Optimizing Antibody Performance Through Molecular Dynamics Software

Matthew T. N. Kenney, under the mentorship of Ivan Gregoretti, PhD

Cell Signaling Technology, Inc., 3 Trask Lane, Danvers MA 01923

Project Goals

1. Accelerating Antibody Design:

CST's methodologies for designing highly specific and structurally sound monoclonal antibodies (mAbs) are primarily based in scientific theory and bench work trial-and-error.

SOLUTION:

- While this approach eventually produces quality results, we believe computational tools, specifically, ***molecular dynamics simulations***, could help to accelerate and complement the process of antibody design.

2. Computing for the Future:

Currently, 100% of CST's heavy lifting bioinformatics work is performed entirely on central processing units (CPUs).

SOLUTION:

- In today's world, new, parallelized computing architectures are emerging at lightning speed. These new architectures have the potential to speed up existing applications up to 10-fold or more.
- This project aims to transport heavy-lifting computing tasks to highly parallel processors. We hope to use the knowledge acquired through this project to accelerate even more CST computing work in the future to come.

3. Extreme Portability

Traditionally, provisioning a new machine (laptop, desktop, server, cloud service, etc.) with a complicated application required hard work and expertise.

SOLUTION:

- Through the use of newly emerged containerization technologies, setting up these same machines can be as easy as entering a single command. This project implements containerization technology to build a suite of applications that is incredibly simple to setup and use.

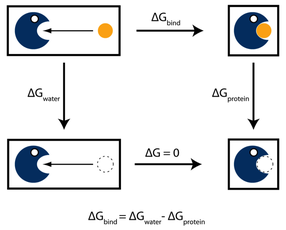
Molecular Dynamics

- ***Molecular Dynamics (MD):*** simulating molecular interactions through computer simulation (in-silico)
- "Statistical mechanics by numbers" (Maddox, 1988): Molecular force fields + Newton's Laws of Motion

NAMD Scalable Molecular Dynamics VMD

Free Energy Perturbation (FEP):

- Perturbation Theory is a 60-year-old theory of mathematics
- Used to compare the energetic difference between a **reference system**, and a **perturbed system**
- With today's computational tools, FEP can be an be employed for ***in-silico protein engineering***



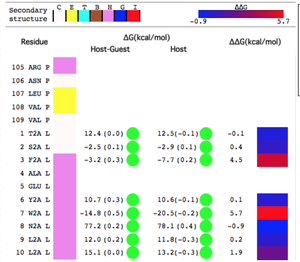
A depiction of an FEP free-energy simulation cycle. In this case, FEP calculations are carried out to determine free energy of binding, where the Pacman figure represents a the protein, and the orange "pill" represents its ligand.

In Silico Alanine Scanning:

- Alanine Scanning: Using mutation as a tool to examine the contribution of a specific residue to the ***stability*** or ***function*** of a given protein"
- Cannot measure contribution of a residue *directly*
- By replacing one residue with another, and measuring *change*, we can discern the importance of the original residue

Envisioning the top chain as the ***reference*** system, and mutated chains as ***perturbed*** systems, FEP can provide information about the energetic relevance of each replaced residue.

Determining binding free energy with Alascan 1.0:



$$\Delta\Delta G_{binding} = \Delta G_{protein+ligand} - \Delta G_{protein}$$

Above: The equation used to determine binding-free energy, given a two-step simulation

Left: The results of an in-silico alanine scanning experiment, displayed using a VMD plugin, "Alascan." Positive **ΔΔG** indicate an unfavorable mutation, thus highlighting the importance of the original residue in binding.

Parallel Computing

- While FEP is the most accurate method of alanine scanning, it is also the most computationally expensive.
- Recently, **parallel computing** has seen incredible performance gains
- Parallel computing is a game changer in the field of molecular dynamics

GPU
Parallelized processor

Optimized for small, concurrent computing tasks

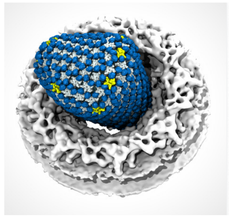
CPU:
Serial Processor

Optimized for arduous / sequential tasks

Molecular Dynamics and Parallelized Computing

- MD = enormous quantities of small, parallel molecular interactions define a system → Perfect candidate for parallel computing
- Major performance increases with GPU on both visualization and MD simulations
- Intel Xeon parallelized CPU also shows 2-3x speedup
- Observed at CST: 10-20x improvement in equilibration and minimization simulations
- ***Caveat: FEP + CUDA is still under development in NAMD software (but when it comes, research suggests up to 50X speedup).***

Parallelized MD in Action: Breakthroughs in HIV research

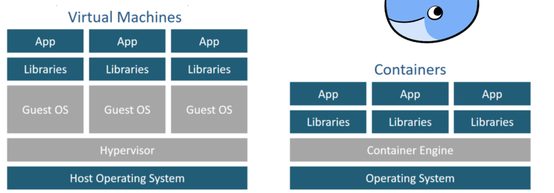


NVIDIA TESLA®

- Identification of 1 formerly unknown binding site
- 1200 Tesla GPUs = 4 petaflops
- 26 Million Atoms

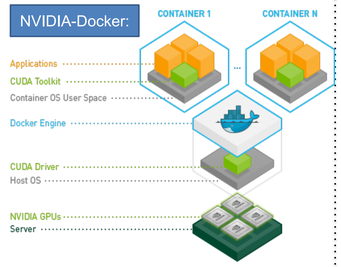
- "NVIDIA® Tesla® GPUs enable researchers to understand the interaction between the dynamic HIV structure and the human protein to help develop new therapeutic drugs" - Headline from NVIDIA corporation.
- Simulation unveiled a new binding site between human protein **CypA** and HIV capsid. This discovery has fueled the development of new HIV drugs

Why Containerization?



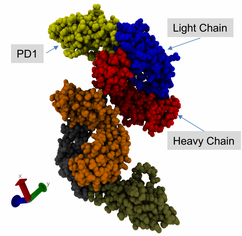
Containers:

- **Host-Agnostic:** Deployable on any machine- Server, Laptop, Window, Mac, Linux, etc.
- **Built-in** dependencies and Modifications
- **Easily deployable** on any machine. Just one command to run.
- Now **compatible with GPU**: NVIDIA (Accelerated Computing) + Docker (Containerization)



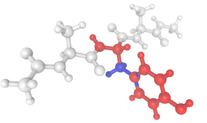
Molecular Dynamics, in Action

1. Start with a molecular structure file, such as the one to the right.
2. Minimization, Hydration, and Equilibration: to enable the system to fall into a natural state:
   - CUDA Accelerated
   - Biological or room temperature
3. On the binding site, select a set of residues to examine
4. Mutate CDR3 residues to alanine
5. Prepare file sets for FEP calculation



Above: PDB File 5GGR: Commercial Monoclonal Antibody, Opdivo, or Nivolumab, in complex with **PD1**. Image taken using high-resolution rendering from Visual Molecular Dynamics (VMD).

λ=0 → λ=1



Left: An example mutated protein structure file, where the residue to be mutated (referential state) is outlined in red, and mutant alanine residue is outlined in blue. To determine **ΔG**, the simulation gradually tunes out interactions from the red residue, and tunes up interactions from the blue. All the while, the system is constantly measuring the free energy changes taking place. Image taken using VMD.

6. SSH to Intel Xeon-based computing server at CST

7. Pull NAMD container with a single command, and server is ready to run. Now, run FEP simulations.
8. Analyze the results, and advise antibody engineers accordingly

Moving Forward

Successes:

- Built a GPU workstation from scratch
- Created a set of containerized molecular dynamics applications to be easily deployed on desktops, servers, and cloud services
- Avoided the luxury tax of MD
- Designed a workflow for FEP calculations, to help aid our researchers in antibody design
- Accelerated molecular dynamics through GPU.
- Bug fixes to enable analysis of FEP

Work-in-Progress:

- Bug fixing Alascan 1.0 Plugin (Automate steps 3-5 above)
- Xeon Optimization
- Streamlining methodology for generating CST-relevant protein structure files
- Uncovering new use-cases for MD
- Continuing to accelerate & containerize CST's bioinformatics capabilities

Cell Signaling TECHNOLOGY

Matthew T. N. Kenney
Cell Signaling Technology, Inc.
matthew.kenney@cellsignal.com