POTENTIAL HETEROGENEITY IN P53/S100B(ββ) COMPLEX

by

CHESTER DALE MCDOWELL

B.S., Kansas State University, 2009

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Graduate Biochemistry Group
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2012

Approved by:

Co-Major Professor
Dr. Jianhan Chen

Approved by:

Co-Major Professor
Dr. Paul E Smith

Copyright

Abstract

Intrinsically disordered proteins have been shown to be important in many physiological processes, including cell signaling, translation, and transcription. They are also associated with cancer, and neurodegenerative diseases. The tumor suppressor p53 contains several disordered regions, including the C-terminal negative regulatory domain (NRD). In cancer the function of p53 has been shown to be repressed by S100B(ββ) binding to p53-NRD. Binding of S100B(ββ) blocks acetylation and phosphorylation sites in the p53-NRD, which leads to tetramer dissociation and prevents p53 activation. NMR studies have shown that p53-NRD binds S100B(ββ) in a stable α-helix conformation. Interestingly, despite the well-converged and apparent rigid nature of the NMR structure ensemble, a majority of intermolecular NOEs used to calculate the NMR ensemble are very weak (≥6 Å). The final NMR structures also contains unsatisfied buried charged residues at the binding interface. It's plausible that the p53-S100B(ββ) complex is more dynamic than previously believed. The goal of the study is to determine the potential conformational heterogeneity in p53-S100B(ββ) complex using molecular modeling. For this, five diverse structures were selected from the 40-member NMR ensemble. For each initial conformation, we performed 100 ns molecular dynamic simulations in explicit solvent to explore the structure and dynamics of the p53-NRD in complex with S100B(ββ). Several analytical tools were used to characterize the p53-NRD conformation, including root-mean squared deviation (RMSD), root-mean squared fluctuation (RMSF), and residue helicity. The accuracy of the simulations was mainly assessed by comparing with experimental NOEs. The results show that, even though the ensemble is heterogeneous it satisfies 82% of the experimental NOEs. Clustering analysis further suggests that many conformational sub-states coexist for this complex, and individual clusters appear to satisfy only subsets of NOE distances. Importantly, the buried surface analysis demonstrates that the heterogeneous ensemble generated from MD provides similar shielding of key residues, which include post-translational modification residues needed for p53 activation. This study also demonstrates that atomistic simulations can provide important insights into structure and dynamics of IDPs for understanding their biological function.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank Dr. Jianhan Chen, and Dr. Paul Smith for their support and guidance throughout my graduate studies. I would also like to thank Elizabeth Ploetz for her GROMACS expertise. Along with, Dr. Debabani Ganguly, Dr. Jian Gao, and Weihong Zhang for their help with CHARMM, my project, and discussions. Computational resources were supplied by Beocat supercomputer at Kansas State University. The project could not have been completed without support from NSF and AES.

Dedication

I would like dedicate this to my parents, for their continuing support over the years.

Chapter 1 - Introduction of IDPs and Molecular Modeling

What Are IDPs?

In the late 1990's a new class of proteins began gaining recognition in the scientific community. These proteins are now known as intrinsically disordered proteins (IDPs). Under physiological conditions, IDPs lack stable tertiary structure (1). Despite their high degree of disorder, they are functionally active in the cell. IDPs are highly abundant in nature, and one third of eukaryotic proteins are now known to contain intrinsically unstructured region spaces (2). IDPs appear to be enriched in cell signaling, cancer associated proteins, Parkinson and Alzheimer disease, making them intriguing targets for therapeutic purposes (3) (Fig. 1.1).

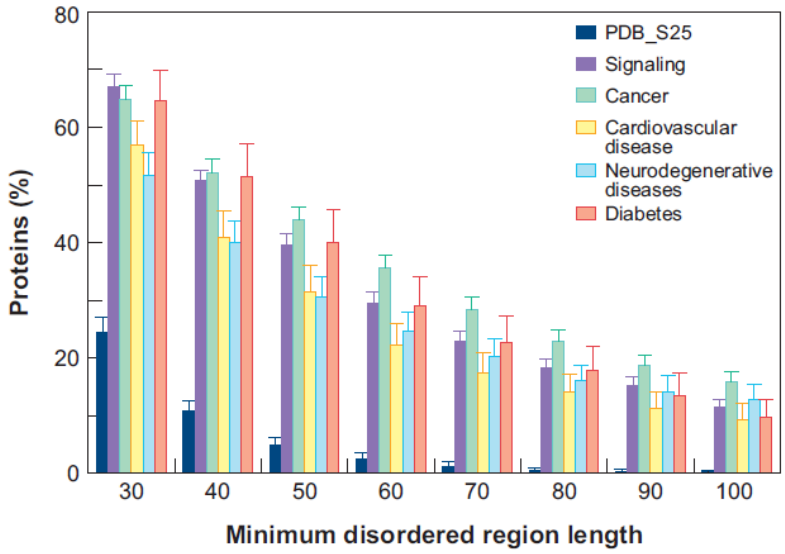

Figure 1.1 The percent of proteins containing disordered regions involved in cellular functions and diseases (4).

Man functional benefits have been proposed for IDPs. IDPs can bind their target proteins with low affinity and high specificity. These proteins also display structural plasticity, which allows them to bind many different proteins. Recently, IDPs have been shown to have heterogeneous conformations in the bound state (5,6). The structure of IDPs are in constant flux between interconvertible conformations. Some of these heterogeneous conformations may aid in the folding mechanism of IDPs. This rapid exchange of heterogeneous conformations makes it

difficult to fully describe IDPs with a single structure (7). In order to gain further mechanistic insights, IDPs need to be studied as conformational ensembles.

Sequence Characteristics of IDPs

The inherit disorder observed in IDPs arises from their amino acid sequence. IDPs are depleted in bulky hydrophobic residues, such as C, W, Y, F, V, L, and I, which are considered structure-inducing residues. These proteins contain increased amounts of hydrophilic residues, such as M, K, R, S, Q, P, and E, which promote extended structures (8). The reduced amount hydrophobic residues prevent IDPs from forming stable hydrophobic cores that are observed in structured proteins. The lack of stable structure in IDPs gives rise to their promiscuous activity in the cell, being able to bind an array of functional proteins.

Structural Plasticity

One of IDPs hallmarks is their structural plasticity. Depending on the binding partner, IDPs may adopt di e structures, involving subsets of amino acids. p53 is a prime example for structural plasticity of IDPs. The negative regulatory domain (NRD) of p53 has many binding partners. The structure of p53 is known for four binding partners, Cyclin A2, S100B(ββ), Sirtuin, and CBP; which includes an extended, α-helix, β-stranded, or random coil structure, respectively (9) (Fig. 1.2). Structural plasticity of IDPs enables them to be involved in one-to-many and many-to-one signaling; where one IDP can cause several different signaling cascades or an IDP can contain several signaling domains and bind one partner, respectively (4).

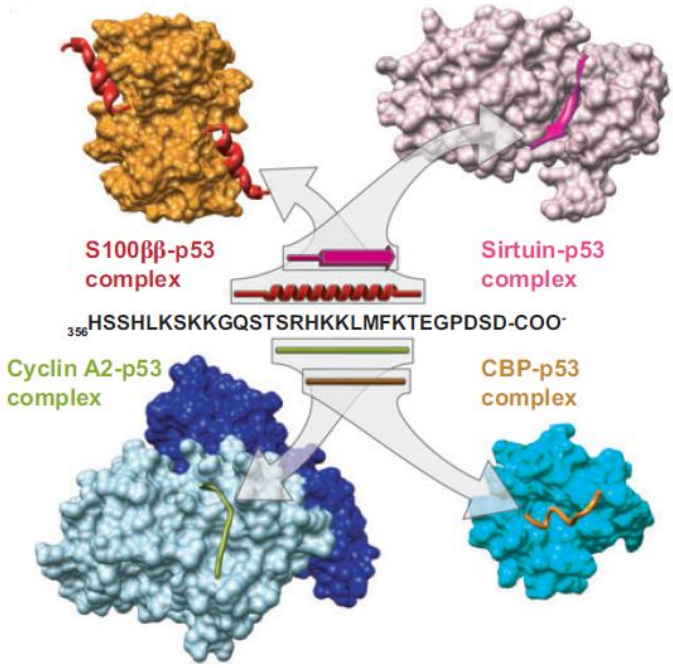

Figure 1.2 The structure of p53's NRD bound to four different binding partners. p53 adopts a different structure for each binding partner, α-helix, β-strand, extended structure, and random coil; when binding S100B(ββ), Sirtuin, Cyclin A2, and CBP, respectively (4).

Current Experimental Techniques for Studying IDPs

There are several experimental techniques currently used in studies of IDPs. These methods include nuclear magnetic resonance (NMR), X-Ray crystallography, and circular dichroism (CD). These are powerful structure determination techniques, but only give us a partial picture of IDP-protein interactions. X-ray crystallography can resolve a static crystal structure of IDP complexes. These rigid structures are insufficient at describing dynamic properties of IDPs. NMR studies have provided a wealth of structure and dynamic information about IDPs (7). These studies give us stable structures, conformational ensembles, and many other observables. These NMR ensembles can be very homogeneous. For IDPs, these experimental observables are measured by averaging over large set of heterogeneous conformational sub-states. The recovery of a heterogeneous ensemble from averaged properties

is a severely underdetermined problem. The major problem with the current methods is that they cannot access or fully describe the heterogeneous conformational ensemble of IDPs. A lack of knowledge of IDP conformational ensembles at the atomistic level has hindered the understanding of the binding and folding mechanism of IDPs.

Molecular Dynamic Investigation of IDPs

Advancements in computing power and quality of biomolecular force fields have increased the accuracy of molecular dynamic (MD) simulations of large biomolecular systems. One of the major appeals of MD simulations is the atomistic insight into the structure-function relationship of biomarcomolecules, which may be difficult to obtain experimentally (10). These insights gained from MD simulations have drawn many groups to employ this technique in their studies of IDPs (4,5,7,11-14). MD simulations of IDPs can allow us to observe the heterogeneous conformational ensembles of IDPs and their functional role in the folding pathway and coupled binding-folding mechanism (7).

A few examples of these studies include unbound residual structure, complex, along with coupled binding and folding studies. There are many examples of MD studies examining the residual structure of IDPs. The goal for these MD studies is to provide molecular detail of IDPs unbound residual structure. The unbound structure of IDPs may have importance in their protein-protein interactions and folding mechanisms. One example is the study of unbound structure of KID and pKID (15). In the study the authors look to determine the effect of phosphorylation of KID and how this effects binding to KIX (2). The conformational substates have marginal differences between KID and pKID structure. Interestingly, the study showed that phosphorylation of KID reduces the flexibility of a linker, which reduces the entropic cost for binding KIX (15). This was not observed in NMR studies because conformational changes were not great enough to appear on the ensemble level (2). Other studies of free IDPs have combined NMR experiments and MD simulations, such as Yoon *et al.* study of p21(16). In this study the authors show that free p21 has residual structure, helical and extended conformation, that resembles complex with either calmodulin or proliferating cell nuclear antigen (16). Studies of the nascent structure of free IDPs may give insights into their folding mechanisms.

MD simulations are used to study the folding mechanisms of IDPs, but there is not an abundance of these studies due to the high cost for these simulations. A study by Verkhivker *et*

al. used high-temperature and unfolding simulations to study the binding mechanism of p27 to Cdk2/Cyclin A (17). The results of this study suggest that p27 binds through an induced folding mechanism. This study pioneered the way for other MD studies of IDPs binding mechanism. Chen *et al.* ran 30 μs REX-MD simulations of NCBD-ACTR complex. In this study, the authors showed that NCBD binds ACTR through a conformational selection mechanism (18,19). Main of these studies use course grained proteins models and modified MD protocols, which reduce the cost of long time simulations and allow for greater conformational sampling. These studies have given us insights into the binding and folding mechanism of IDPs.

The two types of studies mentioned encompass many of the MD studies of IDPs. However, MD studies are also used to study the bound IDPs structure. Bevan *et al.* used MD simulations to study the structure of p53-NRD bound to S100B(ββ) (11). The results from this study show that p53-NRD adopts a stable α-helix. Studies of bound IDPs is important for designing potential inhibitors. Kriwacki *et al.* used MD simulations to study the structural dynamics of p27 bound to Cdk2 (20). Results of this study showed that p27 has a mixture of helical and extended conformations. The extended conformation segments are flexible, which the authors say is needed for p27's signaling processes (20). Many studies employ MD simulations to study the structure, folding and binding mechanisms of IDPs. However, these studies would not be possible without a biomolecular force field.

Empirical Molecular Mechanics Force Fields

The current biomolecular force fields use classical molecular mechanics treatment of atoms and their interactions (21). The potential energy function contains bonded terms (which include bonds, angles, dihedrals, and improper dihedrals energy terms), and a nonbonded terms (including Lennard-Jones potential and electrostatics):

$$V(\vec{R}) = \sum_{bonds} K_d(d - d_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} K_\chi\,(1 + \cos(n\chi - \delta)) +$$

$$\sum_{impropers} K_\varphi(\varphi - \varphi_0)^2 + \sum_{nonbonded} \left\{ \in_{ij} \left[\left(\frac{R_{ij}^{min}}{r_{ij}}\right)^{12} - \left(\frac{R_{ij}^{min}}{r_{ij}}\right)^{6} \right] + \frac{q_i q_j}{\in_l r_{ij}} \right\} \qquad (\mathbf{1})$$

K_d, K_θ, K_χ, and K_φ are bond, angle, dihedral, and improper dihedral force constants, respectively; d, θ, χ, and φ are the bond length, angle, dihedral angle, and improper dihedral angle values of

the current confirmation and the corresponding zero subscript are the reference values (21). The dihedral energy term contains multiplicity term, n, and the phase angle is represented by the δ variable (10). The force constants and reference values in eq.1 are obtained from experimental data and quantum mechanical studies of small molecules or macromolecule fragments. The final term is a combination of Lennard-Jones and Coulombic interactions between nonbonded atom pairs. $\in_{ij}$ corresponds to the well depth of the Lennard-Jones potential, R_{ij}^{min} is the distance where the potential equals zero, and r_{ij} is the distance between atoms *i* and *j*. The Coulombic term consist of the partial atomic charge, q_i and q_j, and $\in_l$ is the effective dielectric constant (21). Nonbonded parameters for atom pairs are calculated based on the Lorentz-Berthelodt combination rules, in CHARMM22, the $\in_{ij}$ values are calculated using the geometric mean between $\in_{ii}$ values and $\in_{jj}$ values, R_{ij}^{min} values are calculated using the arithmetic mean between R_{ii}^{min} values and R_{jj}^{min} values. The use of these combination rules may differ between force fields. The potential energy, *V(r)*, calculated from eq. 1 includes contributions from every bond, angle, and dihedral of the system. This energy is for a single configuration (i.e. snapshot) of the system. A series of snapshots are collected during simulations for subsequent system analysis.

There are several general-purposed biomolecular force fields available, CHARMM22 (22-24), AMBER99SB (25), OPLS-AA (26), and others. The major differences between force fields are the force constants and equilibrium values. Different biomolecular force fields are often optimized with different explicit water models. There are often systematic conformational biases, such as overstabilized helices and protein-protein interactions (1 & references therein). Some force fields, such as CHARMM22, have additional energy terms or use modified energy functions, in order to increase the accuracy of the force field (10). Despite these efforts, the current force fields are not perfect (10).

What is MD?

Molecular Dynamics (MD) is a sampling technique that solves Newton's equation of motion with respect to time. MD studies can give thermodynamic and kinetic information of protein complexes in solution. This information allows us to study dynamic properties of protein structure and folding mechanisms (21). MD simulations are expensive for two reasons. First, MD uses a time step in the femtosecond time scale, which is limited by vibrational frequencies

of chemical bonds. Secondly, for each step in time, the potential energy function is calculated for the biomolecular system. Together, these factors render MD computationally expensive. The computational cost of energy calculation and time step has limited MD to select time scales. Recent developments in MD methodology has allowed many to simulate proteins into the micro- and millisecond time range, by reduce number of atoms in the system and/or improving sampling.

Two primary challenges of atomistic simulations of IDPs are: force field accuracy and conformational sampling. The current biomolecular force fields are not perfect, leading to inherent errors in protein simulations (10). Besides force field problems MD faces another challenge. MD simulations have problems with adequately sampling the conformational space, which reduces accuracy of calculated thermodynamic and kinetic properties of proteins. Inadequate sampling can arise from large entropic and enthalpic energy barriers that hinder spontaneous conformational transitions (21). Adequate sampling of the conformational space often requires long time simulations (milliseconds to seconds timescale), which are computationally expensive or inaccessible in practice. MD simulations using explicit solvent have increased system sizes and atom number, which makes the energy calculation very expensive, due the added water molecules. This increased cost from solvent limits the accessible time scales to nanosecond and low microsecond range. Some these challenges can be overcome due to novel MD methodology and advancements in computer technology (21).

Chapter 2 - The Molecular Dynamic Study of p53-S100B(ββ) complex

Introduction

Over the past couple of decades, a new class of functional proteins lacking a stable structure under physiological conditions has caught the attention of structural biochemists and other related fields. These proteins are now known as intrinsically disordered proteins (IDPs) (12). IDPs have several features of great interest to many. IDPs bind their protein targets with low affinity and high specificity. The structural plasticity of IDPs allows them to bind many proteins with differing structures (1,8,12). Importantly, IDPs have heterogeneous conformations, and cannot be fully described by a single structure (5,7). Studies have shown that IDPs are involved in a wide range of biological function, such as cell signaling, proliferation, and differentiation (1,3,4,8,12). IDPs have also been shown to be involved in cancer and other diseases (4). Many groups across the globe are studying this intriguing newly recognized class of proteins for insights into their functions and therapeutic purposes (4,5,7,11-14).

Due to heterogeneous conformations studying IDPs through conventional structural studies has proven difficult (7). Conventional experimental methods only measure average conformational properties, subsequently suppressing IDPs conformational ensembles that are important in IDP function (5,7). Advances in computing technology and quality of force fields have made molecular dynamics a powerful tool for studying IDPs. Molecular dynamic simulations in principle allow us to observe the conformational evolution of IDPs with respect to time, giving more complete pictures of these proteins.

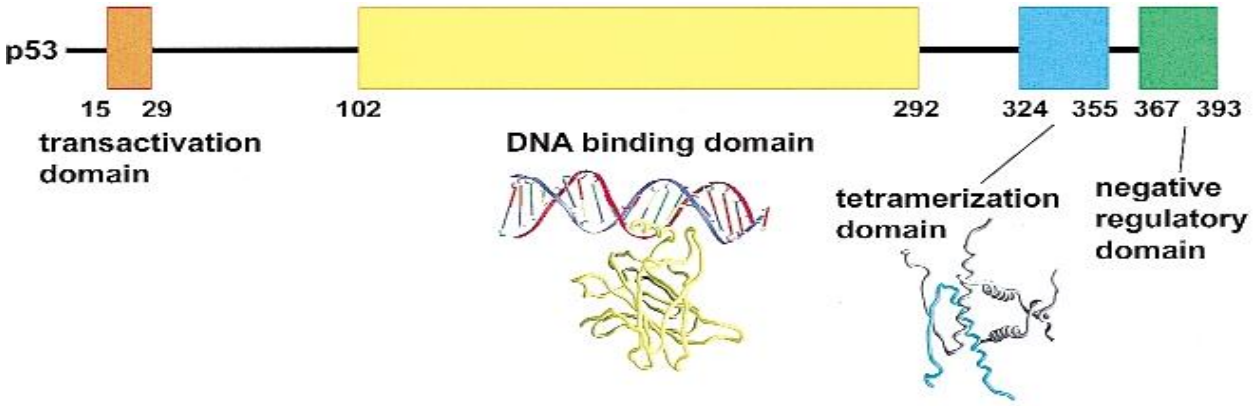

Figure 2.1 The Domains of the Tumor Suppressor p53 (20)

One IDP that has gained a lot of attention over the years is the tumor suppressor p53, due to its potential use in cancer therapy. In normal cells, p53 is present in low levels and serves as a hub for signaling transduction. p53 is activated in response to stress, which causes levels of active p53 to increase rapidly to prevent cell cycling, apoptosis, and repair DNA (27,28). Many studies have shown that the function of p53 is altered in cancer cells, due to mutations of p53 or inhibition by regulatory proteins (3,29,30). p53 is a modular protein that binds DNA as a tetramer, each subunit is comprised of the following domains. The central domain of p53 is a folded DNA binding domain. p53 contains two disorder domains at the termini. The C-terminus is known as the negative regulatory domain (NRD) and the N-terminus is known as the transactivation domain (TAD) (4,8) (Fig. 2.1). These two regions serve as signaling domains; the N-terminus actives transcription and the C-terminus has regulatory functions. The NRD contains phosphorylation and acetylation sites, which serve to control oligomerization, transcriptional activation, and DNA binding. This region has many binding partners but only four complex structures are known (see Fig. 1.2). One of these complexes is between p53-NRD and S100B(ββ).

The normal function of S100B(ββ) is negative regulation of p53. When S100B(ββ) binds p53-NRD, it blocks several post-translational modification sites, which inhibits the function of p53 (29-31). S100B(ββ) binding p53-NRD leads to dissociation of the p53 tetramer. This is needed in normal cells to allow further cellular proliferation if required. However, S100B(ββ) levels are elevated in cancer cells. This increase severely reduces functional p53, allowing the

cells to proliferate without control. This interaction is of great interest to the biomedical community as a possible cancer therapy target that could restore functional p53 levels.

Previous NMR studies show that p53-NRD adopts a stable helix when binding S100B(ββ) (PDB ID: 1DT7) (32). Despite the apparent well converged and rigid ensemble of NMR structures, there are several features that raise questions about its stable structure. A majority of the assigned intermolecular NOEs from NMR fall in the weak range (6-7 Å) (32), which is suggestive of a heterogeneous conformational state (7). Many of the NMR structures have buried charged residues and bulky side-chains lying across the helix to satisfy buried NOEs (Fig. 4). These characteristics from this NMR study lead us to believe that p53-NRD may sample a heterogeneous conformational ensemble when bound to S100B(ββ).

In this study we use extensive all-atom simulations to study the potential heterogeneity in the p53-S100B(ββ) complex. The NMR structures from previous study are used as initial conformations. Two well-known force fields are used for the simulations of p53-S100B(ββ) complex. The simulated ensemble satisfies a vast majority (82%) of the experimental NOEs. However, the ensemble is heterogeneous particularly in the p53 moiety. Clustering analysis of the simulated ensemble suggest that many conformations are sampled by bound p53, satisfying different sub-sets of intermolecular NOE restraints.

Methods

Explicit solvent simulation

The NMR ensemble consists of 40 NMR structures (PDB ID: 1DT7) and each contains two p53-NRD (369-388) in complex with a S100B(ββ) dimer (32). Five diverse structures were selected to initiate the independent simulations to facilitate conformational sampling (see Fig. 2.2). These conformers were identified based on mutual root mean square deviation (RMSD). The first NMR structure serves as a reference of the NMR ensemble, along with initial conformation for the first CHARMM (22-24,33) and GROMACS (34,35) simulation. The other conformers are the 21st, 23rd, 30th and 31st models in the NMR ensemble.

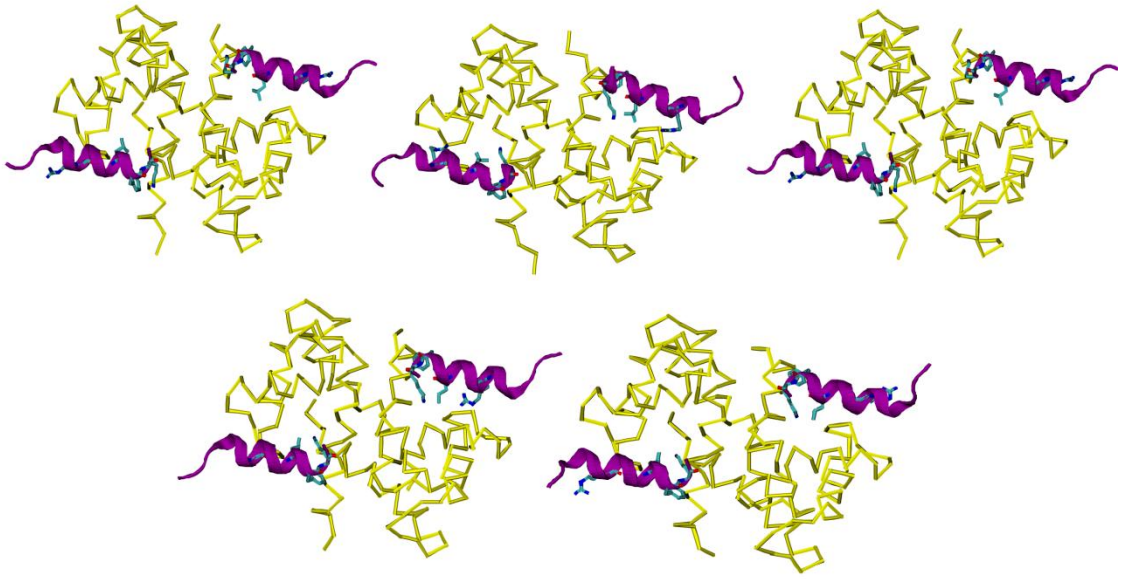

Figure 2.2 Five models from the NMR structure ensemble that were used as initial conformations for CHARMM simulations. From top left to right, the model numbers are 1st, 21nd, 23rd, 30th, and 31st. Structure 1 was also used in the Gromacs simulation. The residues shown are R379, L383, and F385.

The initial conformations require several modifications before proceeding with the simulations. The p53 segment consists of residues 369-388; N-terminal residues (369-373) are in an extended conformation and appear to have little involvement in the interaction. These residues are truncated to reduce size (and computational cost) of the simulations. The five models are solvated in a ~70×60×50 Å water box containing ~9011 TIP3P waters; overlapping waters are removed. Six calcium ions were added to neutralize the system.

The five systems are simulated in CHARMM (24,33). All molecules are described by CHARMM22/CMAP all-atom force field. Each system is equilibrated for 50 ps under NPT conditions, followed by a 100 ns NPT simulation. Weak harmonic positional restraints were placed on S100B(ββ) backbone to prevent large conformational deviations or dissociation of the S100B(ββ) dimer. SHAKE algorithm was used to constrain the bonds to allow for a two femtosecond (fs) time step (36). Particle Mesh Ewald (PME) was used to treat long-range electrostatics (37), and van der Waals (vdW) interactions use a 15.0 Å cutoff. For validation,

model 1 was simulated in GROMACS using AMBER99SB (25) following the same protocol as in CHARMM simulations.

Structural Analysis

All analysis was performed using CHARMM, Multiscale Modeling Tools for Structural Biology (MMTSB) toolset (38), and/or in house scripts, unless otherwise noted. NOE like distances were calculated from the trajectories for direct comparison with experimental values. For this, coordinates of all the atoms involved in the NOE assignments were first extracted from simulation. For NOEs that involve more than a single pair of atoms, the NOE distances were calculated using the r^{-6} summation (39),

$$r_{NOE}(t) = \left(\sum \frac{1}{r_{ij}^6}\right)_t^{-\frac{1}{6}} \quad \textbf{(2)}$$

where r_{ij} is the distance between atoms and t is time. The time-averaged NOE distances are calculated as:

$$r_{NOE} = \{\langle r_{NOE}(t)^{-6}\rangle_t\}^{-\frac{1}{6}} \quad \textbf{(3)}$$

Clustering

The last 60 ns of each 100 ns independent production simulations were included to extract an ensemble of p53-S100B(ββ). Each simulated model contains a dimer complex, and is seperated into two independent monomer conformations. When these five simulations are combined, there is a total of 600 ns of production sampling of the monomer complex. A conformation was extracted every 100 ps, generating a conformational ensemble of 6000 p53-S100B(ββ) monomer complex structures. This ensemble was clustered using the MMTSB toolset (38), using k-mean algorithm, where members of the cluster are assigned based on distances to the cluster center. The distance is based on p53 C_α RMSD, calculated by first aligning the complex using the backbone atoms of S100B(ββ). The calculated RMSD values as defined reflect both internal conformational or binding mode changes of p53. Several cutoff radii were tested in clustering; with a final value of 5 Å that appeared to give clusters with the cleanest separation. NOE and helicity analysis was performed for each of the ten most populated clusters identified. These clusters presumably represent sub-states sampled by p53 when bound to S100B(ββ).

Results and Discussion

Structural and NMR evidence for fuzziness in the p53/S100B(ββ) complex

NMR structures of p53/S100B(ββ) have several features that suggest possible conformational heterogeneity. As mentioned in the introduction, these features include unsatisfied buried charges, bulky residues and large number of weak intermolecular NOEs. For example, R379 of p53-NRD is buried within the hydrophobic binding site of S100B(ββ) (Fig. 2.3). These residues prefer hydrophilic environments, which are not observed in the NMR structures. We propose that these residues may adapt their conformation to a suitable environment.

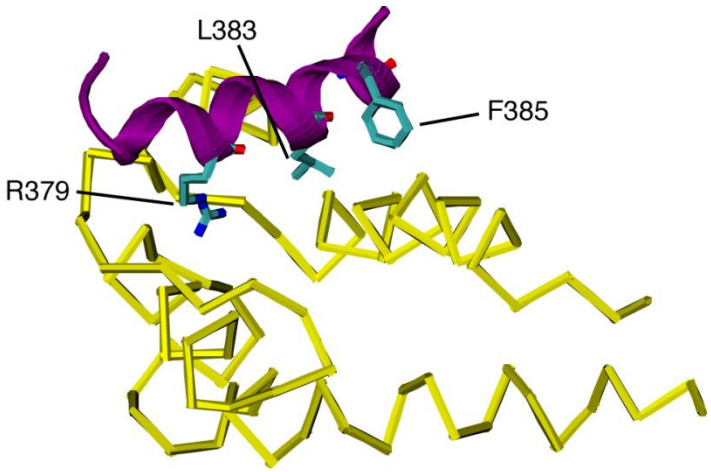

Figure 2.3 Monomer of p53-S100B(ββ) complex extracted from the NMR structure (PDB ID: 1DT7). Three key residues are shown above F385, R379, and L383. L383 is the proposed anchoring residue from NMR studies. R379 is an example of an unsatisfied buried charged residue observed in the NMR structure. The other residue of interest is F385, which is seen to lay across the helix in the NMR structure, but is involved in several intermolecular NOEs to S100B(ββ).

If the structure of p53 is stable in complex with S100B(ββ), then we would except to observe a larger number of strong NOEs. However, a only a set of very weak NOEs were assigned for this complex (see Table 2.1). Several of the experimental intermolecular NOEs involve buried atom pairs that are not satisfied by these stable NMR structures. These

observations lead us to believe that p53 is heterogeneous and samples a series of conformations that satisfy sub-sets of intermolecular NOEs.

Stability of the p53/S100B(ββ) complex

During simulation, large changes are observed in conformation and binding pose of p53-NRD. As shown in Fig. 2.4 A, The C_α RMSD rapidly increases to ~5.5 Å in simulation. These large deviations in C_α RMSD are correlated with a change in both conformation of p53-NRD and its relative position with respect to S100B(ββ) (Fig. 2.5).

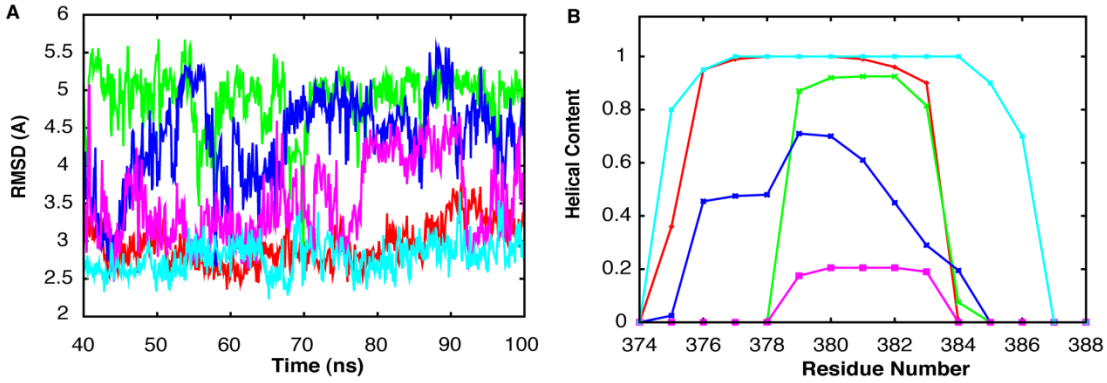

Figure 2.4 (A) Cα RMSD of p53 from the five simulated models. Model 1 (red), Model 21 (green), Model 23 (blue), Model 30 (pink), Model 31 (cyan). **(B)** Residue helicity plot shows four of the simulated models and experimental helicity (cyan); several models had similar helicity values to the red line.

The residue helicity of simulations shows a reduction in the helical content of p53 (Fig. 2.4B). Greater conformational changes in p53-NRD are observed in models with large RMSDs. The large conformational changes are correlated to previously unsatisfied charges (R379) become solvent exposed. However, models that retain their helicity display changes in the binding position of p53-NRD. The large RMSDs and changes in residue helicity suggest that p53-NRD is heterogeneous.

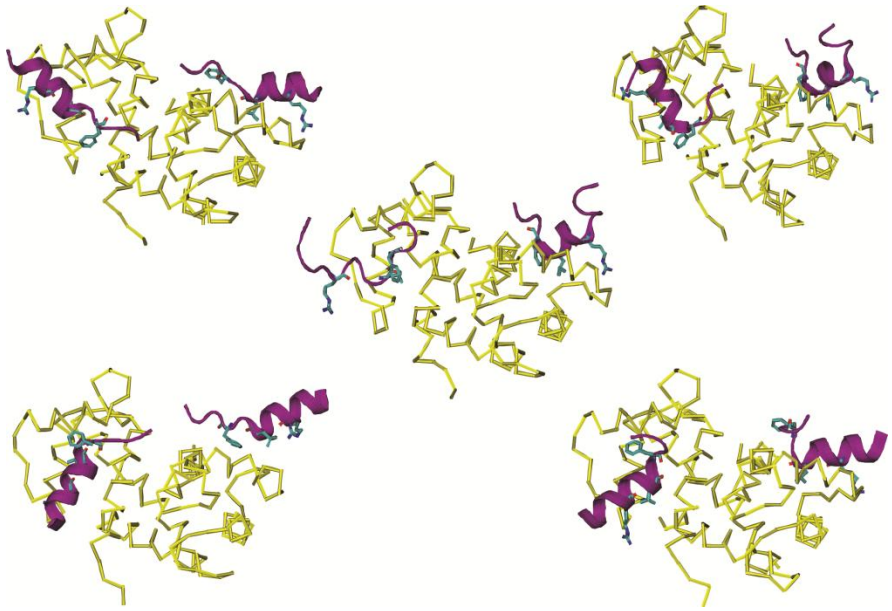

Figure 2.5 Final snapshots of the five independent simulations; top left to right, simulated models 1st, 21st, 23rd, 30th, and 31st.

Previous simulations have shown that p53 remains very stable (11), and many of the changes we observed in our simulations are absent in these simulations. To confirm that our findings were not an artifact of the CHARMM22/CMAP force field, the system was also simulated using the AMBER99SB force field in GROMOS. The GROMOS simulation also showed a loss in helicity along with large RMSDs (Fig. 2.6).

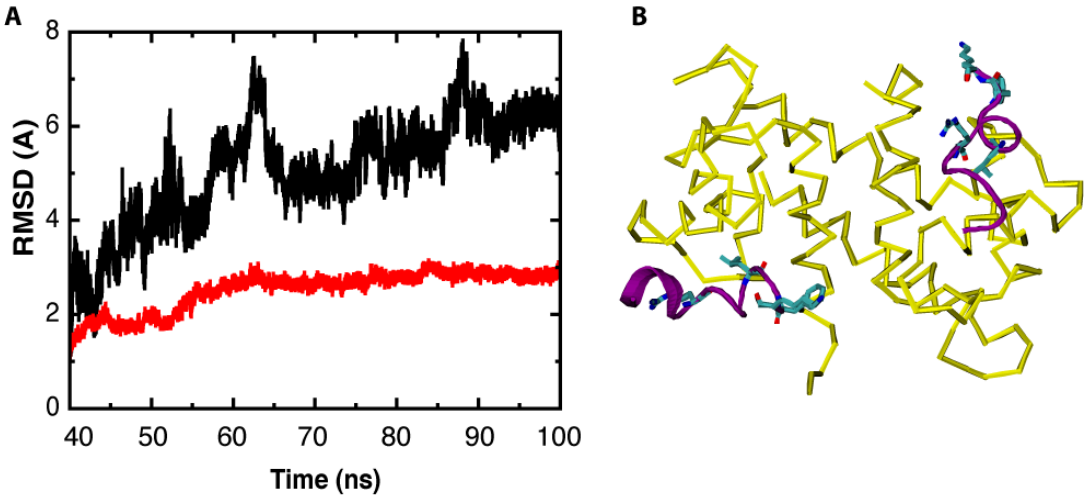

Figure 2.6 (A) C_α RMSD of p53 and S100B(ββ), black and red respectively. **(B)** Final snapshot of p53-S100B(ββ) from GROMACS simulations.

The GROMACS simulation of p53-S100B(ββ) complex confirms our results from the CHARMM22/CMAP simulations. The stable nature of p53 observed in previous simulations may arise from differences in initial conditions or duration of simulation, which may suppress the dynamic features of p53 that are observed in this study. Taken together, the current extensive simulations clearly support that p53-NRD remains highly dynamic when bound to S100B(ββ).

NOE analysis of simulated models

In order to examine how realistic the simulated ensemble is, we back calculated NOE-like distances and compared them with experimental values. The NOE-like distances obtained from simulation underwent summation and time averaging (see Methods). As shown in Table 2.1, the majority of NOEs calculated from the five independent simulations are satisfied, with only 6 or 18% of the 33 intermolecular NOEs violated. The violated NOEs involve previously buried charged residues and/or bulky side chains that become solvent exposed. Figure 2.7 and 2.8 show the histograms and time-traces of a few representative NOE distances. There is significant heterogeneity in the instantaneous NOE distances, $r_{noe}(t)$ as defined by Eq. 2, frequently fluctuating above the experimental upper bound. Note that these NOE-average distances can be misleading when comparing to NOE distances of the individual models. These values obtained from r^{-6} averaging of the whole ensemble emphasize contributions from conformers with shorter distances.

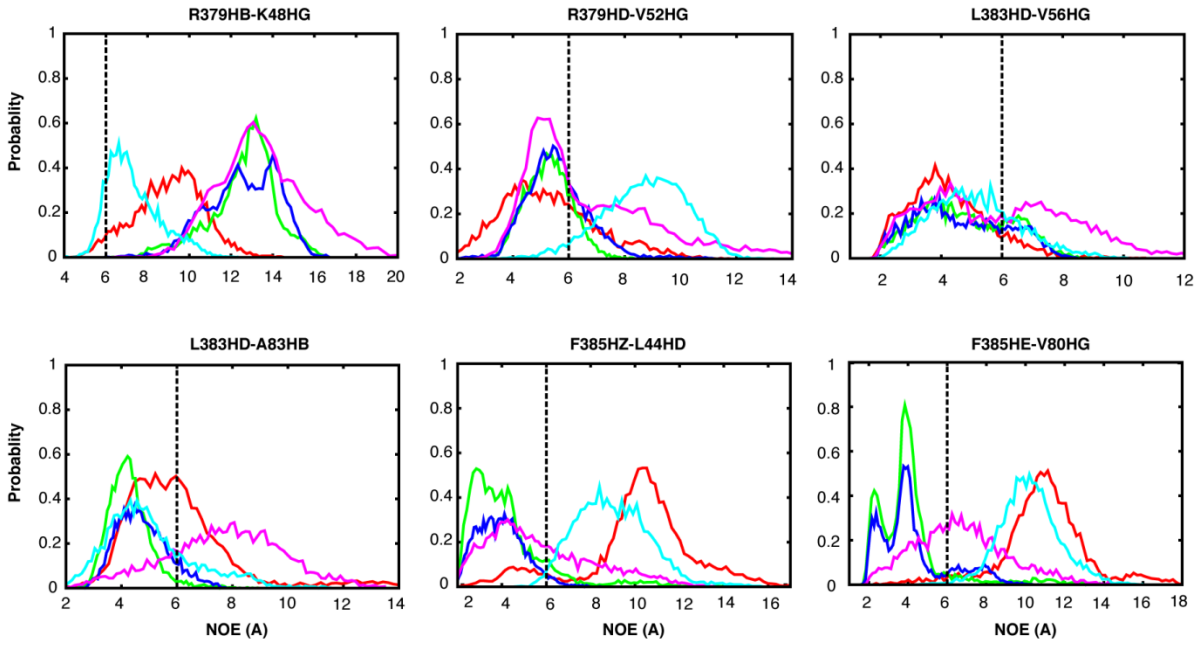

Figure 2.7 Histograms of six NOEs, two for each of these residues R379, L383, and F385, from the five independent simulations. The same color code is used as in Fig. 2 A. The black dashed line represents the experimental NOE upper bound.

Table 2.1 The first column contains the atom pairs involved in intermolecular NOEs. Groups of atoms are denoted by HX*. Second column has the upper bound from experiment. Third column is the r^{-6} summation of the ten monomers from simulations. The final column is the violation between simulation and experiment. Zeroes correspond to satisfied NOEs.

NOE	Exp. Upper Bound (Å)	Sim. Distance (Å)	Violation (Å)
Q375 HB* - E51 HG*	6	7	1
R379 HB* - K48 HG*	5	7	2
R379 HD* - V52 HG2*	6	3	0
L383 HD* - L3 HD*	6	6.2	0
L383 HD* - V52 HG2*	6	3.1	0
L383 HD* - V56 HG*	5	2.2	0
L383 HD* - M7 HE*	5	6.2	1.2
L383 HD* - V77 HG2*	6	7.3	1.3
L383 HD* - M79 HE*	3.3	3.1	0
L383 HB* - M79 HE*	5	4	0
L383 HD* - V80 HG2*	4.2	2.9	0
L383 HD* - V80 HG1*	5	3.4	0
L383 HD* - T82 HG2*	6	7.1	1.1
L383 HB* - A83 HB*	5	3.1	0
L383 HD* - A83 HB*	5	4.7	0
M384 HE* - L44 HD*	6	3.4	0
M384 HB* - K55 HG*	6	4.5	0
M384 HG* - V56 HG*	6	2.3	0
M384 HE* - V56 HG*	6	2.8	0
F385 HZ - L44 HD*	6	3	0
F385 HE* - M79 HE*	6	4.9	0
F385 HZ - M79 HE*	6	6.5	0
F385 HE* - V80 HG2*	6	3	0
F385 HZ - V80 HG2*	6	4.3	0

F385 HD* - V80 HG1*	6	5	0
F385 HE* - A83 HB*	5	4.1	0
F385 HZ - A83 HB*	6	3.3	0
K386 HD* L44 HD*	6	8.4	2.4
K386 HD* - T82 HG2*	5	3.4	0
K386 HG* - T82 HG2*	6	4.4	0
T387 HG* - K55 HG*	5	4.2	0
E388 HG* - M79 HE*	5	4.6	0
E388 HG* - T82 HG2*	6	4.2	0

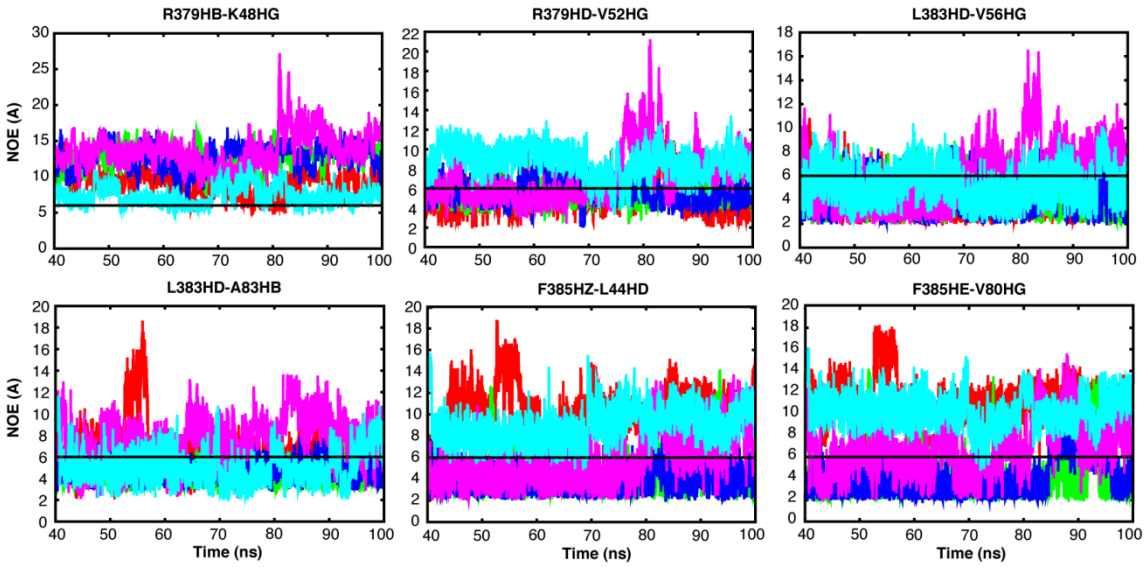

Figure 2.8 Time averaged NOE distances over the final 60ns of simulations; these are the same NOEs from the histograms above. The black line is the experimental upper bound.

The ability of the heterogeneous ensemble derived from extensive MD simulations to reproduce experimental NOEs strongly supports our hypothesis that p53 is dynamic when bound to S100B(ββ).

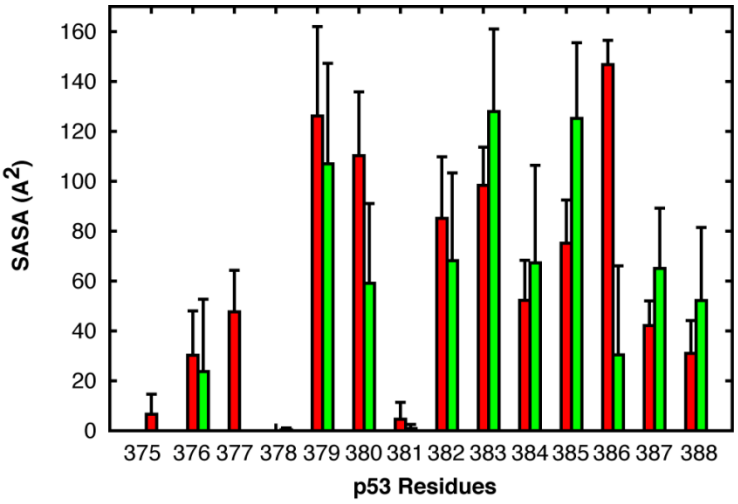

Figure 2.9 Solvent accessible surface area (SASA) of p53 residues from simulation and experiment. The red bars are the average SASA of 40 NMR structures with error bars. The green bars are the average SASA of the 5 simulations with error bars.

More importantly, can this heterogeneous ensemble still inhibit activation of p53? To answer this we calculated the buried solvent accessible surface area (SASA) of p53. As shown in Figure 2.9, the SASA profiles between experimental and simulated ensembles are very similar. There are several residues in simulation (R379 and K386) that have increased SASAs compared to experiment. These changes are caused by the residues flip out of the binding site during simulation to satisfy their unpaired charges. However, key residues that need to be block from post-translational modifications (Phosphorylation sites: S376 and T377, Acetylation sites: R379 and K386) are similarly buried. This suggests that despite the dynamic nature of p53 in simulation, the inhibition of p53 is maintained.

Clustering analysis

Data collected from the five independent simulations show that the structure of p53 is highly dynamic when interacting with S100B(ββ). NOE analysis of simulation showed varying sets of satisfied NOEs over the simulations. From these results, we suspect that the complex samples a distinct set of conformational sub-states, which could give rise to different NOE subsets. The ensemble of sub-states together samples all intermolecular NOEs assigned

experimentally. Ensemble clustering is used to identify the largest sampled sub-states. The highest populated sub-states are shown in figure 2.10.

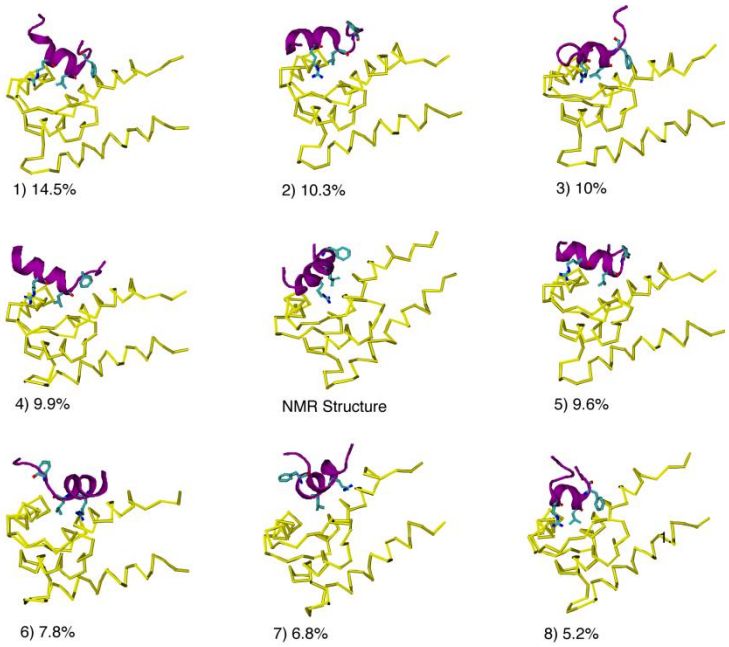

Figure 2.10 Top ranked cluster conformations with NMR structure for reference; yellow trace is S100B(ββ) and purple cartoon p53. The value next to the rank is the percent of structures occupying that cluster.

Many of the proposed sub-states are highly helical; exceptions to this are clusters 3 and 8, which have coiled conformations at the termini (Fig. 2.10). The largest difference between sub-states is the binding poses of p53 with respect to S100B(ββ). These features appear to arise from the increased solvation of R379 and K386 side-chains. Interestingly, p53 segments pivot around L383 (Fig. 2.10). We have also analyzed NOE violation for each cluster. As shown in figure 13, individual clusters satisfy different NOE subsets. All of the sub-states satisfy a majority of the L383 NOEs, which are important for anchoring of p53. The clusters also violate many of the NOEs involving charged terminal residues (R379, E388, K386, and Q375) (Fig. 2.11), which have the most dynamic residual structure. These differences in NOE subsets arise from conformation and positional differences in p53. When these sub-states are observed as an ensemble, all experimental NOE distances are satisfied. This shows that p53 samples a heterogeneous ensemble of many sub-states that fulfill the intermolecular NOEs, and inhibit the activation of p53.

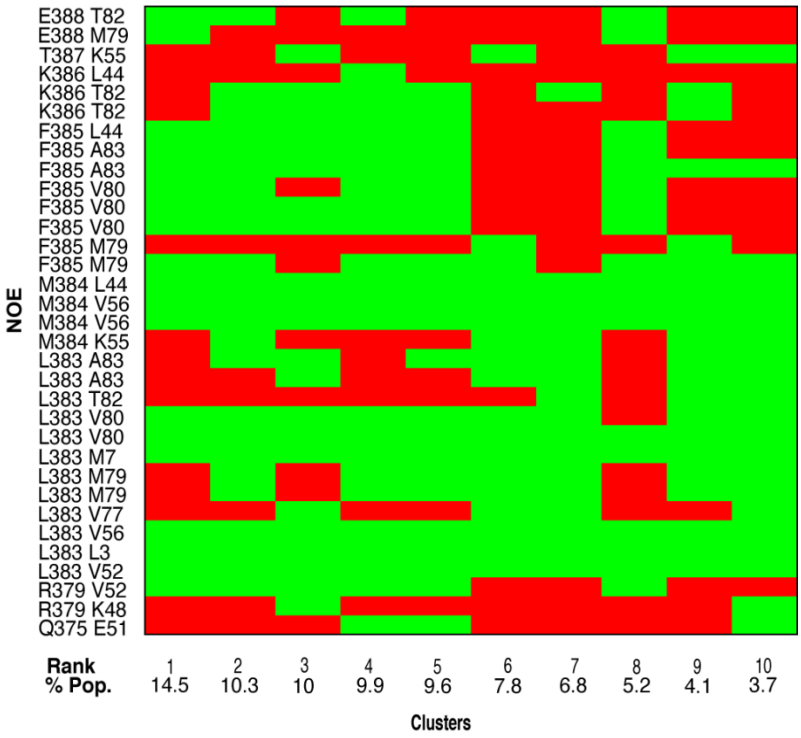

Figure 2.11 NOE violations of the top ten highest populated clusters. Intermolecular NOEs are shown on the y-axis. The green blocks are satisfied and red are violated.

Conclusions

We use extensive MD simulations to study the p53-S100B(ββ) complex, in which p53 was previously thought to adopt a stable alpha-helix. However, the calculated structures and NOEs provide evidence that p53-NRD may not have a stable structure. Extensive MD simulations of the p53-S100B(ββ) complex show that p53-NRD is highly dynamic, while satisfying 82% of the experimental NOEs. We propose a conformational ensemble of p53-NRD that fulfills the intermolecular NOEs. These sub-states still block key post-translational modification sites needed for p53 activation. In this study, we show that p53 is heterogeneous, which samples an ensemble of conformations that still prevent p53 activation. We also show MD is a good technique for studying the structure and dynamics of IDPs, further understanding the biological function of these newly recognized proteins.

References

1. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., and Obradovic, Z. (2002) *Biochemistry* **41**, 6573-6582
2. Click, T. H., Ganguly, D., and Chen, J. (2010) *International journal of molecular sciences* **11**, 5292-5309
3. Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradović, Z., and Dunker, A. K. (2002) *Journal of molecular biology* **323**, 573-584
4. Uversky, V. N., Oldfield, C. J., and Dunker, A. K. (2008) *Annual review of biophysics* **37**, 215-246
5. Tompa, P., and Fuxreiter, M. (2008) *Trends in biochemical sciences* **33**, 2-8
6. Mittag, T., Marsh, J., Grishaev, A., Orlicky, S., Lin, H., Sicheri, F., Tyers, M., and Forman-Kay, J. D. (2010) *Structure* **18**, 494-506
7. Mittag, T., and Forman-Kay, J. D. (2007) *Current opinion in structural biology* **17**, 3-14
8. Dunker, A. K., Silman, I., Uversky, V. N., and Sussman, J. L. (2008) *Current opinion in structural biology* **18**, 756-764
9. Oldfield, C. J., Meng, J., Yang, J. Y., Yang, M. Q., Uversky, V. N., and Dunker, A. K. (2008) *BMC genomics* **9 Suppl 1**, S1
10. Mackerell, A. D., Jr. (2004) *Journal of computational chemistry* **25**, 1584-1604
11. Allen, W. J., Capelluto, D. G., Finkielstein, C. V., and Bevan, D. R. (2010) *The journal of physical chemistry. B* **114**, 13201-13213
12. Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., and Obradovic, Z. (2001) *Journal of molecular graphics & modelling* **19**, 26-59
13. Haupt, Y., Maya, R., Kazaz, A., and Oren, M. (1997) *Nature* **387**, 296-299
14. Chen, J. (2009) *Journal of the American Chemical Society* **131**, 2088-2089
15. Ganguly, D., and Chen, J. (2009) *Journal of the American Chemical Society* **131**, 5214-5223
16. Yoon, M. K., Venkatachalam, V., Huang, A., Choi, B. S., Stultz, C. M., and Chou, J. J. (2009) *Protein science : a publication of the Protein Society* **18**, 337-347
17. Verkhivker, G. M. (2005) *Proteins: Structure, Function, and Bioinformatics* **58**, 706-716
18. Ganguly, D., Zhang, W., and Chen, J. (2012) *Molecular bioSystems* **8**, 198-209
19. Zhang, W., Ganguly, D., and Chen, J. (2012) *PLoS computational biology* **8**, e1002353
20. Galea, C. A., Nourse, A., Wang, Y., Sivakolundu, S. G., Heller, W. T., and Kriwacki, R. W. (2008) *Journal of molecular biology* **376**, 827-838
21. Adcock, S. A., and McCammon, J. A. (2006) *Chemical reviews* **106**, 1589-1615
22. MacKerell, A. D., Bashford, D., Bellott, Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) *The Journal of Physical Chemistry B* **102**, 3586-3616
23. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) *Journal of computational chemistry* **4**, 187-217

24. MacKerell, A. D., Brooks, B., Brooks, C. L., Nilsson, L., Roux, B., Won, Y., and Karplus, M. (2002) CHARMM: The Energy Function and Its Parameterization. in *Encyclopedia of Computational Chemistry*, John Wiley & Sons, Ltd. pp
25. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) *Proteins* **65**, 712-725
26. Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996) *Journal of the American Chemical Society* **118**, 11225-11236
27. Kubbutat, M. H. G., Jones, S. N., and Vousden, K. H. (1997) *Nature* **387**, 299-303
28. Symonds, H., Krall, L., Remington, L., Saenz-Robles, M., Lowe, S., Jacks, T., and Van Dyke, T. (1994) *Cell* **78**, 703-711
29. Baudier, J., Delphin, C., Grunwald, D., Khochbin, S., and Lawrence, J. J. (1992) *Proceedings of the National Academy of Sciences of the United States of America* **89**, 11627-11631
30. Lin, J., Blake, M., Tang, C., Zimmer, D., Rustandi, R. R., Weber, D. J., and Carrier, F. (2001) *The Journal of biological chemistry* **276**, 35037-35041
31. Fernandez-Fernandez, M. R., Rutherford, T. J., and Fersht, A. R. (2008) *Protein science : a publication of the Protein Society* **17**, 1663-1670
32. Rustandi, R. R., Baldisseri, D. M., and Weber, D. J. (2000) *Nature structural biology* **7**, 570-574
33. Brooks, B. R., Brooks, C. L., 3rd, Mackerell, A. D., Jr., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009) *Journal of computational chemistry* **30**, 1545-1614
34. Lindahl, E., Hess, B., and van der Spoel, D. (2001) *Journal of molecular modeling* **7**, 306-317
35. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005) *Journal of computational chemistry* **26**, 1701-1718
36. Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977) *Journal of Computational Physics* **23**, 327-341
37. Darden, T., York, D., and Pedersen, L. (1993) *The Journal of chemical physics* **98**, 10089-10092
38. Feig, M., Karanicolas, J., and Brooks, C. L., 3rd. (2004) *Journal of molecular graphics & modelling* **22**, 377-395
39. Fletcher, C. M., Jones, D. N. M., Diamond, R., and Neuhaus, D. (1996) *Journal of biomolecular NMR* **8**, 292-310