UNIVERSITY OF CALIFORNIA, SAN DIEGO

Automation of molecular dynamics simulations and progress in drug discovery efforts towards the inhibition of two kinases on the non-canonical NF-kappa B pathway

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

in

Chemistry

by

Garrett J. Chan

Committee in charge:

Professor Rommie E. Amaro, Chair
Professor Gourisankar Ghosh
Professor J. Andrew McCammon

2017

The Thesis of Garrett J. Chan is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2017

Dedication

To my family and to my friends: Apurva, Christina, Jimmy, Melissa, Shannon, and Sushmitha

Table of Contents

List of Figures

List of Tables

List of Supplementary Materials

Acknowledgements

I would like to express my gratitude to all of the members of the Amaro lab, whose support and friendship has guided me through this journey. I would also especially like to thank Drs. Robert Malmstrom and Özlem Demir for their mentorship, and Professor Rommie Amaro for her support as the chair of my committee and an indefatigable model of gumption and optimism.

Chapter 1, in full, has been accepted for publication of the material as it may appear in “A Kepler Workflow Tool for Reproducible AMBER GPU Molecular Dynamics” by Purawat, Shweta; Ieong, Pek U.; Malmstrom, Robert D.; Chan, Garrett J.; Yeung, Alan K.; Walker, Ross C.; Altintas, Ilkay; and Amaro, Rommie E., accepted for publication in *Biophysical Journal* in 2017. This chapter is included with the permission from all of the authors.

Chapter 3, in full, is in preparation for publication of the material as it may appear in “Elucidation of the Dynamics of NF-κB Kinase and the Discovery of a Potential Allosteric Site” by Chan, Garrett J.; Demir, Özlem; Schiffer, Jamie M.; and Amaro, Rommie E. in 2017. This chapter is included with the permission from all of the authors.

ABSTRACT OF THE THESIS

Automation of molecular dynamics simulations and drug discovery efforts towards the inhibition of two kinases on the non-canonical NF-kappa B pathway

by

Garrett J. Chan

Master of Science in Chemistry

University of California, San Diego, 2017

Professor Rommie E. Amaro, Chair

Rational drug design has been revolutionized by computational methods such as molecular dynamics (MD) simulations and virtual screening. From both the methods development and the application standpoints, the entry of computation into chemistry has greatly changed how biophysics and medicinal chemistry are approached. First, we have developed an interactive platform on which a broader array of scientists can efficiently run molecular

dynamics simulations, presenting an opportunity for a more rigorous and more accessible approach to utilizing this method. Second, virtual screening has allowed us to foray into the design of a protein-protein inhibitor for a polymeric form of IκB kinase alpha. And third, the application of molecular dynamics simulations has aided us in understanding the dynamics of proteins. Through MD, our studies on NF-κB-inducing kinase have given us insight into the interplay of interactions among its catalytic domains and revealed a possible allosteric site.

Chapter 1:

A Kepler Workflow Tool for Reproducible AMBER GPU Molecular Dynamics

Abstract

With the drive towards high throughput molecular dynamics (MD) simulations involving ever greater numbers of simulation replicates run for longer, biologically relevant timescales (microseconds), the need for improved computational methods that facilitate fully-automated MD workflows gains more importance. Here we report the development of an automated workflow tool to perform AMBER GPU MD simulations. Our workflow tool capitalizes on the capabilities of the Kepler platform to deliver a flexible, intuitive, and user-friendly environment and the AMBER GPU code for a robust and high-performance simulation engine. Additionally, the workflow tool reduces user input time by automating repetitive processes and facilitates access to GPU clusters, whose high-performance processing power makes simulations of large numerical scale possible. The presented workflow tool facilitates the management and deployment of large sets of MD simulations on heterogeneous computing resources. The workflow tool also performs systematic analysis on the simulation outputs and enhances simulation reproducibility, execution scalability, and MD method development including benchmarking and validation.

1.1 Introduction

Continued advances in computing power and the development of the graphics processing unit (GPU), make the study of proteins, receptors, and other biophysical systems with molecular dynamics simulations increasing accessible[1, 2]. Molecular dynamics (MD) simulations are a powerful computational tool that predicts protein dynamics in physiological conditions and allows researchers to study atomic interactions[3], aiding in improved models of molecular recognition and drug design[4].

1.1.1 An overview of MD simulation packages

Several well-known MD simulation packages available for biomolecular simulations, including Gromacs[5], NAMD[6], Desmond[7], OpenMM[8], and AMBER[9], have made significant contributions to scientific discoveries[3, 10-12]. Each can run calculations on a single central processing unit (CPU) or a single graphic processing unit (GPU), and on clusters using multiple CPUs or GPUs. In contrast to a CPU, a GPU is constructed with significantly reduced cache memory. It is built with thousands of smaller computing cores, designed to handle multipe identical mathematical operations simultaneously. Many GPU MD packages utilize the GPU for the non-bonded force calculation. The AMBER GPU software can perform complex calculations, including semi-isotropic pressure scaling, adaptively biased MD, and thermodynamic integration. The AMBER package utilized in this work can support MD simulations on a single GPU and multiple GPUs[1, 2]

with performances that outstrip even the most powerful conventional CPU based supercomputers[13, 14]. It takes advantage of the GPU to accelerate classical MD on realistic-sized systems, typically up to a maximum of ~3 million atoms per simulation, bound by a size limit due to memory restrictions unique to specific GPU models. Nevertheless, systems of this size include many drug targets, including membrane-embedded G-protein coupled receptors, or a protein trimer with bound antibodies (3).

1.1.2 The issue of reproducibility in MD

Broadly speaking, a major criticism of MD simulations relates to their reproducibility. Reproducing published MD findings (i.e., an independent group reproducing the same set of results for the same system of interest) can be challenging due to many reasons, including complicated run procedures that rely on user-developed scripts, technical numerical reproducibility (which is dependent on a number of factors including the compiler, executable, and dependent libraries, as well as whether FFTW is utilized), statistical reproducibility (which can be achieved through the use of a high number of simulation replicates), and incomplete reporting of methods in the published literature. For example, it is common for an end-to-end MD experiment to require multiple dozens of steps; conceptually we can define the set of required steps as a “workflow.” Typically researchers employ various scripts to piece together each of the different steps. The formal encoding of these many scripts into an automated workflow framework is one

way to address many challenges related to reproducibility, as well as method development and researcher training.

1.1.3 Workflow platforms provide reproducibility to scientific tasks

Multiple workflow platforms are available for scientific research. Kepler is a completely open source option with an extensive set of basic and advanced modules that support various workflow configurations. Kepler is a Java-based platform that delivers an easy-to-navigate graphical user interface (GUI) and user-friendly text file interface. From the GUI platform, the user is given a palette of modules, actors, and directors, with which to visually program their workflow[15]. A director controls the execution of the workflow, and actors carry out specific functions in the workflow. Contributors around the world develop modules and other workflows that perform complementary functions. Within the last decade, Kepler has implemented multiple modules that support the applications of statistical analysis, biological research, adaptations of external software, and options to select execution on different platforms and with different schedulers[15,16]. Consequently, Kepler has become a flexible and broad-purpose platform on which to build a workflow that can perform both MD simulations and subsequent statistical analysis. The user-friendly GUI interface also provides a suitable learning apparatus for beginners to understand the components of MD and for experienced users to customize and extend the workflow to their project needs.

1.1.4 Kepler as an MD simulation gateway

Combining these two computational tools, the AMBER GPU simulation code and the Kepler platform, we have developed a workflow tool that allows scientists to navigate, develop, deploy, analyze, and share MD simulation protocols. The workflow tool merges multiple MD steps that are otherwise scattered across disparate software tools or scripts into one central command that controls and executes all steps, starting from the initial energy minimization to multicopy production dynamics and routine analysis. The workflow tool submits jobs to various hardware architectures and copies back the output data to the user's local machine, helps to systematically run MD simulations, and plots thermodynamic properties to check the integrity of the simulations. The integrated simulation execution and analysis features of the workflow reduces the time needed to monitor the various simulation steps and facilitates handling of the simulation data. Moreover, Kepler is equipped with an automatic execution check function, which will stop and highlight the step if the simulations fail to execute until the end.

The AMBER GPU workflow tool leverages Kepler's advanced provenance capabilities to assist in achieving reproducibility[17]. Each time the workflow tool is executed, a detailed provenance report containing key signatures required to replicate a simulation is generated. The report, which can be used as a basis for method section reporting, contains not only MD configuration parameter information, but also details execution parameters not normally reported, including hardware configuration and software and

library version information (AMBER, CUDA, etc.). The latter set of information is not typically reported, though it has the potential to impact reproducibility as much as MD configuration file information.

1.1.5 Differences between the AMBER GPU workflow and QuikMD

Our AMBER GPU workflow tool provides some similar capabilities as QwikMD, another software that executes MD simulations, but unlike QwikMD[18], which focuses on VMD and NAMD, our workflow tool utilizes AMBER and its well-known GPU computing power. QwikMD also focuses on assisting the user in building the system itself, whereas our workflow tool focuses on MD job execution and analysis, relying on the user to create the system files on their own. Our workflow tool is also able to execute jobs on various clusters including the XSEDE architectures, and the users can easily control the execution options via the GUI or command-line interface. Furthermore, through its integration with Kepler, the AMBER GPU workflow tool can easily utilize Kepler's extensive data analytics capabilities, which are continually being developed. CHARMM-GUI[19] is another complimentary tool that provides a web interface for system construction. It is capable of building and generating MD input files for a magnitude of biological systems, including solvated proteins, solvated proteins with ligands attached and membrane proteins, etc. The inputs produced by CHARMM-GUI could be used in the AMBER GPU workflow tool to actually execute the MD

simulations. To our knowledge CHARMM-GUI does not yet provide job execution capabilities.

Finally we note that the workflow tool presented here is freely accessible, open source, and easy to modify and reuse. Each of the modules (a.k.a. Kepler actors) can be reconfigured with different components (e.g., including automated preparation and simulation of receptor-ligand complexes), and additional analysis steps can be developed and integrated into customized experiments. Workflows developed with KNIME can be called from within Kepler and functionality of different components can be extended via interoperability with packages such as HTMD[20], MSMBuilder[21], and PyEMMA[22]. The AMBER GPU workflow tool can be found at http://nbcr.ucsd.edu/products/workflow-distribution along with the user guide.

1.2 Materials and Methods

The workflow tool requires an input directory containing two subdirectories. One subdirectory, labeled confDir, contains all the template simulation input files. The template AMBER input scripts provided in the default workflow tool were converted from the same parameters used in Wassman et al.'s NAMD MD simulations[4], which consisted of four stages of minimization, one step of heating, three stages of equilibrations and one step of production. Users can also configure their own input files if desired (see user manual for details), e.g., if they would like to run another type of MD simulation such as accelerated or replica exchange MD. The second

subdirectory, labeled with the system's stem name, contains the system's AMBER topology and coordinate files. For example, the system's topology and coordinate files are named under r174h_stictic.top and r174h_stictic.crd, respectively; then, the subdirectory folder name will be r174h_stictic.

1.2.1 Workflow submission options

There are three submission options implemented to run the AMBER GPU MD simulation workflow: LocalExecution on user's local machine, CometGPUCluster on the San Diego Super Computer (SDSC) comet cluster, and PrivateGPUCluster on a remote private cluster. The workflow will always start from a user's local machine, but it can connect to remote clusters, submit MD simulation jobs, and transfer simulation results. Moreover, users can connect the workflow to a GPU cluster of their choice. Regardless of which submission option a user chooses, all outputs at the end of a simulation can be found on the local machine, where users started the workflow.

The workflow tool automates MD simulation processes and generates analytical plots that help analyze the system of interest and validate simulation quality. There are two parts to this workflow: simulation and analysis. Simulations can be further broken down into four components executed by four actors: minimization, heating, equilibration, and production. The analysis components are small actors working collectively to call locally implemented software, including cpptraj from the AmberTools suite and R,

the programming language (23, 24) to generate time-evolution energy, temperature, pressure, and RMSD plots. The resulting plots are stored in the system subdirectory for analysis and to verify simulations quality.

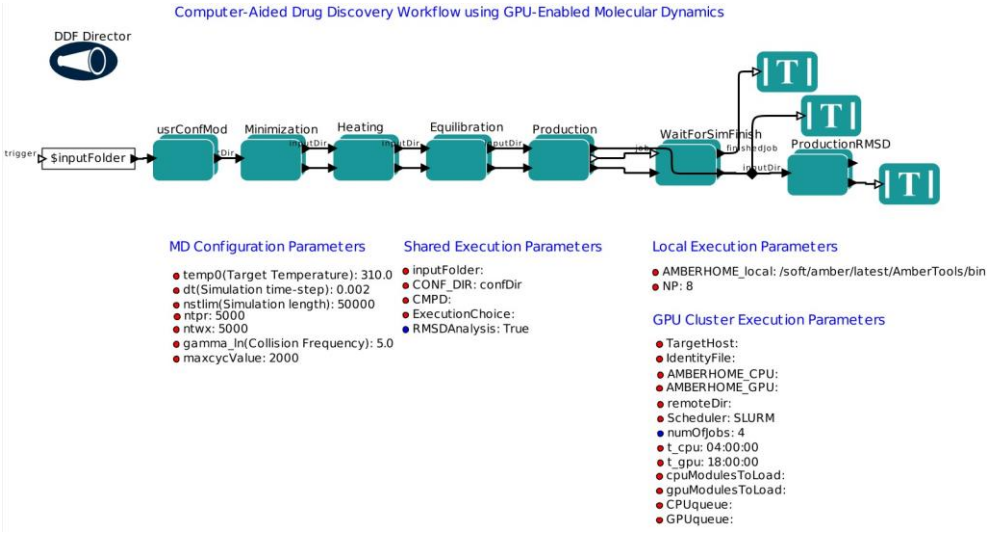


Figure 1.1 AMBER GPU MD simulation workflow, showing the uppermost level of the workflow construct.

1.2.2 Components of the workflow: minimization to production

The first simulation component is the minimization (**Fig. 1.1**). In this step, a system is energetically minimized to a local minimum. After the minimizations, the workflow plots the energy for each step and an overall diagram combining all the plots. Since energy is the largest changing parameter during minimization, the energy plots are good indicators of whether minimizations were performed successfully.

Minimizations are useful for finding local minima but typically one would want to simulate a system in its physiological environment or at room

temperature. The next heating step (**Fig. 1.1**), thus, adds kinetic energy or "heats" the system from 0 K to a target temperature defined by the user. For the heating step, the temperature plot over time will indicate whether the system did indeed heat from 0 K to a user-defined value. After heating, the equilibration steps (**Fig. 1.1**) shift the system in gradual steps toward thermal equilibrium under NPT or NVT conditions.

Both the heating and equilibration steps output several thermodynamic property plots including kinetic energy, potential energy, total energy, pressure, and temperature. The workflow also calculates and plots the time-dependent root-mean-square-deviation (RMSD) of backbone α−carbons of the system. RMSD analyzes the time evolution of a selected set of atoms against some reference frame, in this case the default is the initial frame. The time series backbone α−carbon RMSD analysis provides information about the overall global change of the protein backbone structure over time. A collection of these property plots would indicate whether stable equilibration was attained.

The last component, production (**Fig. 1.1**), runs non-constrained NPT or NVT MD simulations (based on the parameters set on users' input scripts) and derives a dynamic atomic trajectory of the system in a simulated physiological environment. In this step, users can choose to run the workflow on a local machine with one GPU card or on a remote cluster with multiple GPUs; in the latter case it is possible to deploy multiple copies of the production simulations concurrently. As with heating and equilibration, time

series plots for temperature, pressure, energy, and RMSD plots are also generated in the production stage. Regardless of how many copies of production simulation users would like to run, they will all start with the same starting structure from the last equilibration frame. The best solution to check whether the simulations ran successfully is to visualize the trajectories and examine if there are any abnormalities. In addition, analysis techniques presented here, such as the RMSD plots, can also serve as an investigational tool to identify and quantify any outstanding abnormal structural deviations from the starting structure. At the end of each execution, the workflow prints out a detailed report, which lists out the input parameters, hardware specification, the AMBER version used, the random seeds for the equilibrations and productions, and file locations. This report can be appended to a publication as Supporting Information, and would contain all the requisite details to ensure technical (and statistical, if achieved) reproducibility. Or, it can be used as a guide to assist authors in writing more comprehensive Methods sections.

1.3 Application

We show the utility of the workflow tool on multiple protein systems of variable size, and with and without bound ligands (**Fig. 1.2**). Here we present in detail one of the systems, the p53 protein bound to a small molecule named stictic acid (4). P53 is considered to be a potential drug target for anticancer therapy (25-27), making its simulation with our workflow a relevant

test case to perform and scrutinize. Non-activated and mutant p53 is seen in more than half of human cancer cases (28), and stictic acid is a known compound that activates mutant p53 (4). We are therefore interested in understanding its dynamics when stictic acid is bound.

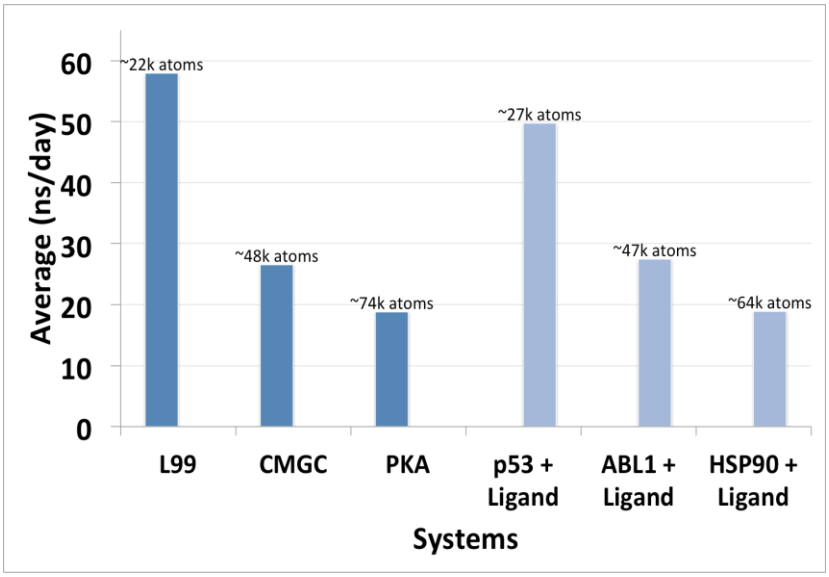


Figure 1.2 Performance of the AMBER GPU workflow on various systems on a local NVIDIA Tesla K20 cluster.

1.3.1 Demonstration of capability with a p53-stictic acid system

Courtesy of Wassman et al., we obtained the AMBER topology and coordinate files of the p53-stictic acid complex to run MD simulations using the workflow. The simulation consisted of five minimization steps, one heating step, three equilibration steps and one production step that branched out to perform three independent production simulations. We set up the simulations through the command prompt with Kepler on our local machine

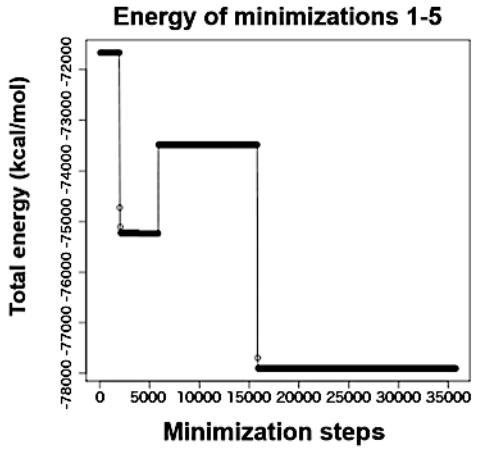
(**Fig. 1.3**), and Kepler submitted the simulation jobs to the Comet cluster at SDSC. This single command communicates with Kepler to run all the simulation steps consecutively without interruption. For more information on each parameter, please refer to the user manual.

```
./kepler.sh -runwf -nogui -ExecutionChoice CometGPUCluster -TargetHost user@comet.sdsc.edu -
remoteDir /oasis/scratch/comet/user/temp_project AMBERHOME_CPU /opt/amber/bin/ -
AMBERHOME_GPU /share/apps/gpu/amber cpuModulesToLoad "module load amber" -
gpuModulesToLoad "module load cuda/6.5" -t_cpu 04:00:00 -t_gpu 18:00:00 -CPUqueue compute -
GPUqueue gpushared -inputFolder /home/user/Desktop/MD_Test_Oct -IdentityFile /root/.ssh/id_rsa -
CMPD r175h_stictic -numOfJobs 3
/home/user/Desktop/MD_Test_Oct/MDCADD_WF_SingleCmpd_2016.xml
```

Figure 1.3 Command line for initiating MD simulations.

The resulting output data was then copied back to the local machine. We first checked the energy plot generated from the minimization steps (**Fig. 1.4**). As shown in **Fig. 1.4**, the system did move toward an energy minimum after five rounds of minimization, which suggested that the system was adequately minimized to a local/global minima. Next, we studied the temperature plot from the heating step. We assigned the desired temperature to 310 K to mimic physiological conditions. As shown in **Fig. 1.4**, the temperature gradually ramped from 0 K to 310 K and plateaued at 310 K.

This indicated that the system was successfully heated from 0 K to 310 K and held steady at 310 K.



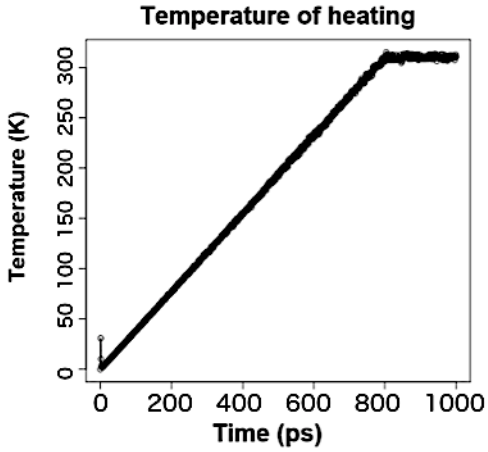


Figure 1.4 Energy minimization plot.

Figure 1.5 Heating plot.

As per the default workflow parameters, for our p53-stictic acid system, the protein backbone was constrained during equilibration, so it was not necessary to study the RMSD during these steps. We looked into the total energy, pressure, and temperature from the last step of equilibration. During production runs, the temperature and pressure oscillated around the defined system values of 310 K and 1.0 bar, respectively, as one would expect. Total energy increased but plateaued around -59000 kcal/mol, which suggested the system achieved at least a metastable equilibrium for the duration of the run (**Fig. 1.5**).

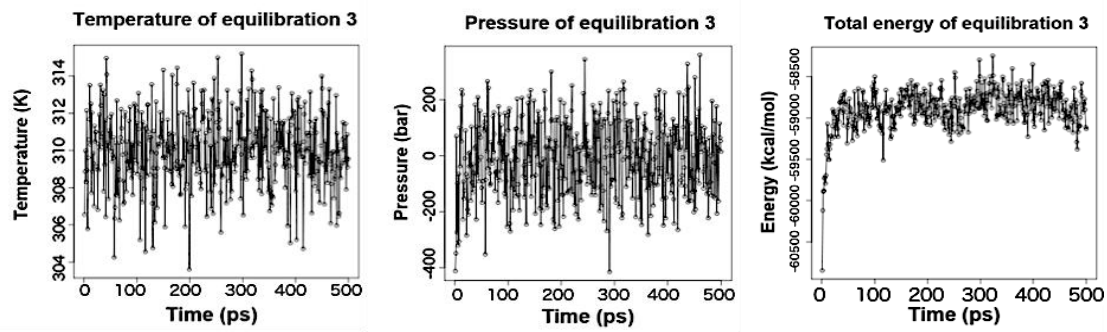


Figure 1.6 Time-series analyses (left to right) of the temperature, pressure, and total energy from the third and last stages of equilibration.

We generated three copies of 30 nanoseconds (ns) production dynamics for our system. Since the system had already gone through constant pressure simulations in a step-wise equilibration steps, we did not expect any drastic global change during production. We were able to verify that the behavior of our system was as expected from the RMSD plot produced by the workflow (**Fig. 1.7**). Shown in **Fig. 1.7**, we see small deviations (< 3Å) from the starting structure in all three production runs, indicating that the system sampled local conformational space without outstanding abnormal structural deviation.

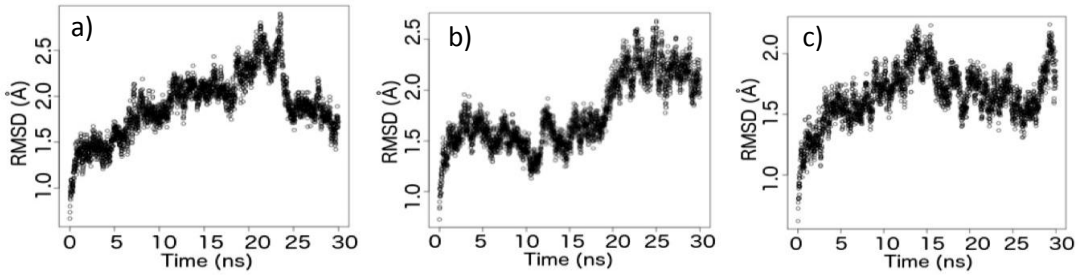


Figure 1.7 Time evolution of the RMSD of **a)** copy 1 **b)** copy 2, and **c)** copy 3 of production MD.

1.3.2 Built-in workflow report files on MD parameters

At the end of the execution, the workflow tool outputs a provenance report file consisting of input parameters, hardware specifications, the AMBER version, file locations, and the random seeds used for the simulations. Using the report, we were able to configure the workflow with the same exact random seed to test for reproducibility. For the next execution, we input the random seeds from the report, and the workflow used the values to generate a new set of input files. We tested the workflow on a local machine, on the SDSC Comet supercomputer, and on a private GPU cluster. The three runs resulted in trajectories that were highly similar, but not exactly the same because they were executed on different GPU clusters. Even with the same random seeds, all three runs sample slightly different yet similar conformational spaces (**Fig. 1.8**). This result is as expected because in order for the runs to be exactly reproducible, we must use the exact same hardware (29). For the local execution, AMBER was compiled with the FFTW algorithm, which introduces uncontrollable variability (29). Nevertheless, we

were able to re-create identical input files with the workflow and successfully run additional MD simulations of different trajectories, enhancing statistical and methodological reproducibility.

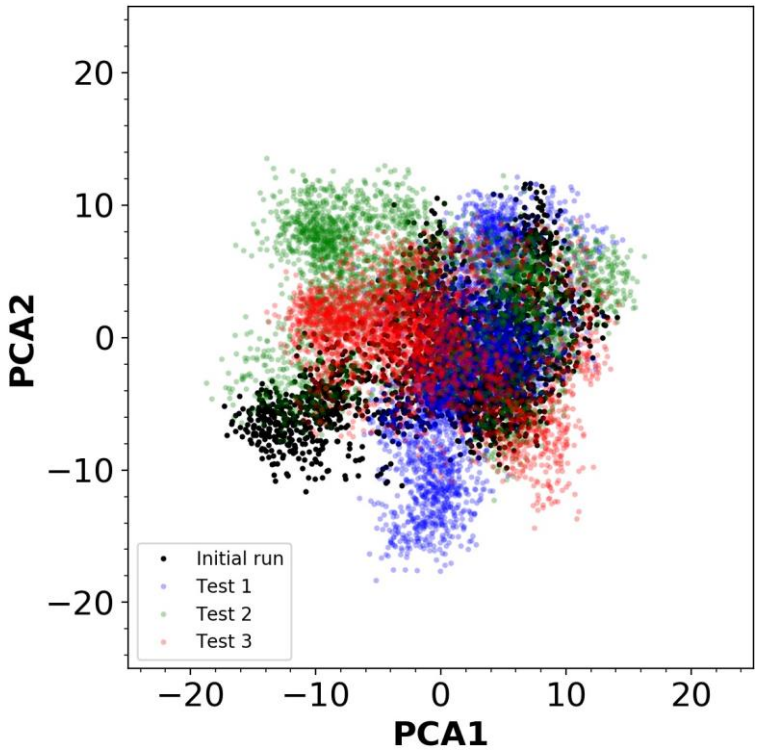


Figure 1.8 Principal component analysis (PCA) plot of the p53 simulation results. All of the trajectories from the initial run (computer generated random seed) and the three simulation tests (used the same random seed from the initial run) with 1) Comet (Test 1), 2) private GPU cluster (Test 2), 3) local machine (Test 3) were first aligned. The first two principle components were then calculated and the trajectories from each test were projected onto the PC space.

1.4 Software Availability

The AMBER GPU MD Simulation workflow is available under the GNU Lesser General Public License (LGPL), version 1.0 or later. Full documentation and examples are available through the National Biomedical Computation Resource download page, http://nbcr.ucsd.edu/products/workflow-distribution, and development is hosted on GitHub at https://github.com/nbcrrolls/workflows/tree/master/Production/AmberGPUMDSimulation. This workflow is also available for download in Zenodo (DOI 10.5281/zenodo.192490) and SciCrunch (RRID SCR_014389).

An open online training course on the Kepler-MDCADD workflow will be available on the BBDTC (https://biobigdata.ucsd.edu) in 2017. The course will include lecture contents, videos and hands-on training utilizing VM toolboxes. The BBDTC is a collaborative platform to encourage and facilitate training and education of the Biomedical Big Data community. The platform provides an intuitive portal to upload, distributed and discover new content about developments in the biomedical big data domain (30).

This chapter, in full, has been accepted for publication of the material as it may appear in “A Kepler Workflow Tool for Reproducible AMBER GPU Molecular Dynamics” by Purawat, Shweta; Ieong, Pek U.; Malmstrom, Robert D.; Chan, Garrett J.; Yeung, Alan K.; Walker, Ross C.; Altintas, Ilkay; and

Amaro, Rommie E., accepted for publication in *Biophysical Journal* in 2017.

1.5 References

1. Salomon-Ferrer, R., A. W. Gotz, D. Poole, S. Le Grand, and R. C. Walker. 2013. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. J Chem Theory Comput 9:3878-3888.

2. Gotz, A. W., M. J. Williamson, D. Xu, D. Poole, S. Le Grand, and R. C. Walker. 2012. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. J Chem Theory Comput 8:1542-1555.

3. Ieong, P., R. E. Amaro, and W. W. Li. 2015. Molecular dynamics analysis of antibody recognition and escape by human H1N1 influenza hemagglutinin. Biophys J 108:27042712.

4. 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.

5. 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.

6. 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.

7. Kevin J. Bowers, E. C., Huafeng Xu, Ron O. Dror, Michael P. Eastwood, Brent A. Gregersen, John L. Klepeis, Istvan Kolossvary, Mark A. Moraes, Federico D. Sacerdoti, John K. Salmon, Yibing Shan, and David E. Shaw. 2006. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In ACM/IEEE Conference on Supercomputing (SC06), Tampa, Florida.

8. Eastman, P., M. S. Friedrichs, J. D. Chodera, R. J. Radmer, C. M. Bruns, J. P. Ku, K. A. Beauchamp, T. J. Lane, L. P. Wang, D. Shukla, T. Tye, M. Houston, T. Stich, C. Klein,

M. R. Shirts, and V. S. Pande. 2013. OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. J Chem Theory Comput 9:461-469.

9. 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.

10. Allen, T. W., S. Kuyucak, and S. H. Chung. 1999. Molecular dynamics study of the KcsA potassium channel. Biophys J 77:2502-2516.

11. Tai, K., T. Shen, U. Borjesson, M. Philippopoulos, and J. A. McCammon. 2001. Analysis of a 10-ns molecular dynamics simulation of mouse acetylcholinesterase. Biophys J 81:715-724.

12. Awasthi, M., N. Jaiswal, S. Singh, V. P. Pandey, and U. N. Dwivedi. 2015. Molecular docking and dynamics simulation analyses unraveling the differential enzymatic catalysis by plant and fungal laccases with respect to lignin biosynthesis and degradation. J Biomol Struct Dyn 33:1835-1849.

13. Walker, R. AMBER 16 GPU ACCELERATION SUPPORT - BENCHMARK.

14. Le Grand, S., A. W. Gotz, and R. C. Walker. 2013. SPFP: Speed without compromise-A mixed precision model for GPU accelerated molecular dynamics simulations. Comput Phys Commun 184:374-380.

15. 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.

16. Altinas, I., J. W., Daniel Craw, Weizhong Li. 2012. Challenges and Approaches for Distributed Workflow-Driven Analysis of Large-Scale Biological Data. In EDBT Extending Database Technology. I. A. Divesh Srivastava, editor. ACM. 73-78.

17. Altintas, I., O. Barney, and E. Jaeger-Frank. 2006. Provenance collection support in the kepler scientific workflow system. In Proceedings of the 2006 international conference on Provenance and Annotation of Data. Springer-Verlag, Chicago, IL. 118-132.

18. Ribeiro, J. V., R. C. Bernardi, T. Rudack, J. E. Stone, J. C. Phillips, P. L. Freddolino, and K. Schulten. 2016. QwikMD - Integrative Molecular Dynamics Toolkit for Novices and Experts. Sci Rep 6:26536.

19. Jo, S., X. Cheng, J. Lee, S. Kim, S. J. Park, D. S. Patel, A. H. Beaven, K. I. Lee, H. Rui, S. Park, H. S. Lee, B. Roux, A. D. MacKerell, Jr., J. B. Klauda, Y. Qi, and W. Im. 2016. CHARMM-GUI 10 years for biomolecular modeling and simulation. J Comput Chem.

20. Doerr, S., M. J. Harvey, F. Noe, and G. De Fabritiis. 2016. HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. J Chem Theory Comput 12:1845-1852.

21. Harrigan, M. P., M. M. Sultan, C. X. Hernandez, B. E. Husic, P. Eastman, C. R.

    Schwantes, K. A. Beauchamp, R. T. McGibbon, and V. S. Pande. 2017. MSMBuilder: Statistical Models for Biomolecular Dynamics. Biophys J 112:10-15.

22. Scherer, M. K., B. Trendelkamp-Schroer, F. Paul, G. Perez-Hernandez, M. Hoffmann, N. Plattner, C. Wehmeyer, J. H. Prinz, and F. Noe. 2015. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. J Chem Theory Comput 11:5525-5542.

23. Roe, D. R., and T. E. Cheatham, 3rd. 2013. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. J Chem Theory Comput 9:3084-3095.

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

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

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

27. 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.

28. 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.

29. Duke, R. 2007. AMBER: reproducibility between software. AMBER mailing list.

30. Purawat, S., C. Cowart, R. E. Amaro, and I. Altintas. 2016. Biomedical Big Data Training Collaborative (BBDTC): An effort to bridge the talent gap in biomedical science and research. Procedia Comput Sci 80:1791-1800.

Chapter 2:

Progress Made on the Development of an Inhibitor of Polymerization of IκB Kinase Alpha

Abstract

IκB kinase alpha (IKKα) lies on the non-canonical NF-κB pathway, a cascade of signaling proteins that regulates the transcription of certain oncogenesis-related genes. Leveraging the crystal structure of a newly discovered hexameric form of IKKα, we have begun to develop an inhibitor that binds to this kinase and inhibits its polymerization at a protein-protein interface. We started with the characterization of druggable “hot spots” on the IKKα surfaces. Subsequently, we have utilized virtual screening and scoring techniques to nominate compounds with predicted favorable binding activities for biochemical assays. Further optimization of hits from those assays, through a similarity search and additional virtual screenings, have allowed us to explore additional regions of chemical space, from which we can continue to take preliminary steps towards the development of a potent and selective inhibitor.

2.1 Introduction

2.1.1 An overview of the non-canonical NF-κB pathway

The non-canonical NF-κB pathway regulates the transcription of the *APOBEC3B* gene, whose product, the protein APOBEC3B (A3B), is the second largest source of mutation across all cancers[1,2]. Specifically, the overexpression of A3B has been linked to head and neck cancer, and has been found in over 50% of breast cancer tumors[3]. The NF-κB transcription mechanism of the *APOBEC3B* gene is regulated by two pathways: the canonical and non-canonical NF-κB pathways.

IκB kinase alpha (IKKα), also called IKK1, is a signaling protein on both the canonical and non-canonical pathways[4]. Once in its homodimer form and phosphorylated, IKKα is known to facilitate the phosphorylation of the downstream protein complex p100-RelB. Furthermore, IKKα forms a hexamer on the non-canonical pathway. This form of the kinase has been found by Ghosh et al. to be necessary for the upstream signaling protein NF-κB-inducing kinase[5].

2.1.2 Domains and a novel hexameric form of IKKα

IKKα contains three domains: the kinase domain (KD), where the active site is located, the scaffold dimerization domain (SDD), which aids in the dimerization of IKKα, and the ubiquitin-like domain (ULD). The trimerization of dimers occurs at the KD-SDD interface, while dimerization of the trimers occurs at the SDD-SDD interface. Additionally, a recently solved

crystal structure[5], reveals a previously unknown hexameric form of IKKα, which our collaborators determined to be essential in p100 processing **(Fig. 2.1)**. More specifically, they determined that this hexamer was a trimer of dimers, presenting another potential protein-protein interface to target.

This study aims to develop a drug that will impede the polymerization of IKKα through structure-based drug design. Starting with the identification of regions of the protein surface favorable to binding, we and our collaborators performed iterative cycles of virtual screening and biochemical assays. From these first efforts, we have begun to characterize potentially favorable ligand-IKKα relationships and develop a potent and selective inhibitor.

2.2 Materials and Methods

We obtained the IKKα hexamer crystal structure (PDB: 5EBZ), which had been solved by cryo-electron microscopy (cryo-EM) at average global resolution of approximately 5.9 Å [5]. Fragment-based identification of hot spots was performed with FTMap[6,7] to determine sites of favorable ligand binding. A hot spot was then identified in each of the two regions of interest: the KD-SDD interface and the SDD-SDD interface.

Virtual screening, using the 3943 small molecules catalogued in the Center for Drug Discovery Innovation library, was done with Schrödinger's Glide software package[8–11]. The crystal structure first was minimized with the Protein Prep Wizard at pH 7.4 to assign protonation states and determine

hydrogen bonding. The ligands from the CDDI library were prepared with LigPrep[12], also at pH 7.4 and with stereochemistry retained. Cubic receptor grids 10 Å in length were centered on the residues found in promising hot spots identified in FTMap. Rigid docking was then performed.

From the docked and scored compounds, we identified 100 promising compounds to assay experimentally through a variety of methods. We took two factors into account when choosing compounds: the ligand's docking score (a predictive parameter that incorporates several factors of ligand binding free energy) and its ligand efficiency (the docking score divided by the number of heavy atoms). Although not all of the compounds chosen based on ligand efficiency were necessarily compounds with the best docking scores, our rationale for choosing these compounds was that they were generally smaller in size than compounds with the best docking scores, therefore imparting them greater potential for later optimization.

Immunoprecipitation (IP) assays were performed by our collaborators in Gourisankar Ghosh's group, also at UC San Diego, to determine the level of inhibitory activity of each compound on IKKα. Of the compounds that were virtually screened. The IP assays were performed by incubating the inhibitor, at a concentration of 100 μM, with MEF cells, and then exposing them to the agonist α-LTβR. The assays were then monitored for changes in the level to which p100 was processed to p52.

Those assays produced four hits. From those, we used the similarity search tool on the ChemBridge website to find derivative and analog

compounds. With this second cohort of compounds, flexible virtual screening, in which the ligand bonds were allowed to rotate, was performed to score the compounds. This second round of virtual screening hits is currently available for biochemical assays to test for inhibition of IKKα activity.

2.3 Results and Discussion

2.3.1 Druggable hot spots located at IKKα-IKKα interfaces were targeted

To begin the design of an inhibitor of IKKα polymerization, we first needed to identify druggable hot spots on the IKKα surface. We decided to focus our search efforts on regions of the surface outside of the conserved IKKα active site for better ligand selectivity. Furthermore, since we wanted to discover an inhibitor of protein-protein interaction, we decided to focus on the dimerization and trimerization interfaces of hexameric IKKα complex, at the SDD-SDD and KD-SDD interfaces, respectively. As the crystal structure contained two hexamers, there were a total of twelve protein chains, and FTMap analysis was run separately for each chain.

After performing hot spot-seeking analysis on each IKKα dimer, we counted the number of probes in each pocket identified by FTMap using Visual Molecular Dynamics[13,14]. With our attention directed towards hot spots located at protein-protein interfaces, we considered hot spots with more probes to be located in pockets that would have a better ligand-receptor binding free energy. Furthermore, using the Pose Viewer tool in

Schrödinger's Maestro software package[15], we were able to identify hydrogen bonds, salt bridges, and pi interactions between the probes and the residues. Locations of pockets heavily populated with probes were also compared across all twelve chains to find pockets common to multiple chains. The residues defining the pockets that were most common were noted.

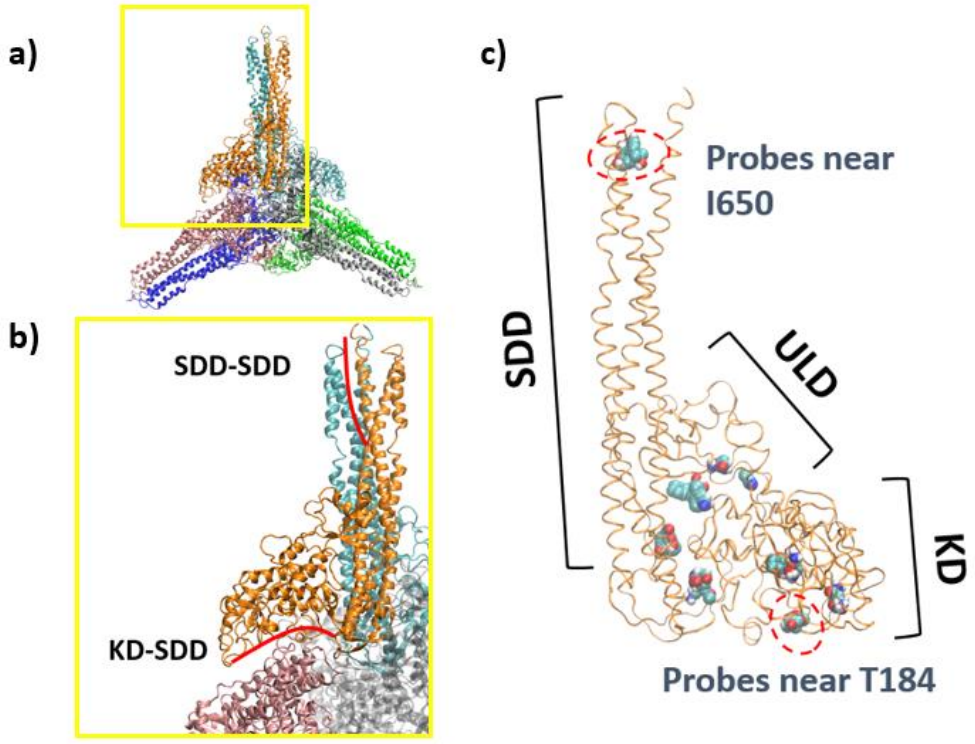


Figure 2.1 Structure of the IKKα hexamer and domains of a single IKKα chain. a) Overview of the IKKα hexamer in ribbon representation, with each of the six chains colored differently. **b)** Close-up of the interfaces between chains, demarcated in red. An interface between two scaffold dimerization domains is found between the orange and blue chains (SDD-SDD), and the interface between the kinase and scaffold dimerization domains is found between the orange and pink chains (KD-SDD). **c)** Locations of probes docked to the IKKα surface after performing FTMap analysis. The probes in the two pockets later used in virtual screening are circled in red. The ubiquitin-like domain is labeled as ULD.

When deciding which pockets to ultimately target, we first excluded pockets found in the IKKα active site. We then took probe populations, probe interactions, and recurrence of that pocket across multiple chains into account. From our census of the pockets, we chose the two pockets with the highest number of probes and that were common to multiple chains, and designated them with a name based on a representative residue. The T184 pocket is at the KD-SDD interface, while the I650 pocket is at the SDD-SDD interface (**Fig. 2.1**).

Virtual screening of the CDDI library was performed at each pocket using Glide, following the protocol described above. Scores from the virtual screening performed in Glide are tabulated below.

Table 2.1 Virtual screening of three IKKα pockets. The pockets are denoted by an identifying residue. The interface at which the pocket is located is given. The docking scores and ligand efficiencies of the 3900 compounds screened are listed for each pocket, and the best in each category are listed. No further work was done on pocket Y218 because it was determined to be too far from a neighboring chain to be an effective site of protein-protein interactions.

POCKET	INTERFACE	BEST DOCKING SCORE	BEST LIGAND EFFICIENCY
T184	KD-SDD	-5.751	-0.422
Y218	KD-SDD	-7.527	-0.509
I650	SDD-SDD	-9.112	-0.725

For both docking score and ligand efficiency, a more negative score indicates more favorable binding. As it can be seen from the table, the I650

pocket at the SDD-SDD interface, where two IKKα chains interact to form a dimer, had the most promising docking score and ligand efficiency. What is less certain, however, is if targeting this pocket will also inhibit the hexamerization process in some way. We also noted the interactions formed between the docked compounds and the ligands. One common motif that we noticed at the I650 pocket was the tendency of ligands to interact with Y494, usually through pi-pi interactions. We therefore believe that the pharmacophore of a potent inhibitor targeting this site should have an aromatic group positioned near to the tyrosine. Other residues with the potential for additional interactions include the hydrogen bond and polar interactions made by D490 and R493, S504, and E505.

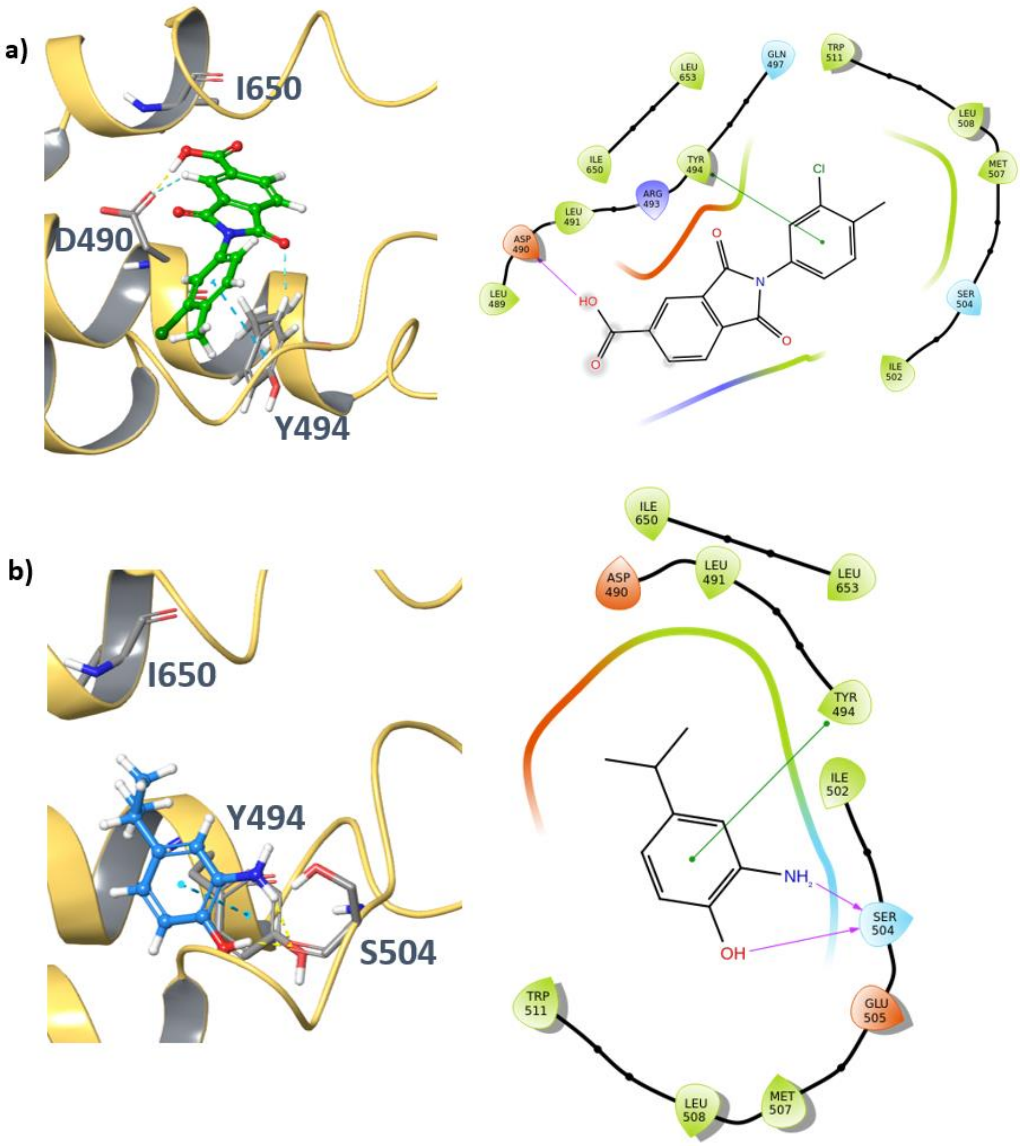


Figure 2.2 Interactions between virtually screened ligands and IKKα at the I650 pocket. a) Interactions shown here are for the top-scoring compounds. In the left panel, yellow dashed lines are hydrogen bonds, and blue dashed lines are pi interactions. In the right panel, green lines represent pi interactions, while purple arrows represent hydrogen bonds or polar interactions. The arrows point towards the electron donor. **b)** Interactions shown for the compound with the best ligand efficiency. Color scheme is the same as in **a)**.

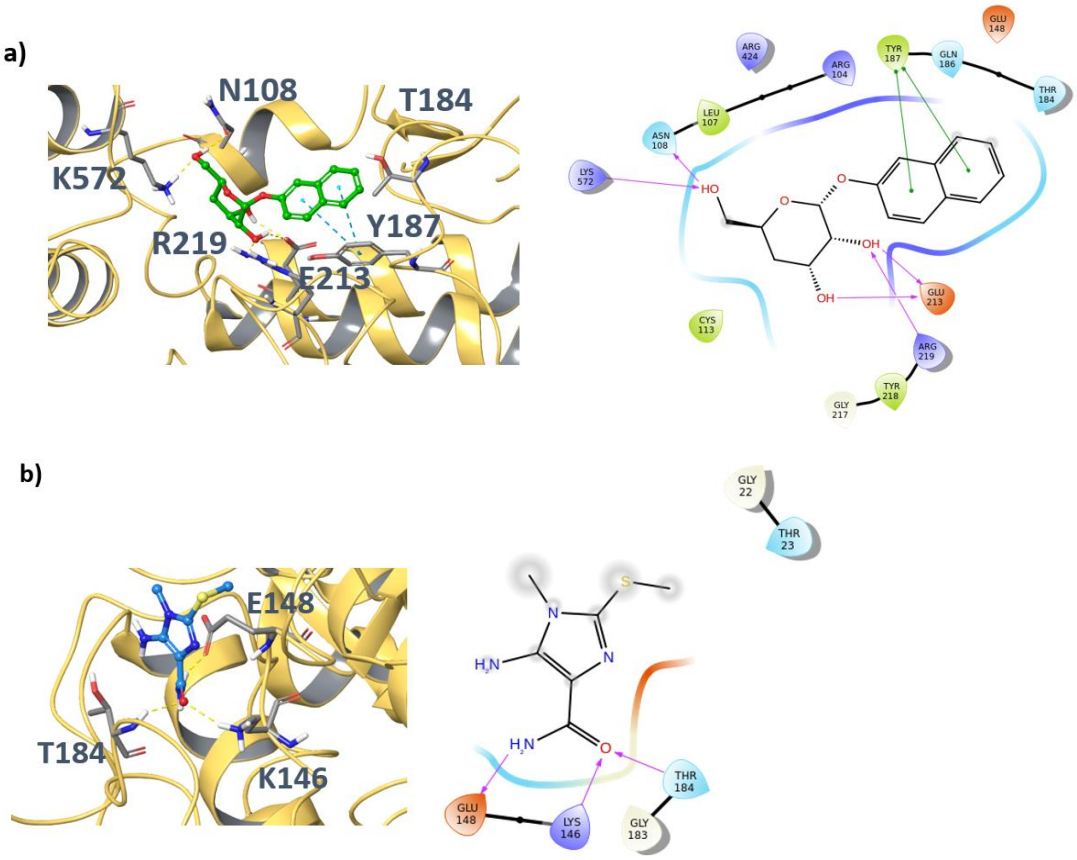


Figure 2.3 Interactions between virtually screened ligands and IKKα at the T184 pocket. a) Interactions shown here are for the top-scoring compound. In the left panel, yellow dashed lines are hydrogen bonds, and blue dashed lines are pi interactions. In the right panel, green lines represent pi interactions, while purple arrows represent hydrogen bonds or polar interactions. The arrows point towards the electron donor. **b)** Interactions shown for the compound with the best ligand efficiency. Color scheme is the same as in **a)**.

In the T218 pocket, T184 is a hydrogen bond donor in many ligand-receptor interactions, while E148 is also available for polar interactions. Furthermore, Y187 forms pi-pi interactions with ligand aromatic groups. However, there were fewer interactions common to many ligands in the T218

pocket than in the I650 pocket, which would explain its lower top docking score.

2.3.2 Virtual screening led the eventual identification of four hits for IKKα

From our virtual screenings, 100 compounds were chosen, based on the methods described above, for immunoprecipitation (IP) assays to be performed by our collaborators in Gourisankar Ghosh's group. The compounds chosen for assaying were evenly divided between the three pockets.

The IP assays revealed that these four compounds showed inhibitory properties. These assays were performed in the presence of α-LTβR, a signal for IKKα processing of p100 to p52, thereby distinguishing the inhibitors that were capable of attenuating that processing (**Fig. 2.4).** Subsequent rounds of IP assays narrowed down the initial cohort of possible hits to four confirmed hits. The virtual screening results of the four compounds from their docking to the I650 pocket are also tabulated below.

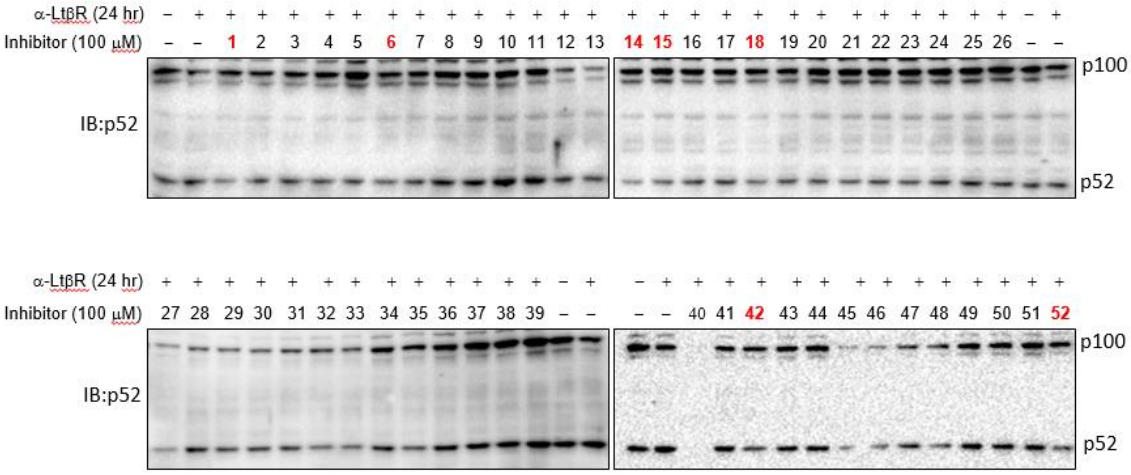


Figure 2.4 Selected results from immunoprecipitation assays. After exposing inhibitor-incubated MEF cells to the agonist signal α-LTβR, the expression of p52 for each inhibitor was compared. Inhibitors with significant diminishments in p52 expression are in red typeface. Not all IP assay results are shown. Image courtesy of Anup Mazumder, Gourisankar Ghosh, et al.

Table 2.2 Virtual screening results of the four hits from assay 1. The docking scores and ligand efficiencies of these compounds were taken from Glide after being docked into pocket I650. All of the compounds have relatively similar docking scores, but ligand efficiencies have a greater range of variation. Ligand 4027697 has both the best docking score and best ligand efficiency.

LIGAND ID	DOCKING SCORE	LIGAND EFFICIENCY
4027697	-7.814	-0.601
4029496	-7.058	-0.504
4032547	-6.395	-0.492
7790626	-7.037	-0.440

2.3.3 A second of virtual screening was pursued based on analogs to the four hits

We wished to optimize the first cohort of four hits we obtained by finding derivative and analogous compounds. Derivative compounds are those with the same core scaffold as the parent compound but with differences in their functional groups and outer moieties. Analogous compounds are those that have a core scaffold with chemical properties similar to those of the parent compound, but may have entirely different chemical connectivities or compositions.

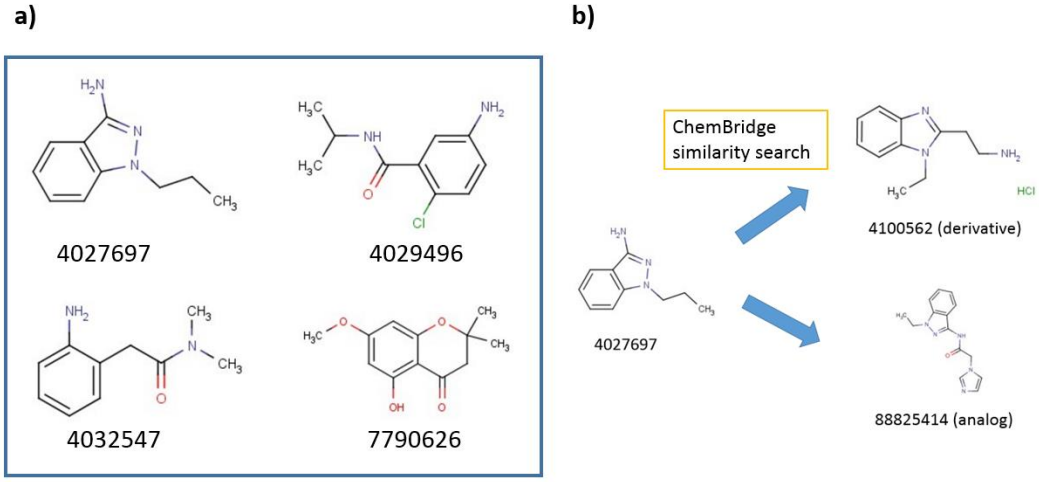


Figure 2.5 Actives obtained from immunoprecipitation assays and second round virtual screening. a) Chemical structures of the four compounds determined to be actives in IKKα inhibition assays. **b)** Schematic of the ChemBridge similarity search protocol detailing the differences between a derivative and an analog of a given compound.

For each of the four parent compounds, we first searched for these types of similar compounds on the ChemBridge database. We also believed that it we wanted to maintain the chemistry of the core scaffold of the four

parent compounds, so we decided to search for derivatives before searching for analogs. In order to obtain between 50 and 200 derivatives of each parent compound, we set the similarity search level at 60%. For parent compound 4032547, a similarity search at that level yielded only 10 results, making it necessary to search for analogs of that compound. Our search produced a total of 401 compounds.

Table 2.3 Similarity searches from the four hits. The four hits from assay 1 are the first round of hits, and from those, we obtained additional compounds to virtually screen. The method and number of compounds for obtained from each search are shown.

LIGAND ID	METHOD OF SELECTION	NUMBER OF COMPOUNDS
4027697	Similarity search	71
4029496	Similarity search	152
4032547	Analogs	120
7790626	Similarity search	58

Since we had found the I650 pocket to be the most promising, we decided to continue prioritize that pocket over the T184 pocket. We performed another round of flexible virtual screening on the pocket with the new set 401 compounds, using the same methods as above.

We then examined the virtual screening results of these second-round compounds. The docking score of each of the best second round compounds was better than that of the parent compound. Therefore, we can see that the top docking scores were improved by finding these derivative and analogous

compounds. This set of compounds represents one step in the process of optimizing our hits into leads.

Furthermore, we observed the major interactions of these second-round compounds. We found that they bonded to several of the same residues as the parent compounds did: D490, R493, Y494, and S504. In addition, though, they also formed new interactions. The new interactions that came up most frequently were W511, Q497, and Q657. Selected data of the interactions is shown in the table below. Therefore, our future assays will be focused on comparing that interact with those residues in the virtual screens, and comparing them against those without such interactions, to determine if the predicted interactions play a significant role in ligand binding.

Table 2.4 Virtual screening results of round 2 of screening. The best docking score of each set of children compounds from the first four hits is shown. Additionally, residues with which those best scoring compounds interact are shown.

PARENT COMPOUND	DOCKING SCORE OF PARENT COMPOUND	BEST DOCKING SCORE OF SECOND-ROUND COMPOUNDS	INTERACTING RESIDUES
4027697	-7.814	-8.032	Y494, Q497, W511, Y651
4029496	-7.058	-8.616	D490, R493, Y494
4032547	-6.395	-8.354	Y494, S504
7790626	-7.037	-8.325	D490, Q647

2.4 Conclusions

The two rounds of virtual screening we completed allowed us to focus on two pockets in IKKα to develop a targeted inhibitor. As we continue this

work, we will go also return to the T184 pocket. Targeting this pocket would be especially exciting because it would allow us to go after an interface that would be directly involved in the formation of the IKKα hexamer. Although the discovery of four hits from the first round of assays was promising, we believe that there are many other methods which could be applied to give us further insight into IKKα-ligand interactions. For example, broad-scale insights into the dynamics of IKKα and, on the canonical pathway, IKKβ[16], have been recently published could provide us with additional conformations of the protein to target[5]. Even more precisely, we could run molecular dynamics simulations on an IKKα chain, dimer, or the hexamer to access additional conformations.

2.5 References

1. Burns, M. B., L. Lackey, M. A. Carpenter, A. Rathore, A. M. Land, B. Leonard, E. W. Refsland, D. Kotandeniya, N. Tretyakova, J. B. Nikas, D. Yee, N. A. Temiz, D. E. Donohue, R. M. McDougle, W. L. Brown, E. K. Law, R. S. Harris. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **494,** 366–370 (2013).

2. Albin, J. S. & R. S. Harris. Interactions of host APOBEC3 restriction factors with HIV-1 in vivo: implications for therapeutics. *Expert Rev. Mol. Med.* **12,** (2010).

3. Roberts, S. A., M. S. Lawrence, L. J. Klimczak, S. A. Grimm, D. Fargo, P. Stojanov, A. Kiezun, G. V. Kryukov, S. L. Carter, G. Saksena, S. Harris, R. R. Shah, M. A. Resnick, G. Getz, D. A. Gordenin. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet.* **45,** 970–976 (2013).

4. Ghosh, G., V. Y.-F. Wang, D. B. Huang, A. Fusco. NF-κB regulation: lessons from structures. *Immunol. Rev.* **246,** 36–58 (2012).

5. Polley, S., D. O. Passos, D.-B. Huang, M. C. Mulero, A. Mazumder, T. Biswas, I. M. Verma, M. Lyumkis, G. Ghosh. Structural Basis for the Activation of IKK1/α. *Cell Rep.* **17,** 1907–1914 (2016).

6. Brenke, R., D. Kozakov, G.-Y. Chuang, D. Beglov, D. Hall, M. R. Langdon, C. Mattos, S. Vajda. Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics* **25,** 621–627 (2009).

7. Kozakov, D., D. Hall, G.-Y. Chuang, R. Cenci, R. Brenke, L. E. Grove, D. Beglov, J. Pelletier, A. Whitty, S. Vajda. Structural conservation of druggable hot spots in protein–protein interfaces. *Proc. Natl. Acad. Sci.* **108,** 13528–13533 (2011).

8. Fehrer, V., M. Gilson, T. Hermann. UCSD Center for Drug Discovery Innovation: Resourcing and Educating Scientists and Students for Academic Discovery Translation. in

9. Friesner, R. A., J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, P. S. Shenken. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **47,** 1739–1749 (2004).

10. Friesner, R. A. R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin, D. T. Mainz. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein- ligand complexes. *J. Med. Chem.* **49,** 6177–6196 (2006).

11. Halgren, T. A. R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard, J. L. Banks. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **47,** 1750–1759 (2004).

12. *Schrödinger Release 2017-1: LigPrep, Schrödinger, LLC, New York, NY, 2017.*

13. Humphrey, W., A. Dalke, K. Schulten. VMD – Visual Molecular Dynamics. *J. Mol. Graph.* **14,** 33–38 (1996).

14. Eargle, J., D. Wright, Z. Luthey-Schulten. Multiple Alignment of protein structures and sequences for VMD. *Bioinformatics* **22,** 504–506 (2006).

15. *Schrödinger Release 2017-1: MS Jaguar, Schrödinger, LLC, New York, NY, 2017.*

16. Schiffer, J. M., T. T. Nguyen, Ö. Demir, G. J. Chan, R. E. Amaro, G. Ghosh. Construing the Dynamic Complexity at a Plausible IKK2-Nemo Interface. in *Biophysical Journal* **112,** 352a (2017).

Chapter 3

Elucidation of the Dynamics of NF-κB-Inducing Kinase and the Discovery of a Potential Allosteric Site

Abstract

NF-κB-inducing kinase (NIK) is a signaling protein in the non-canonical NF-κB pathway whose up-regulation has been linked to dysregulation of cell death. This work presents all-atom molecular dynamics (MD) simulations initiated from crystal structures of NIK bound to active site inhibitors. These simulations reveal that the active site adopts two major conformations. Furthermore, simulations reveal that the active site inhibitors not only obstruct the ATP binding pocket, but also alter the conformations in NIK domains critical to ATP binding and catalysis. In addition, in our simulations, a previously developed lead compound was serendipitously displaced from the active site into a previously unexplored pocket, near R509. We initiated further trajectories of NIK from the frames containing the R509 pocket and reveal here that this pocket to be a potential allosteric site.

3.1 Introduction

3.1.1 The role of NIK in the non-canonical NF-κB pathway

The non-canonical nuclear factor κB (NF-κB) pathway regulates the transcription of the protein APOBEC3B (A3B), whose overexpression has been linked to oncogenic behavior in cells and has been found in over 50 percent of breast cancer tumors[1–3]. Inhibition of the proteins on this pathway

could diminish overexpression of A3B and prevent tumor evolution. NF-κB inducing kinase (NIK) is a serine/threonine protein kinase, also known as mitogen-activated kinase kinase kinase 14 (MAP3K14), that binds to IKKα to activate the cleavage of p100 to p52. The p52-RelB complex then transcribes the A3B gene. An inhibitor of NIK is then desirable for attenuating overactivity of this pathway.

While active site, or type I, inhibitors of NIK, e.g. 6-alkynylindoline, and modified type I inhibitors such as those developed by Castanedo et al. exist, type III kinase inhibitors would be more attractive because they do not target the active site and can therefore be designed to target this kinase specifically[4,5]. So far, trametinib is the only MAP kinase inhibitor approved by the FDA[6,7]. However, several other MAP kinase inhibitors, such as BIRB 796[8] and XL518/cobimetinib[9], are currently in development or were so in the past.

3.1.2 An overview of NIK domains implicated in ATP catalysis

We performed all-atom molecular dynamics (MD) simulations in explicit solvent. The preliminary focus of our analysis of the simulations was on residues that have been identified to be important for ATP binding of NIK, including R408, F411 and G412, which are found on the NIK phosphate loop and form hydrogen bonds with the ATP phosphates[10–12]. There are the catalytically important E440, on the α_C helix, which forms a salt bridge with K429, and N520, which orients D515, a catalytic base residue[11,13,14]. In its active conformation, when Mg^{2+} is bound to NIK, the ion chelates with D534

and N520. Finally, the activation loop (A-loop) has been reported to play a role in substrate binding[15]. The DFG motif (residues 534-536) at the N-terminal end of the A-loop, has been scrutinized in past work, as it has been hypothesized that when it is in the out conformation, kinases are usually held in their inactive state, due to the steric hindrance of the phenylalanine with phosphates in ATP[8,12,16,17]. Furthermore, in the out conformation, the aspartate in the DFG motif points outward and away from the ATP-binding site[7].

In our studies, we looked at which of these residues and loops the inhibitors T28 and 13V interacted with or displaced. From this finding, we set out to determine if T28 could act as an allosteric inhibitor when bound to the R509 pocket.

To our knowledge, there are no leads for allosteric inhibitors of NIK. The inhibitors targeting the ATP binding site or the active site shared by all kinases have been struggling to achieve target specificity due to the large number of kinases present in human body[7,18,19]. Therefore, we desired to determine if NIK contained allosteric druggable pockets, and if the R509 pocket was one of those sites.

3.2 Materials and Methods

3.2.1 Molecular dynamics simulations and trajectory analysis

We simulated a total of four systems. Two co-crystal structures of inhibitors discovered by Li et al. bound to human NIK served as the basis for

our MD simulations: one of the NIK-T28 complex (PDB 4IDT) and one of the NIK-13V complex (PDB 4IDV)[20]. We wished to elucidate the dynamics of these holo NIK systems, along with a ligand-less, apoenzyme NIK system derived from crystal structure 4IDT, in which we omitted ligand T28 from the PDB file. The systems were first parameterized in LEaP using the Generalized Amber Force Field[21].

These systems were solvated in cubic TIP3P water boxes using AMBER14. The systems were then minimized over 13,000 steps, heated to the physiological temperature of 310 K over a period of ps, and allowed to equilibrate for 750 ps. When simulation setup was complete, we realized that sometime during the parameterization process that in the simulation of the NIK-T28 complex, ligand T28 had been serendipitously relocated to the R509 pocket. Although the ligand was no longer in accordance with its previously reported binding pose in the active site, we decided to proceed with the simulation of this NIK-T28 R509 system anyway. We later added a simulation of T28 when bound to the NIK active site, designated the NIK-T28 system.

We performed all-atom molecular dynamics (MD) simulations on all four of these systems: the apoenzyme, NIK-T28, NIK-13V, and NIK-T28 R509. For each system, triplicate simulations, each 150 ns, were performed to sample the conformational landscape of NIK and of any inhibitor bound to it[22–25]. The active site volume was analyzed with the Pocket Volume Measurer (POVME), and clustering was performed based on the active site volume and root-mean-squared-deviation (RMSD)[26].

Further analysis was performed using the GROMACS tool g_covar to perform principal component analysis, and several tools in Visual Molecular Dynamics to measure atom-atom distances and dihedral angles, and find salt bridges[27–29]. Lastly, FTMap was used to identify changes in druggable hot spots in the NIK-T28 R509 system[30,31].

3.3 Results and Discussion

3.3.1 Analysis of the active site suggests differences in behavior based on presence and type of ligand

While both ligands T28 and 13V are known to act as type I inhibitors by occupying the active site and blocking ATP from binding, we wanted to know in greater detail how the presence of these inhibitors would change the binding pocket, and to look for differences between the volumes and dynamics of the apo and holo pockets. Analysis of the active site with POVME revealed that the active site changes conformation differently depending on whether an inhibitor is present, and on inhibitor structure and position.

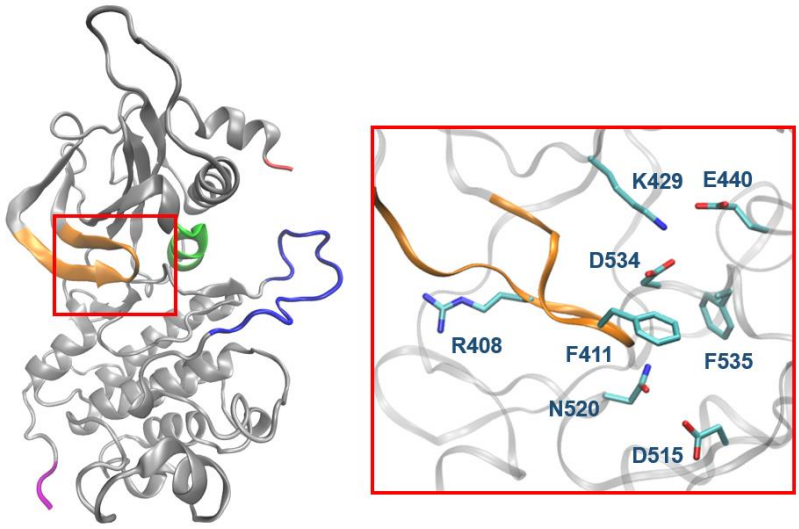


Figure 3.1 An overview of the domains of interest of NF-κB inducing kinase. In orange is the phosphate loop (P-loop), in green, the α_C helix, and in blue, the activation loop (A-loop). The N-terminus is at the top right in red; the C-terminus in the bottom left in magenta. The inset shows the P-loop in orange and important catalytic residues in sticks.

As previously mentioned, the volume of the active site of each system was measured over the duration of their MD trajectories (**Fig. 3.2a**). Both the NIK-13V and NIK-T28 R509 systems were fairly stable in their volumes over the simulation time. We note that the NIK-13V system has a larger average volume than that of the apo system. As expected, the NIK-13V active site volume was greater than that of NIK-T28 because ligand 13V is larger than T28. The active sites of all three inhibitor-bound systems, furthermore, were on average smaller in volume than that of the apo system. Additionally, the active site volumes of the NIK-13V and NIK-T28 R509 systems demonstrated a bimodal distribution. This is an indication that these two systems may therefore have two distinct major conformations.

Principal component (PC) analysis was performed based on the active site conformation for each of the four systems (**Fig. 3.2b**). We find that the existence of two conformations suggested in the volume histogram for the NIK-T28 R509 system is also supported by the PC data. Brighter areas on the PC plots indicate a greater number of MD frames with an active site assigned to those PC values. Therefore, those brightest areas represent active sites with stable conformations.

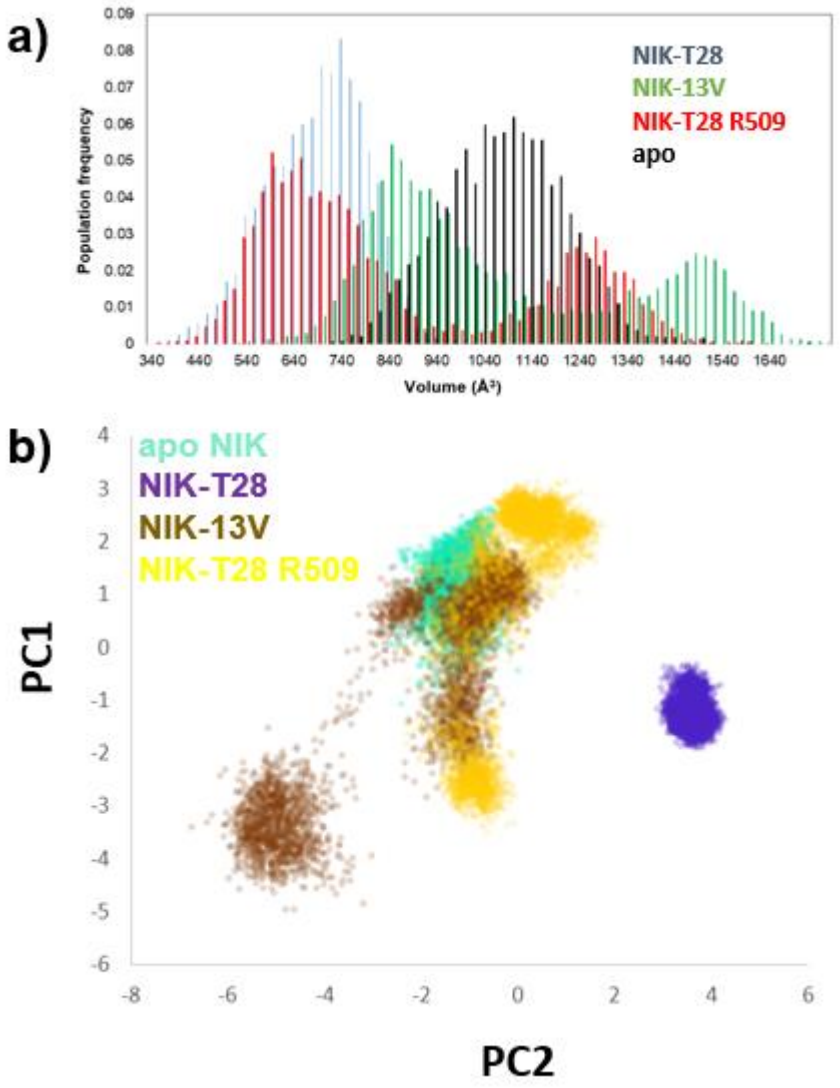


Figure 3.2 Quantitative analysis of the NIK active site. a) Histograms of active site volumes for the four NIK systems. The apo system is black, NIK-13V is green, NIK-T28 is blue, NIK-T28 R509 is red. **b)** Principal component analysis of NIK active site volume. The NIK-13V and NIK-T28 R509 systems display two distinct regions of PC density.

When we first consider how each system behaves in principal component space, we see the relative diffusivity of the graph of the NIK-13V system compared to that of the NIK-T28 and apo systems, and we conclude

that the active site in the NIK-13V system samples a wider range of conformations. The NIK-13V PC plot shows that its active site has sampled two distinct areas of PC space, circled in red in **Fig. 3.2b**. This supports our hypothesis formulated from the active site volume histogram: the bimodal distribution of those volumes in the NIK-13V system is symptomatic of the existence of two major active site conformations. In contrast, the PC space covered in the NIK-T28 system is compact, indicating that, compared to the apo and NIK-13V systems, relatively little conformational space is covered.

Lastly, the PC behavior of NIK-T28 R509 system is similar in some ways to that of the NIK-13V system. It displays diffusive behavior and two clusters of PC density, thereby also indicating at least two active site conformations. Changes in ATP-catalyzing NIK domains near the active were explored in our analysis, and their relevance to the NIK active site, will be discussed later.

Therefore, from the preliminary observations about the active site volume histogram and the PC graphs, we focused our analyses to further explore causes of two phenomena: the differences between the orthosteric ligand-bound systems and the apo system, and the differences between the NIK-T28 R509 and apo systems.

3.3.2 When bound to the active site, inhibitors NIK-T28 and 13V induce different changes in the phosphate loop

In searching for reasons for the differences between the orthosteric ligand systems (NIK-T28 and 13V) and the apo system, we first analyzed the dynamics of the phosphate loop (P-loop). Also called the glycine-rich loop, the P-loop is a turn of 11 residues fond between the β1 and β2 sheets across a variety of kinase families[10,26]. In NIK, the P-loop consists of residues from H402 to E413. Its glycine backbone interacts with the phosphates in ATP, stabilizing the substrate's position in the active site[32,33].

The P-loop revealed itself to be one of the most noteworthy motifs in NIK. We quantified changes in the bending of the P-loop through its distance from the C_α of K517 (**Fig. 3.3a**), which, being on a relatively immobile loop on the C-terminal end of the α_E helix, served as a reference. We also quantified those changes through P-loop torsion by measuring the dihedral of the C_α of four residues at the apex of the loop, R409, S410, F411, and E413 (**Fig. 3.3b**). In our studies, the P-loop was found to undergo a conformational change to varying degrees, depending on whether or not the system was ligand-bound, and if so, on the type and position of the ligand. In the NIK-13V system, the P-loop adopts an outward-facing conformation in which the apex of loop bends away from the active site (**Fig. 3.3c**). Additionally, these displaced conformations were found in the representative frames of the largest active site volume-based clusters of the system trajectories. Therefore, we believe that this conformation is not a rare event, and that the

P-loop remains in this conformation for a significant portion of the MD trajectories.

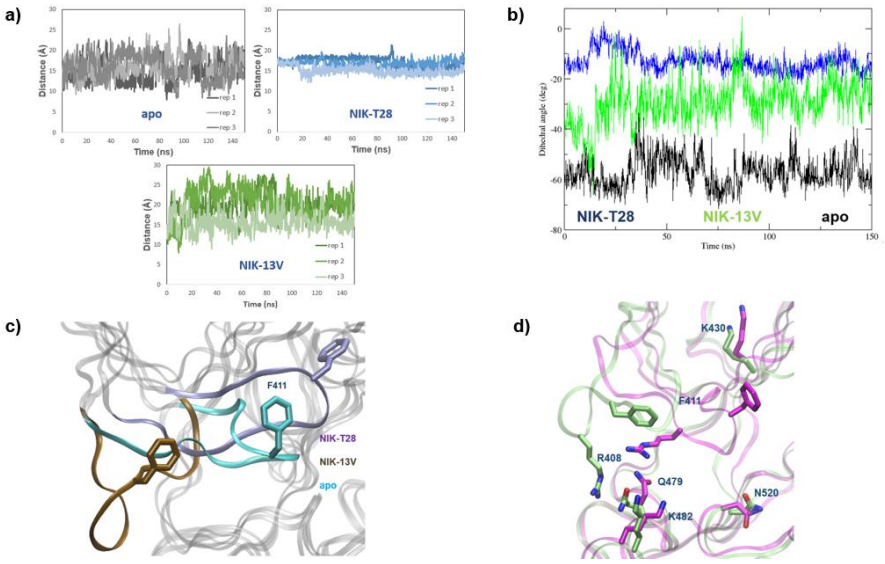


Figure 3.3 The phosphate loop changes conformation in different ways depending on the ligand bound to the active site. a) Change in distance from P-loop residue F411 to K517 across individual trajectories. **b)** Dihedral angle of four residues at the turn of the P loop: R408, S410, F411, and E413. **c)** The orthosteric ligands T28 and 13V induce different conformational changes in the P-loop, causing it to adopt displaced or twisted conformations and opening up the active site. **d)** The NIK active site adopts two conformations when bound to ligand 13V. Conformation 1 is represented with lime-colored carbons, and conformation 2 is represented with magenta carbons.

From the quantitative results and visual observations, what we find interesting is that the P-loop is in this bent conformation when inhibitor 13V, but not T28, is bound to NIK. Ligand T28, when placed in the active site, tends to force the P-loop away from the active site and consistently increase the F411-K517 distance compared to the distance in the apo system, the loop in the NIK-13V system is much more dynamic, alternating between the

upward bent P-loop and a downward bent P-loop that traps the ligand in the active site. P-loop bending away from the active site, either distal to the active site, would be ideal for an inhibitor because the loop glycines, along with F411, would be displaced from their position around the active site, therefore mitigating their role in stabilizing the phosphates in ATP from entering the active site.

3.3.3 The two active site conformations in the NIK-13V system mainly result from the bending of the P-loop

Observation of the MD trajectories revealed further behavior about the P-loop. The loop also appears to pinch or twist, as illustrated by the dihedral angle of the loop. This results in the P-loop taking a half-folded conformation, although this conformation does not exactly the conformation adapted MAP4K4 in Guimaraes et al.'s study[10]. This folded conformation may be due to the hydrogen bond that forms between F411 and E413 only in the apo system, which would be consistent with studies on the correlation between having an aromatic residue at that position and P-loop folding[10]. This hydrogen bond may be disrupted when an inhibitor binds in the active site. Instead, in the inhibitor-bound systems, a new hydrogen bond forms between the backbones of G412 and K430, which then bends the P-loop upwards away from the active site.

Furthermore, we hypothesize that ligand 13V causes conformational changes in the P-loop and other residues in the active site area (**Fig. 3.3d**).

These changes lead to the presence of two conformations in the NIK-13V system, which were suggested by the active site volume histogram and the PC plots. In the first conformation, R408 and M469, the gatekeeper residue, move away from the active site pocket. This leads to the opening of pocket when both the methionine side chains swing out of the active site. Similarly, the side chain of R408, which is needed for nucleotide binding[4], is displaced and, as such, allows Q479 and K482 to close over the active site. Furthermore, we observed that the methoxy tail of 13V interacts with Q479, and the side chain of R408 closes over the inhibitor, preventing its exposure to solvent. Cowan-Jacob et al. reported that a hydrophobic cage forms around inhibitor, holding it in place in the binding pocket[32]. We believe that it is the interaction between inhibitor 13V and R408 that most contributes to its higher potency. The ether oxygens of the methoxy tail on 13V make their closest approaches to the arginine side chain nitrogens at 4.31 and 5.93 Å, and may be able to hydrogen bond with their hydrogens.

In the second conformation, the side chains of R408 and F411 point in opposite directions, preventing R408 from shielding the active site. Q479 and R408 have a head-on interaction that allows the space below the active site to remain open. Furthermore, the guanidinium side chain of N520 swings from an outward-facing to and inward, active site-facing conformation, which would allow the side chain to be appropriately oriented to chelate with Mg^{2+} during ATP catalysis.

In the NIK-T28 system, the presence of inhibitor T28 causes several changes to the residues in the active site. The long side chain of K429 moves out of the way of the active site, allowing a pocket above the P-loop to open. Being blocked by the inhibitor, Q479 is unable to form a salt bridge with R408. Overall, though, the active site volume is less in the NIK-T28 system than in the apo system because the residues in its active site form interactions with ligand T28 that draw residues in the active site closer together: N520 is another such residue that points towards the active site, and whose side chain nitrogen forms a salt bridge with S410 at a distance of 3.75 Å, which can be compared to the S410-N520 distance in the apo system, which never approaches any closer than 6.48 Å.

3.3.4 Interactions between the activation loop and α_C helix change when an inhibitor is bound

The activation loop (A-loop) can be delimited by residues Q542 to T559 in NIK. It consists of a chain of residues found between the α_E and α_F helices[12,32]. Several residues before its N terminus is the Asp-Phe-Gly (DFG) motif, located at positions 534-536 in human NIK.

Observation of the MD trajectories of the A-loop led to the conclusion that the A loop is less mobile in inhibitor-bound systems, while being less ordered in the apo conformation. This observation is supported by RMSD analysis performed on the residues of the A-loop, as compared to the starting conformation. The steady level of the plots of the inhibitor-bound systems in

Figs. 3.4a, and **3.4b** show that the A-loop stays in one major conformation. However, compared to the system in which the ligand binds to the active site, the A-loop does not show as much stability in the allosteric system.

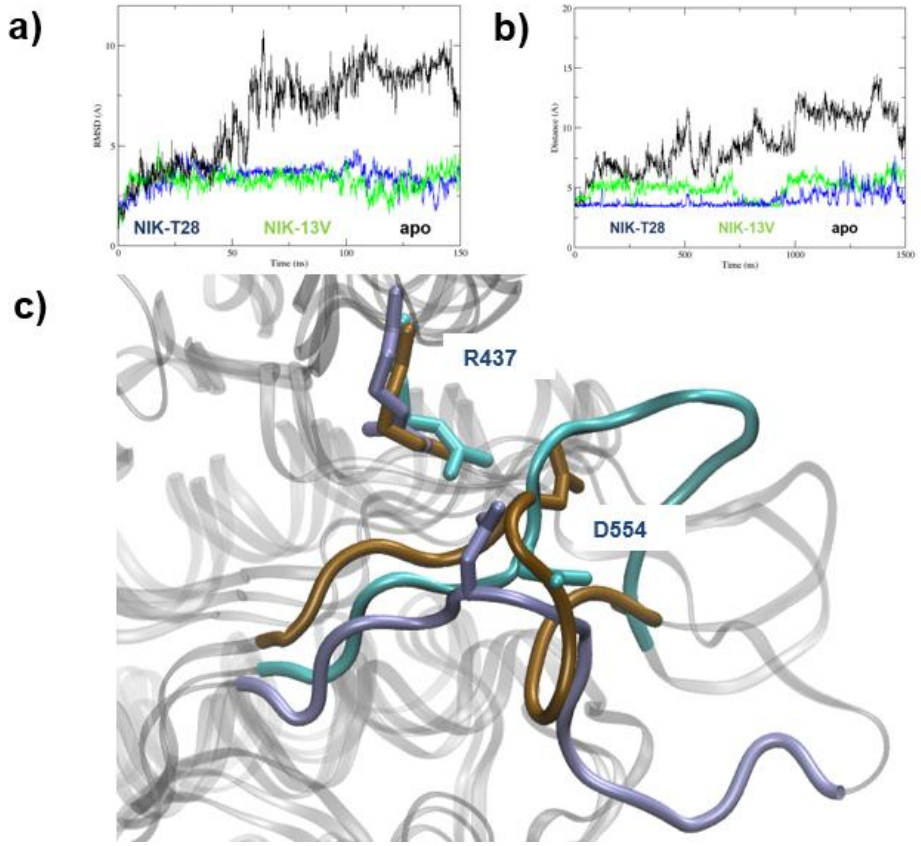


Figure 3.4 Changes in the activation loop. a) RMSD of the A loop over 150 ns of MD simulation. The apo system is black, NIK-13V is green, NIK-T28 is blue, NIK-T28 R509 is red. **b)** Distance of the D554-R437 salt bridge in the MD simulations of four NIK systems. **c)** The activation loop can adopt a pinched, in conformation, as in the apo system (cyan), or a relaxed conformation, as in the holoenzyme systems. The salt bridge is between D554 and R437.

The A-loop can adopt in and out conformations[4,34]. When the A-loop is in the in conformation, the portion of the loop between residues K548 to G558 is closer to the active site, while when it is in the out conformation, the

loop farther from the active site[17]. The three ligand-bound systems the A-loop exhibited a conformation unlike those published in the literature; whereas the apo system A-loop was often extended away from the active site in the in conformation, the ligand-bound systems, the A-loop would un-pinch and flatten, bringing D554 towards the α_C helix. Therefore, the A-loop in the inhibitor-bound systems appears to moving away from the in conformation.

In addition, crucial catalytic residues, when compared to the apo system, were found to be in positions and orientations that held the A-loop in this conformation. A portion of the loop, located from L549 to P555, forms a single helical turn, and this turn has been shown to stabilize the position of the α_C helix by packing underneath it[32].

We found a residue on this turn to be of special importance, D554. This aspartate is crucial because it forms a salt bridge with R437, which stabilizes the loop in its hairpin form. Additionally, R437 is on the α_C helix, and this salt bridge stabilizes the conformations of both this helix and the A-loop. The α_C helix needs to be oriented so that ATP can bind to the enzyme[4]. In the NIK-inhibitor systems, these two residues appear to stay in fairly close proximity consistently, while this is not seen in systems lacking a ligand. Therefore, we are led to believe that the salt bridge may be stabilized by the presence of an inhibitor. Compared to the apo system, D554 was found to be consistently closer to R437 in the inhibitor-bound systems (**Fig. 3.4b**). This stabilizing effect may prevent ATP binding by keeping the channel between the α_C helix and activation loop more constrained.

Furthermore, we observed F535, a residue of the DFG motif, which is located at the N-terminal end of the activation loop. In the NIK-13V system, F535 consistently extended into the hydrophobic pocket adjacent to the active site occupied by 13V but not T28, and hydrogen bond analysis showed that a hydrogen bond formed between the hydroxy hydrogen of ligand 13V and the backbone of F535, this being the same hydroxy group that bonds to E440 on the α_C helix.

Lastly, it was found that a long-lasting hydrogen bond is found between F535 and E440 in the NIK-13V system. We also found that hydrogen bonds formed between the hydroxyl group and both E440 and K429, which also helps stabilize the salt bridge. This combination of electrostatic interactions stabilizes conformation of the back pocket and the binding pose of ligand 13V. Compared to ligand T28, which does not have this hydroxyl group extension into the back pocket, it therefore makes sense that 13V is a more potent inhibitor. Therefore, ligand 13V not only binds to these two back pocket ligands, but disrupts the hydrogen bond between the two residues, thereby opening up the pocket.

3.3.5 Binding of T28 to the R509 pocket induces changes in previously discussed domains

As previously discussed, the binding of the NIK-T28 to the R509 system intrigued us, and we wished to understand its effect on the conformational space sampled by NIK. We began by searching for druggable

hotspots on the NIK surface with FTMap to see if there would be any changes the locations of druggable pockets on NIK. As controls, we used the 4IDT, 4IDV crystal structures. Additionally since we also wanted an additional non-ligand-bound crystal structure to compare against the NIK-T28 and NIK-13V systems, we chose 4G3D, the apo structure solved from human NIK with the best resolution[4].

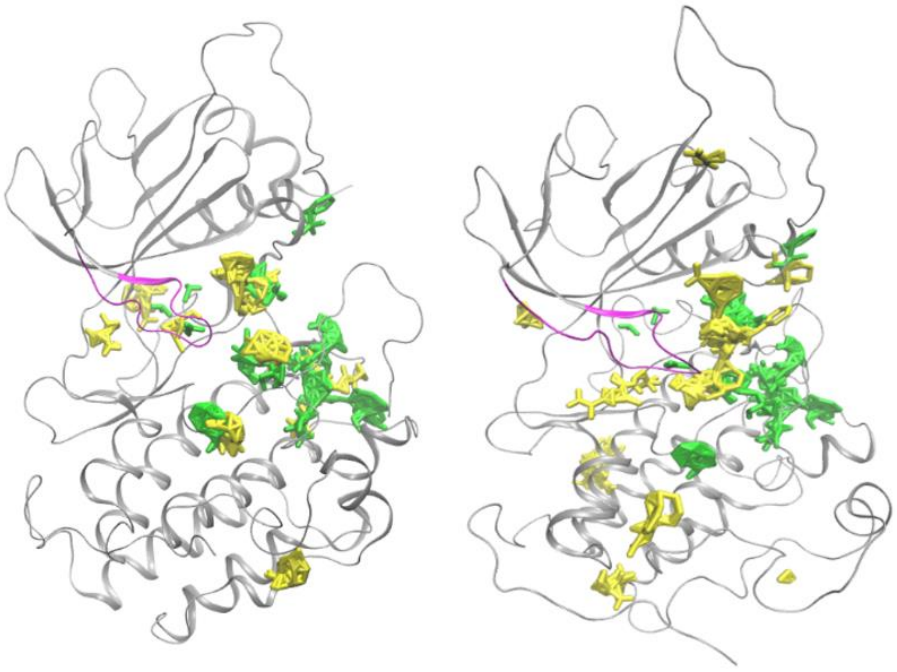


Figure 3.5 Comparison of FTMap probes docked to NIK MD simulation structures. Probes docked to the apo system are in yellow, and probes docked to NIK-T28 R509 system are in green. The P-loop is highlighted in purple. The protein traces are taken from apo NIK. Probes docked to the centroid frame of the **a)** most-populated active site volume-based cluster and **b)** most-populated RMSD-based cluster.

We then searched the conformations sampled in our MD simulations for druggable regions on NIK. To find representative conformations sampled in our MD trajectories, we used RMSD- and active site volume-based

clustering. All structures were searched for druggable hot spots using FTMap[30,31]. In the FTMap results obtained from the crystal structures, we found that small-molecule probes docked to the active site (or ATP binding site) in all NIK-T28, NIK-13V, and 4G3D NIK crystal structures, as can be expected.

From our examinations of the NIK-13V and 4G3D NIK crystal structures with FTMap, the region of NIK corresponding to the MEK allosteric binding site was also populated by many probes. Furthermore, this pocket had probes binding to it in all three of the representative frames from the most-populated RMSD-based cluster of each of the 4IDT, 4IDV, and apo system MD trajectories, indicating that it would be druggable across a wide area of the conformational landscape of NIK. In the active site volume-based results, meanwhile, the apo and NIK-13V systems showed druggability in that site.

While comparing the apo and NIK-T28 R509 systems from their FTMap results, we noticed that there was a shift in which NIK surface regions were druggable hotspots when T28 was bound to the R509 pocket. Our first observation was that there was a noticeable shift in the druggable regions of the NIK-T28 R509 system compared to the apo system. Even more excitingly, in both the RMSD- and active site volume-based clustering, several probes from the apo FTMap were usually found in the active site, indicating that it would be expected for molecules to bind there.

However, in the NIK-T28 R509 system, no probes were bound to the active site. This would indicate that the active site was no longer amenable to having molecules bind there, and suggests that the binding of T28 to the R509 pocket makes binding to the active site less likely.

3.3.6 The phosphate loop, and other catalytic residues, undergo conformational changes when NIK-T28 binds to a non-orthosteric pocket.

There is further evidence that the binding of T28 to the R509 pocket induces conformational changes in NIK. The NIK-T28 R509 active site volume histogram shows that there is a bimodal distribution of active site volumes in that system. Moreover, the PC plot, compared to the PC graph of NIK-13V, more prominently displays two closely-set clusters of PC density. As with the NIK-13V system, we set out to explore the reasons for this behavior.

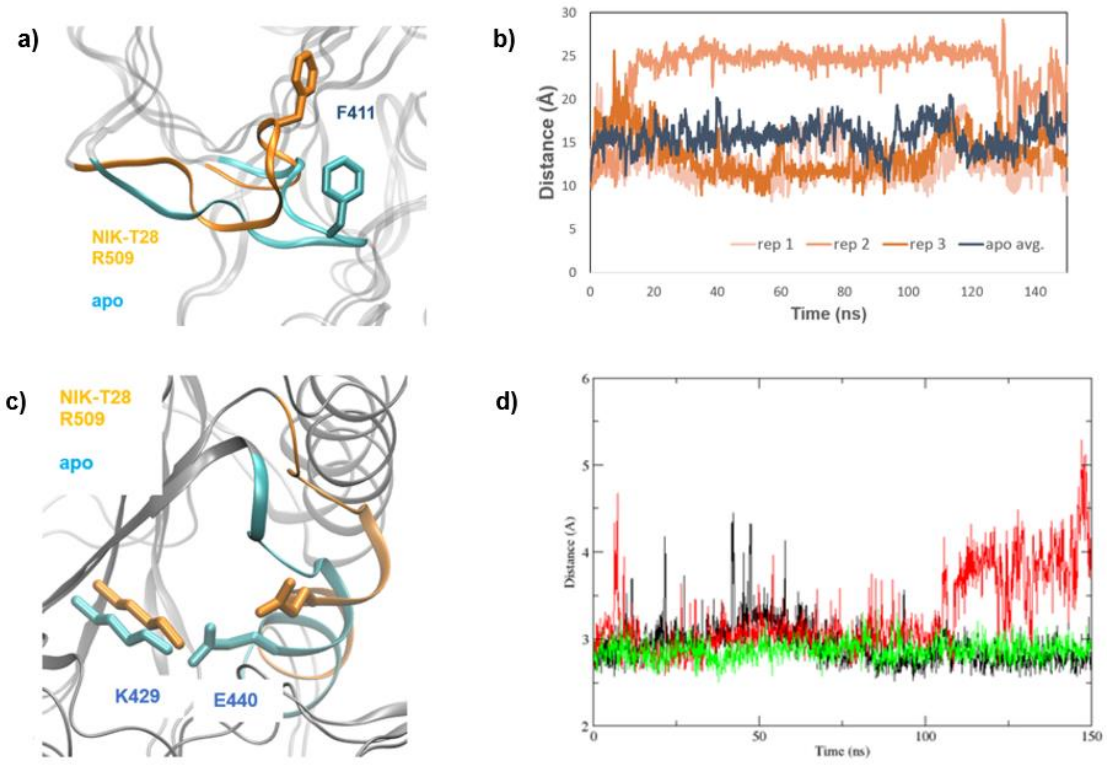


Figure 3.6 Comparison of the P-loop in the NIK-T28 R509 and apo systems. (a) The P-loop adopts a bent conformation when ligand T28 is bound to the non-orthosteric R509 pocket. Residue F411 is depicted in sticks to serve as a reference for comparison between the NIK-T28 R509 and apo systems. **(b)** Comparison of the P-loop to α_D helix distance in the NIK-T28 R509 simulation separated by trajectory, with the average distance in the apo simulation as reference. **(c)** Comparison of position of residues K429 and E440. K429 normally stabilizes the α_C helix, shown at right. **(d)** The significant change in the K429-E440 distance in the NIK-T28 R509 trajectory suggests that the more mobile α_C is displaced in the NIK-T28 R509 system.

In our observations of the P-loop, we noticed that NIK-T28 R509 has a dihedral similar to that of the NIK-13V system (**Fig. 3.6a**), providing further evidence that the P-loop undergoes similar conformational changes in those two systems. The distance between the P-loop and the stable α_D helix is also noteworthy. The distance is more stable during a given trajectory in the NIK-13V system, indicating less fluctuation and bending in the P-loop. However,

the amount of P-loop bending varies widely between trajectories (**Fig. 3.6b**). Therefore, while the P-loop appears to rapidly change conformation—from bent to unbent—in the NIK-13V system, it more steadily remains in one conformation in the NIK-T28 R509 system, and can stay in the bent or unbent conformation for the entire 450 ns-duration of simulation.

Following the conformational changes we saw in certain catalytic residues in the NIK-13V systems, we wanted to know if those changes also occurred in the same residues but with an allosteric inhibitor. Because of its role in ATP catalysis and its differences in orientation in the NIK-13V system, we decided to examine the behavior of N520 in the NIK-T28 R509 system as well. We tracked this residue in the NIK-T28 R509 system as well and compared it to its behavior in the apo system. We saw that its guanidinium group shifted away from the active site. This shift also coincides with the increase in distance between K429 and E440.

In addition to changes in the A-loop, allosteric kinase inhibitors have been reported to shift the α_C helix (Jia) into an inactive conformation because the correct positioning of E440 on this helix is necessary for ATP catalysis (de Leon-Boenig). This glutamate aids in the catalysis of ATP through its salt bridge with K429, and positions the lysine to form dipole-dipole interactions with the alpha and beta phosphates of ATP[12]. In our simulations, the lengths of the salt bridge in the NIK-T28 and NIK-13V systems appear to be relatively similar to each other. However, the distance between E440 and K429 fluctuates more in the NIK-T28 R509 system than in the orthosteric ligand

and apo systems (**Fig 3.6c**), which would preclude salt bridge formation between those two residues. This, in turn, would prevent the α_C helix from being in a position amenable for ATP binding. In our NIK-T28 R509 simulations, because of the shifting of the N-terminal end of the α_C helix, E440 moves farther away from K429, displacing it from the position required to catalyze ATP binding (**Fig. 3.6d**). Therefore, we see this as another piece of evidence that the placing an inhibitor in the R509 pocket may induce NIK to be an allosteric state, and that R509 is a potential allosteric pocket.

3.4 Conclusions

In order to rationally design a NIK inhibitor, we strived to find NIK conformations that are not represented in crystal structures. By running MD simulations of NIK, we have found changes in the apo active site that tell us about the nature of NIK.

Although both ligands 13V and T28 occupy the active site, they cause dissimilar changes in NIK motifs necessary for ATP-binding. Most notably, T28 causes the P-loop to be drawn in towards the active site, while 13V does the opposite.

We have also found a potential allosteric site. Targeting this allosteric site yielded changes to the P-loop, A-loop, and α_C helix that are suggestive of an inactive NIK conformation. We therefore hypothesize that T28 can allosterically inhibit NIK. Further work needs to be done to understand the mechanisms behind the allosteric effects of ligand T28 on the active site.

This chapter, in full, is in preparation for publication of the material as it may appear in “Elucidation of the Dynamics of NF-κB Kinase and the Discovery of a Potential Allosteric Site” by Chan, Garrett J.; Demir, Özlem; Schiffer, Jamie M.; and Amaro, Rommie E. in 2017. This chapter is included with the permission of all of the authors.

3.5 Supplementary Material

Supporting Movie 3.1 Location of inhibitor T28 when bound to the NIK R509 pocket

Supporting Movie 3.2 Bending of the NIK P-loop due to the binding of R509

3.6 References

1. Odqvist, L. M., Sanchez-Beato, S. Montes-Moreno, E. Martin-Sanchez, R. Pajares, L. Sanchez-Verde, P. L. Ortiz-Romero, J. Rodriguez, S. M. Rodriguez-Pinilla, F. Iniesta-Martinze, J. C. Solera-Arroyo, R. Ramos-Asensio, T. Flores, J. M. Palanca, F. G. Bragado, P. D. Franjo, M. A. Piris. NIK Controls Classical and Alternative NF- B Activation and Is Necessary for the Survival of Human T-cell Lymphoma Cells. *Clin. Cancer Res.* **19,** 2319–2330 (2013).

2. Ranuncolo, S. M., S. Pittaluga, M. O. Evbuomwan, E. S. Jaffe, B. A. Lewis. Hodgkin lymphoma requires stabilized NIK and constitutive RelB expression for survival. *Blood* **120,** 3756–3763 (2012).

3. Thu, Y. M., Y. Su, J. Yang, R. Splittgerber, S. Na, A. Boyd, C. Mosse, C. Simons, A. Richmond. NF-κB inducing kinase (NIK) modulates melanoma tumorigenesis by regulating expression of pro-survival factors through the β-catenin pathway. *Oncogene* **31,** 2580–2592 (2012).

4. de Leon-Boenig, G. K., K. Bowman, J. A. Feng, T. Crawfod, C. Everett, Y. Franke, A. Oh, M. Stanley, S T. Staben, M. A. Starovasnik, H. J. Wallweber, J, Wu, L. C. Wu, A. R. Johnson, S. G. Hymowitz. The Crystal Structure of the Catalytic Domain of the NF-κB Inducing Kinase Reveals a Narrow but Flexible Active Site. *Structure* **20,** 1704–1714 (2012).

5. Castanedo, G. M., N. Blaquiere, M. Beresini, B. Bravo, H. Brightbill, J. Chen, H.-F. Cui, C. Eigenbrot, C. Everett, J. Feng, R. Godemann, E. Gogol, S. Hymowitz, A. Johnson, N. Kayagaki, P. B. Kohli, K. Knüppel, J. Kraemer, S. Krüger, P. Loke, P. McEwan, C. Montalbetti, D. A. Roberts, M. Smith, S. Steinbacher, S. Sujatha-Bakhsar, R. Takahasi, X. Wang, L. C. Wu, Y, Zhang, S. T. Staben. Structure-Based Design of Tricyclic NF-κB Inducing Kinase (NIK) Inhibitors That Have High Selectivity over Phosphoinositide-3-kinase (PI3K). *J. Med. Chem.* **60,** 627–640 (2017).

6. Abe, H., S. Kikuchi, T. Iida, N. Nagahashi, K. Maeda, J. Sakamoto, N. Matsumoto, T. Miura, K. Matsumura, N. Seki, T. Inaba, H. Kawasaki, T. Yamaguchi, R. Kakfuda, T. Nanayama, H. Kurachi, Y. Hori, T. Yoshida, J. Kakegawa, Y. Watanabe, A. G. Gilmartin, M. C. Richter, K. G. Moss, S. G. Laquerre. Discovery of a Highly Potent and Selective MEK Inhibitor: GSK1120212 (JTP-74057 DMSO Solvate). *ACS Med. Chem. Lett.* **2,** 320–324 (2011).

7. Wu, P., T. E. Nielsen, M. H. Clausen. FDA-approved small-molecule kinase inhibitors. *Trends Pharmacol. Sci.* **36,** 422–439 (2015).

8. Pargellis, C. L. Tong, L. Churchill, P. F. Cirillo, T. Gilmore, A. G. Graham, P. M. Grob, E. R. Hickey, N. Moss, S. Pav. Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat. Struct. Biol.* **9,** 268–272 (2002).

9. Rice, K. D. N. Aay, N. K. Anand, C. M. Blazey, O. J. Bowles, J. Bussenius, S. Costanza, J. K. Curtis, S. C. Defina, L. Dubenko, S. Engst, A. A. Joshi, A. R. Kennedy, A. I. Kim, E. S. Koltun, J. C. Lougheed, J. L. Manalo, J. Martini, J. M. Nuss, C. J. Peto, T. H. Tsang, P. Yu, S. Johnston. Novel Carboxamide-Based Allosteric MEK Inhibitors: Discovery and Optimization Efforts toward XL518 (GDC-0973). *ACS Med. Chem. Lett.* **3,** 416–421 (2012).

10. Guimarães, C. R. W., B. K. Rai, M. J. Munchof, S. Liu, J. Wang, S. K. Bhattacharya, L. Buckbinder. Understanding the Impact of the P-loop Conformation on Kinase Selectivity. *J. Chem. Inf. Model.* **51,** 1199–1204 (2011).

11. Liu, J. A. Sodom, X. Min, Z. Cao, X. Gao, M. Ayres, F. Lee, P. Cao, S. Johnstone, O. Plotnikova, N. walker, G. Chen, Z. Wang. Structure of the Nuclear Factor B-inducing Kinase (NIK) Kinase Domain Reveals a Constitutively Active Conformation. *J. Biol. Chem.* **287,** 27326–27334 (2012).

12. Mol, C. D., D. R. Dougan, T. R. Schneider, R. J. Skene, M. L. Kraus, D. N. Scheibe, G. P. Snell, H. Zou, B.-C. Sang, K. P. Wilson. Structural Basis for the Autoinhibition and STI-571 Inhibition of c-Kit Tyrosine Kinase. *J. Biol. Chem.* **279,** 31655–31663 (2004).

13. Demchenko, Y. N., L. A. Brents, Z. Li, L. P. Bergsagel, L. R. McGee, M. W. Kuehl. Novel inhibitors are cytotoxic for myeloma cells with NFkB inducing kinase-dependent activation of NFkB. *Oncotarget* **5,** 4554–4566 (2014).

14. Johnson, L. N., M. E. Noble, D. J. Owen. Active and inactive protein kinases: structural basis for regulation. *Cell* **85,** 149–158 (1996).

15. Nolen, B., S. Taylor, G. Ghosh. Regulation of protein kinases: controlling activity through activation segment conformation. *Mol. Cell* **15,** 661–675 (2004).

16. Nagar, B. c-Abl tyrosine kinase and inhibition by the cancer drug imatinib (Gleevec/STI-571). *J. Nutr.* **137,** 1518S–1523S (2007).

17. Treiber, D. K. & N. P. Shah. Ins and Outs of Kinase DFG Motifs. *Chem. Biol.* **20,** 745–746 (2013).

18. Roskoski, R. A historical overview of protein kinases and their targeted small molecule inhibitors. *Pharmacol. Res.* **100,** 1–23 (2015).

19. Zhao, Z. H. Wu, L. Wang, Y. Liu, S. Knapp, Q Liu, N. S. Gray. Exploration of Type II Binding Mode: A Privileged Approach for Kinase Inhibitor Focused Drug Discovery? *ACS Chem. Biol.* **9,** 1230–1241 (2014).

20. Li, K. L. R. McGee, B. Fisher, A. Sudom, J. Liu, S. M. Rubenstein, M. K. Anwer, T. D. Cushing, Y. Shin, M. Ayres, F. Lee, J. Eksterowicz, P. Fauler, B. Waszkowycz, O. Plotnikova, E. Farrelly, S. Xiao, G. Chen, Z Wang. Inhibiting NF-κB-inducing kinase (NIK): Discovery, structure-based design, synthesis, structure–activity relationship, and co-crystal structures. *Bioorg. Med. Chem. Lett.* **23,** 1238–1244 (2013).

21. Case, D.A.; D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman. *AMBER 2017, University of California, San Francisco.*

22. Cheng, L. S., R. E. Amaro, D. Xu, W. W. Li, P. W. Arzberger, J. A. McCammon. Ensemble-Based Virtual Screening Reveals Potential Novel Antiviral Compounds for Avian Influenza Neuraminidase. *J. Med. Chem.* **51,** 3878–3894 (2008).

23. Lin, J.-H., N. A. Baker, J. A. McCammon. Bridging implicit and explicit solvent approaches for membrane electrostatics. *Biophys. J.* **83,** 1374–1379 (2002).

24. Schames, J. R., J.-H. Lin, A. L. Perryman, J. A. McCammon. Computational Drug Design Accommodating Receptor Flexibility: The Relaxed Complex Scheme. *J. Am. Chem. Soc.* **124,** 5632–5633 (2002).

25. Tan, Y. S., P. Sledz, S. Lang, C. J. Stubbs, D. R. Spring, C. Abell, R. B. Best. Using Ligand-Mapping Simulations to Design a Ligand Selectively Targeting a Cryptic Surface Pocket of Polo-Like Kinase 1. *Angew. Chem. Int. Ed.* **51,** 10078–10081 (2012).

26. Durrant, J. D., L. Votapka, J. Sørensen, R. E. Amaro. POVME 2.0: An Enhanced Tool for Determining Pocket Shape and Volume Characteristics. *J. Chem. Theory Comput.* **10,** 5047–5056 (2014).

27. Abraham, M. J., T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2,** 19–25 (2015).

28. Eargle, J., D. Wright, Z. Luthey-Schulten. Multiple Alignment of protein structures and sequences for VMD. *Bioinformatics* **22,** 504–506 (2006).

29. W. Humphrey, A. Dalke, K. Schulten. VMD – Visual Molecular Dynamics. *J. Mol. Graph.* **14,** 33–38 (1996).

30. Brenke, R., D. Kozakov, G.-Y. Chuang, D. Beglov, D. Hall, M. R. Langdon, C. Mattos, S. Vajda. Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics* **25,** 621–627 (2009).

31. Kozakov, D., D. Hall, G.-Y. Chuang, R. Cenci, R. Brenke, L. E. Grove, D. Beglov, J. Pelletier, A. Whitty, S. Vajda. Structural conservation of druggable hot spots in protein–protein interfaces. *Proc. Natl. Acad. Sci.* **108,** 13528–13533 (2011).

32. Cowan-Jacob, S. W. G. Fendrich, A. Floerscheimer, P. Furet, J. Liebetanz, G. Rummel, P. Rheinberger, M. Centeleghe, D. Fabbro, P. W. Manley. Structural biology contributions to the discovery of drugs to treat chronic myelogenous leukaemia. *Acta Crystallogr. D Biol. Crystallogr.* **63,** 80–93 (2006).

33. Patel, R. Y. & R. J. Doerksen. Protein Kinase−Inhibitor Database: Structural Variability of and Inhibitor Interactions with the Protein Kinase P-Loop. *J. Proteome Res.* **9,** 4433–4442 (2010).

34. Müller, S., A. Chaikuad, N. S. Gray, S. Knapp. The ins and outs of selective kinase inhibitor development. *Nat. Chem. Biol.* **11,** 818–821 (2015).