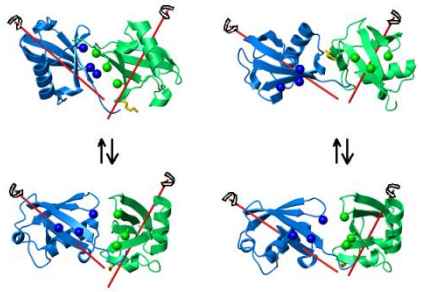
Ярослав Рябов

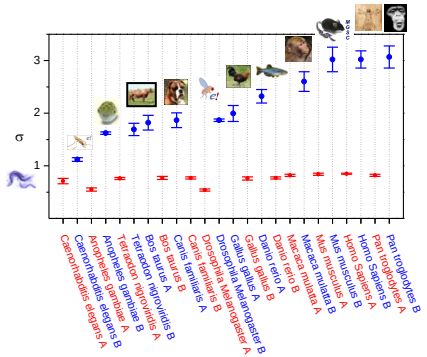
Структурная и Вычислительная Биология

Соотношение между Структурой и Динамикой Белков

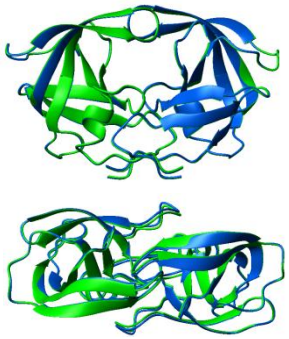
Protein Domain Dynamics



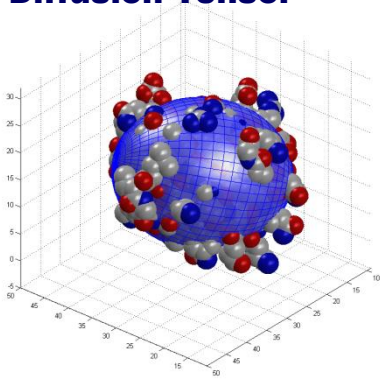
Statistics of genomes



Assembling Protein Structures



Predictions of Protein Diffusion Tensor



Headlines

- **Protein Domain Dynamics**

  a model with applications to NMR

- **Diffusion Properties of Proteins**

  from ellipsoid model

- **Assembling Structures of Multi-Domain Proteins and Protein Complexes**

  guided by protein diffusion tensor

- **Genome Evolution**

  from statistics of proteins' properties to Exon size distribution

Protein Domain Dynamics

a model with applications to NMR

a very General Concept

Theoretical model

$C(t)$

Also a function of model parameters like

Protein structure,
Protein diffusion tensor
Parameters of internal motions
etc.

Experimental data

NMR relaxation R_1, R_2, NOE, RDC
Spin Labeling data
etc.

Ryabov & Fushman, Proteins (2006), MRC (2006), JACS (2007)

Time Correlation Function

$$C(t) = \left\langle D^{(2)*}_{q,0}(\Omega^{0}_{L\to I}) D^{(2)}_{q,0}(\Omega^{t}_{L\to I}) \right\rangle_{L\to I}$$

Wigner rotation matrixes

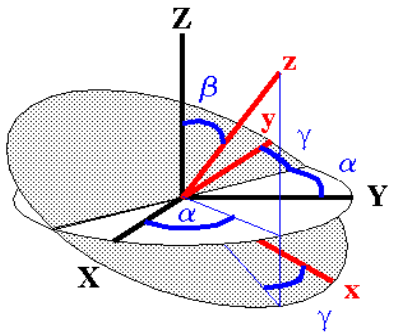
$$D^{(l)}_{m,n}(\Omega) = \exp(-im\alpha) d^{(l)}_{m,n}(\beta) \exp(-in\gamma)$$

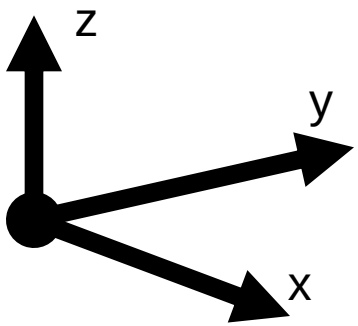
***Abragam**, 1961 Principles of Nuclear Magnetism*

Euler Angles

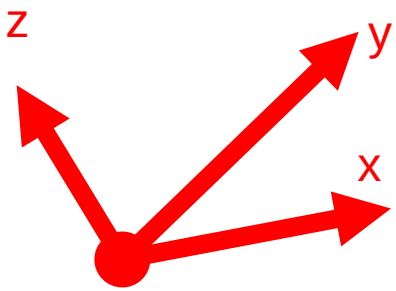
$$\Omega(t) \Rightarrow \{\alpha(t), \beta(t), \gamma(t)\}$$

$$0 \le \alpha \le 2\pi$$
$$0 \le \beta \le \pi$$
$$0 \le \gamma \le 2\pi$$





Ω_{LI}



Laboratory frame (L)

Instantaneous Residue orientation (I)

Laboratory frame (L)

$\downarrow \Omega_{LP}$

Protein orientation (P)

$\downarrow \Omega_{PD}$

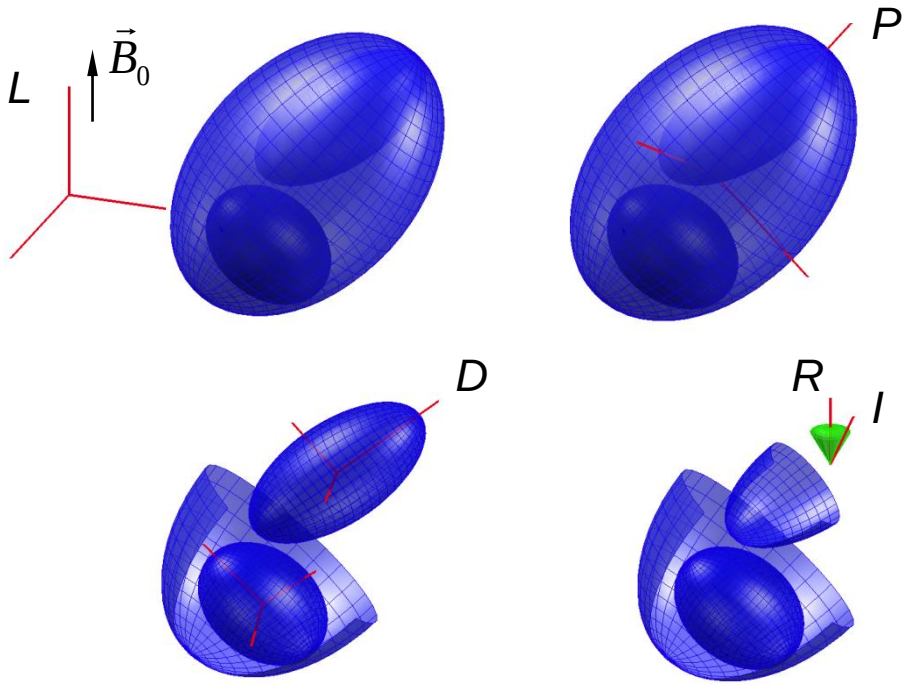
Domain orientation (D)

$\downarrow \Omega_{DR}$

Averaged Residue orientation (R)

$\downarrow \Omega_{RI}$

Instantaneous Residue orientation (I)



$$C(t) = < \sum_{l,l'=-2}^{2} \sum_{m,m'=-2}^{2} \sum_{n,n'=-2}^{2} \sum_{k.k'=-2}^{2} D^{2*}_{ql}(\Omega_{LP}(0)) D^{2}_{ql'}(\Omega_{LP}(t)) \times$$
$$\times D^{2*}_{lm}(\Omega_{PD}(0)) D^{2}_{l'm'}(\Omega_{PD}(t)) \times$$
$$\times D^{2*}_{mn}(\Omega_{DR}(0)) D^{2}_{m'n'}(\Omega_{DR}(t)) \times$$
$$\times D^{2*}_{nk}(\Omega_{RI}(0)) D^{2}_{n'k',}(\Omega_{RI}(t)) >$$

Assumption

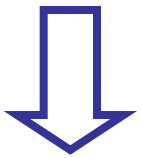
All dynamic modes are statistically independent from each other

$$C(t) = < \sum_{l,l'=-2}^{2} \sum_{m,m'=-2}^{2} \sum_{n,n'=-2}^{2} \sum_{k.k'=-2}^{2} D^{2*}_{ql}(\Omega_{LP}(0)) D^{2}_{ql'}(\Omega_{LP}(t)) \times$$

$$\times D^{2*}_{lm}(\Omega_{PD}(0)) D^{2}_{l'm'}(\Omega_{PD}(t)) \times$$

$$\times D^{2*}_{mn}(\Omega_{DR}(0)) D^{2}_{m'n'}(\Omega_{DR}(t)) \times$$

$$\times D^{2*}_{nk}(\Omega_{RI}(0)) D^{2}_{n'k',}(\Omega_{RI}(t)) >$$



$$C(t) = \sum_{l,l'=-2}^{2} \sum_{m,m'=-2}^{2} \sum_{n,n'=-2}^{2} \sum_{k.k'=-2}^{2} < D^{2*}_{ql}(\Omega_{LP}(0)) D^{2}_{ql'}(\Omega_{LP}(t)) > \times$$

$$\times < D^{2*}_{lm}(\Omega_{PD}(0)) D^{2}_{l'm'}(\Omega_{PD}(t)) > \times$$

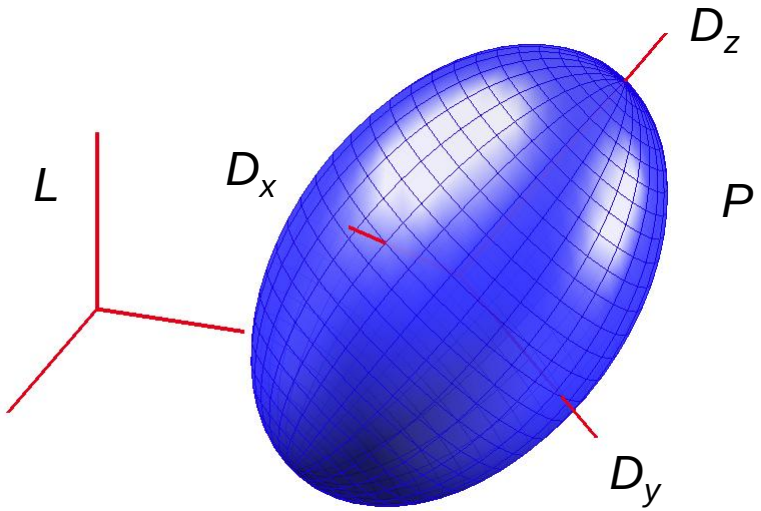
$$\times < D^{2*}_{mn}(\Omega_{DR}(0)) D^{2}_{m'n'}(\Omega_{DR}(t)) > \times$$

$$\times < D^{2*}_{nk}(\Omega_{RI}(0)) D^{2}_{n'k',}(\Omega_{RI}(t)) >$$

Laboratory frame (L)

$\downarrow \ \Omega_{LP}$

Protein orientation (P)



Rotation diffusion of an ellipsoid

***Favro**, Phys. Rev. 1960*
***Woessner**, J. Chem. Phys. 1962*

$$C^{LP}_{qm,qn}(t) = \left\langle D^{(2)*}_{q,m}(\Omega^{0}_{L\to P}) D^{(2)}_{q,n}(\Omega^{t}_{L\to P}) \right\rangle_{L\to P} = \frac{1}{5}\sum_{z=-2}^{2} e^{-E_z t} a^{*}_{z,m} a_{z,n}$$

$a_{z,m}$ Decomposition coefficients and

E_z Eigenvalues are

$D_x,\ D_y,\ D_z$ Functions of Diffusion tensor Principal values only

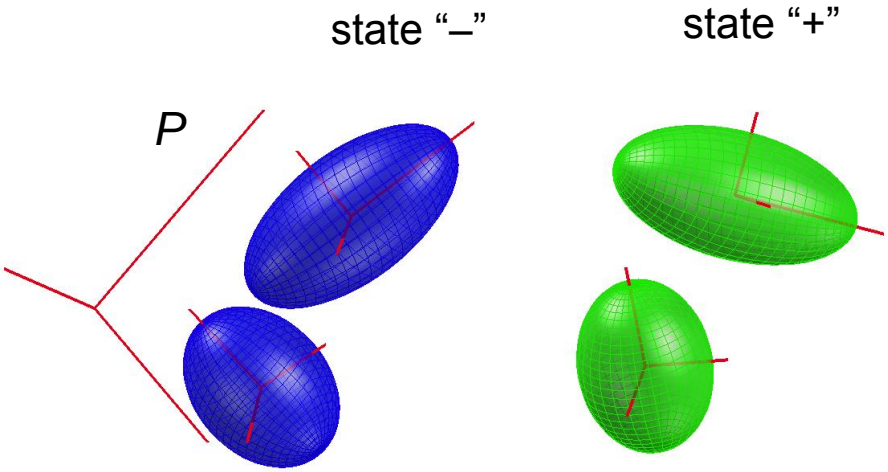
P coincides with Diffusion tensor principal vectors

***Huntress Jr**, Adv. Magn. Reson. 1970*

Protein orientation (P)

$\downarrow \ \Omega_{PD}$

Domain orientation (D)



***Model* of Interconversion between Two States**

$$p_{eq}(\Omega^{-}_{P\to D}) = \frac{\tau_{-+}}{\tau_{-+}+\tau_{+-}} \equiv p_{-}$$

$$p_{eq}(\Omega^{+}_{P\to D}) = \frac{\tau_{+-}}{\tau_{-+}+\tau_{+-}} \equiv p_{+}$$

$$p(\Omega^{-}_{P\to D}\,|\,\Omega^{-}_{P\to D},t) = p_{-} + p_{+}\exp\{-t/\tau_j\}$$

$$p(\Omega^{-}_{P\to D}\,|\,\Omega^{+}_{P\to D},t) = p_{+}\left(1-\exp\{-t/\tau_j\}\right)$$

$$p(\Omega^{+}_{P\to D}\,|\,\Omega^{-}_{P\to D},t) = p_{-}\left(1-\exp\{-t/\tau_j\}\right)$$

$$p(\Omega^{+}_{P\to D}\,|\,\Omega^{+}_{P\to D},t) = p_{+} + p_{-}\exp\{-t/\tau_j\}$$

Domain mobility correlation time

$$\tau_j = \tau_{-+}\tau_{+-}/(\tau_{-+}+\tau_{+-})$$

$$C^{PD}_{mk,nl}(t) = \left\langle D^{(2)*}_{m,k}(\Omega^{0}_{P\to D}) D^{(2)}_{n,l}(\Omega^{t}_{P\to D})\right\rangle_{P\to D} =$$

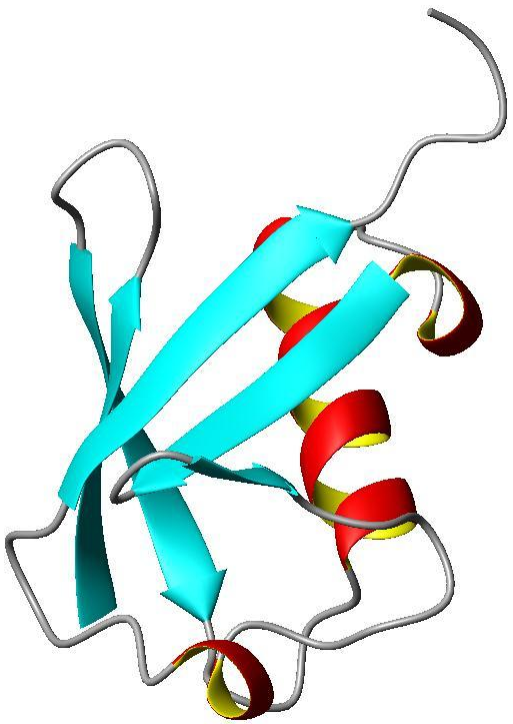
$$= \sum_{\Omega^{0}_{P\to D},\,\Omega^{t}_{P\to D}=\Omega^{-}_{P\to D},\,\Omega^{+}_{P\to D}} p_{eq}(\Omega^{0}_{P\to D})\, p(\Omega^{0}_{P\to D}\,|\,\Omega^{t}_{P\to D},t)\, D^{(2)*}_{m,k}(\Omega^{0}_{P\to D}) D^{(2)}_{n,l}(\Omega^{t}_{P\to D})$$

Domain orientation (D)

$\downarrow \ \Omega_{DR}$

Averaged Residue orientation (R)

PDB 1D3Z: Ubiquitin



Assumption

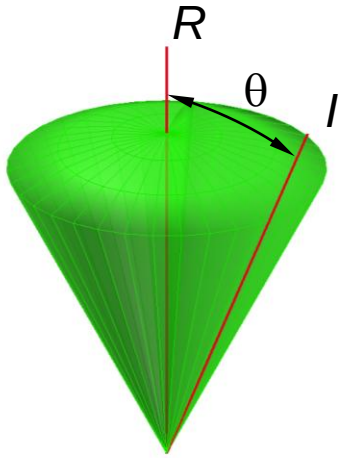
Averaged orientation of a residue within a domain is constant and Ω_{DR} is given in *PDB* convention

$$C^{DR}_{ks,lh}(t) = \left\langle D^{(2)*}_{k,s}(\Omega_{DR}(0)) D^{(2)}_{l,h}(\Omega_{DR}(t)) \right\rangle_{DR} = D^{(2)*}_{k,s}(\Omega_{DR}) D^{(2)}_{l,h}(\Omega_{DR})$$

Averaged Residue orientation (R)

$\downarrow \quad \Omega_{RI}$

Instantaneous Residue orientation (I)



Lipari & Szabo *– "Model Free"*
J. Am. Chem. Soc. 1982

$$C_{s0,h0}^{RI}(t) = \delta_{s,0}\delta_{h,0}\left[S^2 + (1-S^2)\exp\{-t/\tau_l\}\right]$$

Order Parameter $$S = \left\langle D_{0,0}^{(2)*}(\Omega_{R\to I}^{0}) \right\rangle = \left\langle \frac{1}{2}(3\cos^2(\beta_{R\to I}) - 1) \right\rangle$$

$$S = \frac{\cos(\theta)}{2}\left(1+\cos(\theta)\right)$$

Wobbling in a Cone model

NMR ^{15}N Relaxation and *NOE*

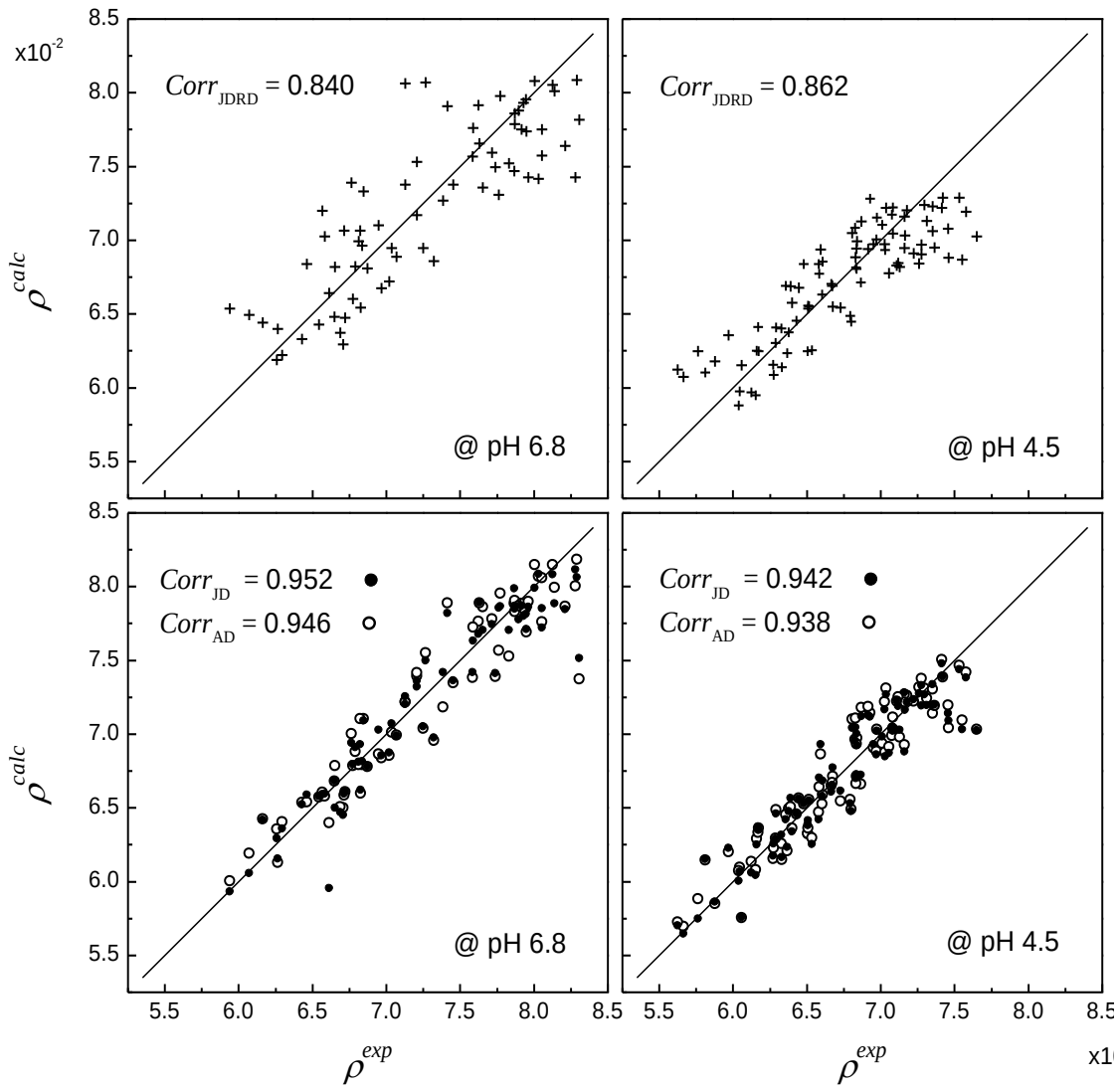
for Ubiquitin dimer

17 fitting parameters
including

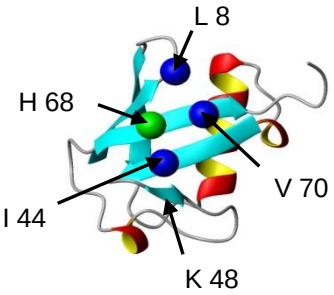
$D_x,\ D_y,\ D_z$

$p_+,\ p_-$

And all 12 Euler angles for both
Domain orientations
in two conformations



Derived Structures and Dynamics Parameters

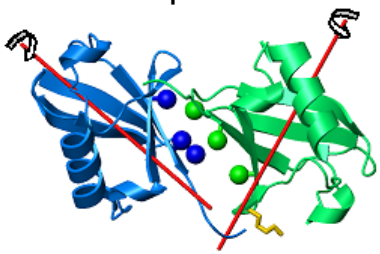


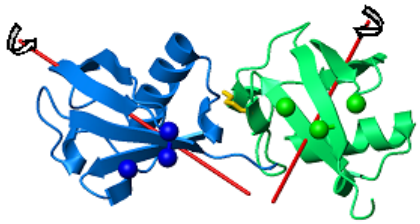
pH 6.8

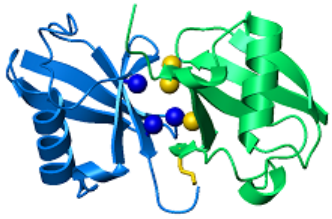
pH 4.5

p_+=0.9

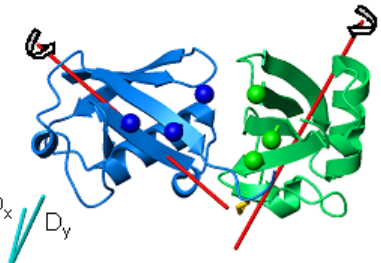
p_+=0.82

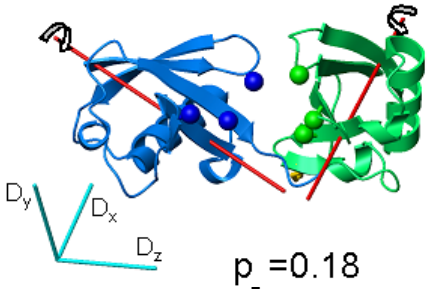


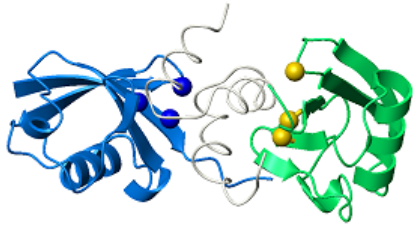




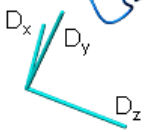
Crystal Ub_2
PDB 1AAR







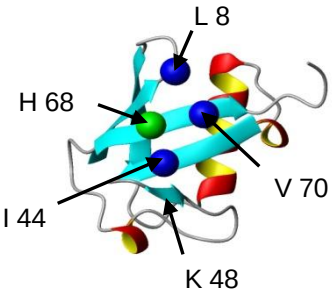
Docked:
Ub_2 + UBA



p_-=0.1

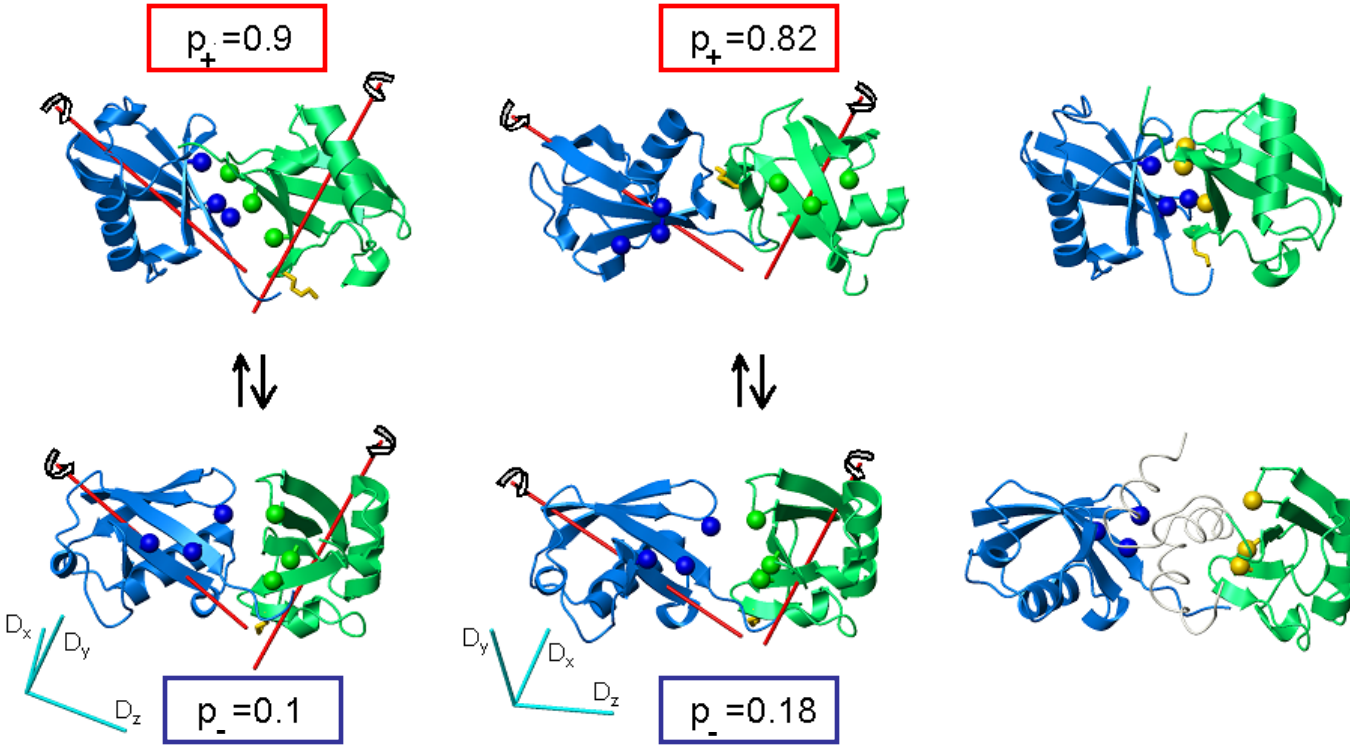
p_-=0.18

Derived Structures and Dynamics Parameters



pH 6.8

pH 4.5



Crystal Ub_2
PDB 1AAR

Docked:
Ub_2 + UBA

His 68 pKa=5.5
***Fujiwara et al**, J. Bio. Chem. 2004*

	pH 6.8	pH 4.5
Both protonated	**0.002**	**0.826**
One protonated	**0.091**	**0.165**
Non protonated	**0.907**	**0.009**

Conclusions

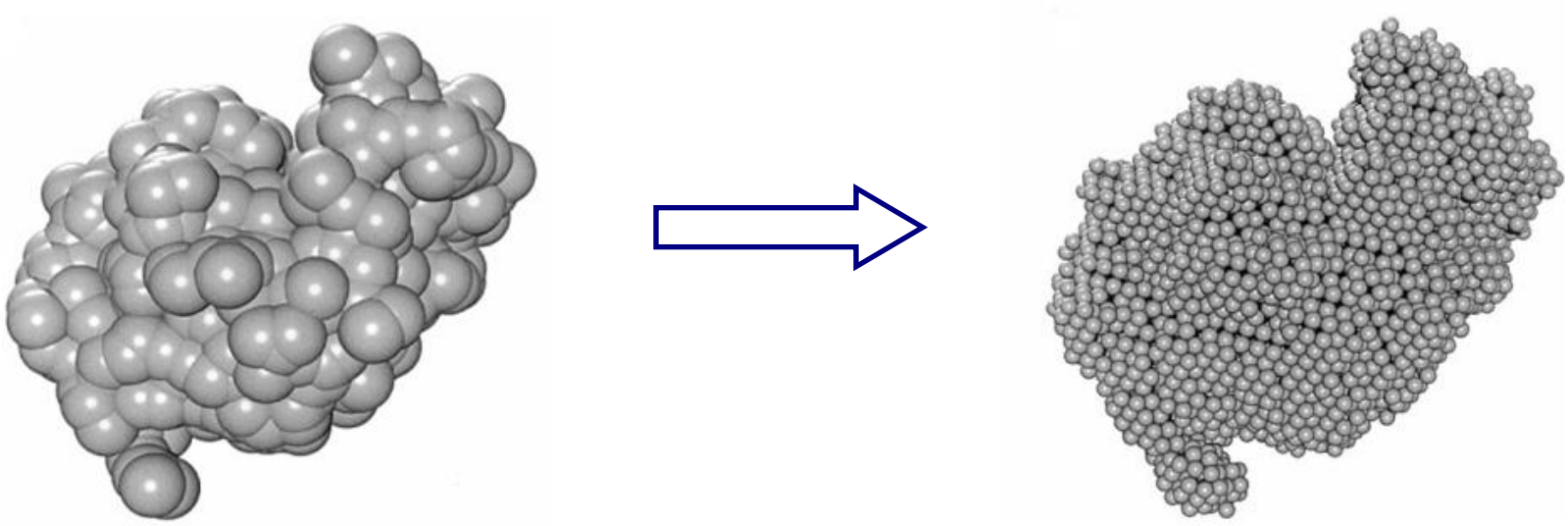
- Suggested model captures essential features of domain dynamics in Ubiquitin Dimer and provides information about domain mobility and orientations in this protein that cannot be derived from other approaches
- The results derived from this model agree with independent experimental data such as crystal and docked structures, chemical shift perturbation, and spin labeling data
- However, it should be noted that this approach can report only about mutual domain orientations - not positioning - in multidomain proteins

Connections with the next Project

- Is it possible to use derived information about protein dynamics for further characterization of protein structure?

Modeling **Diffusion Properties of Proteins**

State of the art: Bead algorithms



Extrapolation for Bead size ⟹ zero

10 000 beads

J. Garcia de la Torre et al., J Mag Res **147,** 138–146 (2000)

Diffusion Properties of Proteins

from ellipsoid model

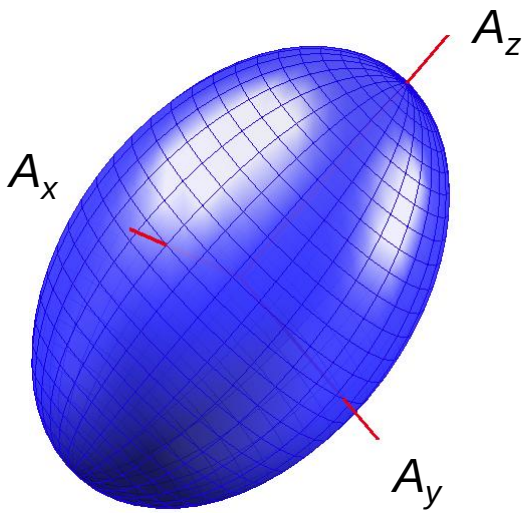
Why an ellipsoid model ?

Diffusion Tensor

$$\begin{bmatrix} D_x & 0 & 0 \\ 0 & D_y & 0 \\ 0 & 0 & D_z \end{bmatrix}$$

One-to-**One**
mapping

Ellipsoid Shell



3 Euler angles for
Diffusion Tensor PAF

3 Euler angles for
Ellipsoid orientation

Diffusion Properties of Proteins

from ellipsoid model

Main problem:

How to build the ellipsoid shell for a protein structure ?

State of the art: Inertia-equivalent ellipsoids

$$Re = \frac{\rho V L}{\mu} = \frac{\text{Inertia Forces}}{\text{Friction Forces}} \approx 0.01$$

$$\tau_{inertial\ relaxation} \approx 10^{-13} s \qquad d_{inertial relaxation} \approx 0.1\ \text{Å}$$

Inertia is irrelevant for protein diffusion

Diffusion Properties of Proteins

from ellipsoid model

Diffusion process related to **friction**

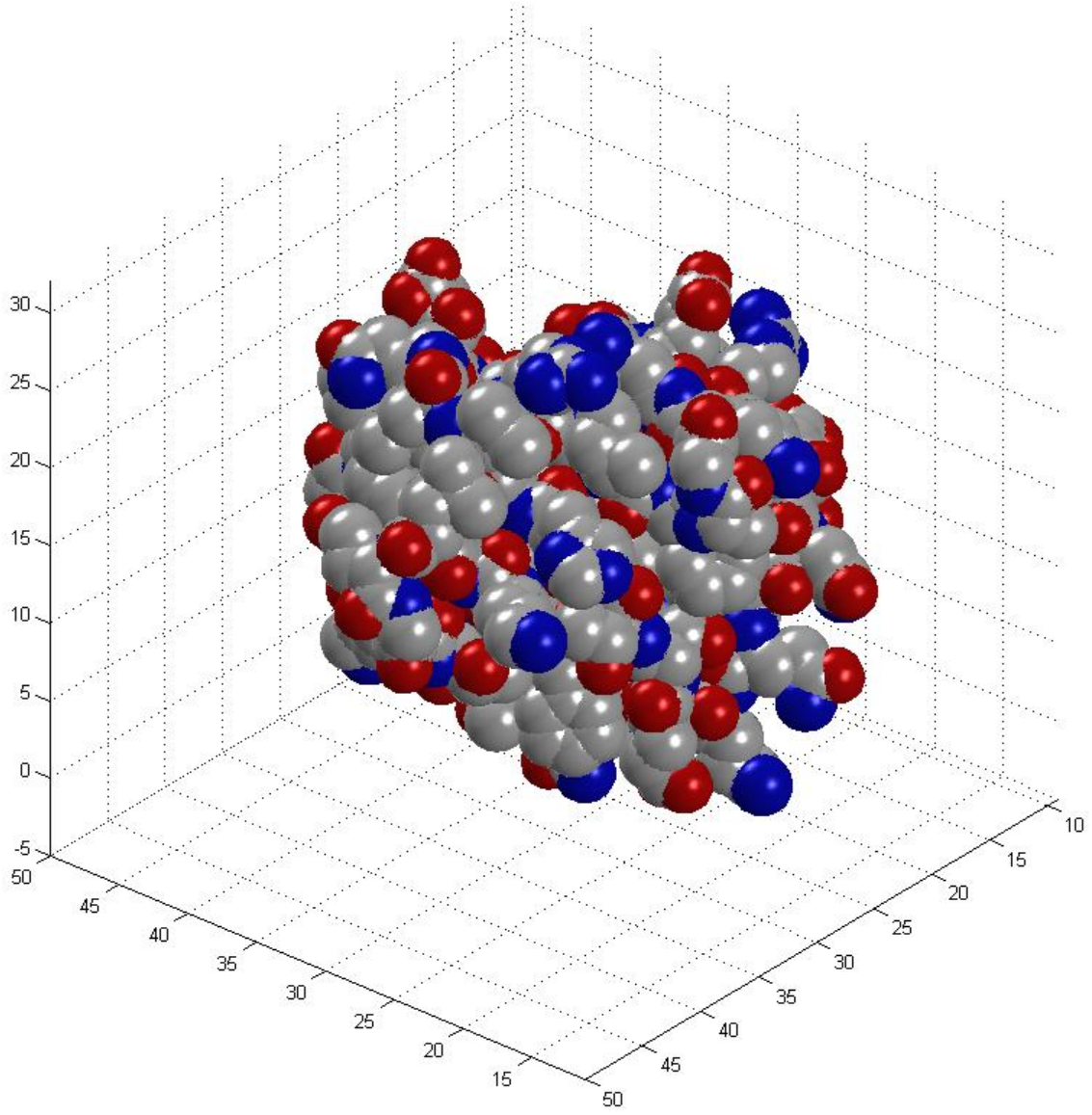
The **friction** occurs at protein **surface**

Proposal

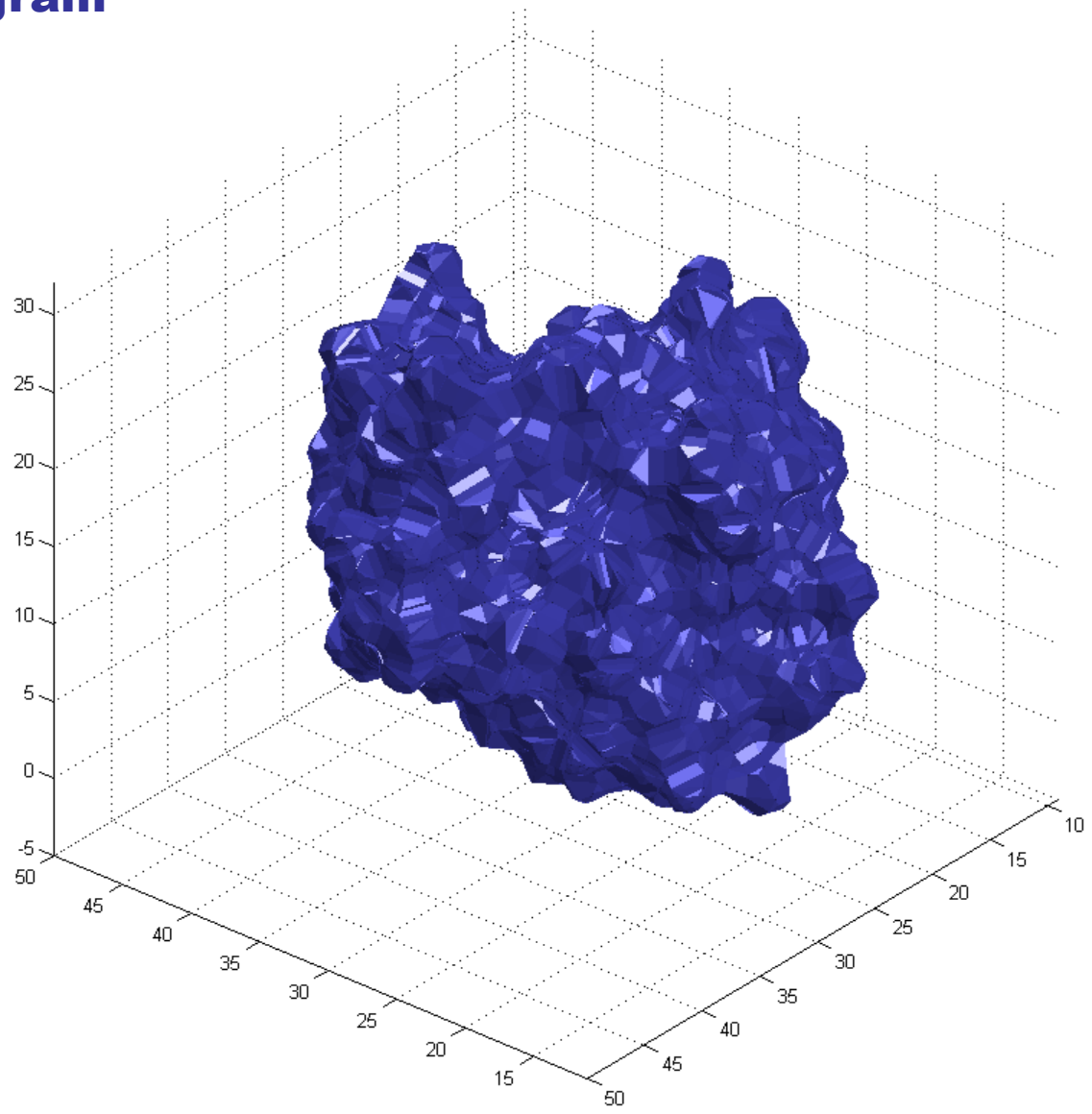
Let us use topology of protein surface to derive
Equivalent ellipsoid for protein

Ryabov & Fushman, JACS (2006)

Mapping protein surfaces



Mapping protein surfaces
SURF program



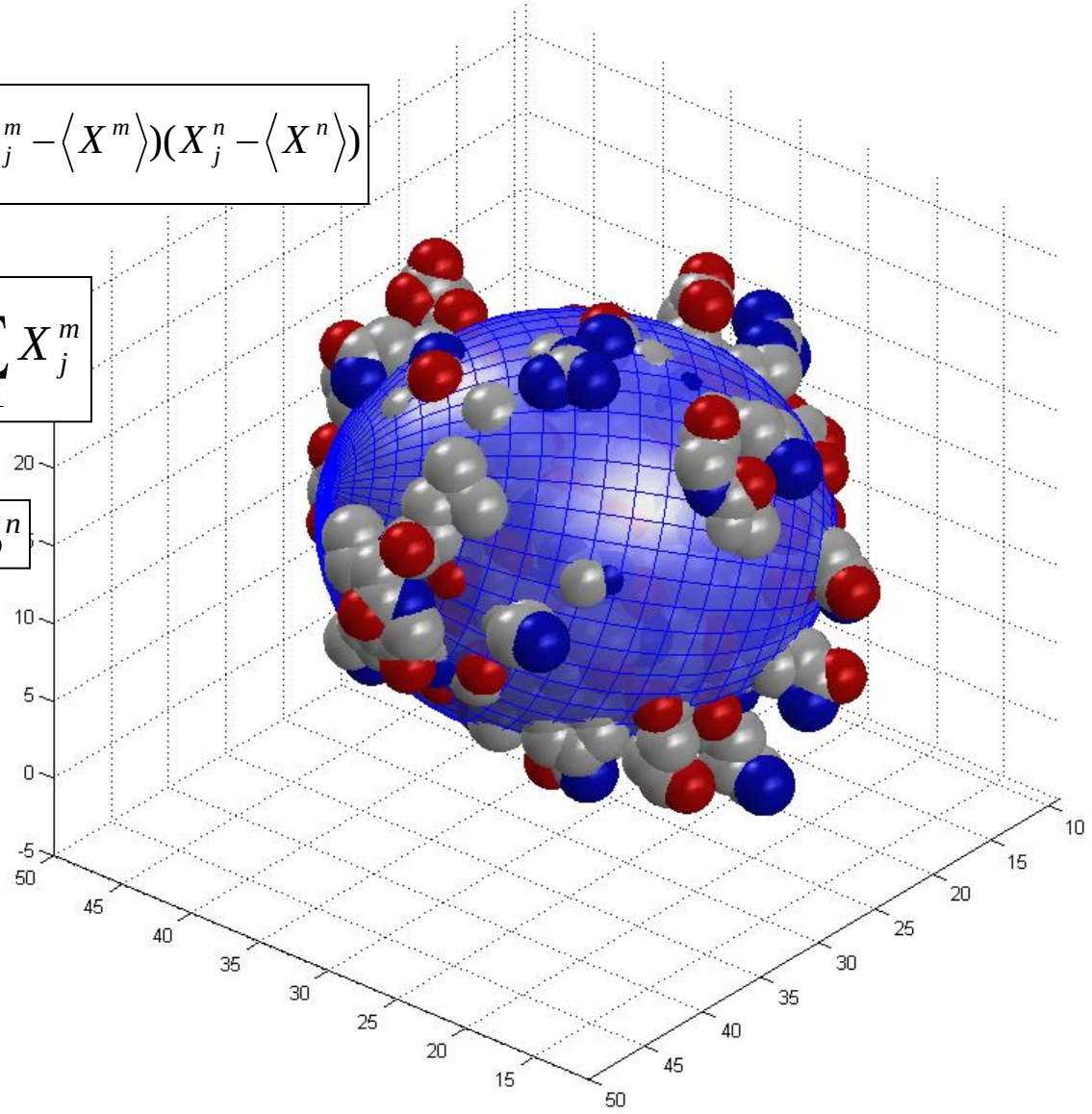
Build equivalent ellipsoid
Principal Component Analysis (PCA)

$$Cov_{m,n} = \frac{1}{N}\sum_{j=1}^{N}(X_j^m - \langle X^m \rangle)(X_j^n - \langle X^n \rangle)$$

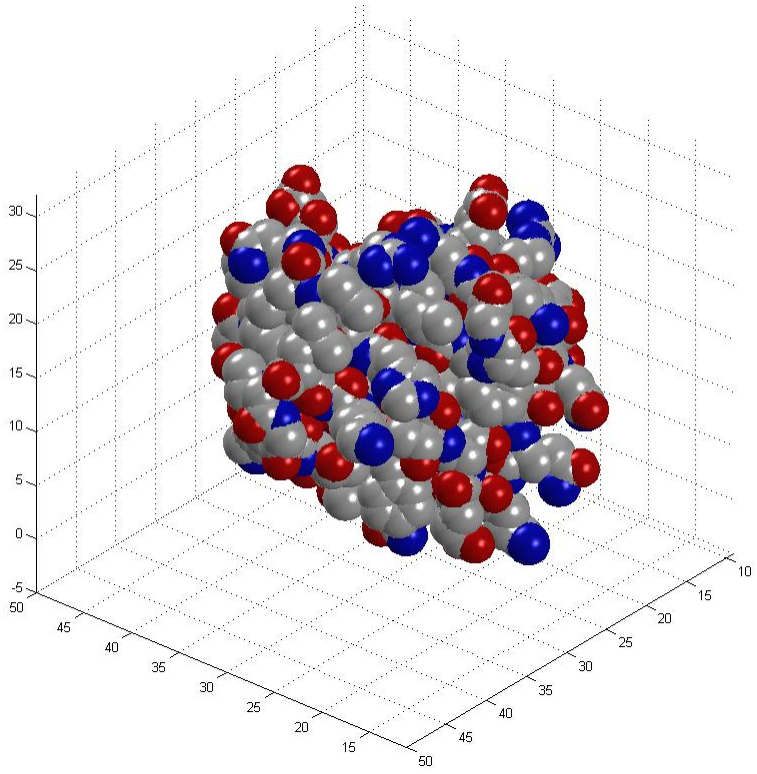
$$< X^m >= \frac{1}{N}\sum_{j=1}^{N} X_j^m$$

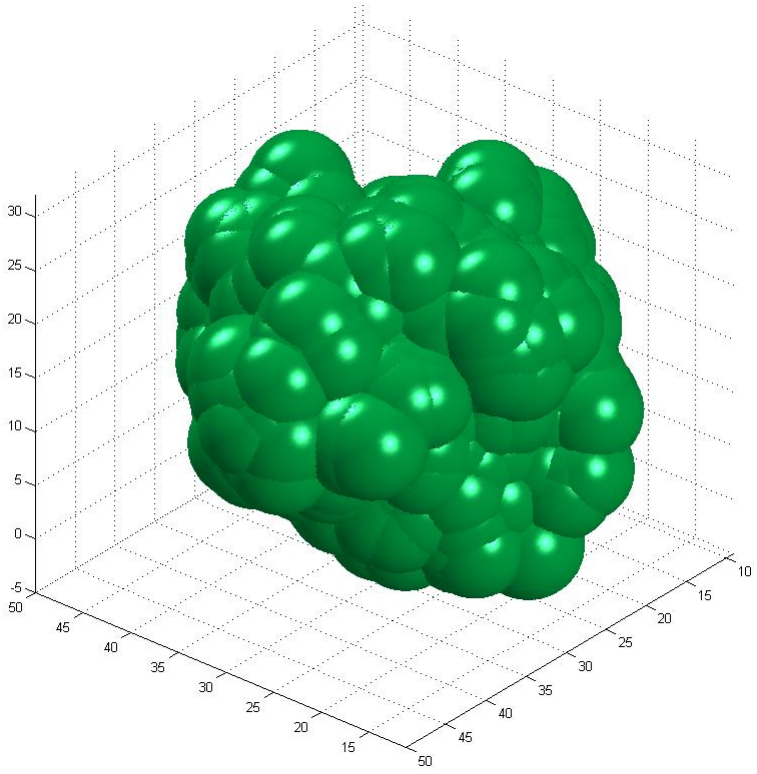
$$\mathbf{Cov}\,\mathbf{S}^n = E^n \mathbf{S}^n$$

$$a_n = \sqrt{3E^n}$$

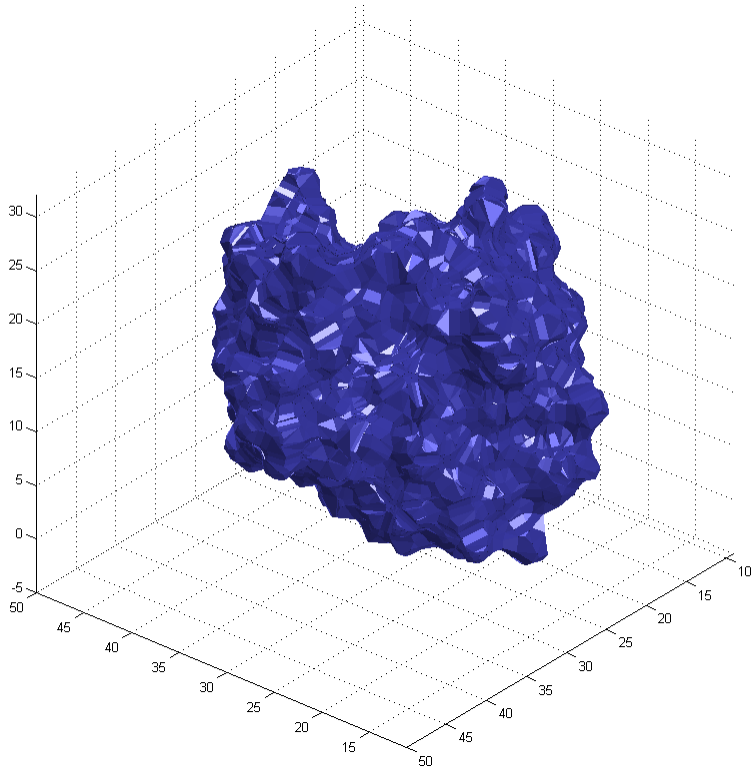


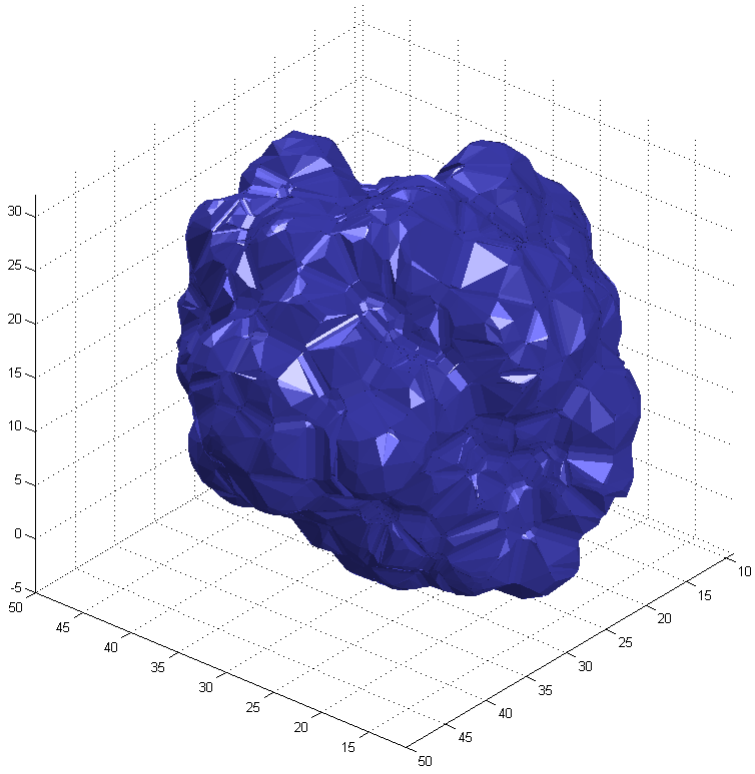
Hydration shell



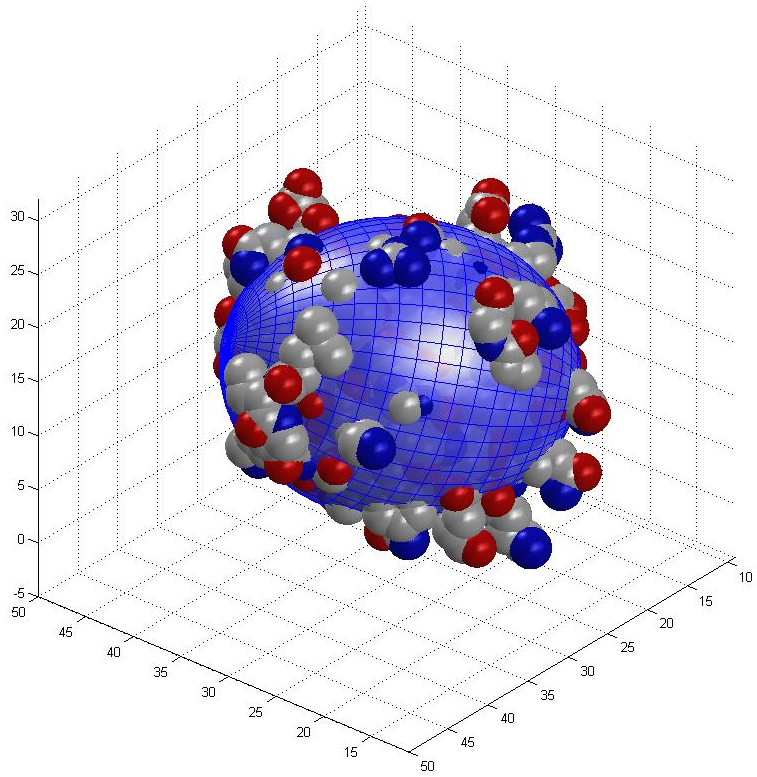


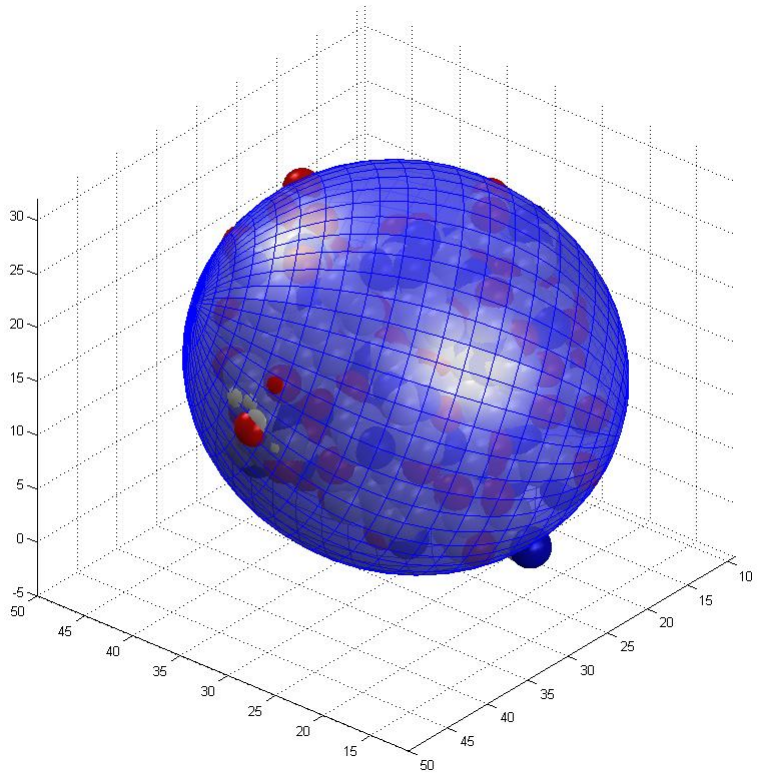
Hydration shell





Hydration shell
Equivalent ellipsoid is approximately twice bigger





ELM Algorithm

Build equivalent ellipsoid with PCA

$$Cov_{m,n} = \frac{1}{N}\sum_{j=1}^{N}(X_j^m - \langle X^m \rangle)(X_j^n - \langle X^n \rangle) \qquad < X^m >= \frac{1}{N}\sum_{j=1}^{N} X_j^m$$

$$a_n = \sqrt{3E^n} \qquad \mathbf{Cov}\,\mathbf{S}^n = E^n\mathbf{S}^n$$

Evaluate diffusion tensor components with Perrin's Equations

$$D_l = \frac{kT}{C_l} \qquad C_x = \frac{16\pi\eta(a_y^2 + a_z^2)}{3(a_y^2 Q + a_z^2 R)} \qquad C_y = \frac{16\pi\eta(a_x^2 + a_z^2)}{3(a_z^2 R + a_x^2 P)} \qquad C_z = \frac{16\pi\eta(a_x^2 + a_y^2)}{3(a_x^2 P + a_y^2 Q)}$$

$$P = \int_0^\infty \frac{ds}{\sqrt{(a_x^2+s)^3(a_y^2+s)(a_z^2+s)}} \qquad Q = \int_0^\infty \frac{ds}{\sqrt{(a_y^2+s)^3(a_z^2+s)(a_x^2+s)}} \qquad R = \int_0^\infty \frac{ds}{\sqrt{(a_z^2+s)^3(a_x^2+s)(a_y^2+s)}}$$

Perrin *J. Phys. Radium* (1934, 1936)

Complexity of the algorithms

ELM N_{at}

HYDRONMR N_{at}^2

ELM : HYDRONMR

1 : 500

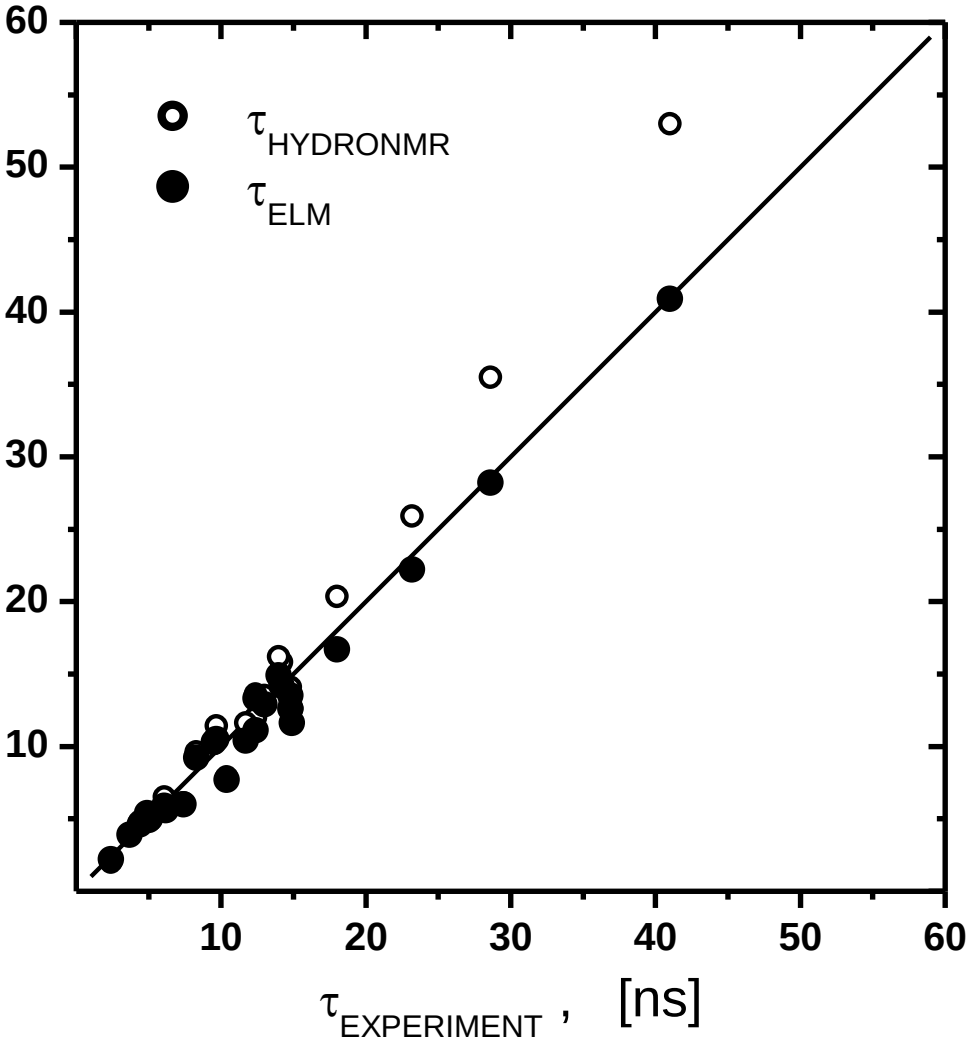
Comparison with experimental data

proteins from 2.9 to 82 kDa

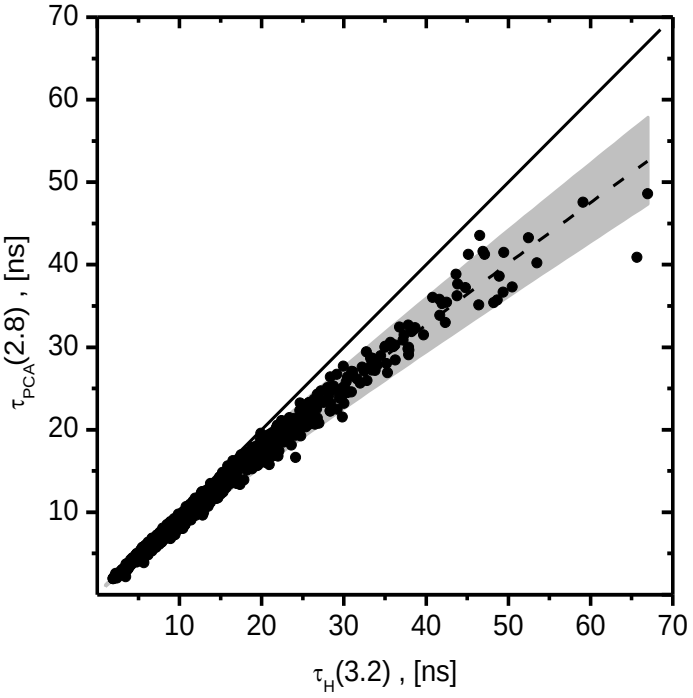
protein	PDB code	experiment[b] τ_c exp. [ns]	PCA variable HLT HLT Å	PCA variable HLT τ_c calc. [ns]	HYDRONMR AER = 3.2 Å τ_c calc. [ns]	HYDRONMR AER = 3.2 Å diff. %	PCA HLT = 2.8 Å τ_c calc. [ns]	PCA HLT = 2.8 Å diff. %	PCA HLT = 0.0 Å $2\tau_c$ calc. [ns]	PCA HLT = 0.0 Å diff. %
malate synthase G[67]	1y8b	55.3 (N)	4.6	55.38	54.6	−1	45.7	−17	47.7	−14
human serum albumin[68]	1ao6	41.0 (F)	2.8	40.87	53.0	29	40.9	0	45.7	11
maltose binding protein[69]	1ezp	28.6 (N)	2.9	28.65	35.5	24	28.2	−1	27.6	−3
β-lactoglobulin a (dimer)[61]	1bsy	23.2 (F)	3.1	23.13	25.9	12	22.2	−4	23.0	−1
Δ^5-3 ketosteroid isomerase[70]	1buq	18.0 (F,N)	3.2	17.96	20.4	13	16.7	−7	17.6	−2
leukemia inh. factor [44]	1lki	14.9 (N)	4.4	14.93	12.6	−15	11.6	−22	11.8	−21
trypsin[61]	2blv	14.8 (F)	4.1	14.76	13.0	−12	12.6	−15	14.4	−3
yellow fluorescent protein[71]	2yfp	14.8 (F)	3.4	14.9	14.1	−5	13.5	−9	13.6	−8
green fluorescent protein[72]	1w7s	14.2 (F)	2.8	14.24	15.8	11	14.2	0	14.6	3
carbonic anhydrase[61]	2cab	14.0 (F)	2.4	13.9	16.2	16	14.9	6	16.5	18
HIV-1 protease[73]	1bvg	13.0 (N)	2.8	12.94	13.5	4	12.9	−1	14.2	9
savinase[44]	1svn	12.4 (N)	2.3	12.38	13.6	10	13.3	7	15.2	23
interleukin-1β[44]	6i1b	12.4 (N)	3.5	12.43	12.0	−3	11.1	−10	11.4	−8
ribonuclease H[58]	2rn2	11.7 (N)	3.5	11.68	11.6	−1	10.4	−11	11.3	−3
cytochrome c_2[74]	1c2n [c]	10.4 (N)	4.6	10.40	7.9	−24	7.7	−26	7.5	−28
β-lactoglobulin A (mono)[61]	1bsy	9.7 (F)	2.4	9.79	11.4	18	10.5	8	10.7	10
apomyoglobin[61]	1bvc	9.5 (F)	2.4	9.55	10.2	9	10.3	10	10.5	12
lysozyme[44]	1hwa	8.3 (N)	2.2	8.26	9.6	16	9.2	11	9.2	11
barstar C40/83A[44]	1bta	7.4 (N)	3.9	7.33	6.1	−18	6.0	−19	6.1	−18
eglin c[44]	1egl [d]	6.2 (N)	3.4	6.23	6.0	−3	5.6	−10	5.4	−13
cytochrome b_5[44]	1wdb	6.1 (N)	3.0	6.10	6.5	7	5.9	−3	5.4	−11
calbindin-D9k+Ca^{2+} [44]	2bca	5.1 (N)	2.9	5.07	5.1	0	5.0	−2	4.9	−4
ubiquitin[14]	1ubq [e]	5.0 (N)	2.9	4.98	5.0	0	4.9	−2	4.8	−4
calbindin-D9k[44]	1clb [d]	4.9 (N)	2.3	4.88	5.2	6	5.4	10	5.3	8
BPTI[44]	1pit [d]	4.4 (N)	2.6	4.41	4.8	9	4.6	5	4.7	7
Protein G[9]	1igd [f]	3.7 (N)	2.6	3.74	3.9	5	3.9	5	3.4	−8
Xfin-zinc finger DBD[44]	1znf [d]	2.4 (N)	3.2	2.37	2.0	−17	2.2	−8	1.9	−21
mean abs. value			3.1(0.7)[g]			11%		8%		10%

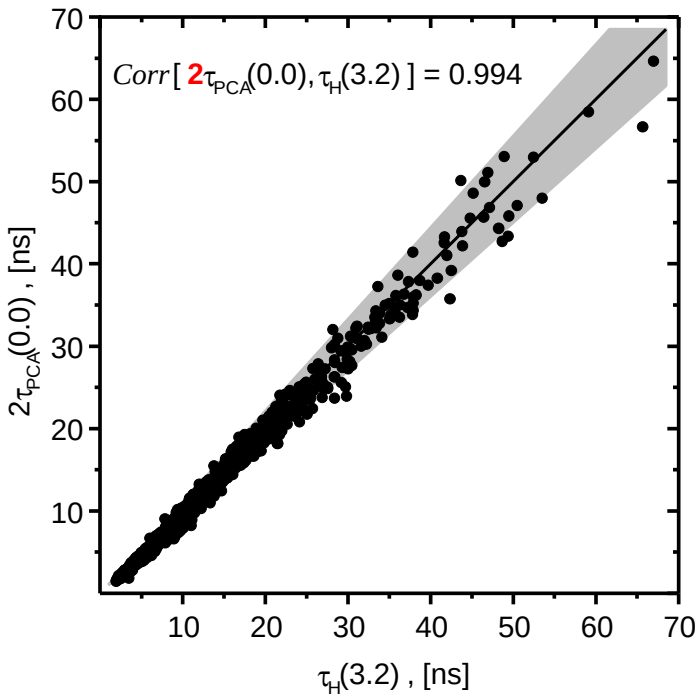
Ryabov & Fushman, JACS (2006)

Comparison with the experimental data



Correlation times for 841 protein structures
Hydration shell effect

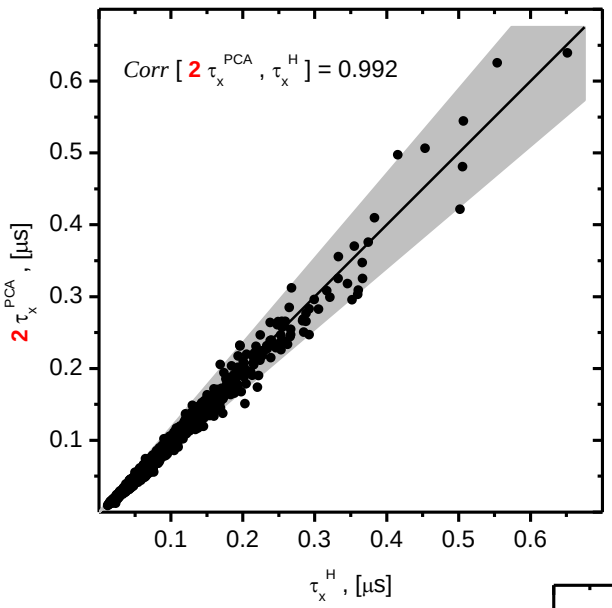


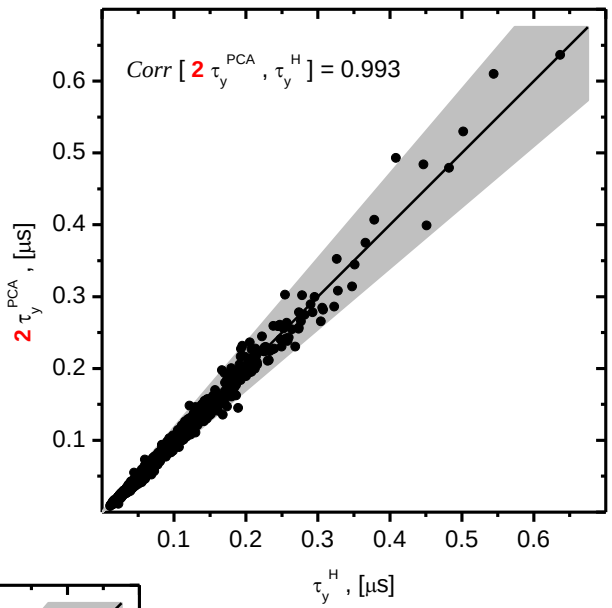


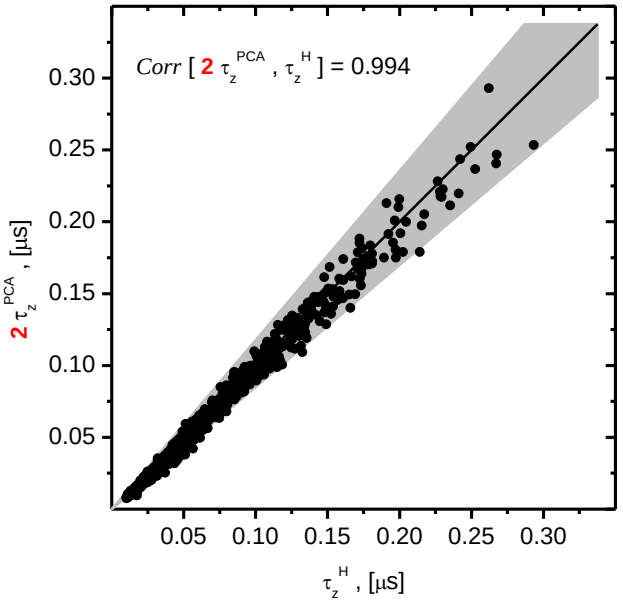
Power law $\tau_{PCA} = M\tau_H^q$ $q \sim 0.923$

Fractal surface dimension $d_f = 2/q$ $\sim 2.2 \div 2.3$

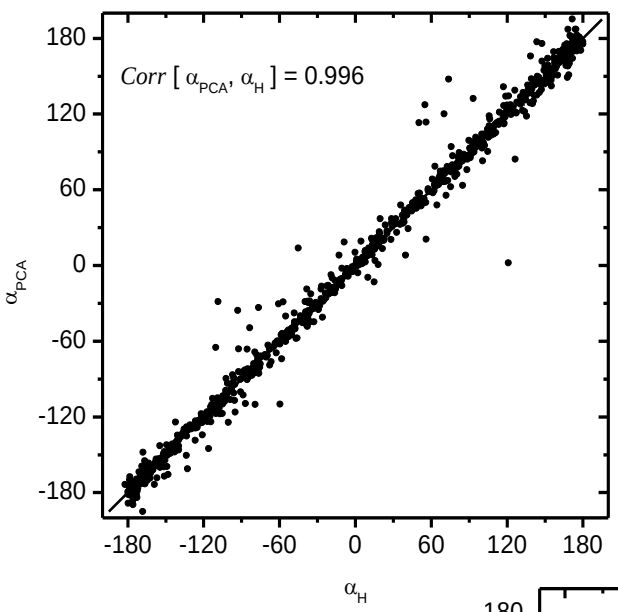
Correlations for diffusion tensor components

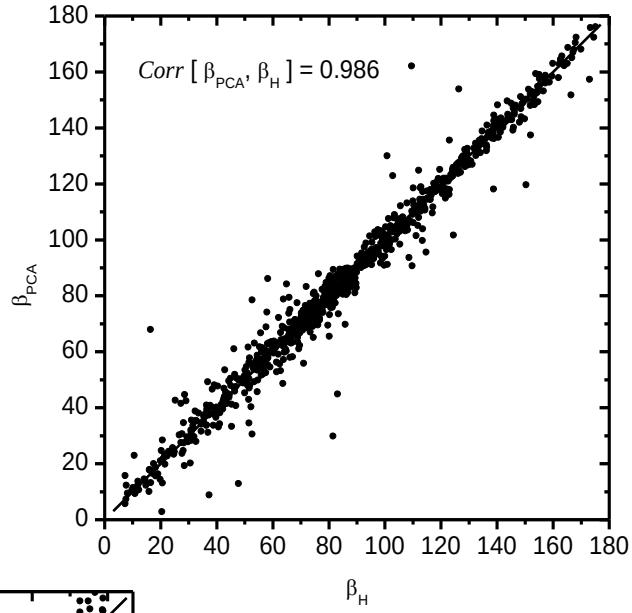


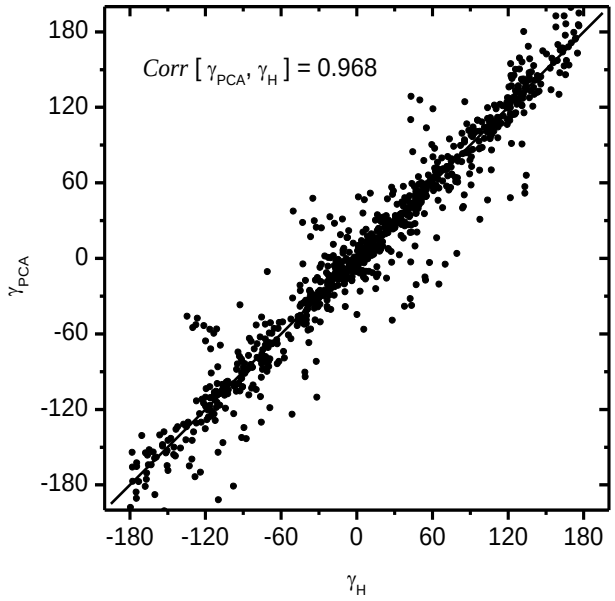




Correlations of diffusion tensors orientations







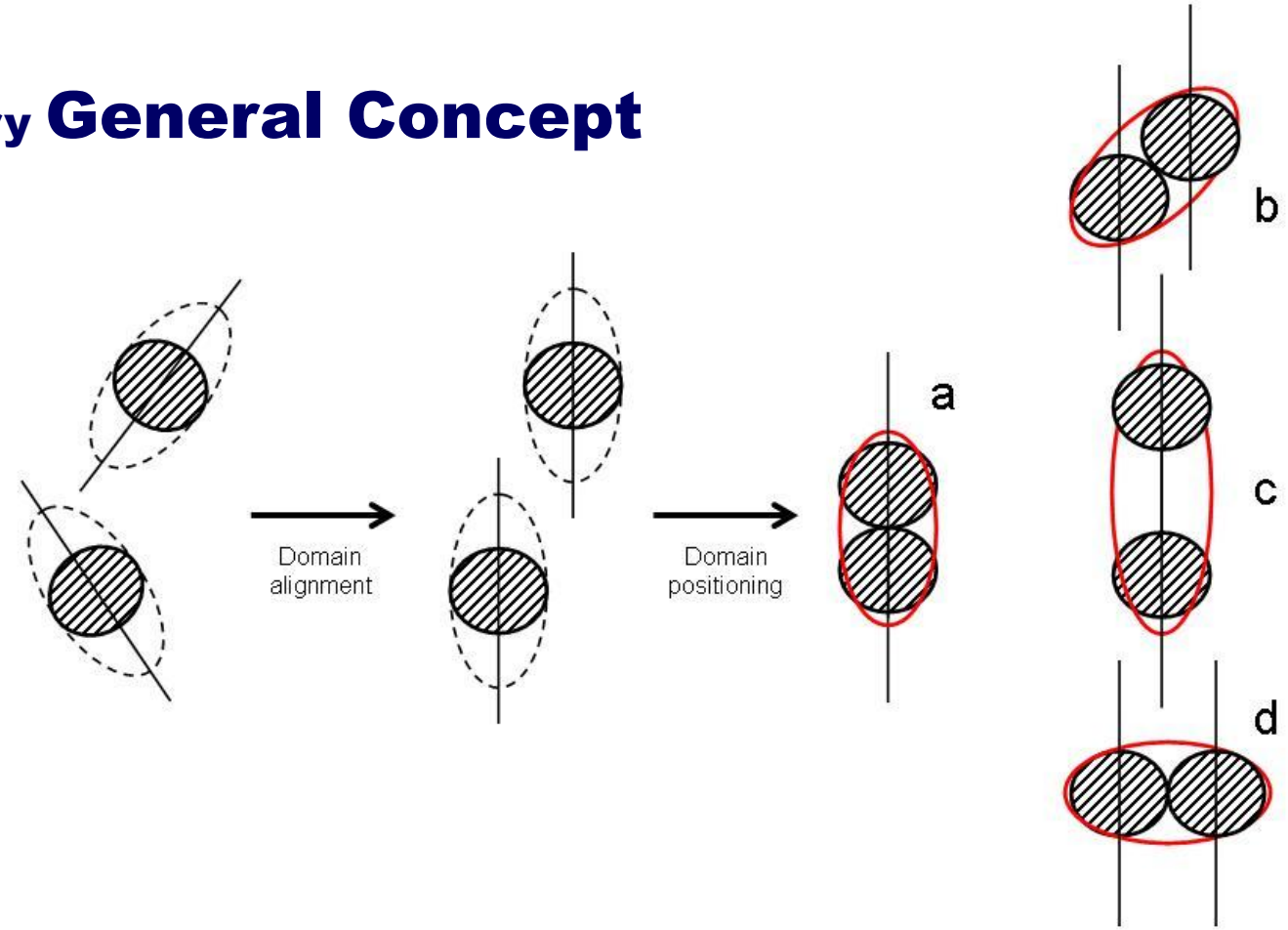
Summary

- ELM provides at least the same level or accuracy and precision in description of experimental data as the other approaches, which represent protein surface with large number of friction elements (HYDRONMR). This results 500 times speed up in calculation time
- Hydration shell makes apparent diffusion correlation time of real proteins approximately twice longer
- Most of the diffusion tensors of real protein structures analyzed in our work are axially symmetric. In general rhombicities of diffusion tensors for real proteins are small and below the precisions of existing calculation methods predicting protein diffusion properties

Assembling Structures of Multi-Domain Proteins and Protein Complexes

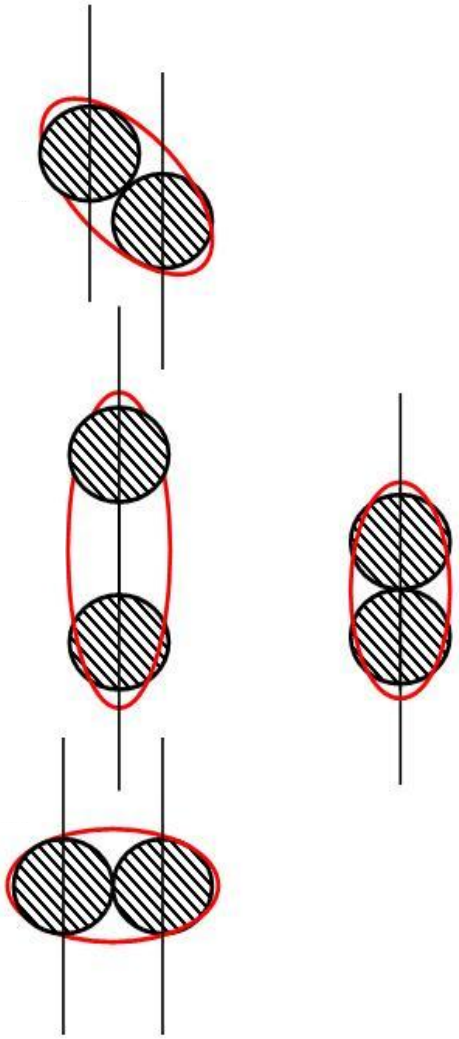
guided by protein diffusion tensor

a very General Concept



Ryabov & Fushman, JACS (2006)

Positioning Procedure



Matches components of diffusion tensor

Minimizing $$\chi^2 = \sum_{\substack{i=1,3 \\ j=i,3}} \left(D_{i,j}^{calc} - D_{i,j}^{exp} \right)^2$$

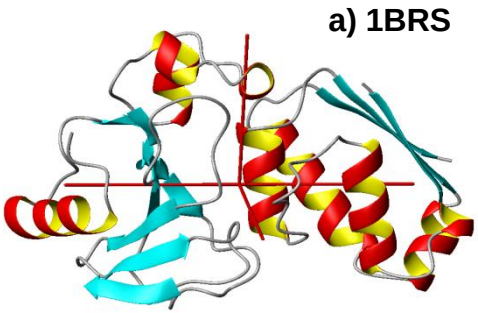
Use to be impossible with
bead and shell algorithms like HYDRONMR

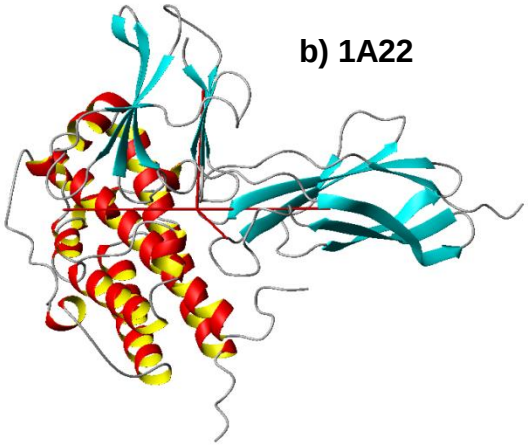
NOW possible with ELM

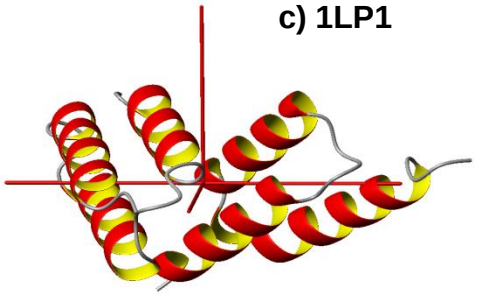
Tests for Proteins with Known Structure

and Predicted Diffusion Tensor

Fit with components of Diffusion tensor Predicted with ELM







$\chi^2 = 7.5 * 10^{-8}$

RMSD=0.0034 [A]

$\chi^2 = 4.4 * 10^{-9}$

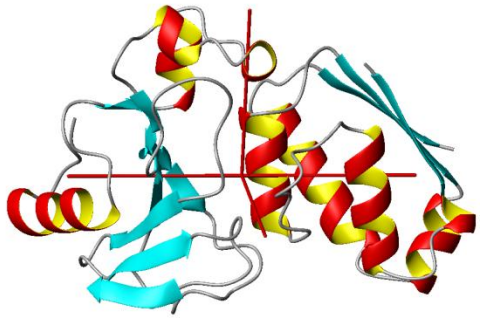
RMSD=0.0085 [A]

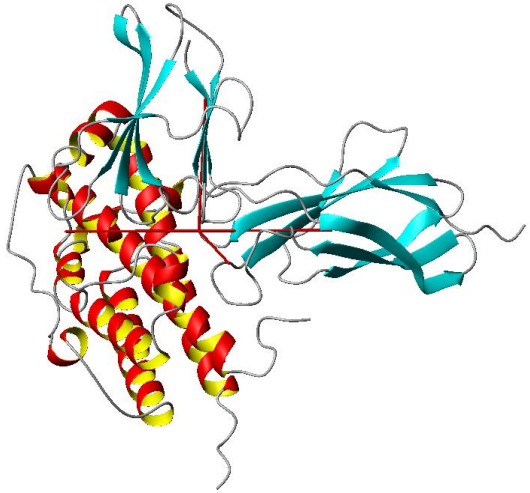
$\chi^2 = 6.1 * 10^{-9}$

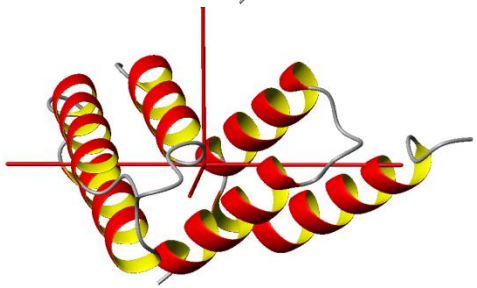
RMSD=0.0008 [A]

Tests for Proteins with Known Structure

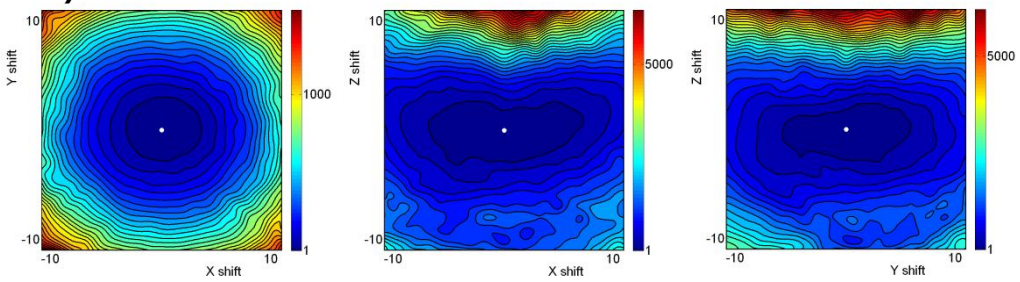
and Predicted Diffusion Tensor : Mapping χ^2 space



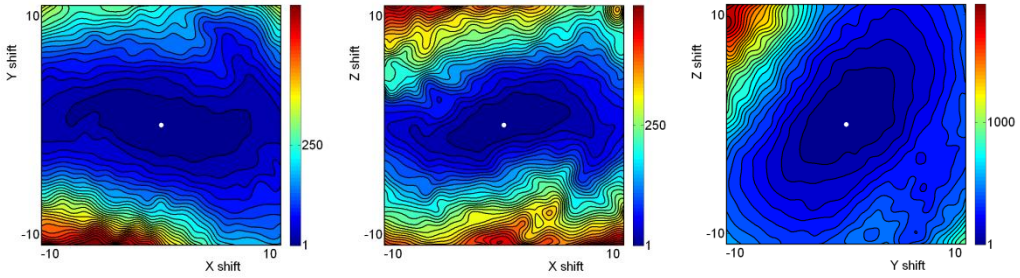




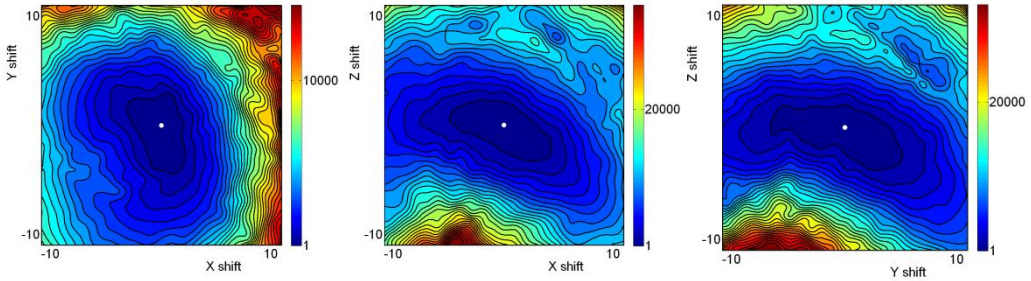
a) 1BRS



b) 1A22

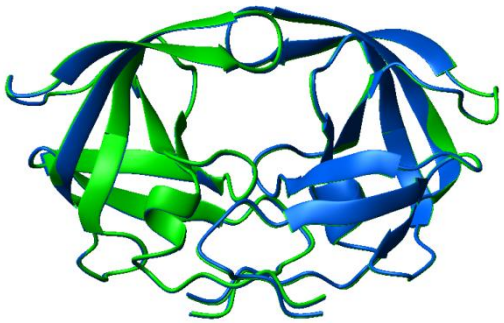


c) 1LP1

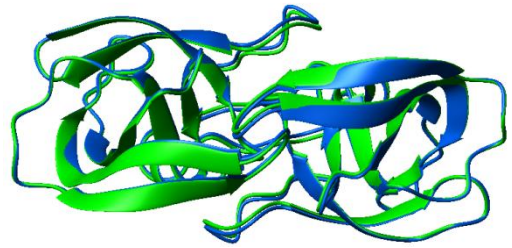


Tests for Proteins with Known Structure

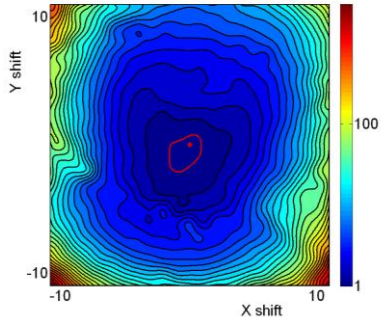
and Experimental Diffusion Tensor : HIV-1 protease

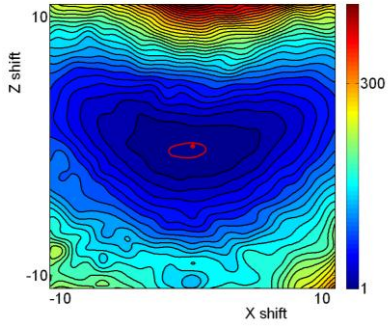


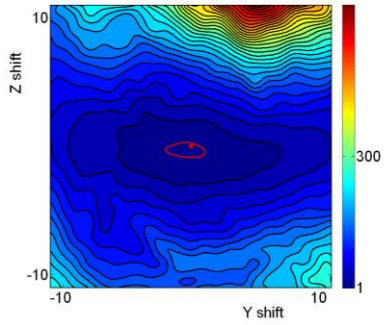
RMSD=0.36 [A]



$X \in [-1.8;1]$
$Y \in [-2;1]$
$Z \in [-1;0.3]$

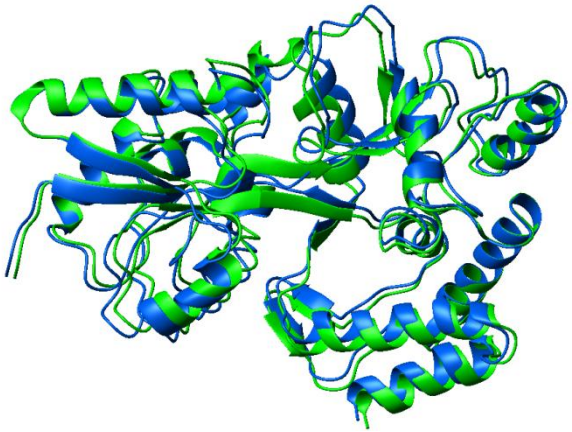




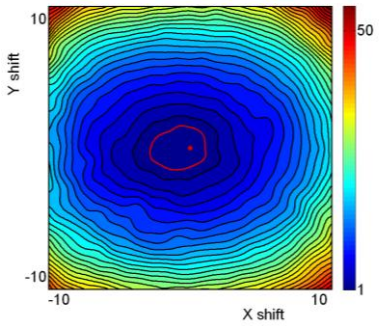


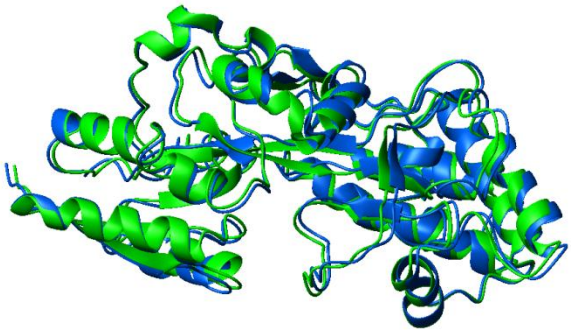
Tests for Proteins with Known Structure

and Experimental Diffusion Tensor : MBP

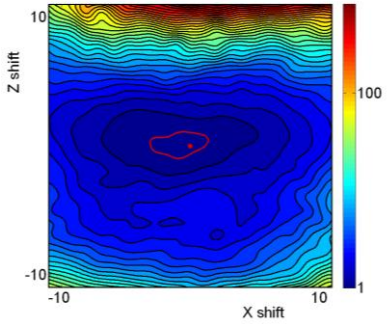


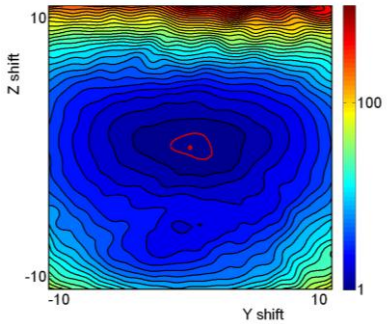
RMSD=1.34 [A]





$X \in [-3.1;1.3]$
$Y \in [-1.7;1.7]$
$Z \in [-1;1.1]$

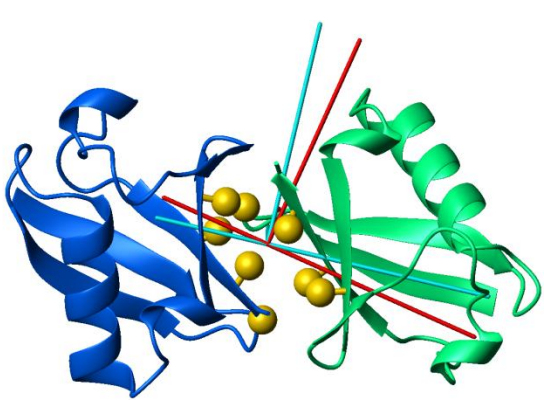


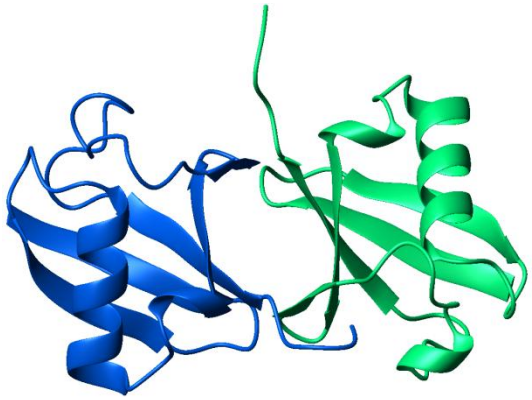


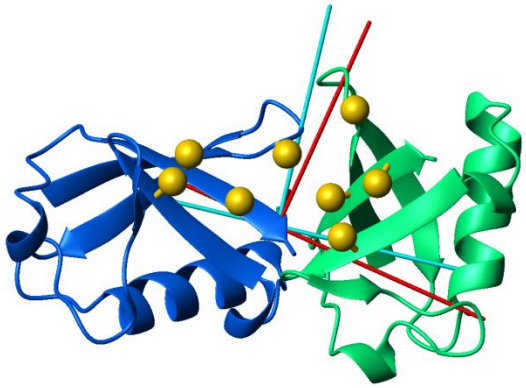
Protein with UnKnown Structure

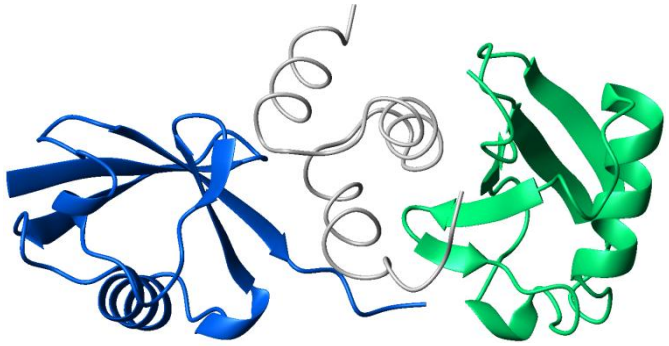
$X \in [-2.5;4]$
$Y \in [-3.9; 3.7]$
$Z \in [-2;1.9]$

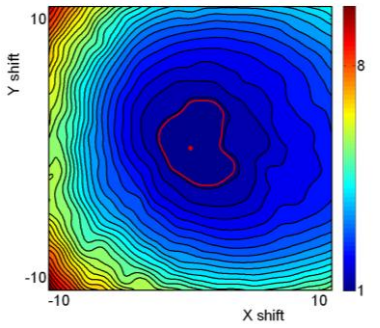
Ub_2

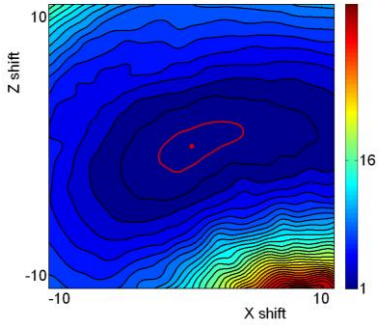


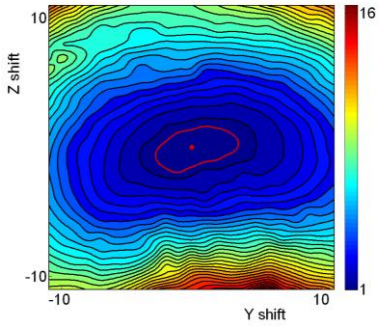












Conclusions

- Diffusion tensor can provide significant portion of new information encoded in principal values of its components. This information can be successfully used for assembling of multi domain protein structures

- Tested for a number of examples the suggested procedure of multi-domain proteins assembling showed very high performance and accuracy

- The method provides the unique possibility to assemble dynamic protein structures when traditional X-ray, NMR NOE based, or docking methods are inappropriate

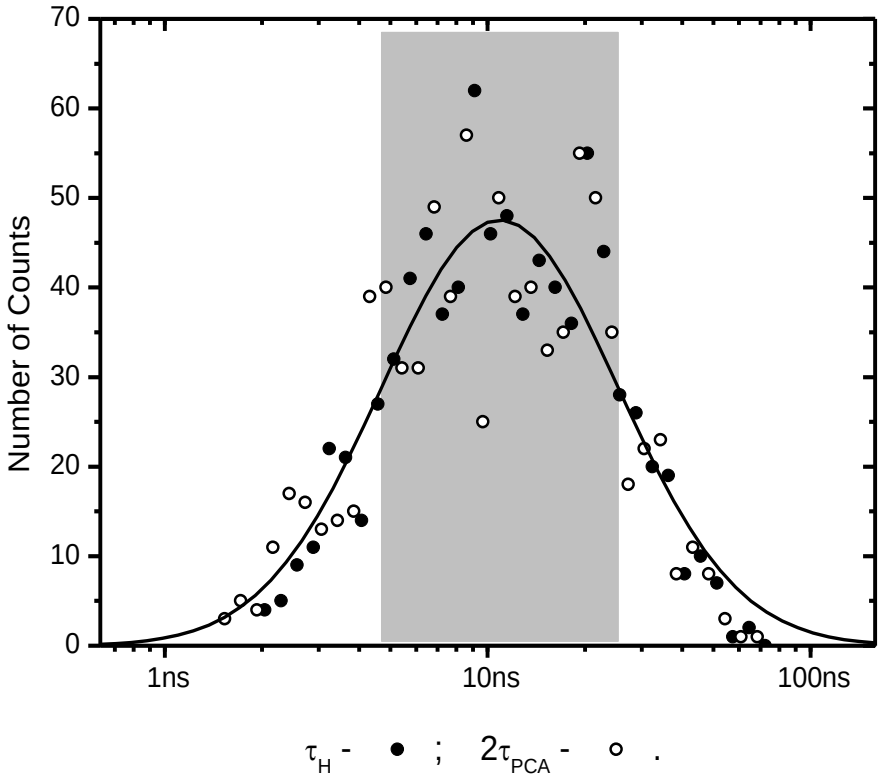
- Method shall be further developed to take explicitly into account rotational degrees of freedom, potentials of intermolecular interaction and improved method for evaluation the hydration shell effect

Genome Evolution

from statistics of proteins' properties to Exon size distribution

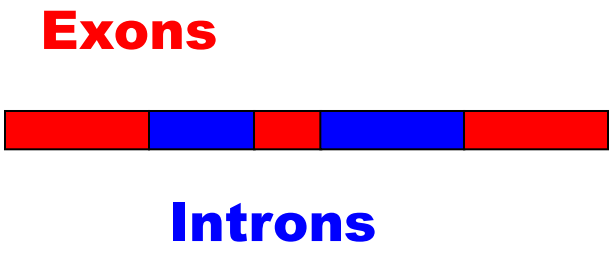
τ obeys Lognormal Distribution

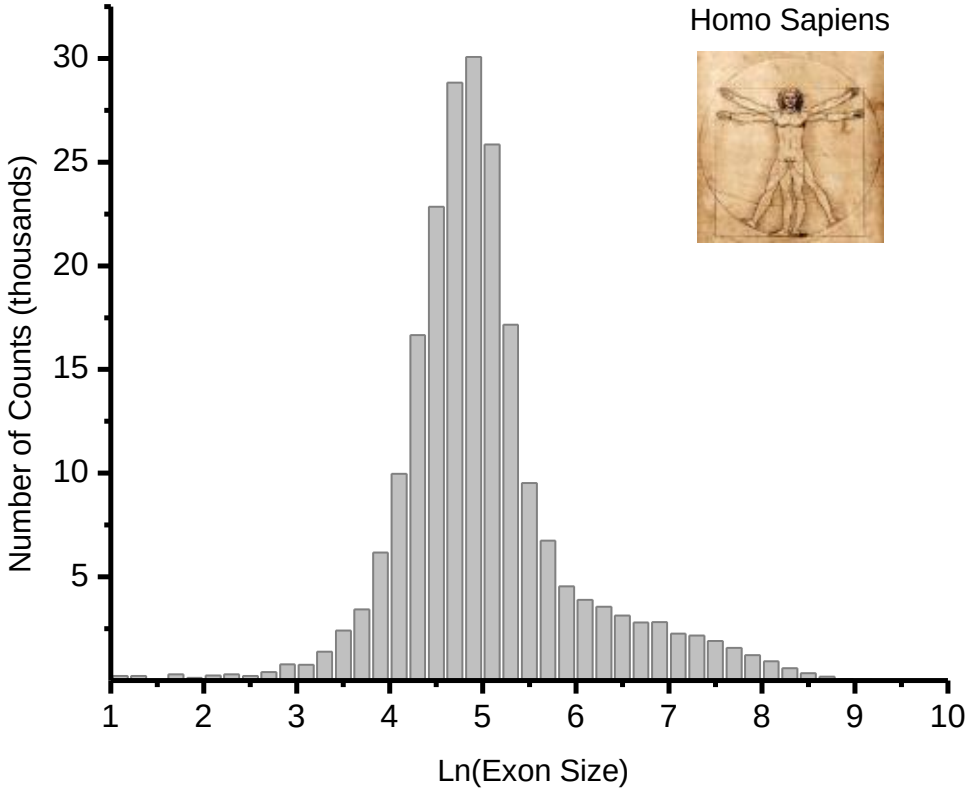
Why ?



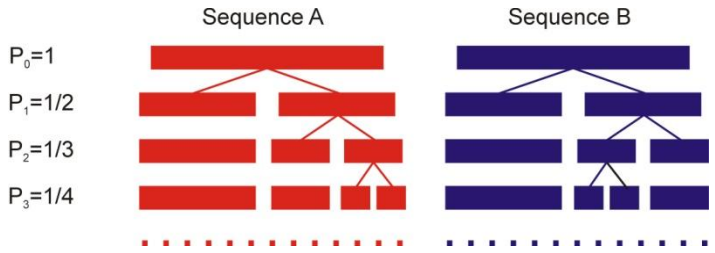
Hypothesis

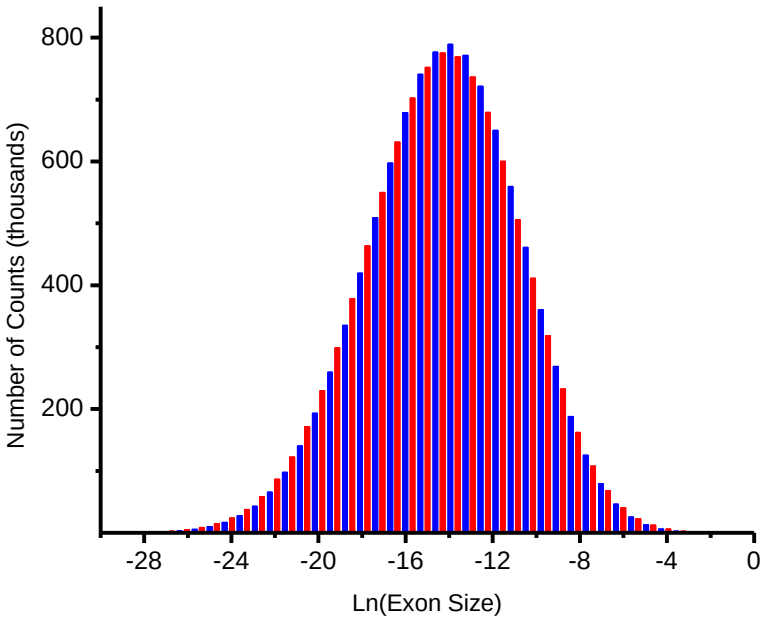
τ ~ Mass ~Length ~ Gene





Kolmogoroff process (1941)

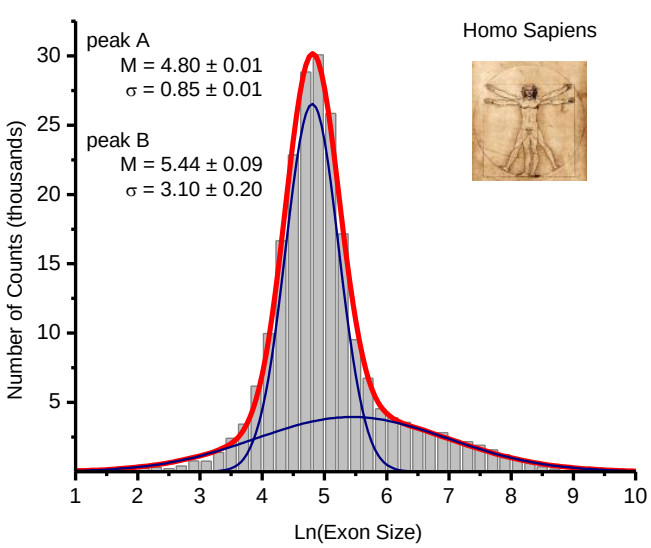


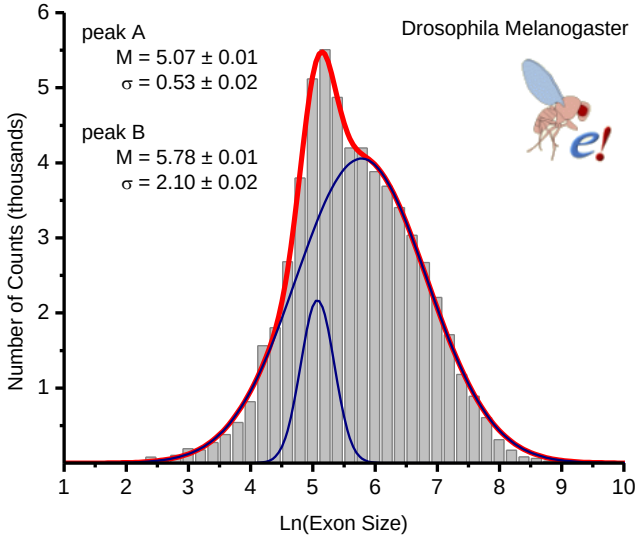


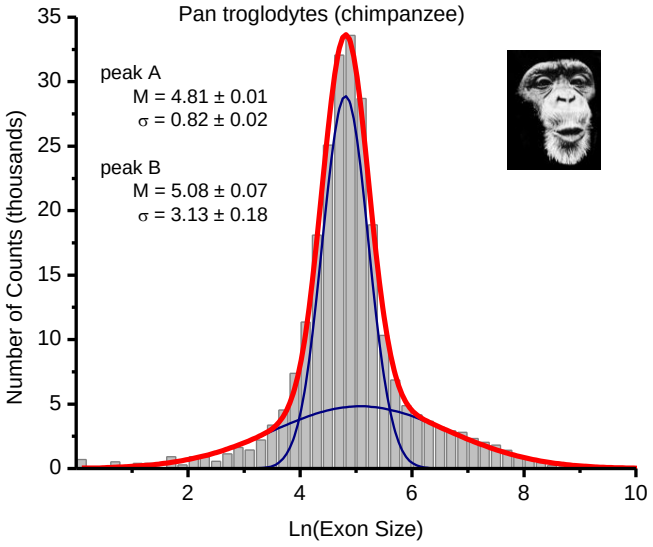
$$\ln E - M \sim \ln N \sim t$$

$$\sigma \sim \sqrt{\ln N} \sim \sqrt{t}$$

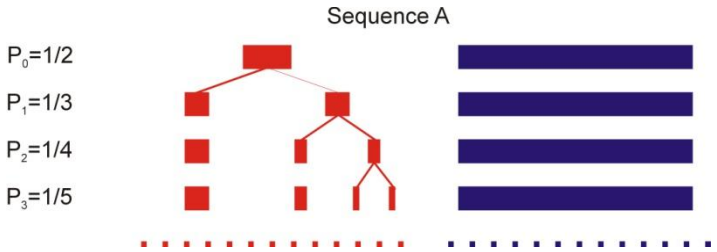
Real Genomes: Two peaks Pattern

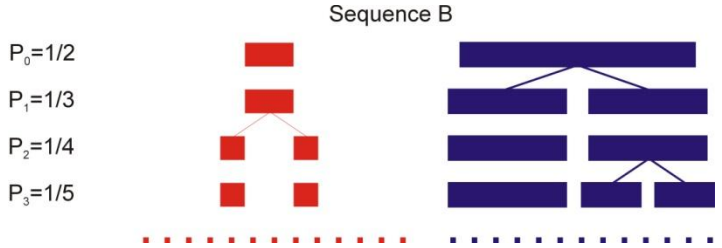


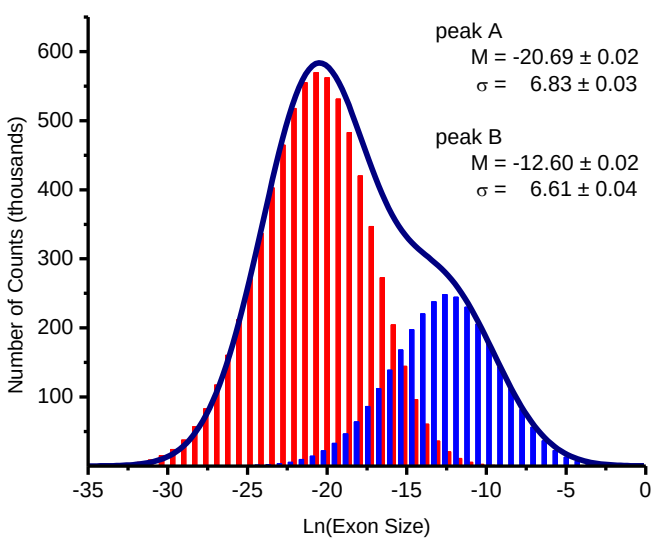


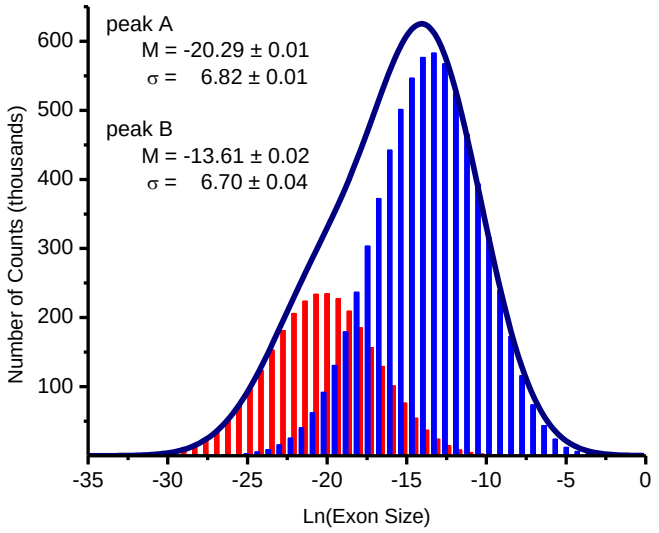


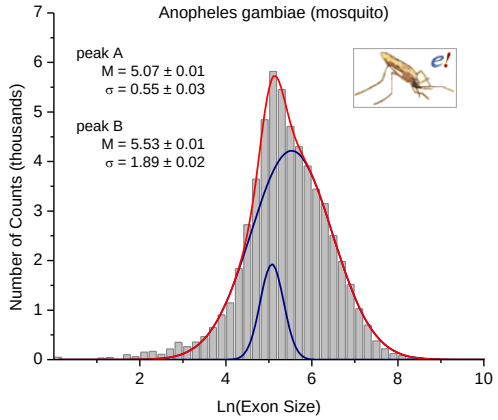
Two peaks Pattern from Soft Bifurcation

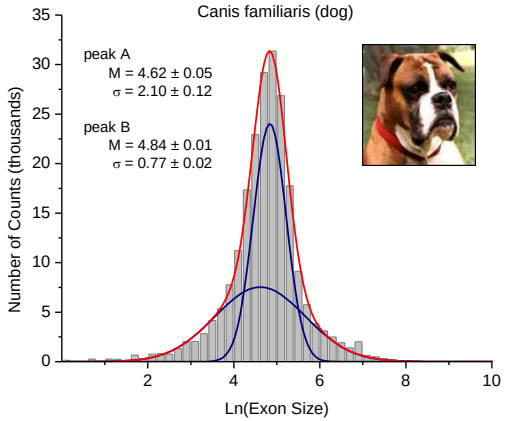


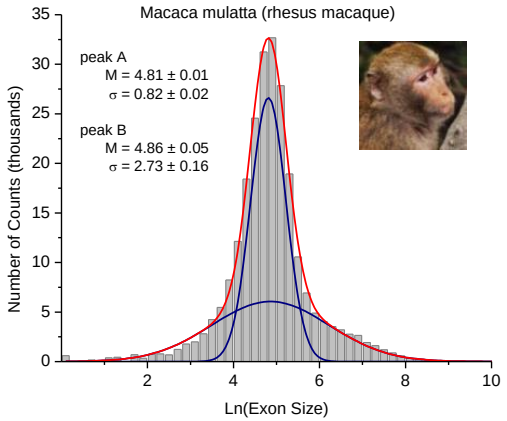


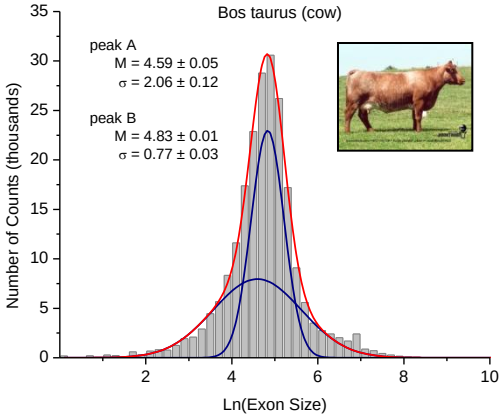


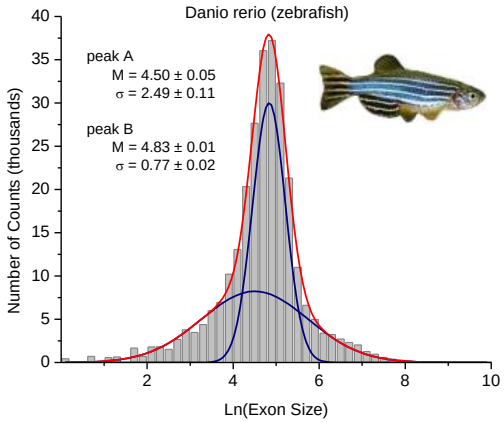


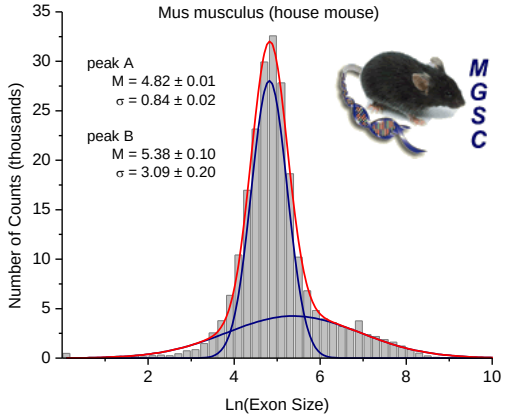

Anopheles gambiae (mosquito)
peak A
M = 5.07 ± 0.01
σ = 0.55 ± 0.03
peak B
M = 5.53 ± 0.01
σ = 1.89 ± 0.02
Number of Counts (thousands)
Ln(Exon Size)

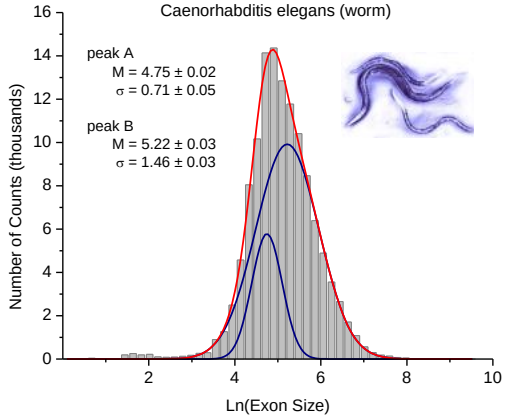

Canis familiaris (dog)
peak A
M = 4.62 ± 0.05
σ = 2.10 ± 0.12
peak B
M = 4.84 ± 0.01
σ = 0.77 ± 0.02
Number of Counts (thousands)
Ln(Exon Size)

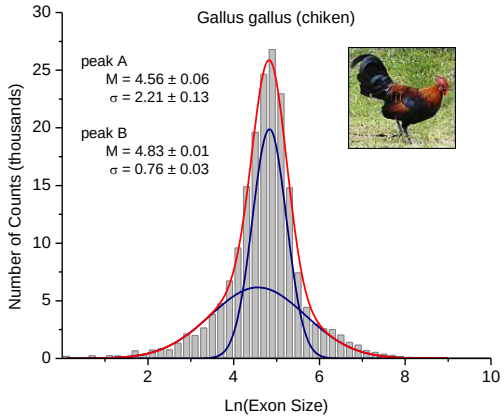

Macaca mulatta (rhesus macaque)
peak A
M = 4.81 ± 0.01
σ = 0.82 ± 0.02
peak B
M = 4.86 ± 0.05
σ = 2.73 ± 0.16
Number of Counts (thousands)
Ln(Exon Size)

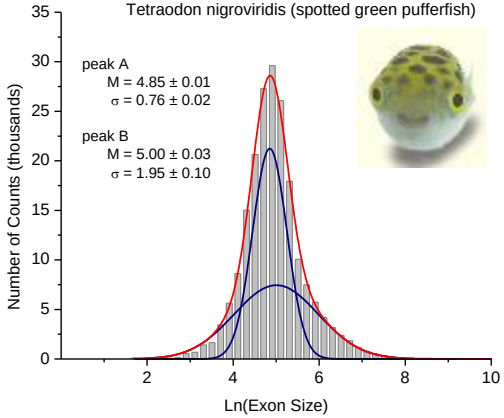

Bos taurus (cow)
peak A
M = 4.59 ± 0.05
σ = 2.06 ± 0.12
peak B
M = 4.83 ± 0.01
σ = 0.77 ± 0.03
Number of Counts (thousands)
Ln(Exon Size)


Danio rerio (zebrafish)
peak A
M = 4.50 ± 0.05
σ = 2.49 ± 0.11
peak B
M = 4.83 ± 0.01
σ = 0.77 ± 0.02
Number of Counts (thousands)
Ln(Exon Size)

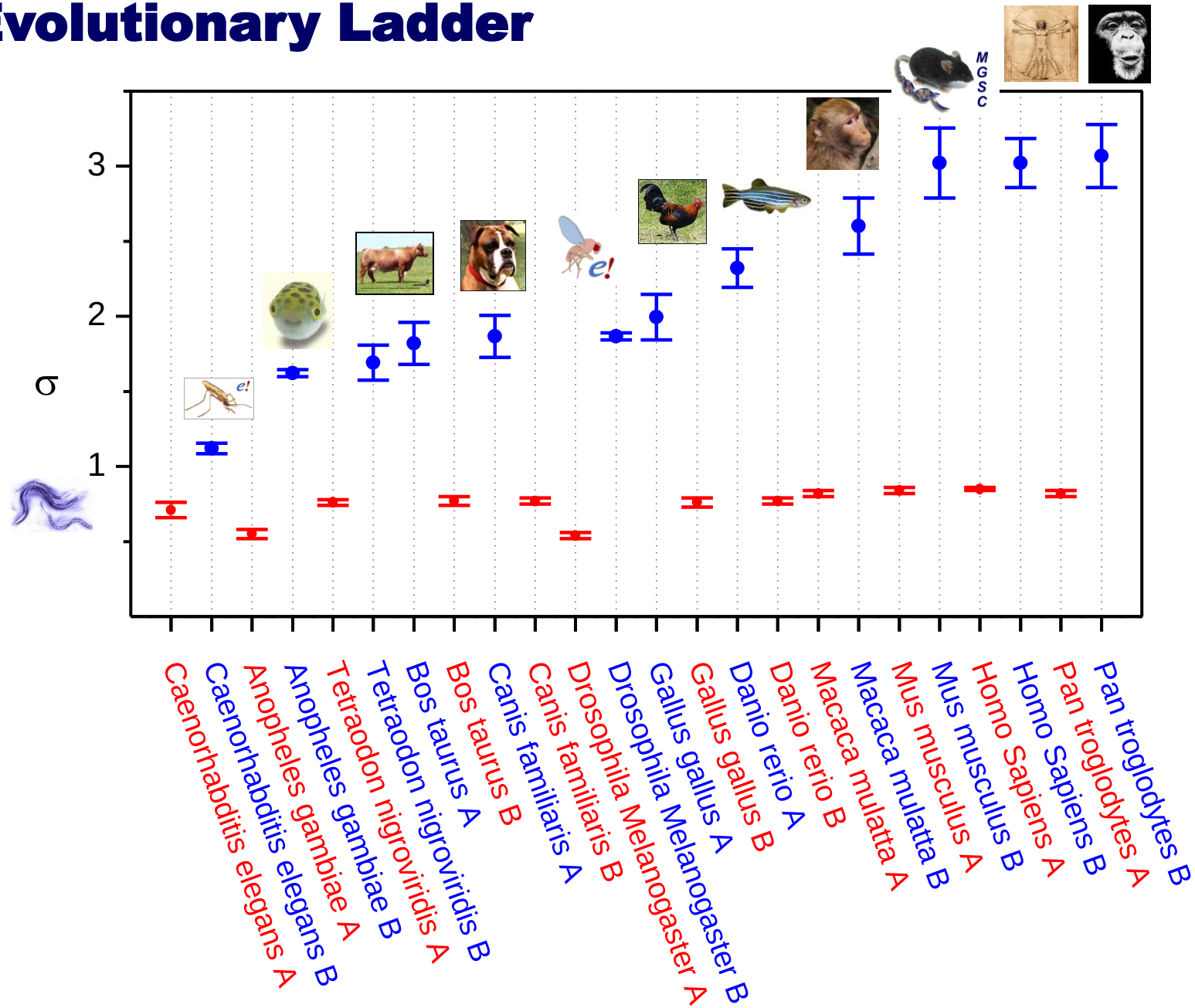

Mus musculus (house mouse)
peak A
M = 4.82 ± 0.01
σ = 0.84 ± 0.02
peak B
M = 5.38 ± 0.10
σ = 3.09 ± 0.20
MGSC
Number of Counts (thousands)
Ln(Exon Size)


Caenorhabditis elegans (worm)
peak A
M = 4.75 ± 0.02
σ = 0.71 ± 0.05
peak B
M = 5.22 ± 0.03
σ = 1.46 ± 0.03
Number of Counts (thousands)
Ln(Exon Size)

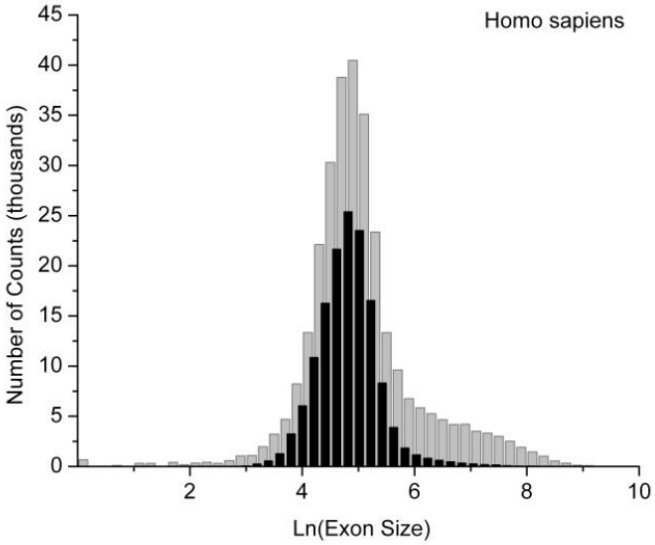

Gallus gallus (chiken)
peak A
M = 4.56 ± 0.06
σ = 2.21 ± 0.13
peak B
M = 4.83 ± 0.01
σ = 0.76 ± 0.03
Number of Counts (thousands)
Ln(Exon Size)

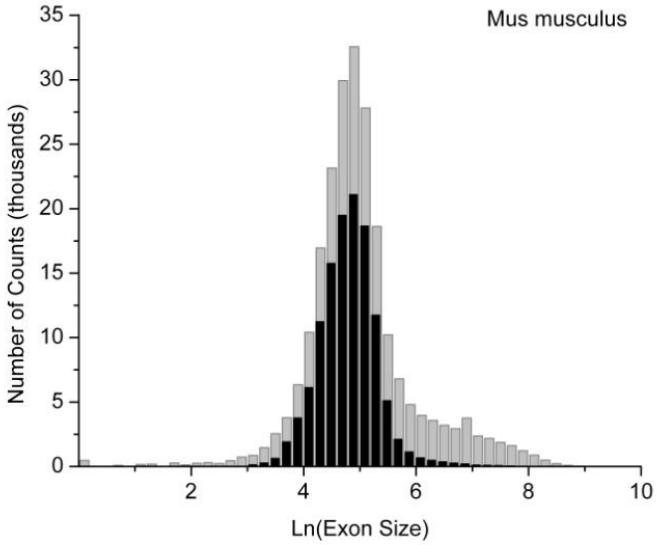

Tetraodon nigroviridis (spotted green pufferfish)
peak A
M = 4.85 ± 0.01
σ = 0.76 ± 0.02
peak B
M = 5.00 ± 0.03
σ = 1.95 ± 0.10
Number of Counts (thousands)
Ln(Exon Size)

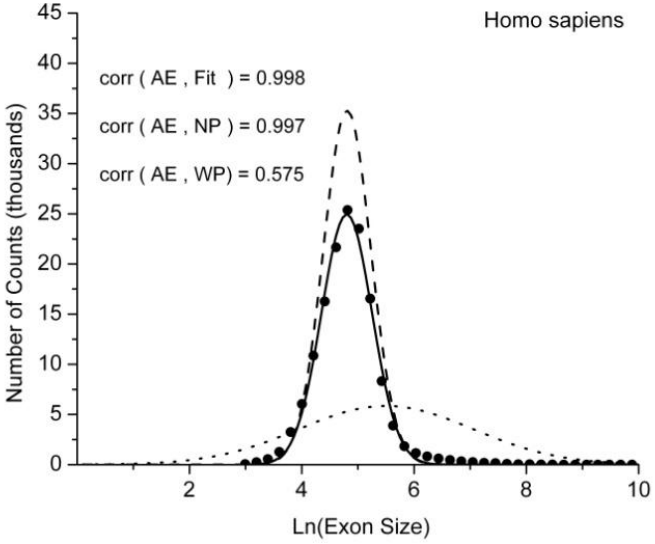
Evolutionary Ladder

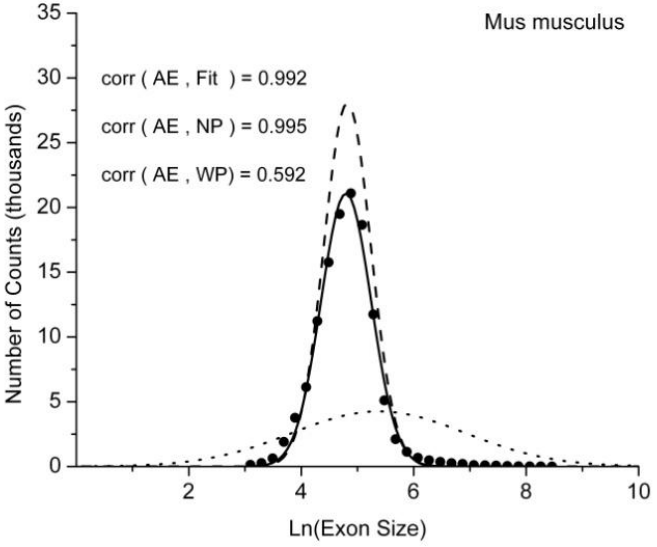


The role of alternative splicing









Conclusions

- Statistical distribution of exon sizes in general obey the log-normal law that can be originated from random Kolmogoroff fractioning process

- Concept of Kolmogoroff process suggests that the process of intron gaining is independent of exon size but, hypothetically, related to a mechanism dependent of intron-exon boundaries

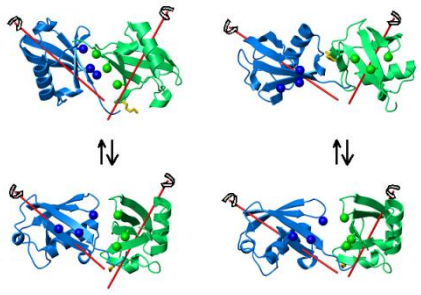
- The concept of random exon breaking supports the hypothesis that at the initial stages of evolution simpler organisms had lower fraction of introns – Introns Late Hypothesis

- Two distinctive classes of exons presented in exon length statistics are still not fully understood. However, it is clear that one of these classes contain exons that are in general conserved during evolution while another class undergoes intensive diversification

- Distribution of exons between these classes correlates with phenomenon of alternative splicing which is, presumably, responsible for the existing variety of living species.

Directions for Future Researches

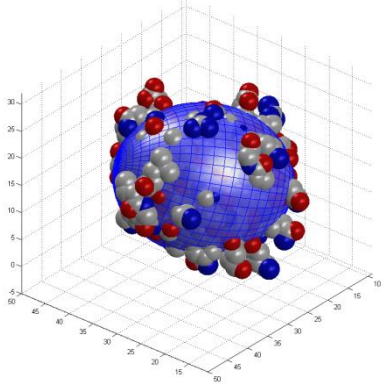
Protein Domain Dynamics



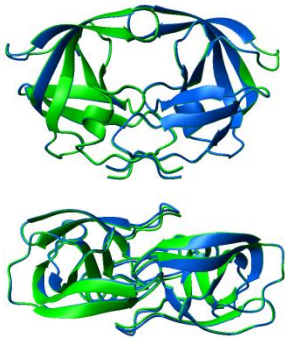
- Theoretical Models for the case of explicit coupling between overall tumbling and domain reorientations
- Search for the other possible applications like for the case of Insulin
- Molecular Dynamics investigations of domain mobility. The first target probe correlations between charged state of His 68 in UB_2 and domain dynamics for MD trajectories
- Elaboration of data treatment software for other than NMR experimental techniques

Directions for Future Researches

Predictions of Protein Diffusion Tensor

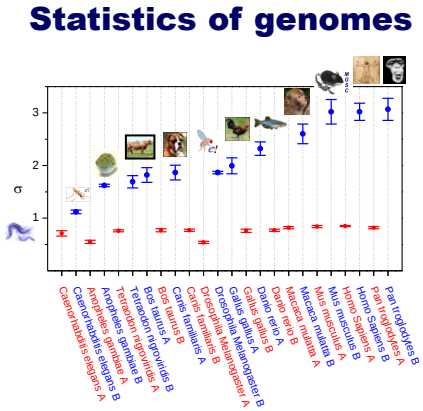


Assembling Protein Structures



- Improvements of the method to include exhaustive search for all degrees of freedom including rotation angles.
- Adding parameters of intermolecular interactions with implementations in docking algorithms
- Investigations of Hydration Layer effect
- Molecular Dynamics investigations of coupling between overall tumbling and large scale mobility in proteins

Directions for Future Researches



- Statistics of other parts of genomes: Introns is the first target
- Investigations of correlations between observed two exon peaks and other known classes of genes: ortologous and paralogous gene families, major and minor spliceosomal pathways are the first targets
- Plants and other classes of organisms
- Alternative splicing from the point of view of statistical distributions
- Correlations with real evolutionary mechanisms involving exon-intron boundaries

Acknowledgements

David Fushman

Nikolai Skrynnikov

Amitabh Varshney

Ranjanee Varadan

Jenifer Hall

Yuri Feldman

Alexei Sokolov

Michel Gribskov

Stephen Mount

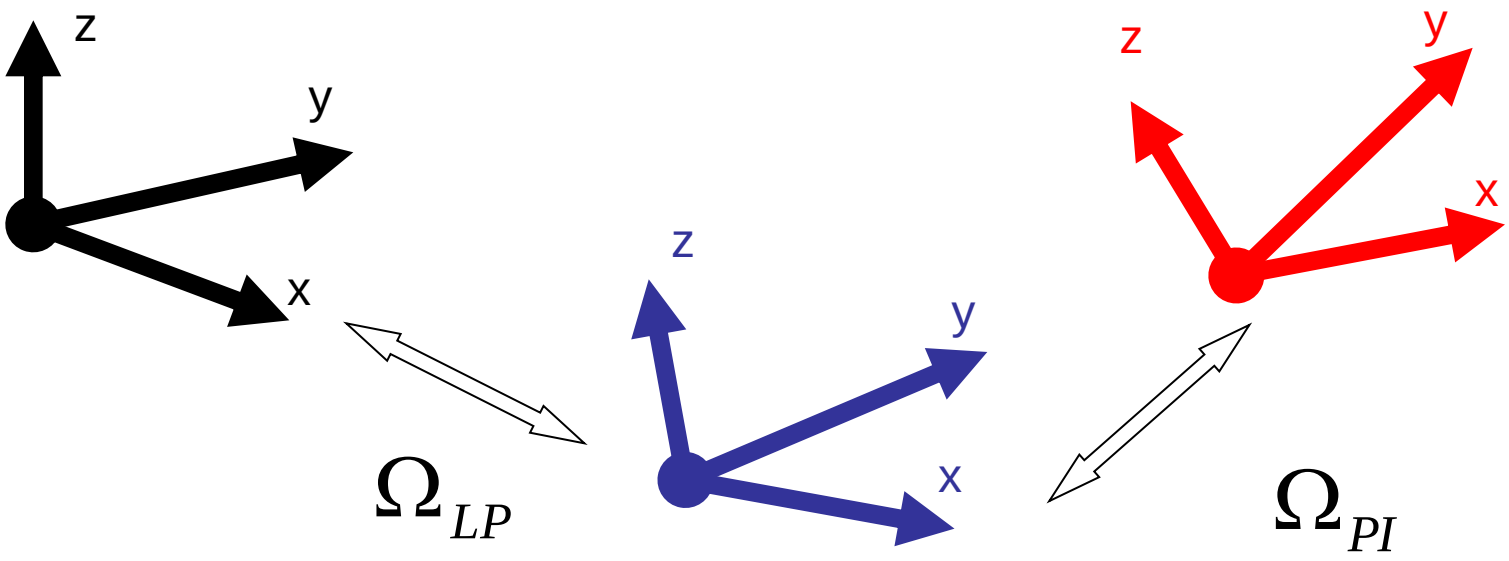
Natalia Grishina

Appendix

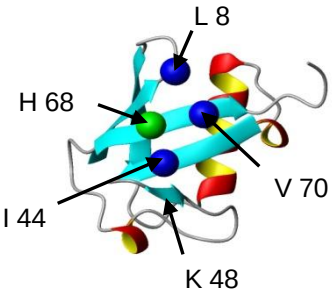
$$C(t) = < \sum_{l,l'=-2}^{2} D_{ql}^{2*}(\Omega_{LP}(0)) D_{ql'}^{2}(\Omega_{LP}(t)) \times$$

$$\times D_{l0}^{2*}(\Omega_{PI}(0)) D_{l'0}^{2}(\Omega_{PI}(t)) >$$

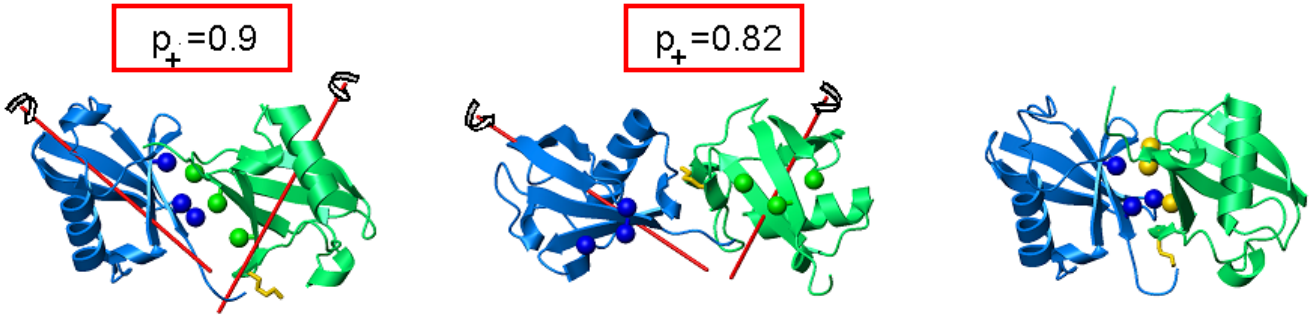
***Wigner**, 1959*
Group theory and its application to the quantum mechanics of atomic spectra



Derived Structures and Dynamics Parameters

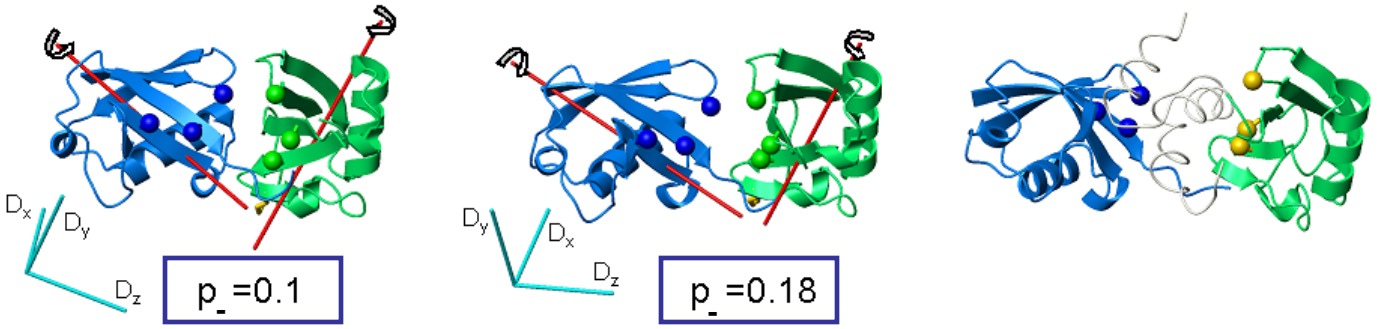


pH 6.8 $\tau_d \approx 9.2\,\text{ns}$ pH 4.5 $\tau_d \approx 9.1\,\text{ns}$



Crystal Ub_2
PDB 1AAR

$\tau_j \approx 9.25\,\text{ns}$ $\tau_j \approx 31.95\,\text{ns}$



Docked:
Ub_2 + UBA

His 68 pKa=5.5
***Fujiwara et al**, J. Bio. Chem. 2004*

	pH 6.8	pH 4.5
Both protonated	0.002	0.826
One protonated	0.091	0.165
Non protonated	0.907	0.009

Appendix

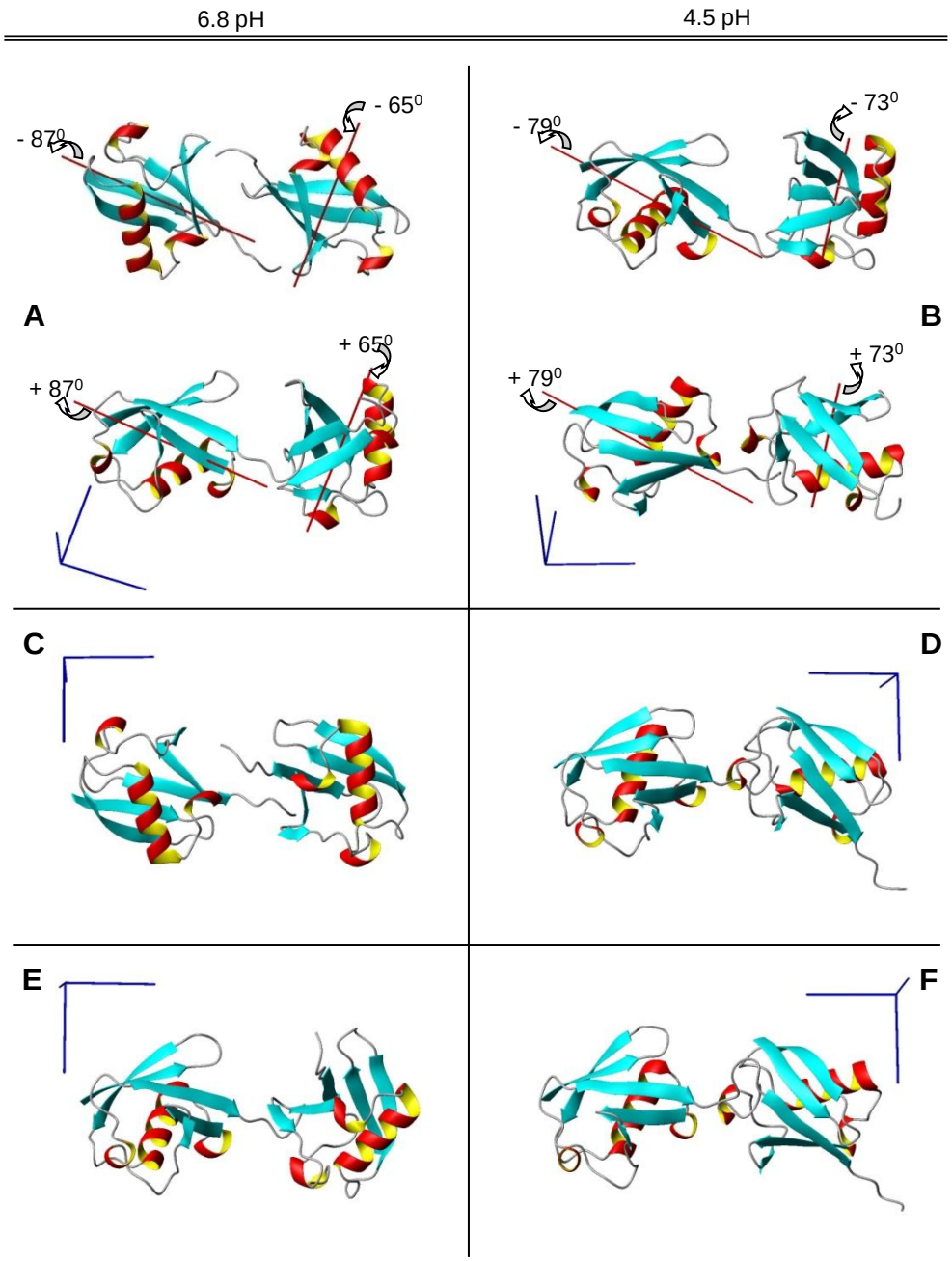
Structures from relaxation fit

Ub2 conformations. For all presented dimers proximal domains are on the right. Orientations of diffusion tensor principle axes are shown on the left. On the left panel of the figure are conformations obtained for 6.8 pH data; on the right panel are conformations for 4.5 pH data.

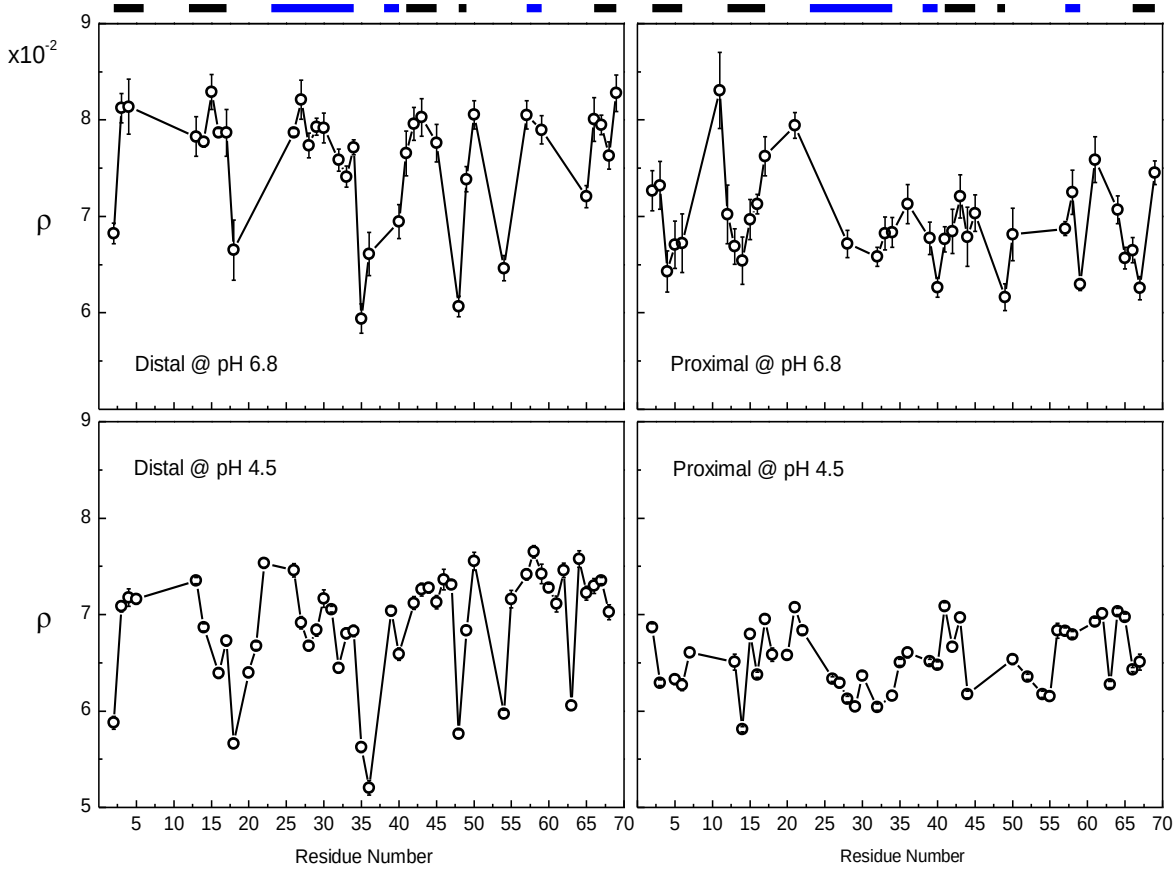
A) and B) each show two distinct conformations of a dimer obtained by jumping domain model: red axis are rotation axis for every domain.

C) and D) show conformations obtained for anisotropic diffusion model applied separately to each domain.

E) and F) represent conformations obtained by jumping domain model with artificially repressed dynamics.



Appendix NMR ^{15}N Relaxation and *NOE*
for Ubiquitin dimer



Local fast motions subtracted by relaxation rates ratio approach

$$\rho = \frac{R'_1}{2R'_2 - R'_1} \cong \frac{3J(\omega_N)}{4J(0)}$$

***Fushman et al**, Prog NMR 2004*

Experimental data from: ***Varadan et al**, JMB 2002*

Appendix

NMR ^{15}N Relaxation, *NOE* and *RDC*

$$d = d_{\max}^{NH} \sum_{i,j=x,y,z} S_{ij}\left(p_+ \cos\varphi_i^+ \cos\varphi_j^+ + p_- \cos\varphi_i^- \cos\varphi_j^-\right)$$

Only at pH 6.8

Assumption

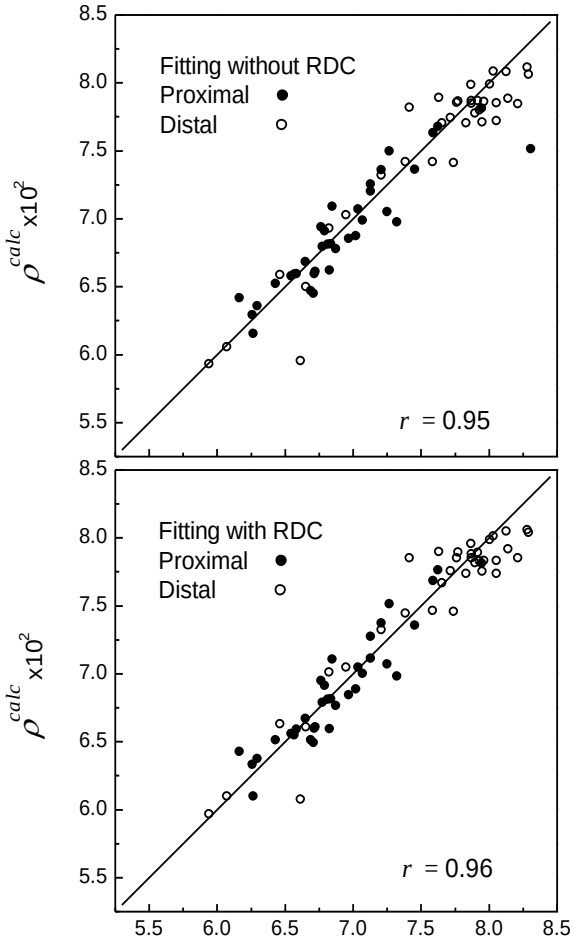
Alignment process
occurs on the time scale
that is much longer
than the time scales of
all considered dynamic modes

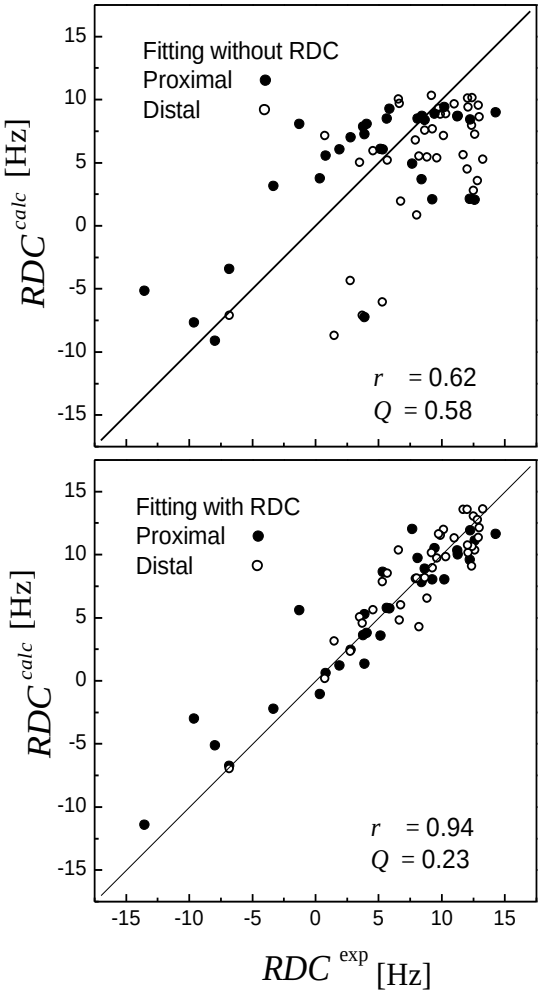
22 fitting parameters
including

$D_x,\ D_y,\ D_z$

$p_+,\ p_-$

12 Euler angels for both
Domain orientations
in two conformations
And
5 components of alignment tensor





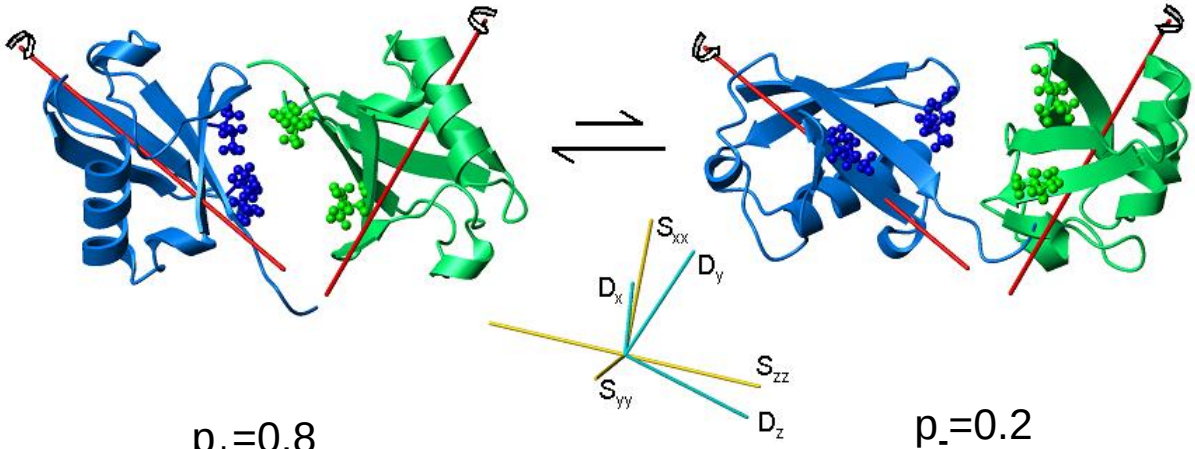
Experimental data from: ***Varadan et al***, *JMB 2002*

Appendix

Derived Structures and Dynamic Parameters

Relaxation VS Relaxation + RDC

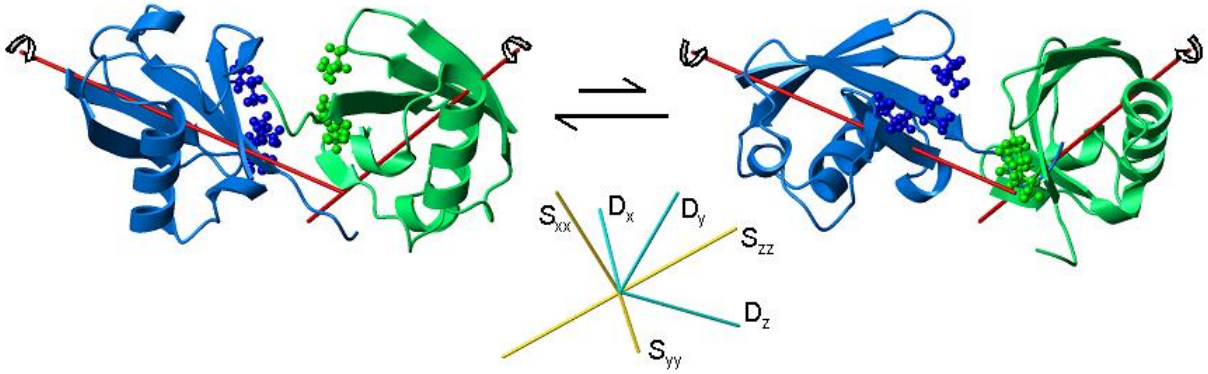
p_+=0.9 p_-=0.1



Relaxation data only

$\tau_j \approx 9.25\,\text{ns}$

p_+=0.8 p_-=0.2

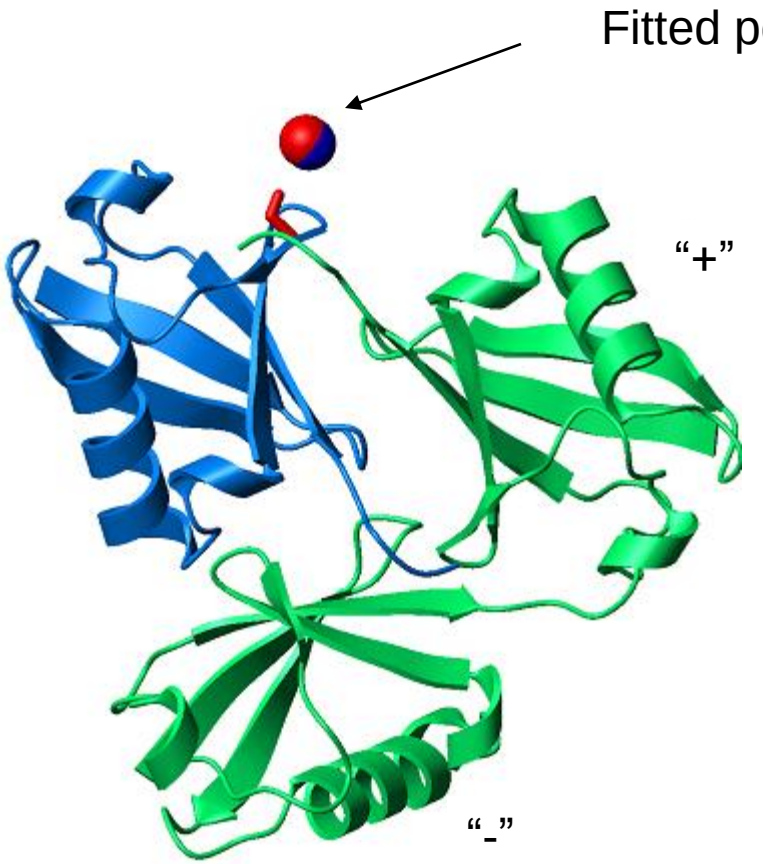


Relaxation + RDC data

$\tau_j \approx 36.0\,\text{ns}$

Appendix

Validation with Spin Labeling data



Blue ball for “+” conformation only

Red ball for “+” and “-” together

Appendix

Spectral density for the model of Interconversion between Two States

$$J_\rho(\omega) = \frac{2}{5}\sum_{m=-2}^{2}\sum_{n=-2}^{2}\sum_{k=-2}^{2}\sum_{l=-2}^{2}\sum_{z=-2}^{2} a^*_{z,m}a_{z,n} \times$$

$$\times\left\{\frac{E_z}{E_z^2+\omega^2}\left[p_-^2\, D^{(2)*}_{m,k}(\Omega^-_{P\to D})D^{(2)}_{n,l}(\Omega^-_{P\to D}) + p_+^2\, D^{(2)*}_{m,k}(\Omega^+_{P\to D})D^{(2)}_{n,l}(\Omega^+_{P\to D}) + \right.\right.$$

$$\left. + p_-p_+ D^{(2)*}_{m,k}(\Omega^-_{P\to D})D^{(2)}_{n,l}(\Omega^+_{P\to D}) + p_-p_+ D^{(2)*}_{m,k}(\Omega^+_{P\to D})D^{(2)}_{n,l}(\Omega^-_{P\to D})\right] +$$

$$+\frac{\tau_j(E_z\tau_j+1)p_-p_+}{(E_z\tau_j+1)^2+(\omega\tau_j)^2}\left[D^{(2)*}_{m,k}(\Omega^-_{P\to D})D^{(2)}_{n,l}(\Omega^-_{P\to D}) + D^{(2)*}_{m,k}(\Omega^+_{P\to D})D^{(2)}_{n,l}(\Omega^+_{P\to D}) - \right.$$

$$\left.\left. - D^{(2)*}_{m,k}(\Omega^-_{P\to D})D^{(2)}_{n,l}(\Omega^+_{P\to D}) - D^{(2)*}_{m,k}(\Omega^+_{P\to D})D^{(2)}_{n,l}(\Omega^-_{P\to D})\right]\right\}\times$$

$$\times D^{(2)*}_{k,0}(\Omega_{D\to R})D^{(2)}_{l,0}(\Omega_{D\to R})$$

Appendix

Definitions of diffusion related parameters

$a_{z,g}$

g \ z	2	1	0	-1	-2
2	$\frac{w}{N\sqrt{2}}$	0	$-\frac{u}{N\sqrt{2}}$	0	$\frac{1}{\sqrt{2}}$
1	0	$\frac{1}{\sqrt{2}}$	0	$\frac{1}{\sqrt{2}}$	0
0	$\frac{u}{N}$	0	$\frac{w}{N}$	0	0
-1	0	$\frac{1}{\sqrt{2}}$	0	$-\frac{1}{\sqrt{2}}$	0
-2	$\frac{w}{N\sqrt{2}}$	0	$-\frac{u}{N\sqrt{2}}$	0	$-\frac{1}{\sqrt{2}}$

$$P(\Omega^0|\Omega^t,t)=\sum_{z=-2}^{2}\sum_{g=-2}^{2}\Psi^*_{z,g}(\Omega^0)\Psi_{z,g}(\Omega^t)\,e^{-E_z t}$$

$$u=\sqrt{3}\left(D_x-D_y\right)$$

$$N=2\sqrt{|\Delta|w}$$

$$w=\begin{cases}2D_z-D_x-D_y+2\Delta & \\ -2D_z+D_x+D_y-2\Delta & \text{for the oblate case, } D_z<D_x=D_y\end{cases}$$

Appendix

Definitions of diffusion related parameters

z	E_z	$\Psi_{z,g}(\Omega)$
2	$6D_s + 2\Delta$	$\Psi_{2,g} = \frac{1}{2\pi N}\sqrt{\frac{5}{2}}\left(uD^{(2)*}_{g,0}(\Omega) + \frac{w}{\sqrt{2}}\left(D^{(2)*}_{g,2}(\Omega) + D^{(2)*}_{g,-2}(\Omega)\right)\right)$
-2	$3(D_z + D_s)$	$\Psi_{-2,g} = \frac{\sqrt{5}}{4\pi}\left(D^{(2)*}_{g,2}(\Omega) - D^{(2)*}_{g,-2}(\Omega)\right)$
1	$3(D_x + D_s)$	$\Psi_{1,g} = \frac{\sqrt{5}}{4\pi}\left(D^{(2)*}_{g,1}(\Omega) + D^{(2)*}_{g,-1}(\Omega)\right)$
-1	$3(D_y + D_s)$	$\Psi_{-1,g} = \frac{\sqrt{5}}{4\pi}\left(D^{(2)*}_{g,1}(\Omega) - D^{(2)*}_{g,-1}(\Omega)\right)$
0	$6D_s - 2\Delta$	$\Psi_{0,g} = \frac{1}{2\pi N}\sqrt{\frac{5}{2}}\left(uD^{(2)*}_{g,0}(\Omega) - \frac{w}{\sqrt{2}}\left(D^{(2)*}_{g,2}(\Omega) + D^{(2)*}_{g,-2}(\Omega)\right)\right)$

$$\Delta = \begin{cases} +\sqrt{(D_x - D_x)^2 + (D_z - D_x)(D_z - D_y)} \\ -\sqrt{(D_x - D_x)^2 + (D_z - D_x)(D_z - D_y)} \end{cases} \quad \text{for the oblate case, } D_z < D_x = D_y$$

$$D_s = \frac{1}{3}\left(D_x + D_y + D_z\right)$$

Appendix

Fitted parameters for the model of Interconversion between Two States

Relaxation data only

pH	D_x	D_y	D_z	τ_j	p_+	Domain	α^+	β^+	γ^+	α^-	β^-	γ^-
6.8	1.53 (0.32)	1.73 (0.06)	2.20 (0.08)	9.3 (4.8)	0.90 (0.06)	Proximal	218 (35)	109 (11)	140 (4)	203 (38)	110 (9)	72 (8)
						Distal	91 (28)	58 (7)	321 (19)	156 (33)	96 (38)	356 (33)
4.5	1.61 (0.13)	1.71 (0.06)	2.20 (0.06)	31.9 (9.8)	0.82 (0.06)	Proximal	147 (30)	112 (16)	322 (8)	191 (25)	122 (14)	45 (13)
						Distal	213 (24)	80 (8)	350 (12)	151 (29)	51 (20)	328 (23)

D in 10^7 s^{-1}

τ in [ns]

Appendix

Analysis of Relaxation data together with RDC

Parameters of the alignment tensor $\chi^2_{relax} = 1.77$

D_x	D_y	D_z	τ_j	p_+	Domain	$\alpha^+_{P\to D}$	$\beta^+_{P\to D}$	$\gamma^+_{P\to D}$	$\alpha^-_{P\to D}$	$\beta^-_{P\to D}$	$\gamma^-_{P\to D}$
1.57 (0.12)	1.75 (0.06)	2.39 (0.08)	36.0 (9.8)	0.76 (0.07)	Proximal	287 (22)	131 (8)	153 (7)	258 (28)	115 (10)	88 (12)
					Distal	72 (15)	52 (8)	327 (8)	147 (21)	87 (21)	356 (16)

Parameters of the ITS model $\chi^2_{RDC} = 0.49$

$d^{NH}_{max} \cdot S_{xx}$	$d^{NH}_{max} \cdot S_{yy}$	$\alpha_{P\to S}$	$\beta_{P\to S}$	$\gamma_{P\to S}$
7.3 (1.3)	18.5 (3.2)	-52 (23)	95 (8)	11 (6)

D in 10^7 s^{-1}

τ in [ns]

Appendix

Model free ideas in application to domain mobility

Correlation function for domain mobility in MF form

$$C^{PD}_{mk,nl}(t) = D^{(2)*}_{m,0}(\Omega_{P\to MF})D^{(2)}_{n,0}(\Omega_{P\to MF})\left(S^2 + (1-S^2)e^{-t/\tau}\right)D^{(2)*}_{0,k}(\Omega_{MF\to D})D^{(2)}_{0,l}(\Omega_{MF\to D})$$

Total correlation function and spectral density

$$C(t) = \frac{1}{5}\sum_{m=-2}^{2}\sum_{n=-2}^{2}\sum_{k=-2}^{2}\sum_{l=-2}^{2}\sum_{z=-2}^{2} e^{-E_z t}a^*_{z,m}a_{z,n}D^{(2)*}_{m,0}(\Omega_{P\to MF})D^{(2)}_{n,0}(\Omega_{P\to MF})\left(S^2 + (1-S^2)e^{-t/\tau}\right)\times$$
$$\times D^{(2)*}_{0,k}(\Omega_{MF\to D})D^{(2)}_{0,l}(\Omega_{MF\to D})D^{(2)*}_{k,0}(\Omega_{D\to R})D^{(2)}_{l,0}(\Omega_{D\to R})$$

$$J_\rho(\omega) = \frac{2}{5}\sum_{m=-2}^{2}\sum_{n=-2}^{2}\sum_{k=-2}^{2}\sum_{l=-2}^{2}\sum_{z=-2}^{2} a^*_{z,m}a_{z,n}D^{(2)*}_{m,0}(\Omega_{P\to MF})D^{(2)}_{n,0}(\Omega_{P\to MF})\left(\frac{S^2E_z}{E_z^2+\omega^2} + \frac{(1-S^2)(E_z\tau+1)}{(E_z\tau+1)^2+(\omega\tau)^2}\right)\times$$
$$\times D^{(2)*}_{0,k}(\Omega_{MF\to D})D^{(2)}_{0,l}(\Omega_{MF\to D})D^{(2)*}_{k,0}(\Omega_{D\to R})D^{(2)}_{l,0}(\Omega_{D\to R})$$

NOTE that for this spectral density relaxation rates ratio, $\rho = R_1^{'}/(2R_2^{'} - R_1^{'})$ becomes independent from the term, $D^{(2)*}_{0,k}(\Omega_{MF\to D})D^{(2)}_{0,l}(\Omega_{MF\to D})D^{(2)*}_{k,0}(\Omega_{D\to R})D^{(2)}_{l,0}(\Omega_{D\to R})$ which means absolute independence of the residue orientation. In other words this representation leads to *quasi-isotropic* form of ρ.

Appendix

Using Extended Model Free for domain motions

Clore et al, *JACS 1990*
Chang and Tjandra*, JACS 2001*

$$C^{RI}_{s0,h0}(t) = \left\langle D^{(2)*}_{s,0}(\Omega^{0}_{R\to I}) D^{(2)}_{h,0}(\Omega^{t}_{R\to I}) \right\rangle_{R\to I} = \delta_{s,0}\delta_{h,0}\left\{(1-S_f^2)\exp(-t/\tau_f) + \right.$$
$$\left. + S_f^2\left[S_s^2 + (1-S_s^2)\exp(-t/\tau_s)\right]\right\}$$

$$C^{PD}_{mk,nl}(t) = D^{(2)*}_{m,k}(\Omega_{P\to D}) D^{(2)}_{n,l}(\Omega_{P\to D})$$

Apparently NO domain motions

$$C(t) = \left\langle D^{(2)*}_{q,0}(\Omega^{0}_{L\to I}) D^{(2)}_{q,0}(\Omega^{t}_{L\to I}) \right\rangle_{L\to I} \cong$$

$$\cong \frac{1}{5}\sum_{m=-2}^{2}\sum_{n=-2}^{2}\sum_{k=-2}^{2}\sum_{l=-2}^{2}\sum_{z=-2}^{2} e^{-E_z t} a^{*}_{z,m} a_{z,n} D^{(2)*}_{m,k}(\Omega_{P\to D}) D^{(2)}_{n,l}(\Omega_{P\to D}) D^{(2)*}_{k,0}(\Omega_{D\to R}) D^{(2)}_{l,0}(\Omega_{D\to R})$$
$$\left[S_s^2 + (1-S_s^2)\exp(-t/\tau_s)\right]$$

$$J_\rho(\omega) = \frac{2}{5}\sum_{m=-2}^{2}\sum_{n=-2}^{2}\sum_{k=-2}^{2}\sum_{l=-2}^{2}\sum_{z=-2}^{2} a^{*}_{z,m} a_{z,n} D^{(2)*}_{m,k}(\Omega_{P\to D}) D^{(2)}_{n,l}(\Omega_{P\to D}) D^{(2)*}_{k,0}(\Omega_{D\to R}) D^{(2)}_{l,0}(\Omega_{D\to R}) \times$$
$$\times\left(\frac{S_s^2 E_z}{E_z^2+\omega^2} + \frac{(1-S_s^2)(E_z\tau_s+1)}{(E_z\tau_s+1)^2+(\omega\tau_s)^2}\right)$$

Appendix

Using Extended Model Free fitting Ubiquitin dimer data

Overall fitting parameters pH 6.8		
D_x 10^7 rad^2/s	D_y 10^7 rad^2/s	D_z 10^7 rad^2/s
0.70	0.88	1.42

Domain fitting parameters (all angles are in degrees)		
	Proximal	Distal
$\tau_s[ns]$	2.0800	11.2334
S_s^2	0.3197	0.0164
$\alpha_{P\to D}$	266.5950	25.6547
$\beta_{P\to D}$	42.6660	54.4322
$\gamma_{P\to D}$	327.6032	348.3614

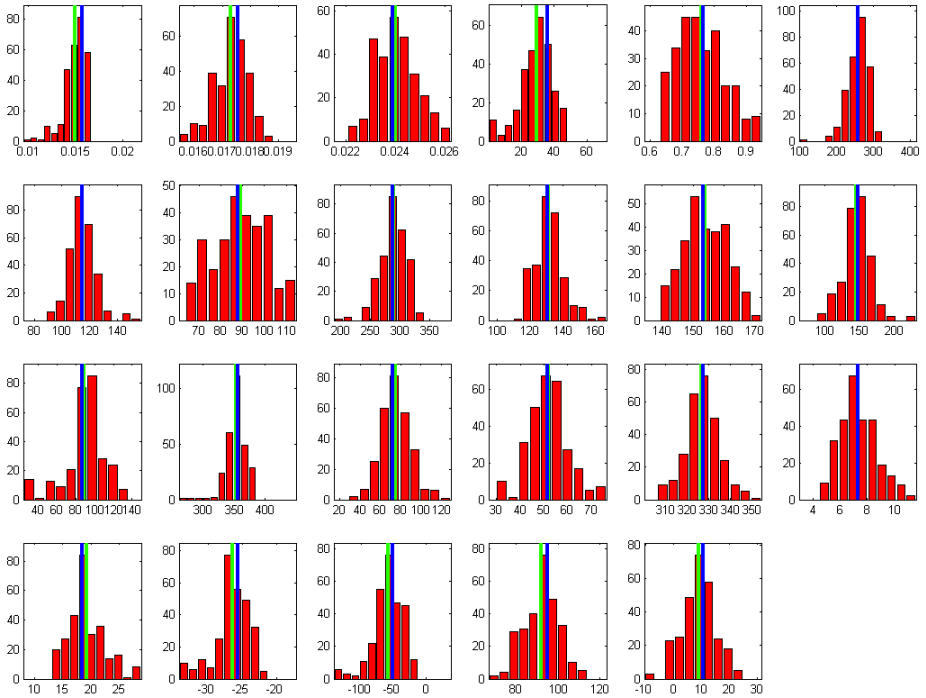
Overall fitting parameters pH 4.5		
D_x 10^7 rad^2/s	D_y 10^7 rad^2/s	D_z 10^7 rad^2/s
1.36	1.50	1.83

Domain fitting parameters (all angles are in degrees)		
	Proximal	Distal
$\tau_s[ns]$	2.5784	8.2130
S_s^2	0.5556	0.1970
$\alpha_{P\to D}$	111.2829	166.3341
$\beta_{P\to D}$	72.0742	111.9480
$\gamma_{P\to D}$	151.4625	176.9133

Appendix

Bootstrap method for estimation of Confidence Intervals

Press et al,
Numerical recipes in C, 1992



$1/e \sim 37\%$
of original data substituted by duplicates of the rest data.

Then

Data fitted in the same way as the original data

Procedure repeated 200 times

Confidence intervals estimated from resulted distributions of fitted parameters

Appendix

Hydrodynamics Calculations HYDRONMR and PCA

De La Torre et al,
J Mag. Res, 2000

Settings for HYDRONMR

Temperature 297 K,

eta=0.0091 [sp], AER=3.2 Å

Settings for Surf-PCA based algorithm

Temperature 297 K,

a) was evaluated for dry protein and then the eigenvalues of diffusion tensor were divided by factor of 2.
b) Sell parameter was 2.8 Å. No any additional factors

	Dx 10^7 [rad/s]	Dy 10^7 [rad/s]	Dz 10^7 [rad/s]	Tau [ns]
Relaxation fit @ **pH 6.8**				
Fitted values	1.53	1.73	2.20	9.1575
Closed Structure used in the paper				
HydroNMR	1.211	1.222	2.107	11.01
Surf-PCA a)	1.2763	1.2782	2.2900	10.32
Surf-PCA b)	1.5034	1.510	2.3432	9.3342
Open Structure used in the paper				
HydroNMR	1.312	1.326	2.213	10.31
Surf-PCA a)	1.2817	1.2876	2.2390	10.399
Surf-PCA b)	1.5088	1.5192	2.288	9.4057
Closed Structure fitted by surf-PCA routine				
HydroNMR	1.415	1.426	2.266	9.796
Surf-PCA a)	1.3769	1.3877	2.2587	9.9536
Surf-PCA b)	1.6011	1.6123	2.331	9.0147
Open Structure fitted by surf-PCA routine				
HydroNMR	1.414	1.438	2.280	9.742
Surf-PCA a)	1.3647	1.3747	2.2353	10.051
Surf-PCA b)	1.628	1.6372	2.3379	8.9236

Appendix

Hydrodynamics Calculations HYDRONMR and PCA

De La Torre et al,
J Mag. Res, 2000

	Dx 10^7 [rad/s]	Dy 10^7 [rad/s]	Dz 10^7 [rad/s]	Tau [ns]
Relaxation fit @ **pH 4.5**				
Fitted values	1.61	1.71	2.20	9.0580
Closed Structure used in the paper				
HydroNMR	1.189	1.201	2.155	11.00
Surf-PCA b)	1.4320	1.4387	2.3087	9.6536
Open Structure used in the paper				
HydroNMR	1.189	1.207	2.172	10.95
Surf-PCA b)	1.3867	1.3934	2.3057	9.8314

Settings for HYDRONMR

Temperature 297 K,

eta=0.0091 [sp], AER=3.2 Å

Settings for Surf-PCA based algorithm

Temperature 297 K,

a) was evaluated for dry protein and then the eigenvalues of diffusion tensor were divided by factor of 2.
b) Sell parameter was 2.8 Å. No any additional factors

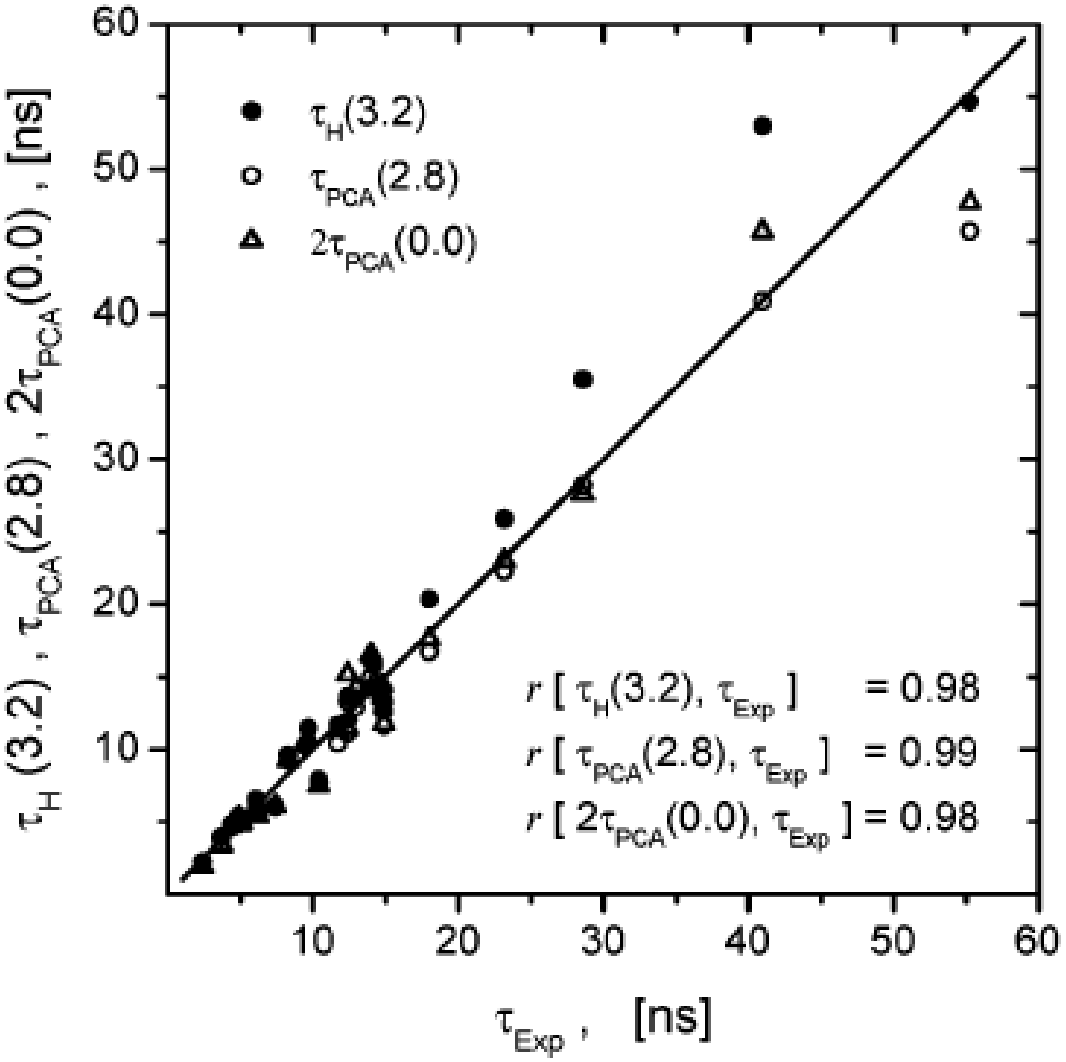
Complexity of the algorithms

ELM	N_{at}
FastHYDRONMR	$N_{at}^{4/3}$
HYDRONMR	N_{at}^{2}

ELM : FastHYDRONMR : HYDRONMR

1 : 2 : 500

Comparison with the experimental data

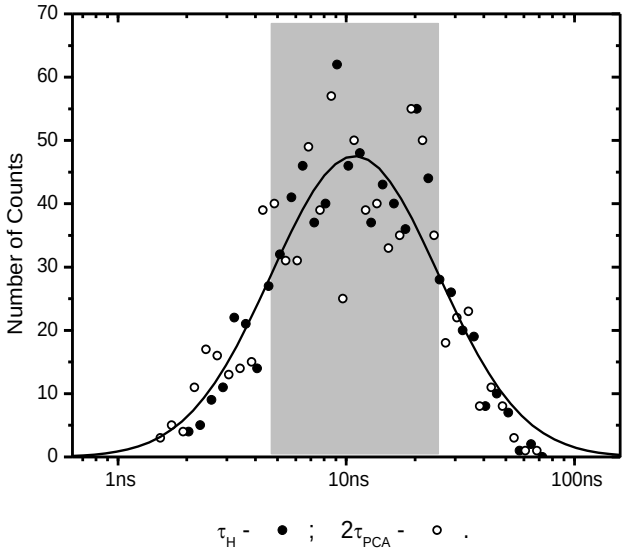


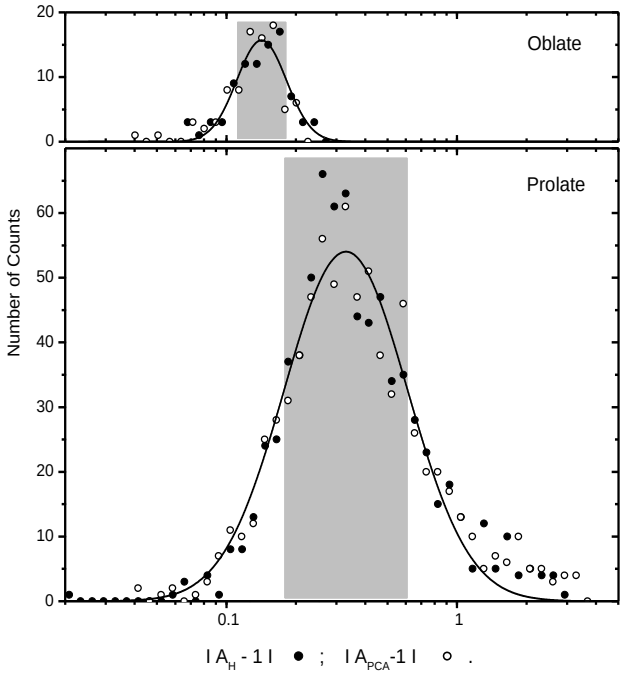
Comparison with experimental data
proteins from 2.9 to 82 kDa

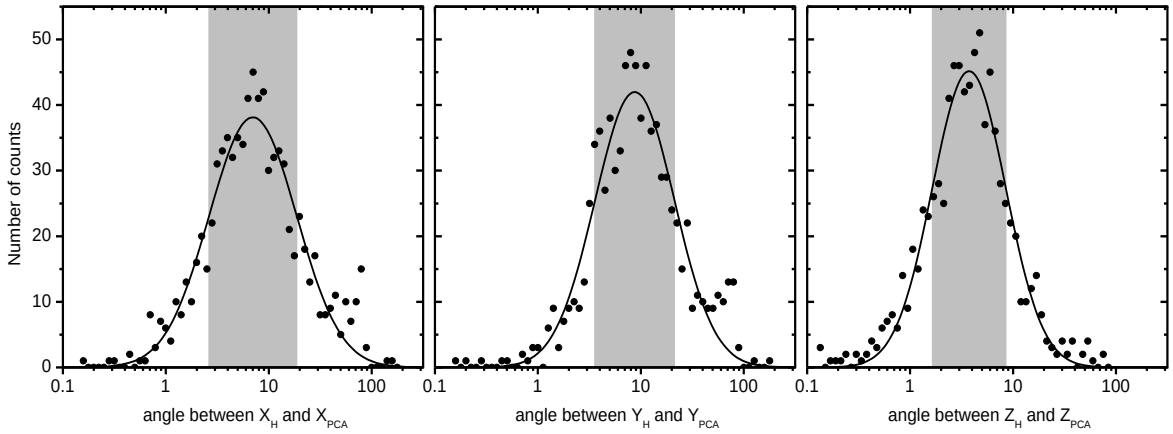
protein	PDB code	experiment[b] τ_c exp. [ns]	PCA variable HLT HLT Å	PCA variable HLT τ_c calc. [ns]	HYDRONMR AER = 3.2 Å τ_c calc. [ns]	HYDRONMR AER = 3.2 Å diff. %	PCA HLT = 2.8 Å τ_c calc. [ns]	PCA HLT = 2.8 Å diff. %	PCA HLT = 0.0 Å $2\tau_c$ calc. [ns]	PCA HLT = 0.0 Å diff. %
malate synthase G[67]	1y8b	55.3 (N)	4.6	55.38	54.6	−1	45.7	−17	47.7	−14
human serum albumin[68]	1ao6	41.0 (F)	2.8	40.87	53.0	29	40.9	0	45.7	11
maltose binding protein[69]	1ezp	28.6 (N)	2.9	28.65	35.5	24	28.2	−1	27.6	−3
β-lactoglobulin a (dimer)[61]	1bsy	23.2 (F)	3.1	23.13	25.9	12	22.2	−4	23.0	−1
Δ^5-3 ketosteroid isomerase[70]	1buq	18.0 (F,N)	3.2	17.96	20.4	13	16.7	−7	17.6	−2
leukemia inh. factor [44]	1lki	14.9 (N)	4.4	14.93	12.6	−15	11.6	−22	11.8	−21
trypsin[61]	2blv	14.8 (F)	4.1	14.76	13.0	−12	12.6	−15	14.4	−3
yellow fluorescent protein[71]	2yfp	14.8 (F)	3.4	14.9	14.1	−5	13.5	−9	13.6	−8
green fluorescent protein[72]	1w7s	14.2 (F)	2.8	14.24	15.8	11	14.2	0	14.6	3
carbonic anhydrase[61]	2cab	14.0 (F)	2.4	13.9	16.2	16	14.9	6	16.5	18
HIV-1 protease[73]	1bvg	13.0 (N)	2.8	12.94	13.5	4	12.9	−1	14.2	9
savinase[44]	1svn	12.4 (N)	2.3	12.38	13.6	10	13.3	7	15.2	23
interleukin-1β[44]	6i1b	12.4 (N)	3.5	12.43	12.0	−3	11.1	−10	11.4	−8
ribonuclease H[58]	2rn2	11.7 (N)	3.5	11.68	11.6	−1	10.4	−11	11.3	−3
cytochrome c_2[74]	1c2n [c]	10.4 (N)	4.6	10.40	7.9	−24	7.7	−26	7.5	−28
β-lactoglobulin A (mono)[61]	1bsy	9.7 (F)	2.4	9.79	11.4	18	10.5	8	10.7	10
apomyoglobin[61]	1bvc	9.5 (F)	2.4	9.55	10.2	9	10.3	10	10.5	12
lysozyme[44]	1hwa	8.3 (N)	2.2	8.26	9.6	16	9.2	11	9.2	11
barstar C40/83A[44]	1bta	7.4 (N)	3.9	7.33	6.1	−18	6.0	−19	6.1	−18
eglin c[44]	1egl [d]	6.2 (N)	3.4	6.23	6.0	−3	5.6	−10	5.4	−13
cytochrome b_5[44]	1wdb	6.1 (N)	3.0	6.10	6.5	7	5.9	−3	5.4	−11
calbindin-D9k+Ca^{2+} [44]	2bca	5.1 (N)	2.9	5.07	5.1	0	5.0	−2	4.9	−4
ubiquitin[14]	1ubq [e]	5.0 (N)	2.9	4.98	5.0	0	4.9	−2	4.8	−4
calbindin-D9k[44]	1clb [d]	4.9 (N)	2.3	4.88	5.2	6	5.4	10	5.3	8
BPTI[44]	1pit [d]	4.4 (N)	2.6	4.41	4.8	9	4.6	5	4.7	7
Protein G[9]	1igd [f]	3.7 (N)	2.6	3.74	3.9	5	3.9	5	3.4	−8
Xfin-zinc finger DBD[44]	1znf [d]	2.4 (N)	3.2	2.37	2.0	−17	2.2	−8	1.9	−21
mean abs. value			3.1(0.7)[g]			11%		8%		10%

Ryabov & Fushman, JACS (2006)

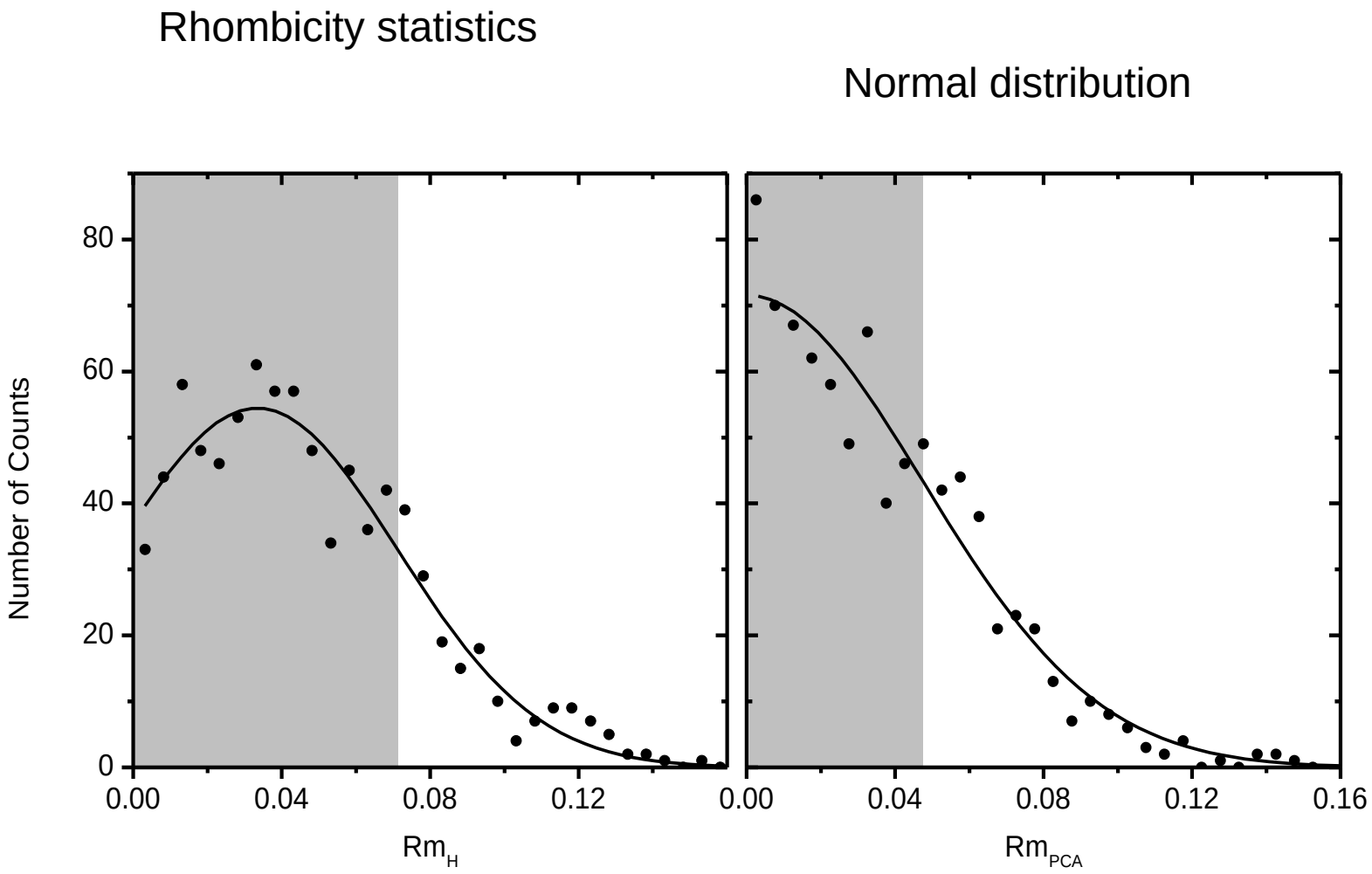
Appendix


Number of Counts
1ns
10ns
100ns
τH - ● ; 2τPCA - ○ .


Oblate
Prolate
Number of Counts
0.1
1
IAH - 1I ● ; IAPCA-1I ○ .


Number of counts
angle between XH and XPCA
angle between YH and YPCA
angle between ZH and ZPCA

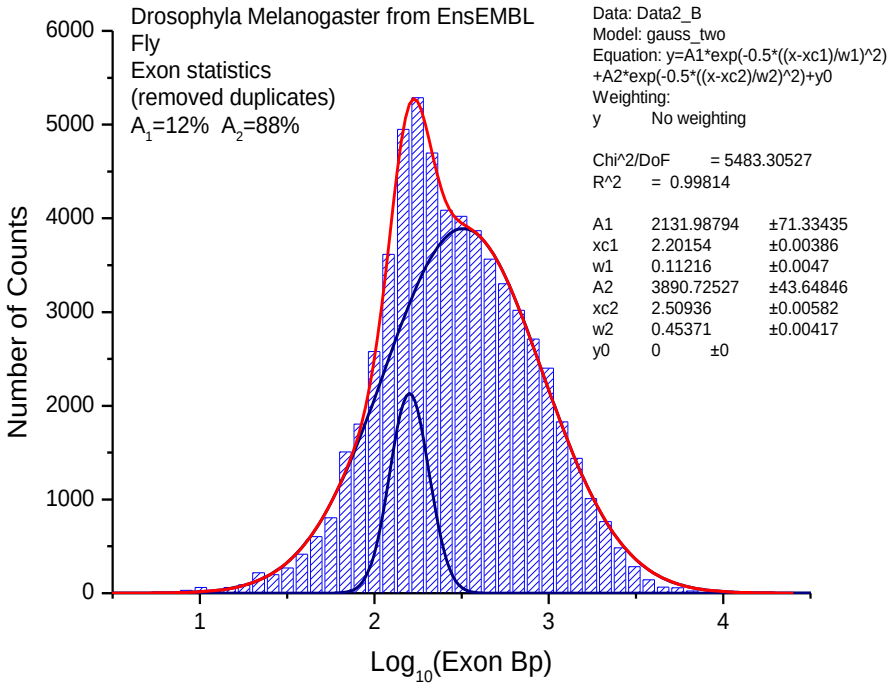
Appendix

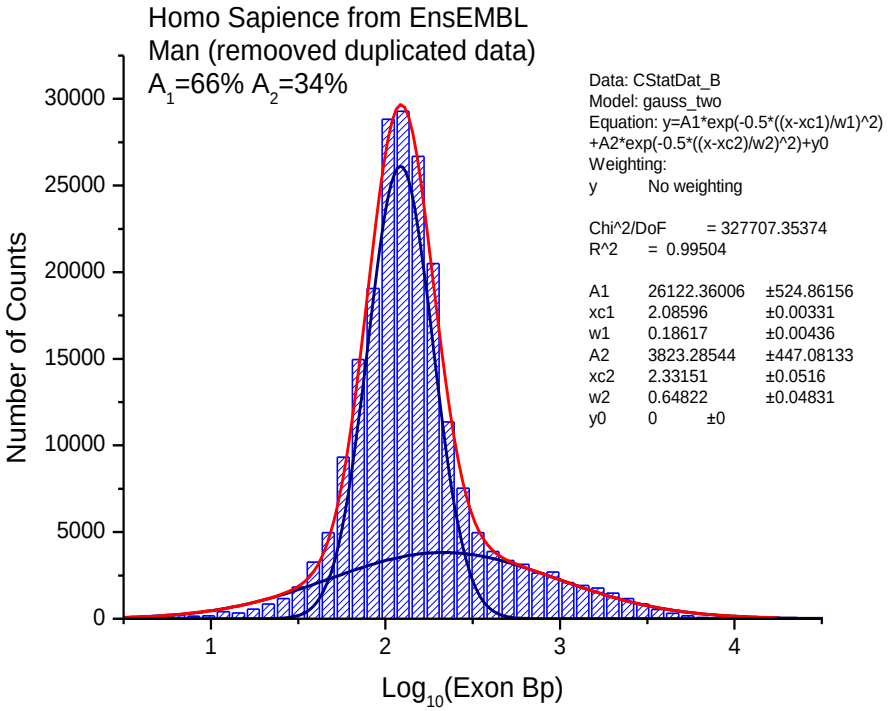

Rhombicity statistics
Normal distribution
Number of Counts
80
60
40
20
0
0.00
0.04
0.08
0.12
0.00
0.04
0.08
0.12
0.16
Rm$_H$
Rm$_{PCA}$

Appendix

Exons

Introns

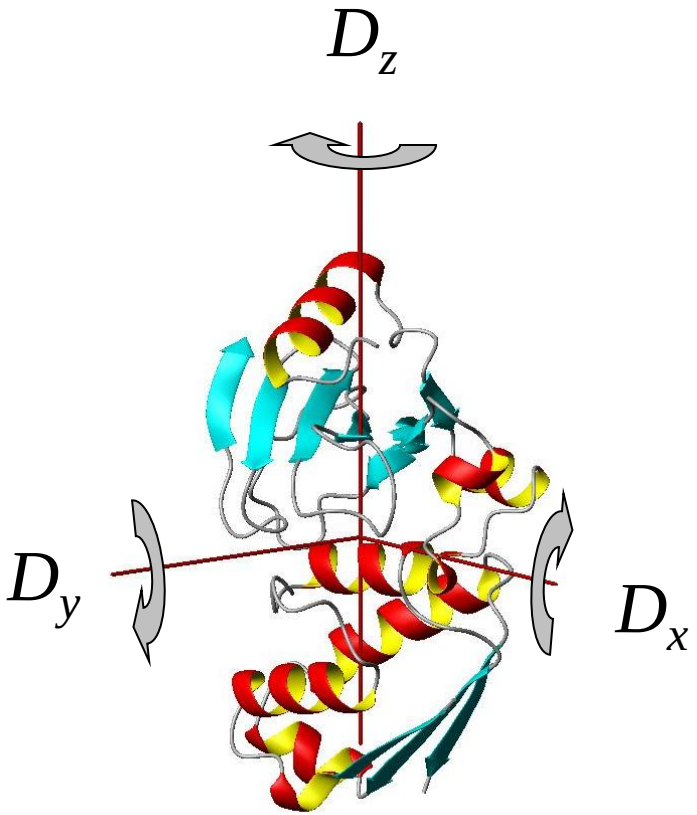




Major and Minor Spliseosomal Pathway ???

Protein Rotation Diffusion Tensor

$$\begin{bmatrix} D_x & 0 & 0 \\ 0 & D_y & 0 \\ 0 & 0 & D_z \end{bmatrix}$$



3 Euler angles for
Diffusion Tensor PAF