UNIVERSITY OF CALIFORNIA, SAN DIEGO

Automation of scientific workflows along with studies of hemagglutinin antibody and p53 DNA complexes

A Thesis submitted in partial satisfaction of the
requirements for the degree Master of Science

in

Chemistry

by

Pek U Ieong

Committee in charge:

Professor Rommie Amaro, Chair
Professor J. Andrew McCammon
Professor Elizabeth Villa

2016

The Thesis of Pek U Ieong is approved and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2016

Dedication

To my parents: Hip Ng and Su Wong Ieong,
And to my late grandfather Mr. Ng

Table of Contents

List of Abbreviation

MD, molecular dynamics

VMD, Visual Molecular Dynamic

RMSD, root mean square deviation

PDB, protein data bank

fl-p53, full length p53

HA, hemagglutinin

CADD, computed-aided drug discovery

NBCR, National Biomedical Computation Resource

HPC, high-performance computing

RCS, relax complex scheme

VS, virtual screening

XSEDE, Extreme Science and Engineering Discovery Environment

SaaS, Scientific Software as a Service

AUC, area under the curve

ROC, receiver operating characteristic

Ig-2D1, Immunoglobulin 2D1

IgL, Immunoglobulin light chain

IgH, Immunoglobulin heavy chain

18HA, 1918 pandemic flu

09HA, 2009 pandemic flu

06HA, 2006 seasonal flu

09HA_mut, a 2009 pandemic flu mutant

MM-PBSA, molecular mechanic - Poisson-Boltzmann surface area

MM-GBSA, molecular mechanic - Generalized Born surface area

SE, standard error

NTD, N-terminal domain

DBD, DNA binding domain

CTD, C-terminal domain

TET, tetrameric domain

RE, response element

PCA, principal component analysis

List of Supplementary Video

List of Figures

List of Tables

Acknowledgements

I would like to acknowledge Professor Rommie Amaro for her support as the chair of my committee. She has been a great mentor and her support is proven to be the key to my research journey.

I would also like to acknowledge the Amarolab members, especially Özlem Demir, Rob Swift, Jeffrey Wagner, Robert Malmstrom, Jacob Durrant and former Amarolab members Jesper Sørensen and Lane Votapka.

Chapter 1, in full, is a reprint of the material as it appears in Ieong, Pek U., Sorensen, Jesper, Vemu, Prasantha L., Wong, Celia W., Demir Özlem, Williams, Nadya, P., Wang Jianwu, Crawl, Daniel, Swift, Rob V., Malmstrom, Robert. D., Altintas, Ilkay, Amaro, Rommie. E., "Progress towards automated Kepler scientific workflows for computer-aided drug discovery and molecular simulations". Procedia Computer Science. 2014. Vol. 29, 1745-1755. The thesis author was the primary investigator and author of this paper.

Chapter 2, in full, is a reprint of the material as it appears in Ieong, Pek U; Li, Wilfred; Amaro, Rommie E., "Molecular Dynamic Analysis of Antibody Recognition and Escape by Human H1N1" Biophys. J., 2015. Vol 108, Issue 11, 2704-2712. The thesis author was the primary investigator and author of this paper.

Chapter 3, in full, has been submitted for publication of the material as it may appear in Full-length p53 Tetramer Bound to DNA and Its Quaternary

Dynamics, 2016. Demir, Özlem; Ieong, Pek U; Amaro, Rommie E., PNAS, 2016. The thesis author and Dr. Özlem Demir were primary co-investigators and authors of this paper.

ABSTRACT OF THE THESIS

Automation of scientific workflows along with studies of hemagglutinin antibody and p53 DNA complexes

by

Pek U Ieong

Master of Science in Chemistry

University of California, San Diego, 2016

Professor Rommie Amaro

Molecular Dynamic (MD) simulation is a powerful computational tool that can be applied to study biological systems at an atomic scale. Antibody 2D1 was isolated from the 1918 influenza virus surface glycoprotein hemagglutinin (HA) and was also known to cross-neutralize the 2009 pandemic influenza HA. Nevertheless, the detailed mechanism is unclear. We have conducted molecular dynamic (MD) simulations to study the interactions

between Ig-2D1 and the HAs from four different strains including its natural binder 1918HA, the 2009 HA, a seasonal 2006 strain and a 2009HA mutant. We found that in 09HA, a serine to asparagine mutation from the 18HA weakened one of the salt bridges, which leaded to the loss of hydrogen bonds and the formation of a water pocket between 09HA and Ig-2D1. Another system involves the cancer suppressor, full-length p53 protein, and its DNA counter-parts. In this system, we observed that the C-terminals contacted DNA and formed direct salt bridges. This observation supported previous research, which reported that the C-terminals interact with DNA nonspecifically to search for the binding sequence. Each of these observations was possible because MD simulations provide atomistic detail, which facilitates the study of protein-protein and protein-DNA interactions. Additionally, MD simulations can furnish refined results, but the simulation and analysis processes can be daunting. To ease the complications, we utilized the Kepler platform and developed several automated workflows that integrated multiple commands into one central process.

Chapter 1

Progress towards automated Kepler scientific workflows for computer-aided drug discovery and molecular simulations

Abstract

We describe the development of automated workflows that support computed-aided drug discovery (CADD) and molecular dynamics (MD) simulations and are included as part of the National Biomedical Computational Resource (NBCR). The main workflow components include: file-management tasks, ligand force field parameterization, receptor-ligand molecular dynamics (MD) simulations, job submission and monitoring on relevant high-performance computing (HPC) resources, receptor structural clustering, virtual screening (VS), and statistical analyses of the VS results. The workflows aim to standardize simulation and analysis and promote best practices within the molecular simulation and CADD communities. Each component is developed as a stand-alone workflow, which allows easy integration into larger frameworks built to suit user needs, while remaining intuitive and easy to extend.

1.1 Introduction

Using computer simulation as an aid in drug discovery is not novel, yet the field is sometimes still considered in its infancy, an opinion that may be due to the relatively complicated processes involved and the lack of community-wide standard procedures. Furthermore, the continuous development of new computer architectures and software parallelization can result in large amounts of data, upwards of 1 terabyte for single computer-aided drug discovery (CADD) projects. Perhaps as a result of this enabling technology, it is common for practitioners to spend more time analyzing the data than generating it. With this in mind, our aim is the development of robust, reusable workflows for simulation preparation, job execution, and analysis that simplify best practices and help the community make the most of their rich data sets.

To develop automated, standardized protocols, we employ Kepler(1), a scientific workflow framework. Kepler is a free, open-source software suite designed for analyzing and modeling scientific data. The Kepler software simplifies the creation of executable models (scientific workflows), even by researchers with little programming background (2). Additionally, it is a platform for users to share and reuse data, workflows, and components for a wide range of scientific and engineering applications. Kepler has powerful support to handle new cyber infrastructure demands (e.g., intelligently

handling/brokering access to Extreme Science and Engineering Discovery Environment (XSEDE) and other simulation-relevant platforms), and it is particularly well suited to handle workflows that cross scales. The flexibility of Kepler makes it an ideal environment for sharing methods among scientists, thus increasing reproducibility and accessibility. Kepler also provides a provenance (e.g., data lineage and the processing history of workflow runs) framework that collects information, which can then be viewed through a molecular modelers' virtual notebook.

1.2 The relaxed complex scheme – main components

Previously, we developed a CADD workflow called the relaxed complex scheme (RCS), (3, 4), an end-to-end CADD experiment that incorporates receptor flexibility into virtual screening (VS) by utilizing molecular dynamics (MD) simulations. As summarized schematically in Figure 1.1, the RCS workflow facilitates all steps of VS, including: 1) generating compound libraries, 2) generating and selecting receptor structures, 3) performing virtual screens, 4) reevaluating and characterizing docked poses, and 5) sharing virtual-screening results. While not illustrated in Figure 1.1, workflow results lend themselves to statistical validation, an extension discussed in section 1.3.7

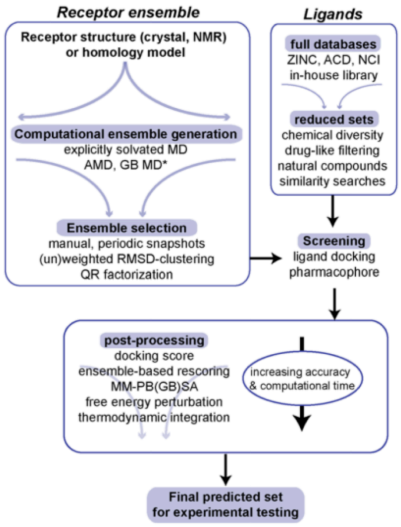


Figure 1.1 General workflow for ensemble-based VS experiment. Blue arrows indicate size of data sets (i.e. increasing or decreasing) at each step; * denotes emerging methods that have not yet been tested. (AMD: accelerated molecular dynamics, GB MD: generalized Born molecular dynamics, RMSD: root-mean-square- deviation, ZINC – ZINC Is Not Commercial, ACD: Available Chemical Database, NCI: National Cancer Institute, MM-PB(GB)SA: Molecular Mechanics – Poisson-Boltzmann (Generalized Born) Surface Area).

Building on our earlier RCS efforts, we are developing individual, stand-alone workflows that are reusable and modular. Collectively, they form a "toolkit" of powerful methods that can be assembled to address challenging VS problems. In particular, to incorporate protein flexibility into rational drug discovery and design, we are constructing a class of workflows to automate the setup, execution, and evaluation of molecular dynamics simulations. The workflows can be assembled in novel ways, creating environments where system-specific MD analysis can be meaningfully conducted, providing extended utility beyond CADD and the RCS. Each Kepler-based reusable workflow module is called an "actor" and is built on an open-source software

platform, or on software that is free to academic groups. The near universal accessibility of the workflows should translate to broad dissemination and use, allowing researchers to handle the challenges inherent in (big) data more effectively.

To prevent each workflow from becoming a “black box”, where appropriate, we are focused on including metrics or analytics that allow the user to judge the quality of the output and make key scientific decisions. As an example, we will focus on providing applications that make conducting and reporting novel MD analysis standard, routine and reproducible (5). Additionally, we plan to build workflows that support data sharing and transportation through cloud and other distributed platforms, using technologies including GlobusOnline and UDT that also facilitate usage of high-speed networks. The combination of these functionalities will provide a simple but powerful way to create and share customizable reports among members of large scientific collaborations.

1.3 CADD workflow – main actors

1.3.1 File management for ligand parameterization

For organization purpose, we developed a Kepler composite actor, which takes a list of PDB files and creates subdirectories using the PDB

names. Subsequently, the PDB files are copied to the corresponding subdirectories. This way, data associated with each PDB is stored consistently, providing better information control. While this actor is small actor, it provides proper file management, a crucial component of CADD.

1.3.2 Ligand Parameterization

A MD simulation of a protein-ligand complex requires development of ligand force field parameters. Parameterization can be cumbersome and is commonly a multi-step process handled by a series of user scripts. To streamline this process, we developed a ligand parameterization composite actor (Figure 1.2), that follows the “gold standard” Amber protocol, using Antechamber (6) and Gaussian (7). For each ligand, Antechamber assigns generalized Amber force field (GAFF) (8) atom types, while Gaussian performs a minimization before calculating the electrostatic potential (ESP), both at the HF/6-31G* level. Atomic partial charges are then assigned to reproduce the Gaussian ESP using the RESP protocol (9) in Antechamber. The only required inputs are the small molecule PDB files. This composite actor subsequently outputs the required FRCMOD and PREPC files containing the force field parameters, which are reusable and easily shared.

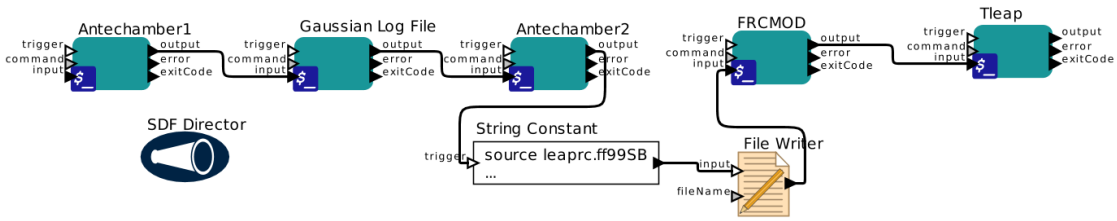


Figure 1.2 Kepler composite actor for the parameterization of small molecule ligands for MD

This composite actor (see Figure 1.2) helps scientists automate the ligand parameterization process by 1) connecting sequential steps and 2) providing input parameters and commands for each parameterization step. Once a user provides a small molecule PDB file to start the workflow, outputs from previous steps will become the inputs for the following steps. This actor will read the PDB file from the assigned workflow parameter settings and allows users to easily modify the location of the PDB file as needed for their simulations.

1.3.3 Receptor-ligand molecular dynamic simulations

The binding of a ligand to a receptor is a dynamic event. Small molecule compounds can assume many different binding poses, and receptor flexibility may change due to ligand binding. Therefore, it is important to consider the dynamic behavior of both ligands and receptors during CADD. The steps to prepare an MD simulation can be routine but lengthy, especially when considering many different ligands in the same target. To standardized and automate the process, we have developed a Kepler composite actor that

simplifies the preparation of MD simulations of ligand-protein complexes (Figure 1.3). This actor takes the outputs generated from the ligand parameterization actor as the inputs. Furthermore, it requires a receptor and a ligand file in order to start the workflow. Once started, the job will run through three major components, described below, that collectively prepare and run an MD simulation of the user's system.

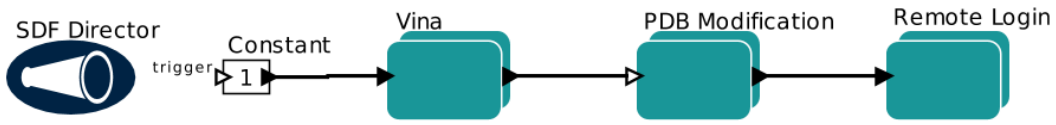


Figure 1.3 Layout of the Receptor-ligand molecular dynamic simulations actor.

Component I – Vina: Given PDB files of a ligand and a receptor, this module prepares the prerequisite files and docks the ligand into the receptor using Autodock VINA. The result is a PDB file that describes the "docked pose" of the ligand, or the conformation of the ligand when bound to the receptor.

Component II – PDB Modification: By concatenating the docked-pose PDB file to the PDB file of the receptor, component II first creates a merged ligand-receptor complex. Next, the receptor-ligand complex is assigned Amber force field parameters, and the topology and coordinate files required for MD are generated. Prior to simulating system dynamics, a restrained minimization is typically carried out to remove steric conflicts, which can cause MD

programs to crash. In a final step, component II prepares the restraint files required during minimization.

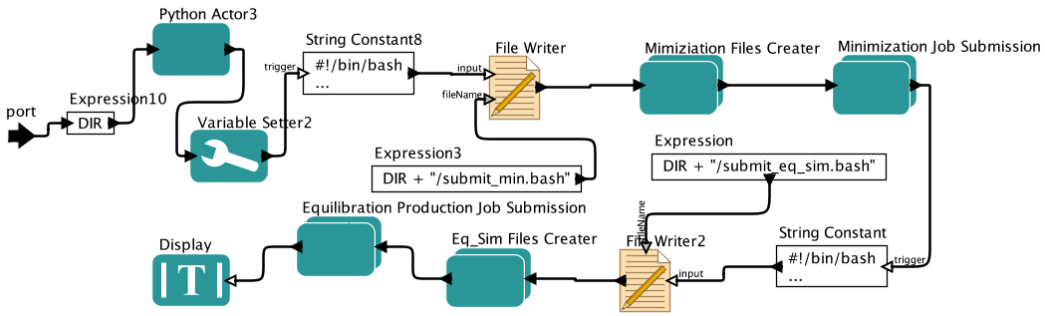


Figure 1.4 Breakdown of the remote login composite actor of the receptor-ligand dynamic simulation actor.

Component III – Remote Login: This module of the composite actor prepares configuration files for MD simulation with NAMD (10) and writes submission scripts for running minimization, equilibration and production jobs on the XSEDE resource Stampede, located at the Texas Advanced Computing Center (**Error! Reference source not found.**). Future developments will enable users to employ alternate HPC resources. In order to take advantage of parallel computing, the files required for MD simulation that were generated in earlier steps must be moved to the HPC platform. Component III performs this operation, moving the prerequisite files to a user specified directory on a remote HPC resource. Once the files are transferred, component III initiates minimization jobs on the HPC resource, generates the files necessary for a restrained MD equilibration, performs the restrained MD equilibration, and finally initiates a production MD simulation.

1.3.4 Receptor structural clustering

1.3.4 Receptor structural clustering

A MD simulation yields a “trajectory,” or a set of coordinates that represent the conformational states of the protein with or without a bound ligand as it evolves through time. With modern HPC resources, these

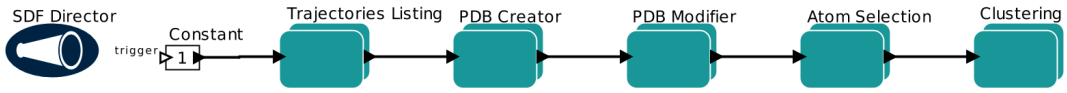


Figure 1.5 Gromos receptor structural clustering actor.

trajectories can consist of thousands or even millions of conformations, which translates into giga- or terabytes of data, making structural analysis challenging. Fortunately, meaningful dataset reduction methods have been devised that extract representative conformations, or structures. These structures, which are generally different than the crystal structure, are often referred to as cryptic binding pockets (11-13), and can be exploited in subsequent VS.

Considering the size of contemporary MD datasets, an effective, integrated platform for studying protein dynamics will require workflow actors that leverage data reduction software in a single, cohesive, user-friendly framework. To that end, we developed a modular set of actors that process MD trajectories by GROMOS cluster analysis (14, 15), a method that categorizes protein conformations based on structural similarity (Figure 1.5). In the first processing step, the trajectory listing composite actor utilizes cpptraj,

implemented in AmberTools, to convert the input trajectory file(s) to the PDB format required by Gromacs (16). It also strips solvent molecules and corrects for periodic boundary conditions, and additionally, removes translational and rotational degrees of freedom by aligning each trajectory conformation to a common reference specified in an atom selection file, provided by the user. The output is then sent to the public NBCR opal server, which clusters the data using the Gromacs. In addition to the GROMOS clustering actor, we have created another web-services-based data reduction actor that performs QR-factorization (11, 17) and can also be performed using the NBCR web services (see section 1.4).

1.3.5 Receptor and ligand preparation for docking

Docking programs, such as the widely used AutoDock (18) and AutoDock Vina (Vina) (19), provide scientists an estimate of the free energy change that occurs when a ligand binds to a receptor. Both AutoDock and Vina require PDBQT files that describe the coordinates, atomic partial charges, and AutoDock atom types of the ligand and the receptor. To streamline the conversion procedure, we have developed an actor that converts a receptor PDB file PDBQT file, which can be used by both AutoDock and Vina. The actor uses the publicly available NBCR opal server to perform the conversion, while Kepler monitors job scheduling and returns the output PDBQT file to the user's local machine. In the future, this actor will be extended to convert PDB to PDBQT for the ligand files as well.

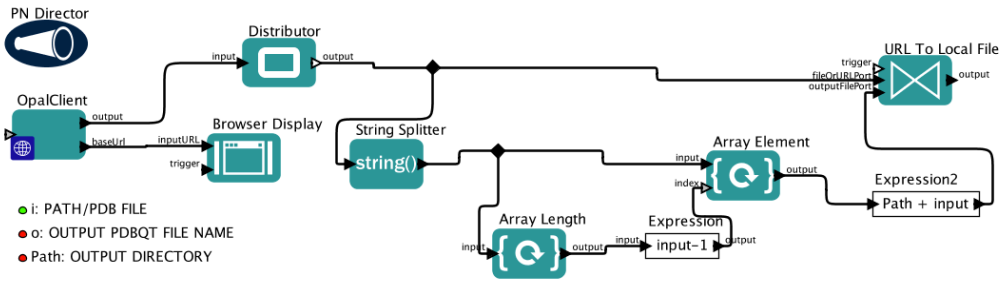


Figure 1.6 Receptor preparation for VS actor.

1.3.6 Ensemble based virtual screening

As previously stated, proteins are dynamic, and static crystal structures offer a poor account of protein flexibility, particularly when it is pronounced. In a drug discovery context, this flexibility is manifest in the observance of so-called cryptic binding pockets (11-13), or ligand binding sites that are absent in a crystal structure but are present during an MD simulation. To incorporate these potential binding sites during VS, it is important to include an ensemble of protein receptor structures that models the flexibility of a receptor in solution. Here, we describe an actor that screens large ligand sets against different receptor conformations using Vina (19) (Figure 1.7). Users supply a directory of receptor PDB files, a directory of ligand PDB files and grid information. Receptor PDB and ligand PDB files are converted to Vina specific PDBQT files. Every ligand is matched with each receptor once in the “Mix&Match” module, which organizes the large number of files generated in this protocol. The combinations are sent to Vina one by one for VS.

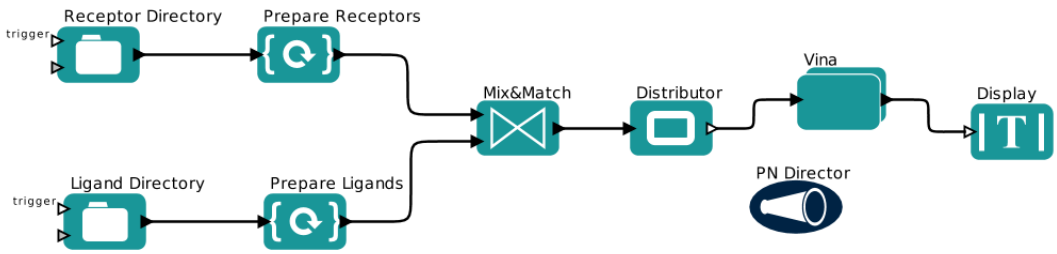


Figure 1.7 Virtual screening actor.

1.3.7 Virtual screening performance statistics

During VS, small molecules are assigned a score, and based on that score, they are classified as either binders or non-binders. For example, during small molecule docking, binding affinity is predicted, and those compounds predicted to bind more favorably receive a higher rank and are more likely to be experimentally assayed.

Performing VS using an ensemble of protein conformations may benefit the discovery effort, but it is also computationally demanding and scales linearly with the number of conformations. To improve computational efficiency, statistical methods can be used to select the ensemble that does the best job of separating known the binders from the known non-binders in a small, experimentally characterized compound database. By carefully selecting the best performing ensemble, this protocol has the potential to reduce the computational expense of screening a much larger database of uncharacterized compounds.

We have developed an actor (see Figure 1.8) that incorporates the experimental status of a compound, *i.e.* binder or non-binder, the docking score of the compound into each receptor ensemble member, and returns the ensemble best able to discriminate known binders from known non-binders. Although there are various VS performance metrics available in the literature (20-22), the area under the curve (AUC) of the Receiver Operating Characteristic (ROC) plot (23) is one of the most popular performance evaluation metrics and is used for our workflow. Part of the AUC's appeal is how easily it is interpreted. It represents the probability that a randomly selected binder will have a higher rank than a randomly selected non-binder (24, 25). Consistent with this interpretation, an AUC value of 0.5 indicates the VS protocol performs randomly, while a value of 1 indicates the protocol ranks all of the binders ahead of all of the non-binders.

In practice, ensemble selection is complicated by the need to evaluate all possible combinations of receptor conformations, a combinatorial process described by the binomial coefficient. The workflow utilizes a series of Matlab (26) scripts to monitor performance of all possible ensembles of conformations. The scripts require an input matrix "total", which is supplied by the user in a comma-separated CSV file format. The first column of total gives ligand identification numbers, compound IDs in a database, for example. The second column is a compound classifier, a 0 or 1, which labels non-binders and binders, respectively. The remaining columns contain the docking scores

for each receptor conformation. After receiving the “total” matrix, the workflow returns the AUC value for all possible ensembles of receptor conformations, as well as the 95% confidence intervals, and p-values, which provide indications of the performance reliability and the statistical significance of the performance of each ensemble.

The calculations in Matlab are designed to utilize the Parallel Computing Toolbox in Matlab (parfor loops), although if the separate license required to use the toolbox is not available, the behavior will default to standard loop iteration. The parallel option is highly recommended particularly for a large number of receptor structures, as these calculations otherwise become very time consuming.

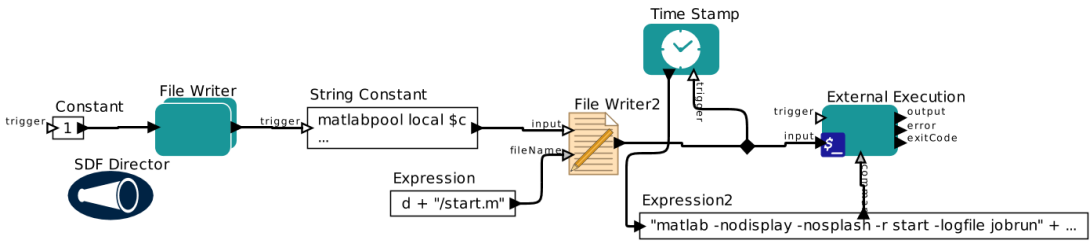


Figure 1.8 VS statistical performance actor utilizing Matlab

1.4 Integrated web-services

The complexity of scientific applications needed in CADD often requires an access to HPC resources. To ensure tasks are completed expediently, scalable and transparent support of distributed computing resources available on both HPC platforms and in the cloud is required of each Kepler workflow

module. To meet this requirement, we use the Opal toolkit (27), which provides Scientific Software as a Service (SaaS) using standard and simple web interfaces. For example, scientific applications executed by the workflows are wrapped as SOAP-based web services that allow for programmatic and web-based application access, which is useful for a wide variety of applications. The programmatic capability allows transparent access of different workflow components, while the web-based service access provides a large number of NBCR applications to our affiliates and collaborators.

Using integrated web-services for scientific applications also aids our objective to develop a modular environment of interchangeable, customizable modules that can be used to create complex scientific workflows. As SaaS providers, we handle software installation configuration and upgrade transparently at the cyber-infrastructure level. With infrastructure complexities replaced by an easy-to-use interface, the full power of the modular workflow environment can be easily applied to pressing scientific problems.

The scientific applications, wrapped as Opal web services (28), can readily be deployed across distributed computing environments to accelerate completion of the scalable computations within the CADD framework. It is easy to access the scientific applications through the Opal web server, which provides a stable, reliable infrastructure for CADD and molecular simulations that can accommodate large throughput in an extensible, reproducible and reusable manner. This approach will allow flexible community resource

sharing and, by providing the framework to incorporate ideas from a broad community of users, it will promote convergence toward a set of standardized best practices.

1.5. Workflow dissemination

CADD workflows, in addition to other NBCR workflow products, are being made available though the NBCR website and GitHub (29). We have enabled the NBCR workflows site to be searched and filtered easily though keywords describing the workflows' application, actors, program dependency, and other relevant terms, enabling the user to select the appropriate workflow for their needs. Upon selecting a desired workflow the user is taken to the workflow documentation and download options. The workflows will be distributed through GitHub to provide transparent version control. The user may either download the workflow itself, requiring a local installation of Kepler and dependent programs, or download the workflow as part of a Rock's Rolls (30) containing dependent programs. The Rock's Rolls facilitate the utilization of workflows in HPC environments. Additionally, we are developing domain specific interfaces for all NBCR workflows. These interfaces will integrate key visualization software, workflow modification, workflow execution management, and electronic lab book functions further optimizing the CADD process.

1.6 Conclusions

We have developed a series of modular actors that can be integrated into a larger CADD framework, or be used as stand-alone tools. The modules described here are have successfully been deployed on a number of different projects and are being optimized based on user feedback. These modules demonstrate the usability of Kepler scientific workflows in CADD with the aim to standardize simulation and analysis, and to promote best practices within the molecular simulation and CADD communities. The workflows demonstrate usability in terms of file-management tasks, molecular simulation including ligand force field parameterization and management of job submission and monitoring on relevant HPC resources, as well as VS elements such as receptor structural clustering, docking and statistical analyses of the VS results. The models are available for download on the NBCR website and have been integrated with NBCR web-services. Our lab is currently developing novel Kepler workflows designed for automation and standardizing of common tasks in CADD and molecular simulation. We will solicit user feedback and use it to guide our efforts, to strengthening an ecosystem that encourages development and distribution of workflows with the simulation and CADD communities.

This chaper, in full, is a reprint of the material as it appears in “Progress towards automated Kepler scientific workflows for computer-aided drug discovery and molecular simulations” by Ieong, Pek U., Sorensen,

Jesper, Vemu, Prasantha L., Wong, Celia W., Demir Özlem, Williams, Nadya, P., Wang Jianwu, Crawl, Daniel, Swift, Rob V., Malmstrom, Robert. D., Altintas, Ilkay, Amaro, R. E., published 2014 in Procedia Computer Science. This chapter is included with the permission from Sorensen, Jesper, Vemu, Prasantha L., Wong, Celia W., Demir Özlem, Williams, Nadya, P., Wang Jianwu, Crawl, Daniel, Swift, Rob V., Malmstrom, Robert. D., Altintas, Ilkay, and Amaro, R. E.

Chapter 2

Molecular dynamics analysis of antibody recognition and escape by human H1N1 influenza hemagglutinin

Abstract

The antibody immunoglobulin (Ig) 2D1 is effective against the 1918 hemagglutinin (HA) and also known to cross-neutralize the 2009 pandemic H1N1 influenza HA through a similar epitope. However, the detailed mechanism of neutralization remains unclear. We have conducted molecular dynamic (MD) simulations to study the interactions between Ig-2D1 and the HAs from the 1918 pandemic flu (A/South Carolina/1/1918, 18HA), the 2009 pandemic flu (A/California/04/2009, 09HA), a 2009 pandemic flu mutant (A/California/04/2009, 09HA_mut), and the 2006 seasonal flu (A/Solomon Islands/3/2006, 06HA). MM-PBSA analyses suggest the approximate free energy of binding (ΔG) between Ig-2D1 and 18HA is -74.4 kcal/mol. In comparison with 18 HA, 09HA and 06HA bind Ig-2D1 about 6 kcal/mol (ΔΔG) weaker, and the 09HA_mut bind Ig-2D1 only half as strong. We also analyzed the contributions of individual epitope residues using the free energy decomposition method. Two important salt bridges are found between the HAs and Ig-2D1. In 09HA, a serine to asparagine mutation coincided with a salt bridge destabilization, hydrogen bond losses and a water pocket formation

between 09HA and Ig-2D1. In 09HA_mut, a lysine to glutamic acid mutation leads to the loss of both salt bridges and destabilizes interactions with Ig-2D1. Even though 06HA has a similar ΔG to 09HA, it is not recognized by Ig-2D1 *in vivo*. Since 06HA contains two potential glycosylation sites that could mask the epitope, our results suggest that Ig-2D1 may be active against 06HA only in the absence of glycosylation. Overall, our simulation results are in good agreement with observations from biological experiments and offer novel mechanistic insights into the immune escape of the influenza virus.

2.1 Introduction

Influenza virus gains entry into the human body through interactions of the viral surface glycoproteins called hemagglutinin (HA) with the sialic acid (Sia) receptors on the human epithelial cell surface (31, 32). Sia is found at the terminals of glycans attached covalently to cell surface glycoproteins or glycolipids. They are also found on the viral surface proteins (33). During viral infection, viral HA binds to Sia receptors on human host cells, and the virus enters through endocytosis. The flu virus then usurps host cell machineries for viral replication (34). There are 18 known HA serotypes: H1 to H18. Within the 18 serotypes, H1 and H5 are more extensively studied. H1 is found to bind preferentially to Sia with an α-2,6 glycosidic bond, whereas H5 prefers Sia with a-2,3 linkage (35).

Humans fight influenza infection through innate and adaptive immune

responses (36) including vaccination or by using pharmaceutical drugs such as *Tamiflu* or *Relenza* (37). The adaptive immune response involves the recognition of HA epitopes by human immune cells, and the production of antibodies against HA. Inactivated, or live attenuated virus, or recombinant HA is often prepared as vaccines, which elicits antibody production seven days after inoculation (38). Antibodies bind HA epitopes, preventing sialic acid binding and endocytosis (39). Four main canonical epitopes on the globular HA head have been identified: Sa, Ca, Sb and Cb (Figure 2.1) (40, 41). More recently, cross-reacting antibodies against multiple HA subtypes have been discovered that target the globular epitopes as well as the conserved stem regions (42-44). Pre-existing antibodies from vaccination or earlier infections may prevent infection by viral strains with similar HA epitopes (38, 45).

Influenza viruses escape from the human immune responses through both antigenic drift and antigenic shift. In antigenic drift, mutations in glycoprotein epitopes render existing antibodies ineffective, a process that is facilitated by the high mutation rate of the influenza RNA genome (46). Thus, the annual vaccines may offer partial protections or fail completely against unanticipated strains. In antigenic shift, abrupt changes in viral RNA genome result when several different viral strains recombine, creating a hybrid virus. The resulting virus is novel and distinctive, sometimes posing lethal threats to the human population. The recent 2009 swine flu, which emerged from a triple

assortment involving swine, human and avian reservoirs, is a good example (47). Since it first appeared in the human population in April 2009, the swine flu quickly spread globally and was declared pandemic by WHO in June 2009 (48).

Although the elderly are particularly susceptible to the seasonal flu, few from this age group have been infected by this pandemic strain (49). Some researchers have hypothesized that they may be immune because of childhood exposure to the 1918 influenza pandemic virus. The 09HA is found to be genetically and structurally very similar to the 18HA; therefore, it is possible that antibodies that recognize the 18HA may also recognize the 09HA (42, 50). Krause et al. showed that monoclonal immunoglobulin (Ig-2D1) against the 18HA appeared to cross- react with the 09HA from the pandemic flu (44).

To identify specific mutations that might affect HA binding, Liu et al. used computational methods to predict hot spots residues on the epitopes of the 09HA, 18HA and 07HA from a 2007 seasonal strain (07HA) that interact with Ig-2D1 (51). They suggested that mutations in the 18HA and 09HA at residues P128, N129, K158, P163, K164 and K167 (using 18HA numbering) could disable Ig-2D1 neutralization. Because N160 was not predicted as a hot-spot residue, Liu et al. proposed that a mutation from S160 in the 18HA to N160 in the 09HA would not affect binding. Their analysis of the binding

between Ig-2D1 and the various HAs was performed on crystal structures and protein flexibility was not explicitly considered.

In this manuscript, we present new findings using the molecular dynamics (MD) simulation technique in four influenza H1N1 systems of Ig-2D1 and HA's: A/South Carolina/1/1918, A/California/04/2009, A/Solomon Island/3/2006 and A/California/04/2009 mutant. We aim to explore the underlying molecular interactions that govern Ig-2D1/HA binding in order to determine how Ig-2D1 is able to elicit a cross-reactive immune response to the 2009 influenza virus. Our results are in good agreement with previous experimental and computational studies; additionally, we have discovered that mutations such as the S160N mutation in 09HA do affect the stability of the antibody-antigen interaction.

2.2 Methods

2.2.1 Simulation Setup

The structures of the 06HA, 09HA, and 18HA were obtained from the Protein Data Bank (PDB) (42, 52) with PDB ID 3SM5, 3LZG, and 3LZF, respectively. Only the 18HA was co- crystalized with Ig-2D1 (42). Seasonal strain HA and 09HA were superimposed on the 18HA structure to model Ig-2D1 binding to these HA variants. The numbering scheme for Ig-2D1 was

adapted from the PDB structure. The trimetric units of HA were built after the superimposition according to the biological unit (53). The 09HA_mut and 09HA model differ by one single residue at position 167. It was prepared from the 09HA system by mutating the residue from K to E (K167E) using Schrödinger (54).

2.2.2 Molecular Dynamic Simulation

Four systems (18HA, 09HA, 06HA and 09HA_mut) were parameterized with the Amber ff99SB force field (55). The systems were neutralized by first adding sodium ions. Additional ions were then added to achieve 20 mM NaCl buffer salt concentration. Histidine charges were assigned using PROPKA from the pdb2pqr web server at pH 7.0 (56, 57). Each system was solvated in a water box of approximately 150×160×210 Å using the TIP3P (58) water model. A total of about 500,000 atoms were in each system (Table S1.1). Molecular dynamics simulations were performed afterward using NAMD 2.9 (59).

The systems were constrained and gradually minimized to reduce the total potential energy in a series of four energy minimizations. The first step of minimization kept all heavy atoms constrained and only hydrogen atoms were allowed to fluctuate. The second step released the constraints on water and ions. The third steps freed the side chains and in the forth step, all atoms were allowed to move without restriction. The non-bonded energy was calculated

every 2 time steps with a cutoff distance of 12 Å. A switching function is applied at 10 Å to abridge the van der Waals potential function. Following minimization, four steps of equilibration were performed, gradually loosening harmonic constraints in 500 ps increments, for a total equilibration time of 2 ns. The first step heated the system up to 310 K while applying a force of 4 kcal/mol to hold the backbone in place. The second steps to forth steps gradually lifted the backbone constraint force from 4 kcal/mol to 1kcal/mol. NPT ensemble was completed. Langevin Dynamic was applied to keep the temperature constant throughout the equilibration, with a damping frequency of 5 picoseconds/terahertz and Langevin Priston barostat helped to maintain the specified one atmospheric pressure. The constraints applied during equilibration were removed for the free simulation of the antigen-antibody complexes. All simulations were run for 69 ns using the XSEDE resources Ranger, Stampede (TACC) and Gordon (SDSC).

2.2.3 Determining Bond Interactions

Barlow and Thornton proposed that salt bridges should be between opposite charge residues ≤4.0 Å (60). Xu et al. reported that three salt bridges were found between the antigen- antibody interfaces in their Ig-2D1 and 18HA co-crystal structure (42). K158 interacts with Ig heavy chain (IgH) D52 and D54, forming two distinct salt bridges, with bond distances of 3.7 Å and 3.0 Å, respectively. Additionally, K167 interacts with D93 on the light chain (IgL) with

a bond distance of 2.8 Å. For the purpose of this investigation, we have defined a salt bridge to be between a pair of oppositely charged group that contain at least one hydrogen bond within 3.5Å of each other, as suggested by other researchers (61). This is because H-bond is important in the stability of salt bridges (61), and further demonstrated later in this manuscript.

The Visual Molecular Dynamics (VMD) software package (62) was used to analyze simulation trajectories. Xu et al. proposed a list of 18HA residues that interacted with Ig-2D1 (42). All the reported interactions between with Ig-2D1 and the different HAs were carefully determined based on their atomic characteristics and a distance matrix. These included hydrogen bonds, salt bridges, dipole-dipole and van der Waals interactions (e.g, Figure S2.2). The distances between contacting atoms were recorded every 100 ps. The distance cutoffs are ≤3.5 Å for hydrogen bond, ≤3.5 Å for salt bridge, 2.6 - 4.6 Å for dipole-dipole interactions and for van Del Waal interactions. To account for the system dynamics and capture the stability of the interactions, an interaction percentage was calculated using the distance matrix for each contact made between the epitope residues and Ig-2D1. Only interactions present in at least two of the three monomers for 75% or more of the simulation time were considered as important for the binding of HAs and Ig-2D1. Similarly, surrounding residues near the reported Sa epitope were also analyzed to identify possible new contacts with Ig-2D1 found only through

simulation.

2.2.4 Free Energy Binding and Decomposition

The free energy of binding (ΔG) was approximated using the Molecular Mechanics - Poisson-Boltzmann Surface Area (MM-PBSA) method using MMPBSA.py implemented in AmberTools 11 (63, 64). The Poisson-Boltzmann (PB) equation was utilized to estimate the polar contribution of the solvation energy. MMPBSA.py stripped the water molecules and ions and carried out the calculation in implicit solvent. The ionic strength for the free energy calculation was also set to 20 mM. Receptor mask was set to the trimetric HA plus two Ig-2D1 and ligand mask was always set to the remaining Ig-2D1. Thus, three free-energy calculations were performed to obtain the mean and standard error (SE) of ΔG for each system.

Per residue free energy decomposition was carried out to determine the energy contribution of epitope residues to the binding with Ig-2D1, using the MMPBSA.py tool from AmberTools 13 (63, 64). The Generalized Born (GB) implicit solvent model was utilized for the decomposition calculation. Since GB was parameterized with the atomic radii mbondi2 (65), all the atom radii were changed from the default mbondi to mbondi2. Saltcon was also set to 20 mM. All other parameters were identical to those in AmberTools11. The decomposition analysis of individual residues was performed using igb2, not the default GB parameter (igb5). In a comparison of igb5 and igb2, we found

that the ΔG calculated from igb2 was closer to the PB energy (data not shown), this is consistent with our previous study that igb2 is well suited to the neuraminidase N1 and N9 systems (66).

2.2.5 RMSD and Surface Pocket Volume Calculations

UCSF Chimera was utilized to calculate the RMSD between crystal structures. Clustal Omega was applied to alignment the HA sequences (53, 67). The POVME 2.0 software (68) was used to calculate the surface pocket volumes at the interface of HA's and Ig-2D1 near residue S/N160 in 18HA and 09HA. A 12Å radius sphere was centered on the S/N160 center of mass fully covering S/N160 and the surrounding Ig-2D1 residues. The volume calculations were done separately for each monomer and an average was reported for each system. A cylinder (25 Å in radius and 4 Å tall) that fully encompassed the interface was used to calculate the volume between the antibody-antigen interface in the 09HA and 09HA_mut systems. The center of the cylinder was positioned at the center of mass of all epitope residues, and oriented towards (-1, 0, 4) for Ig1, (4, 1, 0) for Ig2 and (-1, 4, 0) for Ig3.

2.3 Results

2.3.1 Structure and Sequence Alignments

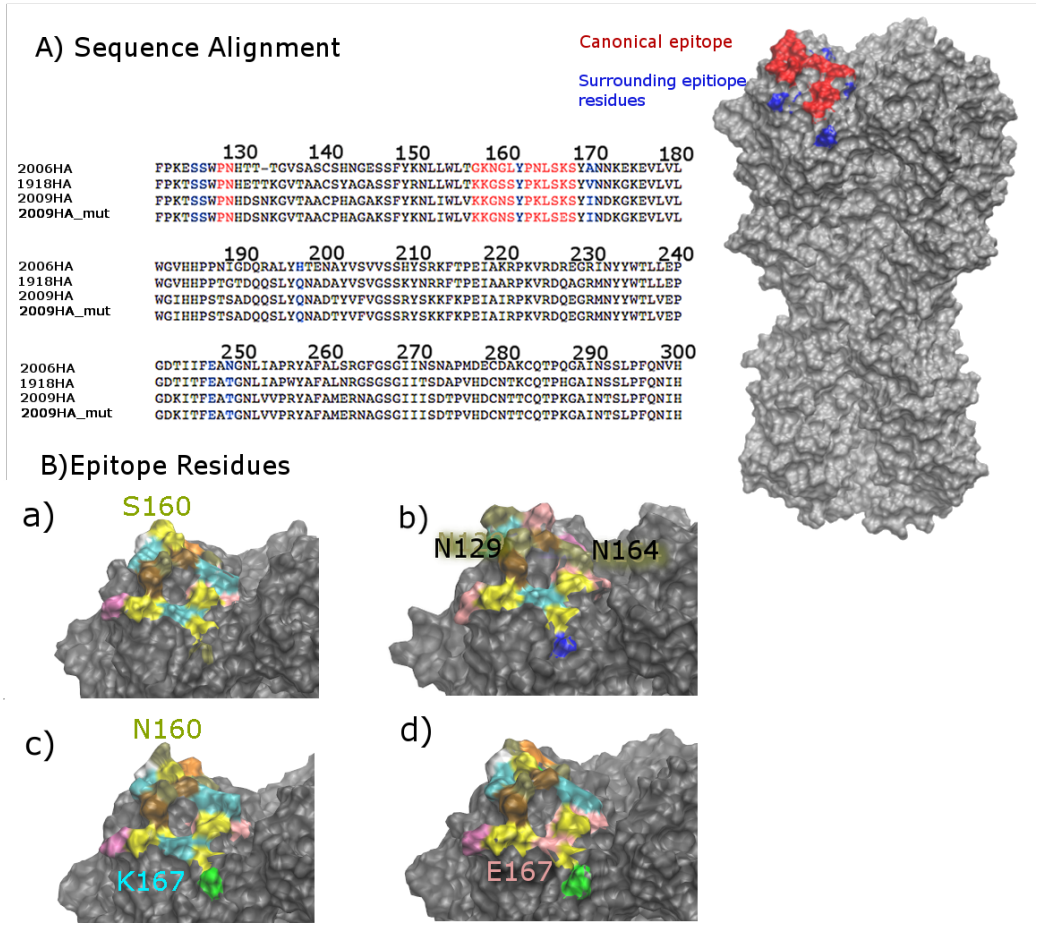


Figure 2.1 Sequence alignment and the epitopes of the four HA glycoprotein. A. Sequence alignment and structural view. The four HA sequences are aligned and numbered using the 18HA numbering convention. On the right, the Sa epitope is colored red and the surrounding residues that form contact with Ig-2D1 (42) are colored blue (top left) in 18HA monomer 1. B. Structural conservation of the HA epitope region and key mutations that affect antibody recognition. The four HA's are shown: a) 18HA, b) 06HA, c) 09HA, and d) 09HA_mut. The epitope residues on monomer 1 are colored by residue names. Several key residues are also labeled to their corresponding residue colors. S160 (18HA) is mutated in N160 in 09HA. K 167 (09HA) is mutated to E167 in 09HA_mut. N129 and N164 in 06HA are potential glycosylation sites.

The 18HA, 09HA, and 06HA selected in this study are structurally

conserved (42), with RMSD values of 0.851Å (09HA) and 0.927 Å (06HA), in comparison to the 18HA (Figure 2.1). Of the twenty HA epitope residues known to interact with Ig-2D1 in the 18HA, two mutations were found in the 09HA and ten mutations in the 06HA. As reported earlier, 06HA contained two new glycosylation sites N129 and N164, (NxS/T, where x is any amino acid other than proline), due to an E131T mutation and a K164N mutation (Figure 2.1) (52).

2.3.2 Bond Interactions

In Table 2.1, we compared the number of bond interactions between HA epitope residues and Ig-2D1. The heavy and light chains of Ig-2D1 are indicated using IgH and IgL respectively. Notably, from the single point mutation from the 09HA to the 09HA_mut, three H-bonds between S126, K167 and S168 (09HA) and D93, N31, and S30 (IgL) were lost respectively, whereas a new H-bond interaction was formed between Y162 (09HA_mut) and S99 (IgH). This is a net loss of two H-bonds between the 09HA and 09HA_mut system.

In the 18HA, K158 and K167 form two distinct salt bridges with D54 (IgH) and D93 (IgL) in the Ig-2D1. These two lysine residues are conserved in all naturally occurring HA's (Figure 2.1).

Table 2.1 Interactions between HA and Ig-2D1 systems categorized by bond types. H-bond stands for hydrogen bond. Dipole-dipole is dipole-dipole interactions. Salt bridge is formed between negatively and positively charged residues within 3.5Å of each other. The van der Waals force described here is the interaction between hydrophobic residues. More details are in Table S2.3-S2.6. Table S3 also indicates interactions found in crystal structure only for 18HA.

	Salt Bridge	H-bond	Dipole-dipole	van der Waals
18HA	2	11	23	4
06HA	1	6	11	4
09HA	1	6	11	5
09HA_mut	0	4	11	5

In contract, the 09HA system has only one salt bridge (K167-D93). An earlier study by Krause et al. suggested that mutations of K167 in the 09HA to either E or N (K167E/N) allow the 09HA to escape neutralization from monoclonal antibody Ig-2D1 (44). We have recreated the 09HA_mut carrying the K167E mutation *in silico*, and the resulting mutant lost both salt bridges. Overall, the Ig-2D1 lost about half of the H-bond and Dipole-Dipole interactions in the 09HA and 09HA_mut systems compared to the original 18HA system. Together, these observations suggest that the Ig-2D1 may not bind as strongly to the newer HA's due to loss of these interactions.

2.3.3 Free Energy of Binding Calculation

The approximate average ΔG for each system was obtained using the MM-PBSA method (63). The 18HA had the lowest ΔG with respect to Ig-2D1 (Table 2) at -74.4 ± 1.1 kcal/mol. The 09HA and the 06HA systems were about

6 kcal/mol lower (ΔΔG), with similar ΔG values of -68.0 ± 1.2 kcal/mol and -67.6 ± 1.6 kcal/mol, respectively (Table 2.2). The calculated ΔG values are consistent with the observed number of interaction shown in Table 2.1. Both 09HA and 06HA had higher ΔDG's, compared to the 18HA by +6.4 ± 1.6 kcal/mol and +6.8 ± 1.9 kcal/mol, respectively. In contrast, the average ΔG of the 09HA_mut was significantly higher, +36.5 ± 1.9 kcal/mol than that of 18HA. Compared with that of 09HA, the ΔG of the 09HA_mut system increased by +30.1 ± 2.0 kcal/mol, even though the only difference is a K167E mutation in 09HA_mut. The 09HA_mut DG result is in agreement with the K167E escape mutant selected by Krause et al (44). The MM-GBSA method also gave very similar results (Table 2.2).

Table 2.2 Estimated DG free energy of binding for each system using MM-PB/GBSA. N is the number of frames used in the calculation. Each frame is ~0.28 ns of the simulations. Average DG energy and standard error (SE) are calculated from three sample runs in each system. All the values are in kcal/mol. Only ΔGsubtotal is reported. The individual components are reported in Table S7 and S8.

	18HA (N=248)		06HA (N=248)		09HA (N=249)		09HA_Mut (N=247)	
	Average	SE	Average	SE	Average	SE	Average	SE
PB $\Delta G_{subtotal}$	-74.4	1.1	-67.6	1.6	-68.0	1.2	-37.9	1.6
GB $\Delta G_{subtotal}$	-74.5	1.0	-63.4	1.5	-62.7	1.1	-39.1	1.1

2.3.4 Free Energy Decomposition

Table 2.3 Free energy decomposition of the epitope residues in the four systems. Only selected key residues 158, 159, 160, 161 and 167 are shown below. All epitope residue based decomposition results are in the supporting materials (Table S2.9). All units are in kcal/mol.

	18HA		06HA		09HA		09HA_mut	
Res ID	Residue	ΔG±SE	Residue	ΔG±SE	Residue	ΔG±SE	Residue	ΔG±SE
158	K	-4.4 ±0.2	-	-2.2 ±0.2	-	-2.3 ±0.2	-	-1.9 ±0.2
159	G	-3.1 ±0.1	N	-2.9 ±0.1	-	-2.2 ±0.1	-	-1.8 ±0.1
160	S	-4.9 ±0.2	G	-0.0 ±0.0	N	-0.5 ±0.1	N	-0.9 ±0.1
161	S	-1.8 ±0.1	L	-4.0 ±0.1	-	-1.8 ±0.1	-	-2.0 ±0.1
167	K	-6.6 ±0.2	-	-6.6 ±0.2	-	-6.3 ±0.2	E	2.9 ±0.1

To further study the contribution of each epitope residues and probe the importance of K167, we performed free energy decomposition using MM-GBSA (Table 2.3, S2.9). The average ΔG of binding energy contribution from K167 was predicted to be -6.6 +/- 0.2 kcal/mol in 18HA, the highest contribution of all the epitope residues. This is also true in 09HA and 06HA systems, with the ΔG of binding energy contributions of K167 determined to be -6.3 +/- 0.2 kcal/mol and -6.6 +/- 0.2 kcal/mol, respectively. Krause et al. identified an escape mutation K167E that prevents Ig-2D1 from neutralizing 09HA (44). In the K167E 09HA_mut system, free energy decomposition results showed that this mutation was highly unfavorable, with a ΔG contribution of +2.9 +/- 0.1 kcal/mol (Table 2.3). The increased the residue

decomposition energy by +9.2 +/- 0.3 kcal/mol. The placement of a negatively charged glutamic acid residue in place a positively charged lysine residue would lead to electrostatic repulsion with structural consequences on the HA-Ig2D1 complex.

Mechanistic Insight from the 09HA Immune Escape Mutation

During the minimization step of the 09HA_mut system, the complementarity determining region (CDR) L3 loop containing D93 (IgL), the salt bridge partner with K167 (09HA), shifted away from E167 (09HA_mut) (Figure 2.2). E167 (09HA_mut) then formed two new hydrogen bonds interacting with S30 and N31 (IgL) in the CDR L1 loop. Initially, the CDR L1 loop maintained these hydrogen bonds (Figure 2.2A). Both L1 and L3 loops had shifted away from E167 and the antigen at the end of the simulation (Figure 2.2B). The number of atoms within 4.6 Å of E167 from the CDR L1 and L3 loops dropped from 13 atoms to around 4 atoms as early as 3 ns into the simulation (Figure 2.2C). The movement of the L1 and L3 loops did not affect the interface volumes between Ig-2D1 and 09HA and 09HA_mut significantly, 5155.2 +/- 12.7 Å and 5382.0 +/- 15.7 Å, respectively. This suggests that key structural changes occurred on the antibody Ig-2D1 L1 and L3 loops.

Destabilization of K158 Salt Bridge in 09HA

The 09HA retained the K167-D93 salt bridge, but lost the K158-D54 salt bridge found between 18HA and Ig-2D1. The DG contribution of K158 was

estimated to be -4.4 +/- 0.2 kcal/mol in the 18HA (Table 2.3). The K158 salt bridge was unstable in the 09HA system and did not meet the 75% occupancy threshold. Of the epitope residues, only two mutations occurred

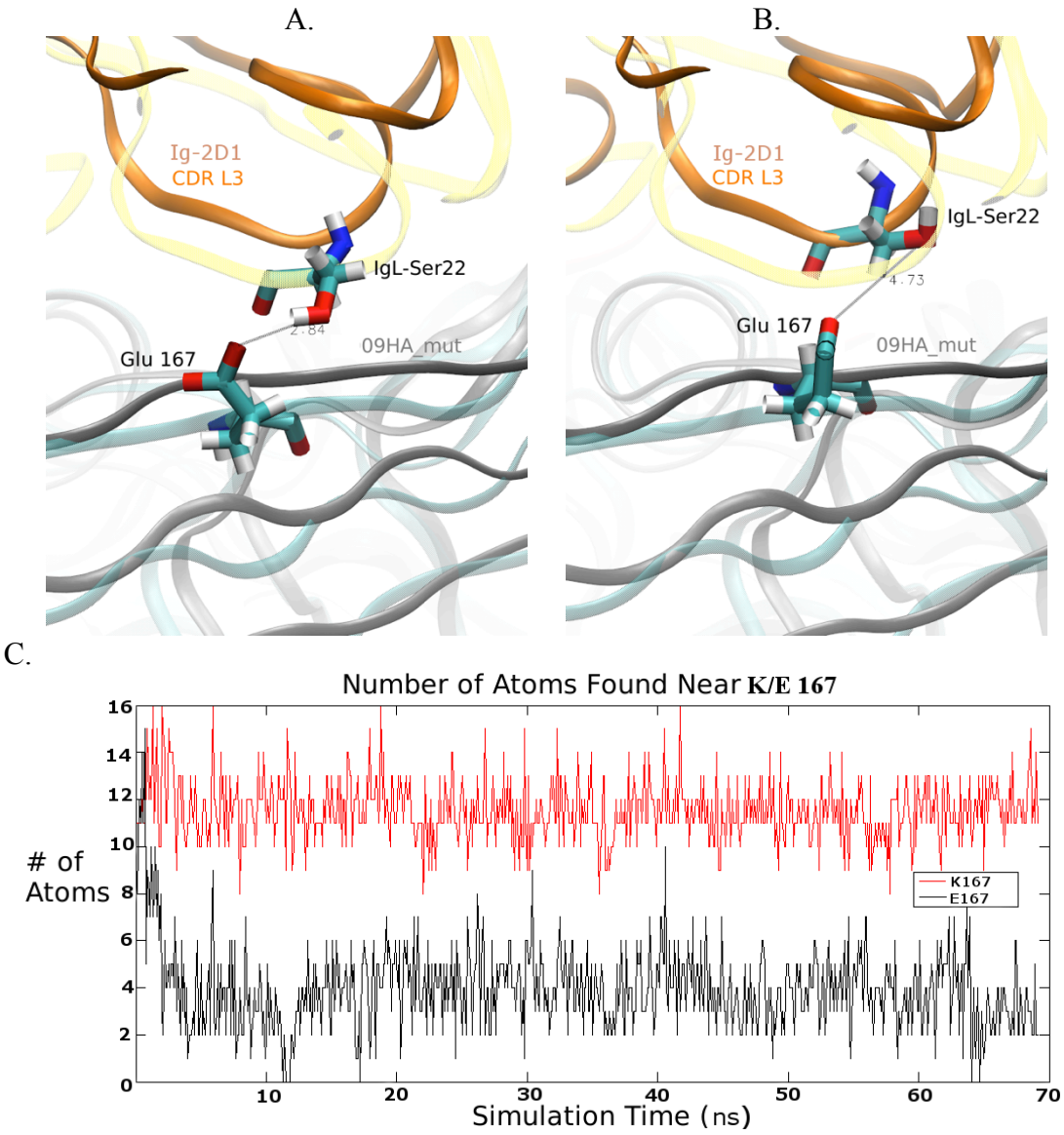


Figure 2.2 Loop motions near Glu 167 in 09HA_mut. A) The position of Glu (E) 167 relative to the CDR L3 loop (colored in yellow), at the beginning of the simulation. B) After the end of the simulation, the loops moved away from the HA and the CDR L3 (orange) was further away from the Glu167. C) The number of Ig-2D1 atoms (y-axis) within 4.6 Å of K167 (09HA, red) and E167 (09HA_mut, black) over time (x-axis).

between 18HA and 09HA, S160N and V170I, respectively. The ΔG of 09HA and Ig-2D1 interaction increased by +6.4 ± 1.6 kcal/mol as a result. The biggest free energy contribution difference between 18HA and 09HA occurred at K158 and N160, with no significant differences observed from residue 161 onwards (Figure 2.3). These include residues K167 and V170I mutation. Thus, the difference observed at S160N may be the major mutation that affects the K158-D54 salt bridge stability.

S160 contributed -4.9 ± 0.2 kcal/mol to 18HA interaction with Ig-2D1, whereas N160 contributed little to the 09HA interaction with the latter (Table 2.3). Throughout the simulation, S160 (18HA) formed two H-bond and two dipole-dipole interactions with Ig-2D1 (Table S2.3), whereas no equivalent interactions were found for N160 (09HA) (Table S2.4). Here we note that only one dipole-dipole interaction for S160 (18HA) is identified through crystal structure examination only, whereas all the interactions at K167 (18HA) are observed in the simulation and the crystal structure (Table S2.3)

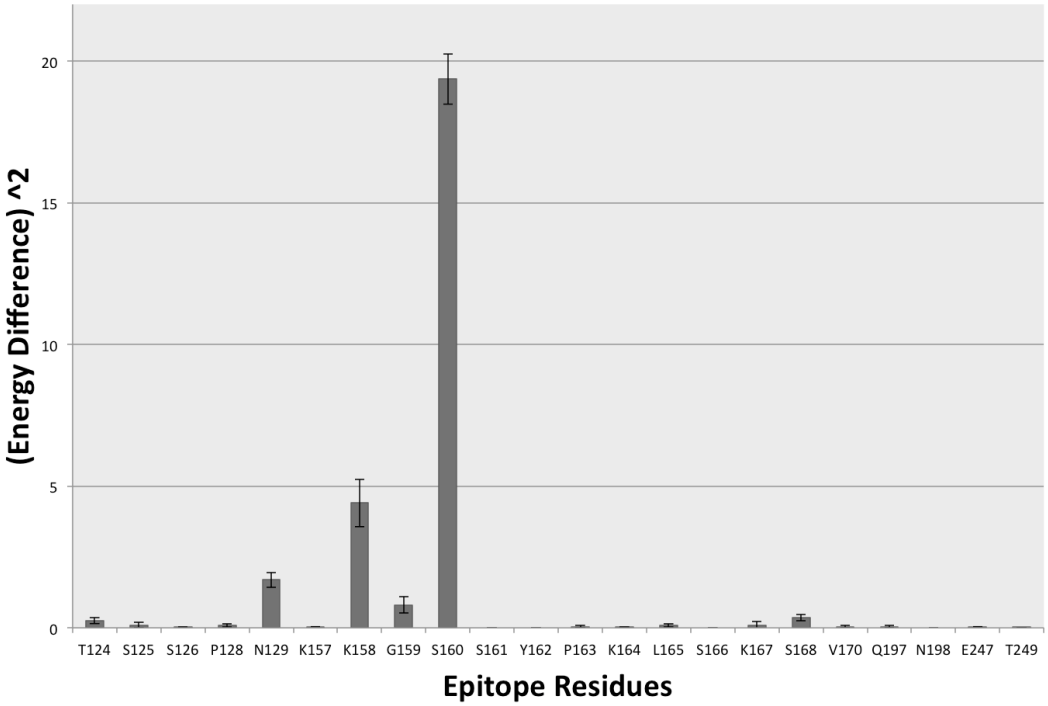


Figure 2.3 The free energy differences squared are shown for all epitope residues between 18HA and 09HA. Y-axis is the energy difference squared, $\Delta\Delta G^{\wedge 2}=(\Delta G_{18HA}-\Delta G_{09HA})^{\wedge 2}$ and x-axis are the residue name and residue ID from 18HA. The two mutations are S160N and V170I from 18HA to 09HA. Detailed data may be found in Table 3 and S9.

Table 2.4 POVME volumes of 18HA and 09HA surrounding S/N160.

	18HA	09HA
Volume (Å^3)	1317.2	1922.9
SE (Å^3)	13.6	17.7

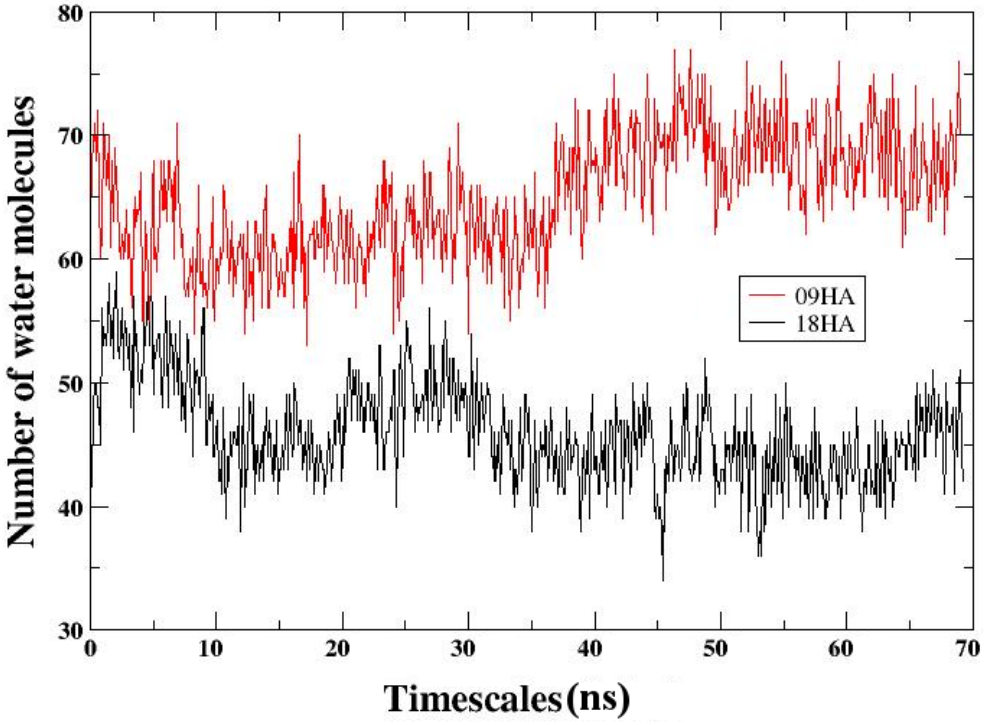


Figure 2.4 The S160N mutation and water pocket formation. The number of water molecules within 5 Å of 18HA S160 (black) and 09HA N160 (red) as a function of simulation time.

We examined the volume between S160 (18HA) and N160 (09HA) and adjacent Ig-2D1 residues (Table 2.4). There was a close to 50% increase in the solvent accessible volume from 18HA to 09HA. The extra volume allowed more water molecules to enter the protein interface (Figure 2.4). Thus, it is likely that the S160N mutation led to increased solvent accessibility and formation of water cavity, with a destabilizing effect on the K158-D54 salt bridge (Table S2.4).

Of the 20 epitope residues, the other mutation between 18 HA and 09HA is V170I. We did not find any significant interactions between HA and Ig-

2D1 at position 170 during the course of simulation (Table S2.3, S2.4). Overall, the S160N could be a major contributing factor to the lower binding affinity 09HA by Ig-2D1, in conjunction with the loss of the K158-D54 salt bridge.

In comparison, 06HA also lost the K158-D54 salt bridge with Ig-2D1, compared to 18HA. The K158 (06HA) contributed similarly to the K158 (09HA), but only half as much as the K158 (18HA) to the ΔG of Ig-2D1 binding. However, ten additional mutations in the epitopes of 06HA resulted in a similar ΔG as 09HA overall (Table 2.3, S2.9). Of these, L161 (06HA) lowered ΔG from the VDW interaction with R97 (IgH) (Figure S2.2). It contributed -4.0 kcal/mol +/- 0.1kca/mol, compared -1.8kcal +/- 0.1kcal/mol from S161 (18HA) (Table 2.3). This decrease of -2.2 kcal/mol +/- 0.1kcal/mol in ΔG suggests that the S161L mutation from a polar amino acid to a hydrophobic residue is favorable within the context of all the other compensatory mutations in 06HA.

2.4 Discussions

2.4.1 Role of salt bridges on stability of antigen-antibody complex

In an earlier study, Krause et al. reported that Ig-2D1 against 18HA may cross react with 09HA, and identified the K167E mutation from experimental screening for escape mutants to Ig-2D1. These escape mutants of 09HA are no longer neutralized by the Ig-2D1 antibody. We have obtained results

consistent with these experimental observations through MD simulation. The 09HA_mut had the lowest binding affinity to Ig-2D1 (Table 2.2). In addition, residue-based free energy decomposition also attested that K167 is a key residue in the binding of Ig-2D1. Others have similarly predicted the importance of this residue using hot spot analysis (51). Xu et al. also determined from crystal structural and experimental mutation studies that the K167 plays a key role in forming a salt bridge between Ig-2D1 and 18HA (42, 44). When K167 was mutated to E, Q or P, the dissociation constant K_d drastically increased. As these studies are based on crystal structures alone or single-residue mutagenesis, they offer limited mechanistic insights on how K167 affects the antigen-antibody interactions.

K167 (09HA) is a positive charged amino acid residue and forms a salt bridge with negatively charged D93 (IgL). When it is mutated to E167 (09HA_mut), the two negative amino acid residues, E167 and D93, were thermodynamically unstable close together. In our MM-PB/GBSA results, the 09HA_mut binds only half as strong as 18HA to Ig-2D1 with a ΔG of -37.9 +/- 1.6 kcal/mol. Since 09HA_mut isn't expected to bind or binds poorly to Ig-2D1, the negative DG could be due experimental setup of the simulation. The 09HA_mut system bound to Ig-2D1 was artificially constructed by superimposing the 09HA_mut and 18HA/Ig-2D1 co-crystal structures (42). Consequently, 09HA_mut was placed in close proximity to Ig-2D1, a state that may not actually occur in nature due to entropic barriers.

Over the length of the simulation, Ig-2D1 remained bound to 09HA_mut, with a number of favorable antigen/antibody interactions that the mutant model inherited from its 18HA/Ig-2D1 crystallographic template. However, the electrostatic repulsion between E167 and D93 eventually led to the loss of other favorable interactions between the antigen and the antibody (Table 1.1). In particular, two H-bond interactions were lost and not replaced between 09HA and 09HA_mut systems.

2.4.2 Relative binding affinities of Ig-2D1 to HA's

Krause et al. reported that the Ig-2D1 concentration taken to neutralize 09HA is 0.04 mg/ml compared to 0.025 mg/ml required for 18HA, suggesting that the binding affinity of Ig-2D1 to 09HA is lower (44). In contrast, Liu et al. (51) predicted, using a single frame reconstructed antigen-antibody system, that six mutations on the 09HA could help the antigen to bind stronger to Ig-2D1. Our simulations determined that the 09HA had weaker binding affinity with Ig-2D1, compared to 18HA, in agreement with the results of Krause et al.

Liu et al (51) also did not identify a role for N160 in 09HA and Ig-2D1 interaction. Our free energy decomposition analysis revealed that S160 in 18HA contributed more to binding than N160 in 09HA. Relatively speaking, the S160N mutation weakened the interaction between Ig-2D1 and HA. The S160N mutation could destabilize the K158-D54 (IgH) salt bridge in the simulations by increasing solvent accessibility and formation of water pockets.

This may have been due to the loss of hydrogen bonds established by S160 from 18HA in 09HA. Even though both S and N are polar amino acid residues, we did not observe any persistent H bonds or other interactions with N160 in 09HA (Table S2.6). The increased water present in the antibody-antigen interface could further weaken the electrostatic interactions from the salt bridges formed. Thus, S160N mutation is likely to have a destabilizing effect on the K158-D54 salt bridge important in the antigen-antibody complex formation. This also suggests that the antibody is possibly recognizing a spatial conformation presented by S160, but not N160. In summary, Ig-2D1 has a higher ΔG toward 09HA compared to its native 18HA antigen primarily due to a S160N mutation and its secondary effects. This demonstrates the advantages of MD simulation, which allows the antigen to change, based on antibody dynamics, and represent a more realistic physiological setting.

The 09HA had only two mutations compared to the 18HA in the epitopes, of which S160N allowed partial escape from the cross-reactive 18HA antibody Ig-2D1. The 06HA, with ten mutations in the epitope residues, was still recognized by Ig-2D1 at a similar affinity as the 09HA *in silico*. We will discuss next how the 06HA might be able to escape from Ig-2D1 immune recognition *in vivo*.

2.4.3 Effect of glycosylation on 06HA recognition

The ΔG's of the 06HA and 09HA systems were -67.6 +/- 1.6 kcal/mol and -68.0 +/- 1.2 kcal/mol, respectively. However, 06HA is not neutralized by Ig-2D1 experimentally. Our sequence analysis revealed that this is likely due to mutations found in 06HA that introduces glycosylation *in vivo*, which provides the necessary immune escape (Figure 2.1B). K164 from the 18HA is mutated to N164 in 06HA. This K164N mutation introduced a new glycosylation site on the Sa epitope. Another E131T mutation added a new glycosylation site at position 129. Glycosylation has been suggested as a defense mechanism against antibody neutralization (35, 49, 69). The carbohydrates attached on epitopes will cause steric crash with antibodies. The 06HA system prepared our simulation was unglycosylated. During the system set up, all carbohydrates were removed. Thus, the Sa epitope was exposed to Ig-2D1 without any steric hindrance. The ΔG free energy calculation results suggest Ig-2D1 could neutralize the unglycosylated 06HA.

Thus, the *in silico* experimental construct allowed Ig-2D1 to "bind/neutralize" the 06HA, even though it may not be feasible under physiological conditions where glycosylation is present. We cannot exclude other mechanisms in play here, since there are more mutations on 06HA among the epitope residues. For example, spatial conformations may be more important in epitope by Ig2-2D1 in this case. Nonetheless, our observations suggest a mechanism for immune augmentation when glycosylation inhibitors may enhance the protection from preexisting antibodies or immune memory.

2.4.4 MM-PB/GBSA analysis for relative binding energy determination

The PB model for implicit solvent was considered more accurate than the GB method for free energy calculations (70). However, recent research has shown that the GB implicit solvent model may be better at predicting relative binding energy than PB (71). In the current work, both the GB and PB models performed similarly, as judged by the relative DG rankings of the four HA/Ig-2D1 systems studied (Table 2). MM-PBSA is relatively fast and could predict relative binding energy very well (72), and our results provided further support for the validity of MM-PBSA in the antibody-antigen systems studied. Other methodologies have been reported in the literature to calculate the free energy of binding, including free energy perturbation (FEP) and thermodynamic integration (TI) (73, 74). Even though both of these methods can produce ΔG values that closely match experiment values, they are computationally more expensive.

Xia et al. used FEP to determine the $\Delta\Delta G$'s between different HAs and monoclonal antibodies. Their results showed that a single escape mutation would increase the $\Delta\Delta G$ by 7.28 to 15.47 kcal/mol (73). Interestingly, our decomposition energy also showed that the K167E mutation contributed 9.2 +/- 0.3 kcal/mol to the total ΔG of the 09HA_mut system. Since ΔG calculation with MM-PB/GBSA for a salt bridge is often overestimated if the salt bridge is buried between two protein interfaces (75), the +9.2 +/- 0.3 kcal/mol difference in energy could be the upper bound limit. Other residues that Xia et al.

suggested would change binding affinity without being detrimental to antibody-antigen interactions had reported ΔΔG values under 4.24 kcal/mol. Our results also predicted that for mutations that do not reverse binding, their ΔΔG's in decomposition energies were also under 2.2 kcal/mol.

2.4.5 Entropy consideration in relative binding energy determination

Conformational entropy has been implicated in the antibody maturation process when mutations in the Fab and Fc regions modulate the binding of antibodies to antigens (76). In our simulation, we did not consider entropic contributions in our specific systems for several reasons. First, the systems are similar in their binding states, as shown through RMSD analyses. HA receptor binding domains are very rigid, with only side chain movements observed. Second, no major conformational changes are observed in the antibody or the antigen. This is in contrast to our previous study, where glycan receptors adopt significant changes upon HA binding, and entropy consideration was necessary to obtain results consistent with experimental studies (77). Third, since we are considering the relative free energy of binding of complexes between highly conserved HA's and the same antibody, it is likely that the entropic differences would cancel out. Hou et al. also discussed in details that the inclusion of entropy consideration is not predictive of accuracy in all systems when relative ΔG (ΔΔG) is calculated. In fact, many previous studies have been successful in ranking relative affinities of ligands

without entropy consideration (71). However, this does not mean entropy considerations may be ignored for relative binding affinities, as shown in (77). It is especially crucial when absolute binding energy is considered (71). Finally, entropy consideration is computationally costly using nmode, and the margin of error may fluctuate widely depending on the choice of frames used in the calculations. Convergence is oftentimes an issue using quasiharmonic analysis (77), especially given the sizes of our systems (data not shown). Therefore, the inclusion of entropic consideration may improve the correlation with experimental data, but beyond the scope of our current hypotheses.

2.5 Conclusions

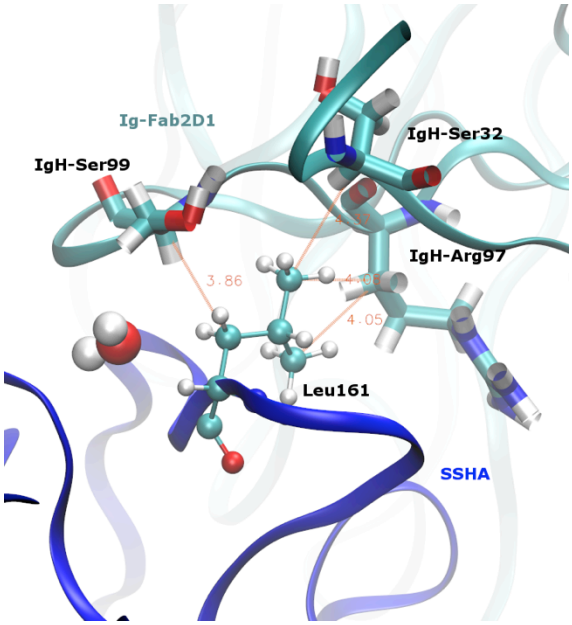
Our results of are consistent with experimental observations using the techniques of MM-GB/PBSA, considering the relative free energy of binding in the systems studied. The formation of two salt bridges plays a key role in the immune recognition of Ig-2D1 of 18HA and cross-reactivity with 09HA. The stability of K158 - D54 (IgH) salt bridges is dramatically weakened by the S160N mutation in the 09HA accompanied by hydrogen bond loss and water pocket formation at the antibody-antigen interface. The immune escape of 09HA may be accomplished through a K167E mutation, which completely disrupts both salt bridges between 09HA and Ig-2D1. On the other hand, 06HA likely achieves immune escape through mutations that introduce glycosylation sites and mask epitope residues. These results provide mechanistic insights to the immune recognition and escape of H1N1 virus, and

could help design better antibodies against this pandemic strain and protection from future threats.

2.6 Appendices

Supporting Table 2.1 Description of each system. The four systems that were simulated in this experiment are shown above. Their system name, PDB ID, strain, simulation time and number of atoms are listed.

Naming Scheme	Crystal Structure	Strain	Simulation Time (ns)	No. of atoms
18HA	3LZF	A/South Carolina/1/1918	69	475,554
09HA	3LZG	A/California/04/2009	69	539,569
06HA	3SM5	A/Solomon Islands/3/2006	69	467,368
09HA_mut	3LZG*	A/California/mutant*	69	565,969



Supporting Figure 2.2 VDW interaction between L161 and R97 (IgH). Ig-2D1 residues are drawn in bonds style and L161 from 06HA is shown in CPK style using VMD. The rest of the Ig-2D1 and 06HA protein backbone are both displayed in ribbon. The 06HA is colored in navy blue and Ig-2D1 is in cyan.

Supporting Table 2.3 Bond interactions between Ig-2D1 and 18HA. All the epitope residues (column I) and the corresponding Ig-2D1 residues (column III) are indicated according to the type of interactions (column II). The interactions that were also found in the crystal structure (PDB ID 3LZF) are marked with $^{+}$ next to Ig residues.

S (125)	Hbond	IgL-W91^{+}
		IgL-D93^{+}
	Dipole	IgL-N95A^{+}
		IgL-G95B^{+}
S (126)	Hbond	IgL-D93^{+}
	Dipole	IgL-W91
P (128)	VDW	IgL-W91^{+}
K (158)	Salt	IgH-D54^{+}
	Hbond	IgH-R97
	Dipole	IgH-D54^{+}
		IgH-T56^{+}
		IgH-R97
	VDW	IgH-D54^{+}
G (159)	Dipole	IgH-D53^{+}
		IgH-R97
S (160)	Hbond	IgH-D53
		IgH-D53
	Dipole	IgH-R97
		IgH-S99^{+}
S (161)	Dipole	IgH-G33
		IgH-R97^{+}
		IgH-S99^{+}
Y (162)	Hbond	IgH-G100^{+}
	Dipole	IgH-V98
		IgH-S99^{+}
P (163)	Dipole	IgH-D100A^{+}
	VDW	IgH-R97^{+}
K (164)	Hbond	IgH-D100A
	VDW	IgH-Y100B^{+}
K (167)	Salt	IgL-D93^{+}
	Hbond	IgL-N31^{+}
	Dipole	IgL-S30^{+}
		IgL-N31^{+}
S (168)	Hbond	IgL-S30^{+}
	Dipole	IgL-S30^{+}
Q (197)	Hbond	IgH- S99
N (198)	Dipole	IgH- S99^{+}
		IgH- G100^{+}
T (249)	Dipole	IgH- G100^{+}
		IgH- G100

Supporting Table S2.4 Bond interactions between Ig-2D1 and 09HA. All the epitope residues (column I) and the corresponding Ig-2D1 residues (column III) are indicated according to the type of interactions (column II).

S (125)	Hbond	IgL-W91
S (126)	Hbond	IgL-D93
	Dipole	IgL-W91
P (128)	VDW	IgL-W91
	Dipole	IgH-R97
	VDW	IgH-D54
G (159)	Dipole	IgH-R97
		IgH-D53
S (161)	Dipole	IgH-R97
Y (162)	Hbond	IgH-G100
P (163)	Dipole	IgH-G100
		IgH-D100A
	VDW	IgH-R97
K (164)	Hbond	IgH-D100A
	Dipole	IgH-W100B or IgH-D100A or IgH-G100
	VDW	IgH-D100A
		IgH-W100B
K (167)	Salt	IgL-D93
	Hbond	IgL-N31
	Dipole	IgL-S30
S (168)	Hbond	IgL-S30
	Dipole	IgL-S30
T (249)	Dipole	IgL-G100

Supporting Table S2.5 Bond interactions between Ig-2D1 and 06HA. All the epitope residues (column I) and the corresponding Ig-2D1 residues (column III) are indicated according to the type of interactions (column II).

E (124)	Hbond	IgH-Y58
S (125)	Hbond	IgL-W91
	Dipole	IgH-Y58
S (126)	Dipole	IgL-W91
P (128)	VDW	IgL-W91
K (158)	Dipole	IgH-R97
N (159)	Hbond	IgH-R97
L (161)	VDW	IgH-R97
Y (162)	Dipole	IgH-G100
P (163)	Dipole	IgH-D100A
	VDW	IgH-V100C
N (164)	Hbond	IgH-D100A
	Dipole	IgH-V100C
		IgH-G100 or IgH-D100A
L (165)	VDW	IgH-Y100B
S (166)	Dipole	IgL-T32
K (167)	Salt	IgL-D93
	Hbond	IgL-N31
	Dipole	IgL-S30
S (168)	Hbond	IgL-S30
	Dipole	IgL-G29
		IgL-S30

Supporting Table S2.6 Bond interactions between Ig-2D1 and 09HA_mut. All the epitope residues (column I) and the corresponding Ig-2D1 residues (column III) are indicated according to the type of interactions (column II).

S (125)	Hbond	IgL-W91
	Dipole	IgL-D93
		IgL-N95A
P (128)	VDW	IgL-W91
K (158)	VDW	IgH-D54
G (159)	Dipole	IgH-D53
		IgL-S24
S (161)	Dipole	IgL-G25
		IgH-S99
Y (162)	Hbond	IgH-S99
		IgH-G100
P (163)	Dipole	IgH-G100
		IgH-D100A
	VDW	IgH-R97
K (164)	Hbond	IgH-D100A
	VDW	IgH-D100A
		IgH-Y100B
S (168)	Dipole	IgL-S30
N (198)	Dipole	IgH-S99
		IgH-G100

Supporting Table S2.7 MM-PBSA energy breakdown of the four systems. All energies are reported in kcal/mol, averages from three independent calculations over a period of 69 ns. ΔE_{vdW} is the van Der Waal term and ΔE_{elec} is the electrostatic energy. ΔE_{PB} is the solvation energy estimated using the Poisson Boltzmann equation. ΔE_{cavity} is a repulsive nonpolar de-solvation energy term. ΔG_{gas} is the energy of the protein complex in vacuum and ΔG_{solv} is the energy takes to add solvent to a system in vacuum.

PB	18HA		09HA		06HA		09HA_mut	
Contribution	Mean	SE	Mean	SE	Mean	SE	Mean	SE
ΔE_{vdW}	-117.7	1.1	-112.2	1.1	-106.4	1.1	-112.2	0.8
ΔE_{elec}	-434.9	4.5	-521.6	4.9	-428.5	5.0	-305.3	5.7
ΔE_{PB}	490.4	4.6	577.2	5.0	479.6	5.0	451.0	5.4
ΔE_{cavity}	-12.2	0.1	-11.4	0.1	-12.3	0.1	-11.5	0.1
ΔG_{gas}	-552.7	4.8	-633.8	5.2	-534.9	5.2	-417.5	5.9
ΔG_{solv}	478.2	4.5	565.8	4.9	497.3	4.9	379.6	5.3
$\Delta G_{subtotal}$	-74.4	1.1	-68.0	1.2	-67.6	1.6	-37.9	1.6

Supporting Table S2.8 MM-GBSA energy breakdown of the four systems. All energies are reported in kcal/mol, averages from three independent calculations over a period of 69 ns. ΔE_{vdW} is the van Der Waal term and ΔE_{elec} is the electrostatic energy. ΔE_{GB} is the solvation energy estimated using the Generalized Born equation. ΔE_{cavity} is a repulsive nonpolar de-solvation energy term. ΔG_{gas} is the energy of the protein complex in vacuum and ΔG_{solv} is the energy takes to add solvent to a system in vacuum.

GB (igb2)	18HA		09HA		06HA		09HA_mut	
Contribution	Mean	SE	Mean	SE	Mean	SE	Mean	SE
ΔE_{vdW}	-117.7	1.1	-112.2	1.1	-106.4	1.1	-112.2	0.8
ΔE_{elec}	-434.9	4.5	-521.6	4.9	-428.5	5.0	-306.1	5.7
ΔE_{GB}	493.5	4.2	585.8	4.7	486.2	4.9	393.2	5.3
ΔE_{cavity}	-15.3	0.1	-14.6	0.1	-14.6	0.1	-14.1	0.1
ΔG_{gas}	-552.7	4.8	-633.8	5.2	-534.9	5.2	-418.3	6.0
ΔG_{solv}	478.1	4.2	571.1	4.7	471.5	4.9	379.2	5.3
$\Delta G_{subtotal}$	-74.5	1.0	-62.7	1.1	-63.4	1.5	-39.1	1.1

Supporting Table S2.9 Energy decomposition breakdown of the four systems. All energies are reported in kcal/mol +/- standard deviation of the mean. The values are obtained from averaging three monomers over 69 ns. All the canonical epitope residues and the surrounding residues are shown for the 18HA. For other systems, only the mutations are listed.

Position		18HA		09HA		06HA		09_mut
124	T	-0.6±0.1		-1.1±0.1	E	0.4±0.1		-1.0±0.1
125	S	-3.4±0.1		-3.1±0.2		-2.6±0.1		-3.5±0.2
126	S	-3.1±0.1		-3.0±0.1		-3.5±0.2		-0.7±0.1
128	P	-2.7±0.1		-3.0±0.1		-2.6±0.1		-2.9±0.1
129	N	0.4±0.1		-0.9±0.1		-0.2±0.1		-0.1±0.1
157	K	0.3±0.0		0.5±0.0	G	0.2±0.0		0.6±0.0
158	K	-4.4±0.2		-2.3±0.2		-2.2±0.2		-1.9±0.2
159	G	-3.1±0.1		-2.2±0.1	N	-2.9±0.1		-1.8±0.1
160	S	-4.9±0.2	N	-0.5±0.1	G	-0.0±0.0	N	-0.9±0.1
161	S	-1.8±0.1		-1.8±0.1	L	-4.0±0.1		-2.0±0.1
162	Y	-2.3±0.1		-2.3±0.1		-0.7±0.1		-2.6±0.1
163	P	-4.5±0.1		-4.7±0.1		-3.3±0.1		-4.7±0.1
164	K	-3.9±0.1		-3.8±0.1	N	-2.7±0.1		-3.3±0.1
165	L	-0.6±0.1		-0.9±0.1		-1.4±0.1		-0.4±0.0
166	S	-0.6±0.1		-0.6±0.1		-1.0±0.1		-0.6±0.1
167	K	-6.6±0.2		-6.3±0.2		-6.6±0.2	E	2.9±0.1
168	S	-1.7±0.1		-1.1±0.1		-2.3±0.1		-0.4±0.1
170	V	-1.3±0.1	I	-1.5±0.1	A	-0.7±0.0	I	-1.3±0.1
197	Q	-0.7±0.1		-0.5±0.1	H	0.1±0.0		-0.6±0.1
198	N	-0.5±0.0		-0.5±0.1	T	-0.4±0.0		-0.4±0.1
247	E	0.8±0.0		1.0±0.0		0.7±0.0		0.8±0.0
249	T	-0.5±0.0		-0.6±0.0	N	-0.6±0.1		-0.6±0.0

This chapter, in full, is a reprint of the material as it appears in “Molecular Dynamic Analysis of Antibody Recognition and Escape by Human H1N1” by Ieong, Pek U; Li, Wilfred; Amaro, Rommie E., published 2015 in Biophysical Journal. This chapter is included with the permission from Li, Wilfred and Amaro, Rommie E.

Chapter 3

Full-length p53 Tetramer Bound to DNA and Its Quaternary Dynamics

Abstract

p53 is a major tumor suppressor that is mutated and inactivated in about 50% of all human cancers. Thus, reactivation of mutant p53 using small-molecules is an attractive anti-cancer therapeutic strategy. p53 is a challenging protein to structurally characterize because of its highly flexible regions. To explore p53 dynamics, we here use molecular modeling and available crystal structures to construct an all-atom model of the full-length p53 (fl-p53) tetramer bound to DNA. Three different DNA sequences (a p21 response element, a puma response element, and a non-specific DNA sequence) are integrated into this model. The simulations yield a final structure that agrees with prior cryo-EM maps(78) and, for the first time, show the direct interaction of the p53 C-terminal with DNA in atomic detail. Through a collective principal component analysis, we identify sequence-dependent differential quaternary binding modes of the p53 tetramer interfacing with DNA. Additionally, L1 loop dynamics of fl-p53 in the presence of DNA is revealed, and druggable pockets of p53 are identified via solvent mapping in order to aid future drug-discovery studies.

3.1 Introduction

Thousands of mutations occur daily in the DNA of each human cell, even at times of perfect health. To prevent tumor formation, the human body has a complex but efficient mechanism for detecting and fixing DNA mutations. p53, also known as “the guardian of the genome”, lies at the heart of this complex tumor-suppression mechanism. Once activated, it signals for cell-cycle arrest, senescence, or apoptosis, either via transcription of various target genes(79, 80) or through non-transcriptional pathways(81-83).

As tumor initiation and maintenance requires the inactivation of p53 pathways, p53 is also the most frequently mutated gene in human cancers. p53 is mutated and non-functional in about 50% of all human cancers; about three-forths of p53 mutations are single point-mutations, and most mutations diminish DNA-binding ability(84). Due to its major role in tumor suppression, many researchers seek small molecules that can reactivate mutant p53 and thereby suppress tumors(85-90). Recent research in transgenic mice demonstrated that p53 reactivation can indeed achieve tumor regression, highlighting p53 reactivation as a very promising anti-cancer therapeutic strategy(91-93).

Full-length p53 (fl-p53) is in part an intrinsically disordered protein (IDP), complicating its complete structural characterization. Due to high flexibility, IDPs such as fl-p53 rarely form crystals and often yield complex NMR spectra, eluding characterization by both X-ray crystallography and NMR

spectroscopy (94). Fl-p53 consists of 393 residues that form a flexible N-terminal domain (NTD), a core DNA binding domain (DBD), a flexible linker region, a tetramerization (TET) domain, and a flexible C-terminal domain (CTD) (95) (Figure 3.1a). The core p53 DBD domain is the most studied because all inactivating p53 mutations occur there, and it possesses definite secondary and tertiary structural elements that are amenable to crystallography and NMR. Also, fl-p53 binds DNA as a tetramer (96) and causes DNA to bend (96-98). To shed some light onto the quaternary structure of fl-p53 tetramer/DNA complex, an integrative medium-resolution 3-dimensional map was constructed by combining data from small-angle X-ray scattering (SAXS), electron microscopy (EM) and NMR spectroscopy (99). A subsequent EM study pointed to multiple DNA binding modes of fl-53 tetramers (78). Recently, others crystalized the tetrameric p53 DBD and TET (with the linker domain truncated) bound to a short strand of DNA, setting the stage for the current study (100-102). Despite many years of work, there are still many questions about p53 structure and function.

Upon activation, p53 needs to efficiently locate and bind to its response elements (REs) on the genome in order to stimulate transcription of target genes (e.g. p21, puma) and subsequently regulate the cell cycle by initiating DNA repair, cell-cycle arrest, or apoptosis (79, 80). However, the mechanism by which p53 searches and recognizes its REs is still under debate. REs consist of four head-to-head nucleotide pentamer repeats or two repeating 10-

nucleotide motifs (RRRCWWGYYY) called the half sites, where R is A or G, W is A or T and Y is C or T (100) (Figure 3.1b). The two half sites are separated by 0 to 13 nucleotides (103, 104). The fl-p53 tetramer forms a dimer of dimers and each p53 dimer binds to a DNA half site (105). The fl-p53 tetramer tightly binds to the REs signaling for cell-cycle arrest (e.g. p21), DNA repair, negative regulation and anti-angiogenesis, while its affinity towards the pro-apoptotic REs (e.g. puma) can be either high or low (106). The role of the DBD has been studied extensively, but the detailed DNA binding mechanism of the fl-p53 tetramer is not well characterized, nor is the effect that DNA sequence has on that binding well understood.

Additionally, the role of the CTD in facilitating p53's DNA search has been controversial (107). Earlier studies showed that the p53 CTD acts as a negative regulator by hindering DBD binding to the short strands of specific response elements (REs) (108). CTD phosphorylation and acetylation alleviate constraints and increase DBD binding to target sites (108). However, further research suggested the opposite: the p53 CTD is needed for the DBD to recognize target sites in long or circular DNA and acts as a positive regulator (109). Using single-molecule experiments, Tafvizi *et al.* explained these two seemingly contradictory observations by proposing that CTDs facilitate target-site search by sliding through the non-specific DNA while the DBDs are immobilized, moving by frequent association and dissociation (110). From a thermodynamic point of view, the CTD hinders DBD binding to its target-site

because during the search process, DBDs are impeded in the non-specific region (110). But from a kinetic point of view, the CTD promotes the search process and helps the DBD find its specific target sequence (110). Consequently, understanding how the p53 CTD behaves at the molecular level is of great interest.

Lastly, the p53 L1 loop is implicated to be an important conformational switch that regulates DNA binding (101, 102). In all p53 crystal structures published before 2011, the L1 loop was captured in an extended conformation. However, the crystal structure of a more recent tetrameric p53 bound to the p21 RE showed that the two inner L1 loops (monomers B and C) adopt an extended conformation, while the two outer L1 loops (monomers A and D) adopt a recessed conformation (Figure 3.5a,c) (100). Lukman *et al.* performed molecular dynamics (MD) simulations of a single p53 DBD monomer in the absence of DNA (111). They observed that the L1 loop was the most flexible region, and that it can adopt both an extended and a recessed conformation, switching from one conformation to another on the nanosecond timescale (111). L1 loop dynamics under physiological conditions is also of great interest because it forms part of the L1/S3 pocket. We previously identified this druggable pocket and discovered a small-molecule ligand that reactivates p53 mutants (85).

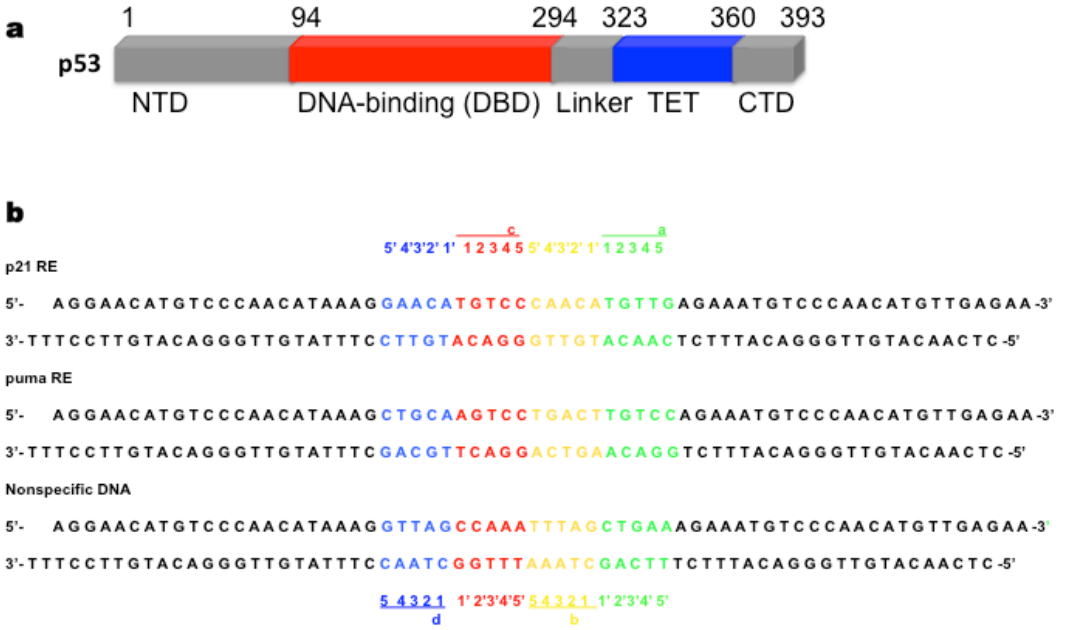


Figure 3.1 Full-length p53 and the different DNA set-up. a) The full-length p53 (fl-p53) sequence with each of the five domains labeled: N-terminal domain (NTD), DNA binding domain (DBD), linker, tetrameric domain (TET), and C-terminal domain (CTD). b) The three different DNA sequences used in the simulations: the two positive response elements (REs), p21 RE and puma RE, and a non-specific DNA sequence. The binding motif consists of two half sites or four pentamer repeats. Pentamers a, b, c, and d are highlighted in blue, red, yellow and green, respectively.

Here, we construct an all-atom model of the fl-p53 tetramer bound to DNA, based on available crystal structures and modeling. We then use this model to explore the structure and dynamics of the fl-p53 tetramer when bound to different DNA sequences including two REs and a non-specific DNA (Figure 3.1b). In our simulations, we observe p53 CTD motion toward DNA,

leading to direct non-specific interactions. The final fl-p53 tetramer structures generated at the end of the simulations are comparable to published 3-dimensional cryo-EM maps of the protein (Figure 3.2a). We capture multiple binding modes of the fl-p53 tetramer/DNA complex that differ depending on the DNA sequence. We also explore the L1 loop dynamics of the fl-p53 tetramer and identify p53 druggable regions based on our DNA-bound tetramer model.

3.2 Results

3.2.1 Steady decrease of the radius of gyration

We simulated three full-length p53 systems (fl-p53) with three different DNA sequences: p21 RE, puma RE, and non-specific DNA. (Figure 3.1b) In order to monitor the global changes in the p53 tetramer, we calculated the radius of gyration values during simulations. Radius of gyration reflects how far the protein stretches from its center of mass, and thus, a small radius of gyration indicates a more compact structure, while a large radius of gyration indicates a more elongated one. Our initial model had all flexible loops in extended conformation, and the radius of gyration decreased steadily during simulations in all systems. However, among the 3 systems with different DNA sequences, the non-specific DNA-bound system reached the lowest radius of gyration at the end of the simulations while the p21-RE-bound system reached

the largest value. (Supplementary Figure 3.1 and Figure 3.2b) Our results implicate that fl-p53 tetramer adopts the most compact form when bound to a non-specific DNA sequence, and the most elongated form when bound to the p21 RE. The fl-p53 tetramer interacting with the puma RE fell in the middle. The final structures yielded in the simulations agree with prior cryo-EM maps(78).

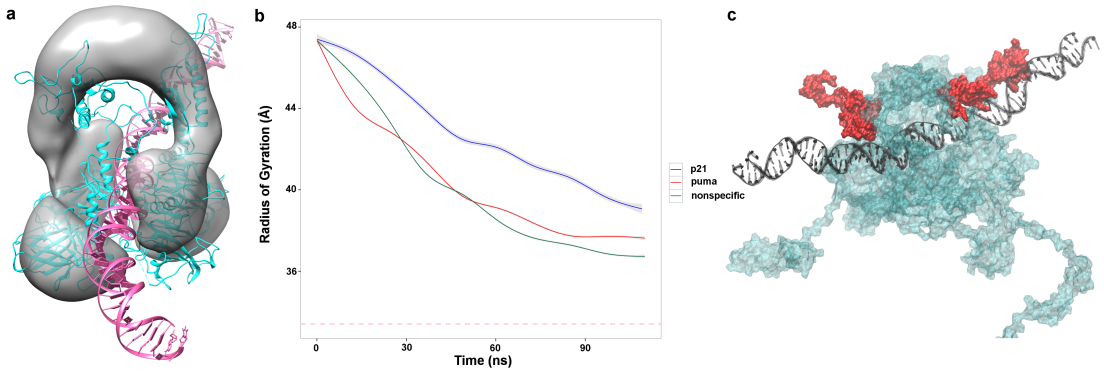


Figure 3.2 Full-length p53 global conformational change. a) Cartoon drawing of the DBD, linker and CTD fitted into the classIII cryo-EM map of p53 from *Melero et al*(78) using Chimera program.(53) DNA colored in magenta is drawn to show relative orientation. b) The time evolution of the average radius of gyration for the Cαs of DBD and TET of each system. c) The p53 C-terminals, highlighted in red, interacting with the DNA, highlighted in gray. The rest of the fl-p53 is colored in cyan.

3.2.2. C-terminal domain directly contacts the DNA

Next, we inspected the flexible loops, which contributed to the steady decrease radius of gyration. Visualization of the MD trajectories revealed that the C-terminal domains of the fl-p53 tetramer approached and directly contacted the DNA in all of our simulations independent of the DNA sequence.

(Figure 3.2c) This is a remarkable observation in 110 ns simulations given that the C-terminal domains had very extended conformations and were quite distant from the DNA initially. Especially the C-terminal of monomer C ended up contacting the DNA in every single MD simulation. In our initial model system, the C-terminal of monomer C was unintentionally built slightly closer to the DNA. Yet, the C-terminals of other monomers including monomers A and B also ended up interacting directly with the DNA.

We also performed a principal component analysis (PCA) including all fl-p53 Cαs in all simulations. The first principal component (PC1) was a motion of the CTDs becoming more compact and approaching the DNA, in line with the steady decrease of the radius of gyration. All three systems sampled the motion described by PC1.

Lastly, we carried out an interaction footprint analysis and identified the salt bridges between the p53 CTDs and DNA. (Supplementary Table 3.1) The key p53 residues that participated in the salt bridge interactions were Lys370, Lys372, Lys373, Arg379, Lys381, Lys382 and Lys386. The p53 CTDs interacted with the DNA only non-specifically via the DNA backbone atoms, and the interactions were variable/dynamic with different parts of the DNA segments at different times. The motions we observed at the molecular level directly support the previously suggested idea that p53 CTDs do not participate in specific DNA recognition and binding, but rather participate in dynamic DNA search (110).

3.2.3 Quaternary binding modes of p53 DBD tetramer to different DNA sequences

Besides the PCA analysis on all the Cα atoms of fl-p53, we performed another PCA including only the Cα atoms of the DBDs (resid 89-291) of the p53 tetramer. Interestingly, this time PC1 showed a clamping/unclamping motion of the tetrameric p53 DBDs around DNA. This quaternary motion can be described in more detail as going from a more asymmetric form of the p53 DBD tetramer clamped around the DNA with 2 DBDs curved inward (low PC1 values) to a more symmetric and flat form of the p53 DBDs in which all 4 monomers are in plane (high PC1 values). (Figure 3.3a and Supplementary Movie 3.1) In our simulations, the p21 RE system only sampled low PC1 values while the puma RE and the non-specific DNA systems extended beyond and sampled both low and high PC1 values. (Figure 3.3a and Supplementary Figure 3.3) In other words, the p21 RE system sampled only the more clamped conformation while the puma RE and non-specific DNA systems sampled both the clamped and the flat conformations. The puma RE system spent more time sampling the more clamped conformation and less time in the flat conformation while it was vice versa in the non-specific DNA system. (Figure 3.3a and Supplementary Figure 3.3) We should also note that all these model systems were constructed by mutating the DNA in the crystal structure of p21-bound p53 tetramer system and thus the initial conformation was the same for all of them. The fl-p53 tetramer is known to have much

higher binding affinity to the p21 RE compared to the puma RE, and a minimal binding is expected in the case of the non-specific DNA sequence. PC1 indicated that fl-p53 tetramer adopts different DBD tetramer conformations to accommodate tighter DNA binding as in p21 RE, weak DNA binding as in puma RE, and minimal binding as in non-specific DNA.

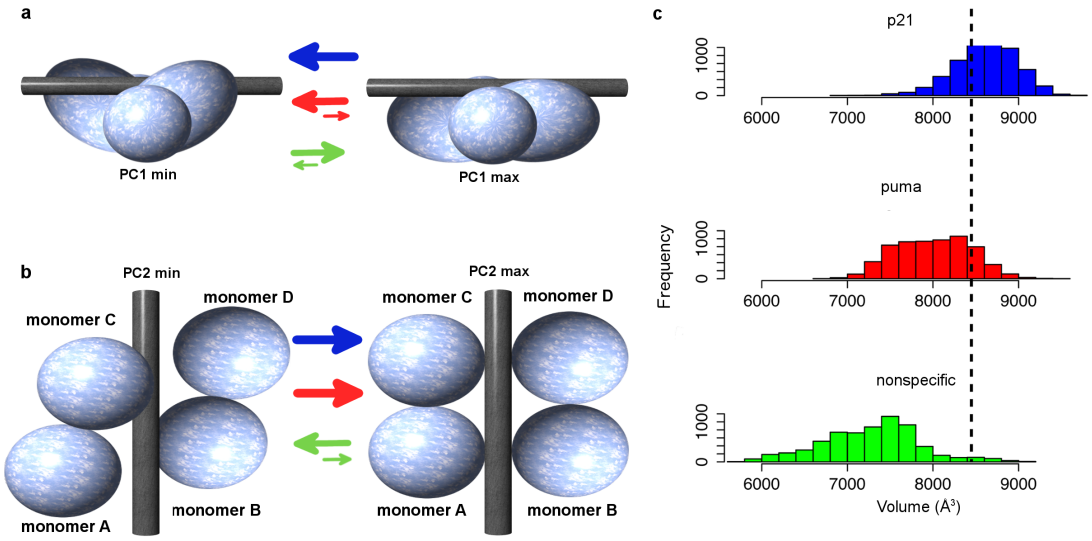


Figure 3.3 Quaternary DBD binding modes. Panels a and b show the DBD binding modes corresponding to the minimum and maximum values of PC1 and PC2, respectively. DBDs and DNA are represented as spherical marbles and a cylinder, respectively. The arrows aim to highlight the conformation that each system samples, and the size of the arrow is proportional to the frequency of sampling. The blue, red and green arrows represent the p21 RE, the puma RE and the non-specific DNA systems, respectively. c) Histogram showing the distribution of the volume gap between the DBDs during simulations of each system. The black dashed line demonstrates the initial volume gap value prior to the simulations.

On the other hand, PC2 described a cooperative binding motion of p53 DBD tetramer upon specific DNA recognition. This quaternary motion can be

described in more detail as going from an asymmetric non-cooperative mode in which monomers A and D are pushed away from the DNA (low PC2 values) to a symmetric cooperative binding mode in which all four monomers are in close proximity of DNA (high PC2 values). (Figure 3.3b and Supplementary Movie 3.2) In our simulations, both the p21-RE-bound and puma-RE-bound p53 tetramer systems sampled only high PC2 values while the non-specific DNA-bound p53 tetramer system solely sampled low PC2 values. (Supplementary Figure 3.3) We should again note that all of our MD simulations started from a cooperative binding mode observed in the p21-RE-bound p53 tetramer crystal structure.

To further inspect the effect of different quaternary binding modes of p53 tetramer on DNA binding interface, we measured the volume gap between four p53 DBD monomers where DNA would be accommodated. (Supplementary Figure 3.2) The results for our simulations revealed that the p21-RE-bound system provided the largest volume gap to accommodate DNA followed by puma-RE-bound and then non-specific DNA-bound systems. (Figure 3.3c) The binding mode seen in p21-RE-bound simulations (characterized by low PC1 and high PC2 values) provided the largest volume gap available for DNA accommodation, while the binding mode seen in non-specific DNA-bound simulations (characterized by high PC1 and low PC2 values) provided the smallest volume gap for DNA accommodation.

Furthermore, hydrogen-bond footprint analysis complemented our DBD PCA analysis. (Supplementary Table 3.2) There were 10 direct hydrogen bonds we identified between fl-p53 DBDs and DNA, but here we focused on the most significant ones. The two most persistent H-bonds we observed in MD were 1. between Ala276 backbone and DNA and 2. between Arg280 side chain and DNA. These two H-bonds were observed highly persistently only in the case of the positive REs, p21 and puma, but were very scarcely seen in the case of non-specific DNA system. Arg273 side chain also formed a more persistent H-bond to DNA in the p21 and puma systems compared to the non-specific DNA system, but the difference was less pronounced. Ser241 was the only residue that binds to all three DNA sequences persistently, indicating that Ser241 is a sequence-independent H-bond donor/acceptor for DNA. This hydrogen bond interaction could be important for the fl-p53 during the DNA search process. We should also note that the most persistent H-bond between the Lys120 and DNA was seen only in the inner monomers (monomers B and C) in the p21 RE system followed by the puma RE system. The H-bond between Lys120 and DNA was not persistent in the non-specific DNA system. Overall, more persistent H-bonds between the fl-p53 tetramer and DNA were observed in the p21 RE and puma RE systems compared to the non-specific DNA-bound system, in line with the clamped binding modes observed for the positive RE systems.

This is the first time to our knowledge that p53 DBD tetramers are explicitly shown to adopt different conformations upon binding to various DNA sequences. We observed that the p53 DBD tetramer adopts a cooperative, clamping, tight-binding mode to bind positive REs (e.g. p21 and puma) while it adopts a non-cooperative, flat, loose-binding mode to bind non-specific DNA sequences. (Figure 3.2a,b) These two binding modes can be related to the two-state mechanism of DNA search and recognition suggested for p53 by Tafvizi *et al* based on their single-molecule experiments (110). The non-cooperative, loose-binding mode we observed in the non-specific DNA case can represent the DNA search mode, while the cooperative, tight-binding mode in the p21 RE case can represent the DNA recognition mode.

3.2.4 DNA Distortion

Next, we examined the effects of fl-p53 tetramer binding on the structure of the DNA. It has been shown previously that DNA gets bent by 27° upon binding to fl-p53 (98). Among our three systems, we observed that only the p21-RE-bound p53 tetramer system achieved significant DNA bending and intercepted with the experimental value (Figure 3.4, Supplemental Figure 3.4).

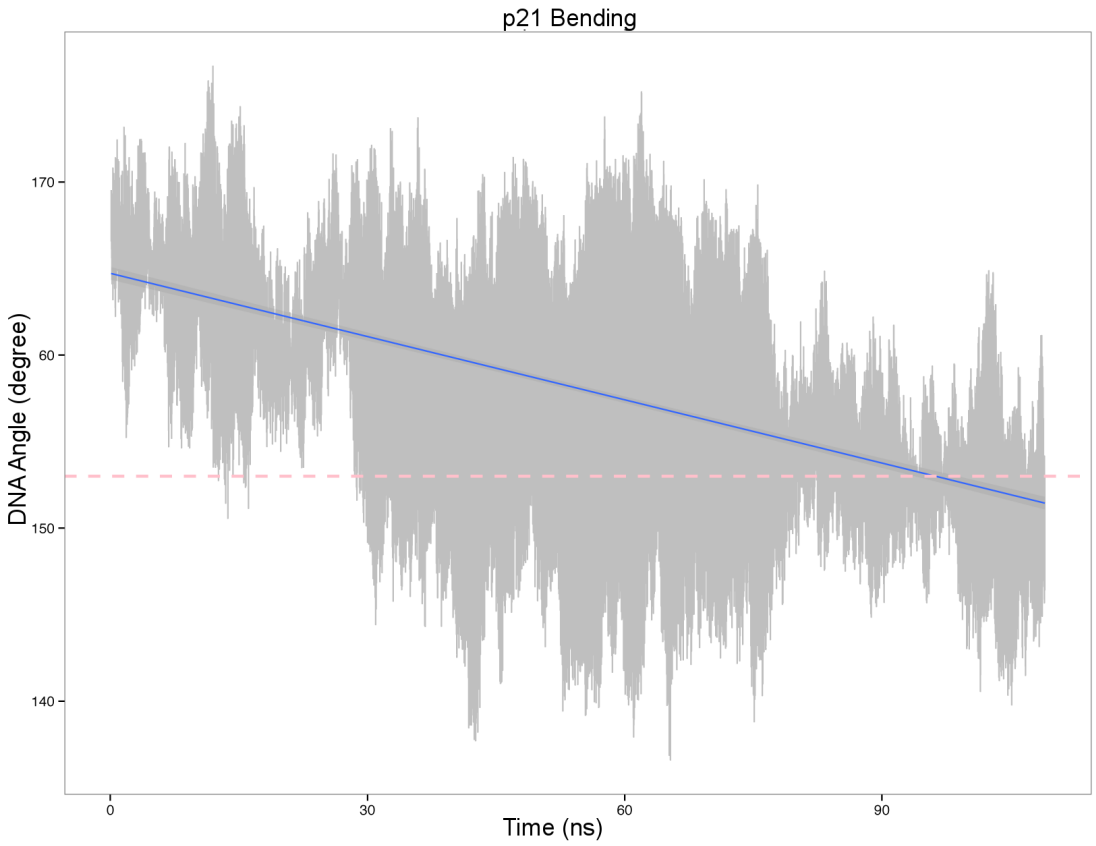


Figure 3.4 DNA bending angle. Time evolution of the DNA bending angle for the three MD copies of the p21 RE system are shown in light grey. A regression line is drawn in blue to show the global trend and, a pink dashed line shows the experimental bending angle(98).

Beyond DNA bending, we further examined average DNA properties in simulations and revealed several that were significantly sequence-dependent. (Table 3.1) Minor groove width was 7.28Å ± 1.12Å, 7.78Å ± 1.13Å and 4.79Å ± 1.20Å for the p21 RE, puma RE, and non-specific DNA systems, respectively. The minor groove width for DNA is typically 4.2 Å, which is within the standard deviation of the non-specific DNA minor groove width, but significantly lower

than the values observed in the positive REs. The much wider DNA minor grooves observed in p21-RE- and puma-RE-bound p53 tetramer systems suggest that the minor groove widens up when the fl-p53 tetramer binds these positive REs in a tight DNA recognition mode.

Table 3.1 Comparison of average DNA properties in MD simulations of the three systems. Average and standard deviation of the values in 3 copies of MD simulations are reported.

	p21	puma	non-specific DNA
Minor groove width	7.28 ± 1.12 Å	7.78 ± 1.13 Å	4.79 ± 1.20 Å
h-twist	29.2 ± 5.7°	30.4 ± 7.4°	34.0 ± 4.1°

Another DNA property we found to be sequence-dependent was the h-twist, which corresponded to the rotation between base pairs (112). The h-twist in our simulations were 29.2° ± 5.7°, 30.4° ± 7.4°, and 34.0° ± 4.1° for p21 RE, puma RE and non-specific DNA systems, respectively. (Table 3.1) The two smaller values observed in the two positive REs, p21 and puma, suggested that in the DNA recognition mode, p53 binding caused a slight DNA untwisting.

3.2.5 L1 loop dynamics

To monitor the L1 loop dynamics of each p53 monomer in our simulations, we measured the time-dependent RMSD of the L1 loop Cα atoms with respect to both the extended and the recessed conformations of the L1 loop in the crystal structure. We did not observe a complete transition of the L1 loop from recessed conformation to extended conformation or vice versa in any of the simulations. The L1 loops of the inner p53 monomers conserved the extended L1 loop conformation throughout the simulations independent of the DNA sequence. (Supplementary Figure 3.5-3.7, Figure 3.5b) However, the recessed L1 loops of the outer p53 monomers were more flexible and sampled intermediate conformations in addition to the recessed conformation. (Supplementary Figure 3.5-3.7, Figure 3.5b) These intermediate L1 conformations were sampled by at least one monomer in the p21 RE and non-specific DNA systems, while it was not observed in the puma RE system. (Supplementary Figure 3.5-3.7) Based on our simulations, the L1 loop conformation of a p53 monomer in DNA-bound fl-p53 tetramer system is dictated by the position of the monomer with respect to the DNA. (Figure 3.5b)

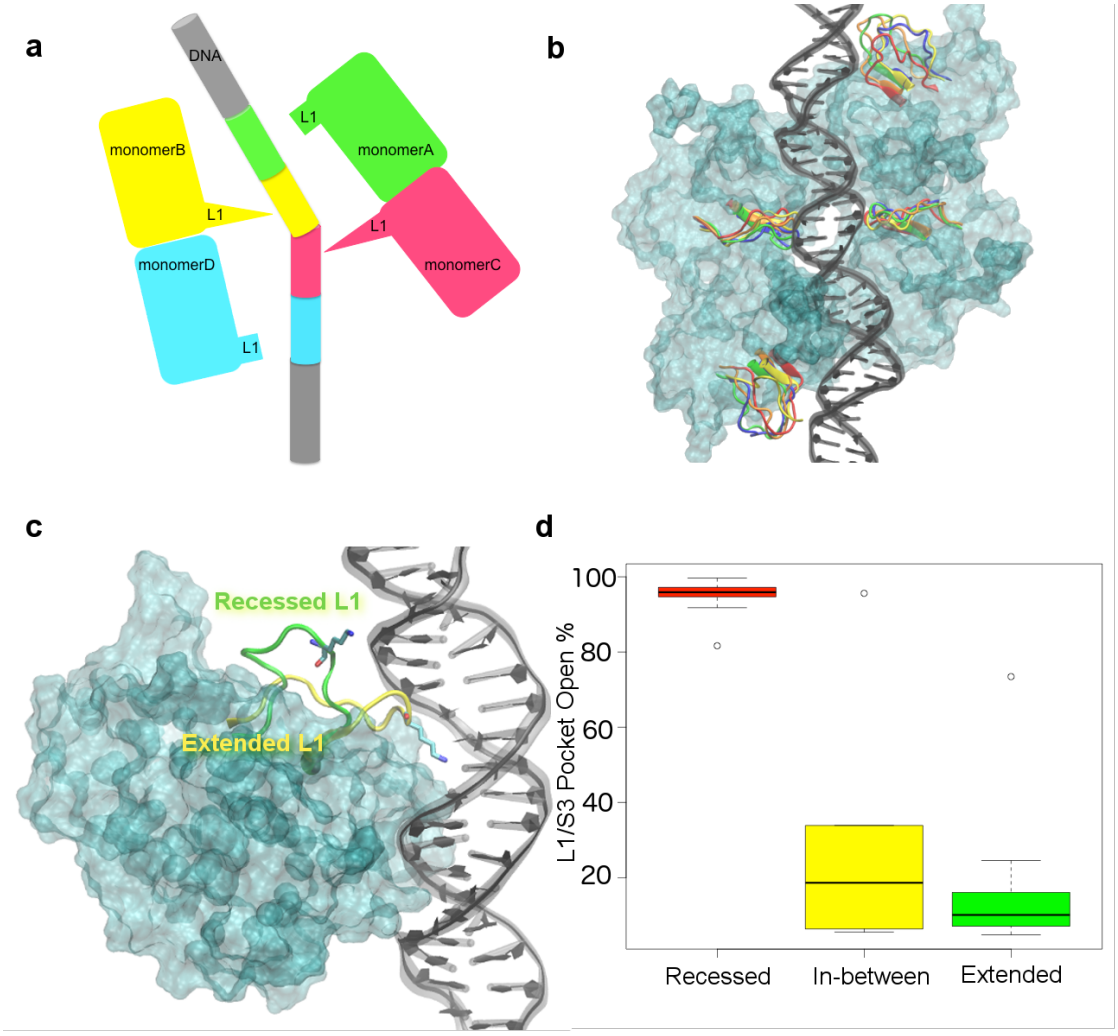


Figure 3.5 L1 loop conformations in the p53 tetramer. a) Cartoon figure that represents the two different L1 conformations and the regions in DNA that each monomer interacts with. The extended L1 loops are seen in monomer B and C while the recessed L1 loops are seen in monomer A and D. b) The conformational space the L1 loop can sample in each monomer. The L1 loops in different conformations are colored in red, orange, yellow, green and blue. The DBD monomer surfaces are colored in cyan and the DNA is colored in black. c) Close-up view of the extended L1 conformation interacting with the DNA. The recessed L1 loop conformation from another monomer is also shown for comparison. d) Boxplot that shows the median, 1st and 3rd quantiles, and the outliers of the time percentage of the L1/S3 pocket being open at various L1 conformations

We were interested in the L1 dynamics also because it directly effects the druggable L1/S3 pocket we previously identified (85) near the L1 loop and found strong evidence to be the binding site for PRIMA-1 in clinical trials (87). In the same study, we also discovered stictic acid to be a novel p53 reactivation compound by using the MD-generated L1/S3 pocket-open conformation in virtual screening (85). Using several geometric criteria as a filter for the pocket-open state, we found the L1/S3 pocket was open only about 6% of the time in MD simulations of the DBDs of wild-type p53 and various p53 mutants (85). In order to investigate the L1/S3 pocket dynamics in fl-p53 tetramer systems, we calculated the percent time the L1/S3 pockets were open in MD simulations using the same geometric criteria established in Ref. (85). In the L1 loops with an initial extended conformation (monomers B and C), we found that the L1/S3 pocket was open only 7% to 15% of the time with only one exception. (Figure 3.5d, Table 3.2) On the other hand, we found that in most of the L1 loops with an initial recessed conformation (monomers A and D), the L1/S3 pocket was open 80% to 99% of the simulation time. (Figure 3.5d, Table 3.2) Much lower L1/S3 pocket-open percentages were computed only for the monomers A and D whose L1 loops spent a significant amount of time sampling the intermediate conformations. (Figure 3.5d, Table 3.2) Overall, we found that a recessed L1 loop conformation correlated with a mostly open L1/S3 pocket while an extended L1 loop conformation correlated with a rarely open L1/S3 pocket. (Figure 3.5d, Table 3.2)

Table 3.2 Percentage of L1/S3 pocket open conformations for each monomer during MD simulations of the three systems.

p21

	monA	monB	monC	monD
Copy1	94.77	7.35	10.06	94.67
Copy2	94.97	23.14	9.05	95.61
Copy3	33.87	6.68	6.04	98.63

puma

	monA	monB	monC	monD
Copy1	81.66	10.08	4.77	99.64
Copy2	96.78	8.08	73.45	96.6
Copy3	99.64	16.07	11.03	97.19

nonspecific DNA

	monA	monB	monC	monD
Copy1	95.9	6.33	11.65	18.62
Copy2	93.19	21.83	7.08	6.35
Copy3	5.54	13.94	24.57	91.78

3.2.6 FTMAP provides insight into druggable pockets found in fl-p53

Besides the L1/S3 pocket, we were also interested in locating novel cryptic druggable sites in fl-p53. The solvent-mapping results from FTMAP identified multiple druggable sites in the DBD as well as in the transactivation (TAD) domain and the tetramerization (TET) domain. The TAD and TET pockets were not consistent because of the high flexibility of these two regions. On the contrary, the druggable pockets on the DBD surface were consistent among all systems and all monomers. The DBD druggable pockets

we identified on the p53 DBD are the L1/S3 pocket, the L1 back pocket, the Tyr220 pocket, the Met160 pocket and the Gln192 pocket. (Figure 3.6) The L1/S3 pocket was previously identified as druggable by us, and can be possibly used by tunneling into the L1 back pocket for drug discovery purposes. The Tyr220 pocket consisted of two pockets with a loop in between them, and was recently shown to bind a ligand that can exploit one of the Tyr220 pockets as well as the transient tunnel between the two pockets(90). The Tyr220 pockets in monomers B and C (the inner monomers) were at the interface of two monomers and may not be easily accessible. (Figure 3.5b) The Gln192 pocket predicted by FTMAP falls between the DBD and the NTD of p53. Thus, this druggable pocket can only be observed in p53 DBDs with a non-truncated N-terminal region as in our fl-p53 systems. Further studies are required to validate these predicted druggable pockets experimentally.

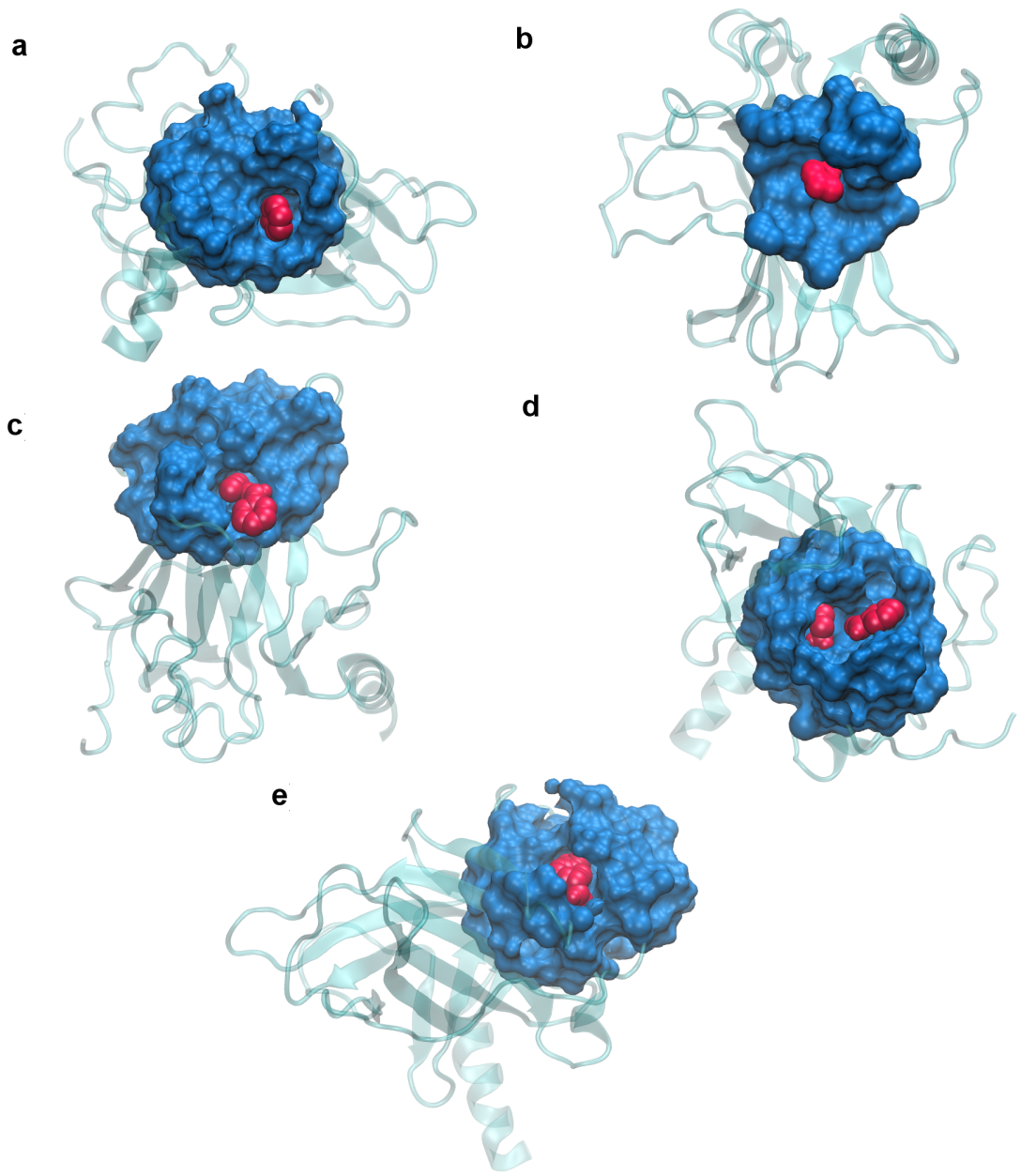


Figure 3.6 Computationally predicted druggable pockets. a) the L1/S3 pocket, b) the L1 loop back pocket, c) the Tyr220 pocket, d) the Met160 pocket, and e) the Gln192 pocket are all drawn in blue surfaces. Small organic probe molecules are shown in each pocket to highlight the cavity.

3.3 Discussion

We report here an all-atom model of the full-length p53 tetramer bound to DNA and its dynamics via simulations that add up to almost 1 μs in total. At the end of the simulations, the resulting fl-p53 structures agree and fit well into the previously determined cryo-EM maps(78). (Figure 3.2a) The closest all-atom model to a full-length p53 tetramer generated in a very recent study included only the residues between 91 and 359 lacking the NTD and CTD regions and was only simulated for approximately 100 ns in total (113). In our simulations, the TET domains moved about 40 Å toward the DBD center of mass, and the extended CTDs became more compact and approached the DNA, while the NTDs remained mostly extended and very dynamic throughout simulations. In every single simulation we performed, the CTDs of the p53 monomers ended up approaching and directly interacting with the DNA phosphates non-specifically via the positively charged residues including Lys372, Lys373, Lys381 and Lys382, which were also implicated to have a role in DNA binding in previous experimental studies (114, 115). Arlt *et al* showed the p53 CTDs to be very flexible by chemical cross-linking followed by mass spectroscopy(114).

They also found CTD Lys381 is capable of cross-linking with TET domain Lys357, suggesting that CTD is in close proximity to the TET domain (114). Their experimental data is in line with our observations. Furthermore,

Friedler *et al*. showed that acetylation at each of these CTD residues would weaken DNA binding(115). Acetylation neutralizes the positively charged CTD residues and hinders the interaction with the negatively charged DNA backbone. Our simulations explicitly demonstrate at the molecular level that CTDs directly interact with the DNA via non-specific electrostatic interactions. Thus, CTD-DNA interactions are very dynamic likely helping the fl-p53 tetramer slide on the DNA in the absence of tight binding of DBDs to DNA.

Through virtual mutations of the initial DNA structure, we also integrated three different DNA sequences, namely p21 RE, puma RE and non-specific DNA, into the fl-p53 tetramer model in order to search for different quaternary DNA binding modes. p53 tetramer is known to bind tightly to the p21 RE, and only weakly to the puma RE (106). Modeling the case with the non-specific DNA sequence, we wanted to explore whether the non-specific quaternary binding mode of the p53 is similar to the case of BamHI, where the non-specific DNA is not fully enclosed by the protein (116). DNA binding affinity of a single p53 DBD monomer is in the micromolar range (106). Under physiological conditions, K_D values of fl-p53 for known REs are in the range of 1.1 to 4.2 nM while K_D values of non-specific DNAs are in the range of 29.8 to 88.6 nM (96). The small affinity differences between specific and non-specific DNAs suggest that binding affinity is not solely sufficient for p53 to recognize its REs in the genome. Other factors such as conformational selection could play a part in DNA recognition.

Using single-molecule imaging tools, the diffusive motion of individual p53 proteins were recently monitored and quantitatively characterized(110). Comparing the diffusive motions of different p53 constructs (namely fl-p53, NTD+DBD+TET and TET+CTD), they revealed that p53 TET+CTD construct can translocate on DNA much faster than fl-p53, while NTD+DBD+TET construct remains immobile due to the absence of CTD (110). Based on these single molecule experiments, a two-state search mechanism was proposed for p53 search on DNA: a search state with mostly nonspecific binding and fast sliding, and an immobile recognition state with sequence-specific binding (110, 117). In the search state, the CTDs accelerate the p53 sliding motion on DNA. And in the recognition state, the p53 DBD binds tightly and sliding is minimal (110). Through principal component analysis (PCA) of p53 DBD tetramer dynamics from MD simulations of all three systems, we identified differential binding modes depending on the DNA sequence. The clamped, symmetrical, cooperative mode uniquely sampled by the DBDs binding to p21 RE can represent the recognition state while the flat, asymmetrical, non-cooperative mode sampled mostly by the DBD binding to non-specific DNA sequence can represent the search state. Including both the positive REs and non-specific DNA sequence in our various simulations and performing a collective PCA, we were able to capture a conformational change motion that is indicative of multiple DNA binding modes of fl-p53, even though the timescale of our MD

simulations are probably not long enough to observe the entire event in a single system.

We also analyzed the conformation and dynamics of the p53 L1 loop that directly interacts with DNA and forms part of the druggable L1/S3 pocket. The first crystal structure of a p53 tetramer bound to a natural RE (pdbID: 3TS8) showed that L1 loops have adopted an extended conformation in the two inner monomers (monomers B and C) and a recessed conformation in the two outer monomers (monomers A and D) (100). Prior to 2011, all p53 L1 loops in the crystal structures were found in extended conformation. Lukman *et al* (111) performed multi-copy 100 ns simulations of a single p53 monomer DBD (without any DNA bound) and observed that the L1 loop samples both the extended and recessed conformations regardless of the starting conformation of the L1 loop. In our 110 ns simulations of the DNA-bound full-length p53 tetramer, we did not observe any full transitions of a recessed L1 loop into an extended L1 loop or vice versa. Especially, the extended L1 loop conformations were persistent in inner monomers B and C of all the simulations. Nevertheless, we have observed flexibility in both conformations with a significantly greater mobility in the recessed L1 loop conformation as shown in Figure 3.5b. When one of the inner p53 monomers with an extended L1 loop conformation is superimposed onto an outer p53 monomer with a recessed L1 loop conformation in the crystal structure, there is a clear steric clash between the DNA and the extended L1 loop. (Supplementary Figure 3.8)

Thus, the outer p53 monomer can't adopt an extended L1 loop conformation when fl-p53 tetramer binds tightly to DNA. Our data and previous data from Lukman *et al (111)* together indicate a conformational selection for the p53 tetramer L1 loops upon DNA binding, dictated by the position of the monomer with respect to DNA. In the inner monomers, L1 loop adopts an extended conformation due to favorable DNA interactions, while in the outer monomers; L1 loop cannot sample extended conformation due to steric clash, but samples the recessed conformation as well as some intermediate ones.

The correlation between the L1 loop dynamics and the behavior of the druggable L1/S3 pocket in the DNA-bound fl-p53 tetramer system is also interesting. Table 2 shows that the L1/S3 pocket is open at least 80% of the time on the two outer monomers, in which L1 loops adopt the recessed conformation. Whereas, the L1/S3 pocket is found open less than 15% of the time on the inner monomers, in which L1 loops strictly stick to the extended conformation. Based on our data, the L1 loop conformations are constrained by DNA binding, and profoundly alter the L1/S3 pocket open percentage. We should also note that the L1/S3 pocket open percentage is much higher compared to the previous values found in single p53 DBD monomer simulations of Wassman et al (85) despite using the exact same set of criteria to define an open pocket. In these previous simulations, the L1/S3 pocket was found to remain open less than 10% of the simulation time (85). The difference

between the two results could be explained by the effect of DNA binding as well as the inclusion of fl-p53 tetramer instead of a single p53 DBD monomer.

3.4 Methods

3.4.1 Model construction

Fl-p53 consists of 393 residues. As the main scaffold, we used the tetrameric p53 crystal structure pdbID:3TS8 (100), which is bound to a p21 RE. This structure includes both the catalytic DBD (residues 94-291) and the TET domain (residues 321-356), but lacks a flexible-linker region (residues 291-321) (100). We modeled the missing linker region using MOE (118) and VMD (62), based on chain A, residues 176-199, of pdbID:1MT6 (119), which the Schrodinger suite identified as having the highest sequence similarity to the linker region in a BLAST search (54). We also modeled residues 91-94 using the crystal structure pdbID:2XWR (120). These residues have recently been identified as components of the core domain.

For each monomer, the N-terminal domain was modeled by superimposing residue 35 of crystal structures pdbID:2K8F (121) (chain B, residues 1-35) and pdbID:2B3G (122) (chain B, residues 35-56), and then connecting the two complexes. Within residues 59-91, residues 66-86 are known to adopt a poly-proline-II (PPII) structure, so we modeled that region appropriately while using an extended conformation for the remainder of the

flexible linker. After modeling the four flexible N-terminal domains and integrating these into the tetrameric p53 model, relative conformations of the N-terminal domains were adjusted by optimizing dihedral angles to prevent steric clashes, using MOE.

For each monomer, the C-terminal domain was modeled by connecting the non-alpha-helical parts of pdbID:1DT7 (123) (residues 367-378) and pdbID:1H26 (124) (residues 378-386). The missing residues 356-367 and 386-393 are modeled in an extended conformation using MOE and integrated with VMD (62). After modeling the four flexible CTDs, relative conformations of these were again adjusted by optimizing the dihedral angles using MOE.

The double-stranded DNA (dsDNA) sequence (p21) was extended on both sides to obtain a 65-nucleotide dsDNA (Figure 3.1b). To explore the effect of binding to different DNA sequences, we modified the DNA sequence manually to a puma RE sequence, as seen in Figure 3.1b. Furthermore, in an attempt to capture the loose p53-DNA binding mode, we used a DNA sequence comprised of the most unlikely nucleotide at each RE position. We obtained this non-specific DNA sequence based on a sequence logo that depicts the frequency of each nucleotide to be at each of the 20 positions of the response element as a result of analyzing 100 known p53 REs (125). (Figure 3.1b).

Once the systems were built, Na^+ ions were added to neutralize each system. In addition to conserving all crystallographic water molecules, a 12 Å

TIP3P (126) water buffer was used to solvate the system explicitly, using the Amber12 suite (127). Zinc and its coordinating residues were modeled using the cationic dummy atom model (128). Each system consisted of 1,592,100 atoms and was built using the Amber FF14SB force field (129).

3.4.2. Molecular Dynamic Simulations

All molecular dynamics (MD) simulations were performed using NAMD2.10 (130). Energy minimization was first performed on the p21 RE, puma RE, and non-specific DNA systems. Each system was restrained, and atom positions were gradually relaxed to allow atomic fluctuations. System relaxation was performed gradually in five steps. In the first 2,000 steps, we constrained all non-hydrogen atoms. In the second 2,000 steps, we constrained the zinc ions, protein, DNA, and non-hydrogen atoms while letting the hydrogen atoms, water molecules, and ions move freely. In the third 2,000 steps, we constrained the zinc ions, protein, and DNA heavy atoms, but set the hydrogen atoms, water molecules, ions, and the zinc-coordinating residues free. During the fourth 10,000 steps, only the protein and DNA backbone were constrained. During the final 20,000 steps, all atoms were set free. The non-bonded energy was calculated at every step. Long-range interactions were calculated using the Particle Mesh Ewald method with a cut-off distance of 10 Å (131). At 8 Å, a switching function was applied to improve energy conservation.

After minimizations, equilibrations were performed on the three systems. Throughout the equilibration, we held the water bonds rigid while slowly decreasing the harmonic constraints on the heavy atoms in 0.25-nanosecond (ns) increments that ultimately totaled 1 ns. Following the equilibrations, an NPT ensemble was performed with no positional constraints. Langevin dynamics kept the temperature constant at 310 K with a gamma value of 5 picoseconds/terahertz. A Langevin piston barostat held the pressure constant at 1 atm with an oscillation period of 100 femtosecond (fs) and a damping time scale of 50 fs. Three independent MD copies were run for each of the three systems, generating a total of nine separate simulations of ~110ns each. In total, we simulated almost 1 μs of DNA-bound fl-p53.

3.4.3 Radius of Gyration

The radius of gyration was calculated using cpptraj, a component of the Amber suite (127). The average radii of gyration with respect to time were calculated for two regions: the fl-p53 Cα atoms and the Cα atoms of the DBD and TET. The results were then plotted using the ggplot2 package (132) in R, shown in Figure 3.2a and Supplementary Figure 3.1.

3.4.4 Principle Component Analysis

Our PCA analysis followed these steps. 1) We concatenated and aligned all the trajectories using the α-carbons of residues 89 to 291 for the DBD PCA analysis, and all α-carbons for the alpha-carbon PCA analysis. 2)

We created a covariance matrix using the α-carbons of the residues of interest. 3) We diagonalized the co-variant matrix to obtain the eigenvectors and the corresponding eigenvalues. 4) We projected the trajectories onto the first and second eigenvectors. 5) We generated pseudo trajectories from the first and second eigenvector to study the motion decomposed by each principal component. The above steps were performed using cpptraj(127). The resulting projections were plotted using gnuplot (133).

3.4.5 Salt Bridge Formation

The positive residues of the fl-p53 and the negative DNA phosphate backbone often formed salt bridges. To inspect these salt bridges, we first generated a list of fl-p53 Lys and Arg residues that were positioned within 5 Å of DNA, using a tool command language (tcl) script executed in VMD. We then loaded the trajectories into VMD and visually identified salt bridges between the DNA and the selected Arg/Lys residues. To quantify the analysis of the salt bridges, we manually extract the distance between the positive nitrogen atom and the negative oxygen atom and used a python script to calculated the percent bond occupancy using a distance cutoff of 3.5 Å (134).

3.4.6 Volume Calculation

The volume between the four DBD monomers was calculated using POVME 2.0.(68) An inclusion sphere centered at Cartesian coordinates [128, 135, 115.5] with a radius of 17 Å fully engulfed the gap between the four

monomers, as shown in Supplementary Figure 3.2. A seed was planted in the center of the sphere and extended for 4 Å. POVME 2.0 calculated the volume starting from the seed and continued until it reached the boundary of the inclusion region. Volumes were calculated for every fifth simulation frame. The resulting volume distribution was plotted in the R program (132) as a histogram.

3.4.7 Hydrogen-Bond Analysis

Using the VMD Hbond plugin, we generated a list of direct hydrogen bonds between the fl-p53 and various DNA sequences, using 3.5Å and 20° distance and angle cut-offs, respectively(62). We further condensed the list by considering only those hydrogen bonds with at least 10% occupancy. A single list was compiled by combining the data from the three copies of each system using a Kepler workflow (135). Finally, we assigned a score to each p53 residue in the list by summing its occupancy with one or multiple DNA residues.

3.4.8 DNA Bending Angle and Properties Analysis

The DNA bending angle was calculated using cpptraj (127). We manually selected phosphate atoms from nucleotides 1653, 1676, and 1697 for angle calculation. The results were plotted with the ggplot2 package in R (112, 132). Additional DNA properties were analyzed using the Canal software for MD trajectories, part of the Curves+ package (112). As we were particularly

interested in how the DNA conformation changed upon fl-p53 binding, we set the search sequence to the four pentamer repeats (highlighted in blue, red, yellow, and green in Figure 3.1b) and calculated only the DNA properties of this region.

3.4.9 L1 Loop Analysis

The L1 loop (residues 113 to 126) was analyzed using a tcl script in VMD(62). First, two monomers were extracted from the crystal structure (PDBID:3TS8 (100)), containing distinct examples of the extended and recessed L1 loop, respectively. They were used as a reference for simulation trajectory alignment and subsequent L1-loop root-mean-square-deviation (RMSD) calculations. The L1 RMSD was calculated twice, with respect to the extended and recessed conformations in the crystal structure, respectively. The atom selections for the alignment and the RMSD calculations were different. The trajectory alignment was performed using the DBD Cα atoms, and the RMSD calculations were performed using the L1 loop Cα atoms. The results were plotted using gnuplot (133). Over the course of the simulations, the L1 loop was said to have adopted the recessed or the extended conformation if the extended and recessed RMSD values did not show overlap. (Supplementary Figure 3.5) If the two RMSD values overlapped for more than 20 ns, the L1 loop was said to have adopted an intermediate structure. Furthermore, we were interested in visualizing the conformational space that the L1 loop sampled. To this end, we used cpptraj to perform

RMSD clustering using a hierarchical agglomerative (bottom-up) approach (127). We generated five clusters and extracted the centroid frame using VMD (62).

3.4.10 C124 Pocket Analysis and Pocket Prediction

Using the distance and angle criteria of Wassman *et al.* (85), we calculated the open C124 pocket percentage with respect to time. There are two main steps in this process: input generation and the actual calculation. We first used cpptraj (127) to generate the four distances and one dihedral angle that serve as inputs for the calculation. We then wrote a python script to pinpoint the frames that satisfy these distance and angle criteria, from which the L1/S3 pocket opening percentage was calculated. The results were plotted using basic plotting in R (136). To identify druggable pockets in the p53 protein, we submitted the fl-p53 monomers from the final frames of the simulations to the FTMap web server (137). FTMAP server floods protein surfaces with various small solvent molecules to find druggable hot spots (137). We then loaded the collective results into VMD and aligned them to the whole fl-p53 system in order to locate each predicted druggable pocket.

3.4.11 Fitting the Density Map Generated from MD Trajectories into the p53 EM Maps

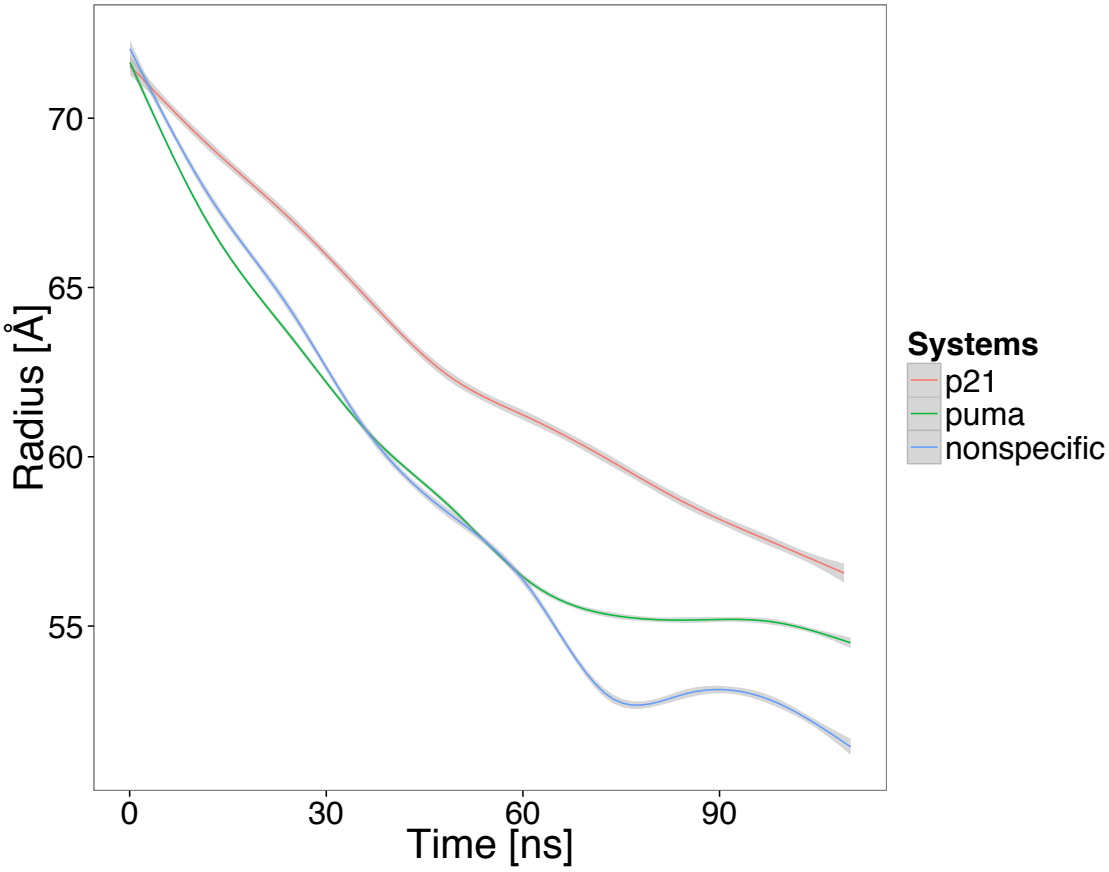
Density maps of the fl-p53 bound to three different DNA sequences were generated using MDFF from the second half (55 ns to 110 ns) of the trajectories (138). The mdff sim function was utilized and the resolution and

spacing were set to 30.0 Å and 2.2 Å, respectively, according to experimental data (78). Chimera program is used to fit the ensemble-averaged density maps from p21-RE-bound, puma-RE-bound and nonspecific-DNA-bound p53 tetramers into each of the four EM maps in Melero *et al*(53, 78). The best correlation is consistently obtained while fitting into the class III EM map.

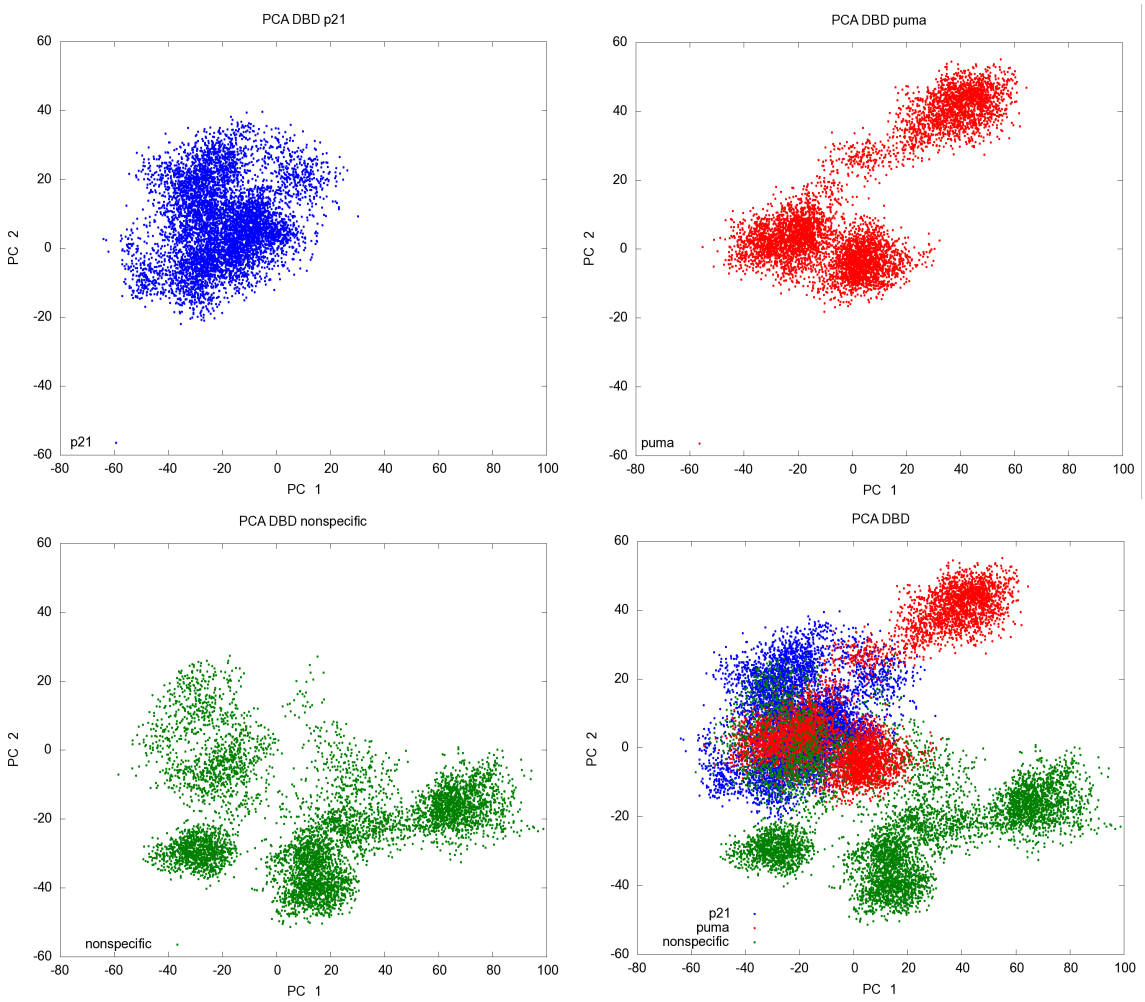
3.4.12 Ensemble averaged electrostatic map calculation

The fl-p53 ensemble averaged electrostatic maps of the three systems were calculated with DelPhi Ensemble Electrostatics. Each ensemble was comprised of 30 trajectories from 80 ns to 110ns (139). The calculations were performed under zero salt concentration, and using 2.0 and 80 as the solute and solvent dielectrics.

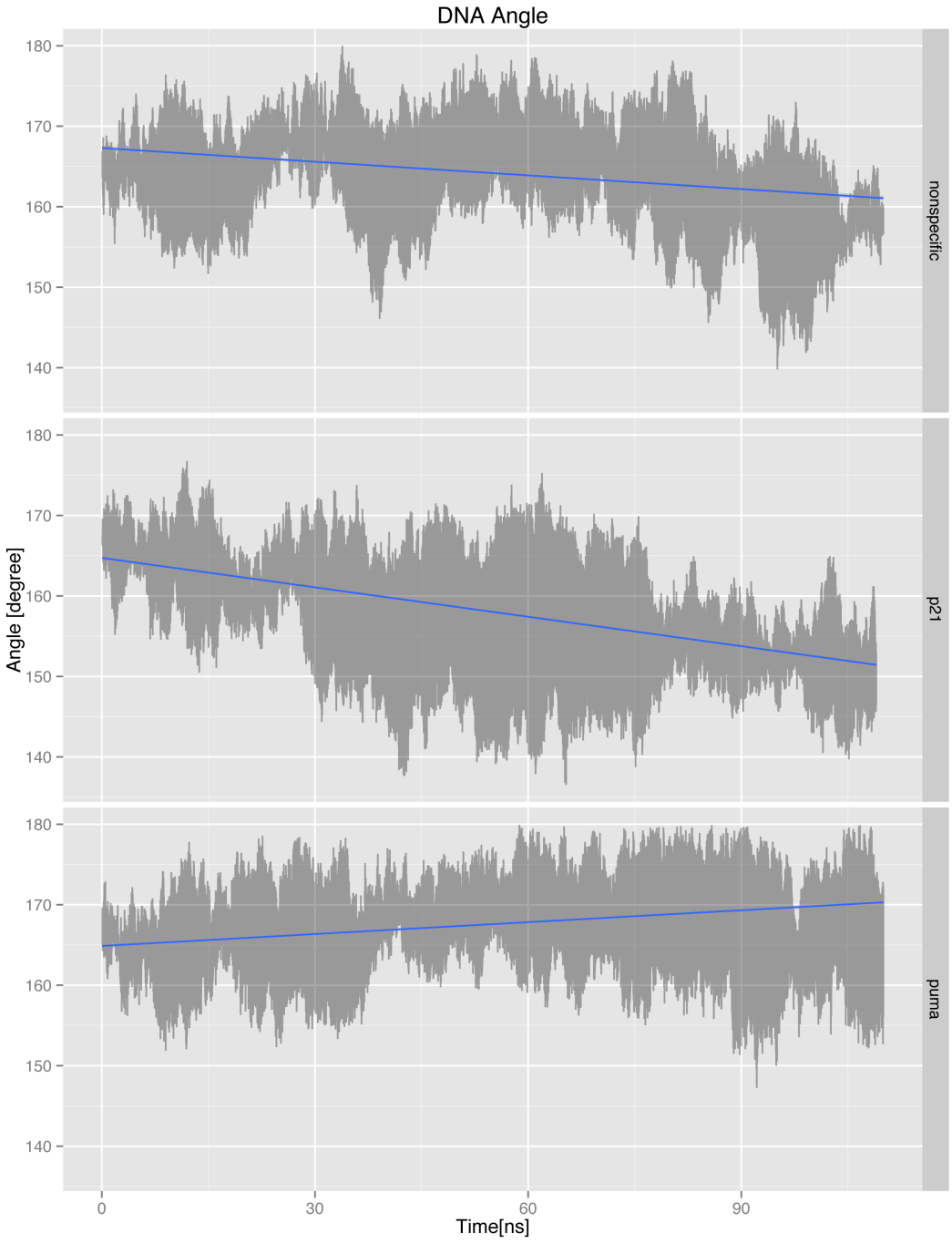
3.5 Appendices



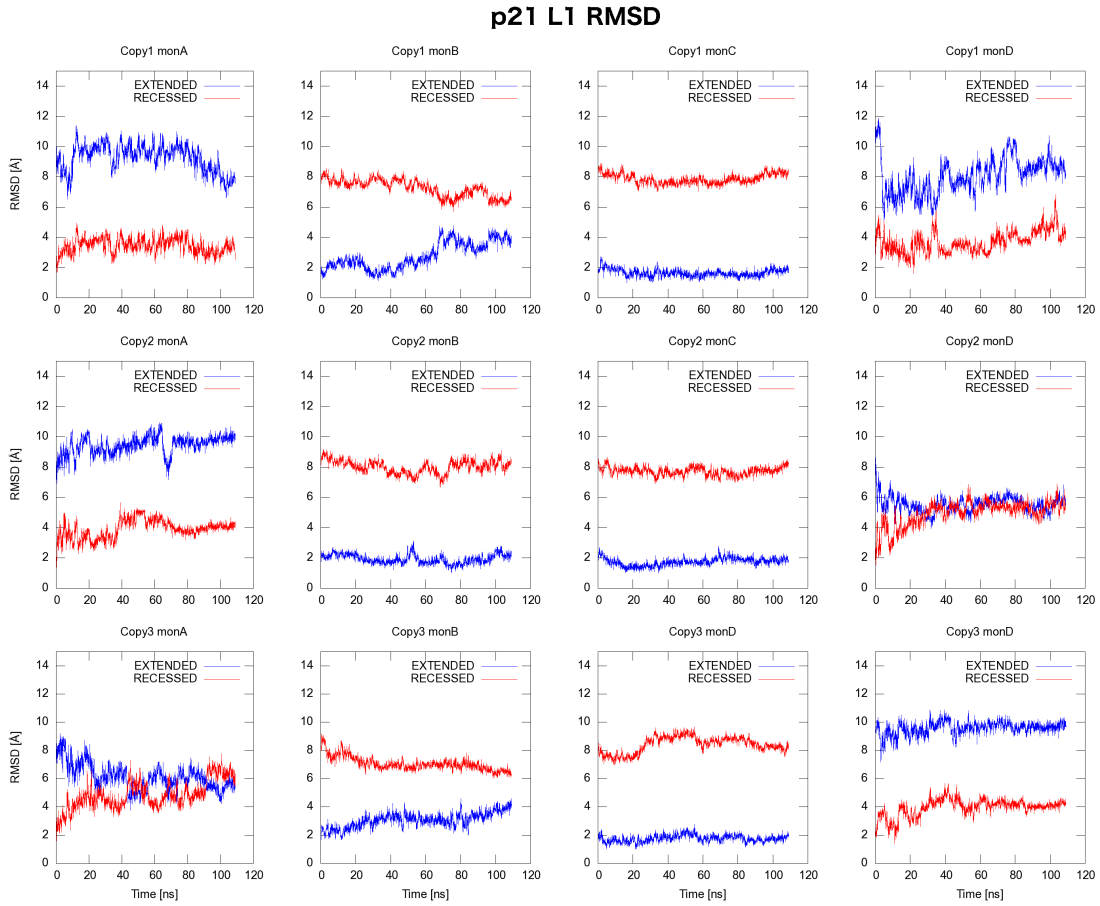
Supporting Figure 3.1 Time evolution of the radius of gyration for the full-length p53.



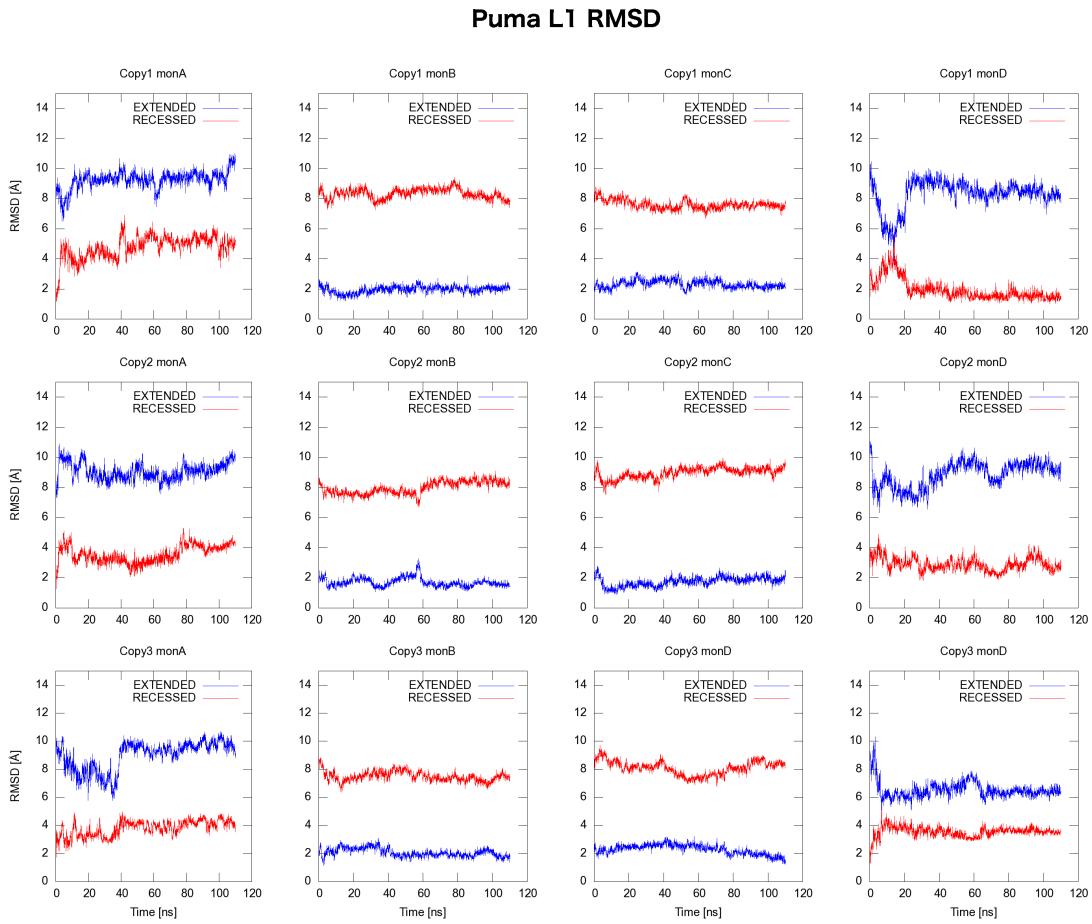
Supporting Figure 3.2 Principle component analysis of the p53 DBD tetramer for p21, puma and nonspecific systems. Panel on the top left corner shows the PCA of the p21 system DBD. Panel on the top right corner shows the PCA of the puma system DBD. Panel on the bottom left shows the PCA of the nonspecific system DBD. Panel on the bottom right shows the combination of the other three panels. PC1 and PC2 are labeled in the x-axis and y-axis, respectively.



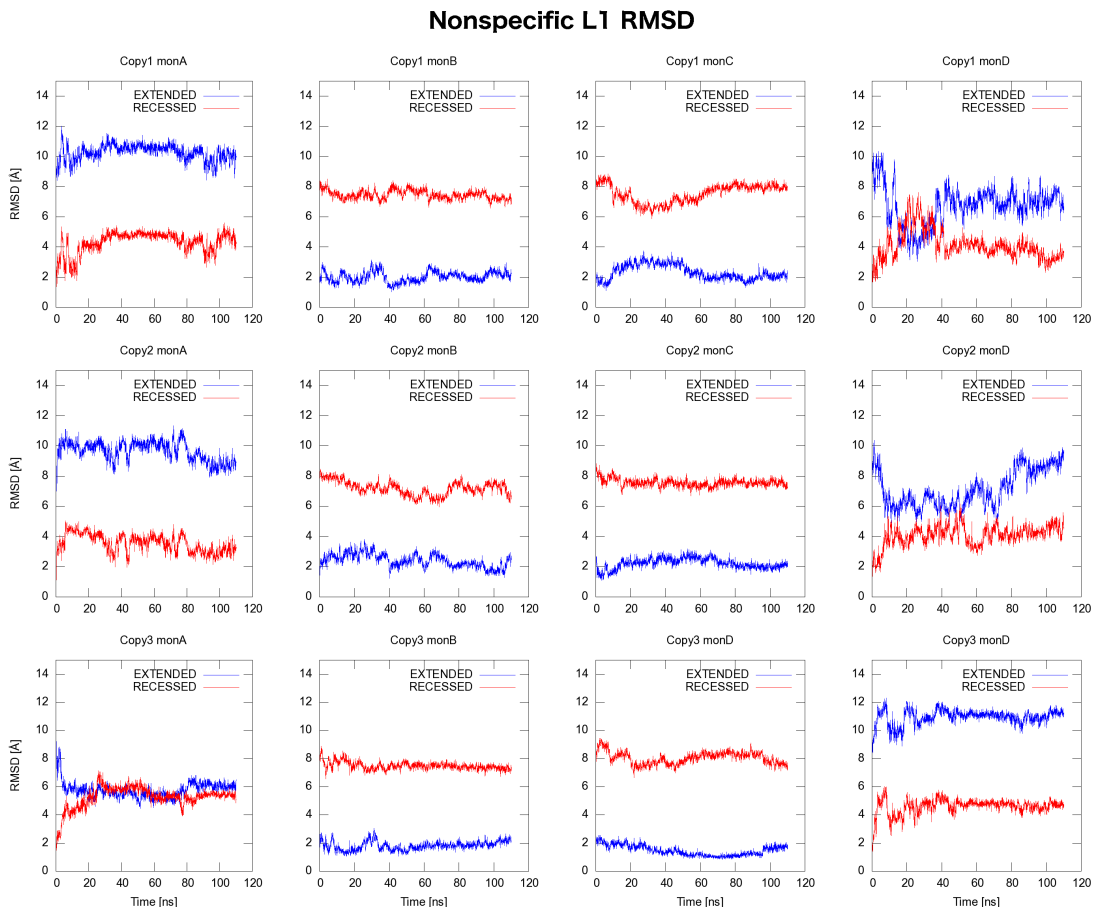
Supporting Figure 3.3 Time evolution of the DNA bending angle in the three systems. The DNA bending angles of the three copies are shown as lines in dark grey. Linear regression lines are drawn in blue to show the bending trend for each system.



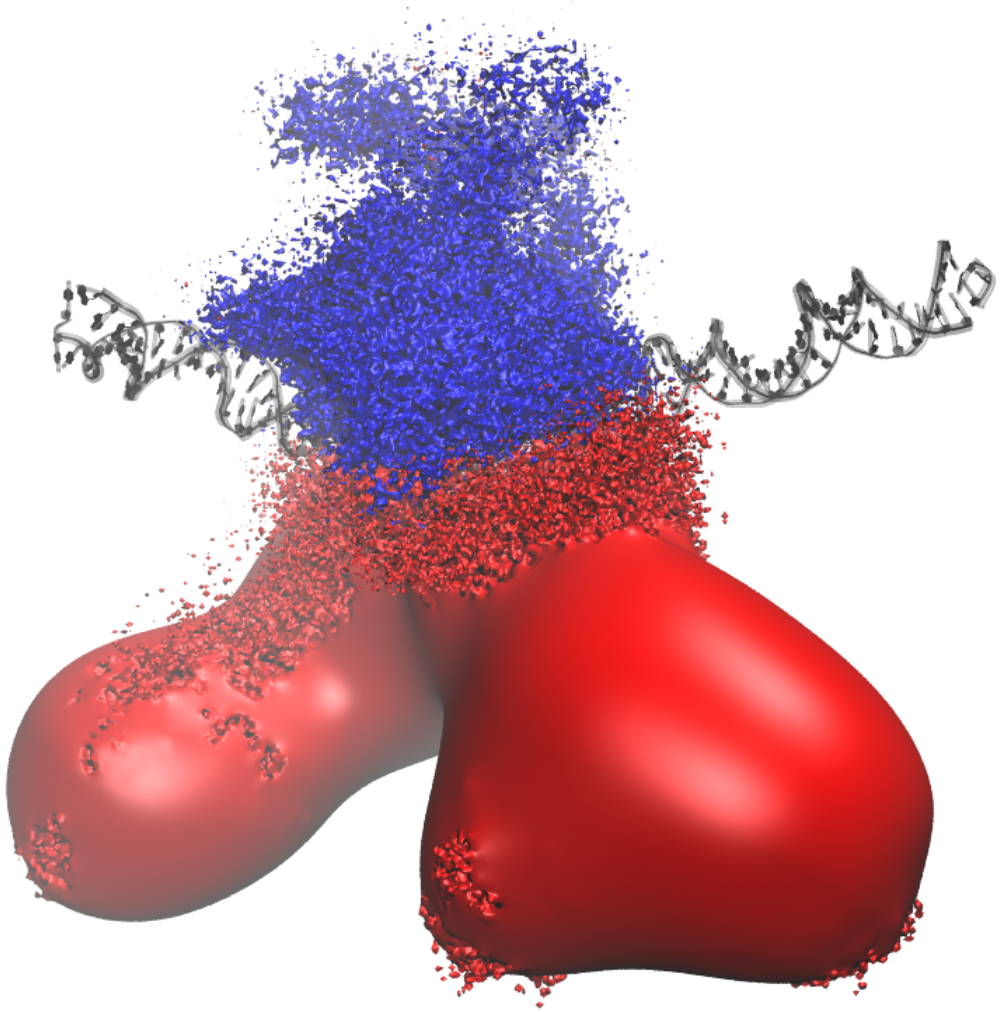
Supporting Figure 3.4 L1 loop RMSD with respect to extended and recessed loop conformations in the p21 system. Time evolution of the rmsd of the L1 loop (of each p53 monomer in each MD copy) calculated with respect to both the extended and the recessed L1 loop conformations. The rmsd values calculated with respect to the extended and recessed L1 conformations are colored in blue and red, respectively.



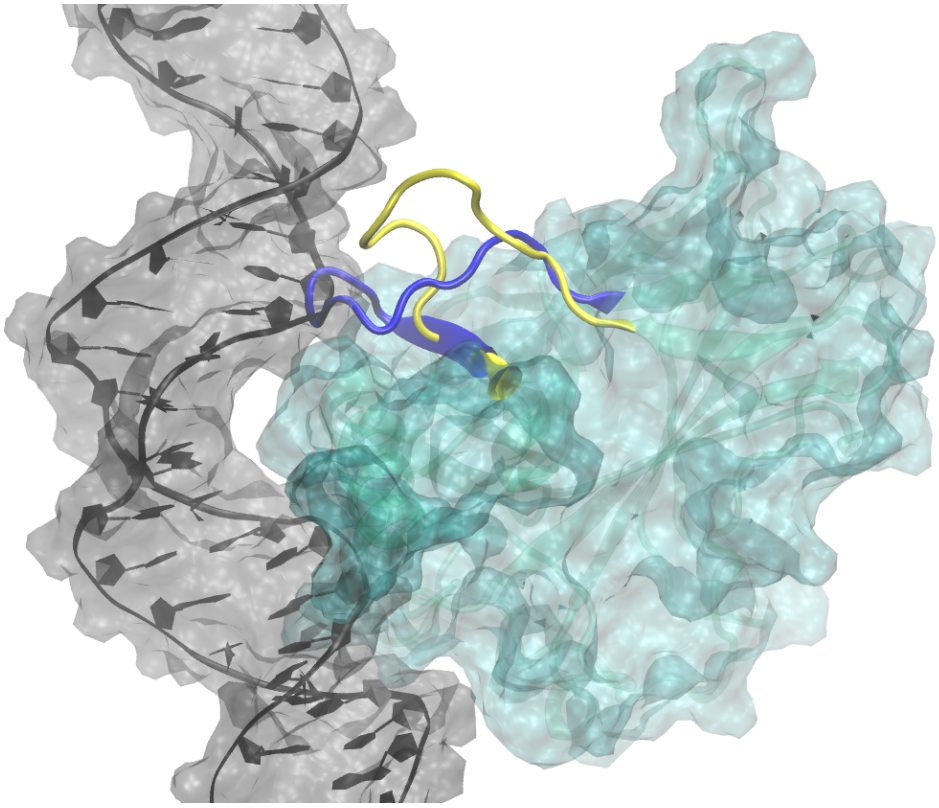
Supporting Figure 3.5 L1 loop RMSD with respect to extended and recessed loop conformations in the puma system. Time evolution of the rmsd of the L1 loop (of each p53 monomer in each MD copy) calculated with respect to both the extended and the recessed L1 loop conformations. The rmsd values calculated with respect to the extended and recessed L1 conformations are colored in blue and red, respectively.



Supporting Figure 3.6 L1 loop RMSD with respect to extended and recessed loop conformations in the nonspecific DNA system. Time evolution of the rmsd of the L1 loop (of each p53 monomer in each MD copy) calculated with respect to both the extended and the recessed L1 loop conformations. The rmsd values calculated with respect to the extended and recessed L1 conformations are colored in blue and red, respectively.



Supporting Figure 3.7 Ensemble averaged electrostatic map of the fl-p53 protein. Negative electrostatic (-4 kT/e) isosurface is colored in red and positive electrostatic (+4 kT/e) isosurface is shown in blue. DNA is drawn in black as a reference.



Supporting Figure 3.8 Extended L1 Loop steric clash with the DNA in the outer monomer. The extended L1 loop conformation is superimposed onto an outer DBD monomer, which usually adopted a recessed L1 loop conformation. Only one monomer surface is shown in cyan. The native recessed L1 conformation is colored yellow while the superimposed extended L1 conformation is colored blue. The DNA is drawn in gray surface and the backbone is shown in ribbon representation. The blue extended L1 conformation overlaps with the DNA as shown in the above figure, indicating an unfavorable steric clash that will force the L1 loop to a recessed conformation.

Supporting Table 3.1 Salt bridge footprint analysis. This table lists the percentage of salt bridge interactions between the C-terminal residues and the DNA in each MD copy for the three systems. The residues highlighted in blue, green and orange are from monomers A, B and C, respectively. The DNA nucleotides that each p53 residue interacted with are shown in the DNA counterpart column.

copy1	% interaction	DNA counterpart	copy2	% interaction	DNA counterpart	copy3	% interaction	DNA counterpart
p21								
370	19.3%	DT1686, DT1687						
373	10.5%	DT1687, DT1686						
379	29.8%	DA1689						
381	25.0%	DA1689						
382	17.1%	DT1687						
puma								
370	7.1%	DA1591, DC1592	363	6.9%	DA1623	370	27.7%	DC1685, DT1686
372	21.5%	DT1686				373	10.5%	DC1685, DG1684
373	11.3%	DT1687				379	4.5%	DC1685
379	30.5%, 18.6%	DA1590, DC1589				381	19.2%	DC1685, DT1686
381	23.8%	DA1697						
382	21.8%	DA1591						
nonspecific								
			379	42.6%	DT1600	382	12.3%	DT1665
			381	50.8%	DT1600, DG1599	386	19.3%	DA1622, DT1666
			382	33.9%	DG1599	370	18.3%	DA1622, DA1621
			386	16.6%, 22.7%	DG1679, DG1678	372	6.1%	DT1666
						373	6.2%	DG1656
						363	8.4%	DA1630
						381	25.3%	DT1688
						382	44.2%	DT1600
						386	17.2%	DT1688, DG1598

Supporting Table 3.2 Hydrogen bond footprint analysis between the DNA and the fl-p53 tetramer. The average scores of the key protein residues in the four p53 monomers in each system are shown.

monomer	ResID	p21	puma	nonspecific
A	Ala276	43.0	18.7	0.0
B	Ala276	47.8	0.0	7.4
C	Ala276	60.1	49.1	22.0
D	Ala276	28.7	35.4	8.4
A	Arg280	44.3	37.1	9.1
B	Arg280	56.7	40.1	0.0
C	Arg280	69.9	44.9	16.9
D	Arg280	35.0	48.0	11.5
A	Ser241	61.3	67.9	60.9
B	Ser241	37.4	55.5	33.9
C	Ser241	62.7	69.8	40.8
D	Ser241	49.9	51.5	39.7
A	Arg273	43.8	60.1	26.5
B	Arg273	43.5	32.2	36.0
C	Arg273	52.9	47.5	36.5
D	Arg273	53.1	41.1	21.6
A	Asn239	39.2	23.6	4.7
B	Asn239	42.0	19.2	15.0
C	Asn239	46.6	36.5	25.8
D	Asn239	29.7	15.8	6.4
A	Ser121	26.6	32.4	17.4
B	Ser121	60.1	90.5	29.0
C	Ser121	42.9	28.3	91.7
D	Ser121	19.3	16.2	15.5
A	Lys120	19.3	9.7	0.0
B	Lys120	62.6	5.5	21.1
C	Lys120	61.9	48.1	5.1
D	Lys120	0.0	16.0	0.0
A	Arg248	23.8	28.9	9.7
B	Arg248	13.5	7.2	9.8
C	Arg248	17.6	40.9	4.8
D	Arg248	16.0	14.5	38.3
A	Thr123	8.0	35.8	0.0
B	Thr123	0.0	0.0	0.0
C	Thr123	0.0	0.0	0.0
D	Thr123	0.0	22.1	29.2
A	Asn288	13.8	20.4	10.8
B	Asn288	3.5	22.7	15.9
C	Asn288	5.3	12.1	3.4
D	Asn288	0.0	22.6	7.8

Supporting Movie 3.1 PC1 motion going from min to max. Only p53 DBDs and DNA is depicted.

Supporting Movie 3.2 PC2 motion going from min to max. Only p53 DBDs and DNA is depicted.

This chaper, in full, has been submitted for publication of the material as it may appear in “Full-length p53 Tetramer Bound to DNA and Its Quaternary Dynamics” by Demir, Özlem; Ieong, Pek U; Amaro, Rommie E., submitted to PNAS in 2016. This chapter is included with the permission from Demir, Özlem and Amaro, Rommie E.

References

1. Altintas, I., C. Berkley, E. Jaeger, M. Jones, B. Ludascher, and S. Mock. 2004. Kepler: An extensible system for design and execution of scientific workflows. 16th International Conference on Scientific and Statistical Database Management, Proceedings:423-424.

2. McPhillips, T., S. Bowers, D. Zinn, and B. Ludascher. 2009. Scientific workflow design for mere mortals. Future Gener Comp Sy 25:541-551.

3. Amaro, R. E., R. Baron, and J. A. McCammon. 2008. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. J Comput Aid Mol Des 22:693-705.

4. Amaro, R. E., and W. W. Li. 2010. Emerging Methods for Ensemble-Based Virtual Screening. Curr Top Med Chem 10:3-13.

5. Murdock, S. E., K. Tai, M. H. Ng, S. Johnston, B. Wu, H. Fangohr, C. A. Laughton, J. W. Essex, and M. S. P. Sansom. 2006. Quality assurance for biomolecular simulations. J Chem Theory Comput 2:1477-1481.

6. Wang, J. M., W. Wang, P. A. Kollman, and D. A. Case. 2006. Automatic atom type and bond type perception in molecular mechanical calculations. J Mol Graph Model 25:247-260.

7. Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. 2009. Gaussian 09. Gaussian, Inc., Wallingford, CT, USA.

8. Wang, J. M., R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. 2004. Development and testing of a general amber force field. J Comput Chem 25:1157-1174.

9. Bayly, C. I., P. Cieplak, W. D. Cornell, and P. A. Kollman. 1993. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. J Phys Chem-Us 97:10269-10280.

10. Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten. 2005. Scalable Molecular Dynamics with NAMD. J. Comput. Chem. 26:1781-1802.

11. Amaro, R. E., A. Schnaufer, H. Interthal, W. Hol, K. D. Stuart, and J. A. McCammon. 2008. Discovery of drug-like inhibitors of an essential RNA-editing ligase in Trypanosoma brucei. P Natl Acad Sci USA 105:17278-17283.

12. Amaro, R. E., D. D. Minh, L. S. Cheng, W. M. Lindstrom, Jr., A. J. Olson, J. H. Lin, W. W. Li, and J. A. McCammon. 2007. Remarkable loop flexibility in avian influenza N1 and its implications for antiviral drug design. J Am Chem Soc 129:7764-7765.

13. Landon, M. R., R. E. Amaro, R. Baron, C. H. Ngan, D. Ozonoff, J. A. McCammon, and S. Vajda. 2008. Novel druggable hot spots in avian influenza neuraminidase H5N1 revealed by computational solvent mapping of a reduced and representative receptor ensemble. Chem Biol Drug Des 71:106-116.

14. Christen, M., P. H. Hunenberger, D. Bakowies, R. Baron, R. Burgi, D. P. Geerke, T. N. Heinz, M. A. Kastenholz, V. Krautler, C. Oostenbrink, C. Peter, D. Trzesniak, and W. F. van Gunsteren. 2005. The GROMOS software for biomolecular simulation: GROMOS05. J Comput Chem 26:1719-1751.

15. Daura, X., W. F. van Gunsteren, and A. E. Mark. 1999. Folding-unfolding thermodynamics of a beta-heptapeptide from equilibrium simulations. Proteins 34:269-280.

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

17. Roberts, E., J. Eargle, D. Wright, and Z. Luthey-Schulten. 2006. MultiSeq: unifying sequence and structure data for evolutionary analysis. BMC Bioinformatics 7:382.

18. Morris, G. M., R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, and A. J. Olson. 2009. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem 30:2785-2791.

19. Trott, O., and A. J. Olson. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31:455-461.

20. Truchon, J. F., and C. I. Bayly. 2007. Evaluating virtual screening methods: good and bad metrics for the "early recognition" problem. J Chem Inf Model 47:488-508.

21. Zhao, W., K. E. Hevener, S. W. White, R. E. Lee, and J. M. Boyett. 2009. A statistical framework to evaluate virtual screening. BMC Bioinformatics 10:225.

22. Sheridan, R. P., S. B. Singh, E. M. Fluder, and S. K. Kearsley. 2001. Protocols for Bridging the Peptide to Nonpeptide Gap in Topological Similarity Searches. J Chem Inf Comput Sci 41:1395-1406.

23. Fawcett, T. 2006. An introduction to ROC analysis. Pattern Recogn Lett 27:861-874.

24. Nicholls, A. 2011. What Do We Know?: Simple Statistical Techniques that Help. In Chemoinformatics and Computational Chemical Biology. J. Bajorath, editor. Humana Press. 531-581.

25. Jain, A. N. 2008. Bias, reporting, and sharing: computational evaluations of docking methods. J Comput Aided Mol Des 22:201-212.

26. 2011. Matlab. The MathWorks Inc., Natick, Massachusetts.

27. Krishnan, S., L. Clementi, R. Jingyuan, P. Papadopoulos, and W. Li. 2009. Design and Evaluation of Opal2: A Toolkit for Scientific Software as a Service. In Services - I, 2009 World Conference on. 709-716.

28. Krishnan, S., B. Stearn, B. Karan, K. K. Baldridge, W. W. Li, and P. Arzberger. 2006. Opal: SimpleWeb Services Wrappers for Scientific Applications. In Web Services, 2006. ICWS '06. International Conference on. 823-832.

29. 2014. Github. https://github.com.

30. Bruno, G., M. J. Katz, F. D. Sacerdoti, and P. M. Papadopoulos. 2004. Rolls: modifying a standard system installer to support user-customizable cluster frontend appliances. In Cluster Computing, 2004 IEEE International Conference on. 421-430.

31. Rossman, J. S., and R. A. Lamb. 2011. Influenza virus assembly and budding. Virology 411:229-236.

32. Nicholls, J. M., R. W. Chan, R. J. Russell, G. M. Air, and J. S. Peiris. 2008. Evolving complexities of influenza virus and its receptors. Trends Microbiol 16:149-157.

33. Schwarzer, J., E. Rapp, R. Hennig, Y. Genzel, I. Jordan, V. Sandig, and U. Reichl. 2009. Glycan analysis in cell culture-based influenza vaccine production: influence of host cell line and virus strain on the glycosylation pattern of viral hemagglutinin. Vaccine 27:4325-4336.

34. Vanderlinden, E., and L. Naesens. 2014. Emerging antiviral strategies to interfere with influenza virus entry. Med Res Rev 34:301-339.

35. Wang, C. C., J. R. Chen, Y. C. Tseng, C. H. Hsu, Y. F. Hung, S. W. Chen, C. M. Chen, K. H. Khoo, T. J. Cheng, Y. S. Cheng, J. T. Jan, C. Y. Wu, C. Ma, and C. H. Wong. 2009. Glycans on influenza hemagglutinin affect receptor binding and immune response. Proc Natl Acad Sci U S A 106:18137-18142.

36. Lachmann, P. 2009. Anti-infective antibodies--reviving an old paradigm. Vaccine 27 Suppl 6:G33-37.

37. Townsend, K. A., and L. S. Eiland. 2006. Combating influenza with antiviral therapy in the pediatric population. Pharmacotherapy 26:95-103.

38. Wrammert, J., K. Smith, J. Miller, W. A. Langley, K. Kokko, C. Larsen, N. Y. Zheng, I. Mays, L. Garman, C. Helms, J. James, G. M. Air, J. D. Capra, R. Ahmed, and P. C. Wilson. 2008. Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. Nature 453:667-671.

39. Skehel, J. J., and D. C. Wiley. 2000. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu Rev Biochem 69:531-569.

40. Gerhard, W., J. Yewdell, M. E. Frankel, and R. Webster. 1981. Antigenic structure of influenza virus haemagglutinin defined by hybridoma antibodies. Nature 290:713-717.

41. Igarashi, M., K. Ito, R. Yoshida, D. Tomabechi, H. Kida, and A. Takada. 2010. Predicting the antigenic structure of the pandemic (H1N1) 2009 influenza virus hemagglutinin. PLoS One 5:e8553.

42. Xu, R., D. C. Ekiert, J. C. Krause, R. Hai, J. E. Crowe, Jr., and I. A. Wilson. 2010. Structural basis of preexisting immunity to the 2009 H1N1 pandemic influenza virus. Science 328:357-360.

43. Ekiert, D. C., R. H. Friesen, G. Bhabha, T. Kwaks, M. Jongeneelen, W. Yu, C. Ophorst, F. Cox, H. J. Korse, B. Brandenburg, R. Vogels, J. P. Brakenhoff, R. Kompier, M. H. Koldijk, L. A. Cornelissen, L. L. Poon, M. Peiris, W. Koudstaal, I. A. Wilson, and J. Goudsmit. 2011. A highly conserved neutralizing epitope on group 2 influenza A viruses. Science 333:843-850.

44. Krause, J. C., T. M. Tumpey, C. J. Huffman, P. A. McGraw, M. B. Pearce, T. Tsibane, R. Hai, C. F. Basler, and J. E. Crowe, Jr. 2010. Naturally occurring human monoclonal antibodies neutralize both 1918 and 2009 pandemic influenza A (H1N1) viruses. J Virol 84:3127-3130.

45. Wrammert, J., D. Koutsonanos, G. M. Li, S. Edupuganti, J. Sui, M. Morrissey, M. McCausland, I. Skountzou, M. Hornig, W. I. Lipkin, A. Mehta, B. Razavi, C. Del Rio, N. Y. Zheng, J. H. Lee, M. Huang, Z. Ali, K. Kaur, S. Andrews, R. R. Amara, Y. Wang, S. R. Das, C. D. O'Donnell, J. W. Yewdell, K. Subbarao, W. A. Marasco, M. J. Mulligan, R. Compans, R. Ahmed, and P. C. Wilson. 2011. Broadly cross-reactive antibodies dominate the human B cell response against 2009 pandemic H1N1 influenza virus infection. J Exp Med 208:181-193.

46. Ping, J., L. Keleta, N. E. Forbes, S. Dankar, W. Stecho, S. Tyler, Y. Zhou, L. Babiuk, H. Weingartl, R. A. Halpin, A. Boyne, J. Bera, J. Hostetler, N. B. Fedorova, K. Proudfoot, D. A. Katzel, T. B. Stockwell, E. Ghedin, D. J. Spiro, and E. G. Brown. 2011. Genomic and protein structural maps of adaptive evolution of human influenza A virus to increased virulence in the mouse. PLoS One 6:e21740.

47. Khiabanian, H., V. Trifonov, and R. Rabadan. 2009. Reassortment patterns in Swine influenza viruses. PLoS One 4:e7366.

48. 2009. World Health Organization, http://www.who.int/mediacentre/news/statements/2009/h1n1_pandemic_phase6_20090611/en/index.html.

49. Hancock, K., V. Veguilla, X. Lu, W. Zhong, E. N. Butler, H. Sun, F. Liu, L. Dong, J. R. DeVos, P. M. Gargiullo, T. L. Brammer, N. J. Cox, T. M. Tumpey, and J. M. Katz. 2009. Cross-reactive antibody responses to the 2009 pandemic H1N1 influenza virus. N Engl J Med 361:1945-1952.

50. Itoh, Y., K. Shinya, M. Kiso, T. Watanabe, Y. Sakoda, M. Hatta, Y. Muramoto, D. Tamura, Y. Sakai-Tagawa, T. Noda, S. Sakabe, M. Imai, Y. Hatta, S. Watanabe, C. Li, S. Yamada, K. Fujii, S. Murakami, H. Imai, S. Kakugawa, M. Ito, R. Takano, K. Iwatsuki-Horimoto, M. Shimojima, T. Horimoto, H. Goto, K. Takahashi, A. Makino, H. Ishigaki, M. Nakayama, M. Okamatsu, K. Takahashi, D. Warshauer, P. A. Shult, R. Saito, H. Suzuki, Y. Furuta, M. Yamashita, K. Mitamura, K. Nakano, M. Nakamura, R. Brockman-Schneider, H. Mitamura, M. Yamazaki, N. Sugaya, M. Suresh, M. Ozawa, G. Neumann, J. Gern, H. Kida, K. Ogasawara, and Y. Kawaoka. 2009. In vitro and in vivo characterization of new swine-origin H1N1 influenza viruses. Nature 460:1021-1025.

51. Liu, Q., S. C. Hoi, C. T. Su, Z. Li, C. K. Kwoh, L. Wong, and J. Li. 2011. Structural analysis of the hot spots in the binding between H1N1 HA and the 2D1 antibody: do mutations of H1N1 from 1918 to 2009 affect much on this binding? Bioinformatics 27:2529-2536.

52. Whittle, J. R., R. Zhang, S. Khurana, L. R. King, J. Manischewitz, H. Golding, P. R. Dormitzer, B. F. Haynes, E. B. Walter, M. A. Moody, T. B. Kepler, H. X. Liao, and S. C. Harrison. 2011. Broadly neutralizing human antibody that recognizes the receptor-binding pocket of influenza virus hemagglutinin. Proc Natl Acad Sci U S A 108:14216-14221.

53. Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25:1605-1612.

54. 2015-4, S. R. 2015. Maestro. Schrödinger, LLC, New York, NY.

55. Hornak, V., R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. 2006. Comparison of multiple Amber force fields and

development of improved protein backbone parameters. Proteins 65:712-725.

56. Li, H., A. D. Robertson, and J. H. Jensen. 2005. Very fast empirical prediction and rationalization of protein pKa values. Proteins 61:704-721.

57. Dolinsky, T. J., P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe, and N. A. Baker. 2007. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res 35:W522-525.

58. William L. Jorgensen, J. C., Jeffry D. Madura, Roger W. Impey and Michael L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. J Chem Phys 79.

59. Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. 2005. Scalable molecular dynamics with NAMD. J Comput Chem 26:1781-1802.

60. Barlow, D. J., and J. M. Thornton. 1983. Ion-pairs in proteins. J Mol Biol 168:867-885.

61. Marqusee, S., and R. L. Baldwin. 1987. Helix stabilization by Glu-...Lys+ salt bridges in short peptides of de novo design. Proc Natl Acad Sci U S A 84:8898-8902.

62. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. J Mol Graph 14:33-38, 27-38.

63. Miller, B. R., 3rd, T. D. McGee, Jr., J. M. Swails, N. Homeyer, H. Gohlke, and A. E. Roitberg. 2012. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. J Chem Theory Comput 8:3314-3321.

64. Case, D. A., T. E. Cheatham, 3rd, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods. 2005. The Amber biomolecular simulation programs. J Comput Chem 26:1668-1688.

65. Onufriev, A., D. Bashford, and D. A. Case. 2004. Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins 55:383-394.

66. Amaro, R. E., X. Cheng, I. Ivanov, D. Xu, and J. A. McCammon. 2009. Characterizing loop dynamics and ligand recognition in human- and avian-type influenza neuraminidases via generalized born molecular dynamics and end-point free energy calculations. J Am Chem Soc 131:4702-4709.

67. Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson, and D. G. Higgins. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539.

68. Durrant, J. D., L. Votapka, J. Sorensen, and R. E. Amaro. 2014. POVME 2.0: An Enhanced Tool for Determining Pocket Shape and Volume Characteristics. J Chem Theory Comput 10:5047-5056.

69. Binley, J. M., Y. E. Ban, E. T. Crooks, D. Eggink, K. Osawa, W. R. Schief, and R. W. Sanders. 2010. Role of complex carbohydrates in human immunodeficiency virus type 1 infection and resistance to antibody neutralization. J Virol 84:5637-5655.

70. Feig, M., A. Onufriev, M. S. Lee, W. Im, D. A. Case, and C. L. Brooks, 3rd. 2004. Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. J Comput Chem 25:265-284.

71. Hou, T., J. Wang, Y. Li, and W. Wang. 2011. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. J Chem Inf Model 51:69-82.

72. Xia, Z., T. Huynh, S. G. Kang, and R. Zhou. 2012. Free-energy simulations reveal that both hydrophobic and polar interactions are important for influenza hemagglutinin antibody binding. Biophys J 102:1453-1461.

73. Gouda, H., I. D. Kuntz, D. A. Case, and P. A. Kollman. 2003. Free energy calculations for theophylline binding to an RNA aptamer: Comparison of MM-PBSA and thermodynamic integration methods. Biopolymers 68:16-34.

74. Huo, S., I. Massova, and P. A. Kollman. 2002. Computational alanine scanning of the 1:1 human growth hormone-receptor complex. J Comput Chem 23:15-27.

75. Corrada, D., and G. Colombo. 2013. Energetic and dynamic aspects of the affinity maturation process: characterizing improved variants from the bevacizumab antibody with molecular simulations. J Chem Inf Model 53:2937-2950.

76. Xu, D., E. I. Newhouse, R. E. Amaro, H. C. Pao, L. S. Cheng, P. R. Markwick, J. A. McCammon, W. W. Li, and P. W. Arzberger. 2009. Distinct glycan topology for avian and human sialopentasaccharide receptor analogues upon binding different hemagglutinins: a molecular dynamics perspective. J Mol Biol 387:465-491.

77. Gohlke, H., and D. A. Case. 2004. Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. J Comput Chem 25:238-250.

78. Melero, R., S. Rajagopalan, M. Lazaro, A. C. Joerger, T. Brandt, D. B. Veprintsev, G. Lasso, D. Gil, S. H. Scheres, J. M. Carazo, A. R. Fersht, and M. Valle. 2011. Electron microscopy studies on the quaternary structure of p53 reveal different binding modes for p53 tetramers in complex with DNA. Proc Natl Acad Sci U S A 108:557-562.

79. Vogelstein, B., D. Lane, and A. J. Levine. 2000. Surfing the p53 network. Nature 408:307-310.

80. Vousden, K. H., and X. Lu. 2002. Live or let die: the cell's response to p53. Nat Rev Cancer 2:594-604.

81. Green, D. R., and G. Kroemer. 2009. Cytoplasmic functions of the tumour suppressor p53. Nature 458:1127-1130.

82. Li, T., N. Kon, L. Jiang, M. Tan, T. Ludwig, Y. Zhao, R. Baer, and W. Gu. 2012. Tumor suppression in the absence of p53-mediated cell-cycle arrest, apoptosis, and senescence. Cell 149:1269-1283.

83. Valente, L. J., D. H. Gray, E. M. Michalak, J. Pinon-Hofbauer, A. Egle, C. L. Scott, A. Janic, and A. Strasser. 2013. p53 efficiently suppresses tumor development in the complete absence of its cell-cycle inhibitory and proapoptotic effectors p21, Puma, and Noxa. Cell Rep 3:1339-1345.

84. Olivier, M., R. Eeles, M. Hollstein, M. A. Khan, C. C. Harris, and P. Hainaut. 2002. The IARC TP53 database: new online mutation analysis and recommendations to users. Hum Mutat 19:607-614.

85. Wassman, C. D., R. Baronio, O. Demir, B. D. Wallentine, C. K. Chen, L. V. Hall, F. Salehi, D. W. Lin, B. P. Chung, G. W. Hatfield, A. Richard Chamberlin, H. Luecke, R. H. Lathrop, P. Kaiser, and R. E. Amaro. 2013. Computational identification of a transiently open L1/S3 pocket for reactivation of mutant p53. Nat Commun 4:1407.

86. Joerger, A. C., and A. R. Fersht. 2010. The tumor suppressor p53: from structures to drug discovery. Cold Spring Harb Perspect Biol 2:a000919.

87. Lambert, J. M., P. Gorzov, D. B. Veprintsev, M. Soderqvist, D. Segerback, J. Bergman, A. R. Fersht, P. Hainaut, K. G. Wiman, and V. J. Bykov. 2009. PRIMA-1 reactivates mutant p53 by covalent binding to the core domain. Cancer Cell 15:376-388.

88. Liu, X., R. Wilcken, A. C. Joerger, I. S. Chuckowree, J. Amin, J. Spencer, and A. R. Fersht. 2013. Small molecule induced reactivation of mutant p53 in cancer cells. Nucleic Acids Res 41:6034-6044.

89. Yu, X., A. Vazquez, A. J. Levine, and D. R. Carpizo. 2012. Allele-specific p53 mutant reactivation. Cancer Cell 21:614-625.

90. Joerger, A. C., M. R. Bauer, R. Wilcken, M. G. Baud, H. Harbrecht, T. E. Exner, F. M. Boeckler, J. Spencer, and A. R. Fersht. 2015. Exploiting Transient Protein States for the Design of Small-Molecule Stabilizers of Mutant p53. Structure 23:2246-2255.

91. Ventura, A., D. G. Kirsch, M. E. McLaughlin, D. A. Tuveson, J. Grimm, L. Lintault, J. Newman, E. E. Reczek, R. Weissleder, and T. Jacks. 2007. Restoration of p53 function leads to tumour regression in vivo. Nature 445:661-665.

92. Martins, C. P., L. Brown-Swigart, and G. I. Evan. 2006. Modeling the therapeutic efficacy of p53 restoration in tumors. Cell 127:1323-1334.

93. Wiman, K. G. 2007. Restoration of wild-type p53 function in human tumors: strategies for efficient cancer therapy. Adv Cancer Res 97:321-338.

94. Fink, A. L. 2005. Natively unfolded proteins. Curr Opin Struct Biol 15:35-41.

95. Hupp, T. R. 1999. Regulation of p53 protein function through alterations in protein-folding pathways. Cell Mol Life Sci 55:88-95.

96. Weinberg, R. L., D. B. Veprintsev, and A. R. Fersht. 2004. Cooperative binding of tetrameric p53 to DNA. J Mol Biol 341:1145-1159.

97. Balagurumoorthy, P., H. Sakamoto, M. S. Lewis, N. Zambrano, G. M. Clore, A. M. Gronenborn, E. Appella, and R. E. Harrington. 1995. Four p53 DNA-binding domain peptides bind natural p53-response elements and bend the DNA. Proc Natl Acad Sci U S A 92:8591-8595.

98. Nagaich, A. K., V. B. Zhurkin, S. R. Durell, R. L. Jernigan, E. Appella, and R. E. Harrington. 1999. p53-induced DNA bending and twisting: p53 tetramer binds on the outer side of a DNA loop and increases DNA twisting. Proc Natl Acad Sci U S A 96:1875-1880.

99. Tidow, H., R. Melero, E. Mylonas, S. M. Freund, J. G. Grossmann, J. M. Carazo, D. I. Svergun, M. Valle, and A. R. Fersht. 2007. Quaternary structures of tumor suppressor p53 and a specific p53 DNA complex. Proc Natl Acad Sci U S A 104:12324-12329.

100. Emamzadah, S., L. Tropia, and T. D. Halazonetis. 2011. Crystal structure of a multidomain human p53 tetramer bound to the natural CDKN1A (p21) p53-response element. Mol Cancer Res 9:1493-1499.

101. Emamzadah, S., L. Tropia, I. Vincenti, B. Falquet, and T. D. Halazonetis. 2014. Reversal of the DNA-binding-induced loop L1 conformational switch in an engineered human p53 protein. J Mol Biol 426:936-944.

102. Petty, T. J., S. Emamzadah, L. Costantino, I. Petkova, E. S. Stavridi, J. G. Saven, E. Vauthey, and T. D. Halazonetis. 2011. An induced fit mechanism regulates p53 DNA binding kinetics to confer sequence specificity. EMBO J 30:2167-2176.

103. el-Deiry, W. S., S. E. Kern, J. A. Pietenpol, K. W. Kinzler, and B. Vogelstein. 1992. Definition of a consensus binding site for p53. Nat Genet 1:45-49.

104. McLure, K. G., and P. W. Lee. 1998. How p53 binds DNA as a tetramer. EMBO J 17:3342-3350.

105. Kitayner, M., H. Rozenberg, N. Kessler, D. Rabinovich, L. Shaulov, T. E. Haran, and Z. Shakked. 2006. Structural basis of DNA recognition by p53 tetramers. Mol Cell 22:741-753.

106. Weinberg, R. L., D. B. Veprintsev, M. Bycroft, and A. R. Fersht. 2005. Comparative binding of p53 to its promoter and DNA recognition elements. J Mol Biol 348:589-596.

107. Ahn, J., and C. Prives. 2001. The C-terminus of p53: the more you learn the less you know. Nat Struct Biol 8:730-732.

108. Hupp, T. R., D. W. Meek, C. A. Midgley, and D. P. Lane. 1992. Regulation of the specific DNA binding function of p53. Cell 71:875-886.

109. McKinney, K., M. Mattia, V. Gottifredi, and C. Prives. 2004. p53 linear diffusion along DNA requires its C terminus. Mol Cell 16:413-424.

110. Tafvizi, A., F. Huang, A. R. Fersht, L. A. Mirny, and A. M. van Oijen. 2011. A single-molecule characterization of p53 search on DNA. Proc Natl Acad Sci U S A 108:563-568.

111. Lukman, S., D. P. Lane, and C. S. Verma. 2013. Mapping the structural and dynamical features of multiple p53 DNA binding domains: insights into loop 1 intrinsic dynamics. PLoS One 8:e80221.

112. Lavery, R., M. Moakher, J. H. Maddocks, D. Petkeviciute, and K. Zakrzewska. 2009. Conformational analysis of nucleic acids revisited: Curves+. Nucleic Acids Res 37:5917-5929.

113. D'Abramo, M., N. Besker, A. Desideri, A. J. Levine, G. Melino, and G. Chillemi. 2015. The p53 tetramer shows an induced-fit interaction of the C-terminal domain with the DNA-binding domain. Oncogene.

114. Arlt, C., C. H. Ihling, and A. Sinz. 2015. Structure of full-length p53 tumor suppressor probed by chemical cross-linking and mass spectrometry. Proteomics 15:2746-2755.

115. Friedler, A., D. B. Veprintsev, S. M. Freund, K. I. von Glos, and A. R. Fersht. 2005. Modulation of binding of DNA to the C-terminal domain of p53 by acetylation. Structure 13:629-636.

116. Viadiu, H., and A. K. Aggarwal. 2000. Structure of BamHI bound to nonspecific DNA: a model for DNA sliding. Mol Cell 5:889-895.

117. Leith, J. S., A. Tafvizi, F. Huang, W. E. Uspal, P. S. Doyle, A. R. Fersht, L. A. Mirny, and A. M. van Oijen. 2012. Sequence-dependent sliding kinetics of p53. Proc Natl Acad Sci U S A 109:16552-16557.

118. Inc., C. C. G. Molecular Operating Environment (MOE).

119. Jacobs, S. A., J. M. Harp, S. Devarakonda, Y. Kim, F. Rastinejad, and S. Khorasanizadeh. 2002. The active site of the SET domain is constructed on a knot. Nat Struct Biol 9:833-838.

120. Natan, E., C. Baloglu, K. Pagel, S. M. Freund, N. Morgner, C. V. Robinson, A. R. Fersht, and A. C. Joerger. 2011. Interaction of the p53 DNA-binding domain with its n-terminal extension modulates the stability of the p53 tetramer. J Mol Biol 409:358-368.

121. Lim, K. W., S. Amrane, S. Bouaziz, W. Xu, Y. Mu, D. J. Patel, K. N. Luu, and A. T. Phan. 2009. Structure of the human telomere in K+ solution: a stable basket-type G-quadruplex with only two G-tetrad layers. J Am Chem Soc 131:4301-4309.

122. Bochkareva, E., L. Kaustov, A. Ayed, G. S. Yi, Y. Lu, A. Pineda-Lucena, J. C. Liao, A. L. Okorokov, J. Milner, C. H. Arrowsmith, and A. Bochkarev. 2005. Single-stranded DNA mimicry in the p53 transactivation domain interaction with replication protein A. Proc Natl Acad Sci U S A 102:15412-15417.

123. Rustandi, R. R., D. M. Baldisseri, and D. J. Weber. 2000. Structure of the negative regulatory domain of p53 bound to S100B(betabeta). Nat Struct Biol 7:570-574.

124. Lowe, E. D., I. Tews, K. Y. Cheng, N. R. Brown, S. Gul, M. E. Noble, S. J. Gamblin, and L. N. Johnson. 2002. Specificity determinants of recruitment peptides bound to phospho-CDK2/cyclin A. Biochemistry 41:15625-15634.

125. Ma, B., Y. Pan, J. Zheng, A. J. Levine, and R. Nussinov. 2007. Sequence analysis of p53 response-elements suggests multiple binding modes of the p53 tetramer to DNA targets. Nucleic Acids Res 35:2986-3001.

126. Jorgensen, W. L., J. Chandrasekhar, M. J. D., R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926-935.

127. D.A. Case, J. T. B., R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K.M. Merz, G. Monard, P. Needham, H. Nguyen, H.T. Ngyuen, I. Omelyan, A. Onufriew, D.R. Roe, A. Roitberg, R. Salomon-Ferrer, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker,

J. Wang, R.M. Wolf, X. Wu, D.M. York and P.A. Kollman. 2015. AMBER 2015. In University of California, San Francisco.

128. Pang, Y. P. 1999. Novel zinc protein molecular dynamics simulations: steps toward antiangiogenesis for cancer treatment. J Mol Model 5:196-202.

129. Maier, J. A., C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling. 2015. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput 11:3696-3713.

130. Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. 2005. Scalable molecular dynamics with NAMD. J Comput Chem 26:1781-1802.

131. Darden, T., L. Perera, L. Li, and L. Pedersen. 1999. New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. Structure 7:R55-60.

132. Wickham, H. 2009. Ggplot2 elegant graphics for data analysis. In Use R. Springer,, New York. viii, 212 pages.

133. Janert, P. K. 2010. Gnuplot in action : understanding data with graphs. Manning, Greenwich, Conn.

134. Foundation, P. S. Python Language Reference. http://www.python.org.

135. Ilkay Altintas, C. B., Efrat Jaeger, Matthew Jones, Bertram Ludäscher, Steve Mock. 2004. Kepler: an extensible system for design and execution of scientific workflows. . In 16th International Conference on Scientific and Statistical Database Management, 2004. Proceedings. 423-424.

136. Team, R. C. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

137. Ngan, C. H., T. Bohnuud, S. E. Mottarella, D. Beglov, E. A. Villar, D. R. Hall, D. Kozakov, and S. Vajda. 2012. FTMAP: extended protein mapping with user-selected probe molecules. Nucleic Acids Res 40:W271-275.

138. Trabuco, L. G., E. Villa, K. Mitra, J. Frank, and K. Schulten. 2008. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. Structure 16:673-683.

139. Votapka, L. W., L. Czapla, M. Zhenirovskyy, and R. E. Amaro. 2013. DelEnsembleElec: Computing Ensemble-Averaged Electrostatics Using DelPhi. Commun Comput Phys 13:256-268.