SI^2-SSE: High-Performance Software for Large-Scale Modeling of Binding Equilibria

PI: Emilio Gallicchio; **SI:** Baofeng Zhang, Daniele Di Marino; **Undergrads:** Holly Tancredi, Tony Zhao

Brooklyn College, The City University of New York

egallicchio@brooklyn.cuny.edu
http://sites.google.com/site/emiliogallicchiolab

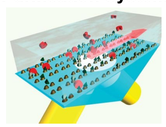
Drug-Protein Interactions, Molecular Recognition, Self-Assembly

Molecular binding is at the basis of many biological and chemical processes

Drug Design & Environmental Toxins

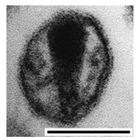
The complex between the Imatinib (red), used to treat leukemia and its target, the BCR-ABL kinase protein (green)

Chemical Sensors & Purification Systems



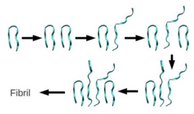
A proposed single molecule biosensor based on recognition receptors and plasmon resonance. *Kravets et. al Nature Materials (2013)*

Self Assembly



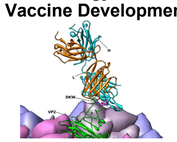
A handful of protein types and RNA self-assemble to create an infectious HIV virus. *Wes Sundquist, U. Utah*

Aggregation



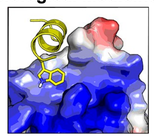
Protein aggregation into amyloid fibers is responsible for neurological disorders such as Alzheimer and Parkinson. *Luca Larini and Joan-Emma Shea, Biophys J. (2012)*

Immunology & Vaccine Development



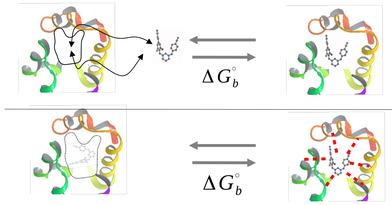
The 2F5 antibody (brown & blue) bound to a HRV-HIV chimeric virus proposed as a potential HIV vaccine. *Lapelosa, Gallicchio et al JMB (2009, 2010)*

Biological Medicines



A peptide inhibitor (yellow) bound to the HIV integrase protein (blue & red). Biological macromolecules are increasingly being considered as therapeutic drugs. Cavalluzzo et al. J. Pept. Science (2013).

Alchemical Molecular Simulation Methods to Study Binding Equilibria

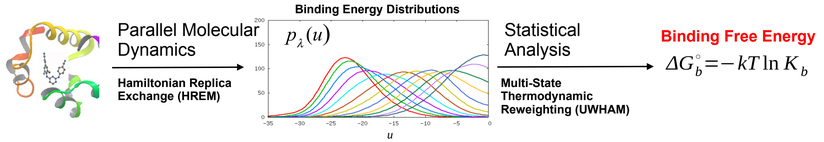


Physical Space
(real world: ligand moves in and out of receptor)

Alchemical Space
(made-up world: ligand is grown into receptor)

Binding Energy Distribution Analysis Method (BEDAM)

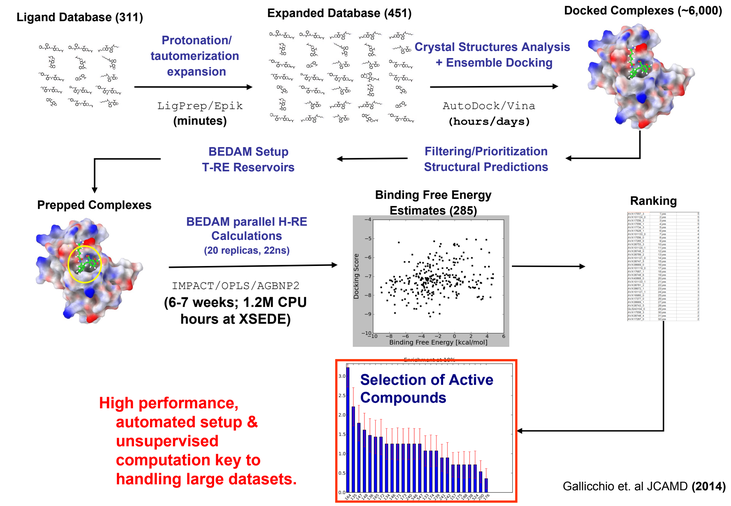
Hybrid Potential Function (introduces interaction gradually): $U_\lambda(x)=U_0(x)+\lambda u(x)$



Gallicchio, Lapelosa, Levy, *JCTC* **(2010)** · Gallicchio & Levy, *Curr. Op. Struct. Biol.* **(2011)** · Gallicchio & Levy, *Adv. Prot. Chem.* **(2011)** · Lapelosa, Gallicchio, Levy, *JCTC* **(2012)** · Gallicchio, Levy *J. Comp. Aid. Mol. Design* **(2012)** · Tan, Gallicchio, Lapelosa, Levy *JCP* **(2012)** · Gallicchio, *Mol. Biosc* **(2012)** · Wickstrom, He, Gallicchio, Levy *JCTC* **(2013)** · Gallicchio et. al JCAMD **(2104)** · Gallicchio et al. *JCAMD* **(2014b)** · Deng, Gallicchio et al. *JPCB* **(2014)**

Large Scale Deployments: Drug Screening

Accurate but limited by software performance



High performance, automated setup & unsupervised computation key to handling large datasets.

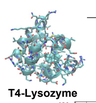
Gallicchio et. al JCAMD **(2014)**

Aim I: Optimization, Vectorization and Parallelization of BEDAM

Implicit solvation speeds up computation of effective binding energy by many orders of magnitude

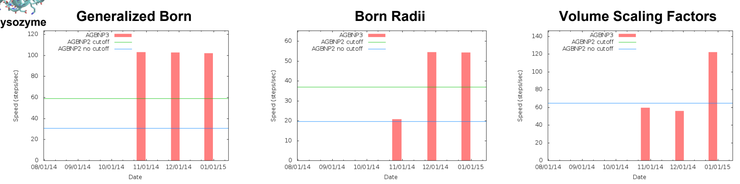
AGBNP2 Implicit Solvation Model

	Electrostatic Model	Non-Polar Model	First-Shell Model
	Volume Scaling Factors	Surface Areas	Hydration Sites Weights
1	$s_i=\frac{V_i^{\rm self}}{V_i}=1-\frac{1}{2}\sum_j\frac{V_{ij}}{V_i}+\frac{1}{3}\sum_{j<k}\frac{V_{ijk}}{V_i}+\ldots$	$A_i=\frac{\partial V}{\partial R_i}$	$w_\alpha=\frac{V_\alpha^{\rm free}}{V_\alpha}=1-\sum_i\frac{V_{\alpha i}}{V_\alpha}+\sum_{i<j}\frac{V_{\alpha ij}}{V_\alpha}-\ldots$
	Born Radii		
2	$\frac{1}{B_i}=\frac{1}{R_i}-\frac{1}{4\pi}\sum_{j\neq i}s_{ji}Q_{ji}$		
	Generalized Born Energy	Non-Polar Energy	First Shell HB energy
3	$\Delta G_{\rm el}=-\frac{1}{2}u_\epsilon\left[\sum_i\frac{q_i^2}{B_i}+2\sum_{i<j}\frac{q_iq_j}{f_{ij}}\right]$	$\Delta G_{\rm np}=\sum_i[\gamma_iA_i+\alpha_iW(B_i)]$	$\Delta G_{\rm HB}=\sum_\alpha h_\alpha S(w_\alpha)$



T4-Lysozyme

Recent Progress in Optimization & Vectorization



Targeting 40x speed-up 16-core CPU systems

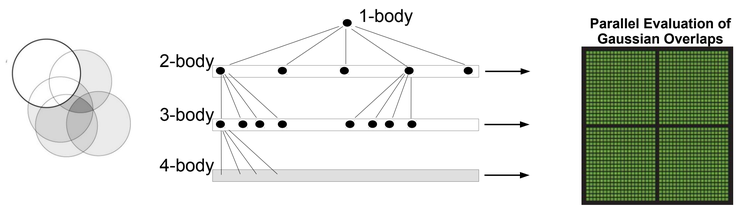
Aim II: Optimized MIC & GPGPU Implementations

- The Intel MIC (Many Integrated Cores) co-processor can provide significant added performance

Combined Throughput Analysis with Current Implementation on Stampede cluster

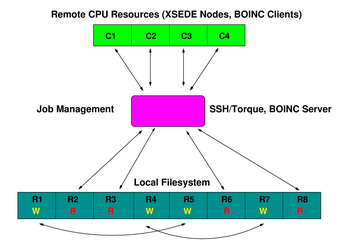
Device	speed[a]	#replicas/device	#devices[b]	throughput[c]	%throughput
CPU (4 threads)	27.1	4	6	84.3 ns/day	64.7%
MIC (24 threads)	5.9	10	6	45.9 ns/day	35.3%
All				130.2 ns/day	100%

- GB-based implicit solvent models generally perform well on GPU hardware, expecting ~100x speed-up
- However Gaussian-based volume scaling factor algorithm of AGBNP2 has never been ported to GPU hardware
- Currently experimenting with breadth-first tree traversal strategies



Aim III: Integration into OpenMM & Interfaces for Molecular Dynamics Engines

- Why OpenMM:
  - A library framework with well-documented API's
  - History of integration into MD engines
  - Multi-architecture design (CPU, CUDA, OpenCL)
  - Ports for the main OS's (Linux, Mac, Windows)
  - GPL licensing
- Initial targets: OpenMM Python API and academic IMPACT
- Asynchronous Replica Exchange for Grid and Heterogeneous computing



- Completed implementation of OPLS2005 force field in OpenMM