

# Prediction of Functional Sites in Predicted Protein Structures using Dynamics Perturbation Analysis

Judith D. Cohn<sup>1</sup> and Michael D. Wall<sup>1,2,3</sup>

<sup>1</sup>Computer, Computational, and Statistical Sciences Division <sup>2</sup>Bioscience Division <sup>3</sup>Center for Non-Linear Studies Los Alamos National Laboratory

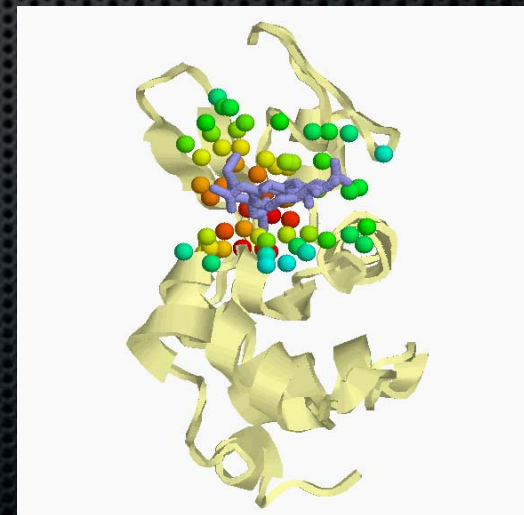
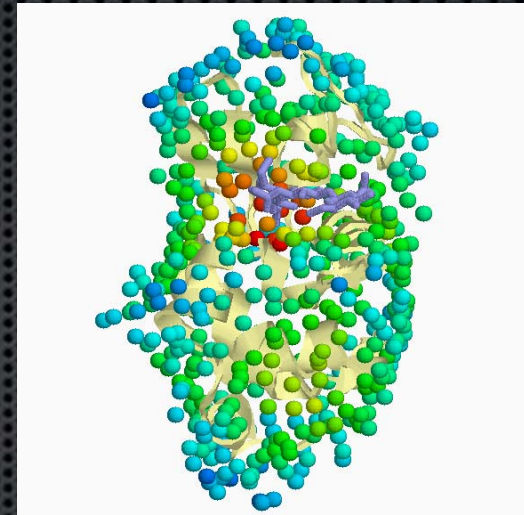
6th Annual Rocky Mountain Bioinformatics Conference (Dec 3-7, 2008)

# Dynamics Perturbation Analysis (DPA)

- **Concept: Determine where a protein is ticklish**
- Based on structure alone
- Decorate protein surface with set of surface test points M
- Calculate change in protein vibrations for each test point

$$D_{\mathbf{x}}^{(m)} = \frac{1}{2} \sum_{i=1 \dots 3N}^{\bar{\lambda}_i^{(m)} \neq 0; \lambda_i \neq 0} \left( \log \frac{\bar{\lambda}_i^{(m)}}{\lambda_i} + \sum_{j=1 \dots 3N}^{\lambda_j \neq 0} \frac{\lambda_j}{\bar{\lambda}_i^{(m)}} |\bar{\mathbf{v}}_i^{(m)} \cdot \mathbf{v}_j|^2 - 1 \right)$$

- Cluster points with with largest change (cluster = functional site)
- Identify protein residues near clusters



Lysozyme (1JEF)

# DPA: Crystal Structures

- Proof of Concept
  - 305 structures from GOLD docking set
    - D Ming & ME Wall (2006). Interactions in native binding sites cause a large change in protein dynamics. *J Mol Biol***358**:213
- High-throughput analysis
  - Fast DPA
    - D Ming, JD Cohn & ME Wall (2008). Fast dynamics perturbation analysis for prediction of functional site. *BMC Structural Biol***8**:5
  - ~50k SCOP domains
    - JD Cohn, D Ming & ME Wall (2008). Prediction of functional sites in SCOP domains using dynamics perturbation analysis. AFP-Biosapiens 2008, Toronto, Canada. Extended abstract available at <http://precedings.nature.com/documents/2209/version/1>

# DPA: Predicted Structures

- Test set: 128 targets x 5 models per target
  - targets from 8th Critical Assessment of Techniques for Protein Structure Prediction (CASP8)
  - models from automated server (Baker lab, University of Washington, Seattle)
- Entered team in CASP8 human category for prediction of binding sites (limited function prediction)
- Method
  - Identify predicted residues for “best” model
  - Assign quality score for each predicted residue = number of models which concur

# CASP Target T0461 (PDB 3DH1)

human tRNA-specific adenosine-34 deaminase subunit ADAT2

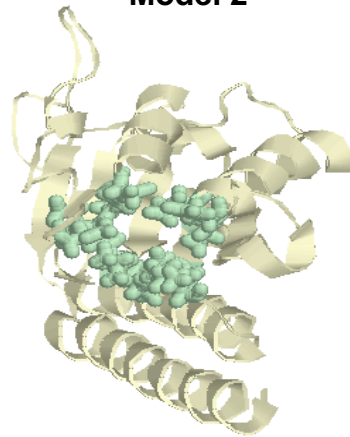
RES NUM	RES NAME	QUAL NUM	STATUS
42	VAL	5	pdb + target
43	PRO	0	pdb only
44	VAL	5	pdb + target
45	GLY	5	target only
60	ASN	4	pdb + target
71	HIS	5	pdb + target
72	ALA	5	pdb + target
73	GLU	5	pdb + target
102	VAL	0	pdb only
103	THR	5	pdb + target
104	VAL	5	pdb + target
105	GLU	5	pdb + target
106	PRO	5	pdb + target
107	CYS	5	pdb + target
110	CYS	2	pdb + target
127	GLN	0	pdb only
128	ASN	5	pdb + target
132	GLY	0	pdb only
167	LEU	0	pdb only

# CASP Target T0461 (PDB 3DH1)

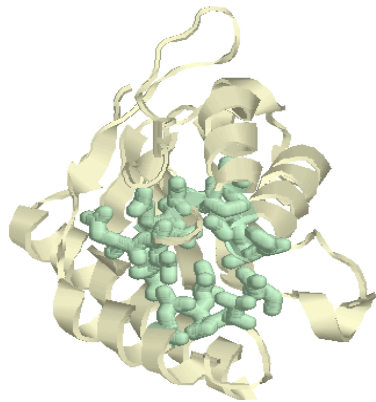
Model1



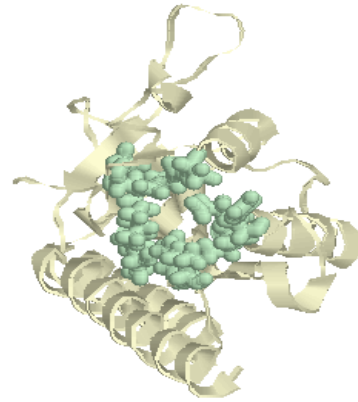
Model 2



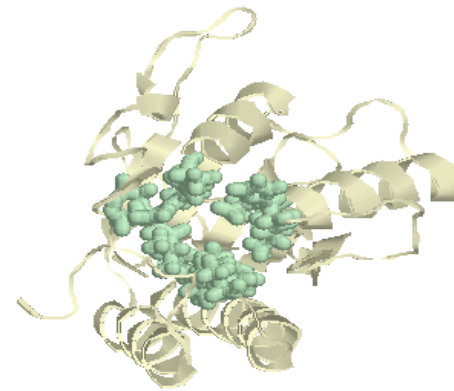
Model3



3DH1chain A



Model4



Model5

DPA can yield predictions which are consistent between experimental and predicted structures.