presence (pink) of an equimolar concentration of the Nrf2 peptide (Ac-^{72}AQFQLDEETGEFLP^{85}-

NH$_2$). **b** Crystal structure of the mouse Kelch domain of Keap1 (grey) in complex with an Nrf2 peptide (red) (PDB id: 1X2R) (7). Residues with traceable assigned resonances are colored based on their combined absolute proton and nitrogen resonance frequency changes (Hz) upon ligand binding (yellow <25 Hz, orange 25-50 Hz and pink >50 Hz).

5.5 Acknowledgements

5.6 References

1. Itoh K et al. (1999) Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain. *Genes Dev* 13:76-86.

2. Baird L, Dinkova-Kostova AT (2011) The cytoprotective role of the Keap1-Nrf2 pathway. *Arch Toxicol* 85:241-72.

3. Tong KI et al. (2006) Keap1 recruits Neh2 through binding to ETGE and DLG motifs: characterization of the two-site molecular recognition model. *Mol Cell Biol* 26:2887-900.

4. Zipper LM, Mulcahy RT (2002) The Keap1 BTB/POZ dimerization function is required to sequester Nrf2 in cytoplasm. *J Biol Chem* 277:36544-52.

5. Itoh K, Tong KI, Yamamoto M (2004) Molecular mechanism activating Nrf2-Keap1 pathway in regulation of adaptive response to electrophiles. *Free Radic Biol Med* 36:1208-13.

6. Lo SC, Li X, Henzl MT, Beamer LJ, Hannink M (2006) Structure of the Keap1:Nrf2 interface provides mechanistic insight into Nrf2 signaling. *EMBO J* 25:3605-17.

7. Padmanabhan B et al. (2006) Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer. *Mol Cell* 21:689-700.

8. Tong KI et al. (2007) Different electrostatic potentials define ETGE and DLG motifs as hinge and latch in oxidative stress response. *Mol Cell Biol* 27:7511-21.

9. Strachan GD et al. (2004) Fetal Alz-50 clone 1 interacts with the human orthologue of the Kelch-like Ech-associated protein. *Biochemistry* 43:12113-22.

10. Lo SC, Hannink M (2006) PGAM5, a Bcl-XL-interacting protein, is a novel substrate for the redox-regulated Keap1-dependent ubiquitin ligase complex. *J Biol Chem* 281:37893-903.

11. Major MB et al. (2007) Wilms tumor suppressor WTX negatively regulates WNT/beta-catenin signaling. *Science* 316:1043-6.

12. Camp ND et al. (2012) Wilms Tumor Gene on X Chromosome (WTX) Inhibits Degradation of NRF2 Protein through Competitive Binding to KEAP1 Protein. *J Biol Chem* 287:6539-50.

13. Padmanabhan P, Nakamura N, Yokoyama Y (2008) *Acta. Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* 64:233.

14. Komatsu M et al. (2010) The selective autophagy substrate p62 activates the stress responsive transcription factor Nrf2 through inactivation of Keap1. *Nat Cell Biol* 12:213-23.

15. Ma J et al. (2012) PALB2 interacts with KEAP1 to promote NRF2 nuclear accumulation and function. *Mol Cell Biol* 32:1506-17.

16. Delaglio F et al. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277-93.

17. Keller R (2004) *The Computer Aided Resonance Assignment Tutorial* (CANTINA Verlag).

18. Marsh JA, Singh VK, Jia Z, Forman-Kay JD (2006) Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci* 15:2795-804.

19. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577-637.

20. Laskowski RA (2009) PDBsum new things. *Nucleic Acids Res* 37:D355-9.

21. Goddard TD, Kneller DG, SPARKY 3, University of California, San Francisco.

6 Effects of Molecular Crowding on the Dynamics of Intrinsically Disordered Proteins

Elio A. Cino[†], Mikko Karttunen[‡], and Wing-Yiu Choy[†*]

Department of Biochemistry[†], The University of Western Ontario, London, Ontario, Canada N6A 5C1, Department of Chemistry[‡], University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

*Corresponding author: Wing-Yiu Choy <jchoy4@uwo.ca

6.1 Abstract

Inside cells, the concentration of macromolecules can reach up to 400 g/L. In such crowded environments, proteins are expected to behave differently than *in vitro*. It has been shown that the stability and the folding rate of a globular protein can be altered by the excluded volume effect produced by a high density of macromolecules. However, macromolecular crowding effects on intrinsically disordered proteins (IDPs) are less explored. These proteins can be extremely dynamic and potentially sample a wide ensemble of conformations under non-denaturing conditions. The dynamic properties of IDPs are intimately related to the timescale of conformational exchange within the ensemble, which govern target recognition and how these proteins function. In this work, we investigated the macromolecular crowding effects on the dynamics of several IDPs by measuring the NMR spin relaxation parameters of three disordered proteins (ProTα, TC1, and α-synuclein) with different extents of residual structures. To aid the interpretation of experimental results, we also performed an MD simulation of ProTα. Based on the MD analysis, a simple model to correlate the observed changes in relaxation rates to the alteration in protein motions under crowding conditions was proposed. Our results show that 1) IDPs remain at least partially disordered despite the presence of high concentration of other macromolecules, 2) the crowded environment has differential effects on the conformational propensity of distinct regions of an IDP, which may lead to selective stabilization of certain target-binding motifs, and 3) the segmental motions of IDPs on the nanosecond timescale are retained under crowded conditions. These findings strongly suggest that IDPs function as dynamic structural ensembles in cellular environments.

6.2   Introduction

Macromolecular crowding and confinement can have significant impacts on the behaviors of proteins in cellular environments. Inside of cells, the concentration of macromolecules can reach up to 400 g/L (1, 2). The cumulative excluded volume from all macromolecules inside of cells is commonly referred to as macromolecular crowding (3, 4). The large volume occupied by macromolecules in the cellular environment exerts nonspecific forces on surrounding molecules (3). It is well documented that these forces can have significant effects on the behaviors of proteins (5-7).

Experimental studies have demonstrated that molecular crowding can affect protein structure and function. For example, at low pH, cytochrome c adopts an unfolded form. When the crowding agent dextran is added to the sample, the protein transitions into a near-native molten globule state (8). Crowding has also been shown to enhance the activity of phosphoglycerate kinase (PGK) in vitro. At a mild concentration of Ficoll 70 (100 g/L), the enzymatic activity of PGK was found to increase by more than 10 fold (after the viscosity effect was taken into account), possibility due to the large-scale of conformational changes induced by the crowders (9). In another study, Stagg et al. (10) investigated effects of crowding on the structure and stability of both the native and denatured states of Flavodoxin. Interestingly, their experimental and computer simulation results indicate that the presence of a high concentration of Ficoll 70 in solution increased the thermal stability and secondary structure content of the native-state ensemble, but had relatively minor effects on the denatured state (10).

The crowded environment in cells also alters the diffusional behavior of proteins, and thus their rates of folding, association with other molecules and intracellular transport (11, 12). A recent work by Leduc et al. (13) suggested that different motor proteins, such as kinesins, process distinct molecular properties in order to operate effectively in the crowded cellular environments. Macromolecular crowding has also been proposed to be one of the possible factors that regulate the phosphorylation of ERK kinase in cells. Aoki et al. (14) demonstrated that under crowded conditions, the phosphorylation of ERK could switch from the distributive to processive mode. Further, experimental and molecular simulation studies suggested that crowding plays a key role in human diseases

that are related to protein aggregation and fibril formation (15-17). For instance, the amyloid formation of human and bovine prion proteins are significantly enhanced even at mild concentration (150-200 g/L) of Ficoll 70. Intriguingly, the amyloid formation of rabbit prion protein is inhibited by crowding agents (17, 18).

The effects of macromolecular crowding on the structure and dynamics of IDPs, on the other hand, are less explored. These proteins lack stable tertiary structures and can be very flexible under non-denaturing conditions. The functions of IDPs are intimately related to their dynamics (19). It has been proposed that proteins with disordered regions have larger capture radius for targets, therefore, enhancing the binding rates by the so-called "fly-casting" mechanism (20). Flexibility of IDPs also governs the affinity of target recognition. The high entropic cost of disorder-to-order transition upon binding needs to be compensated by specific interactions formed in the interface with target. Therefore, IDPs frequently associate with binding partners through low affinity but highly specific interactions, which are important for their functions in signal transduction and cell cycle control (21, 22). Another important link between protein flexibility and function is the rate of inter-conversion between conformers. An IDP exists as an ensemble of conformers in equilibrium (23-25). Different structures in the ensemble can participate in the interactions with distinct targets; therefore, the rate of exchange between conformers can have significant impact on the protein function (26, 27). Further, recent studies show that some IDPs employ multiple linear motifs to engage in a dynamic equilibrium with a target, resulting in ultra-sensitivity of binding (28-30). Undoubtedly, protein flexibility plays a critical role in this polyvalent mode of binding (29).

There are several studies of macromolecular crowding effects on the structure of IDPs. The results, however, are not conclusive. For instance, FlgM is disordered in dilute buffer solutions, but gains structure in its C-terminal half when studied in cells or in solutions with high concentration of glucose (31). On the other hand, Flaugh and Lumb reported that neither the disordered C-terminal activation domain of c-Fos nor the kinase-inhibition domain of p27^{Kip1} undergo any significantly conformational change in the presence of dextran or Ficoll (32). By using small-angle neutron scattering techniques, Johansen et al. (33) demonstrated that the disordered N protein of bacteriophage λ adopts

more compact conformations even in the presence of relatively low concentration of crowding agents (~65 g/L of BPTI protein). A recent work by Tompa and co-worker (34), however, shows that molecular crowding caused only minor structural changes to three IDPs (α-casein, MAP2c and p21^{Cip1}). The authors suggested that retaining dynamics under crowded conditions is a functional requirement of IDPs.

Further experimental studies of the macromolecular crowding effects on IDPs are important for increasing our understanding of how these proteins behave in cellular environments. These studies will also facilitate the development of computational models that can be used to explain and predict the behaviors these proteins under crowded conditions (5, 34, 35). We focus on assessing the effects of macromolecular crowding on the dynamics of IDPs in residue-specific manner using NMR spin relaxation experiments. Three IDPs with different extents of residual structure under dilute buffer conditions were studied. Further, by using one of the IDPs (ProTα) as a representative case, based on an MD simulation, we proposed a model to correlate the observed changes in relaxation rates to the possible alteration in protein motions under crowding conditions. ProTα is a ubiquitously expressed, highly acidic IDP that is involved in multiple biological functions (36-38). Our recent studies demonstrated that ProTα is largely disordered with minimal residual structure present under non-denaturing conditions (39, 40). Although ProTα adopts an extended structure, it can convert to more compact conformations in the presence of zinc ions (40). Another IDP used in this study is Thyroid Cancer 1 (TC-1), which was first found to be overexpressed in thyroid cancer (41, 42). TC-1 is a basic protein and is a positive regulator of the Wnt/β-catenin signaling pathway (42-44). It competes with β-catenin on binding to Chibby (Cby) and therefore inhibits the antagonistic action of Cby on β-catenin mediated transcription (44, 45). Even though TC-1 is classified as an IDP, structural characterization shows that while the N-terminal half of the protein is largely unstructured, high helical propensity is present in the C-terminal part (42, 46). α-synuclein, a well-studied IDP that has been found to be the main structural component of Lewy body fibrils found in patients with Parkinson's disease (47), was also included in this study to add additional depth to our approach. α-synuclein is natively disordered in its soluble form, but is able to self-associate to form insoluble aggregates that have considerable structure (47). In-cell NMR experiments have shown

that the periplasmic environment in *Escherichia coli* prevents α-synuclein from undergoing a conformational change that is detected in dilute buffer conditions, indicating that the crowding acts to keep α-synuclein disordered (48). In addition to the IDPs mentioned above, we also assessed the crowding effect on a well-studied globular protein, Ubiquitin, for comparison. By performing NMR relaxation measurements on these proteins we aim to determine how the dynamics of IDPs with different structural characteristics can be affected by macromolecular crowding.

6.3   Materials and methods

Protein expression and purification

Uniformly ^{15}N labeled ProTα (human isoform 2), TC-1 (human) and α-synuclein (human isoform 1) were expressed in *Escherichia coli* BL21 (DE3) cells grown in minimal M9 medium containing ^{15}NH$_4$Cl (Cambridge Isotope Laboratories) as the sole nitrogen source. ^{15}N/^{13}C labeled TC-1 was expressed as above except with ^{13}C$_6$-D-glucose (Isotec) as the sole carbon source. ProTα was purified using the method described by Yi et al. (39). The N-terminally His tagged TC-1 protein was extracted from inclusion bodies using 6 M guanidine hydrochloride and purified by affinity chromatography using Ni Sepharose™ 6 Fast Flow beads (Amersham Biosciences) (46). The plasmid carrying the α-synuclein cDNA was kindly supplied by Dr. Pielak at the University of North Carolina-Chapel Hill. The protein was purified by osmotic shock, using a procedure similar to the one reported by Shevchik et al. (49), followed by boiling and cooling steps similar to (39). The protein was then precipitated out of solution with 60% saturated solution of ammonium sulfate. Lyophilized ^{15}N labeled human Ubiquitin was kindly supplied by Dr. Gary Shaw's lab at the University of Western Ontario.

NMR spectroscopy

All NMR experiments were performed at 25 ºC on a Varian Inova 600 MHz spectrometer (UWO Biomolecular NMR Facility) with an xyz-gradient triple resonance probe. The experiments were performed in the presence and absence of 160 g/L, and several used 400 g/L, Ficoll 70 (Sigma) or Dextran 70 (Sigma). Each NMR sample contained 10% D_2O and trace sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS, Sigma) for chemical shift referencing. Data was processed with NMRPipe (50) and spectra were visualized with NMRViewJ (51).

^{1}H-^{15}N HSQC spectra were collected using 0.2 mM ^{15}N-labeled ProTα, TC-1 and α-synuclein samples and 1 mM Ubiquitin samples in the presence or absence of crowding agent. Backbone amide resonance assignments of ProTα, TC-1, α-synuclein and Ubiquitin were obtained from (40, 46, 52, 53). The triple-resonance CBCA(CO)NH experiment was carried out using 0.3 mM TC-1 samples in the presence and absence of 160 g/L Ficoll 70 (Sigma) for ^{13}C chemical shift assignments.

Backbone ^{15}N longitudinal relaxation rate (R_1), relaxation rate in rotating frame ($R_{1\rho}$), and steady-state ^{1}H-^{15}N NOE experiments were performed using 0.2 mM of ^{15}N-labeled ProTα, and TC-1 samples and 1 mM Ubiquitin sample in the presence and absence of crowding agent in their corresponding buffers. R_1 experiments were performed with delay times 10-640 ms for ProTα and TC-1 and 10-500 ms for Ubiquitin. $R_{1\rho}$ experiments employed delay times between 10 and 150 ms for all proteins. The relax program (54, 55) was used for two-parameter exponential curve fitting of peak intensities from the R_1 and $R_{1\rho}$ data, and the calculation of R_1 and $R_{1\rho}$ relaxation rates and their associated errors. ^{15}N transverse relaxation rate (R_2) values were calculated using the R_1 and $R_{1\rho}$ rates and the offset between the resonance and carrier frequency ($\Delta\omega$) in hertz, using the equation

$$R_2 = (R_{1\rho} - R_1 \cos^2 \theta)/\sin^2 \theta \qquad (1)$$

where $\tan\theta$ = $B_{SL}/\Delta\omega$. B_{SL} (= 1.5 kHz) was the spin-lock field used in the $R_{1\rho}$ experiments. ^{1}H-^{15}N steady-state NOEs were obtained from the ratio of peak intensities

of spectra recorded with and without proton saturation. Seven and 12 s delays between scans were used for the saturated and non-saturated spectra respectively and 5 s saturation periods were used. Errors were estimated based on the ratios of background noise to the signals in the spectra.

MD simulations

We conducted an atomistic MD simulation of ProTα in its free state in order to help to interpret the NMR relaxation measurements. The starting structure was generated based upon the amino acid sequence of ProTα (human isoform 2) by simulated annealing using the Crystallography & NMR System (CNS) software package (56).

The simulation was performed using GROMACS (GROningen MAchine for Chemical Simulations) version 4 (57) with the GROMOS96 53a6 united atom force-field parameter set (58, 59). This force field has been shown to perform well in simulations of disordered proteins and membrane proteins (60-62). Protonation states of ionizable residues were assigned to their most probable state at pH 7. The starting structure was centered in a cubic box with a side length of 20 nm and periodic boundary conditions were applied. The system was solvated with simple point charge (SPC) water (63). Sodium (Na^+) and chloride (Cl^-) ions were added to make the system charge neutral and bring the salt concentration to 0.1 M. The system contained 265474 water molecules, 525 sodium and 482 chloride ions. MD simulations were performed at constant number of particles, pressure and temperature (NPT ensemble). Protein and non-protein atoms were coupled to their own temperature baths, which were kept constant at 310 K using the Parrinello-Donadio-Bussi algorithm (64). Pressure was maintained isotropically at 1 bar using the Parrinello-Rahman barostat (65). The time constants for temperature and pressure coupling were 0.1 and 0.5 ps, respectively. Prior to the production run, the energy of the system was minimized using the steepest descents method, followed by 2 ps of position-restrained dynamics with all non-hydrogen atoms restrained with a 1000 kJ mol^{-1} force constant. The timestep was set to 2 fs. Initial atom velocities were taken from a Maxwellian distribution at 310 K. All bond lengths were constrained using the LINCS

algorithm (66). Cut-off of 1.0 nm was used for Lennard-Jones interactions and the real part of the long-range electrostatic interactions, which were calculated using the Particle-Mesh Ewald (PME) method (67). For a recent review on the different methods and the importance electrostatics in simulations of biological systems, see (68). Dispersion corrections were applied for energy and pressure. 0.12 nm grid-spacing was used for PME. The MD simulation was run for 427 ns and the last 400 ns were used for analysis. During this time, temperature, pressure and potential energy values remained stable and fluctuated around their averages, without systematic drift, indicating that the system was well equilibrated.

MD Simulation analysis

Autocorrelation functions of backbone ^{1}H-^{15}N bond vectors of ProTα were extracted from the MD trajectory (region 27-427ns) (without the removal of overall tumbling) using the g_rotacf tool in GROMACS (57). Each autocorrelation function was fitted to two-, three-, or four-exponential decay curves (69-71) as shown in equation (2):

$$C(t) = \sum_{i=1}^{n} a_i \exp(-t/\tau_i)$$

(2)

where *C(t)* is the autocorrelation function at time *t*, n=2, 3, or 4, a_i and τ_i are the amplitude and time constant of the ith exponential decay term. The fitted autocorrelation functions were then used to calculate the spectral density *J(ω)* by analytical Fourier transformation (69-71):

$$J(\omega) = \frac{2}{5} \int_{0}^{\infty} \cos(\omega t) C(t) dt = \frac{2}{5} \sum_{i=1}^{n} \frac{a_i \tau_i}{1 + \omega_i^2 \tau_i^2}$$

(3)

To evaluate whether the multi-exponential model j with more parameters statistically outperforms model i in fitting the autocorrelation functions, the F-ratio of statistical F-test were calculated using the following equation:

$$F_{ij} = \frac{(\chi_i^2 - \chi_j^2)/\chi_j^2}{(D_i^2 - D_j^2)/D_j^2} \tag{4}$$

where χ_i^2 (χ_j^2) and D_i (D_j) are the sum of square deviations and degrees of freedom of model i (model j), respectively.

6.4   Results

IDPs remain disordered under crowded environments

To study the effect of macromolecular crowding on the structure and dynamics of IDPs, Ficoll 70, a commonly used crowding agent, was added to the protein samples to mimic the cellular environment (6). First, ^{1}H-^{15}N HSQC spectra of ProTα, TC-1, α-synuclein, and Ubiquitin, acquired in the absence and presence of 160 g/L of Ficoll 70, were compared. Intriguingly, the spectra of the three IDPs all display narrow peak dispersions along their ^{1}H dimension in the presence of Ficoll 70 (Figure 6.1), indicating these proteins remain disordered under this crowded condition. ^{1}H-^{15}N HSQC spectra of ProTα and TC-1 in the presence of 400 g/L crowding agent had similar extents of peak dispersion as those collected in buffer or 160 g/L Ficoll conditions (Figures S6.1 and S6.2). Minor peak shifts between dilute and crowded conditions of some residues in TC-1 were observed (Figure 6.1B). To investigate the possibility that these spectral changes were due to the crowding agents binding to TC-1, we performed isothermal calorimetry (ITC) experiments, titrating 0.1 mM TC-1 into 160 g/L crowder solutions (Figure S6.3). These measurements were not indicative of specific interactions between TC-1 and Ficoll or Dextran 70 (72).
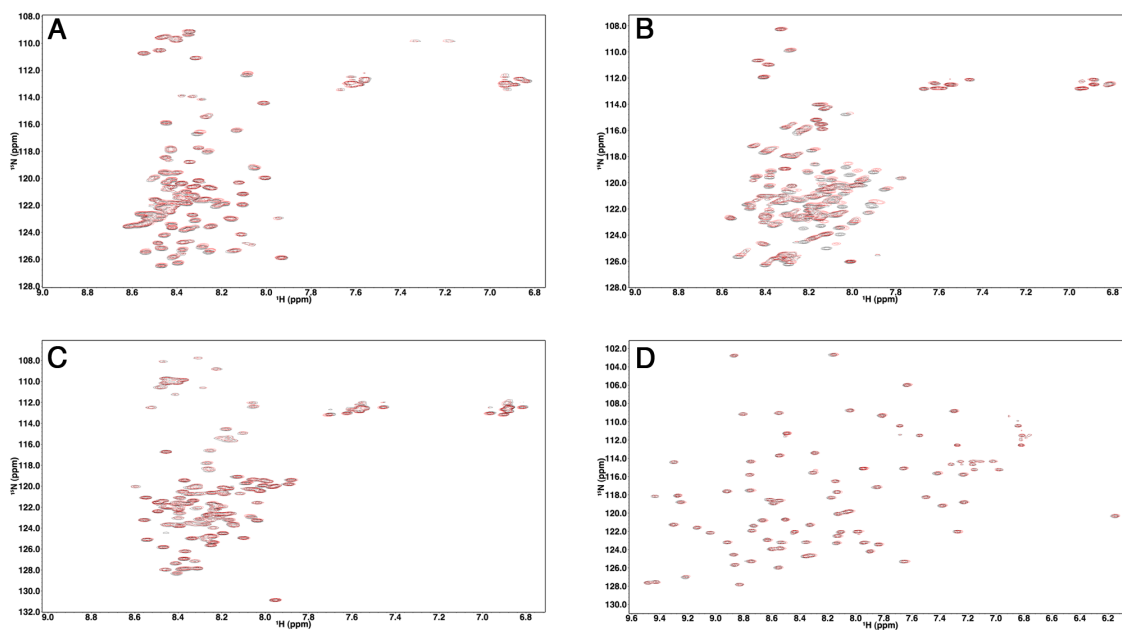
Figure 6.1 ^{1}H-^{15}N HSQC spectra of ProTα, TC-1, α-synuclein and Ubiquitn in the absence and presence of 160 g/L Ficoll 70.

ProTα (A), TC-1 (B), α-synuclein (C) and Ubiquitin (D) spectra were collected in 40 mM HEPES pH 6.8, 10 mM sodium acetate pH 5, 50 mM sodium phosphate pH 7 and 10 mM sodium acetate pH 5 respectively in the absence (black) and presence of 160 g/L Ficoll 70 (red).
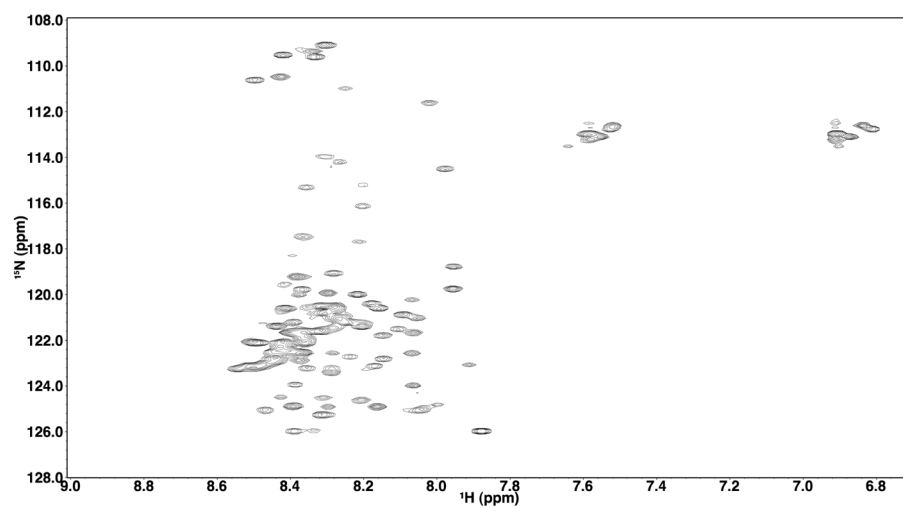
Figure S6.1 ^{1}H-^{15}N HSQC spectrum of ProTα in 400 g/L Ficoll 70.

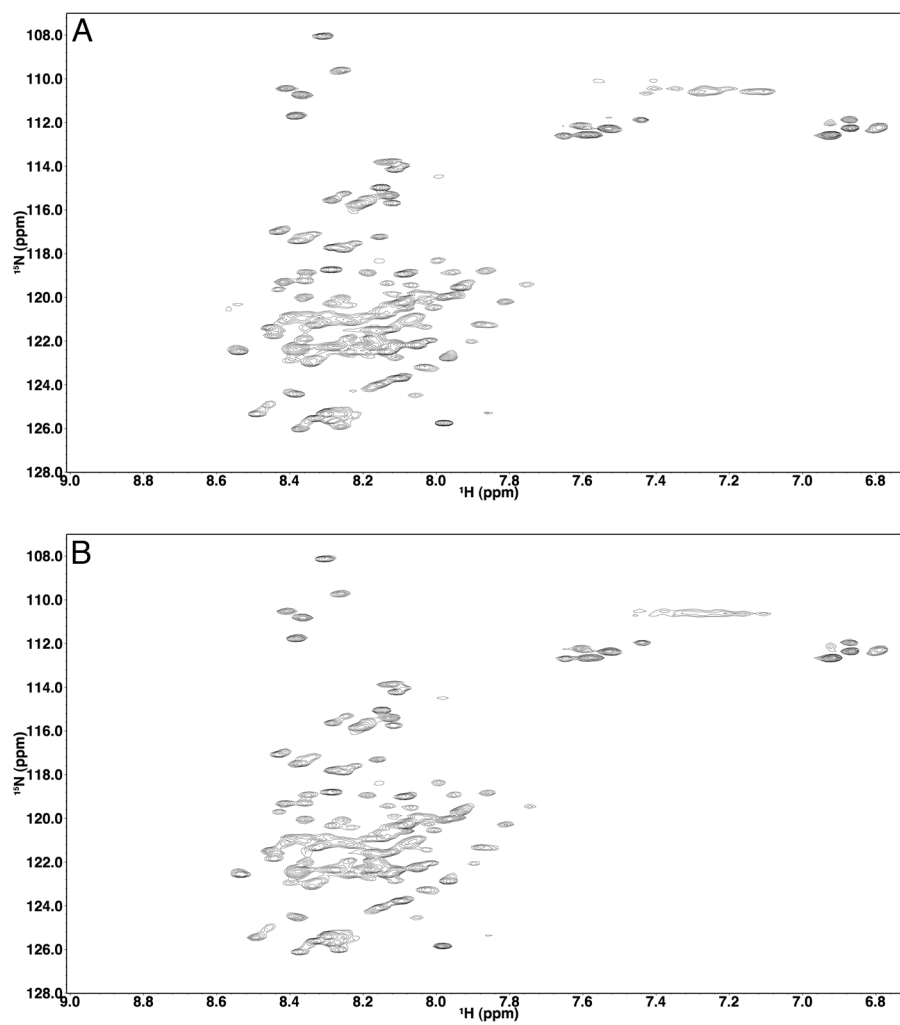The sample contained 0.3 mM ProTα in 50 mM NaPO$_4$ pH 7, 100 mM NaCl and 1 mM DTT.

Figure S6.2 ^{1}H-^{15}N HSQC spectra of TC-1 in 400 g/L Ficoll 70 and Dextran 70.

The samples contained 0.2 mM TC-1 in 10 mM sodium acetate pH 5 and 400 g/L Ficoll 70 (A) or Dextran 70 (B).
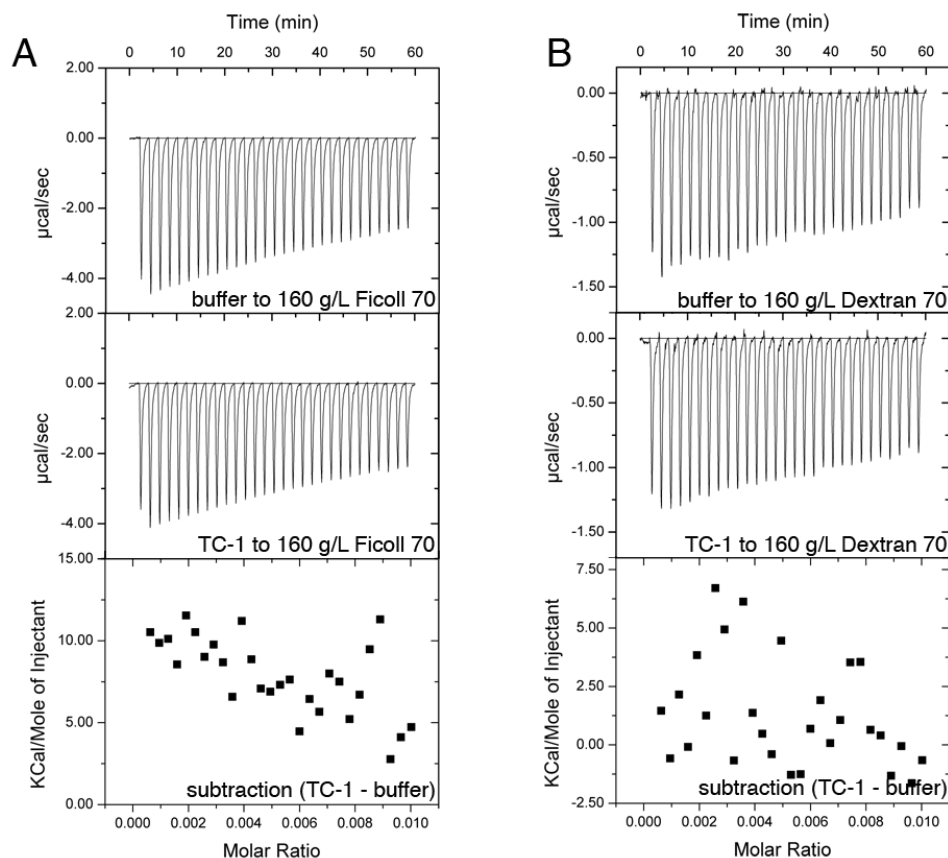
Figure S6.3 ITC profiles of TC-1 titrations into crowded solutions.

Buffer (10 mM sodium acetate pH 5) alone or containing 0.1 mM TC-1 was titrated into the cell, containing 160 g/L Ficoll (A) or Dextran 70 (B) in the same buffer. 10 μL injections were used with 120-second delays.

To determine if the chemical shift changes observed in the ^{1}H-^{15}N HSQC spectrum of TC-1 with 160 g/L of Ficoll 70 were the result of alteration of secondary structure, site-specific secondary structure propensities were determined based on the observed ^{13}Cα and ^{13}Cβ chemical shifts in the absence and presence of crowding agents using the SSP program (46, 73). Residues in well-formed β-strand/extended or α-helical conformations are expected to yield SSP scores close to -1 and 1, respectively. Figure 6.2 shows the SSP score profiles of TC-1. While the N-terminal half of the protein is largely unstructured, three regions (D44-R53, K58-A64 and D73-T88) with high helical

propensities (i.e. SSP scores > 0.2) were found in the C-terminal part under both conditions. The results are consistent with our previous SSP analysis of TC-1 (46). Based on the SSP scores reported here, it is apparent that the presence of crowding agents only leads to a minor increase in the helical propensity of the second helical region (K58-A64), while the other parts of the TC-1 structure are largely unaffected (Figure 6.2).
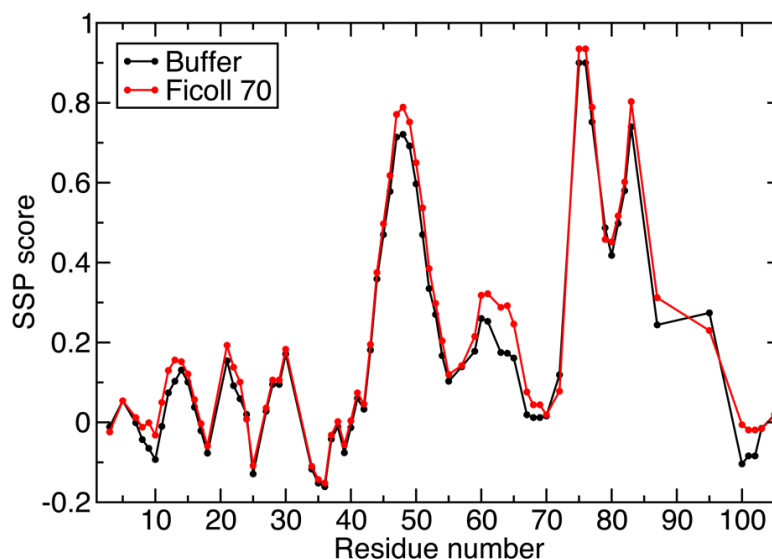

Figure 6.2 Secondary structure propensity (SSP) scores for TC-1 in the absence (black) and presence (red) of 160 g/L Ficoll 70.

SSP scores were calculated on the basis of the assigned $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts (46) using the SSP program (73). The CBCA(CO)NH spectra was collected in 10 mM sodium acetate pH 5 in the absence and presence of 160 g/L Ficoll 70.

Backbone ^{15}N spin relaxation measurements under crowded conditions

The effects of macromolecular crowding on the dynamics of ProTα, TC-1, α-synuclein, and Ubiquitin were investigated with backbone ^{15}N spin relaxation and ^{1}H-^{15}N NOE measurements. The results are shown in Figure 6.3. For the well-folded Ubiquitin, significant increases (decreases) in R_2 (R_1) of residues are observed in the presence of 160

g/L of Ficoll 70. Because crowding does not alter the structure of Ubiquitin, judging from the ^{1}H-^{15}N HSQC spectra (Figure 6.1D), the changes in R_2 and R_1 are expected to be due to the increase in viscosity of the solution. Based on the R_1 and R_2 values, the overall rotational correlation time of Ubiquitin is estimated to increase from 4.3 to 8.0 ns upon addition of crowding agents (74). Even though the molecular tumbling time was increased, crowding does not seem to have significant effects on the fast internal motion of this globular protein since the values of NOE were mostly unaffected by the addition of crowders.

Unlike Ubiquitin, however, the increase in viscosity upon addition of 160 g/L of Ficoll 70 does not lead to dramatic changes in the observed R_1, R_2 and NOE values of ProTα and α-synuclein (Figure 6.3). In particular, the value of R_2, which is sensitive to the rotational correlation time, remains unchanged for most of the residues of ProTα upon addition of crowding agents. On the other hand, residues in different regions of TC-1 show differential responses to crowding. In particular, residues in the high helical propensity regions of TC-1 generally have decreased R_1 and increased R_2 relaxation rates in the presence of 160 g/L Ficoll 70 (Figure 6.3A and B), while R_1 and R_2 values of residues in the flexible N-terminal region show only minor changes. In addition, most of the residues in TC-1 also display slightly higher NOE values in the presence of 160 g/L of Ficoll 70 (Figure 6.3C). To ensure the observed changes in relaxation rates are not due to the particular crowding agent used, ^{15}N relaxation experiments for TC-1 were also repeated with Dextran 70 as a crowder and the results were similar to that aforementioned (Figure 6.4). Figure S6.4 contains the R_1, R_2 and NOE values for TC-1 in buffer and 160 g/L Ficoll and Dextran 70 plotted by residue number.
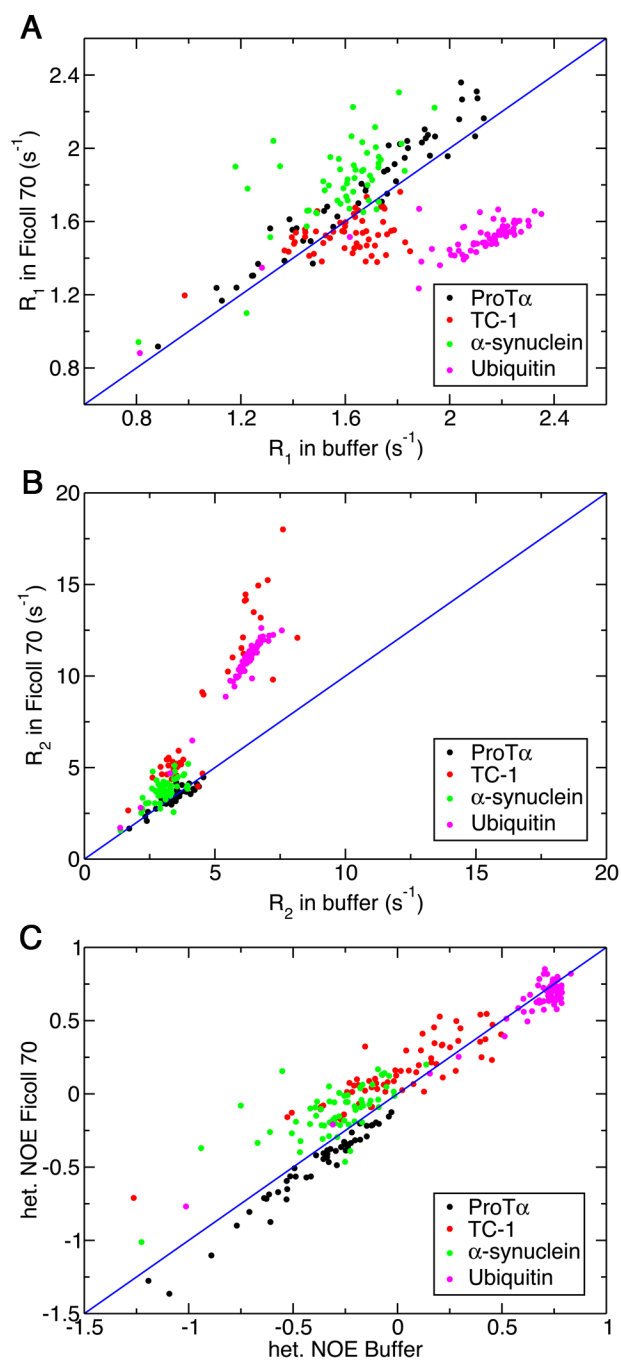
Figure 6.3 Backbone ^{15}N relaxation measurements for ProTα, TC-1, α-synuclein and Ubiquitin in the absence and presence of 160 g/L Ficoll 70.

Longitudinal relaxation rate, R_1 (A), transverse relaxation rate, R_2 (B) and steady-state ^{1}H-^{15}N NOE (C). ProTα (black), TC-1 (red), α-synuclein (green) and Ubiquitin (magenta) relaxation measurements were collected in 40 mM HEPES

pH 6.8, 10 mM sodium acetate pH 5, 50 mM sodium phosphate pH 7 and 10 mM sodium acetate pH 5 respectively in the absence and presence of 160 g/L Ficoll 70. The blue line indicates a unitary slope.
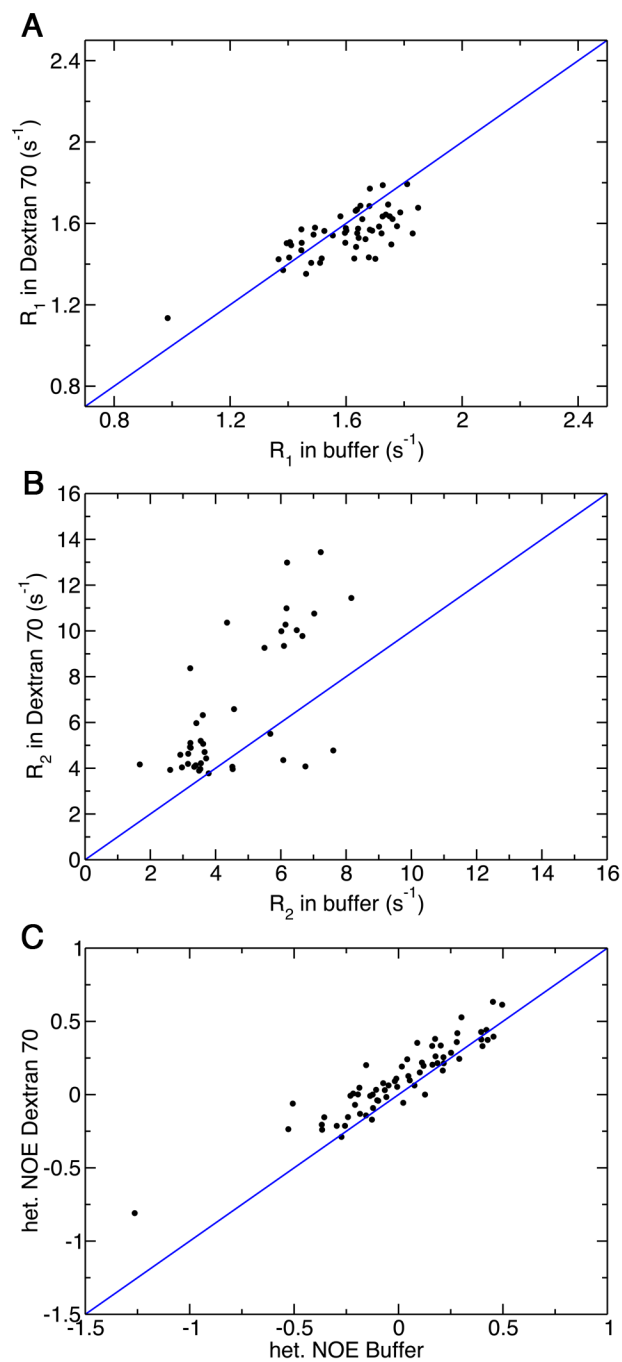
Figure 6.4 Backbone ^{15}N relaxation measurements for TC-1 in the absence and presence of 160 g/L Dextran 70.

Longitudinal relaxation rate, R_1 (A), transverse relaxation rate, R_2 (B) and steady-state ^{1}H-^{15}N NOE (C). The sample contained 10 mM sodium acetate pH 5 in the absence and presence of 160 g/L Dextran 70.
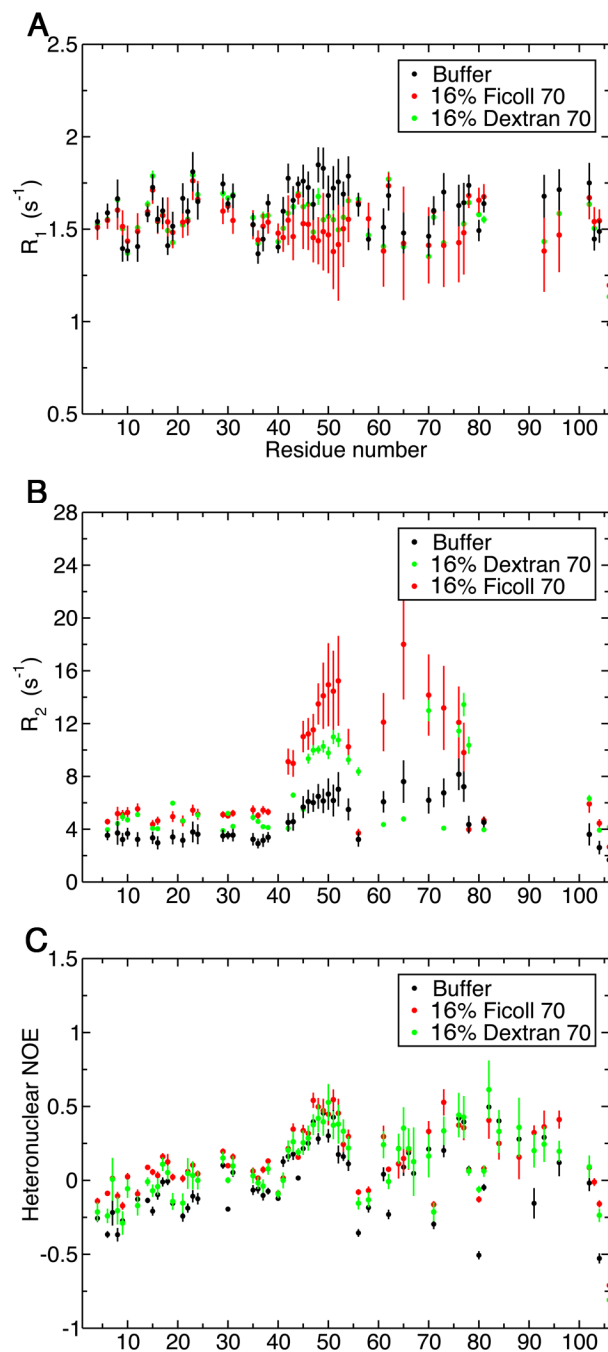
Figure S6.4 R_1, R_2 and NOE values for TC-1 in buffer and 160 g/L Ficoll 70 and Dextran 70 plotted by residue number.

The samples contained 10 mM sodium acetate pH 5 in absence and presence of 160 g/L Ficoll 70 or Dextran 70.

Considerable changes in the relaxation rates were observed for ProTα when the extremely high concentration of crowding agent (400 g/L Ficoll 70) was used (Figure 6.5). In particular, most residues show higher R_2 values in the presence of 400 g/L Ficoll 70 compared to buffer conditions (Figure 6.5B). The largest changes are observed in the region around residues I12-R31. Interestingly, residues in that region also have less negative ^{1}H-^{15}N steady-state NOE values in buffer conditions, suggesting this segment is intrinsically more restricted in motion compared to the rest of the protein in the absence of crowders. Furthermore, NOE values were systematically higher for all residues under this crowded condition (Figure 6.5C).
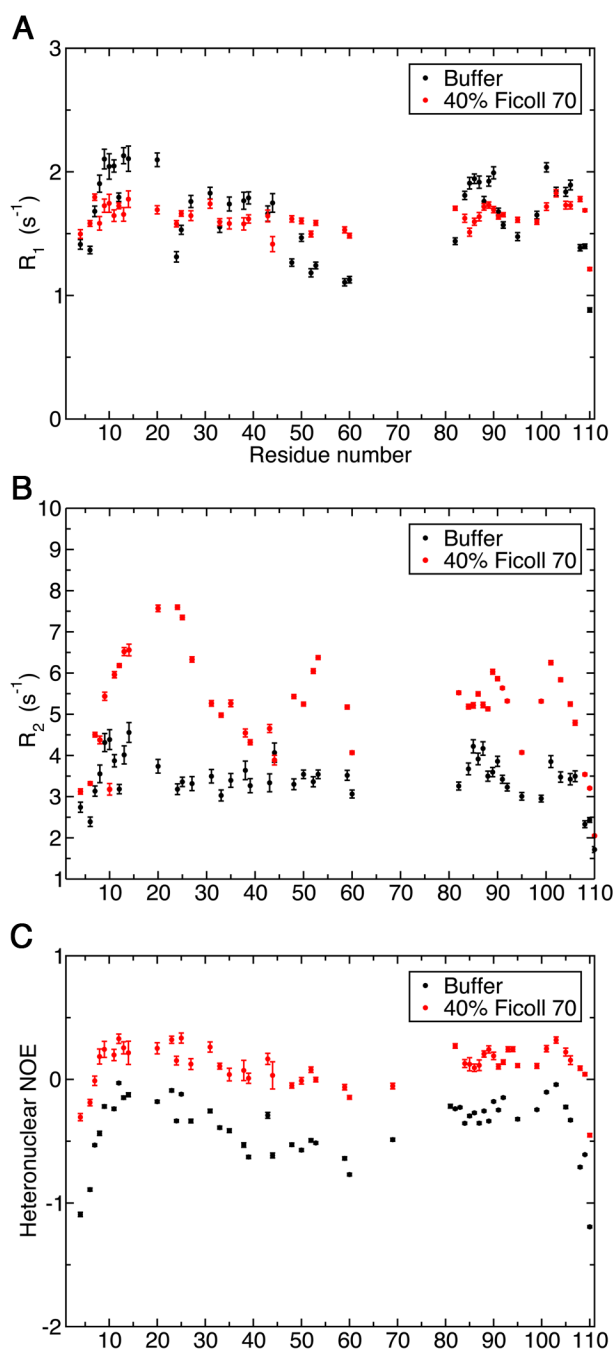
Figure 6.5 Backbone ^{15}N relaxation measurements for ProTα in the absence and presence of 400 g/L Ficoll 70.

Longitudinal relaxation rate, R_1 (A), transverse relaxation rate, R_2 (B) and steady-state ^{1}H-^{15}N NOE (C). The sample contained 0.3 mM ProTα in 50 mM NaPO$_4$ pH

7, 100 mM NaCl and 1 mM DTT in the presence of 400 g/L Ficoll 70. For the
sample without crowder, 40 mM HEPES pH 6.8 was used as the buffer.

Model for interpreting the observed relaxation data

For well-folded globular proteins, the ^{15}N R_1, R_2, and NOE measurements are
commonly fitted to the Lipari-Szabo (LS) model-free model in order to extract the
amplitude and correlation time of internal motion as well as the overall molecular
tumbling time, which are denoted by the order parameter (S^2), τ_e and τ_m in the spectral
density function, respectively (75). A modified LS model was later proposed by Clore
and co-worker to fit the relaxation rates observed from flexible loop regions of a folded
protein (76). In this model, an extra term was introduced to the spectral density function
of the original LS model to describe the internal motion occurring on a slower timescale.
For disordered proteins, however, the timescale of large-amplitude local segmental
motions can be close to the overall tumbling time, making the separation of these two
contributions to the relaxation rates challenging (71, 77).

To establish a simple model to describe the dynamic behaviors of IDPs and
correlate them to the observed relaxation parameters, autocorrelation functions of the
backbone amide bond vectors were extracted from a 427-ns atomistic MD trajectory of
ProTα. Autocorrelation functions of each residue (except the N-terminus and P34) were
fitted to models with different numbers of exponential decay terms. Instead of using these
models to back calculate the observed backbone ^{15}N relaxation rates, which have been
shown by many others to be a challenging task (78, 79), our aim is to establish a simple
model to interpret the relaxation data we obtained.

Autocorrelation functions of individual amide bond vectors extracted from the
MD simulation were fitted to the sum of two, three, or four exponential decay terms
(Equation 2) in order to determine the best LS-like model that can be used to describe the
backbone dynamics of highly disordered proteins such as ProTα. The autocorrelation
functions of several residues are shown in Figure 6.6. In general, quick decreases in the

autocorrelation functions are observed in the beginning, which are likely contributed from the librational motions (fast internal motions) (71, 75). The fast decay is then followed by more gradual decreases in the autocorrelation functions, reflecting the existence of local motions on slower timescales (Figure 6.6). However, it is clear that residues in different positions of the protein display distinct autocorrelation profiles. Figure 6.6 (inset) shows typical fits of the autocorrelation functions to 2-, 3-, and 4-exponential decay terms. We found that for most of the residues, the equation with three exponential decay terms fits the autocorrelation function statistically better than that with only two terms. Increasing the number of exponential decay terms further (*i.e.* n = 4) does not result in dramatic decreases in the root mean square deviation of fitting (Figure S6.5). Additionally, for many residues, different τ_i values obtained from the four-exponential fit are very close, indicating that the motion described by these terms cannot be discriminated. Because of these reasons, our analyses were focused on the three-exponential decay model (LS3 model; n=3 in Equation 3), which is very similar to the modified LS-model described by Clore and coworkers (76).
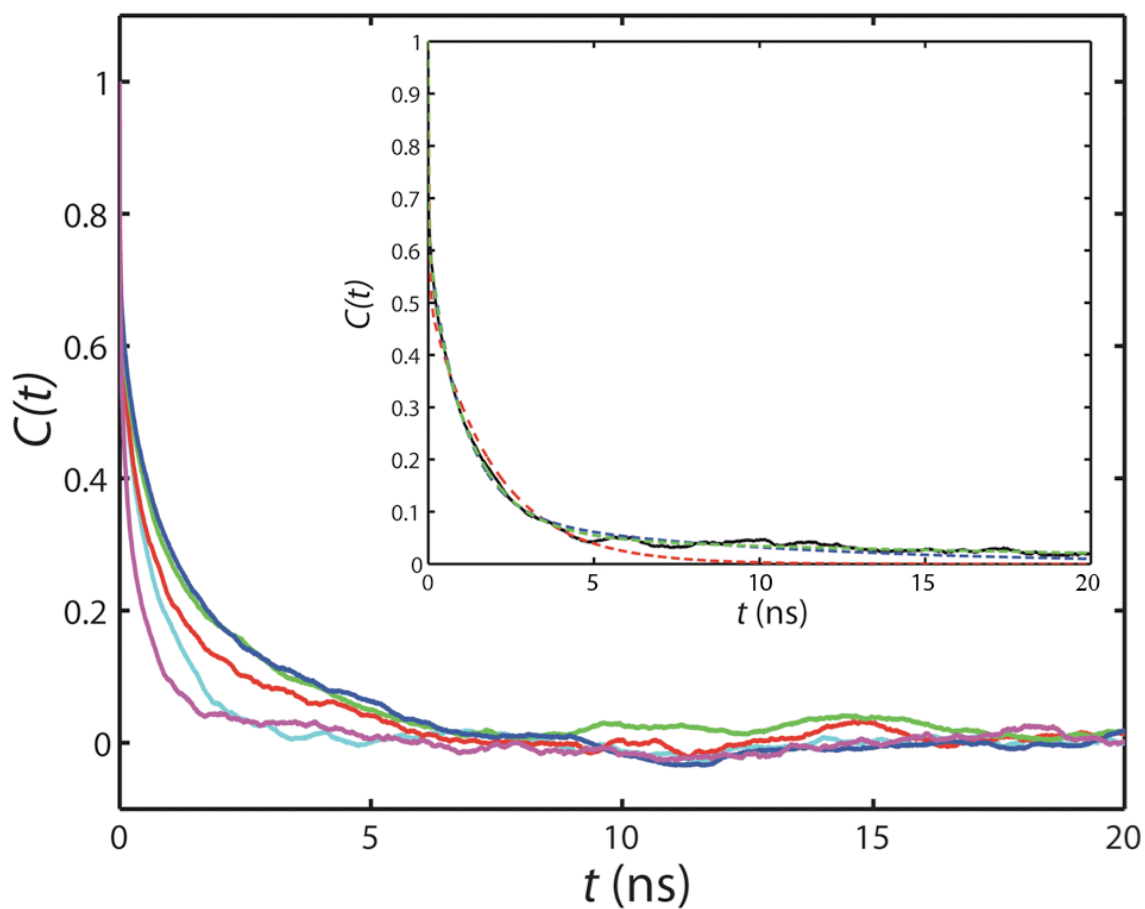
Figure 6.6 Correlation functions of selected backbone ^{1}H-^{15}N amide bond vectors.

(red: residue 2; green: residue 10; blue: residue 48; magenta: residue 57; cyan: residue 102) extracted from a 400 ns MD trajectory of ProTα. The inset shows the fitting of the autocorrelation function (solid black line) of residue 31 to 2- (red dash line), 3- (blue dash line), and 4-exponential decay curves (green dash line) as indicated in Equation 2. The blue and green dash lines overlay remarkably, and only start to deviate when t > 1.5 ns.
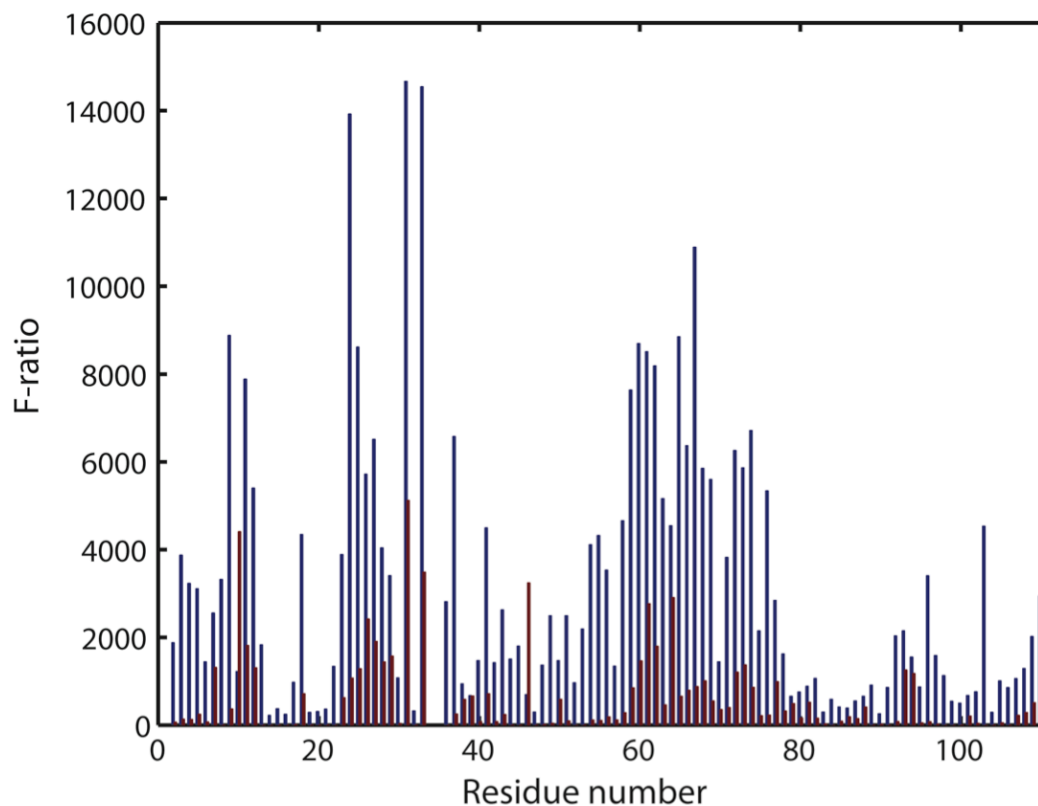
Figure S6.5 Comparison of the fitting of autocorrelations to 2-, 3-, and 4-exponential decay curves.

Blue: *F*-ratios calculated from the χ^2 and degrees of freedom of 2- and 3-exponential models; Red: *F*-ratios calculated from the χ^2 and degrees of freedom of 3- and 4-exponential models (Equation 4).

The results of fitting the amide bond vector autocorrelation functions to three-exponential decay terms are summarized in Table 6.1. To illustrate how the fluctuations in amplitude and timescale of motions translate to the observed relaxation rate changes, ^{15}N R_1, R_2, and ^{1}H-^{15}N steady-state NOE values were calculated using the LS3 model with different values of a_i and τ_i. We first apply this model to Ubiquitin. To simulate the relaxation rates of Ubiquitin, we assumed that the fast internal motion of this rigid protein is not altered upon crowding. By fixing the amplitude and correlation time of fast internal

motion (a_1 and τ_1) to 0.15 and 10 ps, respectively, the significant increase (decrease) in the measured R_2 (R_1) relaxation rates of Ubiquitin in the presence of 160 g/L of Ficoll 70 can be reproduced by changing τ_3 (the overall tumbling time) from 4.3 to 8 ns, assuming that the slower segmental motion can be neglected (i.e. $a_2 \sim 0$; blue arrows) (Figure 6.7).

Table 6.1 Averaged values and the standard deviations of fitted parameters of LS-3 model.

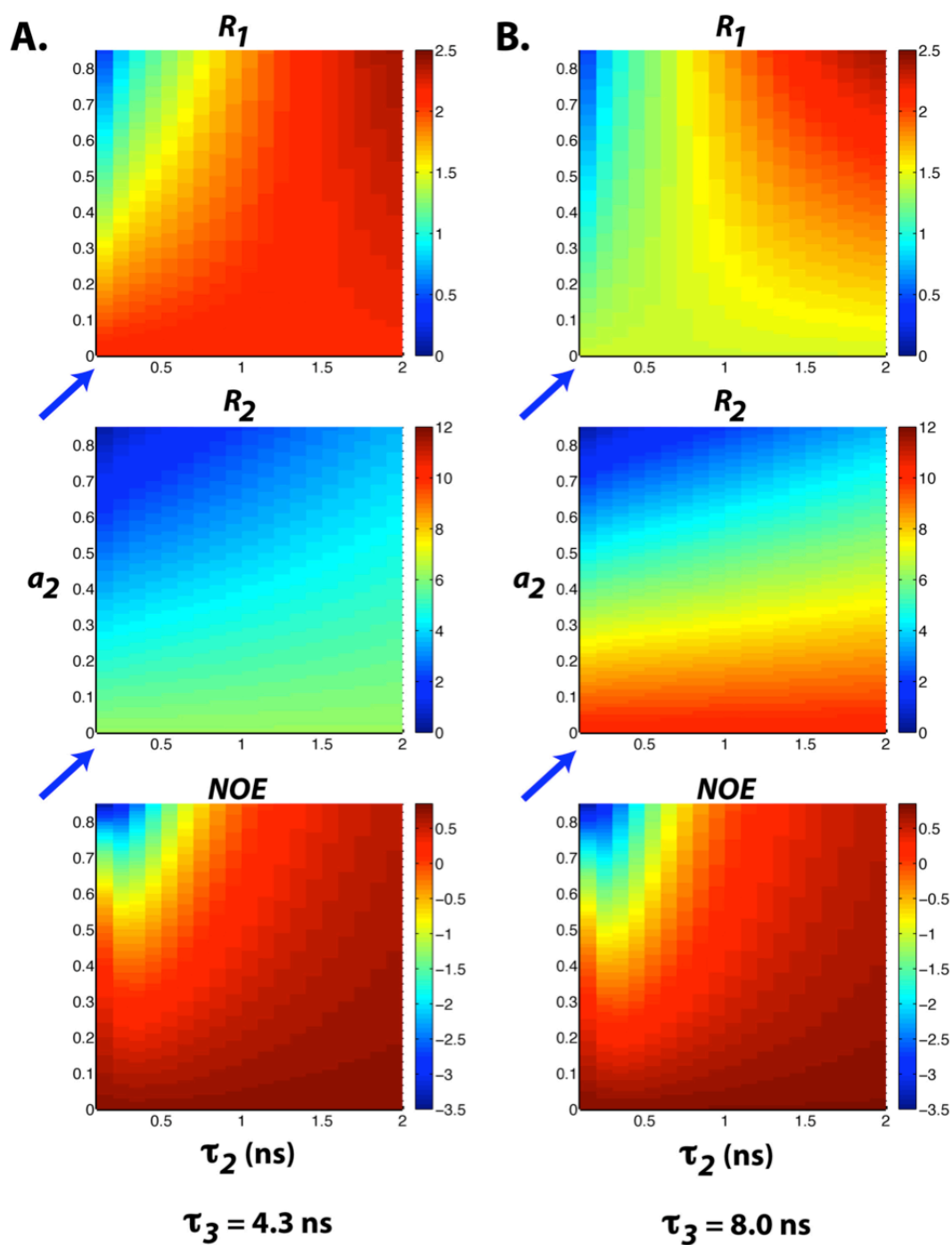|	$i=1$	$i=2$	$i=3$
τ_i (ps)	7±9	419±454	3400±5700
a_i	0.37±0.09	0.36±0.12	0.27±0.17

average±standard

Figure 6.7 ^{15}N Relaxation parameters calculated using the LS-3 model.

(a) a_1=0.15, τ_1=10 ps, τ_3 = 4.3 ns, a_3=1- a_1- a_2 (b) a_1=0.15, τ_1=10 ps, τ_3 = 8.0 ns, a_3=1- a_1- a_2. τ_2 and a_2 values are indicated along the x and y axes, respectively. The slower internal motion is negligible when $a_2 \sim 0$ (blue arrows).

We have also simulated the dependence of the ^{15}N R_1, R_2, and steady-state NOE values of ProTα on the values of a_i and τ_i. Since ProTα remains disordered under crowded conditions and the observed NOEs are significantly smaller than what are expected for a folded protein of similar molecular weight (Figure 6.5), it is reasonable to assume that large amplitude of fast internal motion persists. Figure 6.8A illustrates that with a_1=0.37, τ_1=7 ps, τ_2 ~500 ps, and τ_3 = 3.4 ns, a wide distribution of NOE values can be expected with the variation of the amplitude of segmental motion (value of a_2). Meanwhile, R_2 is predicted to be not very sensitive to the fluctuation in a_2 (R_2 ~ 2 - 4 s^{-1}). These observations agree qualitatively with the distributions of experimental relaxation rates measured under buffer conditions (Figure 6.5).

On the other hand, almost all residues of ProTα have the R_2 and NOE increased at the high concentration of crowding agents (~400 g/L of Ficoll 70), while the variation of R_1 along the protein sequence diminished. Based on the LS3 model, these trends can be explained by the increase in the correlation times of the slow local segmental motions. With τ_2 increases from 500 to 1000 ps and the value of τ_3 doubled (Figure 6.8B), R_2 values can increase to ~6 s^{-1} and many NOEs will turn positive. The simulated relaxation rates further match the experimentally observed values, especially for the R_1 values, if we assume that the amplitude of fast internal motion is reduced in a highly crowded environment (i.e. a_1=0.2; Figure 6.8C).
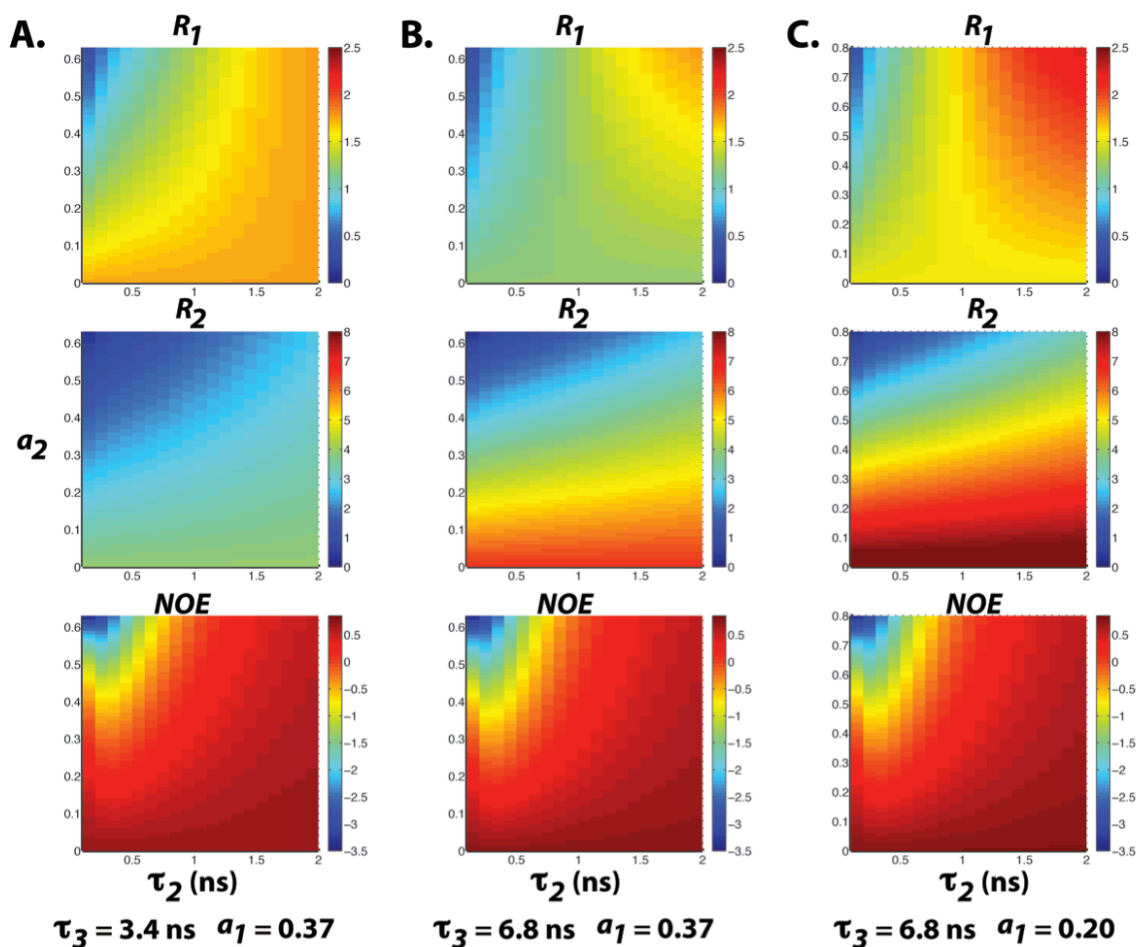
Figure 6.8 ^{15}N Relaxation parameters calculated using the LS-3 model.

(a) a_1=0.37, τ_1=7 ps, τ_3 = 3.4 ns, a_3=1- a_1- a_2 (b) a_1=0.37, τ_1=7 ps, τ_3 = 6.8 ns, a_3=1- a_1- a_2 and (c) a_1=0.20, τ_1=7 ps, τ_3 = 6.8 ns, a_3=1- a_1- a_2, respectively. τ_2 and a_2 values are indicated along the x and y axes, respectively.

Finally, based on the amplitudes and correlation times of motions on different timescales (fitted a_i and τ_i values of autocorrelation functions) extracted from the MD simulation, we have simulated the ^{15}N R_1, R_2, and steady-state NOE values of ProTα. The relaxation parameters in the presence of 160 g/L of Ficoll 70 were then predicted by scaling the correlation time of the slow motions (τ_2 and τ_3) by the same factor (i.e. 1.86) as the Ubiquitin tumbling time changes to account for the increase in viscosity. Figure 6.9

shows the plots of the simulated relaxation data before and after the correlation time adjustments. The result indicates that in the presence of 160 g/L of Ficoll 70, the R_1, R_2, and NOE of ProTα were expected to systematically increase if the correlation times of the slow motions were increased by viscosity. However, these changes were observed experimentally only in the presence of 400 g/L of Ficoll 70. Again, the simulated data suggest that the timescale of local segmental motions were slowed down only at a very high concentration of crowders.
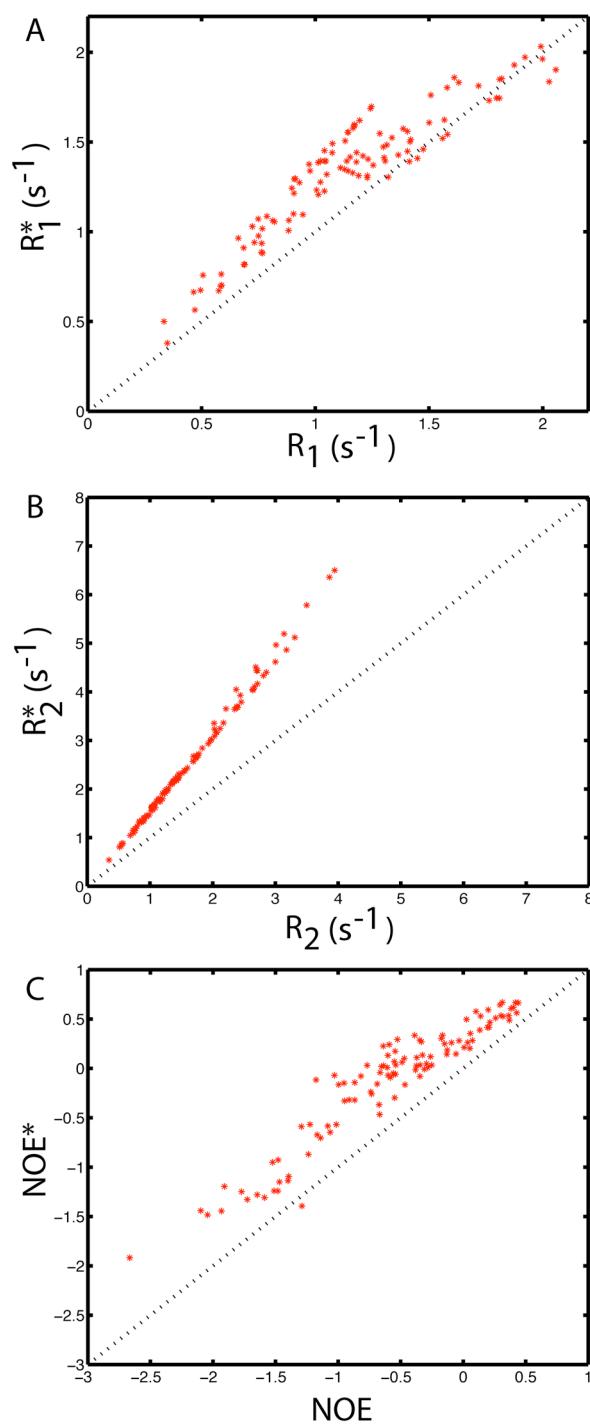
Figure 6.9 Plots of the simulated relaxation data of ProTα before and after correlation time adjustments.

^{15}N R_1, R_2, and steady-state *NOE* values of ProTα were simulated based on the amplitudes and correlation times of motions extracted from the MD simulation

using the LS3 model. R_1*, R_2*, and *NOE** are the relaxation data predicted by scaling the correlation times of the slow motions (τ_2 and τ_3) by the same factor as the Ubiquitin tumbling time changes to account for the increase in viscosity.

6.5   Discussion

We have investigated the effects of macromolecular crowding on the dynamics of three IDPs, ProTα, TC-1 and α-synuclein, with different extents of residual structure using NMR spectroscopy. This complements several recent studies of macromolecular crowding effects on the structure and dynamics of IDPs (34, 35, 80). We used Ficoll 70 and Dextran 70 as crowding agents, which are commonly used to mimic excluded volume effects (7, 17, 18, 72). These polymers are inert and do not interact nonspecifically with proteins. In contrast, the use of polyethylene glycol as a crowding agent is discouraged, due to attractive interactions with proteins (7, 72).

The IDPs examined here all had narrow dispersion of peaks along the ^{1}H dimension in the ^{1}H-^{15}N HSQC spectra compared to the well-folded Ubiquitin, both in the absence and presence of crowding agents, suggesting that they remain disordered in the crowded environments. Interestingly, for the partially disordered TC-1, a minor increase of the helical propensity was observed only in the relatively structured region in the presence of Ficoll 70. This indicates that the crowded environment may have differential effects on the partially structured regions and the highly disordered parts of the protein. Increased helical content in the presence of crowding agent has also been observed for the Flavodoxin (10). Stagg et al. reported that the far-UV CD signal of Flavodoxin at the helical signature wavelength (222 nm) increases by about 10% in the presence of 200 g/L of Ficoll 70; however, a less dramatic effect of crowding in the denatured state was observed.

Site-specific changes in the protein flexibility of ProTα and TC-1 have been characterized by using ^{15}N NMR spin relaxation experiments. In particular, we focused on the highly disordered ProTα since this protein produces NMR data with reasonable signal to noise ratio even at high concentration of Ficoll 70 (400 g/L). It is noteworthy

that besides the excluded volume effect, the presence of high concentrations of crowding agents also inevitably increases the viscosity of the solution (12, 34). This adds a layer of complexity to the interpretation of spin relaxation data. The viscosity effect is reflected in the systematic increase in the ^{15}N R_2 rates of Ubiquitin in the presence of 160 g/L Ficoll 70, while the values of NOE were mostly unaffected. Similar results were obtained by Simorellis & Flynn (81). They showed that encapsulation of Ubiquitin in a confined environment only has very minor effects on the protein backbone dynamics.

Intriguingly, the increase in viscosity did not cause significant changes in the ^{15}N R_2 of intrinsically disordered ProTα under the same conditions. To have a better understanding of our relaxation data, we performed an MD simulation (~ 400 ns) on ProTα to investigate its dynamic behaviors. Although MD simulations in the presence of atomistically represented crowders are not currently practical (because of the large number of atoms these molecules contain and the long time scales such molecules need for diffusion), our simulation facilitated the development of a simple model to correlate the observed changes in relaxation rates to the alteration in protein motions under crowding conditions. While the LS3 model proposed here might not be sufficient to represent the complicated dynamics of IDPs, it provides insights into interpreting the relaxation measurements.

Based on the experimental and simulation results, we conclude that even though crowded environments can slow down the timescale of local segmental motions in the highly disordered ProTα, it still retains a certain level of flexibility at high concentrations of Ficoll 70. Based on the observed R_2 rates (Figure 6.5B), however, it is apparent that a few regions of ProTα become more structured at high concentration of crowders. Interestingly, some of these regions overlap or are close to known target-binding motifs of ProTα. For instance, residues 39-54 are involved in mediating the interaction with the Kelch domain of Keap1 in the oxidative stress response (82) while the caspase-3 cleavage site of ProTα is located around residue 100 (83). Because the dynamics of IDPs can have significant impacts on their target recognitions (60), this observation has a strong biological implication of how this class of proteins functions in crowded cellular environments.

We are aware that while Ficoll and Dextran may be suitable agents to mimic the crowded cellular environment, combining different crowding agents with varying physical characteristics (sizes, shapes, charges, etc) may more accurately represent the *in vivo* environment (2, 7, 84, 85). Therefore, extending the current studies by using other crowding agents with different sizes and chemical properties are required to further our understanding of the macromolecular crowding effects on IDPs. These in vitro studies together with the recently developed in cell NMR techniques (86-92) will hopefully provide further insights into understanding the environmental effects on IDP structure and functions.

6.6   Acknowledgements

6.7   References

1. Zimmerman SB, Trach SO (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *J Mol Biol* 222:599-620.

2. Zimmerman SB, Minton AP (1993) Macromolecular crowding: biochemical, biophysical, and physiological consequences. *Annu Rev Biophys Biomol Struct* 22:27-65.

3. Minton AP (2000) Implications of macromolecular crowding for protein assembly. *Curr Opin Struct Biol* 10:34-9.

4. Minton AP (1997) Influence of excluded volume upon macromolecular structure and associations in 'crowded' media. *Curr Opin Biotechnol* 8:65-9.

5. Elcock AH (2010) Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. *Curr Opin Struct Biol* 20:196-206.

6. Ellis RJ (2001) Macromolecular crowding: obvious but underappreciated. *Trends Biochem Sci* 26:597-604.

7. Zhou HX, Rivas G, Minton AP (2008) Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu Rev Biophys* 37:375-97.

8. Sasahara K, McPhie P, Minton AP (2003) Effect of dextran on protein stability and conformation attributed to macromolecular crowding. *J Mol Biol* 326:1227-37.

9. Dhar A et al. (2010) Structure, function, and folding of phosphoglycerate kinase are strongly perturbed by macromolecular crowding. *Proc Natl Acad Sci U S A* 107:17586-91.

10. Stagg L, Zhang SQ, Cheung MS, Wittung-Stafshede P (2007) Molecular crowding enhances native structure and stability of alpha/beta protein flavodoxin. *Proc Natl Acad Sci U S A* 104:18976-81.

11. McGuffee SR, Elcock AH (2010) Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput Biol* 6:e1000694.

12. Wang Y, Li C, Pielak GJ (2010) Effects of proteins on protein diffusion. *J Am Chem Soc* 132:9392-7.

13. Leduc C et al. (2012) Molecular crowding creates traffic jams of kinesin motors on microtubules. *Proc Natl Acad Sci U S A* 109:6100-5.

14. Aoki K, Yamada M, Kunida K, Yasuda S, Matsuda M (2011) Processive phosphorylation of ERK MAP kinase in mammalian cells. *Proc Natl Acad Sci U S A* 108:12675-80.

15. Batra J, Xu K, Qin S, Zhou HX (2009) Effect of macromolecular crowding on protein binding stability: modest stabilization and significant biological consequences. *Biophys J* 97:906-11.

16. Huang L et al. (2010) Macromolecular crowding converts the human recombinant PrPC to the soluble neurotoxic beta-oligomers. *FASEB J* 24:3536-43.

17. Zhou Z et al. (2011) Fibril formation of the rabbit/human/bovine prion proteins. *Biophys J* 101:1483-92.

18. Ma Q et al. (2012) The contrasting effect of macromolecular crowding on amyloid fibril formation. *PLoS One* 7:e36288.

19. Teilum K, Olsen JG, Kragelund BB (2009) Functional aspects of protein flexibility. *Cell Mol Life Sci* 66:2231-47.

20. Shoemaker BA, Portman JJ, Wolynes PG (2000) Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc Natl Acad Sci U S A* 97:8868-73.

21. Wright PP, Dyson HH (1999) *J. Mol. Biol.* 293:321.

22. Dyson HJ, Wright PE (2002) Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* 12:54-60.

23. Eliezer D (2009) Biophysical characterization of intrinsically disordered proteins. *Curr Opin Struct Biol* 19:23-30.

24. Mittag T, Forman-Kay JD (2007) Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol* 17:3-14.

25. Schneider R et al. (2012) Towards a robust description of intrinsic protein disorder using nuclear magnetic resonance spectroscopy. *Mol Biosyst* 8:58-68.

26. Mittag T, Kay LE, Forman-Kay JD (2010) Protein dynamics and conformational disorder in molecular recognition. *J Mol Recognit* 23:105-16.

27. Smock RG, Gierasch LM (2009) Sending signals dynamically. *Science* 324:198-203.

28. Baker JM et al. (2007) CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. *Nat Struct Mol Biol* 14:738-45.

29. Mittag T et al. (2008) Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proc Natl Acad Sci U S A* 105:17772-7.

30. Tang X et al. (2012) Composite low affinity interactions dictate recognition of the cyclin-dependent kinase inhibitor Sic1 by the SCFCdc4 ubiquitin ligase. *Proc Natl Acad Sci U S A* 109:3287-92.

31. Dedmon MM, Patel CN, Young GB, Pielak GJ (2002) FlgM gains structure in living cells. *Proc Natl Acad Sci U S A* 99:12681-4.

32. Flaugh SL, Lumb KJ (2001) Effects of macromolecular crowding on the intrinsically disordered proteins c-Fos and p27(Kip1). *Biomacromolecules* 2:538-40.

33. Johansen D, Jeffries CM, Hammouda B, Trewhella J, Goldenberg DP (2011) Effects of macromolecular crowding on an intrinsically disordered protein characterized by small-angle neutron scattering with contrast matching. *Biophys J* 100:1120-8.

34. Szasz CS et al. (2011) Protein disorder prevails under crowded conditions. *Biochemistry* 50:5834-44.

35. Uversky VN (2009) Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *Protein J* 28:305-25.

36. Haritos AA, Goodall GJ, Horecker BL (1984) Prothymosin alpha: isolation and properties of the major immunoreactive form of thymosin alpha 1 in rat thymus. *Proc Natl Acad Sci U S A* 81:1008-11.

37. Gast K et al. (1995) Prothymosin alpha: a biologically active protein with random coil conformation. *Biochemistry* 34:13211-8.

38. Piñeiro A, Cordero OJ, Nogueira M (2000) Fifteen years of prothymosin alpha: contradictory past and new horizons. *Peptides* 21:1433-46.

39. Yi S, Brickenden A, Choy WY (2008) A new protocol for high-yield purification of recombinant human prothymosin alpha expressed in Escherichia coli for NMR studies. *Protein Expr Purif* 57:1-8.

40. Yi S, Boys BL, Brickenden A, Konermann L, Choy WY (2007) Effects of zinc binding on the structure and dynamics of the intrinsically disordered protein prothymosin alpha: evidence for metalation as an entropic switch. *Biochemistry* 46:13120-30.

41. Chua EL, Young L, Wu WM, Turtle JR, Dong Q (2000) Cloning of TC-1 (C8orf4), a novel gene found to be overexpressed in thyroid cancer. *Genomics* 69:342-7.

42. Sunde M et al. (2004) TC-1 is a novel tumorigenic and natively disordered protein associated with thyroid cancer. *Cancer Res* 64:2766-73.

43. Polakis P (2000) Wnt signaling and cancer. *Genes Dev* 14:1837-51.

44. Jung Y et al. (2006) TC1 (C8orf4) enhances the Wnt/beta-catenin pathway by relieving antagonistic activity of Chibby. *Cancer Res* 66:723-8.

45. Takemaru K-I et al. (2003) Chibby, a nuclear beta-catenin-associated antagonist of the Wnt/Wingless pathway. *Nature* 422:905-9.

46. Gall C, Xu H, Brickenden A, Ai X, Choy WY (2007) The intrinsically disordered TC-1 interacts with Chibby via regions with high helical propensity. *Protein Sci* 16:2510-8.

47. Yokota O et al. (2002) NACP/alpha-synuclein, NAC, and beta-amyloid pathology of familial Alzheimer's disease with the E184D presenilin-1 mutation: a clinicopathological study of two autopsy cases. *Acta Neuropathol* 104:637-48.

48. McNulty BC, Young GB, Pielak GJ (2006) Macromolecular crowding in the Escherichia coli periplasm maintains alpha-synuclein disorder. *J Mol Biol* 355:893-7.

49. Shevchik VE, Condemine G, Robert-Baudouy J (1994) Characterization of DsbC, a periplasmic protein of Erwinia chrysanthemi and Escherichia coli with disulfide isomerase activity. *EMBO J* 13:2007-12.

50. Delaglio F et al. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277-93.

51. Johnson BA (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* 278:313-52.

52. Lee AL, Wand AJ (1999) Assessing potential bias in the determination of rotational correlation times of proteins by NMR relaxation. *J Biomol NMR* 13:101-12.

53. Wu KP, Kim S, Fela DA, Baum J (2008) Characterization of conformational and dynamic properties of natively unfolded human and mouse alpha-synuclein ensembles by NMR: implication for aggregation. *J Mol Biol* 378:1104-15.

54. d'Auvergne EJ, Gooley PR (2008) Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. *J Biomol NMR* 40:107-19.

55. d'Auvergne EJ, Gooley PR (2008) Optimisation of NMR dynamic models II. A new methodology for the dual optimisation of the model-free parameters and the Brownian rotational diffusion tensor. *J Biomol NMR* 40:121-33.

56. Brunger AT (2007) Version 1.2 of the Crystallography and NMR system. *Nat Protoc* 2:2728-33.

57. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation *J Chem Theory Comput* 4:435-447.

58. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6 *Journal of Computational Chemistry* 25:1656-1676.

59. Oostenbrink C, Soares TA, van der Vegt NF, van Gunsteren WF (2005) Validation of the 53A6 GROMOS force field. *Eur Biophys J* 34:273-84.

60. Cino EA, Wong-Ekkabut J, Karttunen M, Choy WY (2011) Microsecond molecular dynamics simulations of intrinsically disordered proteins involved in the oxidative stress response. *PLoS One* 6:e27371.

61. Cino EA, Choy WY, Karttunen M (2012) Comparison of Secondary Structure Formation Using 10 Different Force Fields in Microsecond Molecular Dynamics Simulations. *J Chem Theory Comput* 8:2725-2740.

62. Wong-Ekkabut J, Karttunen M (2012) Assessment of common simulation protocols for simulations of nanopores, membrane proteins & channels *J Chem Theory Comput*

63. Berendsen HJC, Postma JPM, and Gusteren WF (1981) in *Intermolecular forces*, p 331-342.

64. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126:014101.

65. Parrinello, Rahman (1981) Polymorphic transitions in single crystals: A new molecular dynamics method *J Appl Phys* 52:7182-7190.

66. Hess B, Bekker H, Berendsen HJC, Johannes JGEM (1997) LINCS: A linear constraint solver for molecular simulations *J Comput Chem* 18:1463-1472.

67. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N · log (N) method for Ewald sums in large systems *J Chem Phys* 98:10089-10092.

68. Karttunen M, Rottler J, Vattulainen I, Sagui C (2008) Electrostatics in Biomolecular Simulations: Where Are We Now and Where Are We Heading? *Current Topics in Membranes* 60:49-89.

69. Bremi T, Brüschweiler R, Ernst RR (1997) A protocol for the interpretation of side-chain dynamics based on NMR relaxation: Application to phenylalanines in antamanide *J Am Chem Soc* 119:4272-4284.

70. Istratov AA, Vyvenko OF (1999) Exponential analysis in physical phenomena *Review of Scientific Instruments* 70:1233.

71. Xue Y, Skrynnikov NR (2011) Motion of a disordered polypeptide chain as studied by paramagnetic relaxation enhancements, 15N relaxation, and molecular dynamics simulations: how fast is segmental diffusion in denatured ubiquitin? *J Am Chem Soc* 133:14614-28.

72. Zhang DL, Wu LJ, Chen J, Liang Y (2012) Effects of macromolecular crowding on the structural stability of human α-lactalbumin. *Acta Biochim Biophys Sin (Shanghai)* 44:703-11.

73. Marsh JA, Singh VK, Jia Z, Forman-Kay JD (2006) Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci* 15:2795-804.

74. Kay LE, Torchia DA, Bax A (1989) Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* 28:8972-9.

75. Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity *J Am Chem Soc* 104:4546-4559.

76. Clore GM et al. (1990) Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins *J Am Chem Soc* 112:4989-4991.

77. Choy WY, Kay LE (2003) Model selection for the interpretation of protein side chain methyl dynamics. *J Biomol NMR* 25:325-33.

78. Nederveen AJ, Bonvin AMJJ (2005) NMR Relaxation and Internal Dynamics of Ubiquitin from a 0.2 μs MD Simulation *J Chem Theory Comput* 1:363-374.

79. Showalter SA, Brüschweiler R (2007) Validation of Molecular Dynamics Simulations of Biomolecules Using NMR Spin Relaxation as Benchmarks: Application to the AMBER99SB Force Field *J Chem Theory Comput* 3:961-975.

80. Li C et al. (2008) Differential dynamical effects of macromolecular crowding on an intrinsically disordered protein and a globular protein: implications for in-cell NMR spectroscopy. *J Am Chem Soc* 130:6310-1.

81. Simorellis AK, Flynn PF (2006) Fast local backbone dynamics of encapsulated ubiquitin. *J Am Chem Soc* 128:9580-1.

82. Padmanabhan P, Nakamura N, Yokoyama Y (2008) *Acta. Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* 64:233.

83. Evstafieva AG et al. (2003) Apoptosis-related fragmentation, translocation, and properties of human prothymosin alpha. *Exp Cell Res* 284:211-23.

84. Christiansen A, Wang Q, Samiotakis A, Cheung MS, Wittung-Stafshede P (2010) Factors defining effects of macromolecular crowding on protein stability: an in vitro/in silico case study using cytochrome c. *Biochemistry* 49:6519-30.

85. Minton AP (2001) The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J Biol Chem* 276:10577-80.

86. Selenko P, Serber Z, Gadea B, Ruderman J, Wagner G (2006) Quantitative NMR analysis of the protein G B1 domain in Xenopus laevis egg extracts and intact oocytes. *Proc Natl Acad Sci U S A* 103:11904-9.

87. Ito Y, Selenko P (2010) Cellular structural biology. *Curr Opin Struct Biol* 20:640-8.

88. Pielak GJ et al. (2009) Protein nuclear magnetic resonance under physiological conditions. *Biochemistry* 48:226-34.

89. Sakakibara D et al. (2009) Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature* 458:102-5.

90. Reardon PN, Spicer LD (2005) Multidimensional NMR spectroscopy for protein characterization and assignment inside cells. *J Am Chem Soc* 127:10848-9.

91. Li C et al. (2010) Protein (19)F NMR in Escherichia coli. *J Am Chem Soc* 132:321-7.

92. Schlesinger AP, Wang Y, Tadeo X, Millet O, Pielak GJ (2011) Macromolecular crowding fails to fold a globular protein in cells. *J Am Chem Soc* 133:8082-5.

7   Conclusions and future directions

7.1   Conclusions

Using an experimental and computational approach of NMR, MD simulations and ITC, we have furthered the understanding of the structure and dynamics of IDPs, and their roles in target binding. Importantly, our approach was applied to gain new insights into the interaction mechanisms of proteins with important biological functions. The work illustrates that experimental and computational techniques are complementary tools that can be used to gain new insights into the relationships between IDP structure, dynamics and function. Here, the primary conclusions from this thesis are presented.

7.1.1   MD simulations provide unique insights into the structure and dynamics of IDPs

Conventional methods and conditions for determining structure and dynamics of globular proteins may not be directly applicable to IDPs. NMR is a powerful technique for examining the structure and dynamics of IDPs, but relying solely on this type of data can be limiting. MD simulations can be used in conjunction with NMR and other biophysical techniques to gain a deeper understanding of the structure and dynamics of IDPs. By performing extensive MD simulations and comparing the results to experimental data, we have established suitable parameters and force field choice for accurately simulating this class of proteins. The methodology was shown to be applicable to a diverse set of peptides from 9 IDPs. These MD simulations were in good agreement with our NMR data and also provided unique insights into the mechanisms that IDPs may use to interact with targets. The work emphasizes that MD simulations of IDPs are an excellent complement to experimental techniques and provide new insights that are difficult to obtain by other methods.

7.1.2   Preformed structures are crucial for the interactions between some IDPs and targets

Using a combination of NMR, ITC and MD simulations, we studied the binding of 9 disordered proteins to a common target. The findings illustrated how MoRFs are important for the interaction between IDPs and targets. The IDPs that were examined had free state preformed structures in their MoRF regions that resembled bound state

conformations. The correlations between the binding affinities, amino acid compositions and MoRF stability revealed how different sequence properties can modulate the various protein-protein interactions. Using these findings, the MoRF region of a peptide from one of the IDPs was selectively mutated to enhance its target binding affinity. The designed peptide may be a potentially useful therapeutic agent. Together, the findings provide valuable insights into the mechanisms used by IDPs for target binding and should also help to elucidate the biological roles of the various protein-protein interactions.

7.1.3 Molecular crowding may affect the conformational propensity of distinct regions of an IDP

Using inert polymers as crowding agents to mimic the cellular environment, the effects of molecular crowding on dynamics of IDPs were assessed. We found that IDPs remain at least partially disordered despite the presence of high concentration of other macromolecules and that the crowded environment has differential effects on the conformational propensity of distinct regions of an IDP, which may lead to selective stabilization of certain target-binding motifs. This information will help to understand the behavior of IDPs in cellular environments and to accurately study them *in vitro*.

7.2 Future directions

7.2.1 The origins of molecular crowding effects

In chapter 6, we found that crowding agents may act to selectively stabilize target-binding regions of IDPs. Based upon our experimental NMR data under crowded conditions and an MD simulation of an IDP in the absence of crowders, we proposed a simple model to interpret the observed changes in IDP dynamics between dilute and crowded environments. Although the model qualitatively reproduces some of the experimental data, it is clearly insufficient to fully describe IDP dynamics in crowded environments. MD simulations with accurately parameterized crowding agents would undoubtedly aid in the interpretation of our experimental data. Efforts are underway to modify an existing force field to include crowding agents. Several groups have performed simulations with crowding agents (1-3). However, the agents are often crudely

parameterized as hard spheres with repulsive potentials to all species within the system. Our aim is to provide a more detailed description of the crowding agents that is consistent with experimental data. For example, our modified GROMOS 53a6 force field includes crowding agents with attractive and repulsive non-bonded potentials to different system components. The interaction between crowding agents and proteins is semi-repulsive to insure that physical contact between these species is minimized (eg. non-specific binding does not occur). Additionally, we include an attractive potential between crowding agent and the oxygen atom of water molecules to keep the crowders surrounded by a solvation shell. Extensive testing and tuning of these parameters has been performed. Prior to conducting crowding simulations with IDPs, we have performed MD simulations of Ubiquitin in the absence and presence of crowding agents and checked agreement with our experimental data (chapter 6). Such work should help us to provide a more thorough understanding of crowding on IDP dynamics.

7.2.2   How MoRFs modulate target binding

The identification of MoRFs along an IDP sequence is an area of active bioinformatic research (4). These segments are hot spots for target binding (5, 6) and their identification is an important first step in studying the interactions that IDPs partake in. While MoRF discovery and prediction algorithms have improved in recent years (4), their mechanisms of function are often not well understood. The work presented in chapter 4 illustrated that the regions within and surrounding MoRF sites can modulate interaction affinities with targets. Such information may be useful for improving MoRF identification and prediction of important parameters (e.g., binding affinities). Additionally, the ability to predict how MoRF modifications can affect their binding affinities could possibly be useful for assessing the potential consequences of somatic mutations in these regions. Knowledge of how to selectively modify these sites could be useful for development of targeted therapeutics. By studying the molecular mechanisms used by other IDPs to interact with targets (e.g., using a similar approach as in chapter 4), it should, eventually, be possible to improve predictive methods for assessing important properties of MoRF regions and their relationship to binding.

7.2.3   IDP binding from the perspective of targets

This work has largely focused on the structure and dynamics of IDPs in the absence and presence of a binding target. The assignment of the backbone resonance of the target protein (chapter 5) revealed that considerable changes in the local chemical environments of many residues occur upon binding to an IDP. It is of particular interest that many of these changes are distant from the binding site. It would be interesting to analyze how the dynamics of the target protein change upon IDP binding (e.g., by NMR spin-relaxation measurements). This information would help to interpret our observed thermodynamic parameters of interaction (chapter 3) and further decipher the molecular mechanisms of IDP binding.

7.3 References

1. Cho EJ, Kim JS (2012) Crowding effects on the formation and maintenance of nuclear bodies: insights from molecular-dynamics simulations of simple spherical model particles. *Biophys J* 103:424-33.

2. Li X, Mehler EL (2006) Simulation of molecular crowding effects on an Alzheimer's beta-amyloid peptide. *Cell Biochem Biophys* 46:123-41.

3. Minh DD, Chang CE, Trylska J, Tozzini V, McCammon JA (2006) The influence of macromolecular crowding on HIV-1 protease internal dynamics. *J Am Chem Soc* 128:6006-7.

4. Disfani FM et al. (2012) MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. *Bioinformatics* 28:i75-i83.

5. Mohan A et al. (2006) Analysis of molecular recognition features (MoRFs). *J Mol Biol* 362:1043-59.

6. Vacic V et al. (2007) Characterization of molecular recognition features, MoRFs, and their binding partners. *J Proteome Res* 6:2351-66.

Curriculum Vitae

Name: **Elio Anthony Cino**

Post-secondary education and degrees:

University of Western Ontario
London, Ontario, Canada
2003-2006 B.Sc.

The University of Western Ontario
London, Ontario, Canada
2009-Current PhD.

Honors and awards:

Ontario Graduate Scholarship (OGS)
2012-Current

Ontario Graduate Scholarships in Science and Technology
(OGSST)
2010-2011

Western Graduate Research Scholarships (WGRS)
2009-Current

Gordon Research Conferences (GRC) travel award
Barga, Italy, 2011

A.C. Camargo Global Meeting of Translational Sciences travel
award
São Paulo, Brazil, 2010

University of Western Ontario Graduate Thesis Research Fund
award
2010

University of Western Ontario Graduate Thesis Research Fund
award
2011

Related Work Experience:

        Teaching Assistant - Biochemistry 3381
        The University of Western Ontario
        2009-Current

Publications:

1. Cino, E.A., Killoran, R.C., Karttunen, M., and Choy, W.Y. (2012). Sequence and Structure Analysis of KEAP1 Interacting Proteins. *Submitted (PNAS).*

2. Khan, H., Cino, E.A., Brickenden A., Fan, J., Yang, D. and Choy, W.Y. (2012). Fuzzy Complex Formation between the Intrinsically Disordered Prothymosin α and the Kelch Domain of Keap1 Involved in the Oxidative Stress Response. *Submitted (JMB).*

2. Cino, E.A., Karttunen, M., and Choy, W.Y. (2012). Effects of Molecular Crowding on the Dynamics of Intrinsically Disordered Proteins. *In revision (PLoS ONE).*

3. Cino, E.A., Choy, W.Y., and Karttunen, M. (2012). Comparison of Secondary Structure Formation Using 10 Different Force Fields in Microsecond Molecular Dynamics Simulations. J Chem Theory Comput *8*, 2725-740.

4. Cino, E., Fan, J., Yang, D., and Choy, W.Y. (2012). (1)H, (15)N and (13)C backbone resonance assignments of the Kelch domain of mouse Keap1. Biomol NMR Assign. *In press*.

5. Souza, I.A., Cino, E.A., Choy, W.Y., Cordeiro, M.N., Richardson, M., Chavez-Olortegui, C., Gomez, M.V., Prado, M.A., and Prado, V.F. (2012). Expression of a recombinant Phoneutria toxin active in calcium channels. Toxicon *60*, 907-918.

6. Cino, E.A., Wong-ekkabut, J., Karttunen, M., and Choy, W.Y. (2011). Microsecond Molecular Dynamics Simulations of Intrinsically Disordered Proteins Involved in the Oxidative Stress Response. PLoS ONE, 6(11): e27371.

Presentations:

1. Poster presentation
Conference: Gordon Research Conferences, Computational Aspects of Biomolecular NMR
Location: Barga, Italy
Dates: May 22-27, 2011
Abstract title: Microsecond molecular dynamics simulations of intrinsically disordered proteins
Authors: Elio A. Cino, Jirasak Wong-ekkabut, Mikko Karttunen, and Wing-Yiu Choy

2. Poster presentation
Event: Margaret Moffat Research Day
Location: University of Western Ontario, London Ontario
Date: March 29, 2011
Abstract title: Molecular dynamics simulations indicate that Prothymosin alpha and Nrf2 bind to Keap1 via preformed structural elements and coupled folding and binding
Authors: Elio A. Cino, Jirasak Wong-ekkabut, Mikko Karttunen, and Wing-Yiu Choy

3. Poster presentation
Conference: A.C. Camargo Global Meeting of Translational Sciences
Location: São Paulo, Brazil
Dates: April 19-30, 2010
Abstract title: Structural insights for enhancing the oxidative stress response
Authors: Elio A. Cino, Jirasak Wong-ekkabut, Mikko Karttunen, and Wing-Yiu Choy