Computing the Bacterial Brain:
Molecular dynamics simulations of the bacterial chemosensory array

C. Keith Cassidy
Theoretical and Computational Biophysics Group
University of Illinois at Urbana-Champaign

PRAC: The Computational Microscope
PIs: Klaus Schulten & Emad Tajkhorshid

Blue Waters Symposium
Sunriver, OR
17 May 2017
Outline:

• Background
• Why It Matters
• Key Challenges
• Work on Blue Waters
• Future Directions
Background: Bacterial Chemotaxis

**Bacteria** monitor a wide range of environmental chemical cues and use this information to determine their motile behavior.

Through the **integration** of many complex (and often conflicting) signals the cell is able to efficiently locate **optimal** growing conditions.

*E. coli* uses a **run-tumble** swimming strategy... lengthening runs in “good” directions!
To Tumble, or Not To Tumble

Free-swimming \textit{E. coli} cells

Howard Berg Lab (Harvard U.)

Single, optically immobilized cell

Yann Chemla Lab (UIUC)

Mears et al., (2014), eLife
**The Bacterial Brain**

Large arrays ($10^4 - 10^5$ nm$^2$, $>10^4$ proteins) of **chemoreceptors**, **histidine kinases (CheA)**, and **coupling proteins (CheW)** integrate complex environmental signals.

**E. coli chemotactic network.**
An expanded set of molecules assists in signal regulation and coupling to flagellar motors.

The network is sophisticated!

(I) **Ultra-Sensitivity** - Gradient detection of a ~10 molecule change per cell volume.

(II) **High Gain** - Cells can amplify stimuli over 50-fold.

(III) **Precise adaptation** - Extends the range of concentrations that can be discriminated to five orders of magnitude.
From a molecular perspective this represents a new frontier towards a basic understanding of biological information processing.

• Centerpiece of the most thoroughly studied sensory signal transduction system in biology.

Understanding of molecular underpinnings could enable mass reprogramming and the development of novel antibiotics and fungicides.

• Structural organization is universally conserved.

Transferability of functional mechanisms between distantly related species.

• Critical sensory proteins do not exist in mammalian organisms.

Electron microscopy shows 12-nm hexagonal lattice is universally conserved in chemotactic bacteria and archaea.

Briegel, A. et al., (2009), PNAS.
Briegel, A. et al., (2015), Environmental Microbiology Reports.
Key Challenges

**Goal:** To construct a high fidelity, fully atomistic model of the chemosensory array for the computational investigation of sensory signal transduction.

1. Hi-res structural information is limited…
   - **X-ray structures** of sensory proteins and a few complexes only.
   - Incomplete, not in “array-bound” conformation, local
   - **Electron microscopy densities** of extended complex
   - Non-local but resolution too low (~25 nm) to unambiguously assign to particular proteins and domains.

   **Solution:** Multi-scale modeling techniques

2. The array is necessarily large…
   - Computational ability emerges from the **collective interactions** of many parts.
   - Experiments point to minute structural and dynamical changes in single proteins, requiring **all-atom detail**.

   **Solution:** NAMD on Blue Waters

---

**Chemosensory Array**
- 10 - 100 million atoms
- Runs on Blue Waters (13 PF)

**Ribosome**
- ~3 million atoms
- Ran on Kraken (1.3 PF) 2009

---

Pole of *E. coli* cell
Zhang, P. et al., (2007), PNAS.
NAMD on Blue Waters

Systems over a wide range of length scales perform very well on both XE and XK Blue Waters architectures!

Supports:
- Wide range of user-defined forces (colvars, TCLForces, etc.)
- Enhanced Sampling (REMD, GaMD, String Method, etc.)
- Free Energy (FEP, ABF, etc.)
- QM/MM (interfaces to ORCA and MOPAC) **(NEW!)**
- One-click launch in Amazon Marketplace **(NEW!)**
- Latest release: v2.13 (download: http://www.ks.uiuc.edu/Research/namd)

20 Million atom Array:
~36 ns/day on 2046 XK nodes
Multi-scale Modeling of the Chemosensory Array

1. Begin with high-res structures of portions of *T. maritima* signaling proteins.

2. Computationally model core components of array using existing structural information.

---

**X-ray Structures**

- Tsr trimer-of-dimers (PDB 1QU7)
- Receptor dimer (PDB 2CH7)
- CheA-P34 dimer (PDB 1B3Q)
- CheA-P5/CheW ring (PDB 4JPB)
- CheW monomer (PDB 4JPB)

**12 nm**

- Chemoreceptor trimer-of-dimers (cytoplasmic)
- CheW-only ring

**Coupled CheA/CheW rings**
In vitro Characterization of E. coli Chemosensory Array

Zhang et al. devised reconstitution of purified His-tagged TarCF, CheA, CheW on a Ni$^{2+}$-NTA lipid monolayer

The monolayer is ideal for cryo-ET: thin, crystalline and well defined

Sub-tomogram classification and averaging used to obtain highly-resolved density maps of extended array structure.

MDFF: Computational synthesis of multi-scale structural data

MDFF simulation drives conformations of high resolution structures towards those seen in the intermediate resolution cryoET maps

Rigid docking of core component models produces pseudo-atomic CheA-trimer model

Atomic model reproduces, refines, and identifies novel interactions at key protein-protein interfaces
T. maritima Core Signaling Unit: PDB 3JA6

Cassidy et al. eLife (2015)
All-Atom MD Simulations of Array Unit Cell

<table>
<thead>
<tr>
<th>System</th>
<th>Simulation</th>
<th>Atom #</th>
<th>Length (ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>unit cell</td>
<td>Equilibration</td>
<td>1.25 M</td>
<td>80</td>
</tr>
<tr>
<td>unit cell</td>
<td>Production</td>
<td>1.25 M</td>
<td>9 * 500</td>
</tr>
</tbody>
</table>
Conformational Change in CheA Kinase

RMSD-based K-medoids clustering used to isolated classes of CheA conformation

CheA-P4 “dipping”

“Undipped” CheA dimer (80%)

“Dipped” CheA dimer (12%)

Pick pairs of residues unique to each class, mutate to disrupt salt bridges or stabilize state via cross-linking.
Biochemical Validation in Live *E. coli* Cells.

In vivo swim assay and cross-linking suggest mobility of P3 is key for signaling and validate model.

**kinase-receptor**

D333/K390  
(I304-N405 in *E. coli*)

D345/R79  
(D316-R394 in *E. coli*)

**kinase-kinase**

E390/R379  
(E361/R394 in *E. coli*)

R297/E397  
(R265/E368 in *E. coli*)

MD predicts multiple contacts stabilizing “dipped” dimer state.

**Cassidy et al. eLife (2015)**

Frances Alverez, Zhang Lab

How does this motion affect CheA activity?

How might receptors regulate the relative populations of the two states?
Modeling Transmembrane Serine Receptor

**X-ray structure** of Tsr ligand-binding domain (PDB: 2D4U)

Embed in 3:1 POPE:POPG lipid bilayer (105K atoms) and equilibrate for 3 microseconds

**X-ray structure** of homologous archaeal HAMP (PDB: 2L7H)
*Ferris et al., (2011) Structure*

Cross-linking efficiencies between Tsr TM bundle. *Pakula et al. (1992) PNAS.*

Model reproduces cross-linking distances and **lipid/protein interactions** as well as sheds light on function of **key residues for signaling** (e.g., I214)

Tsr homodimer (200K atoms)
Summary

• Cryo-electron tomography (cryo-ET) used to determine the 3D structure of the *E. coli* chemosensory array in high detail.

• Based on our cryo-ET data and existing crystal structures, Blue Waters was used to computationally construct the first atomic models of an intact chemoreceptor and transmembrane chemosensory array.

• Extended all-atom simulations on Blue Waters revealed a novel conformational change in a key signaling protein that was confirmed to be important for chemotaxis by experiments on mutant *E. coli* cells.

*E. coli* core signaling unit (3 Million atoms)
Acknowledgments

Prof. Emad Tajkhorshid
Department of Biochemistry
TCBG

Prof. Zan Schulten
Department of Chemistry
CPLC

Prof. Peijun Zhang
Department of Structural Biology

Prof. Sandy Parkinson
Department of Biology

Theoretical and Computational Biophysics Group
Beckman Institute, UIUC

Benjamin Himes
Dr. Frances Alvarez

Prof. Klaus Schulten

NSF
NATIONAL INSTITUTES OF HEALTH
Center for the Physics of Living Cells
BLUE WATERS
SUSTAINED PETASCALE COMPUTING
NCSA

University of Pittsburgh Medical Center
University of Utah