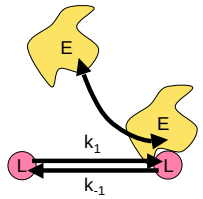
Modeling Allosteric Effects in Proteins

Michael Wall
Computer, Computational, and Statistical Sciences Division
Bioscience Division
Center for Nonlinear Studies
Los Alamos National Laboratory
http://public.lanl.gov/mewall
mewall@lanl.gov

First q-bio Summer School
Los Alamos, NM
31 July 2007

...everything that living things do can be understood in terms of the jigglings and wigglings of atoms.

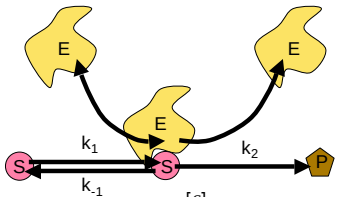
Protein Interactions



$$K_d = \frac{[E][L]}{[E \cdot L]} = \frac{k_{-1}}{k_1}$$

$$f = \frac{[E \cdot L]}{[E] + [E \cdot L]} = \frac{[L]}{[L] + K_d}$$

Enzyme Activity

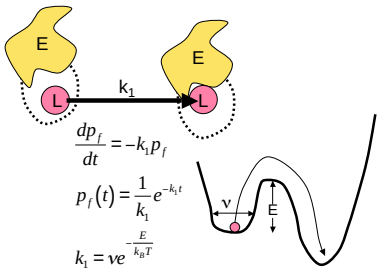


$$v = k_2 [E]_T \frac{[S]}{[S] + K_M}$$

$$[E]_T = [E] + [E \cdot S]$$

$$K_M = \frac{k_{-1} + k_2}{k_1}$$

Stochastic Ligand Binding



$$\frac{dp_f}{dt} = -k_1 p_f$$

$$p_f(t) = \frac{1}{k_1} e^{-k_1 t}$$

$$k_1 = \nu e^{-\frac{E}{k_B T}}$$

Dynamics of Ligand Binding to Myoglobin†

R. H. Austin, K. W. Beeson, L. Eisenstein, H. Frauenfelder,* and I. C. Gunsalus

BIOCHEMISTRY, VOL. 14, NO. 24, 1975 5355

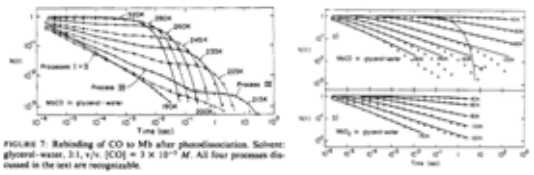
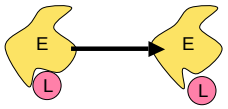


FIGURE 7: Rebinding of CO to Mb after photodissociation. Solvent: glycerol–water, 3:1, v/v. [CO] = 3 × 10⁻⁹ M. All four processes discussed in the text are recognizable.

Protein Fluctuations and Ligand Binding



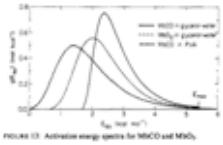
$$k \rightarrow g_k(k)$$

$$N_f(t) = \int_0^\infty dk\ g_k(k) e^{-kt}$$

Dynamics of Ligand Binding to Myoglobin†

R. H. Austin, K. W. Beeson, L. Eisenstein, H. Frauenfelder,* and I. C. Gunsalus

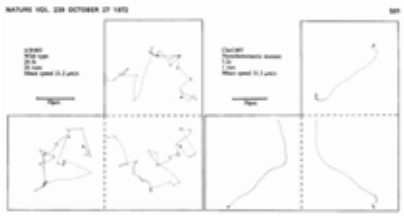
BIOCHEMISTRY, VOL. 14, NO. 24, 1975 5355



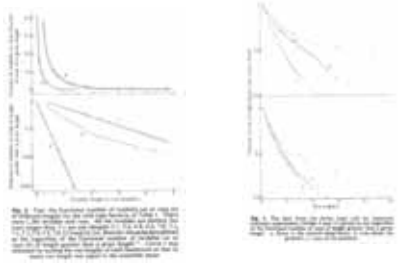
$$k_{ba} = \nu e^{-\frac{E_{ba}}{k_B T}}$$

Chemotaxis in *Escherichia coli* analysed by Three-dimensional Tracking

HOWARD C. BERG & DOUGLAS A. BROWN

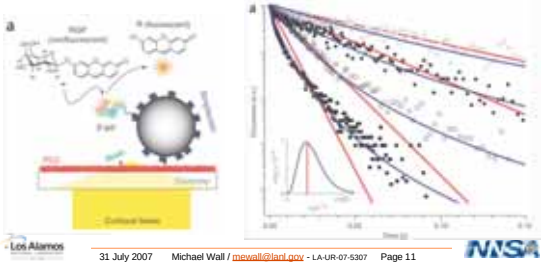


Run Length Distributions

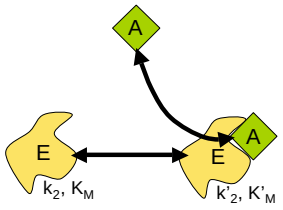


Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited

Brian P English[1], Wei Min[1], Antoine M van Oijen[1,4], Kang Taek Lee[1,4], Guobin Luo[1], Hongye Sun[1,4], Binny J Cherayil[1,2], S C Kou[1] & X Sunney Xie[1]



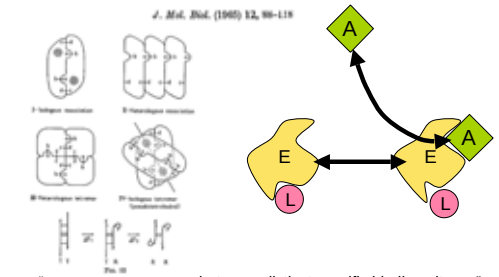
Regulation of Protein Activity



On the Nature of Allosteric Transitions: A Plausible Model

Jacques Monod, Jeffries Wyman and Jean-Pierre Changeux

J. Mol. Biol. (1965) 12, 88–118



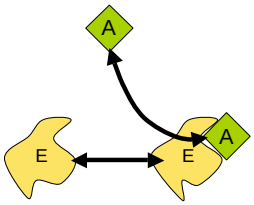
"…*indirect* interactions between *distinct* specific binding-sites…"

A protein is a "kicking and screaming 'stochastic' molecule"

Ligand Binding and Internal Equilibria in Proteins†

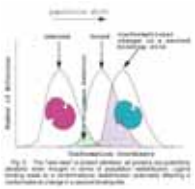
Gregorio Weber*

BIOCHEMISTRY, VOL. 11, NO. 5, 1972



P(x) P(k)

P'(x) P'(k)



Gunasekaran, Ma, & Nussinov 2004

ΔP = ?

Change in energy landscape upon ligand binding

Ligand-Induced Bias

$$\frac{[P_0 \cdot L]}{[P_0][L]} = k_{a0}$$

$$\frac{[P_i \cdot L]}{[P_i][L]} = k_{ai}$$

$$\frac{[P_i]}{[P_0]} = e^{-\Delta G_i^{(0)}/k_B T}$$

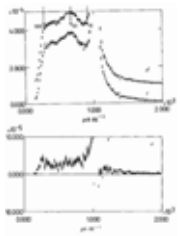
$$\frac{[P_i] + [P_i \cdot L]}{[P_0] + [P_0 \cdot L]} = e^{-\Delta G_i^{(0)}/k_B T} \frac{1 + k_{ai}[L]}{1 + k_{a0}[L]}$$

Freire. Adv Prot Chem 1998

Inelastic neutron scattering analysis of hexokinase dynamics and its modification on binding of glucose

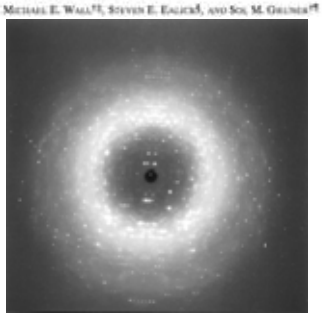
Bernard Jacrot*, Stephen Cusack*, Albert J. Dianoux† & Donald M. Engelman‡

Nature Vol. 300 4 November 1982



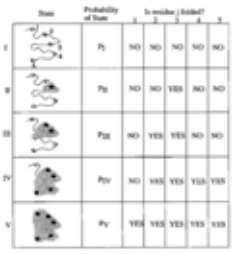
Three-dimensional diffuse x-ray scattering from crystals of *Staphylococcal* nuclease

Michael E. Wall*‡, Steven E. Ealick§, and Sol M. Gruner*¶

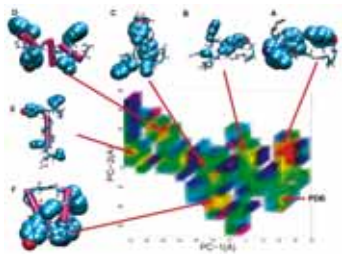


Structure-based Calculation of the Equilibrium Folding Pathway of Proteins. Correlation with Hydrogen Exchange Protection Factors

Vincent J. Hilser and Ernesto Freire*



Exploring the Energy Landscape of a β Hairpin in Explicit Solvent

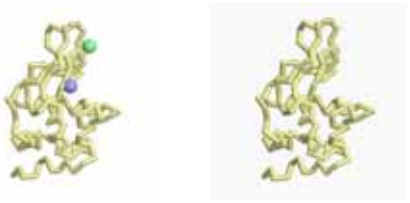


Quantifying Allosteric Effects in Proteins

$$D_{\mathbf{x}} = \int d^{3N}\mathbf{x}\; P'(\mathbf{x}) \ln \frac{P'(\mathbf{x})}{P(\mathbf{x})}$$

$P'(\mathbf{x})$: With ligand; $P(\mathbf{x})$: Without ligand

(Relative Entropy / KL Divergence)



Calculation of D_x for a Normal Distribution

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

$$P'(x) = \frac{1}{\sigma'\sqrt{2\pi}} e^{-(x-\mu')^2/2\sigma'^2}$$

$$D_{KL} = \int dx\; P'(x)\left[\frac{(x-\mu)^2}{2\sigma^2} - \frac{(x-\mu')^2}{2\sigma'^2} + \ln\frac{\sigma}{\sigma'}\right]$$

$$= \ln\frac{\sigma}{\sigma'} + \frac{1}{2\sigma^2}(\mu'-\mu)^2 + \frac{1}{2}\left(\frac{\sigma'^2}{\sigma^2} - 1\right)$$

Calculation of D_x for Normal Modes

$$\mathbf{x} = \mathbf{x}_0 + \sum_{i=1}^{3N} a_i(\mathbf{x})\mathbf{v}_i = \mathbf{x}'_0 + \sum_{i=1}^{3N} a'_i(\mathbf{x})\mathbf{v}'_i$$

$$U(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{3N} \omega_i^2 |a_i(\mathbf{x})|^2$$

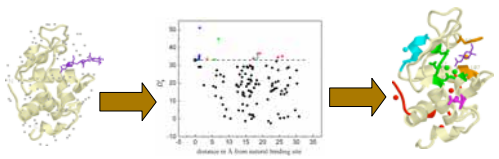
$$Z = \int d^{3N}\mathbf{a}\, e^{\frac{-1}{2k_BT}\sum_{i=1}^{3N}\omega_i^2|a_i|^2} = (2\pi k_B T)^{3N/2}\prod_{i=1}^{3N}\omega_i^{-1}$$

$$D_{\mathbf{x}} = \sum_{i=1}^{3N}\left(\log\frac{\omega'_i}{\omega_i} + \frac{1}{2k_BT}\omega_i^2|\Delta\mathbf{x}_0\cdot\mathbf{v}_i|^2 + \frac{1}{2}\sum_{j=1}^{3N}\frac{\omega_i^2}{\omega_j'^2}|\mathbf{v}'_i\cdot\mathbf{v}_j|^2 - \frac{1}{2}\right)$$

$D_{\mathbf{x}}^{\omega}$ $D_{\mathbf{x}}^{\Delta\mathbf{x}}$ $D_{\mathbf{x}}^{\mathbf{v}}$

Analysis of Allosteric Effects in Lysozyme

- Place tri-NAG ligand at various points on the surface of lysozyme
- Evaluate Dx at each point



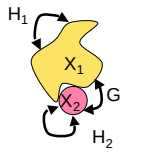
D_x is relatively large in functional sites of lysozyme

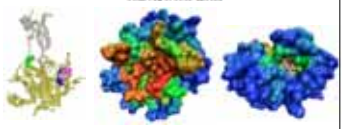
Allostery in a Coarse-Grained Model of Protein Dynamics

Dongming Ming

Michael E. Wall

PRL 95, 198103 (2005)





Allostery in BPTI-trypsinogen

$$\mathbf{Hx} = \begin{pmatrix} \mathbf{H}_1 & \mathbf{G} \\ \mathbf{G}^T & \mathbf{H}_2 \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{H}_1\mathbf{x}_1 + \mathbf{G}\mathbf{x}_1 \\ \mathbf{G}^T\mathbf{x}_2 + \mathbf{H}_2\mathbf{x}_2 \end{pmatrix}$$

$$\overline{\mathbf{H}} = \mathbf{H}_1 - \mathbf{G}\mathbf{H}_2^{-1}\mathbf{G}^T = \overline{\mathbf{V}}|\Omega|^2\overline{\mathbf{V}}^T$$

(Zheng & Brooks. Biophys J 2005)

Allostery in a nicotinic acetylcholine receptor

et al. & Karplus, Changeux. PNAS 2006

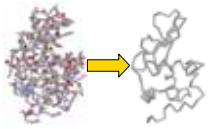
The propagation of binding interactions to remote sites in proteins: Analysis of the binding of the monoclonal antibody D1.3 to lysozyme

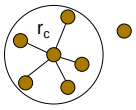
Ernesto Freire*

Proc. Natl. Acad. Sci. USA Vol. 96, pp. 10118–10122, August 1999 Biophysics

$$\kappa_{f,j} = \frac{\sum_i P_{f,j,i}}{\sum_i P_{nf,j,i}}$$

Conditions for Optimal Coarse-Grained Models





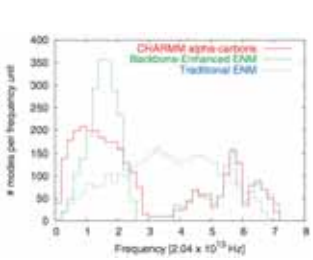
Tirion. PRL 1996

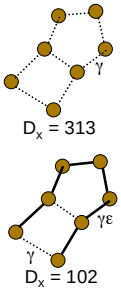
Atilgan et al. Biophys J 2001

$$\overline{D}_x = \sum_{i=1}^{3N_p'} \left(\log\frac{\overline{\omega}_i'}{\overline{\omega}_i} + \frac{1}{2k_BT}\overline{\omega}_i^2 \left|\overline{\mathbf{v}}_i^T \Delta\mathbf{x}_{0,P}\right|^2 + \frac{1}{2}\sum_{j=1}^{3N_p} \frac{\overline{\omega}_j^2}{\overline{\omega}_i'^2} \left|\overline{\mathbf{v}}_i'^T \overline{\mathbf{v}}_j\right|^2 - \frac{1}{2} \right)$$

$$\gamma = \frac{1}{3N_m} \sum_{i=1}^{3N_m} \sum_{j=1}^{3N_m} \frac{\bar{a}^2}{a_i^2} \left|\mathbf{v}_i^{(F)T} \mathbf{v}_j\right|^2$$

Improvement of a Coarse-Grained Model





Thermodynamic Interpretation of D_x

- Why is relative entropy (KL divergence) interesting in this case?
- It has become fashionable in other studies but is often applied in an *ad hoc* way
- In our case, D_x has thermodynamic meaning as an *allosteric free energy*. (Qian 2001)

$$D_{\mathbf{x}} = \frac{1}{k_BT}\left\{G_T' - G_T - \int dx\ P'(x)\left[G'(x) - G(x)\right]\right\}$$
$$= \frac{\Delta G_T}{k_BT} - \int dx\ P'(x)\ln K_d(x)$$

Free energy required to change the conformational distribution of the protein from a ligand-free to a ligand-bound state

ME Wall. AIP Conf Proc 2006

"Proof" of Thermodynamic Interpretation

$$P(x) = \frac{e^{-G(x)/k_BT}}{Z}$$
$$G(x) = E(x) - TS(x) + PV(x)$$
$$D_{KL} = \frac{1}{k_BT}\int dx\ P'(x)\left[-G'(x) + G(x) + k_BT\ln\frac{Z}{Z'}\right]$$
$$k_BT\ln Z = TS_{conf} - \langle G(x)\rangle_x = -G_T$$

$$D_{KL} = \frac{1}{k_BT}\left\{G_T' - G_T - \int dx\ P'(x)\left[G'(x) - G(x)\right]\right\}$$
$$= \frac{\Delta G_T}{k_BT} - \int dx\ P'(x)\ln K_d(x)$$

Normal Distributions

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

$$P'(x) = \frac{1}{\sigma'\sqrt{2\pi}} e^{-(x-\mu')^2/2\sigma'^2}$$

$$D_{KL} = \int dx\ P'(x)\left[\frac{(x-\mu)^2}{2\sigma^2} - \frac{(x-\mu')^2}{2\sigma'^2} + \ln\frac{\sigma}{\sigma'}\right]$$
$$= \ln\frac{\sigma}{\sigma'} + \frac{1}{2\sigma^2}(\mu'-\mu)^2 + \frac{1}{2}\left(\frac{\sigma'^2}{\sigma^2} - 1\right)$$

Normal Modes

$$\mathbf{x} = \mathbf{x}_0 + \sum_{i=1}^{3N} a_i(\mathbf{x})\mathbf{v}_i = \mathbf{x}'_0 + \sum_{i=1}^{3N} a'_i(\mathbf{x})\mathbf{v}'_i$$

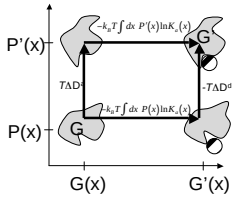
$$U(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{3N} \omega_i^2 |a_i(\mathbf{x})|^2$$

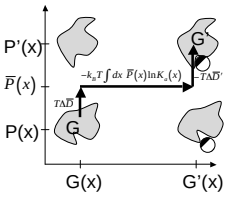
$$Z = \int d^{3N}\mathbf{a} e^{\frac{-1}{2k_BT}\sum_{i=1}^{3N}\omega_i^2|\mathbf{a}_i|^2} = (2\pi k_B T)^{3N/2}\prod_{i=1}^{3N}\omega_i^{-1}$$

$$D_x = \sum_{i=1}^{3N}\left(\log\frac{\omega'_i}{\omega_i} + \frac{1}{2k_BT}\omega_i^2|\Delta\mathbf{x}_0\cdot\mathbf{v}_i|^2 + \frac{1}{2}\sum_{j=1}^{3N}\frac{\omega_j^2}{\omega_i'^2}|\mathbf{v}'_i\cdot\mathbf{v}_j|^2 - \frac{1}{2}\right)$$

Ming & Wall. Proteins 2005

Thermodynamics of Ligand Binding





$$\Delta G = T\Delta D^{\circ} - k_B T\int dx\ P'(x)\ln K_x(x)$$
$$= -T\Delta S_x + \int dx\left[P'(x) - P(x)\right]G(x) - k_B T\int dx\ P'(x)\ln K_x(x)$$

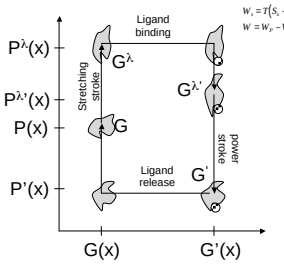
$$\Delta G = -T\Delta D^{\circ} - k_B T\int dx\ P(x)\ln K_x(x)$$
$$= -T\Delta S_x + \int dx\left[P'(x) - P(x)\right]G'(x) - k_B T\int dx\ P(x)\ln K_x(x)$$

$$T\Delta\bar{D} = -T\left(\bar{S}_x - S_x\right) + \int dx\left[\bar{P}(x) - P(x)\right]G(x)$$
$$-T\Delta\bar{D}' = -T\left(S'_x - \bar{S}_x\right) + \int dx\left[P'(x) - \bar{P}(x)\right]G'(x)$$

Thermodynamics of a Molecular Engine

$$W = G^{\lambda\prime} - G' + G - G^{\lambda} = \Delta G^{\lambda} - \Delta G$$

$$W_p = T\left(S'_x - S'^{\lambda}_x\right) - \int dx\left[P'(x) - P'^{\lambda}(x)\right]G'(x)$$
$$W_s = T\left(S_x - S'_x\right) - \int dx\left[P(x) - P'^{\lambda}(x)\right]G(x)$$
$$W = W_p - W_s$$

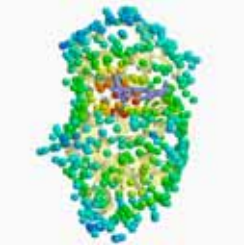


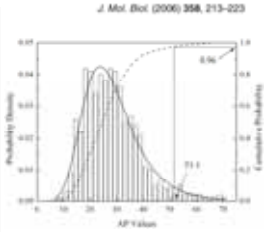
$$P'_\lambda(x) = P(x)\frac{1 + K_x(x)e^{-\Delta G/k_BT}[L]}{1 + Ve^{-\Delta G/k_BT}[L]}$$

Interactions in Native Binding Sites Cause a Large Change in Protein Dynamics

Dengming Ming* and Michael E. Wall*

J. Mol. Biol. (2006) 358, 213–223





Lysozyme in complex with NAG (1JEF)

Dynamics Perturbation Analysis

Analysis of Ligand-Binding Sites

- 305 proteins in GOLD test set
- Select DPA points within 6.0 A of ligand
- P_{kl}^{+} = Fraction of points in protein *k* with D_x value higher than point *l*
- Total score $z_k = \prod_{l \in L} P_{kl}^{+}$

Null Model

Probability of randomly obtaining a score z or smaller

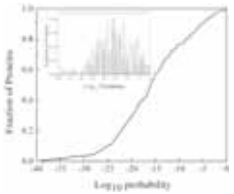
$$\rho_z(z) = \int_0^\infty dP_1 \theta(1-P_1) \int_0^\infty dP_2 \theta(1-P_2) \ldots \int_0^\infty dP_N \delta(z-P_1 \ldots P_N) \theta(1-P_N)$$

$$\rho_z(z) = \int_z^1 \frac{du_{N-1}}{u_{N-1}} \int_{u_{N-1}}^1 \frac{du_{N-2}}{u_{N-2}} \ldots \int_{u_2}^1 \frac{du_1}{u_1}$$

$$\rho_z(z) = \frac{(-\ln z)^{N-1}}{(N-1)!}$$

$$P^-(z') = \int_0^{z'} dz \rho_z(z) = z' \sum_{n=1}^{N} \frac{(-\ln z')^{n-1}}{(n-1)!}$$

Significance of Elevated Values of D_x



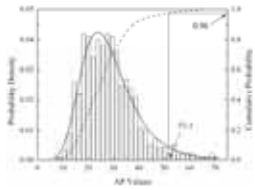
Interactions in native binding sites cause a large change in protein dynamics

Possible Implications for Protein Evolution

- Native, small-molecule binding sites evolve where interactions cause a large change in the conformational distribution
- Are sites where interactions cause a large change in the conformational distribution especially well-suited for controlling protein activity?
- **If so, then perhaps functional sites evolve at locations that are well-suited for controlling protein activity**

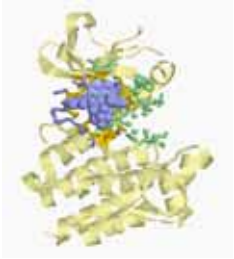
Algorithm to Predict Functional Sites

$$\rho(y) = \frac{1}{\beta} e^{\frac{y-\mu}{\beta}} e^{-e^{\frac{y-\mu}{\beta}}}$$



Typical Results

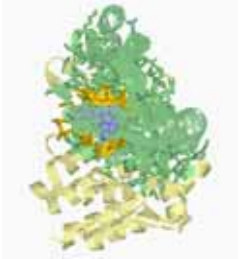
- Ligand and nearby residues
  - Purple
- DPA cluster and nearby residues
  - Green
- Overlap between ligand residues and DPA residues
  - Orange
- Good recall
- Good precision



Kinase domain of CSK in complex with staurosporine (1BYG)

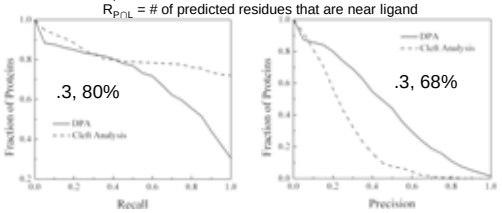
Comparison to Cleft Analysis

- Ligand and nearby residues
  - Purple
- Cleft surface mesh and nearby residues
  - Green
- Overlap between ligand residues and surface mesh
  - Orange
- Good recall
- Bad precision



Recall and Precision

R_L = # of residues near ligand
R_P = # of residues predicted
$R_{P \cap L}$ = # of predicted residues that are near ligand



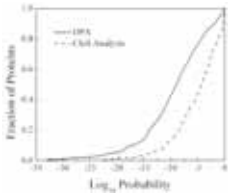
$R_{P \cap L}/R_L$ $R_{P \cap L}/R_P$

Null Model

R_L = # of residues near ligand
R_S = # of surface residues
R_P = # of residues predicted
$R_{P \cap L}$ = # of predicted residues that are near ligand

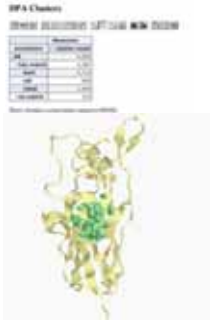
$$P_{null} = \sum_{n=R_{P \cap L}}^{\min(R_P, R_L)} \binom{R_L}{n} \Big/ \binom{R_S}{R_P} = \sum_{n=R_{P \cap L}}^{\min(R_P, R_L)} \frac{R_L! R_P! (R_S - R_P)!}{n!(R_L - n)! R_S!}$$

Significance of Predictions



DPA Applied to 50,000 Protein Domains

- 5020 DPA clusters with high sequence conservation
- 433 predictions with high sequence conservation, and no match to PDB ligand or catalytic site
- Predictions for 118 SCOP families where automated transitive annotation is possible
  - Correct predictions for ligand-free structures
- Predictions for 14 SCOP families where literature search is required
  - E.g. RBD from the pathogenic FeLV-B
    - 79-81,98,100-104
    - Trp101 previously postulated
    - Mediate receptor-cofactor interactions?



Summary

- Ligand binding alters the protein energy landscape
  - "New view" of allostery
- Allosteric effects can be studied computationally using Dynamics Perturbation Analysis (DPA)
- Methods of DPA can be used to improve coarse-grained models of proteins
- Interactions in native binding sites cause a large change in protein dynamics
  - Possible implications for protein evolution
- DPA can be used to predict protein functional sites

Acknowledgments

- LANL
  - Judith Cohn
  - James Faeder
  - William Hlavacek
  - James Howse
  - Dengming Ming, PD (moving to Nanjing U)
  - Ilya Nemenman
  - Jelena Stajic, PD (presently at UCHC)
  - Charles Strauss
  - Thomas Terwilliger
- External
  - Mary Dunlop, CalTech, Krell-DOE/CSGF GS
  - David Markowitz, Princeton, Krell-DOE/CSGF GS
  - Robert G. Martin, NIH/NIDDK
  - J. Lee Rosner, NIH/NDDK
  - Michael Savageau, UC/Davis
- Department of Energy
- National Institutes of Health