BAYESIAN SUPERPOSITIONS WITH MCMC SAMPLING

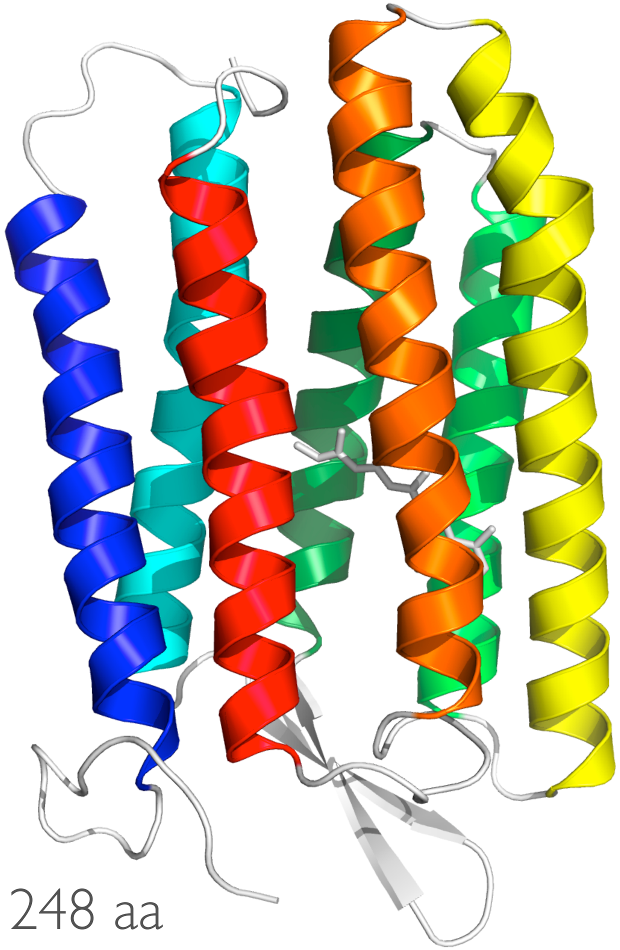
ADISTA 2014
Brussels, Belgium
22 May 2014

Douglas Theobald
Department of Biochemistry
Brandeis University

`theobald.brandeis.edu`

The bacteriorhodopsin/GPCR superfold

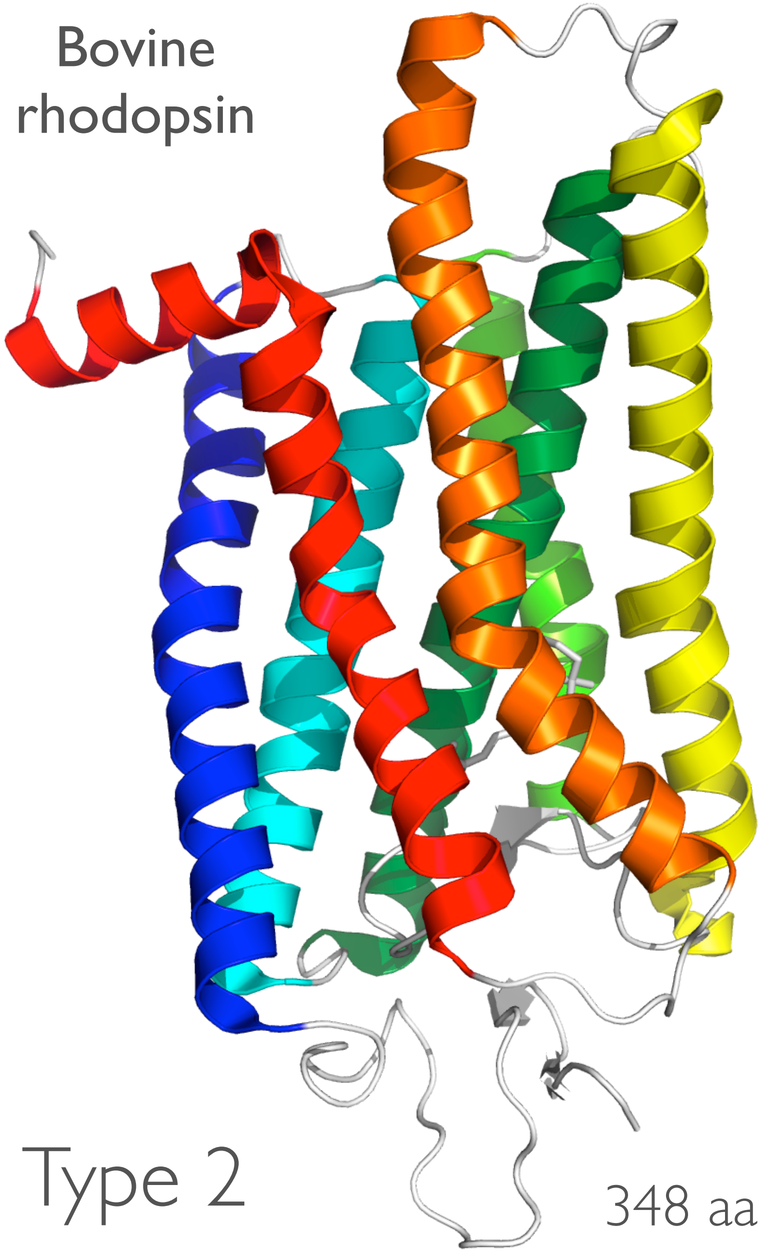
Halobacterium salinarum
bacteriorhodopsin



248 aa

Type 1

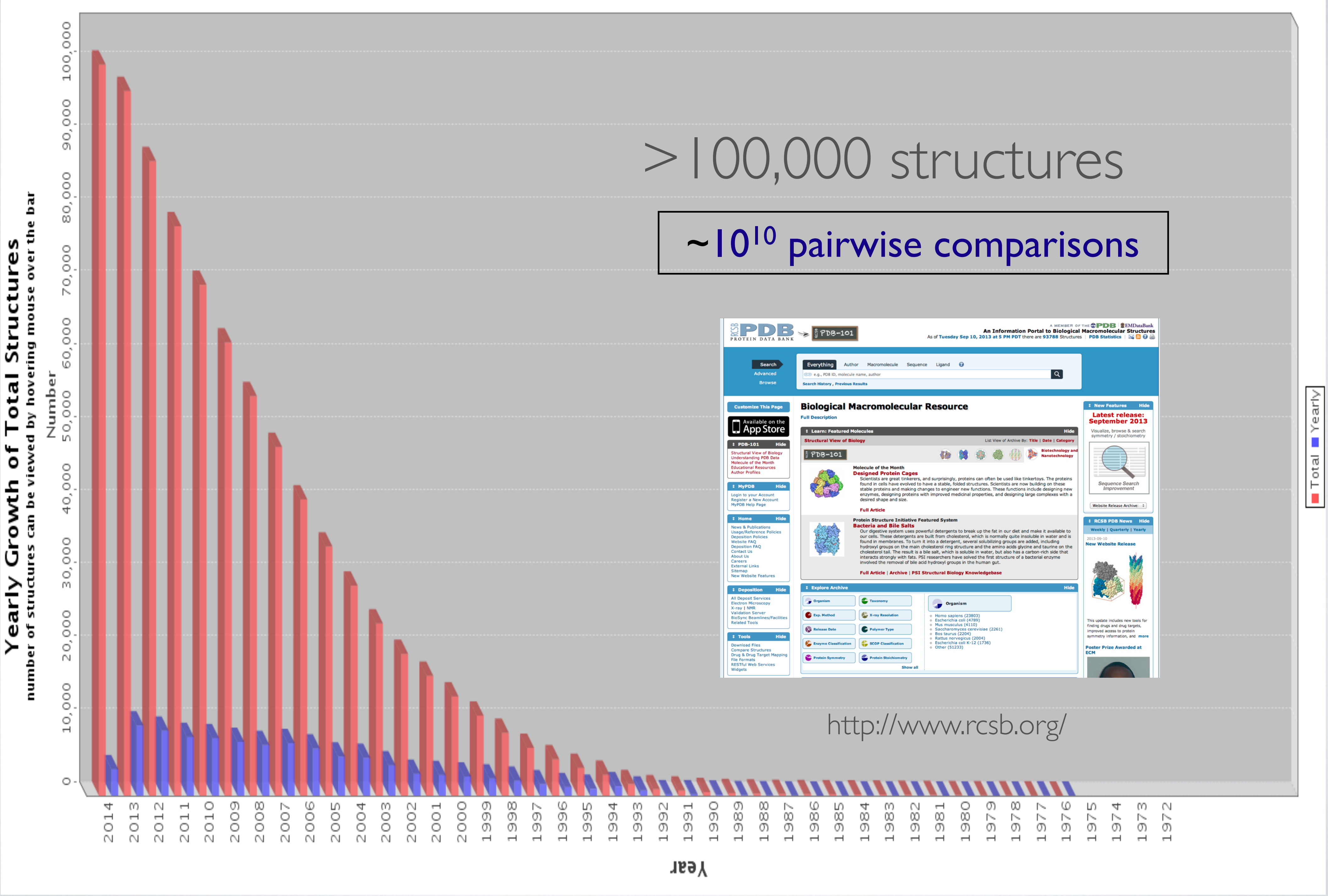
Bovine
rhodopsin



Type 2

348 aa

Historic Protein Data Bank growth



>100,000 structures

~10^{10} pairwise comparisons

RCSB PDB PROTEIN DATA BANK

PDB-101

An Information Portal to Biological Macromolecular Structures

As of Tuesday Sep 10, 2013 at 5 PM PDT there are 93788 Structures | PDB Statistics

Biological Macromolecular Resource

Molecule of the Month

Designed Protein Cages

Protein Structure Initiative Featured System

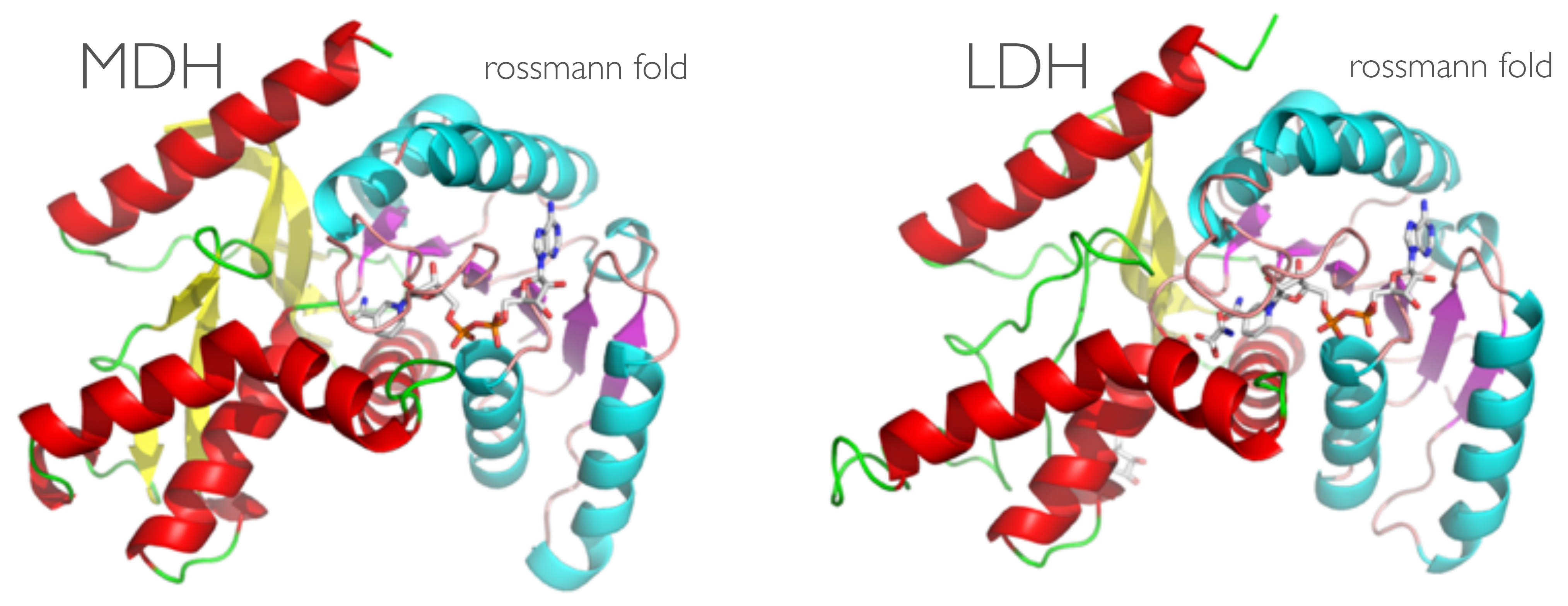
Bacteria and Bile Salts

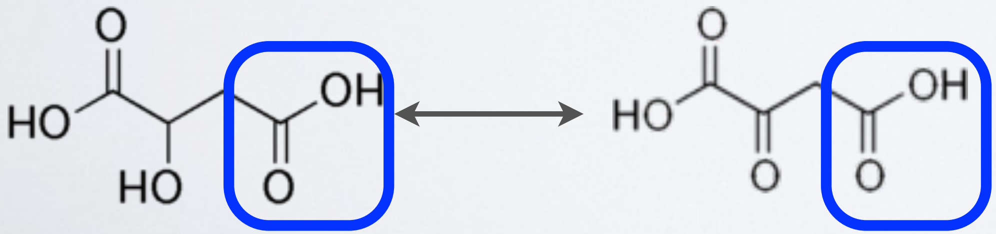
Latest release: September 2013

http://www.rcsb.org/

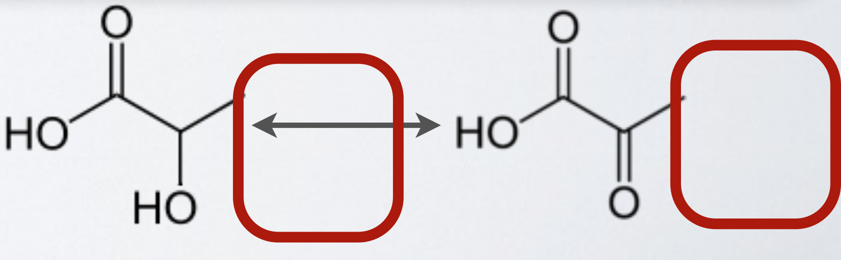
Malate and lactate dehydrogenases

~40% sequence identity over > 300 residues





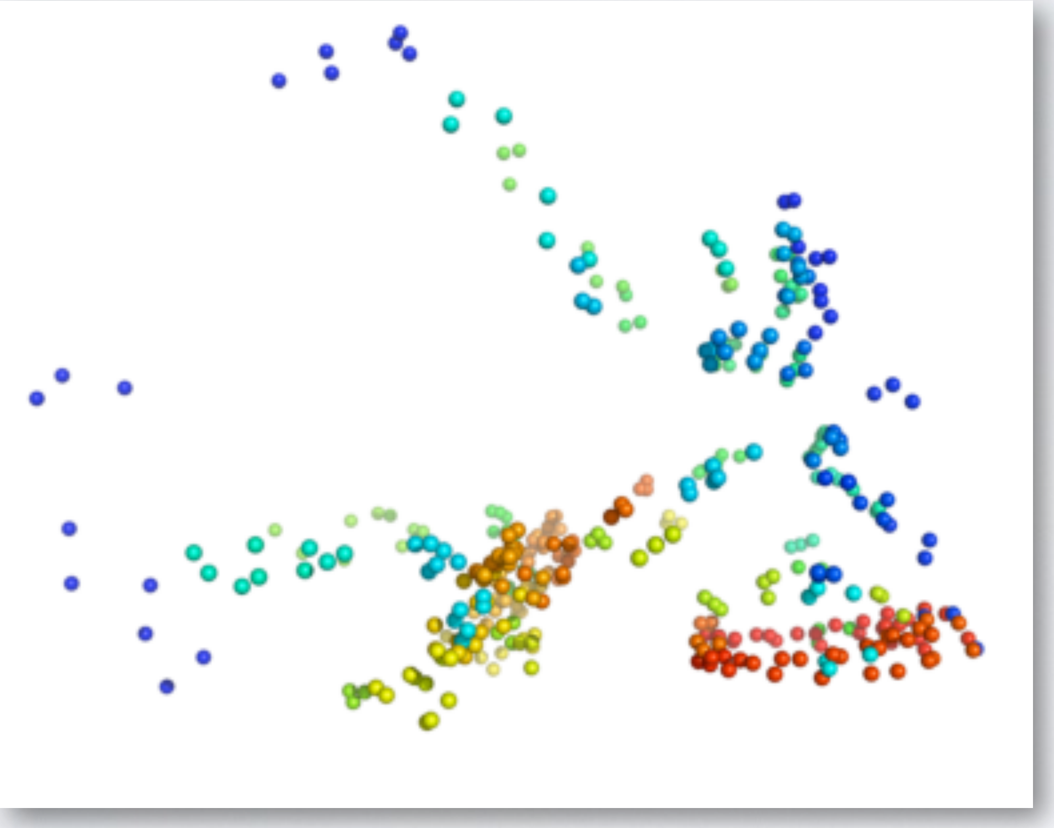
malate oxaloacetate

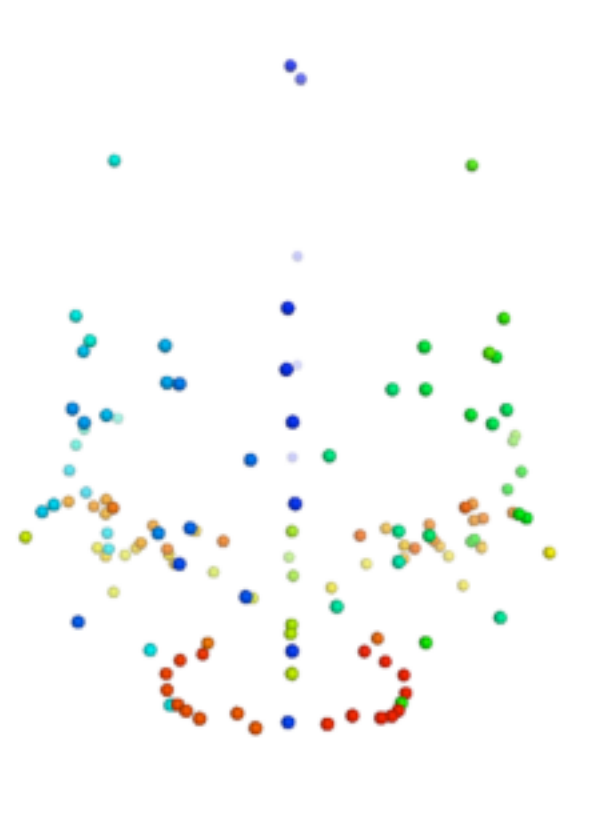


lactate pyruvate

Morphometrics: Hominid crania

Three hominid skulls, 131 landmarks

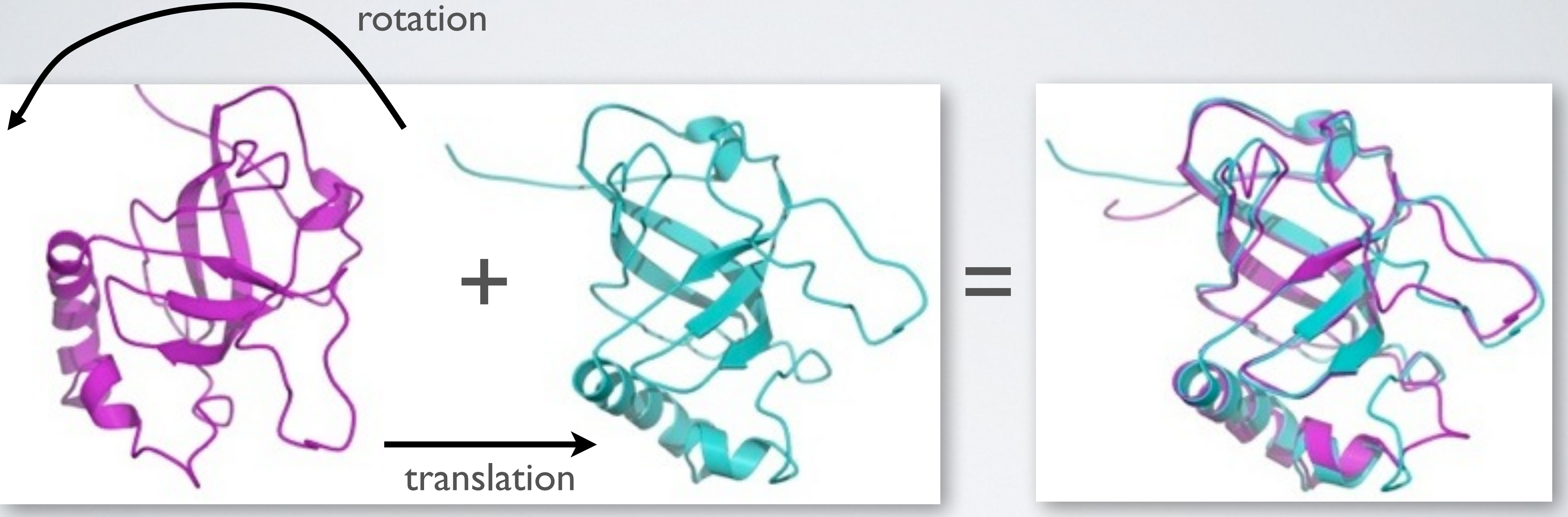




Morphological features can be different sizes, so these involve *scaling*

Superposing fits structures together

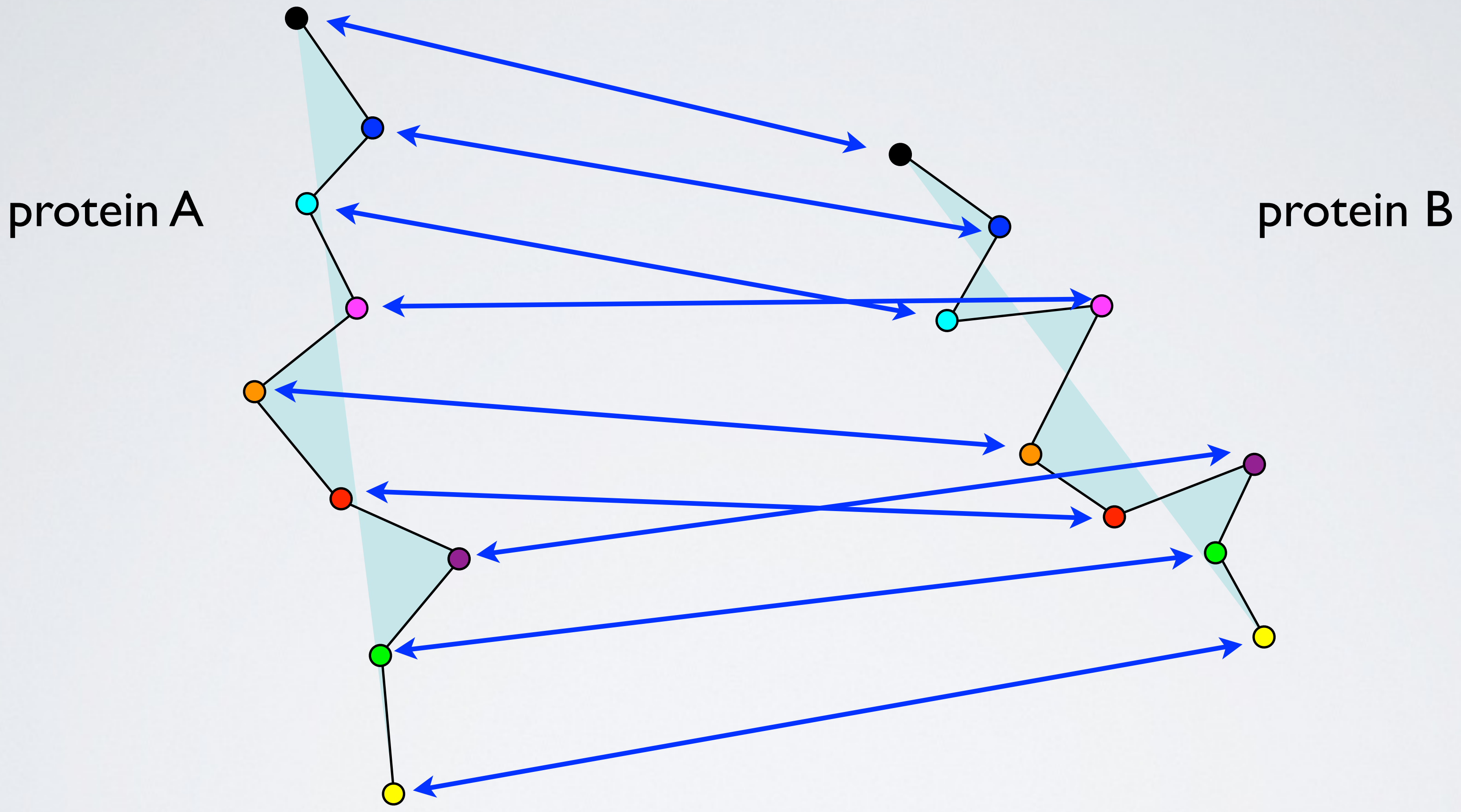
Superposition = "optimal" relative orientation of two or more corresponding sets of atoms



superposition

Least-squares: Find the rotation that minimizes the sum of <u>squared</u> distances between corresponding atoms

Classic superposition method: Least-squares



Least-squares: Find the rotation that minimizes the sum of squared distances between corresponding (labelled) atoms

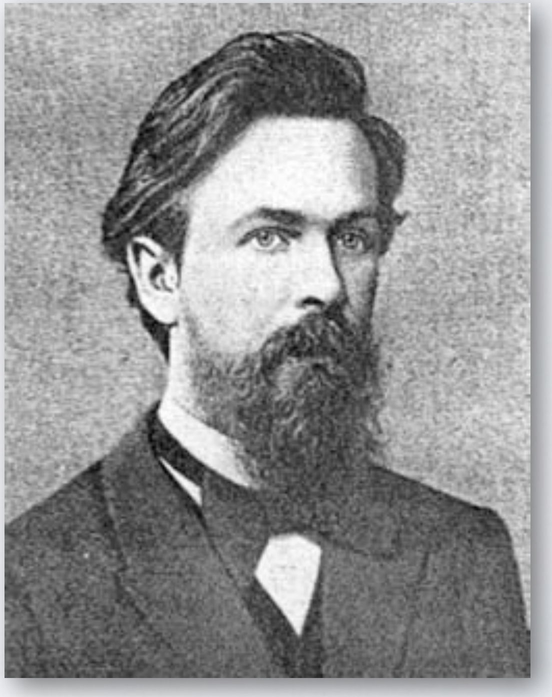
Why least-squares? (and why not?)

Gauss-Markov Theorem

Least-squares gives the "best" (BLUE) answer if:

1. Atoms have equal variance
2. Atoms are uncorrelated

Carl Friedrich Gauss



Andrey Markov

Structural models are imprecise: Experimental error and molecular dynamics

X-ray crystal structures:

B-factors

Pot1, an OB-fold

NMR structures:

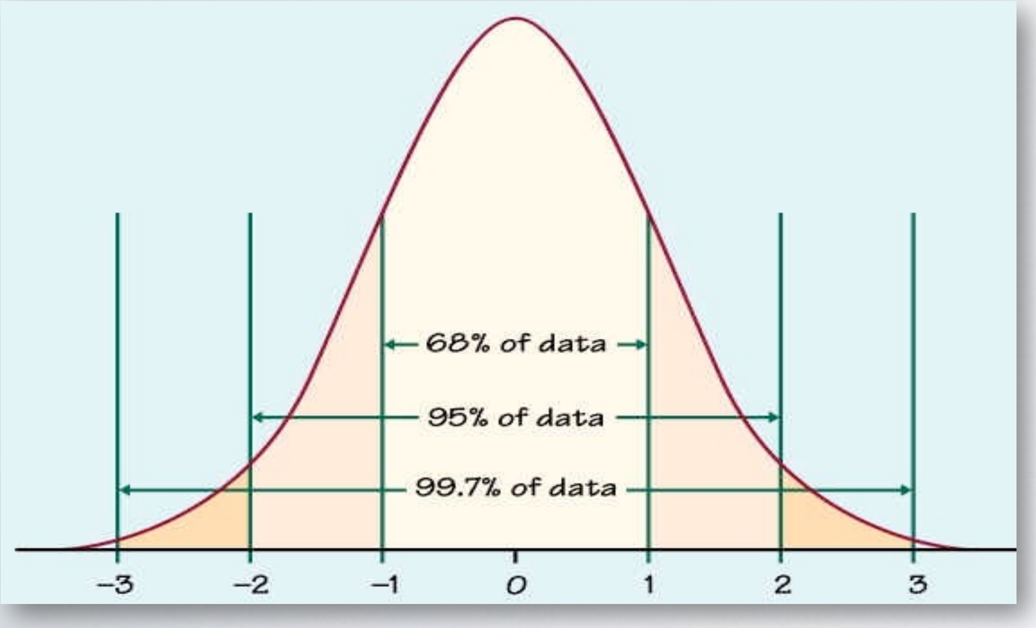
Families

Cdc13, another OB-fold

Model-based alternatives to least-squares: Maximum Likelihood (ML) and Bayes

Model-based methods:

- Assume a statistical model for the data (e.g., a Gaussian distribution).
- Estimate parameters of your model from the observed data
- ML: Find parameters that predict the data with the highest probability
- Bayes: Find distribution of the parameters given the data



$$p\left(x|\mu,\sigma\right) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

Two parameters in Gaussian PDF:
μ = location parameter (mean)
σ = width parameter

Superposition likelihood function: Gaussian probability distribution of the data

X_i is the ith molecular structure (k × 3 matrix) -- n structures, k atoms

Perturbation model

$$\mathbf{X}_i = \frac{1}{\beta_i}\left(\mathbf{M} + \mathbf{E}_i\right)\mathbf{R}_i' - \mathbf{1}_k \boldsymbol{t}_i' \qquad \mathbf{E}_i \sim \mathrm{N}_{k,d}(\mathbf{0}, \boldsymbol{\Sigma}, \mathbf{I})$$

Likelihood

$$\mathrm{p}\left(\mathbf{X}|\boldsymbol{\Sigma}, \mathbf{M}, \mathbf{R}, \boldsymbol{t}, \beta\right) = C \exp\left(-\frac{1}{2}\sum_{i}^{n} \mathrm{tr}\left\{[\mathbf{Y}_i - \mathbf{M}]'\boldsymbol{\Sigma}^{-1}[\mathbf{Y}_i - \mathbf{M}]\right\}\right)$$

$$C = (2\pi)^{-\frac{kdn}{2}}\left(\prod_{i}^{n} \beta_i^{kd}\right)|\boldsymbol{\Sigma}|^{-\frac{dn}{2}} \qquad \mathbf{Y}_i = (\beta_i \mathbf{X}_i + \mathbf{1}_k \boldsymbol{t}_i')\mathbf{R}_i$$

t_i is its translation (3-vector)

R_i is its rotation (3 × 3 matrix)

β_i is its scaling factor

Σ is the overall covariance matrix (k × k matrix)

M is the overall mean structure (k × 3 matrix)

unknown parameters

Superposition likelihood function: Gaussian probability distribution of the data

X_i is the i^{th} molecular structure (k × 3 matrix) -- n structures, k atoms

Perturbation model

$$\mathbf{X}_i = \frac{1}{\beta_i}\left(\mathbf{M}+\mathbf{E}_i\right)\mathbf{R}_i' - \mathbf{1}_k \boldsymbol{t}_i' \qquad \mathbf{E}_i \sim \mathrm{N}_{k,d}(\mathbf{0},\boldsymbol{\Sigma},\mathbf{I})$$

Likelihood

$$\mathrm{p}\left(\mathbf{X}|\boldsymbol{\Sigma},\mathbf{M},\mathbf{R},\boldsymbol{t},\beta\right) = C\exp\left(-\frac{1}{2}\sum_i^n \mathrm{tr}\left\{[\mathbf{Y}_i-\mathbf{M}]'\boldsymbol{\Sigma}^{-1}[\mathbf{Y}_i-\mathbf{M}]\right\}\right)$$

$$C = (2\pi)^{-\frac{kdn}{2}}\left(\prod_i^n \beta_i^{kd}\right)|\boldsymbol{\Sigma}|^{-\frac{dn}{2}} \qquad \mathbf{Y}_i = (\beta_i\mathbf{X}_i + \mathbf{1}_k\boldsymbol{t}_i')\mathbf{R}_i$$

Hierarchical prior on Σ

$$\boldsymbol{\Sigma}\sim \mathrm{IW}\left(\boldsymbol{\Phi},n,k\right)$$

$$\mathrm{p}\left(\boldsymbol{\Sigma}|\boldsymbol{\Phi},n,k\right) = \frac{|\boldsymbol{\Phi}|^{\frac{n}{2}}}{2^{\frac{nk}{2}}\,|\boldsymbol{\Sigma}|^{\left(\frac{n}{2}+1\right)}\,\Gamma\left(\frac{n}{2}\right)^k}\exp\left\{-\frac{1}{2}\,\mathrm{tr}\left(\boldsymbol{\Phi}\boldsymbol{\Sigma}^{-1}\right)\right\}$$

$$\boldsymbol{\Phi}=\phi\mathbf{I}$$

Superposition marginal likelihood function: "Scaled" matrix *t*-distribution

$$\mathbf{X}_i \sim \mathrm{MT}\left(\mathbf{M}, \mathbf{R}, \boldsymbol{t}, \phi, n, k, \beta\right)$$

$$\mathrm{p}\left(\mathbf{X}_i | \mathbf{M}, \mathbf{R}, \boldsymbol{t}, \phi, n, k, \beta\right) = \left(\prod_i^k {\beta_i}^{kd}\right) \Gamma\left(\frac{3n+1}{2}\right)^k \pi^{-k\left(\frac{3n+1}{2}\right)} \phi^{\frac{k}{2}} \; \left|\phi\mathbf{I} + (\mathbf{I} \odot \mathbf{S})\right|^{-\frac{3n+1}{2}}$$

$$\mathbf{\Phi} = \phi\mathbf{I}$$

$$\mathbf{S} = \sum_i^N [\mathbf{Y}_i - \mathbf{M}][\mathbf{Y}_i - \mathbf{M}]'$$

$$\mathbf{Y}_i = (\beta_i \mathbf{X}_i + \mathbf{1}_k \boldsymbol{t}_i')\mathbf{R}_i$$

Empirical Bayes EM ML superposition solutions

$$\hat{\boldsymbol{t}}_i = -\frac{\mathbf{1}'_K \boldsymbol{\Sigma}^{-1} \mathbf{X}_i}{\mathbf{1}'_K \boldsymbol{\Sigma}^{-1} \mathbf{1}_K} \qquad \check{\mathbf{X}}_i = \mathbf{X}_i + \mathbf{1}_K \hat{\boldsymbol{t}}_i$$

Translations

$$\hat{\mathbf{M}}' \hat{\boldsymbol{\Sigma}}^{-1} \check{\mathbf{X}}_i = \mathbf{U} \boldsymbol{\Lambda} \mathbf{V}'$$

$$\hat{\mathbf{R}}_i = \mathbf{V} \mathbf{P} \mathbf{U}'$$

Rotations - found with Singular Value Decomposition

$$\mathrm{E}\left(\boldsymbol{\Sigma}^{-1} | \mathbf{X}, \boldsymbol{t}, \mathbf{R}, \boldsymbol{\beta}, \mathbf{M}, \phi, n\right) = (3N + n)(\mathbf{S} + \phi \mathbf{I})^{-1}$$

Covariance matrix

$$\mathbf{S} = \sum_{i}^{N} [\mathbf{Y}_i - \mathbf{M}][\mathbf{Y}_i - \mathbf{M}]' \qquad \mathbf{Y}_i = (\beta_i \mathbf{X}_i + \mathbf{1}_k \boldsymbol{t}'_i) \mathbf{R}_i$$

$$\hat{\alpha} = \frac{K}{2 \,\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\right)}$$

$$\hat{\mathbf{M}} = \frac{1}{N} \sum_{i}^{N} \check{\mathbf{X}}_i \mathbf{R}_i$$

Mean structure

Must be solved simultaneously (CM & EM algorithm)
Our program THESEUS implements this method (www.theseus3d.org)

Theobald and Wuttke (2006) PNAS 103:1852, Theobald and Wuttke (2006) Bioinformatics 22:2171,
Theobald and Wuttke (2008) PLoS Comput Biol 4:e43

Why Theseus? Superpositioning is a Procrustes problem

Theseus

Procrustes

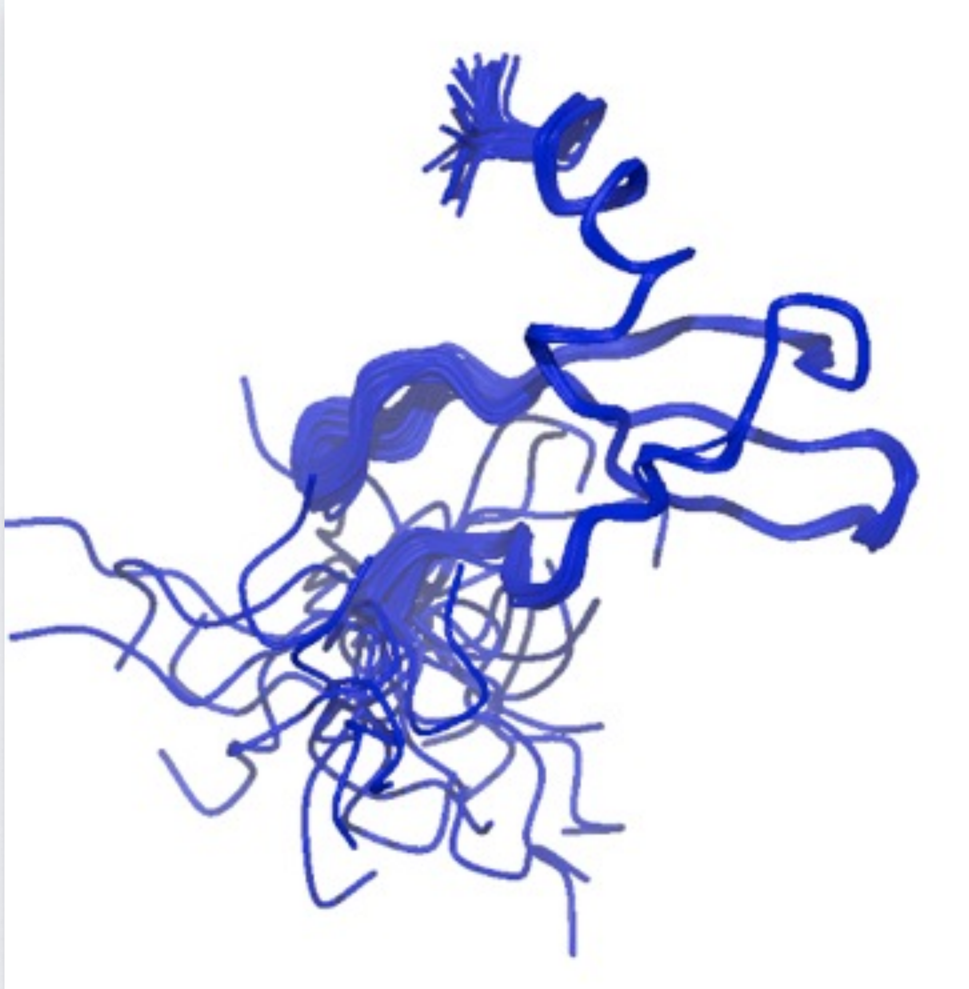
Procrustean
Bed

Maximum likelihood superpositions

2sdf: cytokine stromal cell-derived factor-1 (SDF-1)

67 aa, 30 NMR models

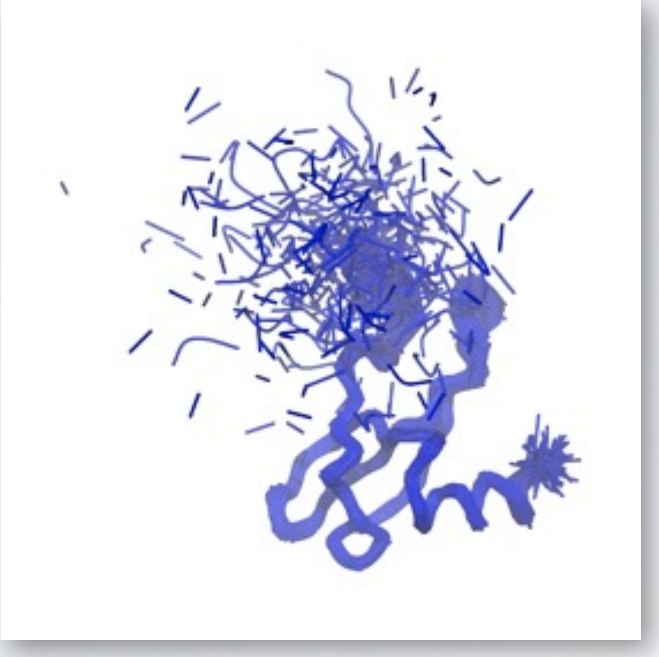
Least-squares



Maximum likelihood

LS vs ML: Maximum likelihood recovers the true covariance matrix accurately

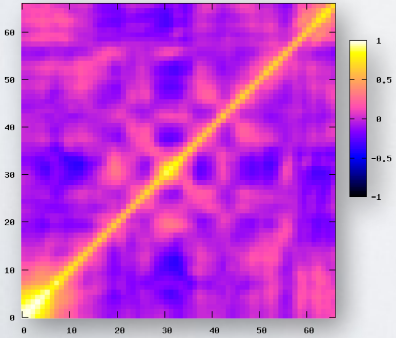
True superposition

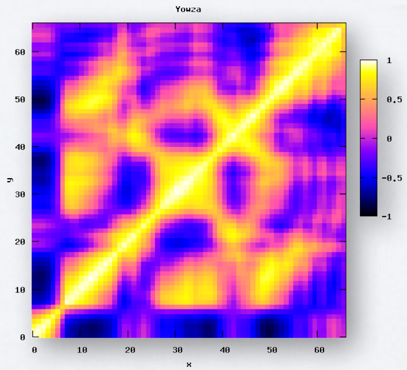


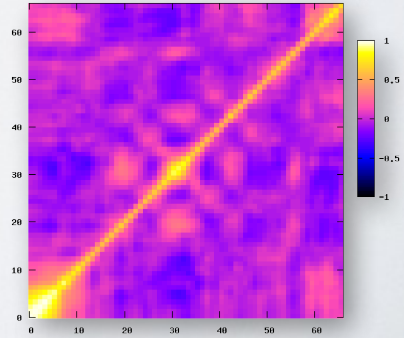
Least-squares

Maximum likelihood

True correlation matrix







Benefits of Bayes

The object of Bayesian inference: Provide a full joint distribution of model parameters

- Complete distribution of parameters, including uncertainty
- Uses all relevant information, including prior info
- Complex models, marginal distributions for "nuisance parameters"
- Solutions for underdetermined and "problematic" models
- Bayesian solutions are exact for finite sample sizes (unlike ML)

Bayes Theorem, easy version

$$\mathrm{p}(\theta|x, M) \propto \mathrm{p}(x|\theta, M)\, \mathrm{p}(\theta|M)$$

posterior
(PDF of parameters, given data)

prior
(PDF of parameters, no data)

likelihood
(PDF of data)

Marginalization: How Bayesians ignore important but uninteresting parameters

$$\mathrm{p}(\theta|x, M) = \int_{\phi} \mathrm{p}(\theta, \phi|x, M)\; d\phi$$

Bayesian analysis of multiple superpositioning

Assume independent priors on each parameter:

- Uniform improper priors on mean M, translations t_i
- Uniform proper prior on rotations R_i
- Exponential prior on scale factor β_i
- Isotropic: Reference prior on isotropic variance ϕ
- Nonisotropic: Vague proper prior on hyper-parameter λ

$$\mathrm{p}\left(\boldsymbol{\Sigma}\right) \propto \mathrm{p}\left(\boldsymbol{\Sigma}|\boldsymbol{\Psi}, n\right)\mathrm{p}\left(\boldsymbol{\Psi}\right).$$

$$\mathrm{p}\left(\boldsymbol{\Sigma}, \mathbf{M}, \mathbf{R}, \boldsymbol{t}, \beta|\mathbf{X}\right) \propto \mathrm{p}\left(\mathbf{X}|\boldsymbol{\Sigma}, \mathbf{M}, \mathbf{R}, \boldsymbol{t}, \beta\right)\underline{\mathrm{p}\left(\beta\right)\mathrm{p}\left(\mathbf{M}\right)\mathrm{p}\left(\mathbf{R}\right)\mathrm{p}\left(\boldsymbol{t}\right)\mathrm{p}\left(\boldsymbol{\Sigma}|\boldsymbol{\Psi}, n\right)\mathrm{p}\left(\boldsymbol{\Psi}\right)}$$

Analytic solutions for all conditional distributions

CEM for Maximum *A Posteriori* estimation

Gibbs/MCMC sampling for full Bayesian solution

Conditional distribution for the mean: Matrix normal

$$\boxed{\mathbf{M} \sim \mathrm{N}_{k,d}(\mathbf{B}, \boldsymbol{\Omega}, \mathbf{I}_d)}$$

$$\boldsymbol{\Omega}_{\text{ref}} = \frac{1}{N}\boldsymbol{\Sigma}$$

$$\mathbf{B}_{\text{ref}} = \frac{1}{N}\sum_{i}^{N}\mathbf{Y}_i$$

$$\mathbf{Y}_i = (\beta_i \mathbf{X}_i + \mathbf{1}_k \boldsymbol{t}_i')\mathbf{R}_i$$

A matrix normal distribution -- with uniform reference prior, it is centered on the sample average (the ML estimate)

Conditional distribution for translations: Multivariate normal

$$\boxed{\boldsymbol{t}_i \sim \mathrm{N}_{d,1}(\tau_i, \theta_i)}$$

$$\tau_i = -\frac{\mathbf{1}_k' \boldsymbol{\Sigma}^{-1} \mathbf{X}_i}{\mathbf{1}_k' \boldsymbol{\Sigma}^{-1} \mathbf{1}_k}$$

$$\theta_i = \frac{1}{\beta_i^2 \left(\alpha + \mathbf{1}_k' \boldsymbol{\Sigma}^{-1} \mathbf{1}_k\right)}$$

A multivariate normal distribution: With uniform reference prior, it is centered on the the ML estimate (the weighted centroid)

Conditional distribution for nonisotropic covariance matrix: Inverse Wishart

Assume a conjugate hierarchical prior for covariance matrix, a diagonal, isotropic inverse Wishart distribution

$$\boxed{\mathbf{\Sigma} \sim \mathrm{IW}\left(\mathbf{\Psi} + \mathbf{S}, k(n+2)\right)}$$

$$\mathbf{S} = \sum_{i}^{n} [\mathbf{Y}_i - \mathbf{M}]' [\mathbf{Y}_i - \mathbf{M}]$$

$$\mathbf{Y}_i = (\beta_i \mathbf{X}_i + \mathbf{1}_k \boldsymbol{t}_i') \mathbf{R}_i$$

$$\mathbf{\Psi} = \lambda \mathbf{I}$$

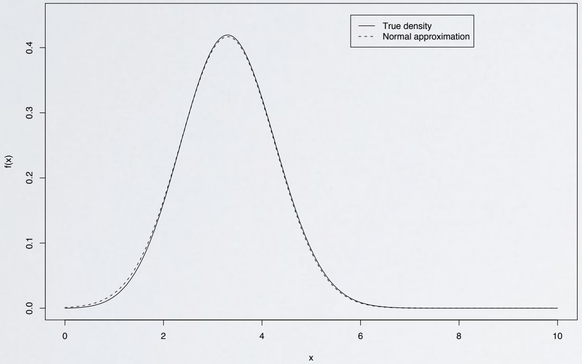
Must assume a proper prior for the hyper-parameter λ, here a conjugate gamma distribution with scale param δ and shape param *p*

$$\boxed{\lambda \sim \mathrm{G}\left(\frac{2}{\mathrm{tr}\left(\mathbf{\Sigma}^{-1} + \frac{2}{\delta}\right)}, \frac{k^2 + 2p}{2}\right)}$$

Conditional scale distribution: Halfnormal-gamma

$$\mathrm{p}\left(\beta_i|\mathbf{X},\boldsymbol{\Sigma},\mathbf{M},\boldsymbol{t}_i,\mathbf{R}\right) = C{\beta_i}^{m-1}\exp\left\{-\frac{\omega_i}{2}{\beta_i}^2+\gamma_i\beta_i\right\}$$

$$C = \frac{2\phi_i^{-\frac{m}{2}}e^{-\frac{\gamma_i^2}{8\omega_i}}}{\Gamma(m)\,\mathrm{D}_{-m}\left(\frac{\gamma_i}{\sqrt{2\omega_i}}\right)}$$



$r = 2,\ \delta = 3$

$$\boxed{\beta_i \sim \mathrm{HNG}(\omega_i,\gamma_i,m)}$$

$$\omega_i = \mathrm{tr}\left(\check{\mathbf{X}}_i'\boldsymbol{\Sigma}^{-1}\check{\mathbf{X}}_i\right)$$

$$\gamma_i = \mathrm{tr}\left(\mathbf{M}'\boldsymbol{\Sigma}^{-1}\check{\mathbf{X}}_i\mathbf{R}_i\right) - 1$$

$$\check{\mathbf{X}}_i = \mathbf{X}_i + \mathbf{1}_k\boldsymbol{t}_i'$$

$$m = kd + 1$$

Mardia et al. (2013) Annals of Applied Statistics 7(2): 989–1009.

Conditional distribution for rotations: Matrix Fisher

$$\boxed{\mathbf{R}_i \sim \mathrm{MF}(\mathbf{A}_i)}$$

$$\mathbf{A}_i = \mathbf{M}'\boldsymbol{\Sigma}^{-1}\check{\mathbf{X}}_i$$

Matrix Fisher centered on the ML estimate (using proper, uniform prior on rotations)

Sampled using:

(1) hybrid Gibbs/Metropolis-Hastings algorithm of Green and Mardia, or

(2) Gibbs using Habeck algorithm, or

(3) Kent's BACG A/R algorithm

Green and Mardia (2006) Biometrika 93:235.
Habeck (2009) Comput Stat (2009) 24:719.
Kent, Ganeiber, and Mardia (2013) arXiv:1310.8110

Gibbs/MCMC sampling for nonisotropic scaling

Initialize chain with ordinary LS superposition

$$\lambda \sim \mathrm{G}\left(\frac{k^2+2p}{2}, \frac{2}{\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}+\frac{2}{\delta}\right)}\right)$$

$$\boldsymbol{\Sigma} \sim \mathrm{IW}\left(\boldsymbol{\Psi}+\mathbf{S}, k(n+2)\right)$$

$$\mathbf{M} \sim \mathrm{N}_{k,d}(\mathbf{B}, \boldsymbol{\Omega}, \mathbf{I}_d)$$

$$\boldsymbol{t}_i \sim \mathrm{N}_{d,1}(\tau_i, \theta_i)$$

$$\mathbf{R}_i \sim \mathrm{MF}(\mathbf{A}_i)$$

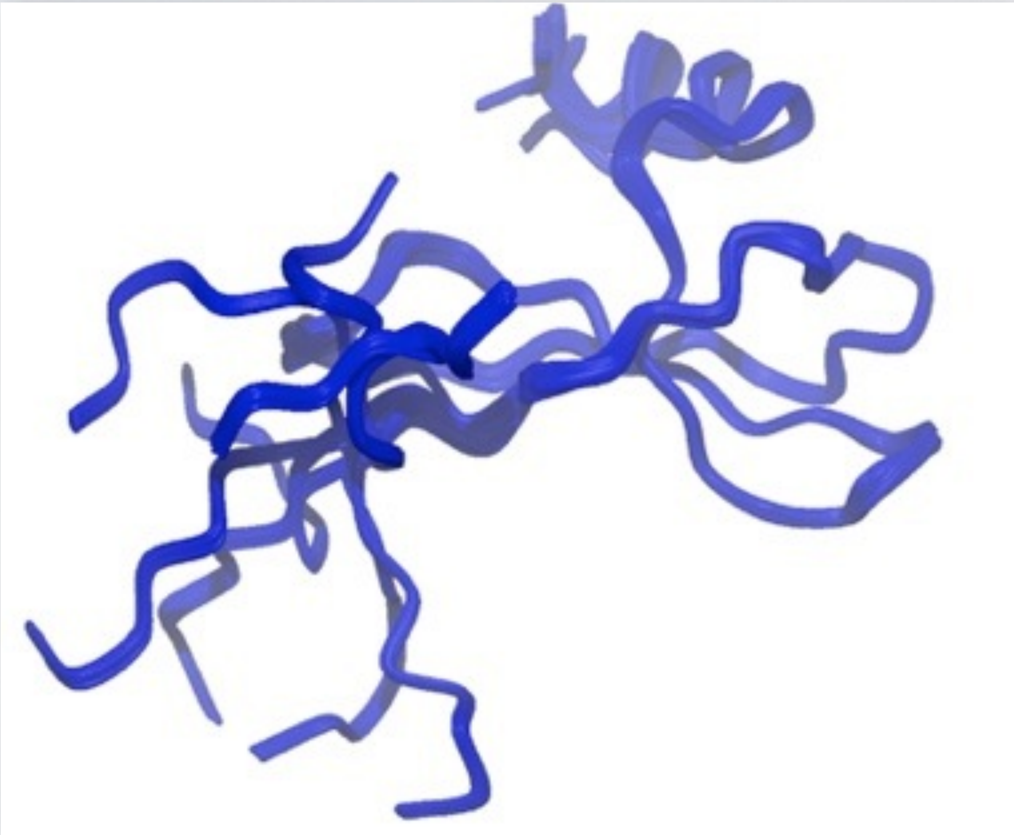
$$\beta_i \sim \mathrm{HNG}(\omega_i, \gamma_i, m)$$

Gibbs/MCMC results for nonisotropic protein superposition, no scaling

2sdf: cytokine stromal cell-derived factor-1 (SDF-1)

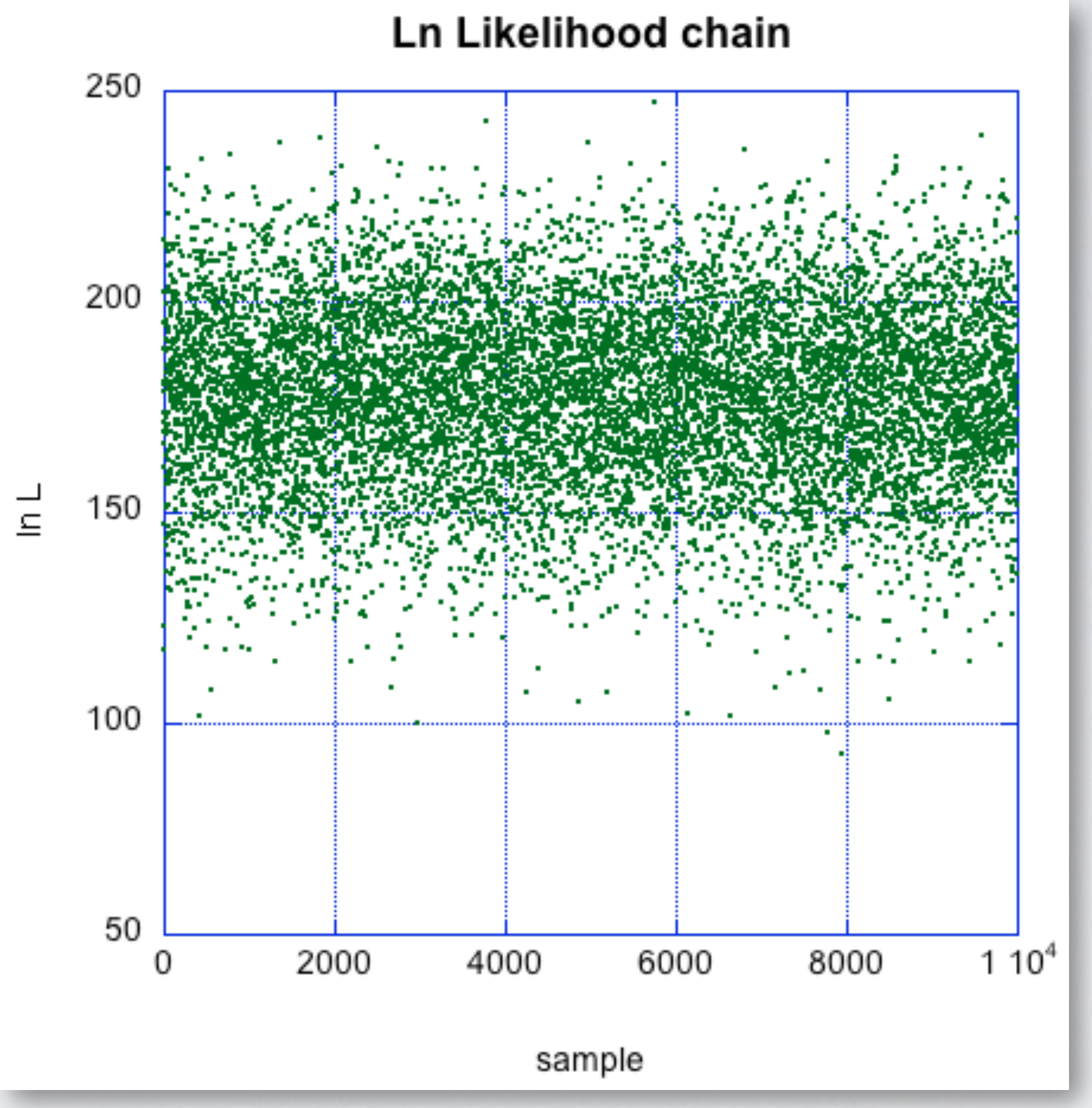
67 aa, 30 NMR models

Bayes: 10,000 subsamples, 1,000,000 generations

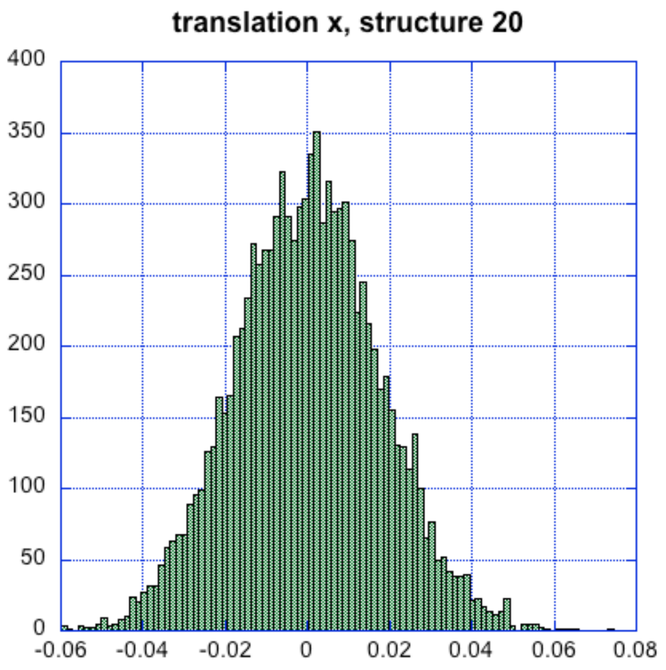


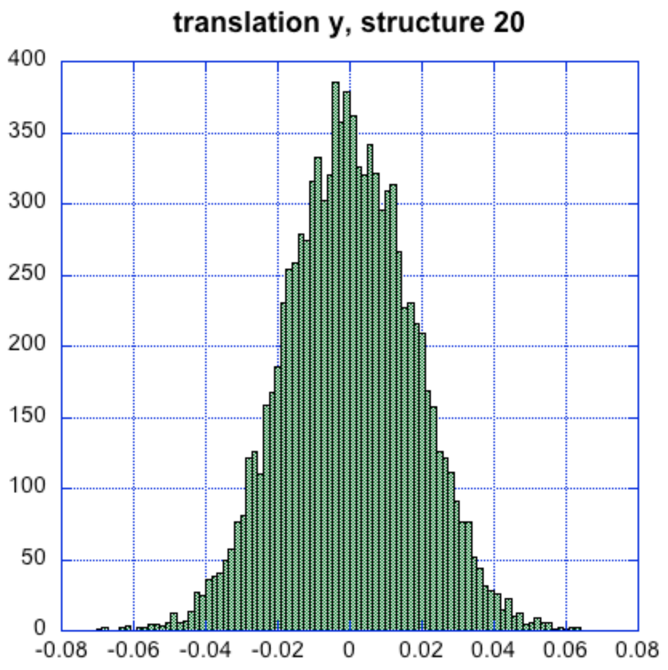
Regular ML

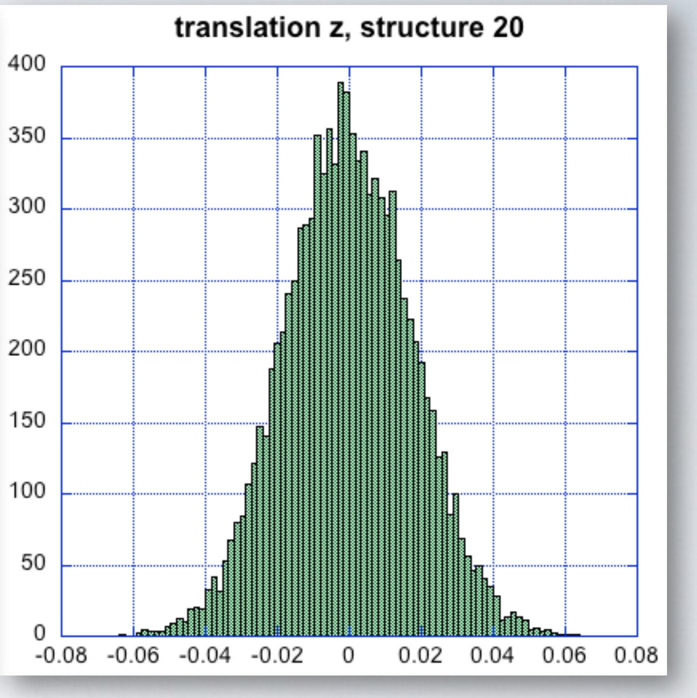
Ln Likelihood across samples

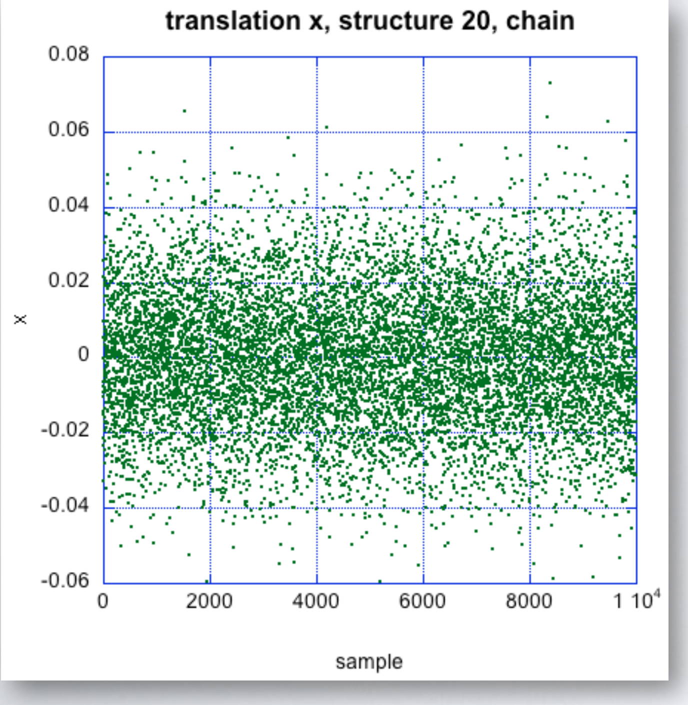


Posterior translations, structure 20

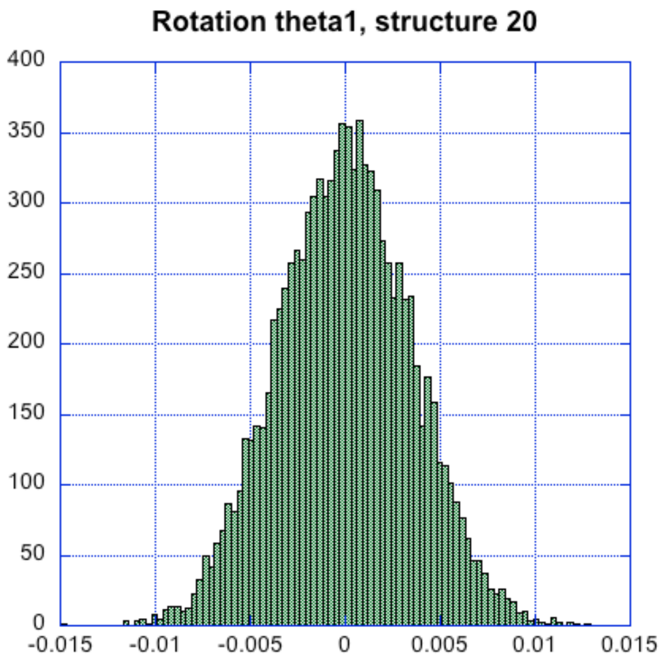


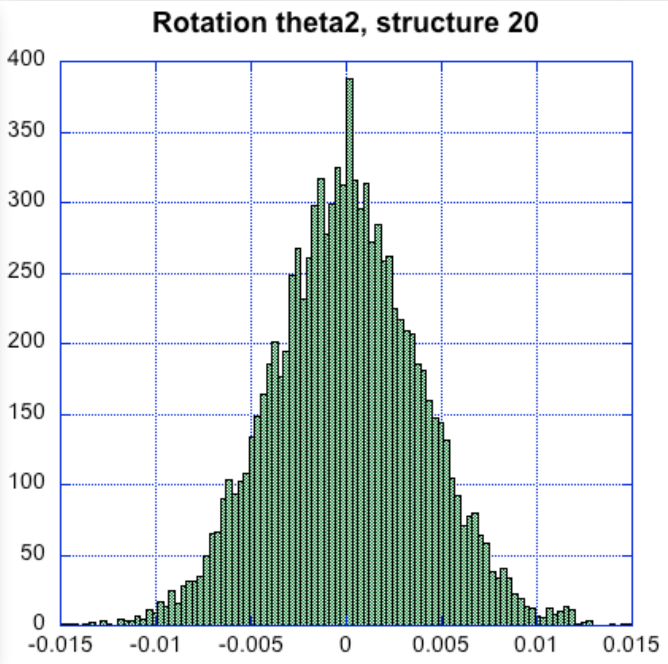


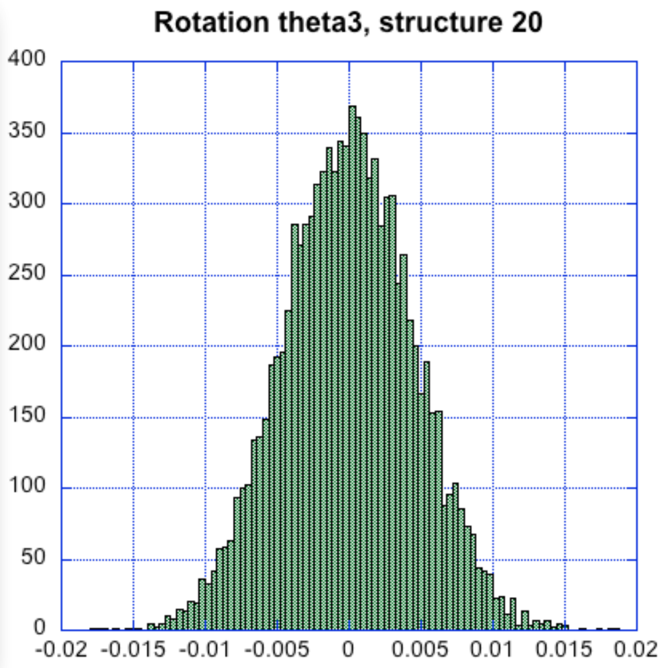


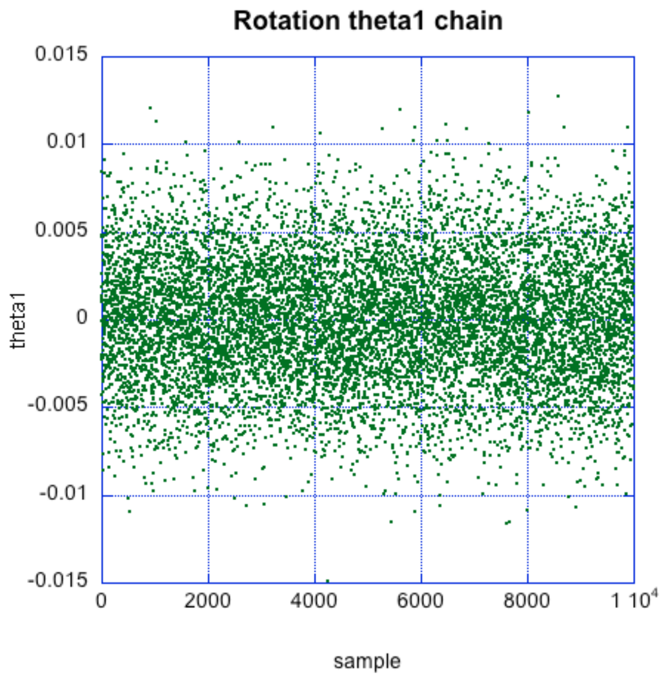


Posterior rotation angles, structure 20

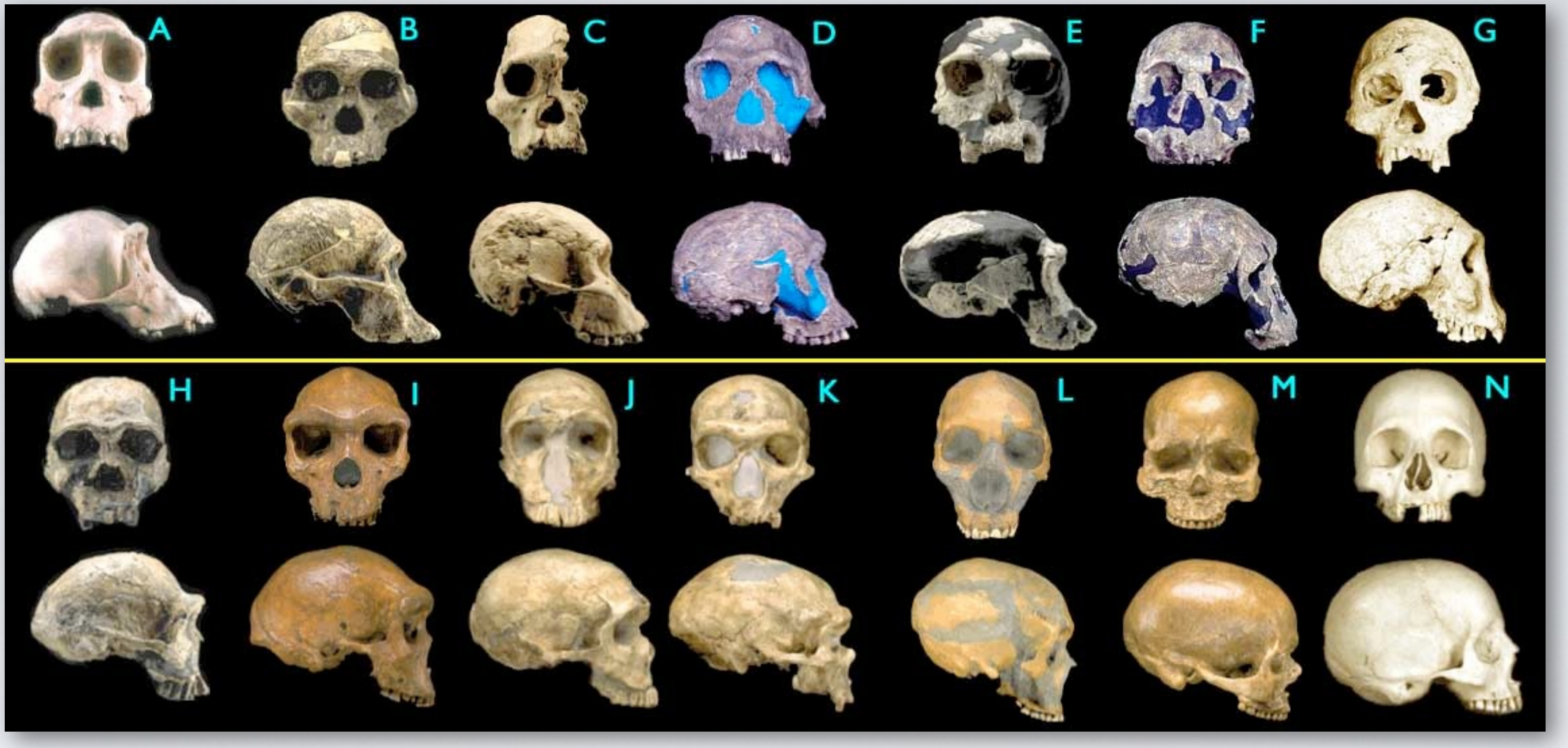

Rotation theta1, structure 20


Rotation theta2, structure 20

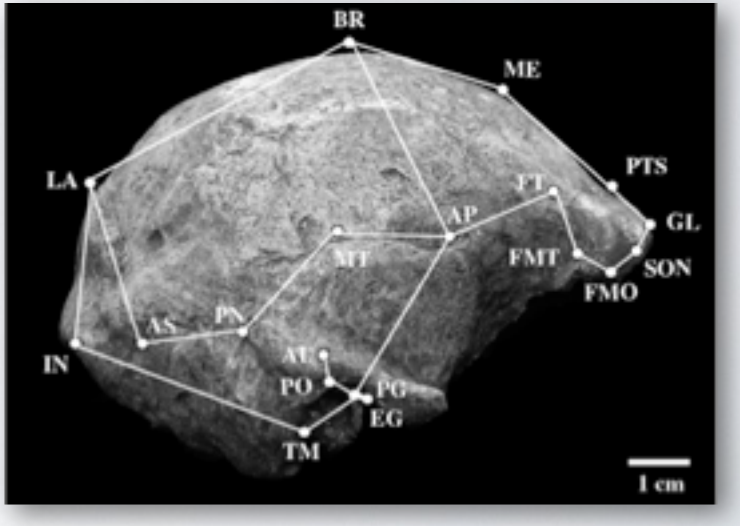

Rotation theta3, structure 20


Rotation theta1 chain

40 hominoid crania, 30 landmarks, w/scaling



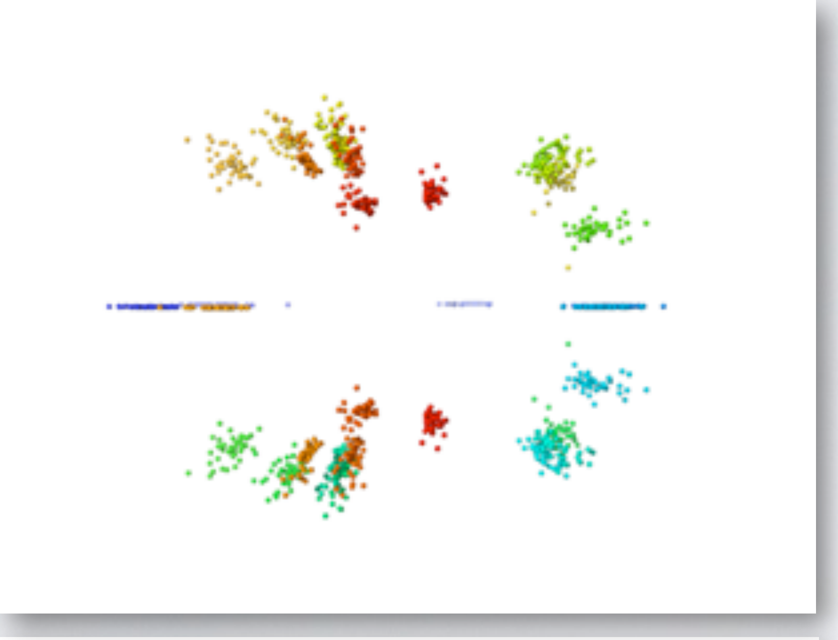
27 modern humans
9 *Homo erectus*
1 *Homo habilis*
1 Neanderthal
2 *Australopithecus bosiei*



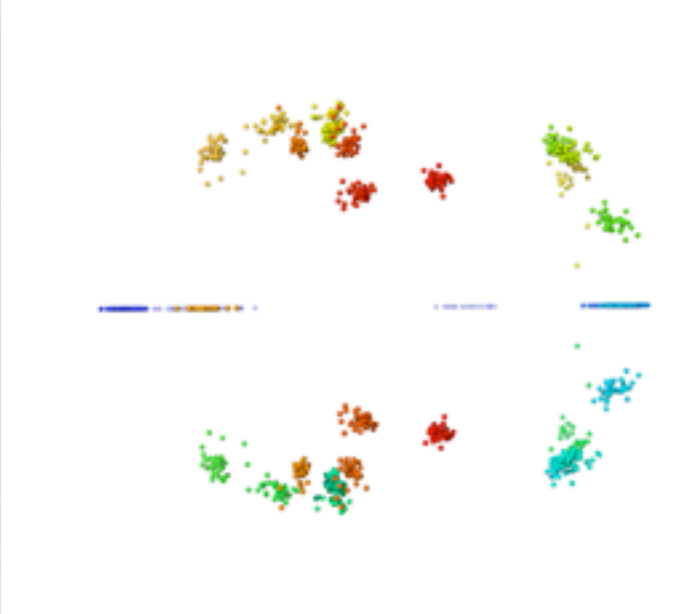
Data set kindly supplied by Karen Baab, SUNY Stony Brook

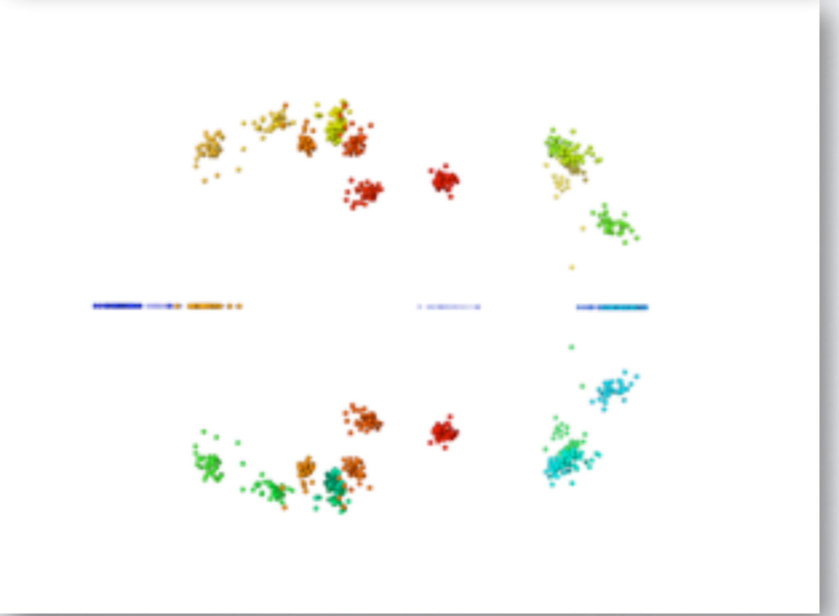
Isotropic superposition: Scaling comparisons

LS, no scaling

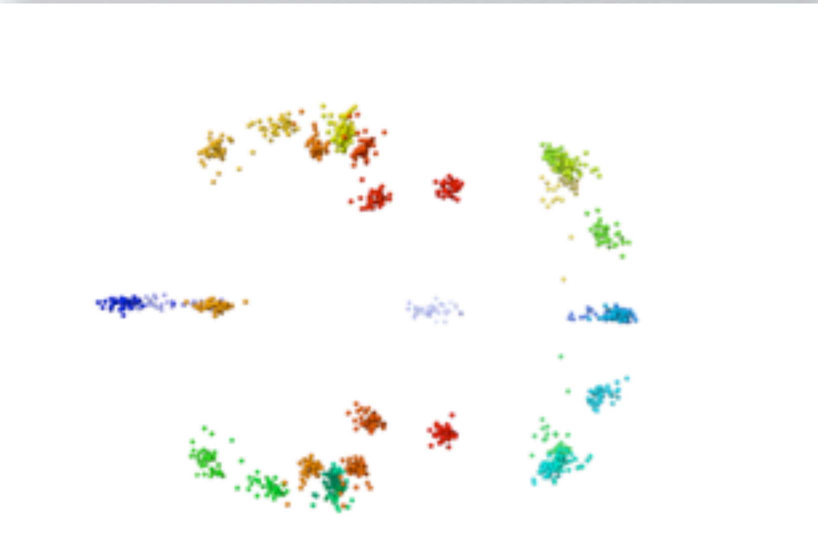


LS, scaling





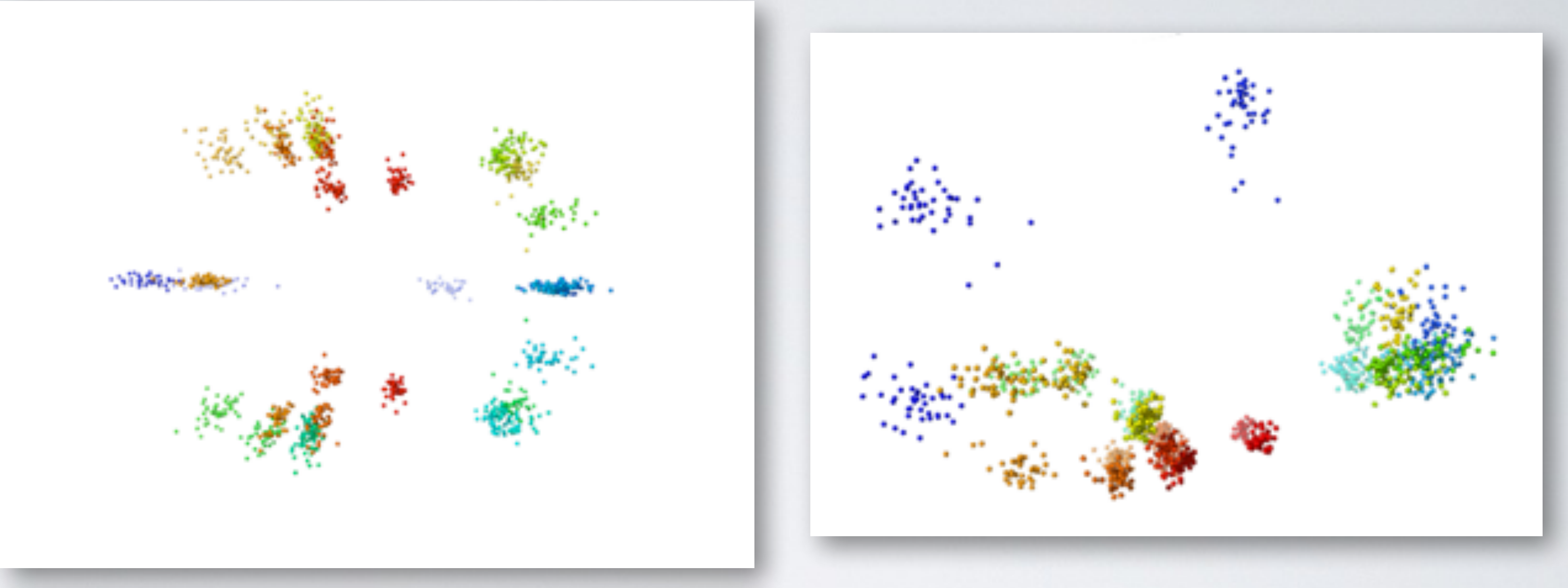
ML, scaling



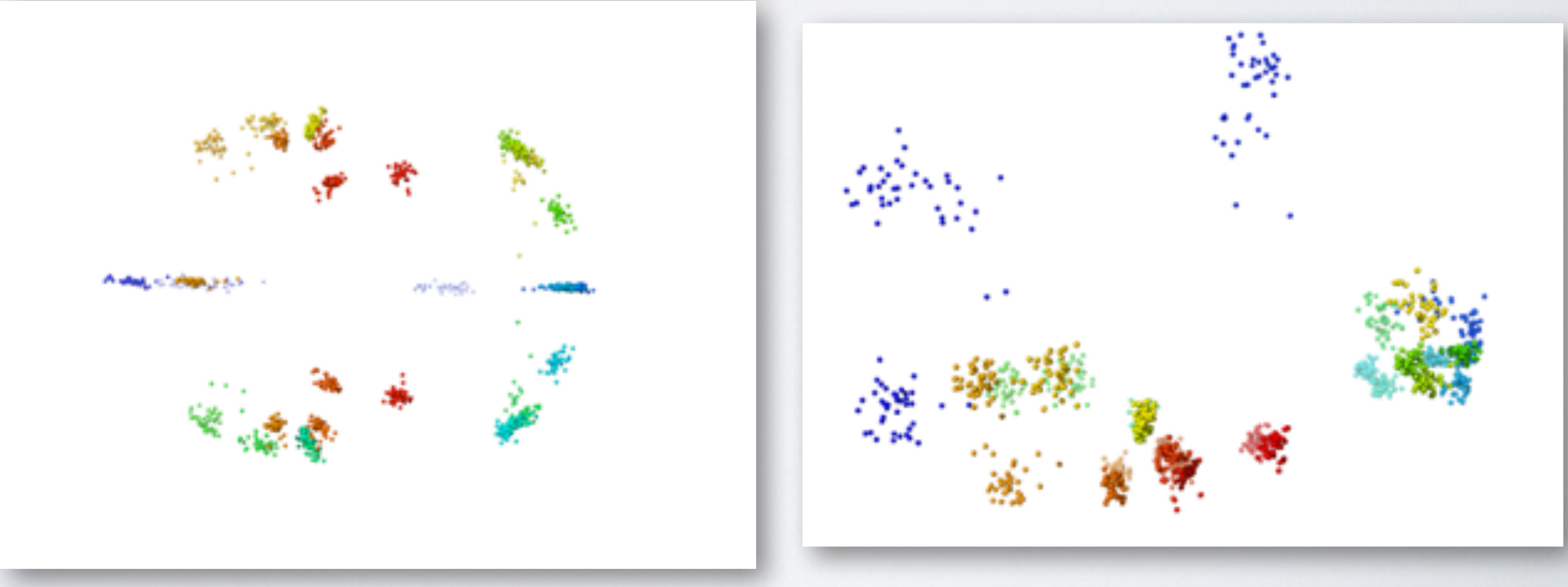
Bayes, scaling

Bayesian superpositions, with and without scaling

Nonisotropic,
no scaling



Nonisotropic,
scaling



The People


Catherine Ackley


Kristine Mackin

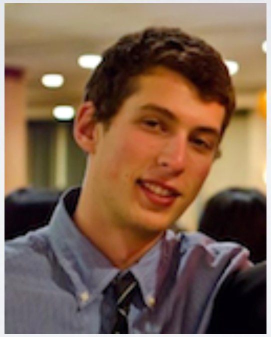

Phil Steindel


Mackenzie Gallegos


Rick Roy


Brian Beckett


Michelle Fry


Joe Jacobowitz


Marion Peyrega

Collaborators

Kanti Mardia, Leeds University

Thomas Hamelryck, University of Copenhagen

Funding

NIH National Institutes of Health

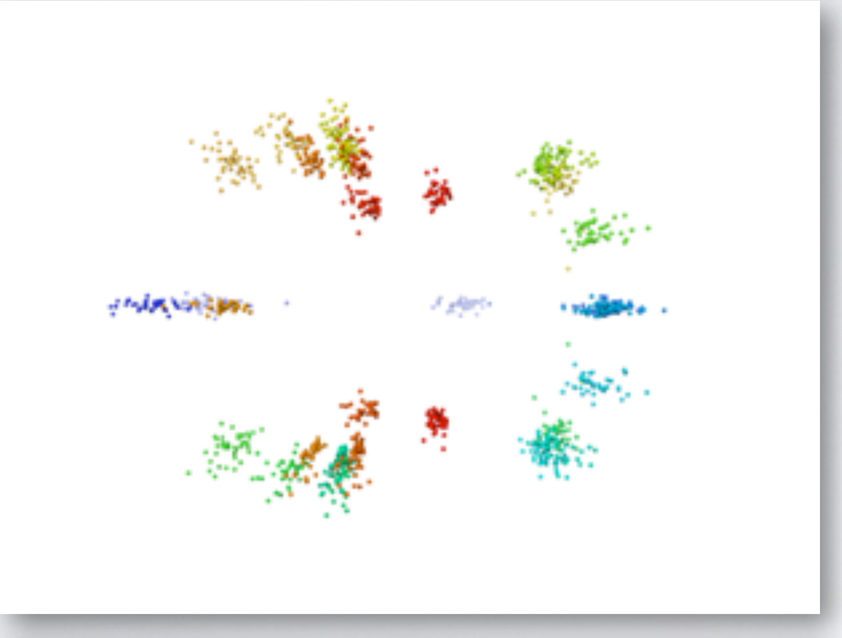
R01-GM094468

R01-GM096053

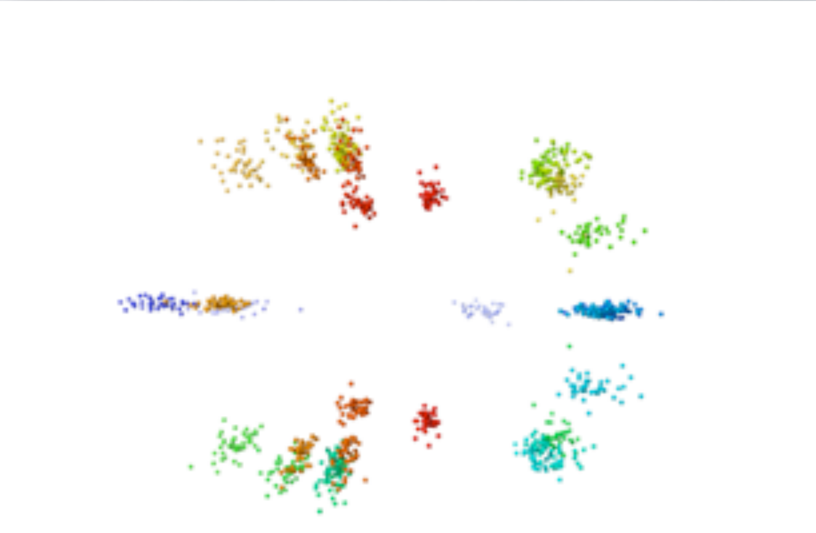
The Camille & Henry Dreyfus Foundation, Inc.

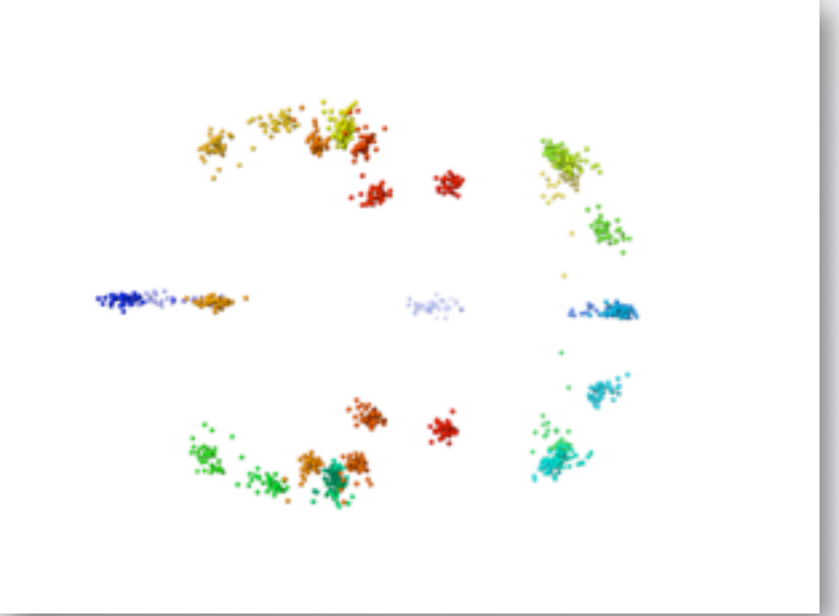
Bayesian superpositions, with and without scaling

Isotropic, no scaling

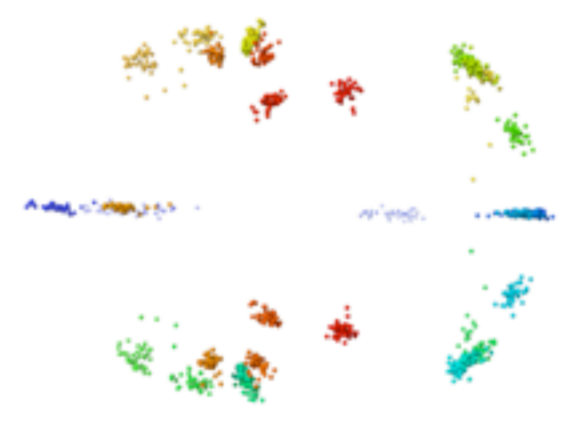


Nonisotropic, no scaling



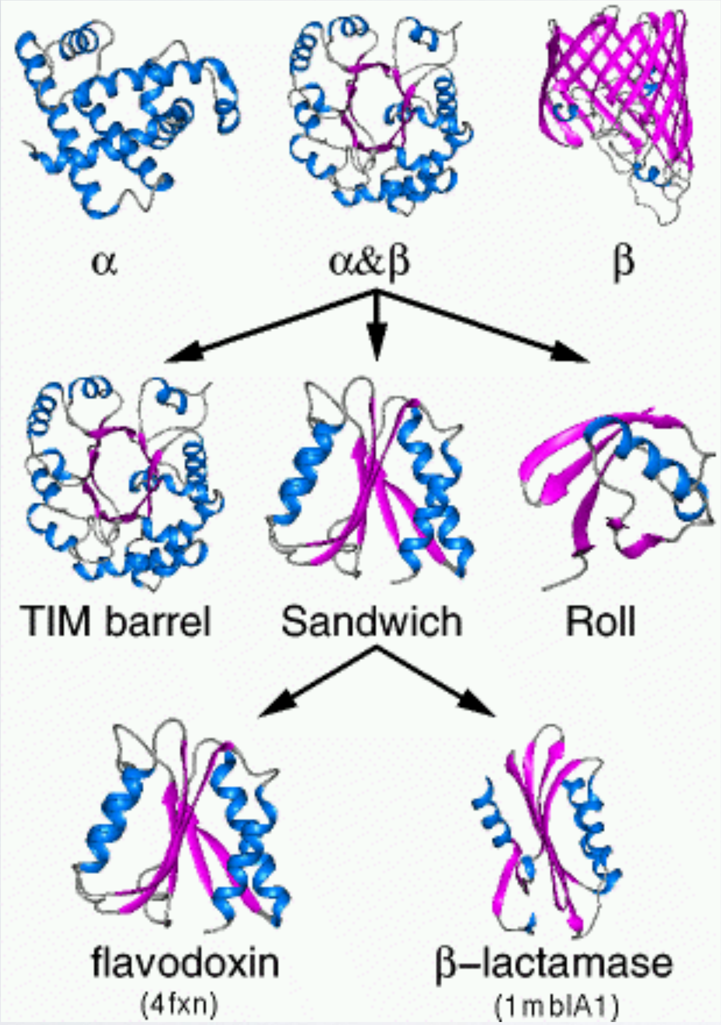


Isotropic, scaling



Nonisotropic, scaling

Protein Folds and Structural Taxonomy

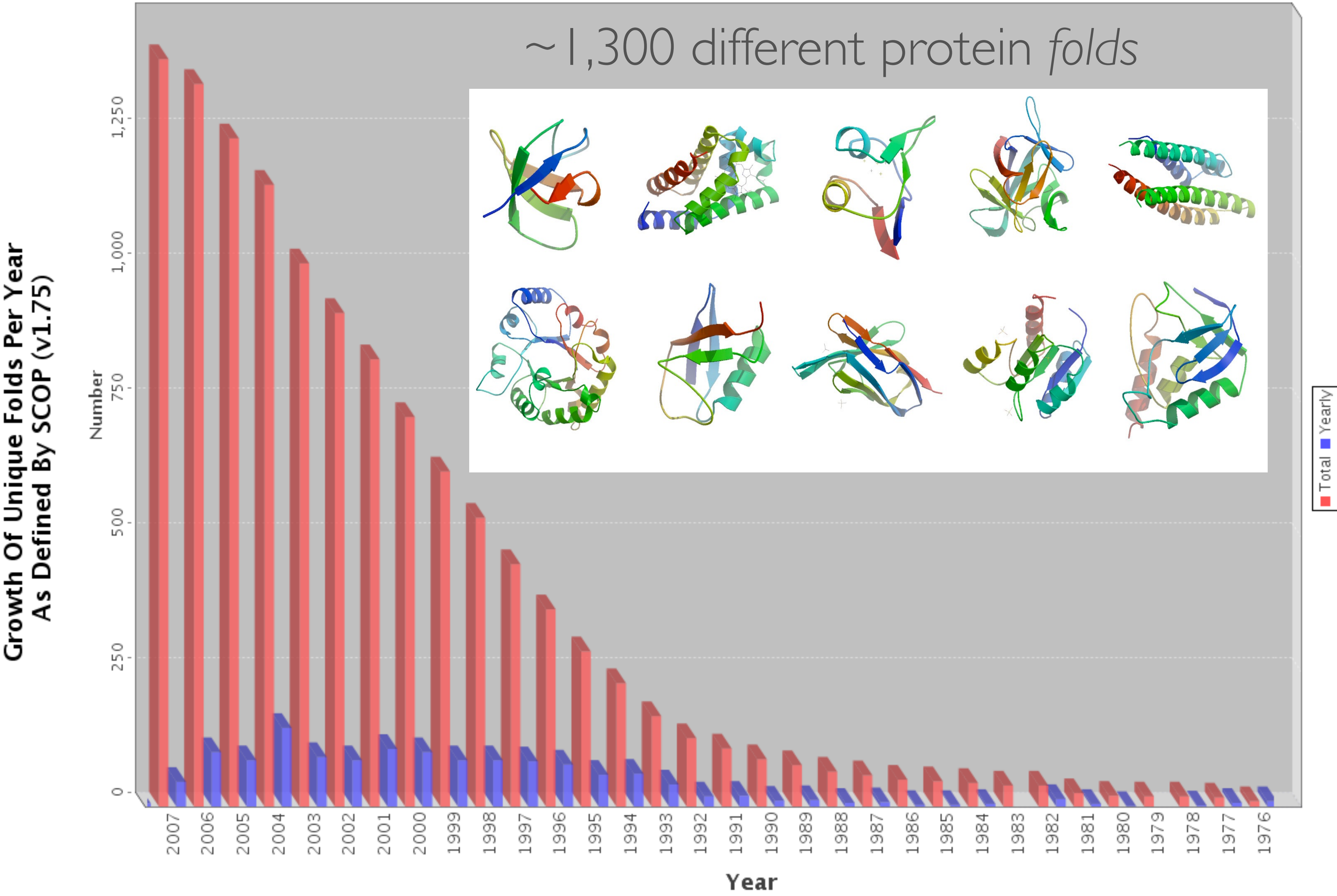


SCOP: Structural Classification Of Proteins

http://scop.berkeley.edu/

FOLD ≡ specific arrangement of secondary structure elements & specific connectivity

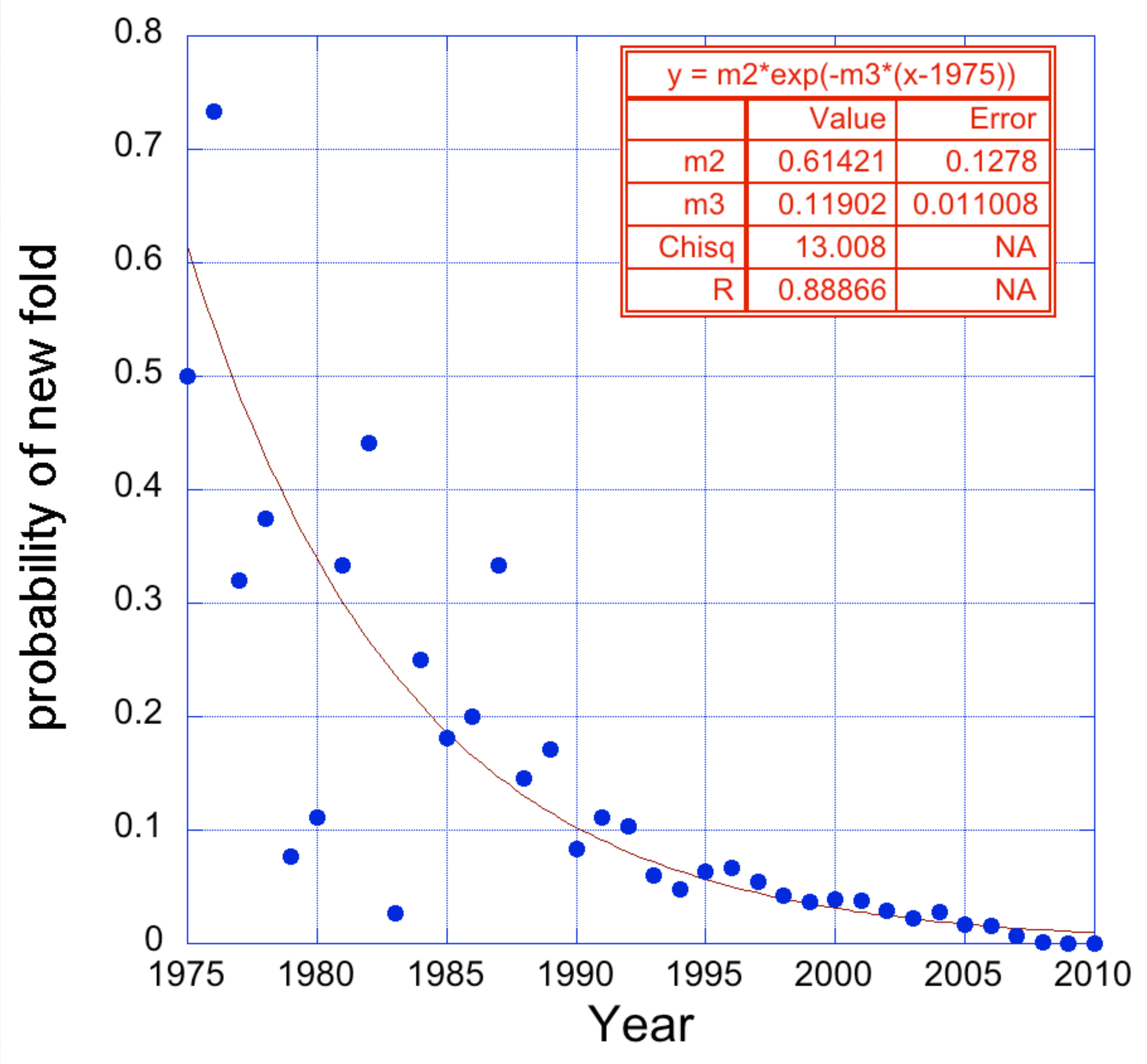
Fold growth in the PDB



~1,300 different protein *folds*

Frequency of novel folds is decreasing

Today’s chance of new fold: 10^{-3} to 10^{-4}



y = m2*exp(-m3*(x-1975))

	Value	Error
m2	0.61421	0.1278
m3	0.11902	0.011008
Chisq	13.008	NA
R	0.88866	NA

Gibbs/MCMC sampling for isotropic scaling

Initialize chain with ordinary LS superposition

$$\phi \sim \mathrm{IG}\left(S, \frac{3nk}{2}\right)$$

$$\mathbf{M} \sim \mathrm{N}_{k,d}(\mathbf{B}, \boldsymbol{\Omega}, \mathbf{I}_d)$$

$$\boldsymbol{t}_i \sim \mathrm{N}_{d,1}(\tau_i, \theta_i)$$

$$\mathbf{R}_i \sim \mathrm{MF}(\mathbf{A}_i)$$

$$\beta_i \sim \mathrm{HNG}(\omega_i, \gamma_i, m)$$

This is the "Bayesian version" of Generalized Procrustes Analysis

Conditional distribution for isotropic covariance matrix: Inverse Gamma

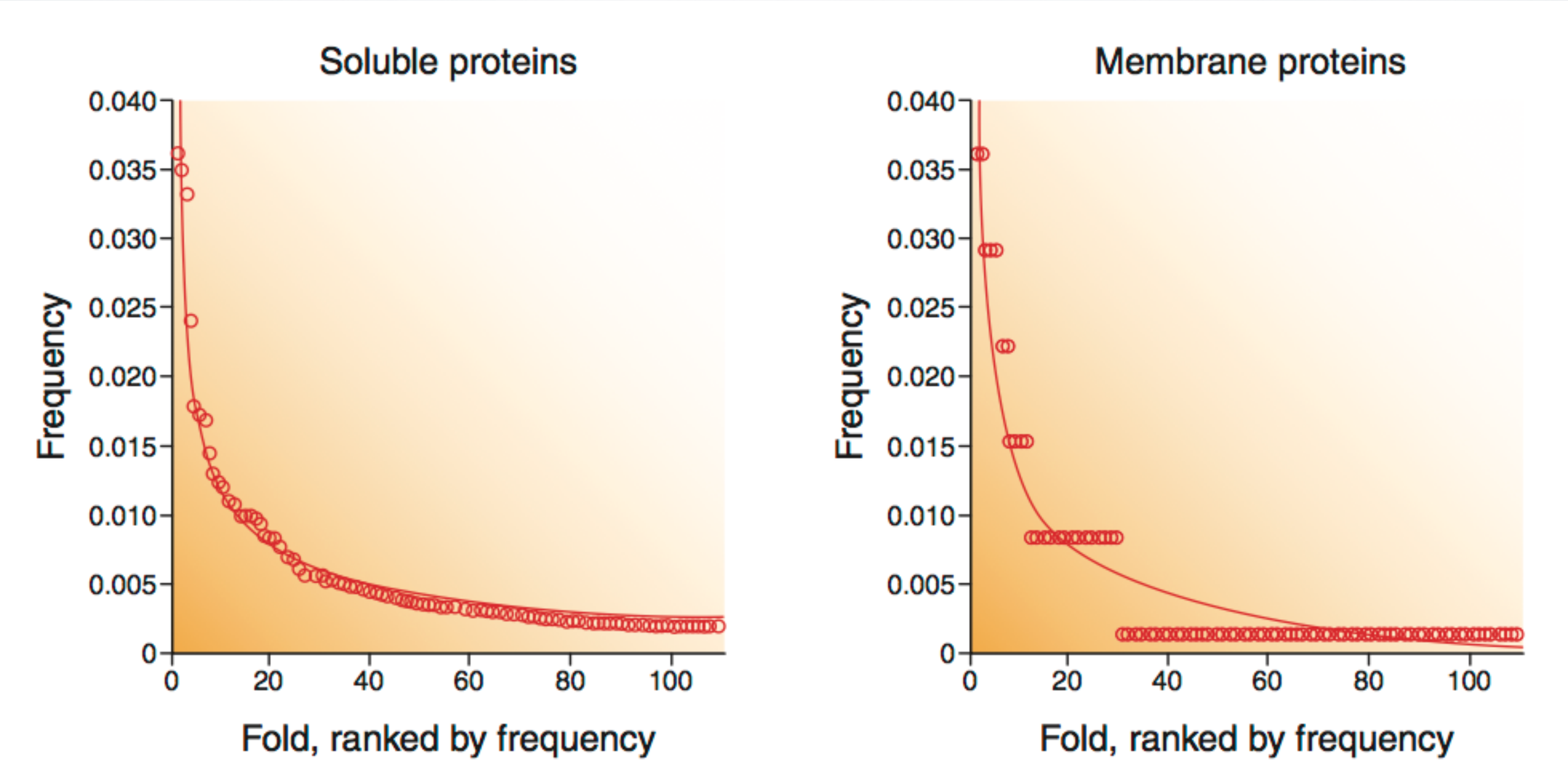
$$\boldsymbol{\Sigma}_{\text{iso}} = \phi\mathbf{I},$$

$$\boxed{\phi \sim \text{IG}\left(S, \frac{3nk}{2}\right)}$$

$$S = \frac{1}{2}\sum_{i}^{n} \text{tr}\left\{[\mathbf{Y}_i - \mathbf{M}]'[\mathbf{Y}_i - \mathbf{M}]\right\}$$

An inverse gamma distribution centered with scale parameter equal to the sum of squares

Membrane protein folds: Same story



Isotropic vs nonisotropic

No correlations, all variances equal = "isotropic"

$$\mathbf{\Sigma}_{\text{iso}} = \phi\mathbf{I},$$

Isotropic ML is equivalent to ordinary least squares

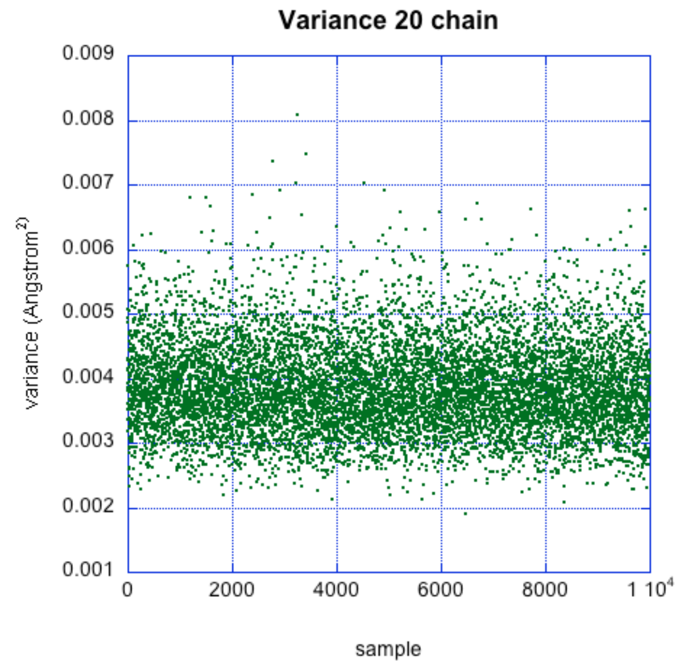
Unequal variances and/or correlations = "nonisotropic"

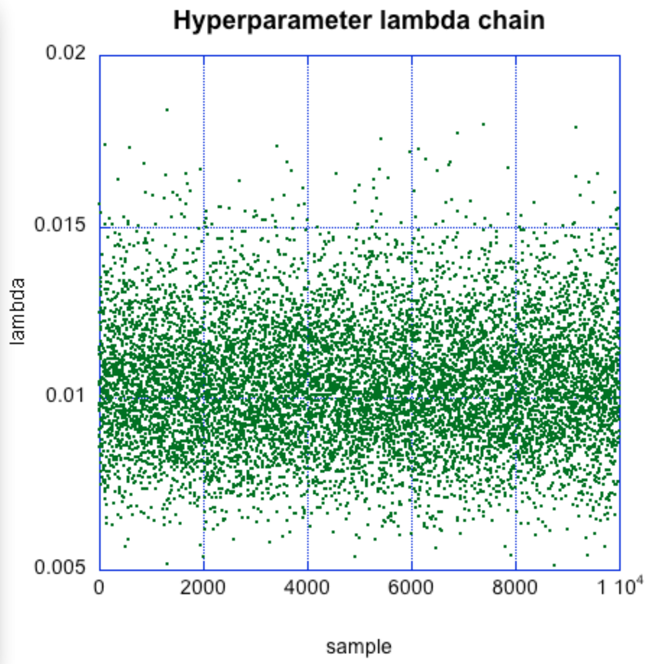
An arbitrary covariance matrix

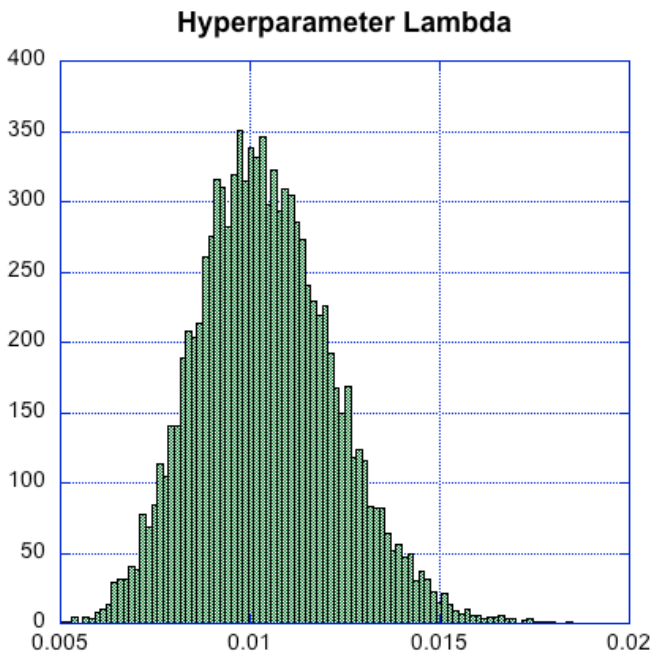
$$\hat{\mathbf{\Sigma}}_U = \frac{1}{3N}\sum_i^N (\check{\mathbf{X}}_i\mathbf{R}_i - \hat{\mathbf{M}})(\check{\mathbf{X}}_i\mathbf{R}_i - \hat{\mathbf{M}})'$$

Covariance matrix

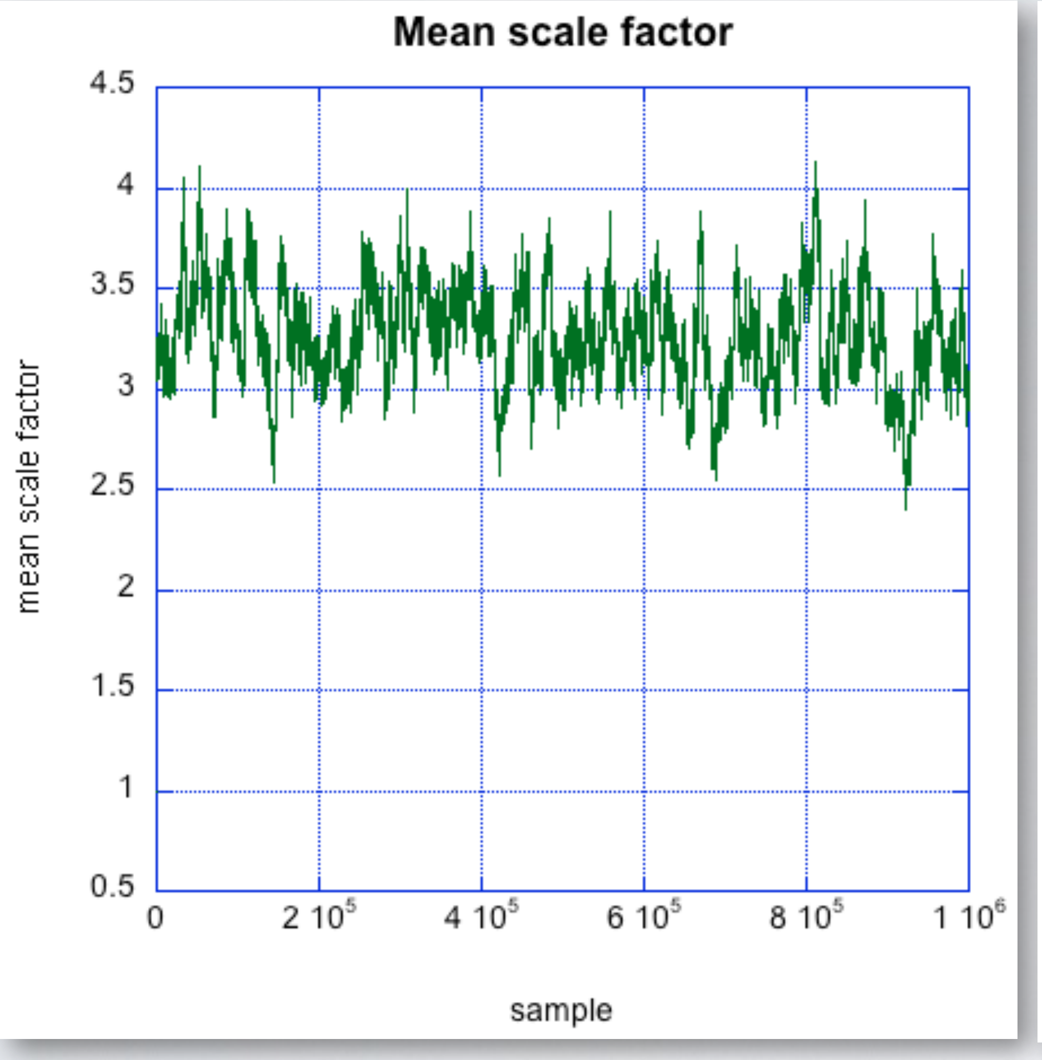
Posterior distribution of covariance matrix and its hyperparameter λ

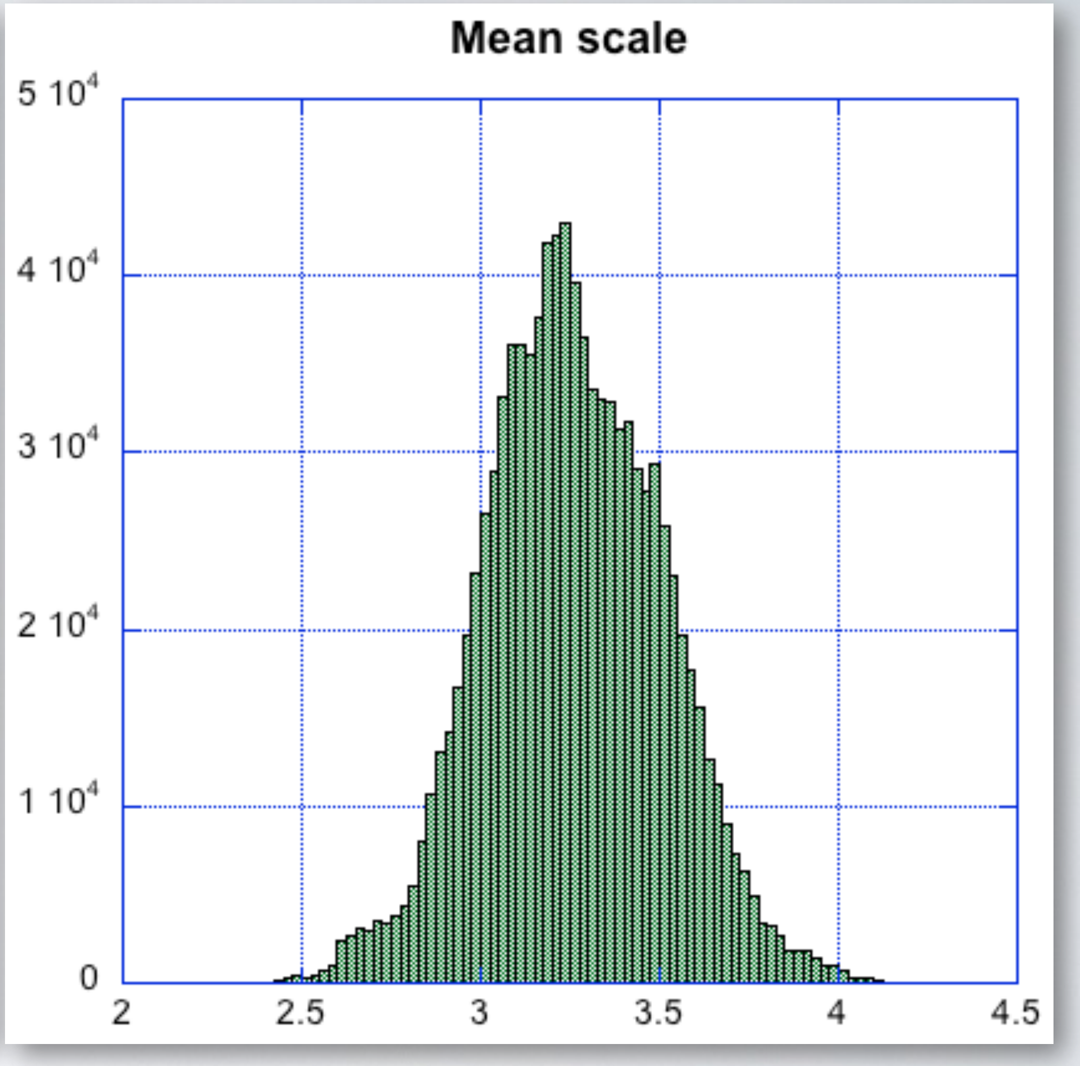






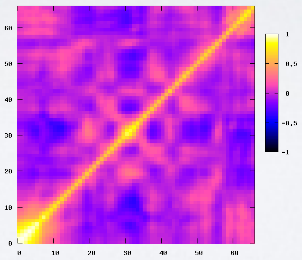
Posterior distribution of mean scale factor

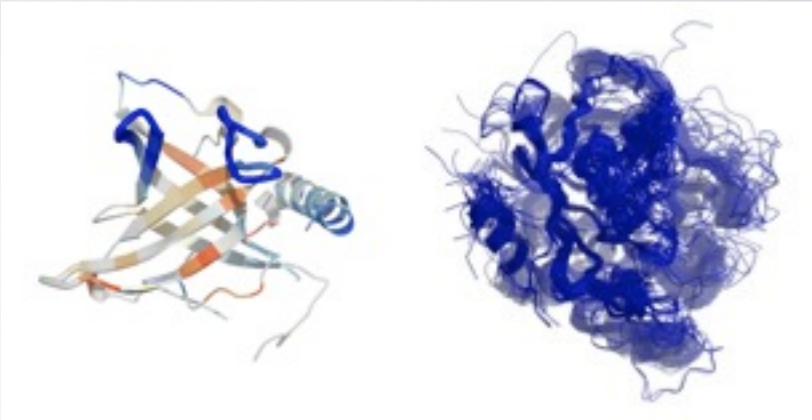




Future directions

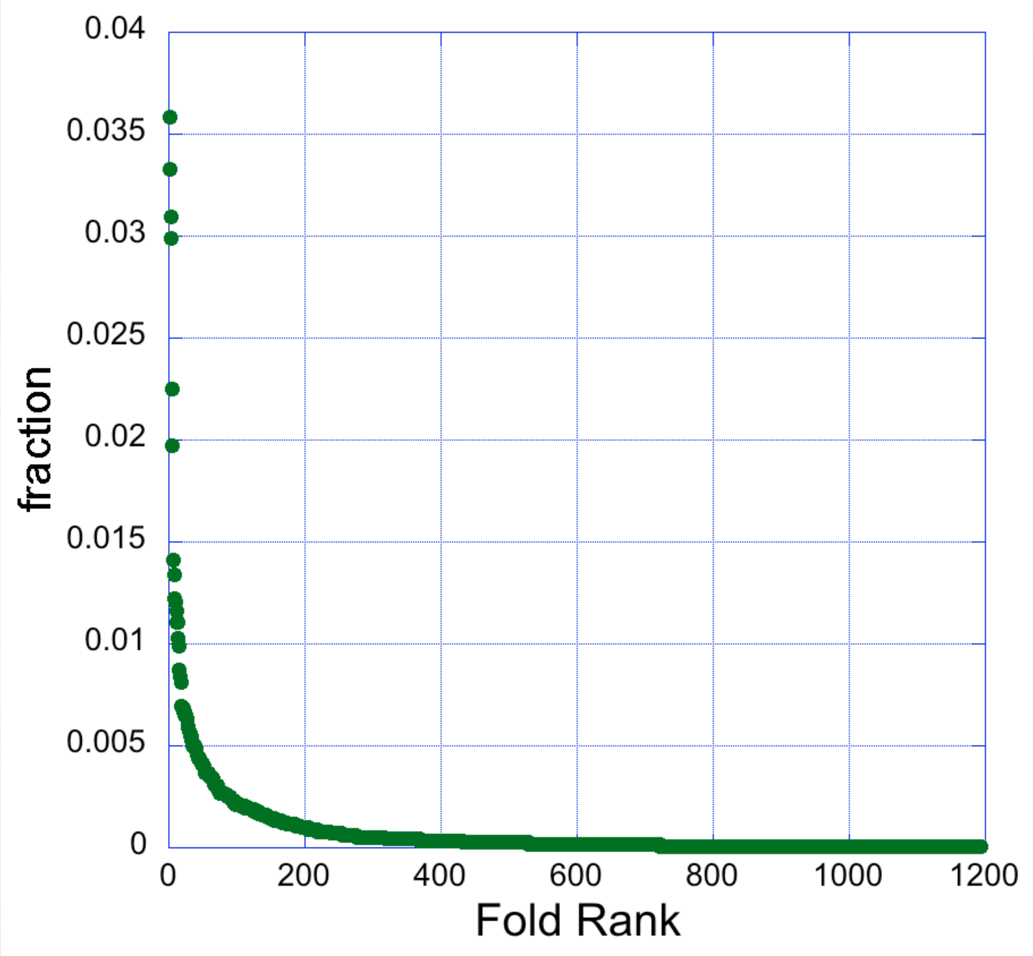
- Better sampling for halfnormal-gamma
- MCMCMC (heated chains)
- Unlabeled problems (match matrix)
- Incorporate sequence information
- Evolutionary models of structural change

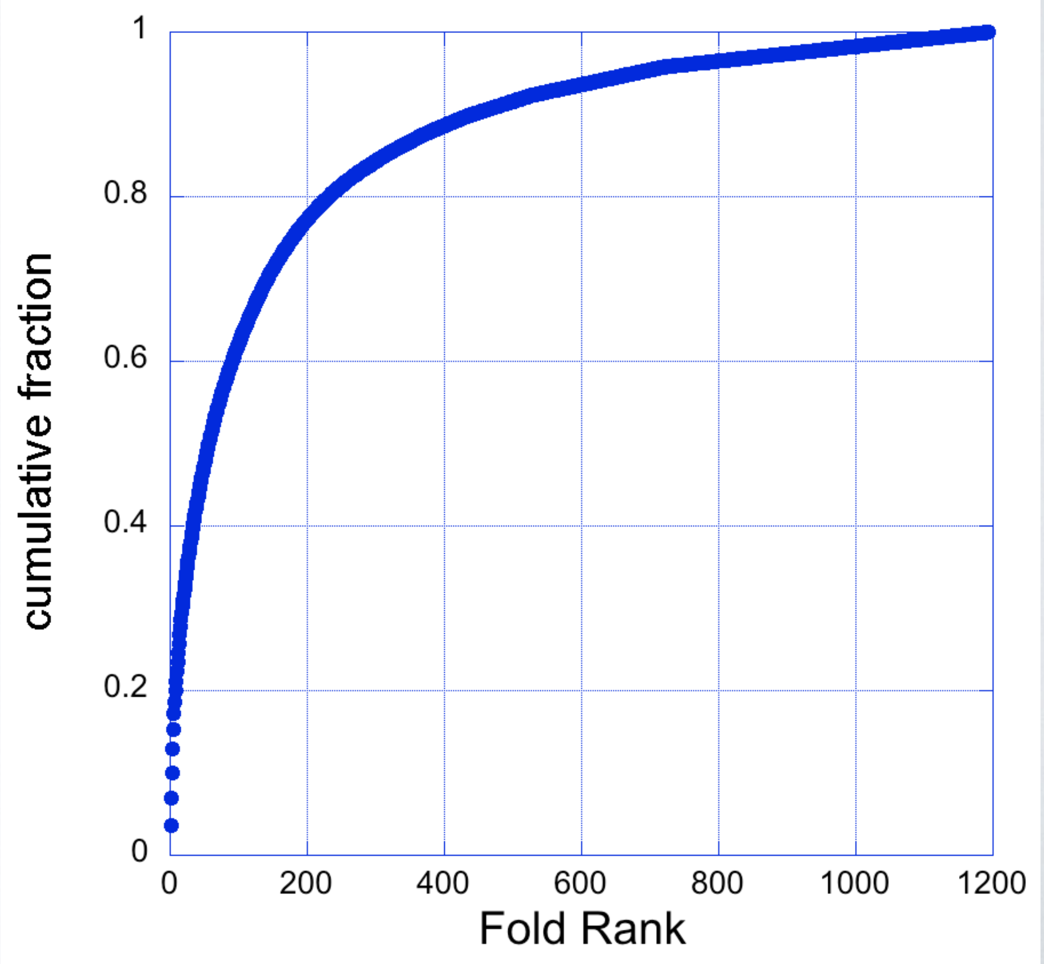




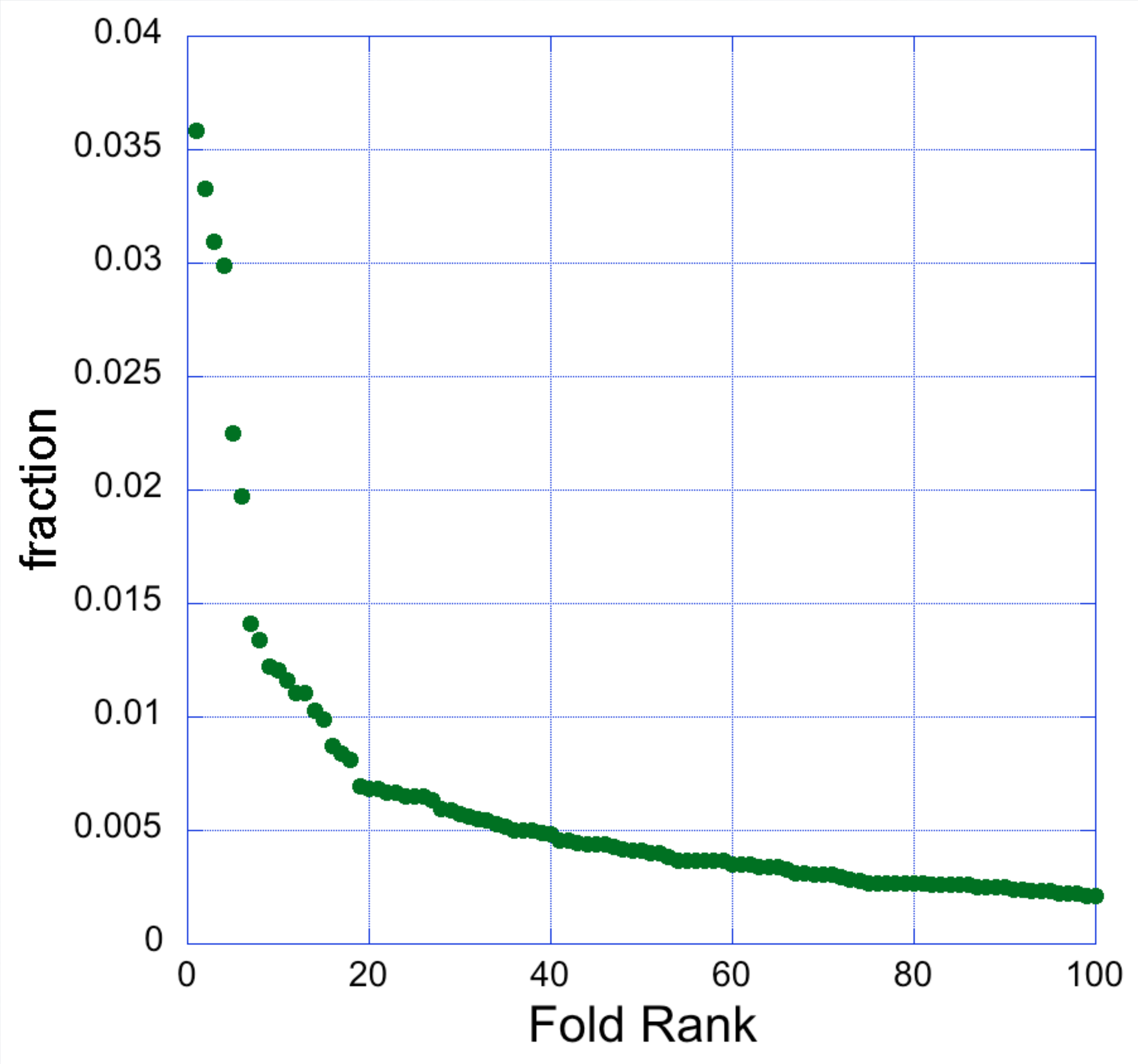
Some folds are much more populated

11,211 domains (no similarity to anything else)
1194 folds
first 12 folds are over 25% of domains
first 56 folds are over 50% of domains

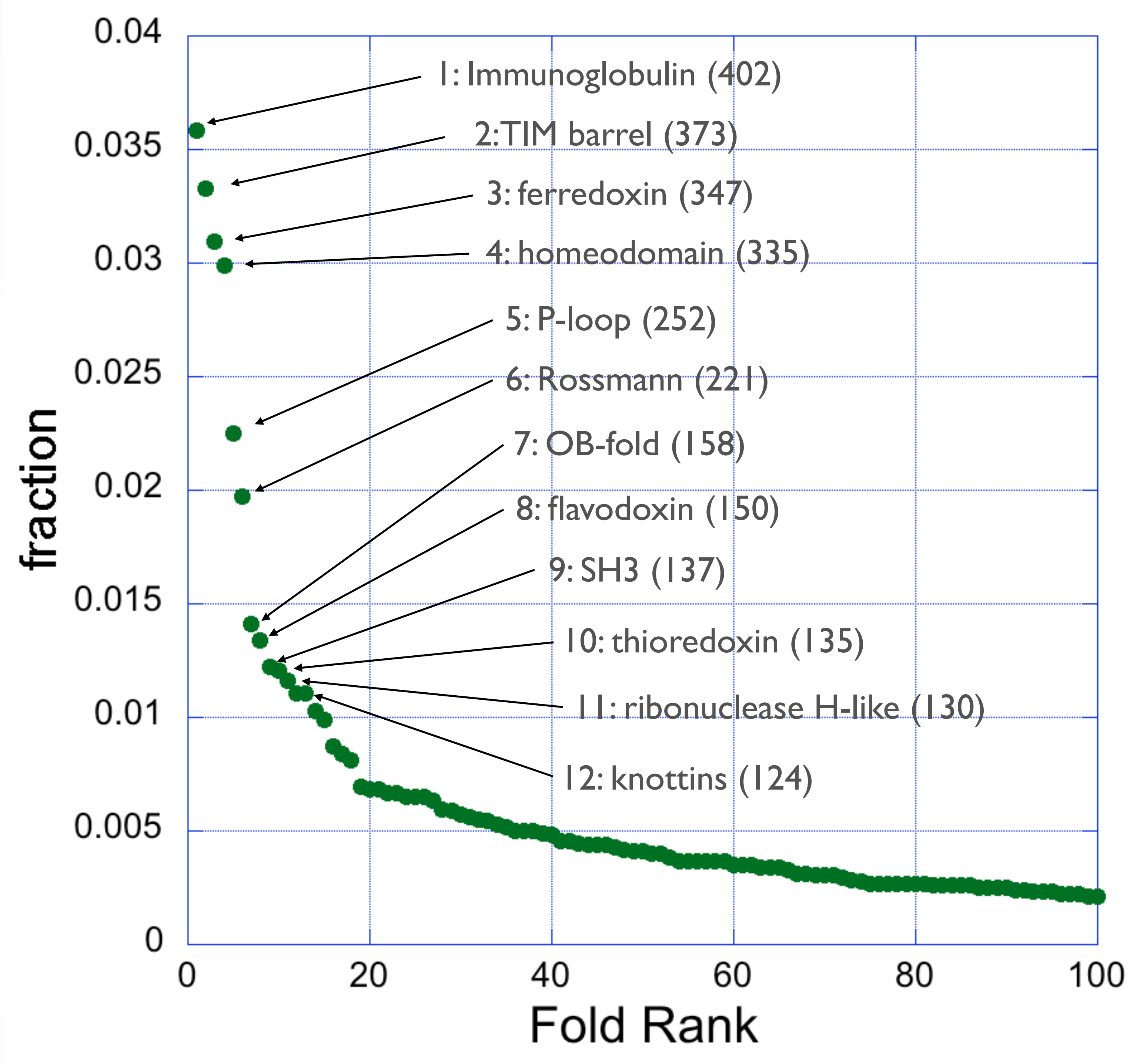




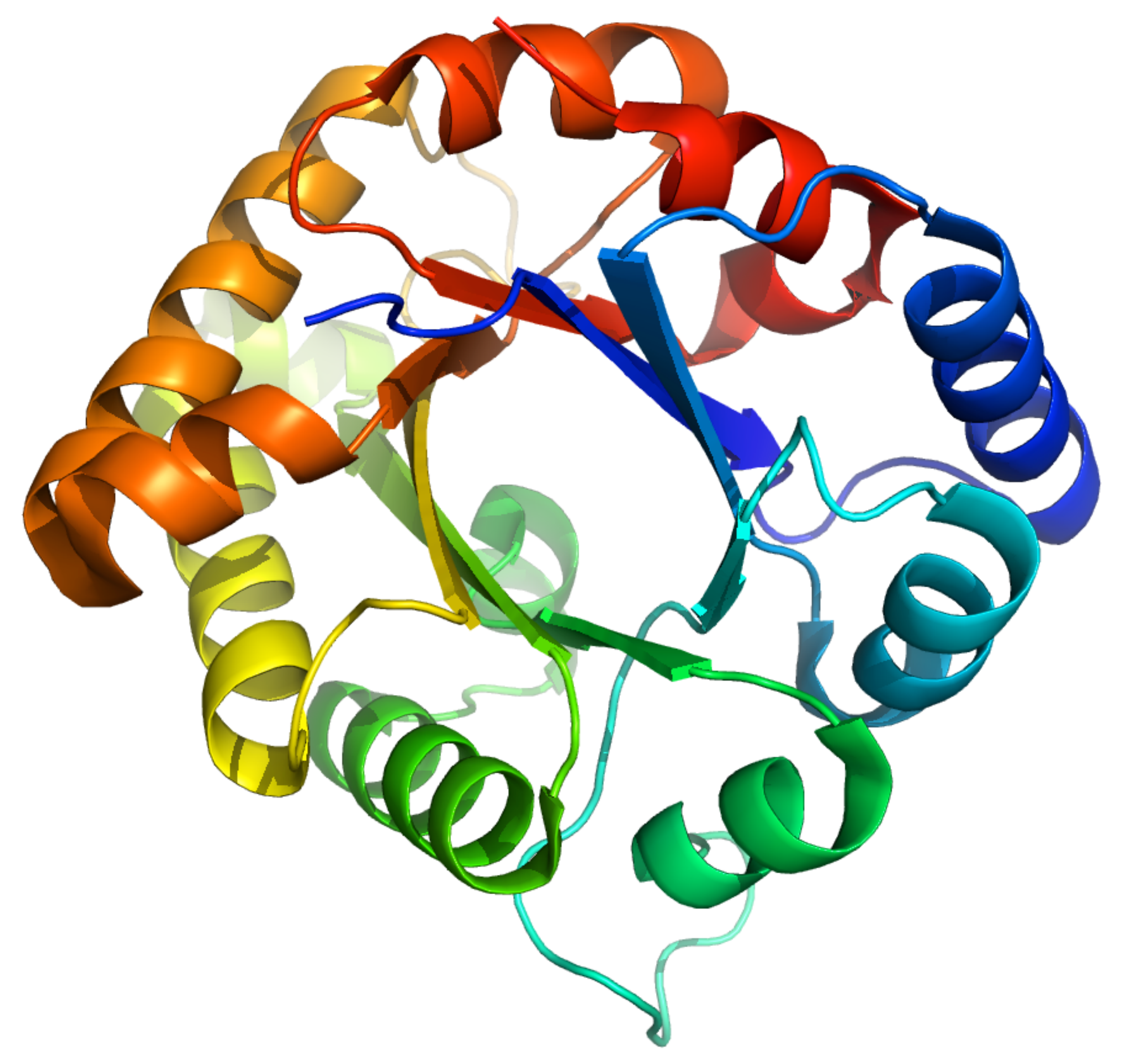
Superfolds


0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
fraction
0
20
40
60
80
100
Fold Rank

Superfolds


1: Immunoglobulin (402)
2:TIM barrel (373)
3: ferredoxin (347)
4: homeodomain (335)
5: P-loop (252)
6: Rossmann (221)
7: OB-fold (158)
8: flavodoxin (150)
9: SH3 (137)
10: thioredoxin (135)
11: ribonuclease H-like (130)
12: knottins (124)
fraction
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
0
20
40
60
80
100
Fold Rank

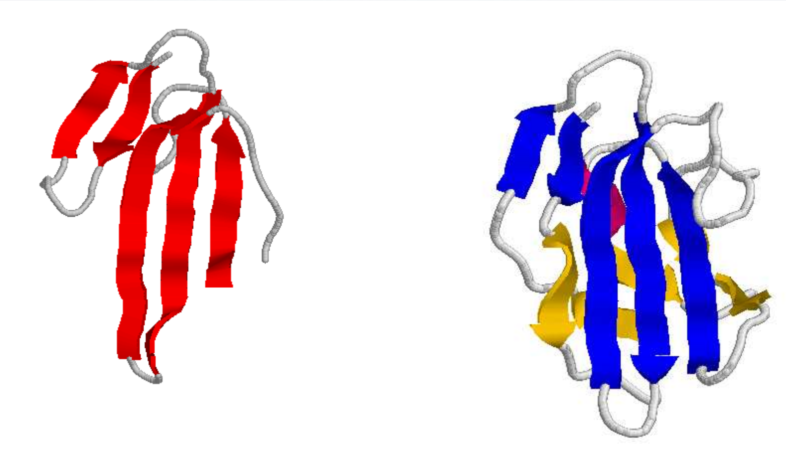
The TIM barrel superfold



- eight α-β repeats
- right handed connectivity

- 1992 different TIM domains
- 373 with $E > 0.01$
- 158 with $E > 10$ (!!)

Bayes: LS vs ML vs Bayes scaling, isotropic



EM for smallest variance from an inverse gamma distribution

$$\hat{\alpha} = \frac{K}{2\,\mathrm{tr}\left(\mathbf{\Sigma}^{-1}\right)}$$

expected inverse of the smallest variances

$$\hat{\alpha} = \frac{K}{2\left(\sum_{i}^{K-1}\left(\lambda_i^{-1}\right) + \mathrm{E}\left(\lambda_K^{-1}\right)\right)}$$

$$\mathrm{E}\left(\lambda_K^{-1}\right) = \mathrm{E}\left(\lambda_K^{-1} | \lambda_K < c = \lambda_{K-1}\right) = \frac{\Gamma\left(\frac{3}{2}, \frac{\alpha}{c}\right)}{\alpha\,\Gamma\left(\frac{1}{2}, \frac{\alpha}{c}\right)}$$

- Must be solved simultaneously

Likelihood function including scaling parameters

$$\mathbf{E}_i \sim \mathrm{N}_{k,d}(\mathbf{0}, \boldsymbol{\Sigma}, \mathbf{I})$$

$$\mathbf{E}_i \sim \mathrm{N}_{k,d}(\mathbf{0}, \boldsymbol{\Sigma}, \mathbf{I})$$

$$\mathrm{p}\left(\mathbf{X} | \boldsymbol{\Sigma}, \mathbf{M}, \mathbf{R}, \boldsymbol{t}, \beta\right) = C \exp\left(-\frac{1}{2} \sum_{i}^{n} \mathrm{tr}\left\{ [\mathbf{Y}_i - \mathbf{M}]' \boldsymbol{\Sigma}^{-1} [\mathbf{Y}_i - \mathbf{M}] \right\} \right)$$

$$\mathbf{Y}_i = (\beta_i \mathbf{X}_i + \mathbf{1}_k \boldsymbol{t}_i') \mathbf{R}_i$$

$$C = (2\pi)^{-\frac{kdn}{2}} \left(\prod_{i}^{n} \beta_i^{kd} \right) |\boldsymbol{\Sigma}|^{-\frac{dn}{2}}$$

$$\mathrm{p}\left(\boldsymbol{\Sigma}\right) \propto \mathrm{p}\left(\boldsymbol{\Sigma} | \boldsymbol{\Psi}, n\right) \mathrm{p}\left(\boldsymbol{\Psi}\right).$$

$$\mathrm{p}\left(\boldsymbol{\Sigma}, \mathbf{M}, \mathbf{R}, \boldsymbol{t}, \beta | \mathbf{X}\right) \propto \mathrm{p}\left(\mathbf{X} | \boldsymbol{\Sigma}, \mathbf{M}, \mathbf{R}, \boldsymbol{t}, \beta\right) \mathrm{p}\left(\beta\right) \mathrm{p}\left(\mathbf{M}\right) \mathrm{p}\left(\mathbf{R}\right) \mathrm{p}\left(\boldsymbol{t}\right) \mathrm{p}\left(\boldsymbol{\Sigma} | \boldsymbol{\Psi}, n\right) \mathrm{p}\left(\boldsymbol{\Psi}\right)$$

Non-isotropic, diagonal, covariance matrix

Assume a conjugate hierarchical prior for covariance matrix, a diagonal, isotropic inverse Wishart distribution

$$\mathrm{p}\left(\boldsymbol{\Sigma}|\boldsymbol{\Psi}=\phi\mathbf{I},n,K\right)=\frac{\phi^{\frac{nK}{2}}}{2^{\frac{nK}{2}}\left|\boldsymbol{\Sigma}\right|^{\left(\frac{n}{2}+1\right)}\Gamma\left(\frac{n}{2}\right)^{K}}\exp\left\{-\frac{\phi}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\right)\right\}$$

$$\mathrm{p}\left(\boldsymbol{\Sigma}|\mathbf{X},\mathbf{M},\mathbf{R},\boldsymbol{t},\phi\right)=\frac{\left|\mathbf{A}\right|^{\frac{k}{2}}}{2^{\frac{kK}{2}}\left|\boldsymbol{\Sigma}\right|^{\left(\frac{k}{2}+1\right)}\Gamma\left(\frac{k}{2}\right)^{K}}\exp\left\{-\frac{1}{2}\operatorname{tr}\left(\mathbf{A}\boldsymbol{\Sigma}^{-1}\right)\right\}$$

$$\mathbf{A}=\mathbf{S}+\phi\mathbf{I}$$

$$k=3N+n$$

Must assume a proper prior for the hyper-parameter ϕ, here a conjugate gamma distribution

$$\mathrm{p}\left(\phi|\alpha,m\right)\propto\phi^{\frac{m-2}{2}}\exp\left\{-\frac{\phi}{2\alpha}\right\}$$

$$\mathrm{p}\left(\phi|\mathbf{X},\boldsymbol{\Sigma},\mathbf{M},\mathbf{R},\boldsymbol{t},n\right)\propto\phi^{\frac{nK+m-2}{2}}\exp\left\{-\frac{\phi}{2}\left[\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\right)+\frac{1}{\alpha}\right]\right\}$$

Bayesian MAP superposition solutions

$$\hat{\boldsymbol{t}}_i = -\frac{\mathbf{1}'_K \boldsymbol{\Sigma}^{-1} \mathbf{X}_i}{\mathbf{1}'_K \boldsymbol{\Sigma}^{-1} \mathbf{1}_K}$$

Translations

$$\check{\mathbf{X}}_i = \mathbf{X}_i + \mathbf{1}_K \hat{\boldsymbol{t}}_i$$

$$\hat{\mathbf{M}}' \hat{\boldsymbol{\Sigma}}^{-1} \check{\mathbf{X}}_i = \mathbf{U} \boldsymbol{\Lambda} \mathbf{V}'$$

$$\hat{\mathbf{R}}_i = \mathbf{V} \mathbf{P} \mathbf{U}'$$

Rotations - found with Singular Value Decomposition

$$\hat{\mathbf{M}} = \frac{1}{N} \sum_{i}^{N} \check{\mathbf{X}}_i \mathbf{R}_i$$

Mean structure

$$\hat{\phi} = \frac{\sum_{i}^{N} \mathrm{tr}\left\{[\mathbf{Y}_i - \mathbf{M}]'[\mathbf{Y}_i - \mathbf{M}]\right\}}{3NK + 2}$$

Isotropic covariance matrix

$$\hat{\boldsymbol{\Sigma}} = \left(\frac{3N}{3N + n + 2}\right)\left(\frac{\phi}{3N}\mathbf{I} + \boldsymbol{\Sigma}_U\right)$$

Nonisotropic covariance matrix

$$\hat{\phi} = \frac{nK - 2}{\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1} + \frac{1}{\alpha}\right)}$$

Conditional distributions: The translations

Assume uniform improper prior on t$_i$

$$\mathrm{p}\left(\boldsymbol{t}_i \middle| \mathbf{X}_i, \mathbf{M}, \boldsymbol{\Sigma}, \mathbf{R}_i\right) =$$

$$\mathrm{p}\left(\boldsymbol{t}_i \middle| \mathbf{X}_i, \boldsymbol{\Sigma}\right) = (2\pi\theta)^{-\frac{3}{2}} \exp\left(-\frac{1}{2\theta}\,\mathrm{tr}\left\{[\boldsymbol{t}_i - \mu_i]'[\boldsymbol{t}_i - \mu_i]\right\}\right)$$

$$\theta = \frac{1}{\mathbf{1}_K' \boldsymbol{\Sigma}^{-1} \mathbf{1}_K}$$

$$\mu_i = -\theta\left(\mathbf{1}_K' \boldsymbol{\Sigma}^{-1} \mathbf{X}_i\right)$$

$$\theta_{\mathrm{iso}} = \frac{\phi}{K} \qquad \boldsymbol{\Sigma}_{\mathrm{iso}} = \phi \mathbf{I}$$

$$\mu_{\mathrm{iso},i} = -\frac{\mathbf{1}_K' \mathbf{X}_i}{K}$$

A multivariate normal distribution centered on the the ML estimate (the weighted centroid)

Conditional distributions: The rotations

Assume uniform proper prior on R_i

$$\mathrm{p}\left(\mathbf{R}_i | \mathbf{X}_i, \mathbf{\Sigma}, \mathbf{M}\right) \propto \exp\left(-\frac{1}{2}\,\mathrm{tr}\left\{\mathbf{A}_i \mathbf{R}_i\right\}\right)$$

$$\mathbf{A}_i = \mathbf{M}' \mathbf{\Sigma}^{-1} \mathbf{X}_i$$

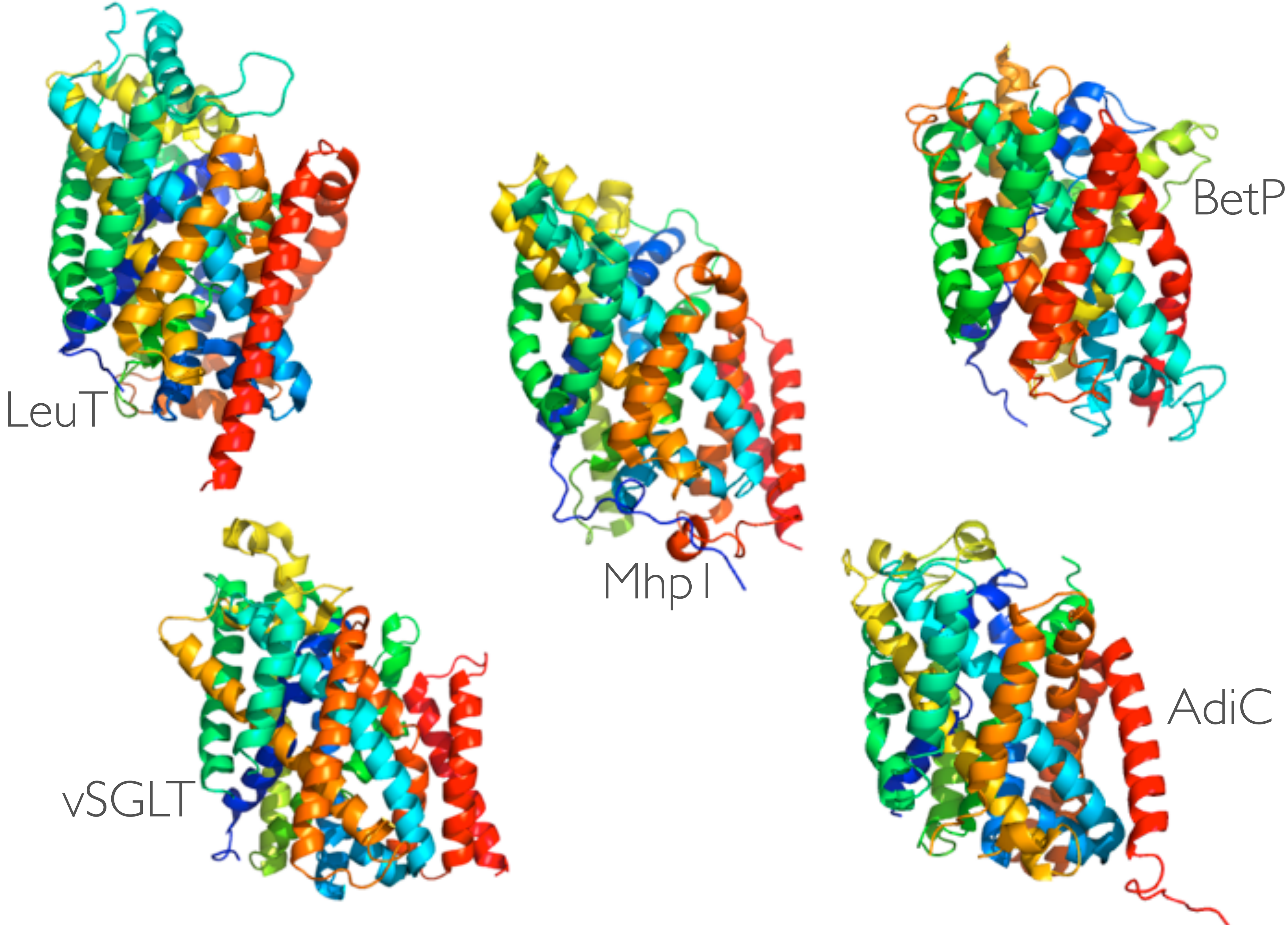
$$\mathbf{A}_{\mathrm{iso},i} = \frac{1}{\phi} \mathbf{M}' \mathbf{X}_i$$

A matrix Fisher-von Mises centered on the ML estimate

Can be sampled using hybrid Gibbs/Metropolis-Hastings algorithm of Green and Mardia

P. Green and K.V. Mardia (2006) "Bayesian alignment using hierarchical models, with applications in protein bioinformatics." Biometrika 93(2):235–254

Five membrane channel structures:
Same fold, no sequence similarity



Acknowledgments

Brandeis University
Chris Miller
Phillip Steindel

University of Colorado
Deborah Wuttke

Brandeis University

Posterior mode solution for inverse Wishart priors on the covariance matrix

$$\mathrm{p}\left(\boldsymbol{\Sigma}|\boldsymbol{\Psi},n,K\right)=\frac{\left(\frac{n}{2}\right)^{\frac{nK}{2}}\left|\boldsymbol{\Psi}\right|^{\frac{n}{2}}}{\left|\boldsymbol{\Sigma}\right|^{\left(\frac{n}{2}+1\right)}\Gamma\left(\frac{n}{2}\right)^{K}}e^{\left\{-\frac{n}{2}\operatorname{tr}\left(\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}\right)\right\}}$$

diagonal inverse Wishart distribution

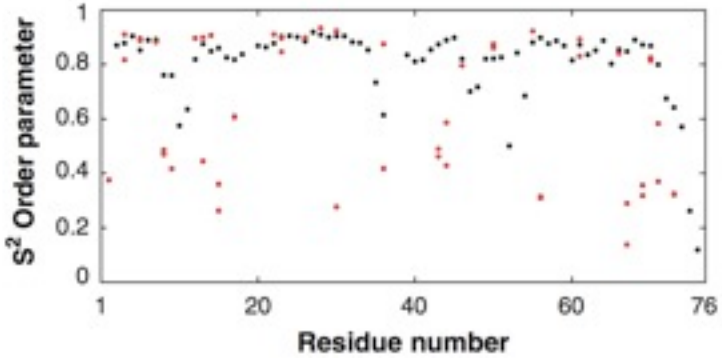
$$\hat{\boldsymbol{\Sigma}}=\left(\frac{3N}{3N+n+2}\right)\left(\frac{n}{3N}\boldsymbol{\Psi}+\boldsymbol{\Sigma}_U\right)$$

$$\hat{\boldsymbol{\Sigma}}_U=\frac{1}{3N}\sum_{i}^{N}(\check{\mathbf{X}}_i\mathbf{R}_i-\hat{\mathbf{M}})(\check{\mathbf{X}}_i\mathbf{R}_i-\hat{\mathbf{M}})'$$

MAP solution: Another “shrunken” covariance matrix:

Examples of Priors:

Crystallographic B-factors



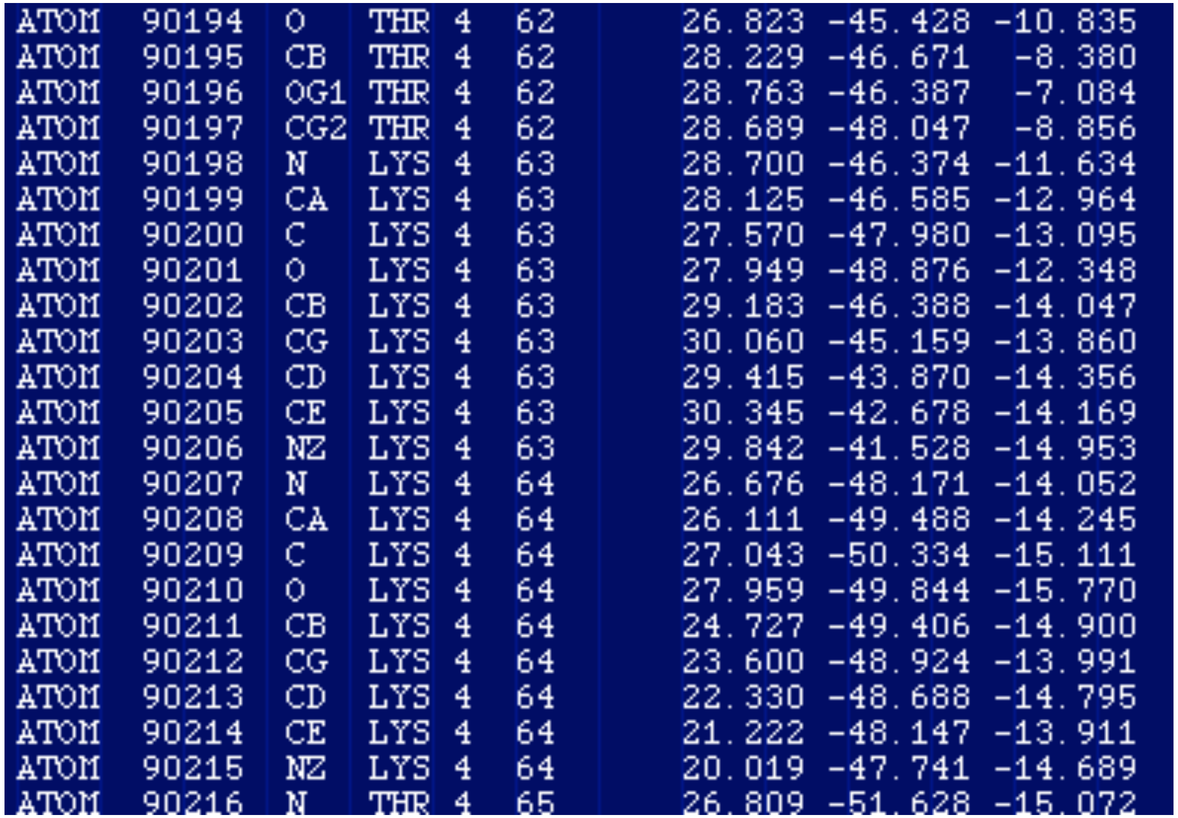
NMR order parameters

Macromolecular structures as matrices

A protein PDB file:

x y z

atoms



```
ATOM  90194  O    THR 4  62          26.823 -45.428 -10.835
ATOM  90195  CB   THR 4  62          28.229 -46.671  -8.380
ATOM  90196  OG1  THR 4  62          28.763 -46.387  -7.084
ATOM  90197  CG2  THR 4  62          28.689 -48.047  -8.856
ATOM  90198  N    LYS 4  63          28.700 -46.374 -11.634
ATOM  90199  CA   LYS 4  63          28.125 -46.585 -12.964
ATOM  90200  C    LYS 4  63          27.570 -47.980 -13.095
ATOM  90201  O    LYS 4  63          27.949 -48.876 -12.348
ATOM  90202  CB   LYS 4  63          29.183 -46.388 -14.047
ATOM  90203  CG   LYS 4  63          30.060 -45.159 -13.860
ATOM  90204  CD   LYS 4  63          29.415 -43.870 -14.356
ATOM  90205  CE   LYS 4  63          30.345 -42.678 -14.169
ATOM  90206  NZ   LYS 4  63          29.842 -41.528 -14.953
ATOM  90207  N    LYS 4  64          26.676 -48.171 -14.052
ATOM  90208  CA   LYS 4  64          26.111 -49.488 -14.245
ATOM  90209  C    LYS 4  64          27.043 -50.334 -15.111
ATOM  90210  O    LYS 4  64          27.959 -49.844 -15.770
ATOM  90211  CB   LYS 4  64          24.727 -49.406 -14.900
ATOM  90212  CG   LYS 4  64          23.600 -48.924 -13.991
ATOM  90213  CD   LYS 4  64          22.330 -48.688 -14.795
ATOM  90214  CE   LYS 4  64          21.222 -48.147 -13.911
ATOM  90215  NZ   LYS 4  64          20.019 -47.741 -14.689
ATOM  90216  N    THR 4  65          26.809 -51.628 -15.072
```

Structure = K x 3 matrix, K rows of atoms, 3 axes

Classic superposition method: Least-squares

Pairwise superposition

$$SS = \|\mathbf{XR} - \mathbf{Y}\|^2$$

$$SS = \mathrm{tr}\left([\mathbf{XR} - \mathbf{Y}]' \, [\mathbf{XR} - \mathbf{Y}]\right)$$

X is the structure to be superpositioned

R is a 3×3 rotation matrix (orthogonal)

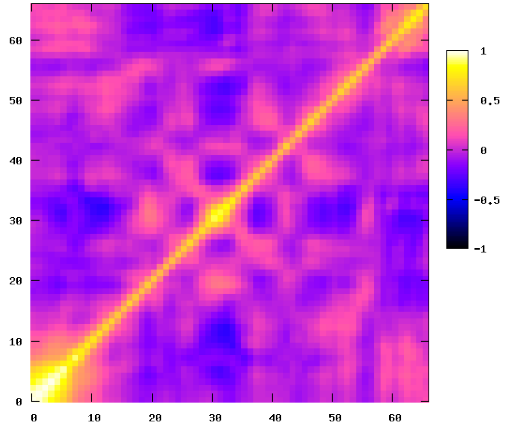
Y is the target structure

F. Boas (1905) "The horizontal plane of the skull and the general problem of the comparison of variable forms." *Science* 21:862
J. von Neumann (1937) "Some matrix-inequalities and metrization of matrix-spaces." *Tomsk Univ Rev* 1:286-300
B.F. Green (1952) "The orthogonal approximation of an oblique structure in factor analysis." *Psychometrika* 17:429
W. Kabsch (1978) "A discussion of the solution for the best rotation to relate two sets of vectors." *Acta Cryst* A34:827

Principal Components Analysis (PCA): Summarize the covariance matrix

- **PCA**: a method to extract the dominant patterns of correlation found in data.
- **Principal component (PC)**: each major mode of correlation
- Multiple PCs, ranked most important to least.
- Each PC is a vector that assigns a measure of correlation to each atom in a structure

Covariance/correlation matrices are information dense



Benefits of Bayes

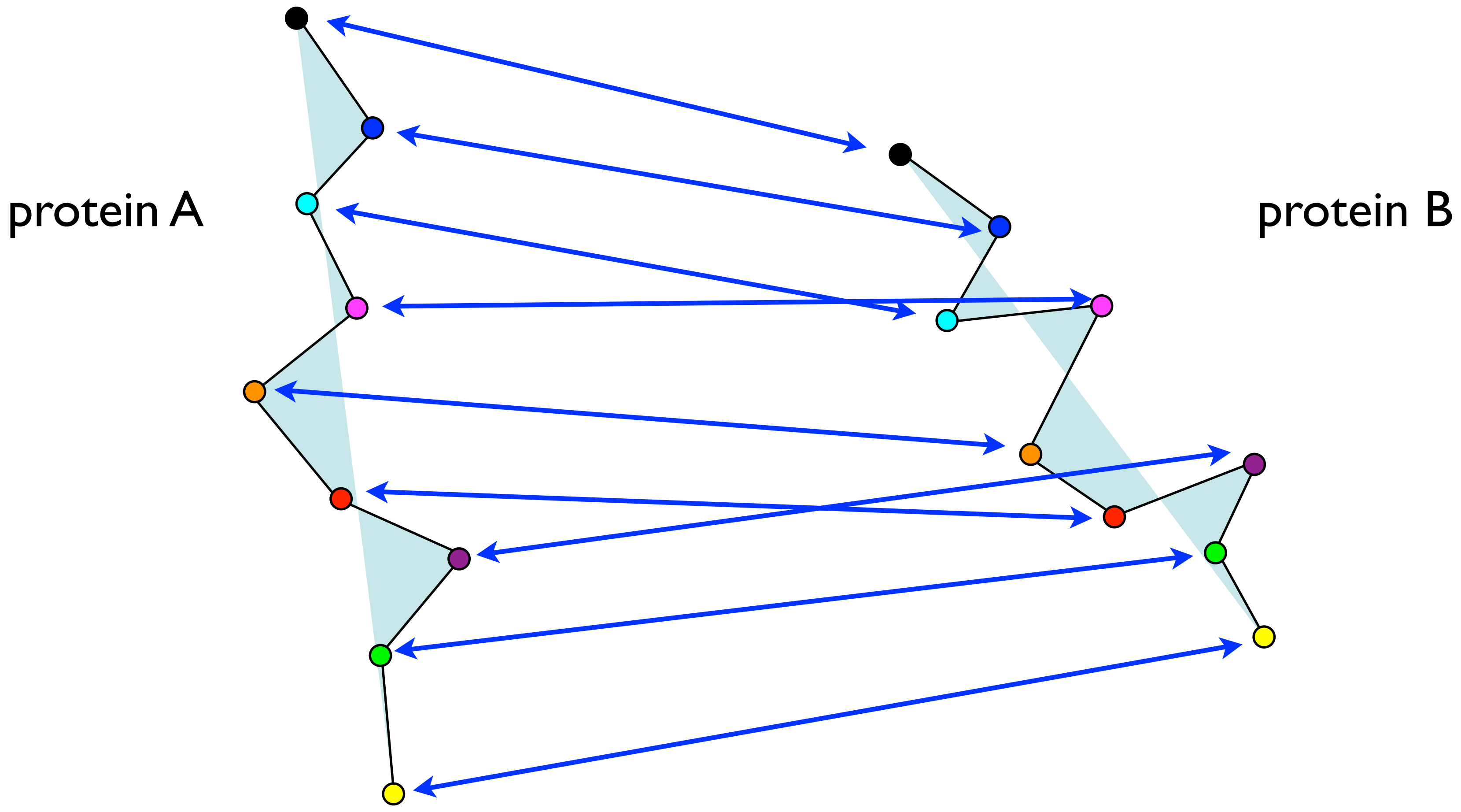
- Complete distribution of parameters
- Uses all relevant information, including prior
- Complex models, marginal distributions for "nuisance parameters"
- Solutions for underdetermined and "problematic" models
- Bayes solutions are exact for finite sample sizes

Bayes Theorem

$$\mathrm{p}(\theta|x, M) = \frac{\mathrm{p}(x|\theta, M)\,\mathrm{p}(\theta|M)}{\mathrm{p}(x|M)}$$

$$\mathrm{p}(\theta|x, M) = \int_{\phi} \mathrm{p}(\theta, \phi|x, M)\; d\phi$$

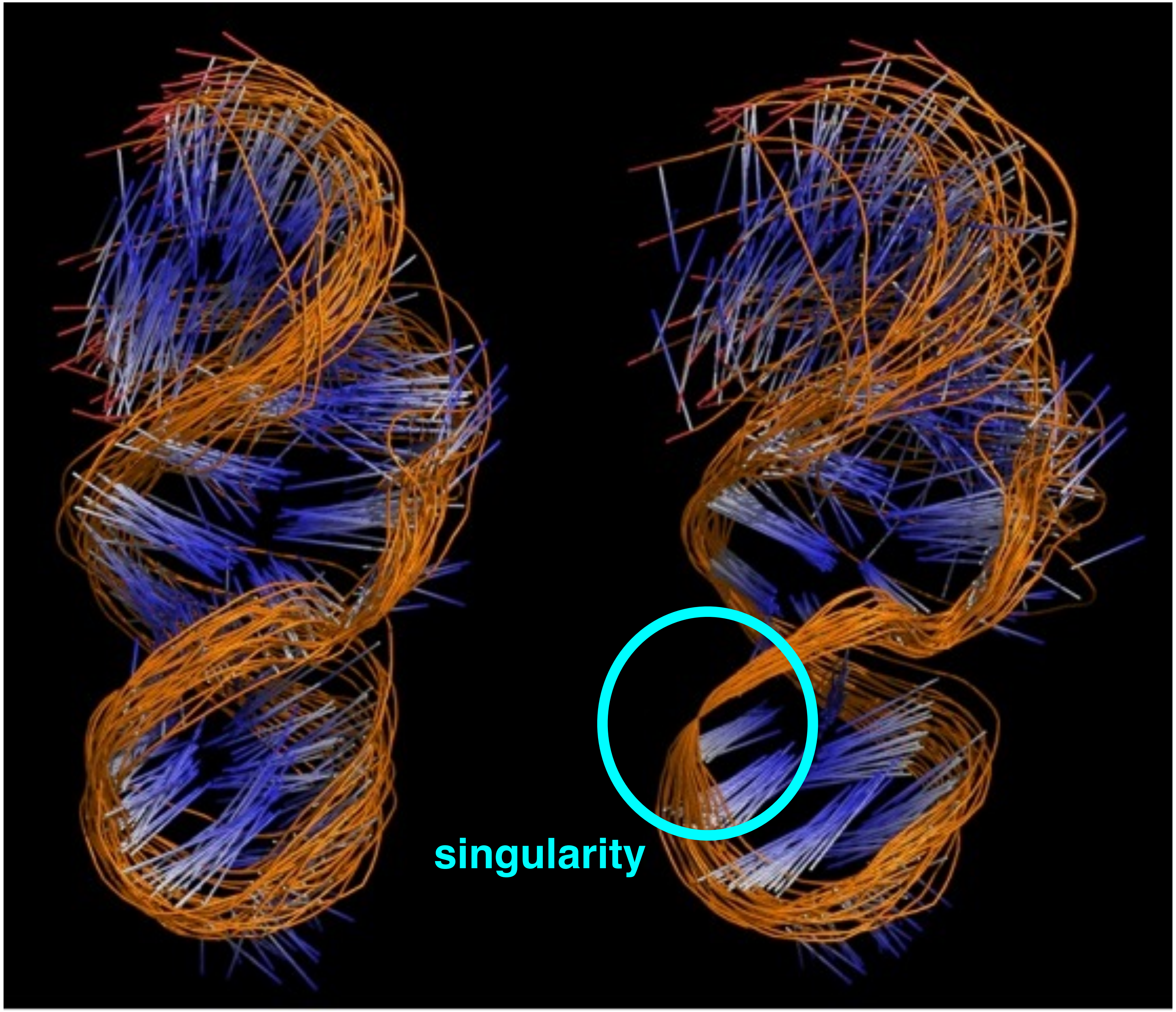
Classic superposition method: Least-squares



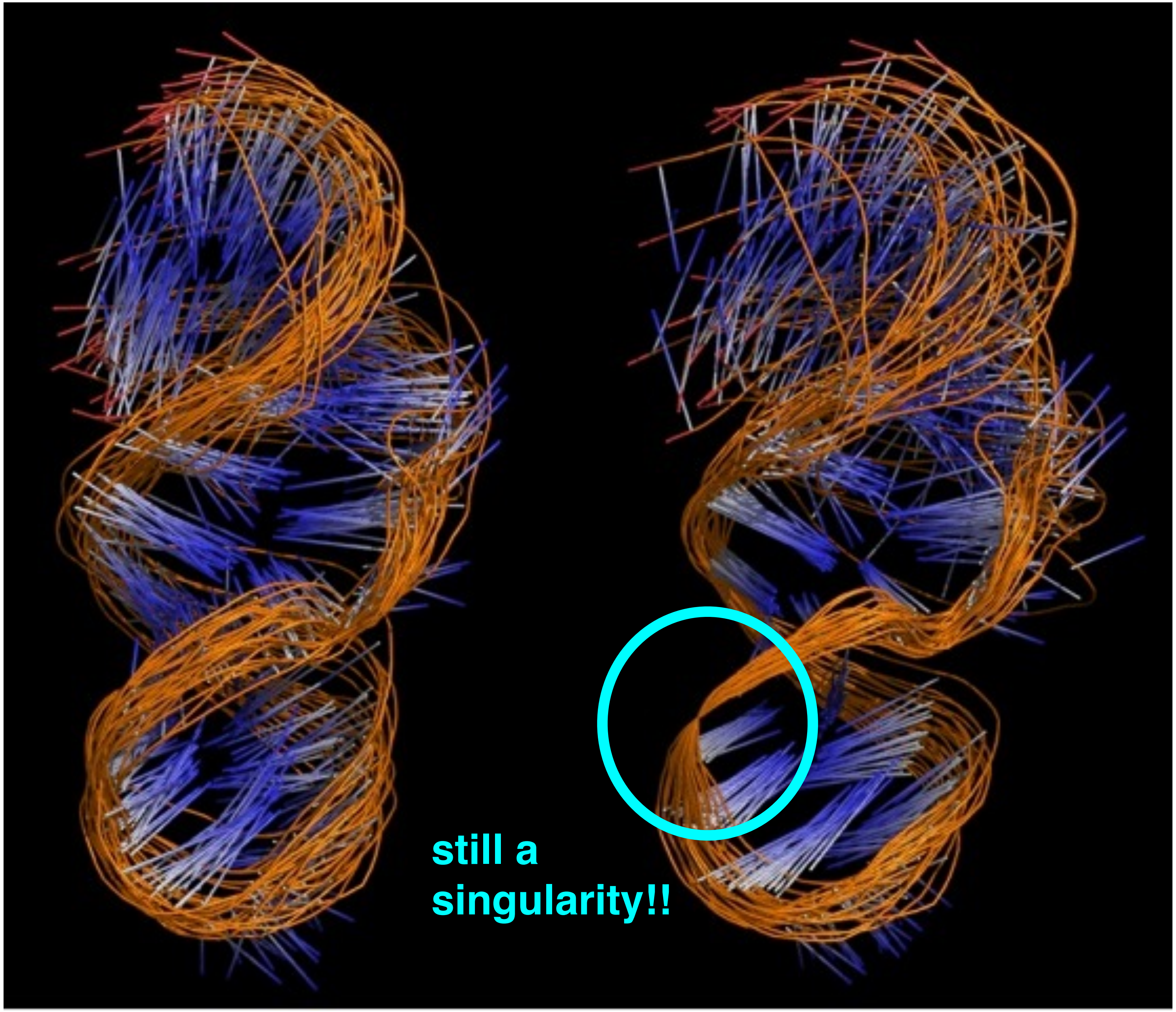
Least-squares: Find the rotation that minimizes the sum of squared distances between corresponding (labelled) atoms

Leadzyme: Ordinary vs Weighted Least-Squares

25 models



Least-Squares vs Maximum likelihood: Variances still a problem



Translations and unconstrained **variances** are "unidentifiable"

Problem: Atoms can be translated so that they perfectly superimpose

Solution: Constrain the variances

Classic superposition method: Least-squares

Multiple simultaneous superpositions for *N* molecules

$$SS = \sum_{i}^{N} \left\| \mathbf{X}_i \mathbf{R}_i - \mathbf{M} \right\|^2$$

$$SS = \sum_{i}^{N} \mathrm{tr}\left(\left[\mathbf{X}_i \mathbf{R}_i - \mathbf{M} \right]' \left[\mathbf{X}_i \mathbf{R}_i - \mathbf{M} \right] \right)$$

X is a structure to be superpositioned

R is a 3×3 rotation matrix

M is the average structure

ML method down-weights variable regions

1adz: Kunitz domain 2 of Tissue Factor Pathway Inhibitor

71 aa, 30 NMR models

Least-squares

Maximum likelihood

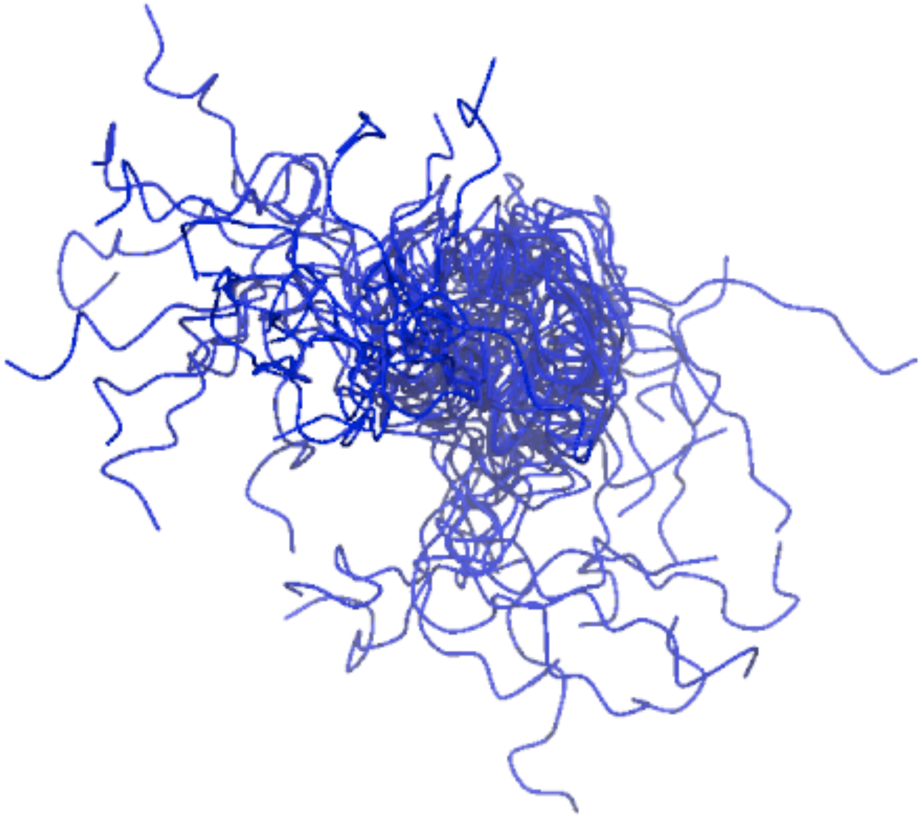
Burgering et al. (1997) JMB 269:395

Maximum likelihood superpositions

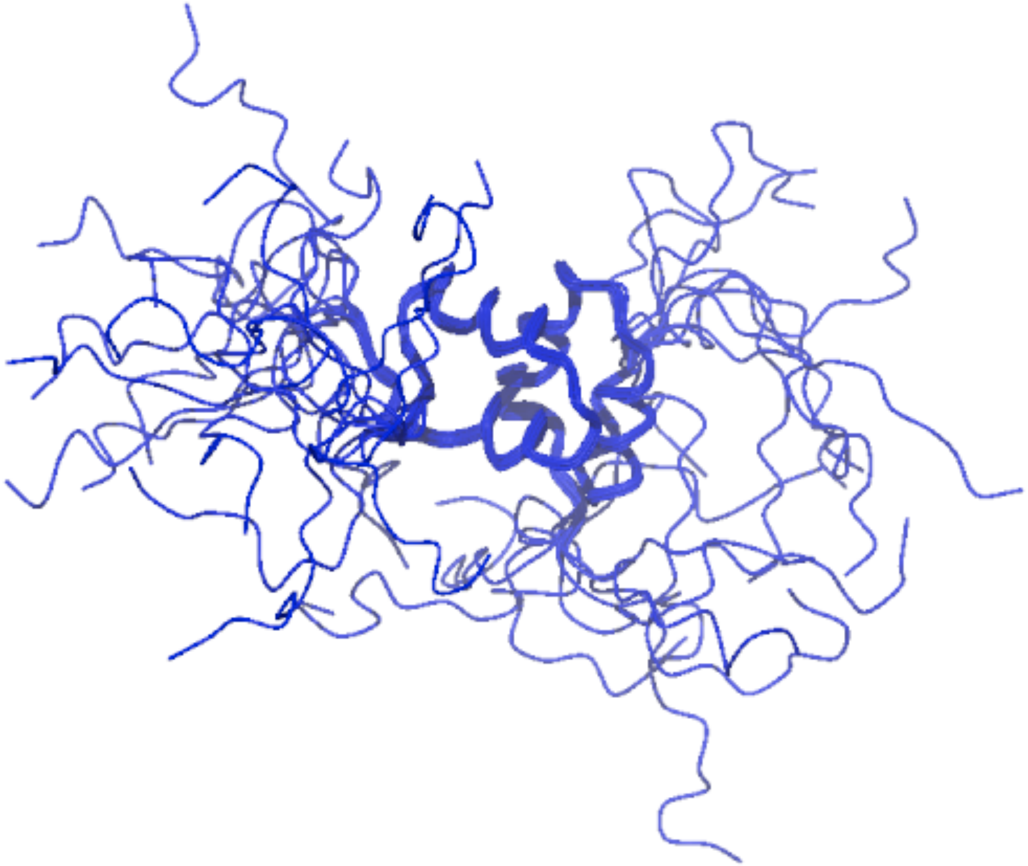
1ng7: poliovirus 3A, soluble domain

60aa, 10 NMR models

Least-squares



Maximum likelihood



Strauss et al. (2003) *JMB* 330:225

Weighted Least-Squares

Correct for unequal variances by weighting by the inverse of the variance:

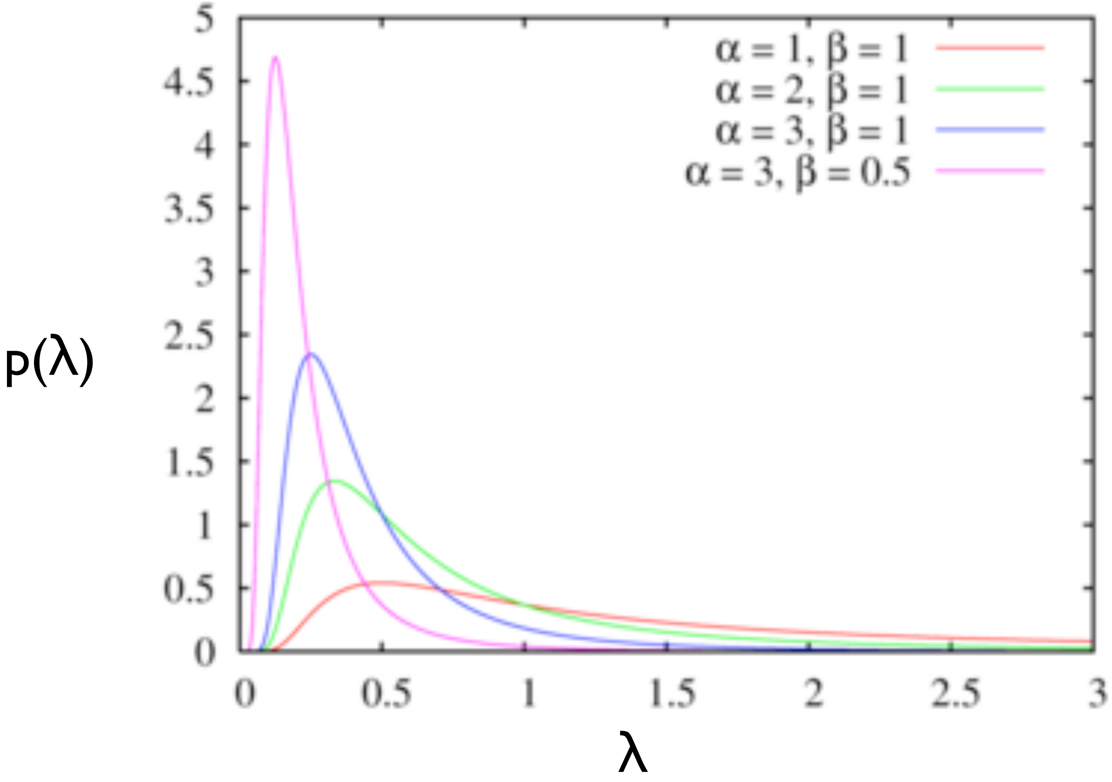
$$SS_w = \sum_i^N \|\mathbf{X}_i\mathbf{R}_i - \mathbf{M}\|^2_{\mathbf{\Sigma}^{-1}}$$

$$SS_w = \sum_i^N \mathrm{tr}\left([\mathbf{X}_i\mathbf{R}_i - \mathbf{M}]'\,\mathbf{\Sigma}^{-1}[\mathbf{X}_i\mathbf{R}_i - \mathbf{M}]\right)$$

Iteratively re-weighted least-squares algorithm:

1. **Calculate weighted superposition with current variances**
2. **Calculate variances from current superposition**
3. **Loop until convergence**

Hierarchical prior for the variances: Inverse gamma distribution



More realistic: variances cannot be zero or infinite

PDF: $$p\left(\lambda_j|\alpha\right) = \frac{\alpha^{\frac{1}{2}}}{\Gamma\left(\frac{1}{2}\right)}\lambda_j^{-\frac{3}{2}}e^{-\frac{\alpha}{\lambda_j}}$$

λ_j = a variance for atom j

Empirical Bayesian solution (hierarchical)

Old, broken solution

$$\hat{\boldsymbol{\Sigma}}_U = \frac{1}{3N}\sum_{i}^{N}(\check{\mathbf{X}}_i\mathbf{R}_i - \hat{\mathbf{M}})(\check{\mathbf{X}}_i\mathbf{R}_i - \hat{\mathbf{M}})'$$

"Shrunken" covariance matrix

$$\hat{\boldsymbol{\Sigma}}_{I\gamma} = \left(\frac{3N}{3N+3}\right)\left(\frac{2\alpha}{3N}\mathbf{I} + \hat{\boldsymbol{\Sigma}}_U\right)$$

$$\hat{\alpha} = \frac{K}{2\,\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\right)}$$

scale parameter of inverse gamma distribution

The difference is the covariance matrix

LS	ML	
$\hat{\boldsymbol{t}}_i = -\frac{\mathbf{1}'_K \mathbf{X}_i}{\mathbf{1}'_K \mathbf{1}_K}$	$\hat{\boldsymbol{t}}_i = -\frac{\mathbf{1}'_K \boxed{\boldsymbol{\Sigma}^{-1}} \mathbf{X}_i}{\mathbf{1}'_K \boxed{\boldsymbol{\Sigma}^{-1}} \mathbf{1}_K}$	Translation
$\hat{\mathbf{M}}' \check{\mathbf{X}}_i = \mathbf{U} \boldsymbol{\Lambda} \mathbf{V}'$ $\hat{\mathbf{R}}_i = \mathbf{V} \mathbf{P} \mathbf{U}'$	$\hat{\mathbf{M}}' \boxed{\hat{\boldsymbol{\Sigma}}^{-1}} \check{\mathbf{X}}_i = \mathbf{U} \boldsymbol{\Lambda} \mathbf{V}'$ $\hat{\mathbf{R}}_i = \mathbf{V} \mathbf{P} \mathbf{U}'$	Rotation

ML weights by the inverse covariance matrix, which down-weights variable regions.

Assume the structures (matrices) have a Gaussian distribution

The usual scalar Gaussian (Bell curve) PDF:

$$\mathrm{p}\left(x_i|\mu,\sigma\right) = \left(2\pi\right)^{-\frac{1}{2}}\sigma^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma}\left(x_i-\mu\right)^2\right\}$$

μ = ***mean***

σ = ***variance***

The matrix Gaussian PDF:

$$\mathrm{p}\left(\mathbf{X}_i|\mathrm{M},\mathbf{\Sigma}\right) = \left(2\pi\right)^{-\frac{KD}{2}}\left|\mathbf{\Sigma}\right|^{-\frac{D}{2}}\exp\left\{-\frac{1}{2}\left\|\mathbf{X}_i-\mathbf{M}\right\|^2_{\mathbf{\Sigma}^{-1}}\right\}$$

The Matrix Gaussian has a *covariance matrix* (Σ) instead of a single variance

Covariance matrix

Each atom has a variance *and* can co-vary with other atoms

Example covariance matrix for five atoms:

$$\begin{array}{cc}
 & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} &
\begin{bmatrix}
10 & 5 & -2 & 1 & 1 \\
5 & 20 & 10 & -5 & -3 \\
-2 & 10 & 15 & 2 & 5 \\
1 & -5 & 2 & 5 & 8 \\
1 & -3 & 5 & 8 & 30
\end{bmatrix}
\end{array}$$

Variances are on the diagonal.

Covariances are off diagonal elements.

PCA structure plots

1. **Superimpose simulated structures**
2. **Do PCA on the covariance/correlation matrix**
3. **Plot PCs on structure**

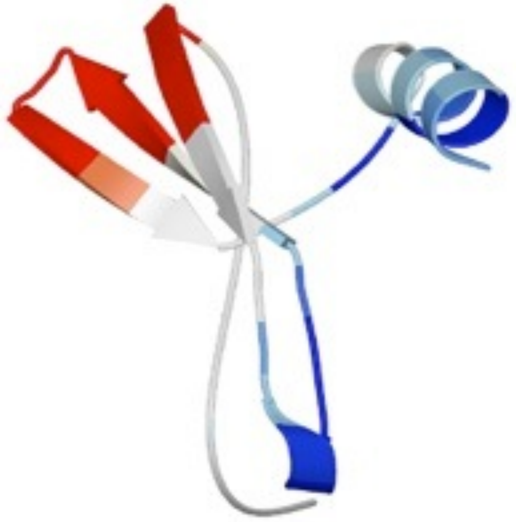
PDB ID: 2sdf
SDF-1, 30 NMR models

Red regions are self-correlated

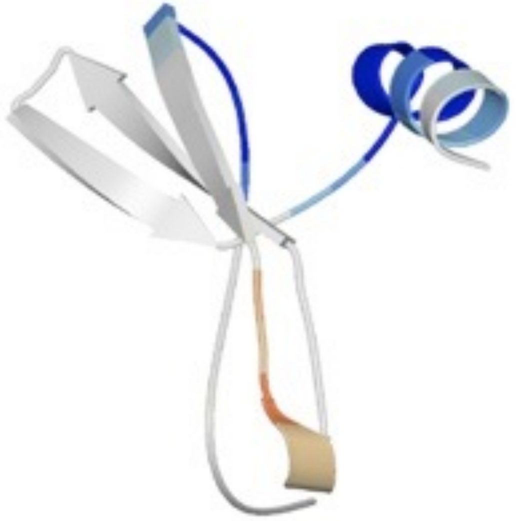
Blue regions are self-correlated

Red and Blue are anti-correlated

PC 1



PC 2

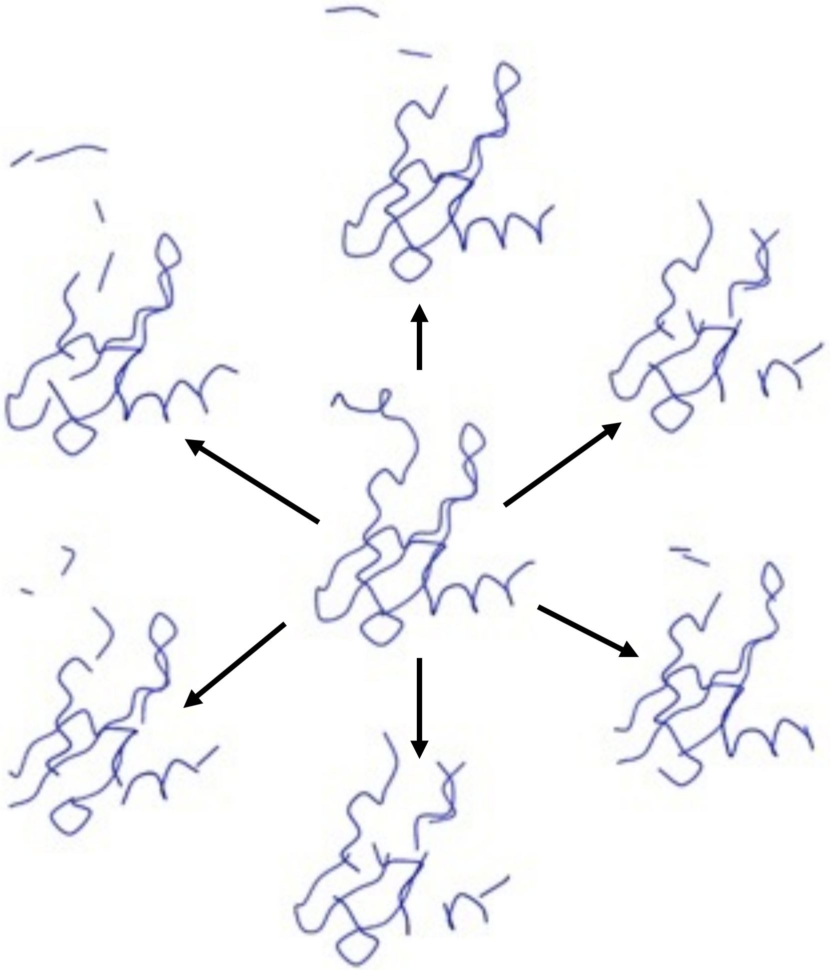


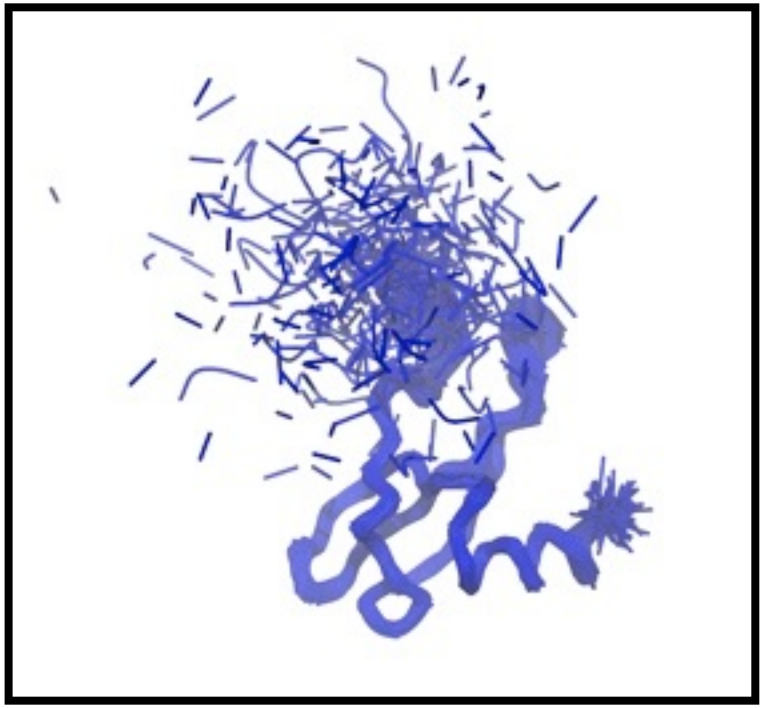
PC 3

Simulation test of the ML method: Generate random structures

Generated 300 random Gaussian structures, with known parameters:

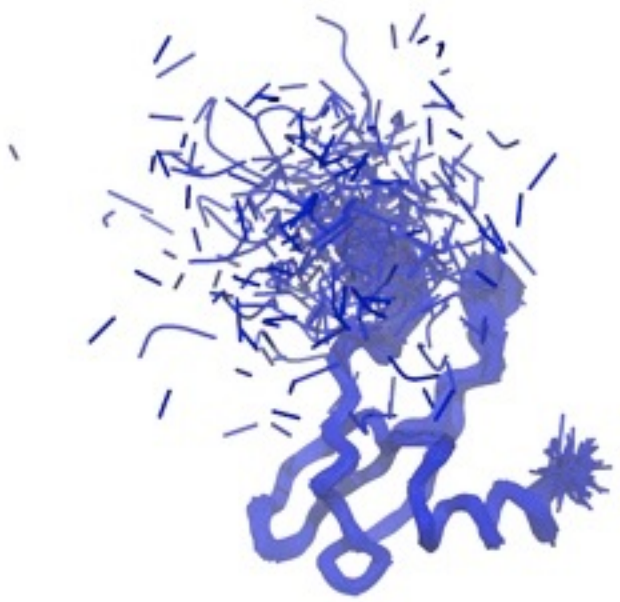
- mean structure
- covariance matrix
- true superposition
- rotations
- translations





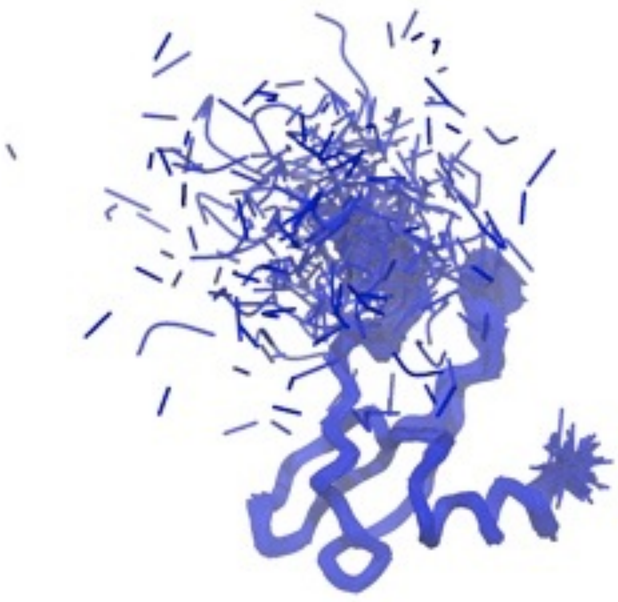
LS vs ML: Maximum likelihood recovers the true superposition accurately

True superposition

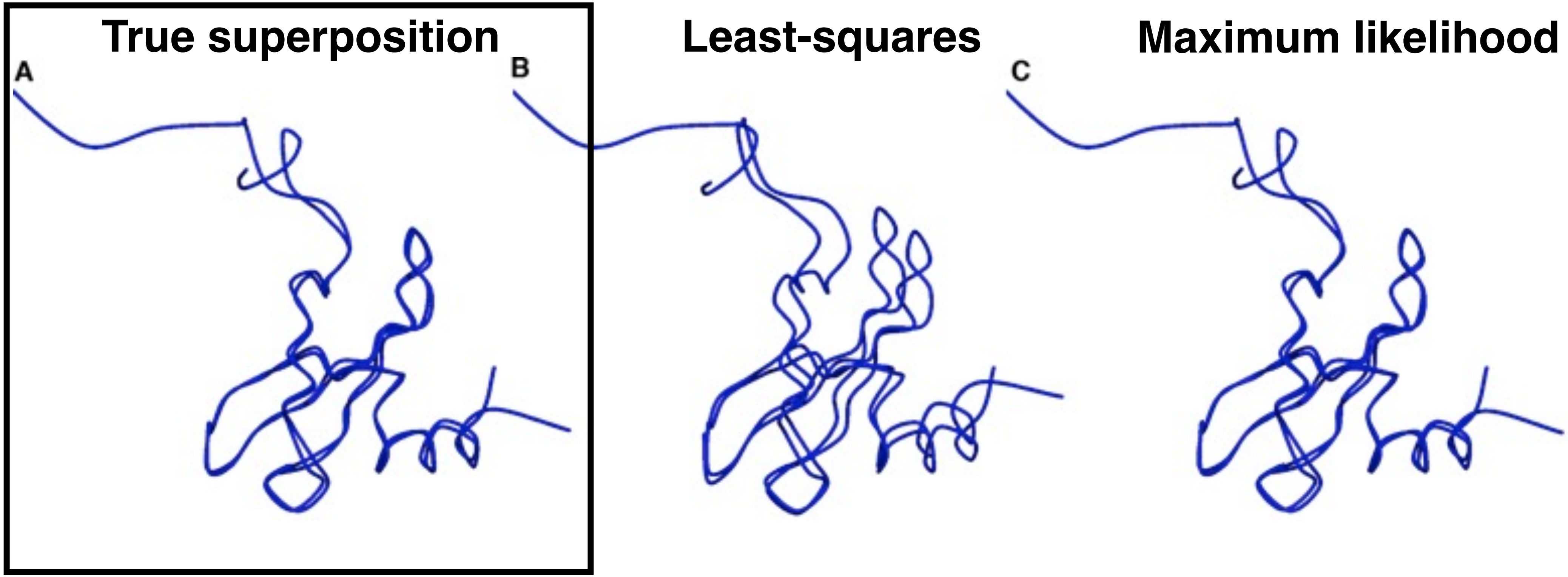


Least-squares

Maximum likelihood



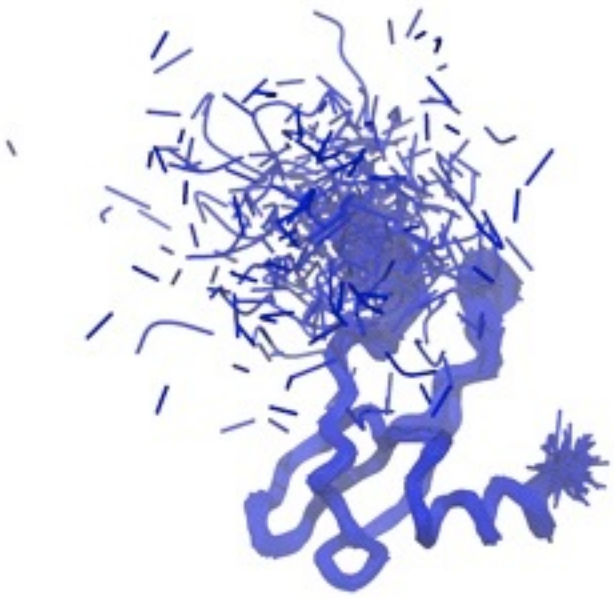
LS vs ML: Maximum likelihood recovers the true superposition accurately



True superposition **Least-squares** **Maximum likelihood**

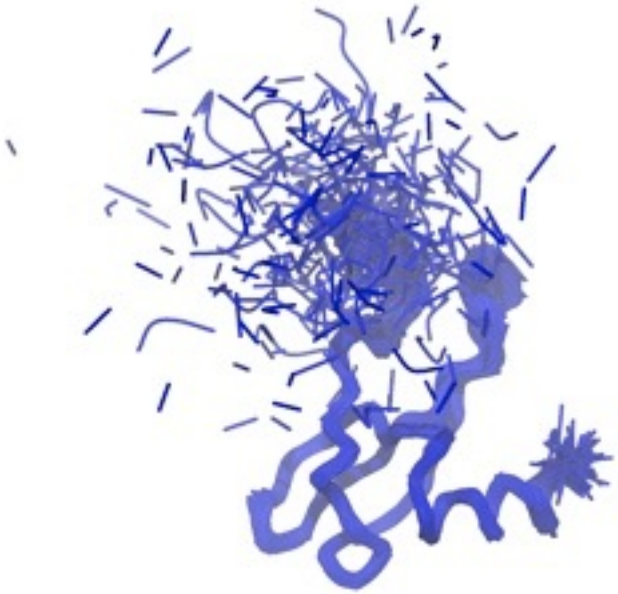
LS vs ML: Maximum likelihood recovers the true superposition accurately

True superposition

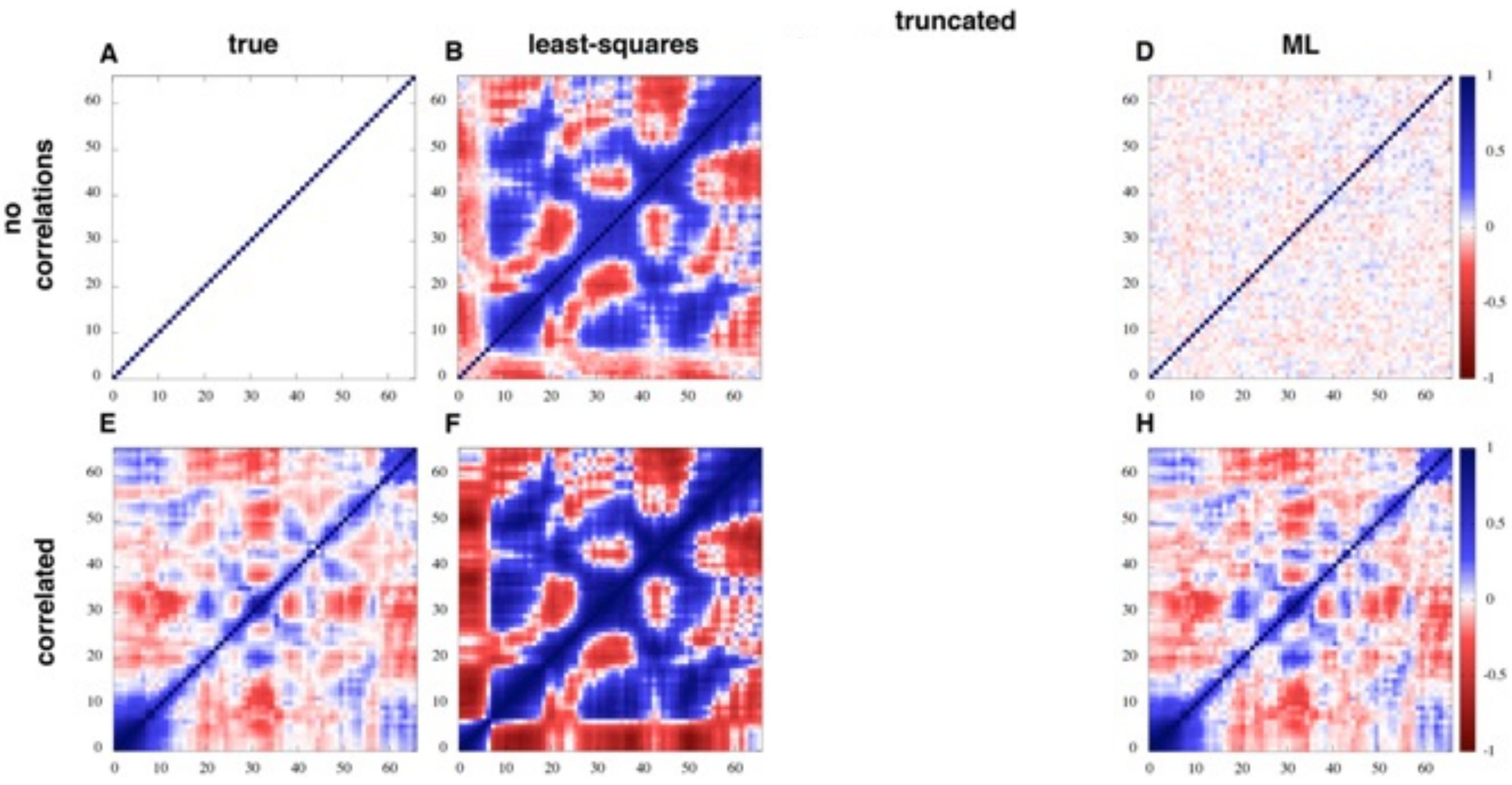


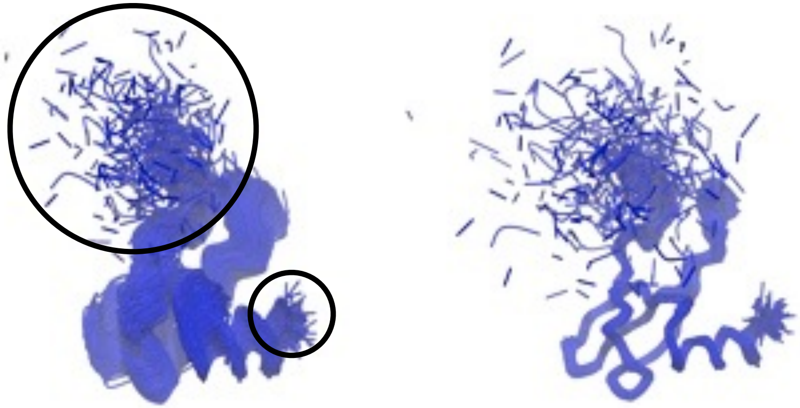
Least-squares

Maximum likelihood



Least-squares gives artifactual correlations

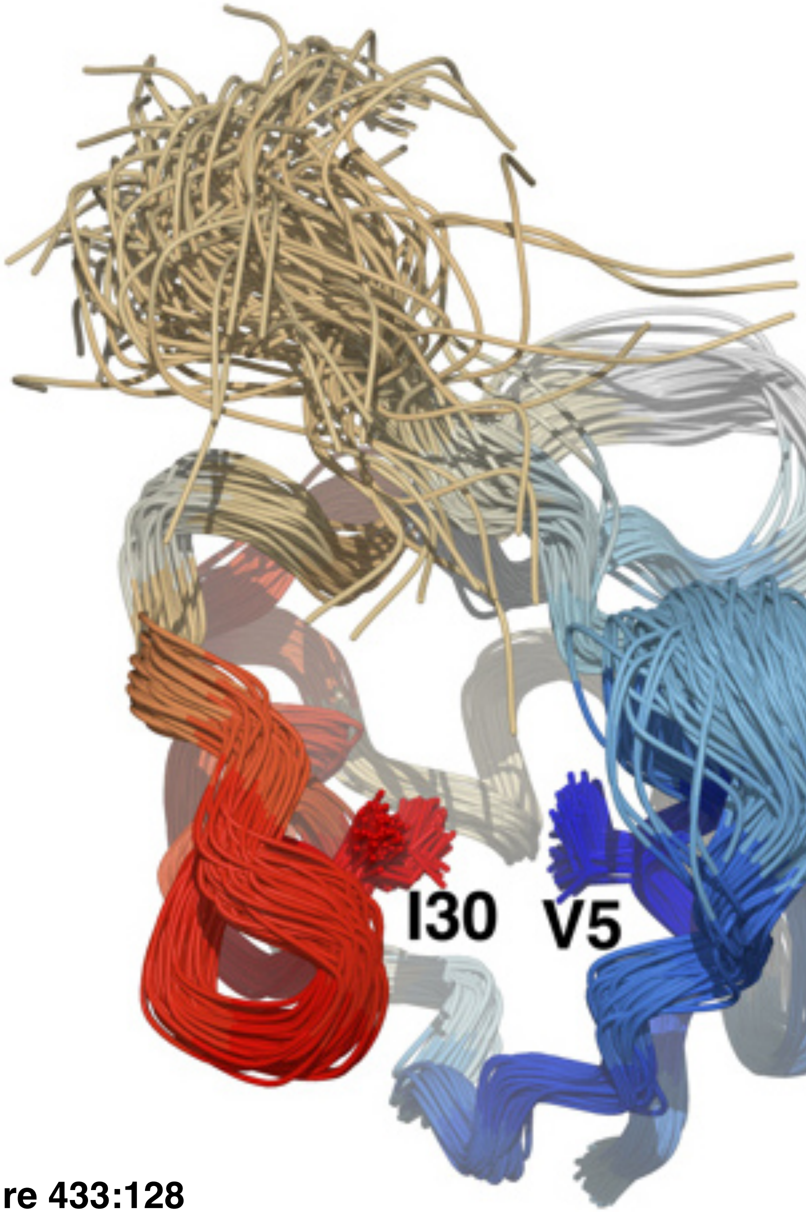




PCA for analyzing correlations in NMR families

Molecular dynamics

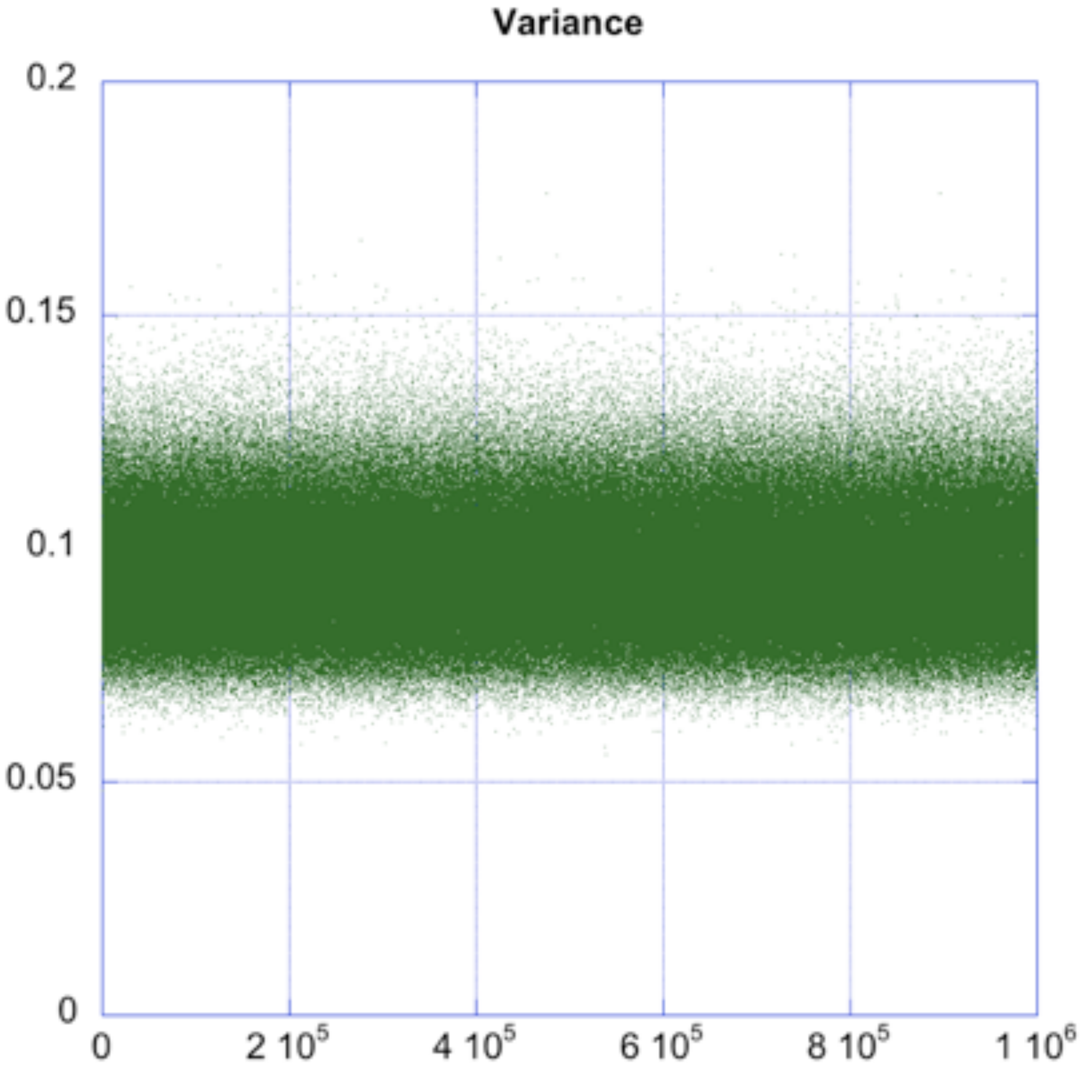
MD trajectory of ubiquitin with NMR constraints



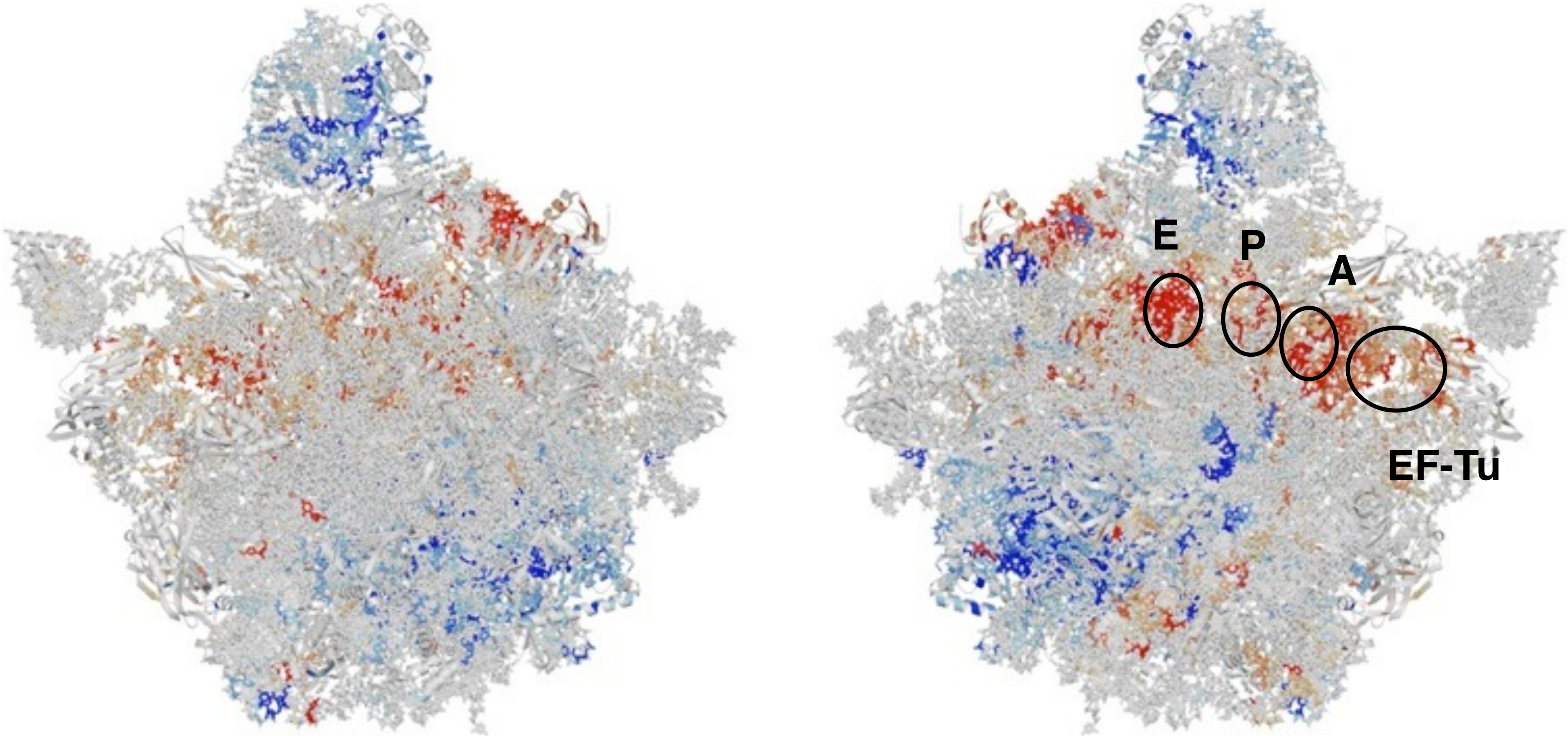
Lindorff-Larsen et al. (2005) Nature 433:128

Posterior mode solution for inverse Wishart priors on the covariance matrix

MCMC chain results for nonisotropic variance hyper-parameter



50S large subunit of the *Haloarcula* ribosome



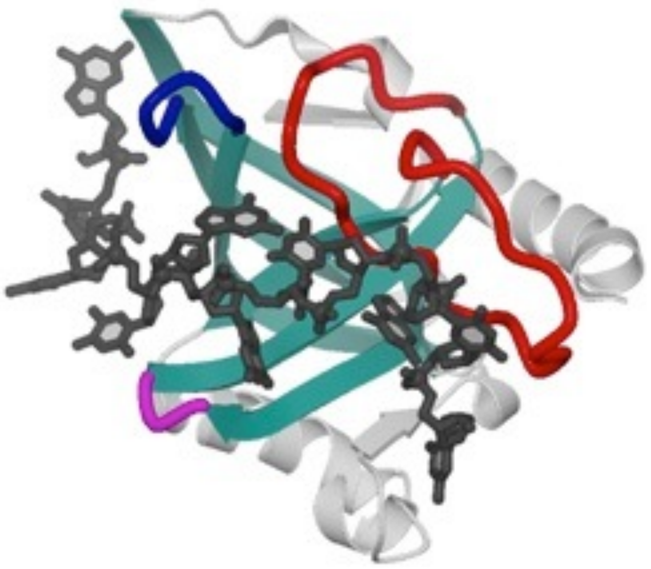
PC2 of ML superposition of 10 ribosome structures

Hansen et al. (2002) *Mol Cell.* 10:117.

Hansen et al. (2003) *J Mol Biol.* 330:1061

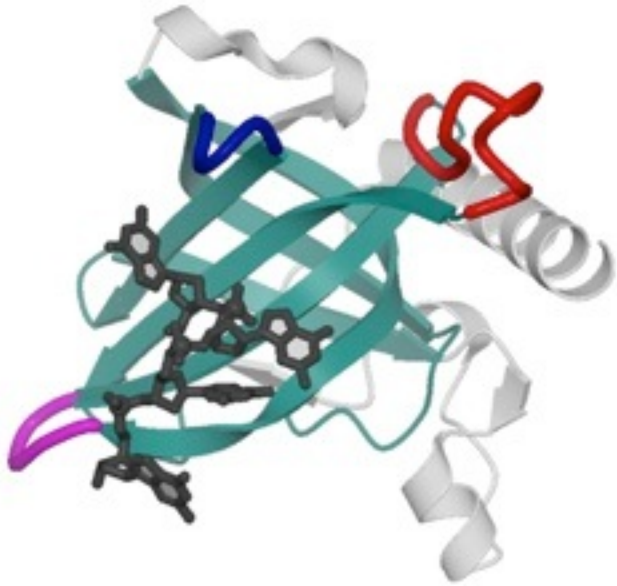
Three OB-fold telomeric domains

Cdc13



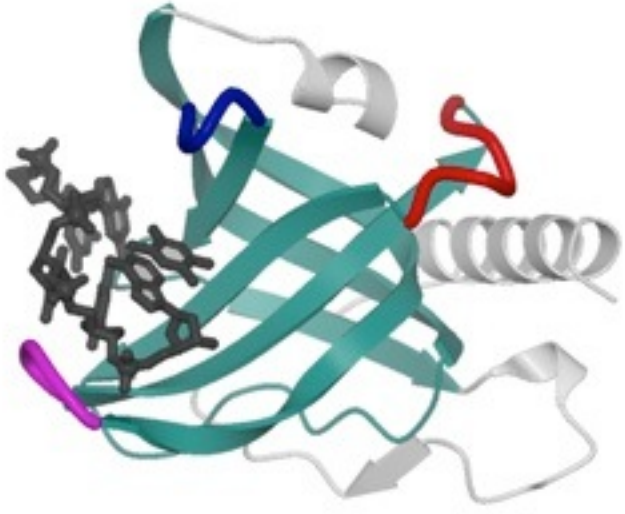
GTGTGGGTGTG

TEBP α1



GGGGTTTTGGGG

Pot1

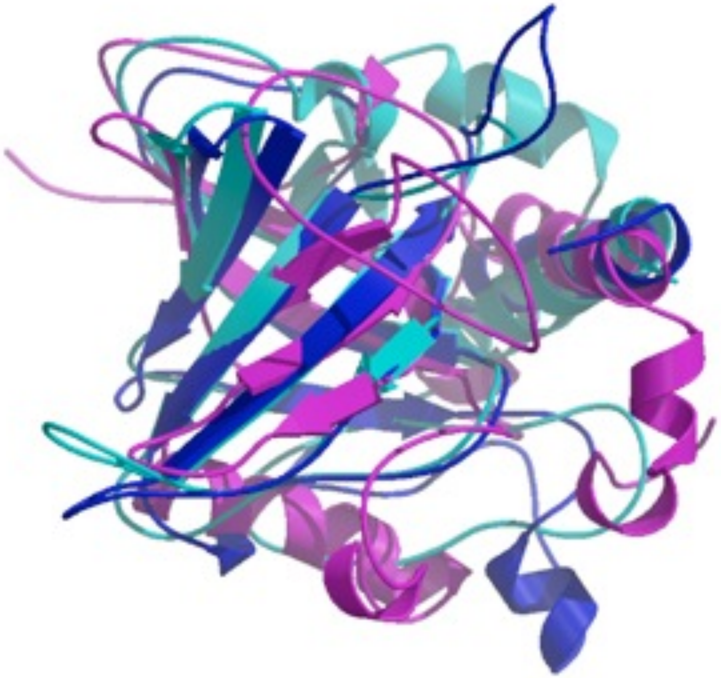


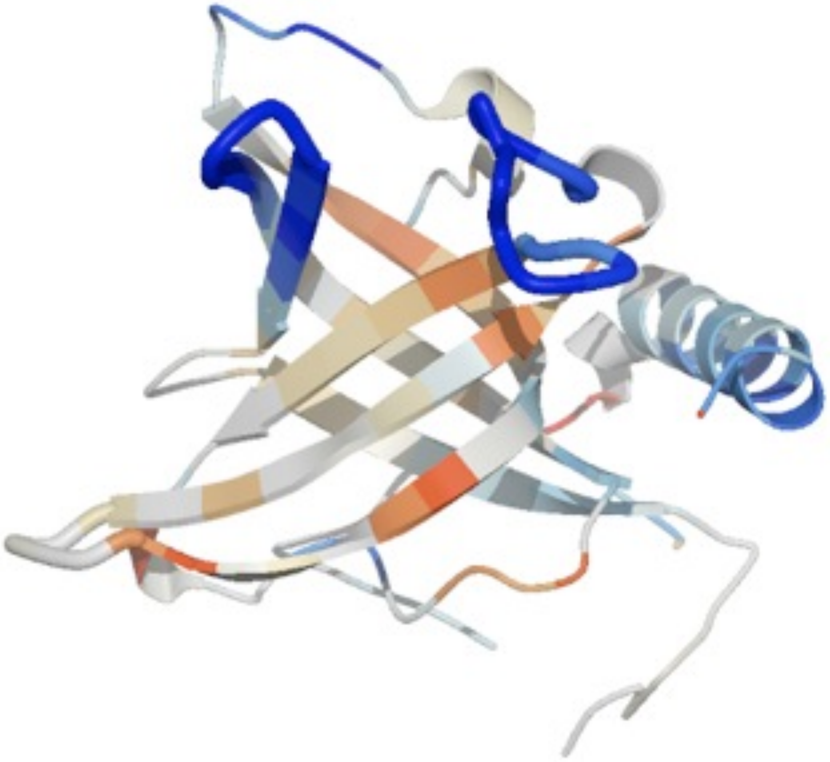
GGGATC

PCA of co-evolving structures

ML superposition of TEBP α1, cdc13, pot1

PC1





Loop conformations have co-evolved

Conditional distributions: The mean

Assume uniform improper prior on M

$$\mathrm{p}\left(\mathbf{M}|\mathbf{X},\mathbf{\Sigma},\mathbf{R},\boldsymbol{t}\right) = (2\pi)^{-\frac{3K}{2}}\left|\mathbf{\Omega}\right|^{-\frac{3}{2}}\exp\left(-\frac{1}{2}\,\mathrm{tr}\left\{[\mathbf{M}-\mathbf{B}]'\mathbf{\Omega}^{-1}[\mathbf{M}-\mathbf{B}]\right\}\right)$$

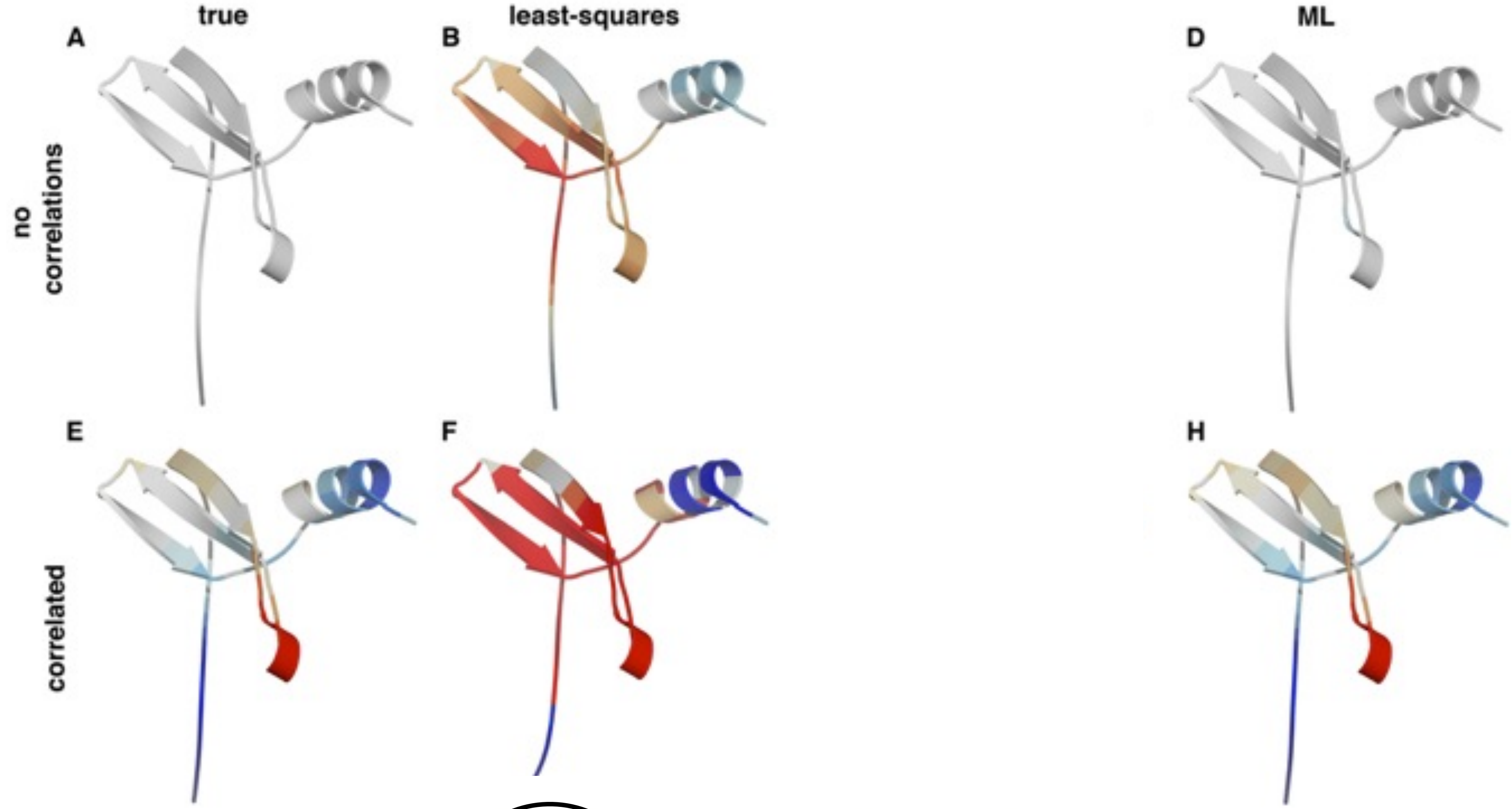
$$\mathbf{\Omega} = \frac{1}{N}\mathbf{\Sigma}$$

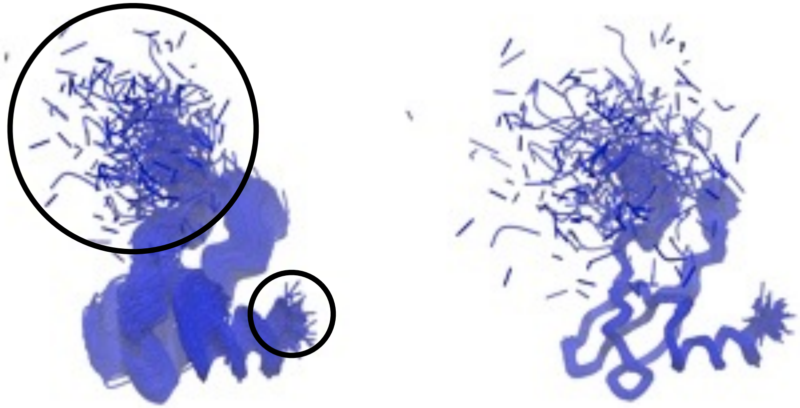
$$\mathbf{\Sigma}_{\text{iso}} = \phi\mathbf{I} \qquad \mathbf{\Omega} = \frac{\phi}{N}\mathbf{I}$$

$$\mathbf{B} = \frac{1}{N}\sum_{i}^{N}\mathbf{Y}_i \qquad \mathbf{Y}_i = (\mathbf{X}_i + \mathbf{1}_K\boldsymbol{t}_i')\mathbf{R}_i$$

A matrix normal distribution centered on the sample average (the ML estimate)

Least-squares produces artifactual PCs

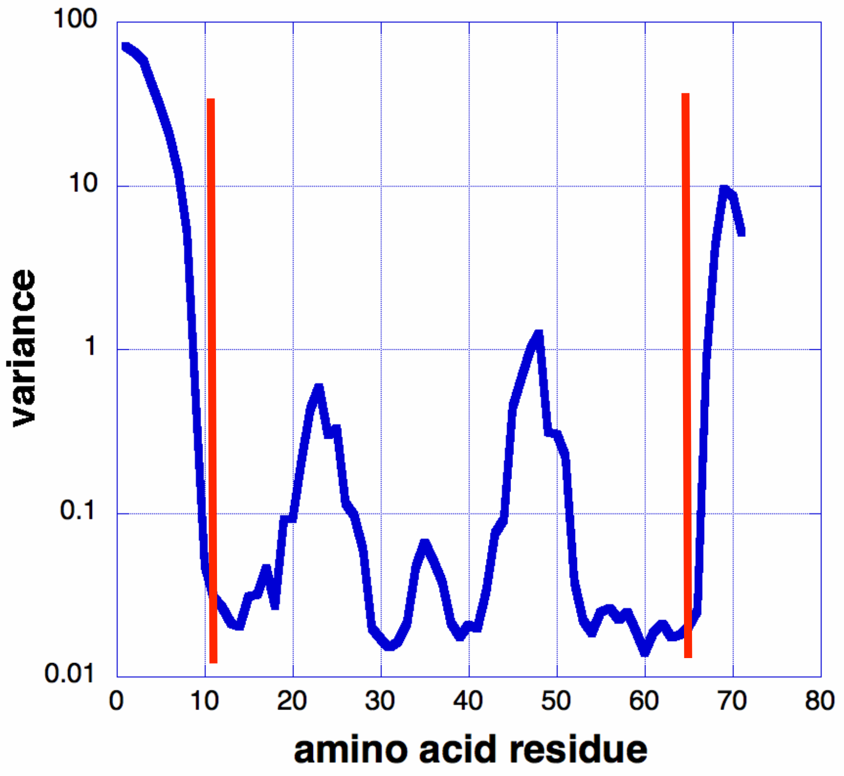




Variance can range 10,000-fold

1adz: Kunitz domain 2 of Tissue Factor Pathway Inhibitor

71 aa, 30 NMR models



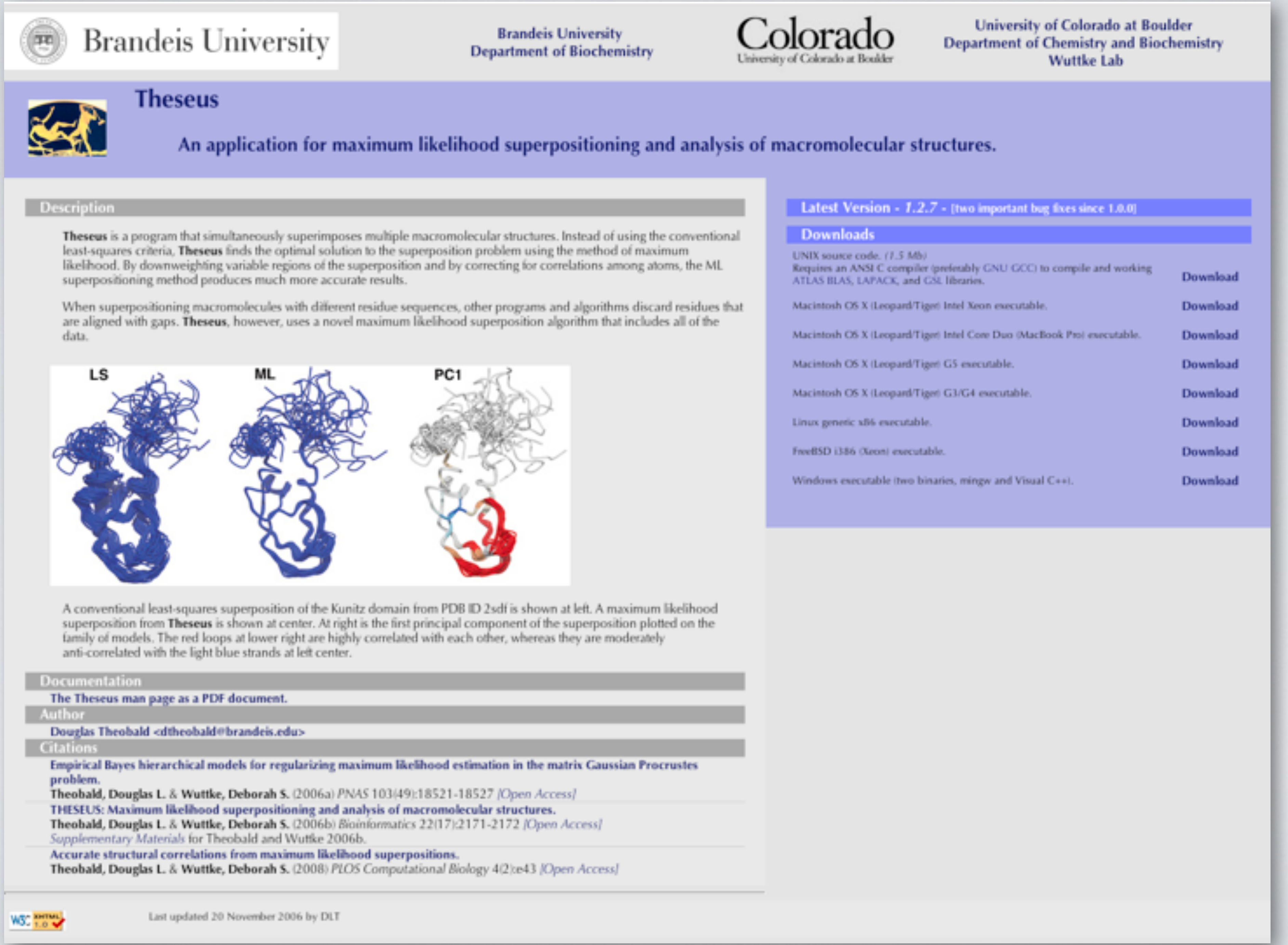
Trim the "un-superimposable", disordered regions

Burgering et al. (1997) JMB 269:395

Hyperparameter for inverse Wishart covariance matrix: Gamma distribution

$$\lambda \sim \mathrm{G}\left(\frac{2}{\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1} + \frac{2}{\delta}\right)}, \frac{k^2 + 2p}{2}\right)$$

THESEUS: http://www.theseus3d.org



Brandeis University

Brandeis University
Department of Biochemistry

Colorado
University of Colorado at Boulder

University of Colorado at Boulder
Department of Chemistry and Biochemistry
Wuttke Lab

Theseus

An application for maximum likelihood superpositioning and analysis of macromolecular structures.

Description

Theseus is a program that simultaneously superimposes multiple macromolecular structures. Instead of using the conventional least-squares criteria, **Theseus** finds the optimal solution to the superposition problem using the method of maximum likelihood. By downweighting variable regions of the superposition and by correcting for correlations among atoms, the ML superpositioning method produces much more accurate results.

When superpositioning macromolecules with different residue sequences, other programs and algorithms discard residues that are aligned with gaps. **Theseus**, however, uses a novel maximum likelihood superposition algorithm that includes all of the data.

A conventional least-squares superposition of the Kunitz domain from PDB ID 2sdf is shown at left. A maximum likelihood superposition from **Theseus** is shown at center. At right is the first principal component of the superposition plotted on the family of models. The red loops at lower right are highly correlated with each other, whereas they are moderately anti-correlated with the light blue strands at left center.

Documentation

The Theseus man page as a PDF document.

Author

Douglas Theobald <dtheobald@brandeis.edu>

Citations

Empirical Bayes hierarchical models for regularizing maximum likelihood estimation in the matrix Gaussian Procrustes problem.
Theobald, Douglas L. & **Wuttke, Deborah S.** (2006a) *PNAS* 103(49):18521-18527 *[Open Access]*

THESEUS: Maximum likelihood superpositioning and analysis of macromolecular structures.
Theobald, Douglas L. & **Wuttke, Deborah S.** (2006b) *Bioinformatics* 22(17):2171-2172 *[Open Access]*
Supplementary Materials for Theobald and Wuttke 2006b.

Accurate structural correlations from maximum likelihood superpositions.
Theobald, Douglas L. & **Wuttke, Deborah S.** (2008) *PLOS Computational Biology* 4(2):e43 *[Open Access]*

Last updated 20 November 2006 by DLT

Latest Version - *1.2.7* - [two important bug fixes since 1.0.0]

Downloads

UNIX source code. *(1.5 Mb)* Requires an ANSI C compiler (preferably GNU GCC) to compile and working ATLAS BLAS, LAPACK, and GSL libraries.	Download
Macintosh OS X (Leopard/Tiger) Intel Xeon executable.	Download
Macintosh OS X (Leopard/Tiger) Intel Core Duo (MacBook Pro) executable.	Download
Macintosh OS X (Leopard/Tiger) G5 executable.	Download
Macintosh OS X (Leopard/Tiger) G3/G4 executable.	Download
Linux generic x86 executable.	Download
FreeBSD i386 (Xeon) executable.	Download
Windows executable (two binaries, mingw and Visual C++).	Download