SUPERPOSING PROTEIN STRUCTURES

Andrea Giansanti

Dipartimento di Fisica, Sapienza Università di Roma
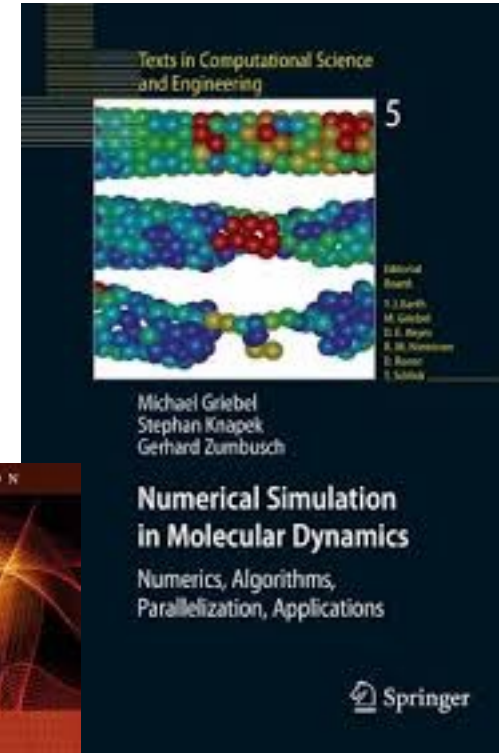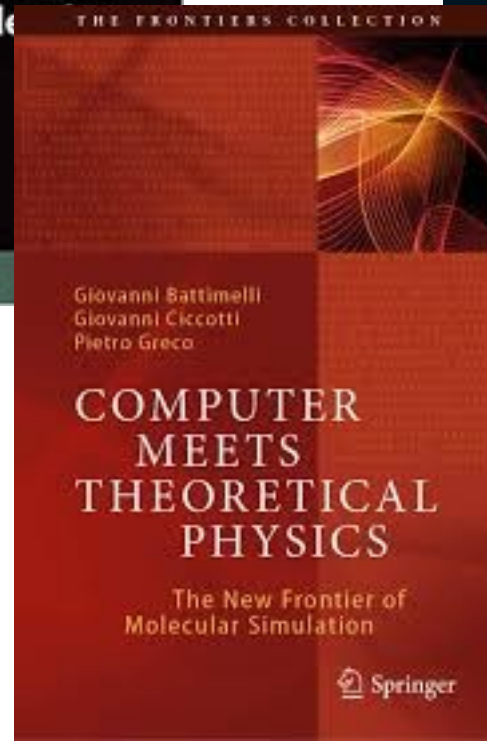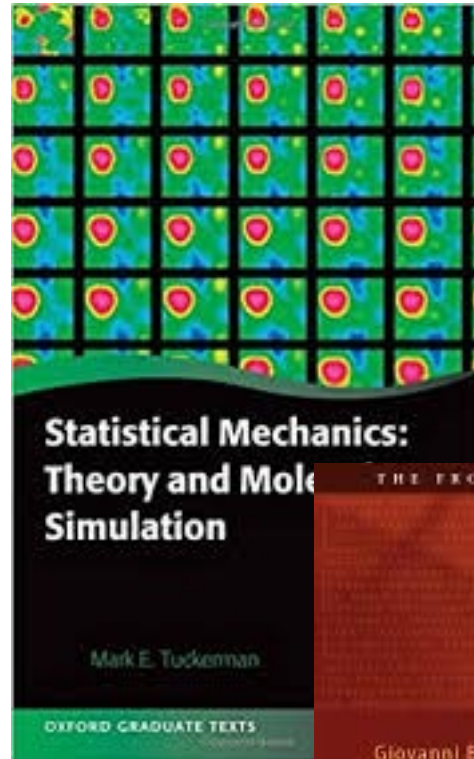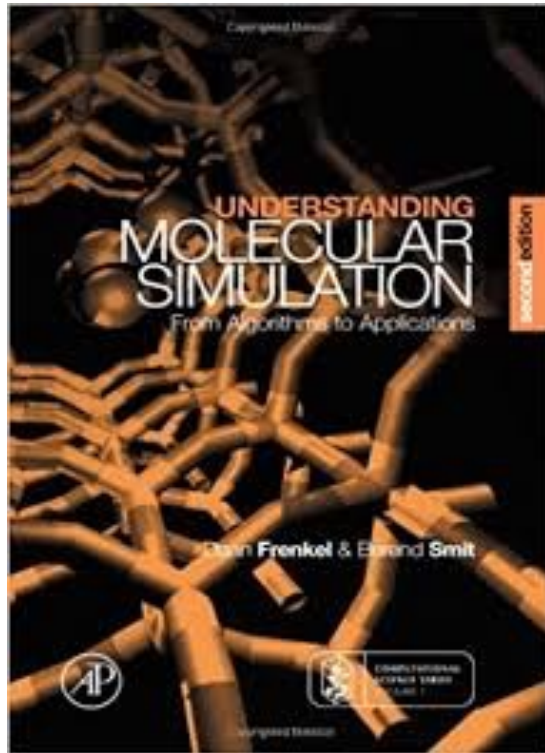
Andrea.Giansanti@roma1.infn.it

CB_23_24 L 15 and 16, Rome 30 and 31 OCT 2023

DIPARTIMENTO DI FISICA

SAPIENZA
UNIVERSITÀ DI ROMA

A NICE MOVIE BY Giff Ransom Strongly recommended as a reca

https://slideplayer.com/slide/4795252/

OUTLINE LECTURES 15 & 16

Textbooks

Tutorials

Time scales

How to superpose 2 molecular structures

Structural superposition of proteins

the problem of roto-translation subtraction (structural superposition)

in protein MD

RMSD , is a not ideal distance

Quaternions

Superpose + Chimera

3)
use of superpose and other on line resources show with images

Homework (due on thursday nov 19 compute the matrix of structural distances between the following set of structures from the PDB taken from the paper Wohlkonig2010
3cqi, 1iee,2dqa,153l, 21zm,, 1am7, 1qgi)
Read the paper and prepare a discussion for the journal club

4) Various vector distance functions
Clustering/Classification

Where to start ?

- GROMACS http://www.gromacs.org

- NAMD VMD https://www.ks.uiuc.edu/Research/namd/

- K. Hinsen MMTK https://github.com/khinsen/MMTK

- . Marchi ORAC http://www.chim.unifi.it/orac/

  - CECAM ICTP SCHOOL https://www.cecam.org/workshop-deta

  MolSim School (http://www.acmm.nl/molsim/molsim2020/)

HOMEWORK

Familiarize with the GROMACS Package: tutorial n.1 at
http://www.mdtutorials.com/gmx/index.html

shared tutorial from last year course L15
askme the access to

tutorial on measuring conformational distances with Chimera

In Chimera https://www.youtube.com/watch?v=oThN3LG8LQU

PBC tutorial (series MOLBIOCHEM at UAM (Madrid))
https://www.youtube.com/watch?v=8iHER6IP6Ds

Homeworks…

- GROMACS TUTORIAL LYSOZYME JA Lemkul

http://www.mdtutorials.com/gmx/lysozyme/

- Mutations in the PDB

- How to mutate a PDB file?

- https://spdbv.unil.ch/mutation_guide.html#:~:text=To%20initiate%20a%20mutation%2C%20simply,the%20new%20amino%20acid%20type.

LEXICON: rotamers

https://www.sciencedirect.com/topics/chemistry/rotamer

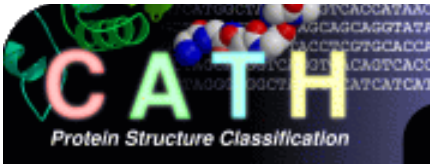Multiple alignment 3d Coffee Clustall omega
Pdb
Pdbsum
Superpose

Eliminate ligand.

Further suggestions on the problem
of structural superposition of
proteins

Adam Zemla's contribution

https://proteopedia.org/wiki/index.p
hp/Calculating_GDT_TS

Protein Structure Classification

CATH - Protein Structure Classification

[http://www.biochem.ucl.ac.uk/bsm/cath_new/]

- UCL, Janet Thornton & Christine Orengo

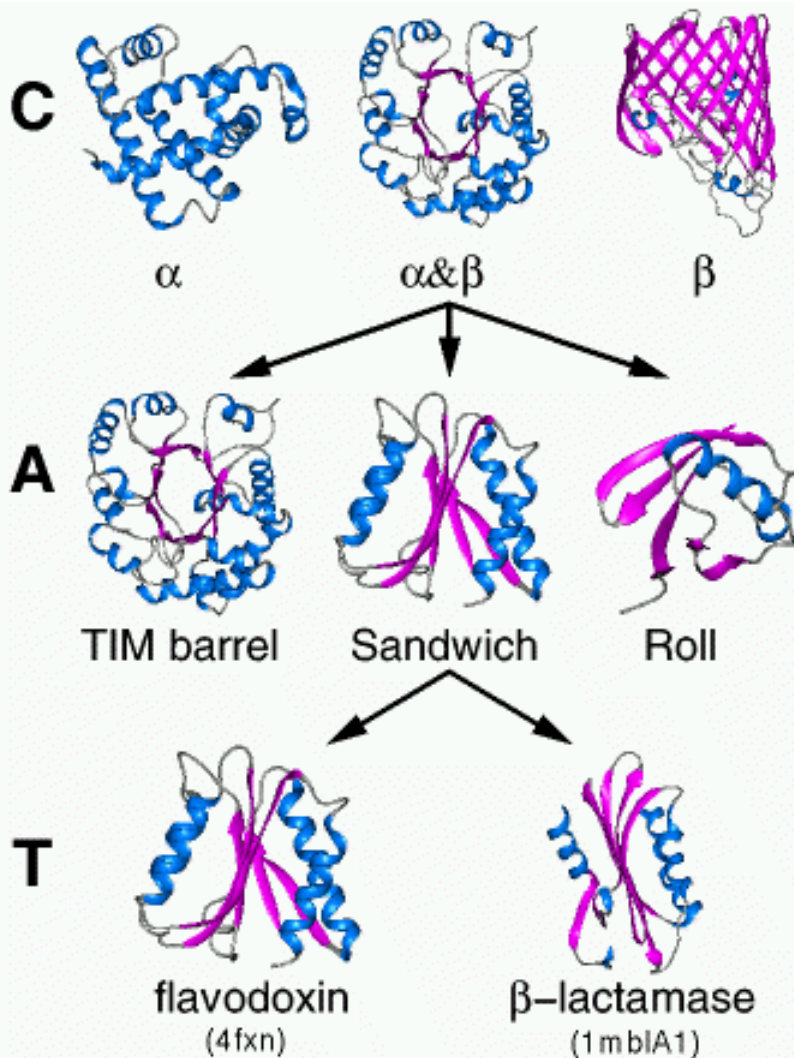- Class (C), Architecture(A), Topology(T), Homologous superfamily (H)

SCOP - Structural Classification of Proteins

- MRC Cambridge (UK), Alexey Murzin, Brenner S. E., Hubb

- created by manual inspection

- comprehensive description of the structural and evolution

[http://scop.mrc-lmb.cam.ac.uk/scop/]

Alejandro Giorgetti
giorget@sissa.it

CATH
Protein Structure Classification

- Class(C)
  derived from secondary structure content is assigned automatically

- Architecture(A)
  describes the gross orientation of secondary structures, independent of connectivity.

- Topology(T)
  clusters structures according to their topological connections and numbers of secondary structures

- Homologous superfamily (H)

Alejandro Giorgetti
giorget@sissa.it

Protein Homology Modeling Resources

SWISS MODEL: http://www.expasy.org/swissmod/SWISS-MODEL.html

Deep View - SPDBV:
homepage: http://www.expasy.ch/spdbv

Tutorials http://www.expasy.org/spdbv/text/tutorial.htm

WhatIf   http://www.cmbi.kun.nl:1100/
Gert Vriend's protein structure modeling analysis program WhatIf

Modeller: http://guitar.rockefeller.edu/modeller
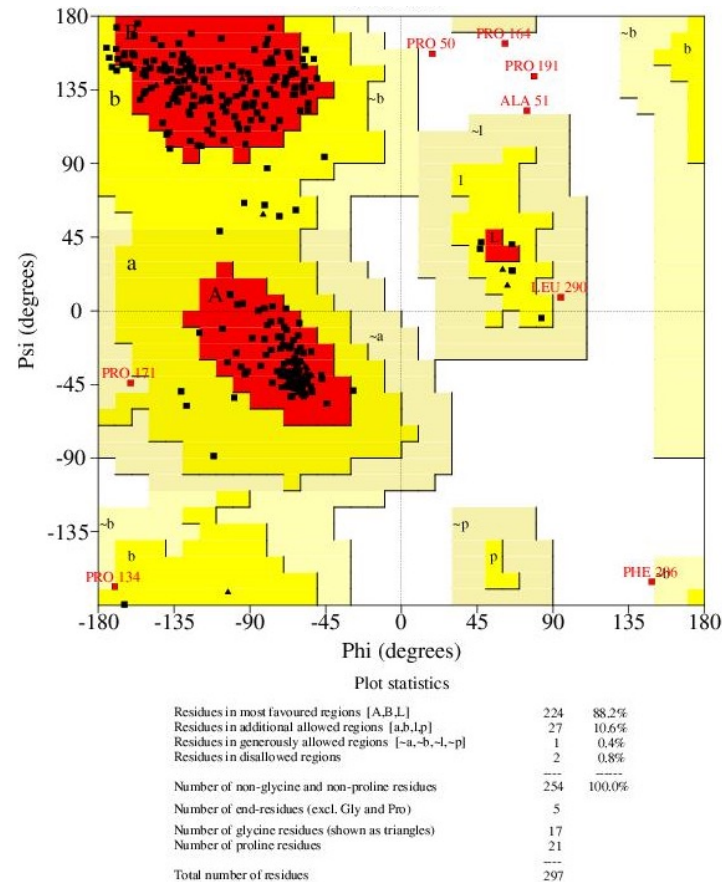Andrej Sali's homology protein structure modelling by satisfaction of spatial restraints

ROBETTA: http://robetta.bakerlab.org/
Full-chain Protein Structure Prediction Server

Programs and www servers very useful in Comparative modeling: http://salilab.org/tools/

Alejandro Giorgetti
giorget@sissa.it

Model Evaluation ?

Topics:

➢ correct fold

➢ model coverage (%)

➢ Cα - deviation (rmsd)

➢ alignment accuracy (%)

➢ side chain placement

➢ Biotech Validation Suite for Protein Structures: http://biotech.ebi.ac.uk:8400

➢ Structure Analysis and Verification Server:
http://shannon.mbi.ucla.edu/Services/SV

➢ WhatIf:
http://www.cmbi.kun.nl:1100/

Alejandro Giorgetti
giorget@sissa.it

Model Accuracy Evaluation

CASP
Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction

http://PredictionCenter.llnl.gov/casp6

EVA

Evaluation of Automatic protein structure prediction

[Burkhard Rost, Andrej Sali, http://maple.bioc.columbia.edu/eva]
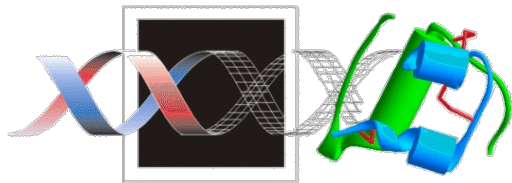
3D - Crunch

Very Large Scale Protein Modelling Project

http://www.expasy.org/swissmod/SM_LikelyPrecision.html

Alejandro Giorgetti
giorget@sissa.it

Protein Structure Resources

PDB   http://www.pdb.org
PDB – Protein Data Bank of experimentally solved structures (RCSB)

CATH   http://www.biochem.ucl.ac.uk/bsm/cath
Hierarchical classification of protein domain structures

SCOP   http://scop.mrc-lmb.cam.ac.uk/scop
Alexey Murzin's Structural Classification of proteins

DALI   http://www2.ebi.ac.uk/dali
Lisa Holm and Chris Sander's protein structure comparison server

SS-Prediction and Fold Recognition

PHD   http://cubic.bioc.columbia.edu/predictprotein
Burkhard Rost's Secondary Structure and Solvent Accessibility Prediction Server

PSIPRED http://bioinf.cs.ucl.ac.uk/psipred/
L.J McGuffin, K Bryson & David T. Jones Secndary struture prediction Server

3DPSSM http://www.sbg.bio.ic.ac.uk/~3dpss
Fold Recognition Server using 1D and 3D Sequence Profiles coupled.

THREADER: http://bioinf.cs.ucl.ac.uk/threader/threader.html
David T. Jones threading program

Alejandro Giorgetti
giorget@sissa.it

Protein Structures from an NMR Perspective

Venn diagram grouping amino acids according to their properties
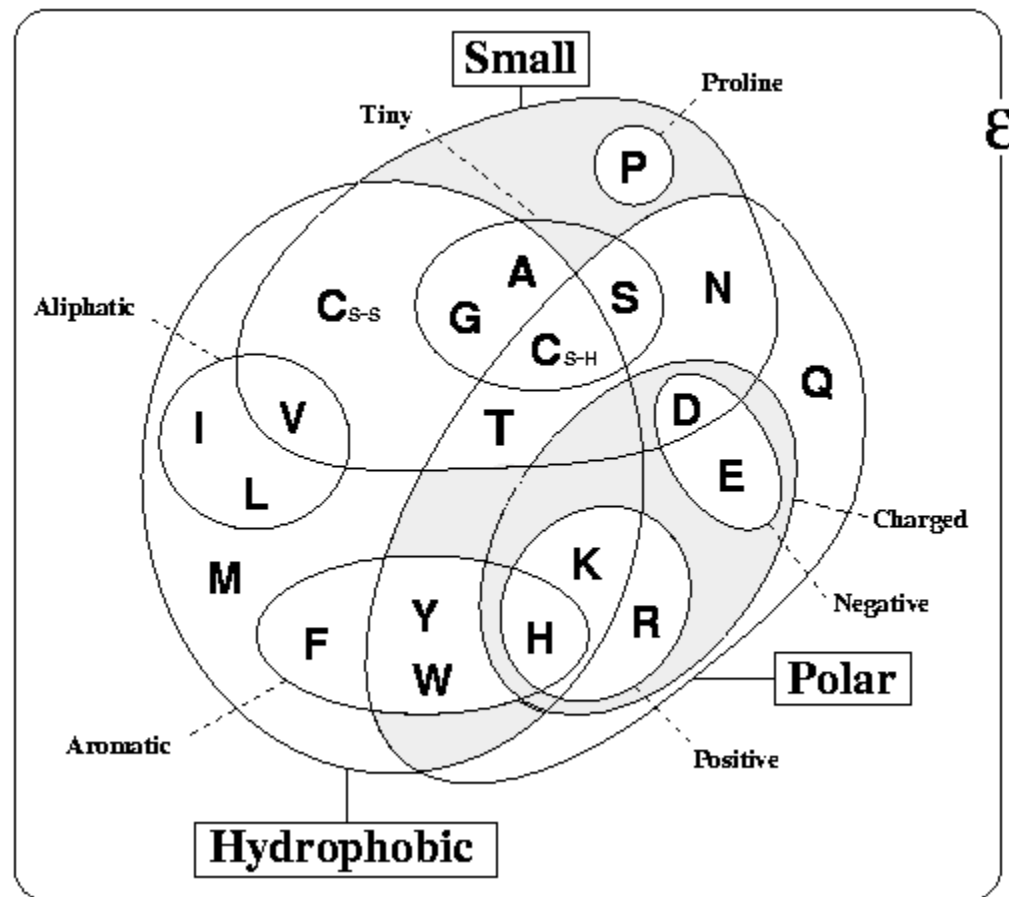
Protein Structures from an NMR Perspective

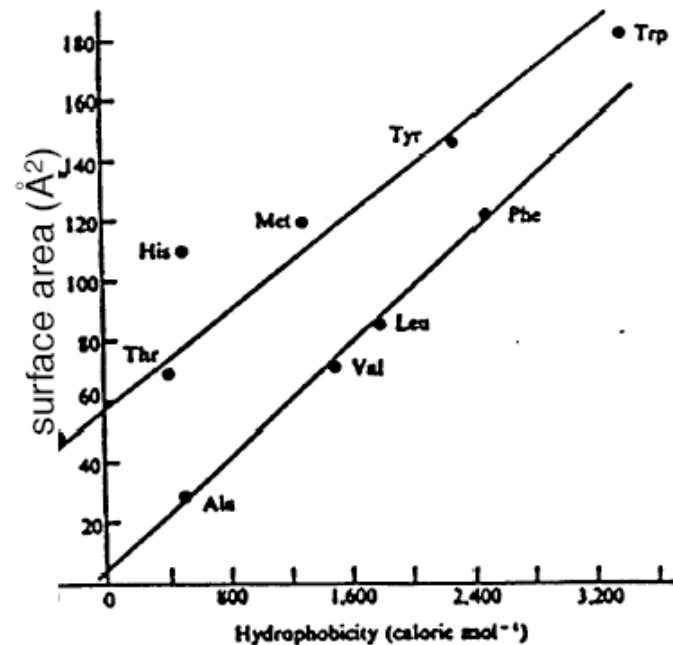| Name	3-Letter Code	Single Code	Relative Abundance	MW	pKa	Residue Volume (Å^3)	Surface Area (Å^2)	Charged, Polar, Hydrophobic
Alanine	ALA	A	13.0	89		88.6	115	H
Arginine	ARG	R	5.3	175	~12	173.4	225	C+
Asparagine	ASN	N	9.9	132		111.1	150	P
Aspartate	ASP	D	9.9	132	4.5	114.1	160	C-
Cysteine	CYS	C	1.8	121	9.1-9.5	108.5	135	P
Glutamate	GLU	E	10.8	146	4.6	138.4	190	C-
Glutamine	GLN	Q	10.8	146		143.8	180	P
Glycine	GLY	G	7.8	75		60.1	75	-
Histidine	HIS	H	0.7	155	6.2	153.2	195	P, C+
Isoleucine	ILE	I	4.4	131		166.7	175	H
Leucine	LEU	L	7.8	131		166.7	170	H
Lysine	LYS	K	7.0	147	10.4	168.6	200	C+
Methionine	MET	M	3.8	149		162.9	185	H
Phenylalanine	PHE	F	3.3	165		189.9	210	H
Proline	PRO	P	4.6	115		112.7	145	H
Serine	SER	S	6.0	105		89.0	115	P
Threonine	THR	T	4.6	119		116.1	140	P
Tryptophan	TRP	W	1.0	204		227.8	255	P
Tyrosine	TYR	Y	2.2	181	9.7	193.6	230	P
Valine	VAL	V	6.0	117		140.0	155	H

Protein Structures from an NMR Perspective

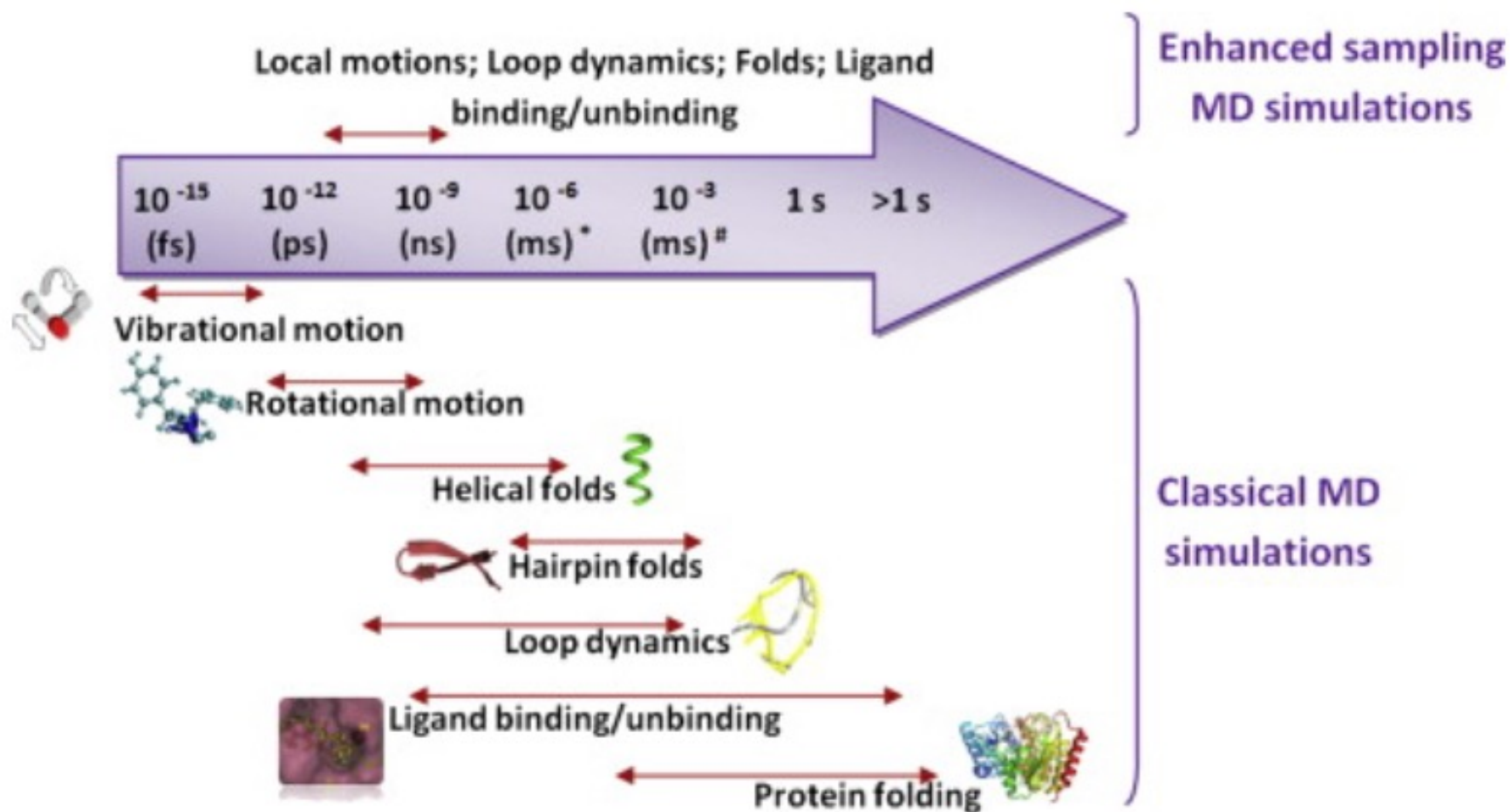*Some General Rules Regarding the Distribution of Amino Acids in Proteins:*

- Charged residues are hardly ever buried.
  - ▸ if buried generally involved in "salt-bridge"
- Polar residues are usually found on the surface of the protein, but can be buried.
  - ▸ if buried generally involved in hydrogen bond
- The inside, or core of a protein contains mostly non-polar

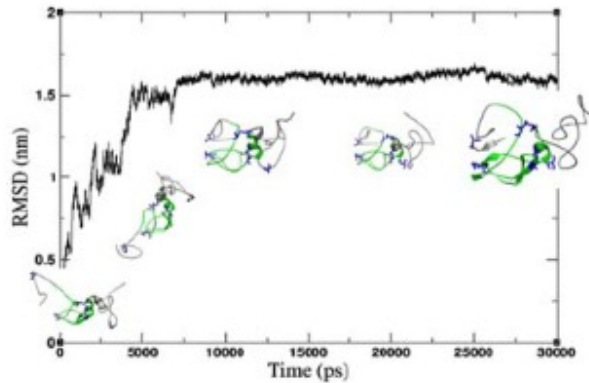Table A.8: Gibbs energy change in kilocalories per mole for transfer from water to 100% ethanol at 25°C[21].

a the outside of proteins

	$\Delta G_{transfer}$	
	Whole molecule	Side chain only
Glycine	+4.63	0
Alanine	+3.90	−0.73
Valine	+2.94	−1.69
Leucine	+2.21	−2.42
Isoleucine	+1.69	−2.97
Phenylalamine	+1.98	−2.65
Proline	+2.06	−2.60
Methionine	+3.33	−1.30
Tyrosine	+1.76	−2.87
Threonine	+4.19	−0.44
Serine	+4.59	−0.04
Asparagine	+4.64	+0.01
Glutamine	+4.73	+0.10

Local motions; Loop dynamics; Folds; Ligand binding/unbinding

10^{-15} (fs) 10^{-12} (ps) 10^{-9} (ns) 10^{-6} (ms)* 10^{-3} (ms)# 1 s >1 s

Enhanced sampling MD simulations

Vibrational motion

Rotational motion

Helical folds

Hairpin folds

Loop dynamics

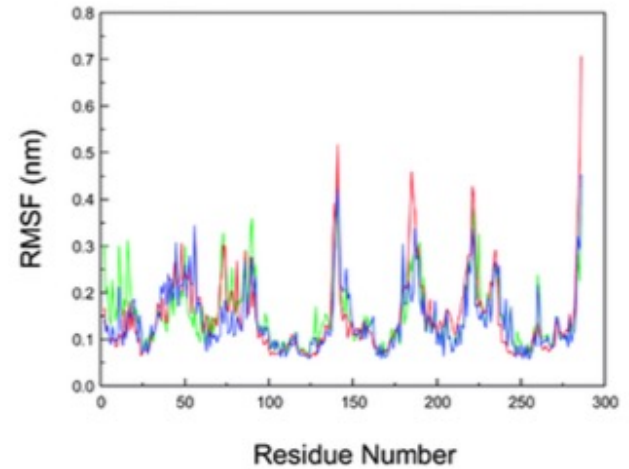Ligand binding/unbinding

Protein folding

Classical MD simulations

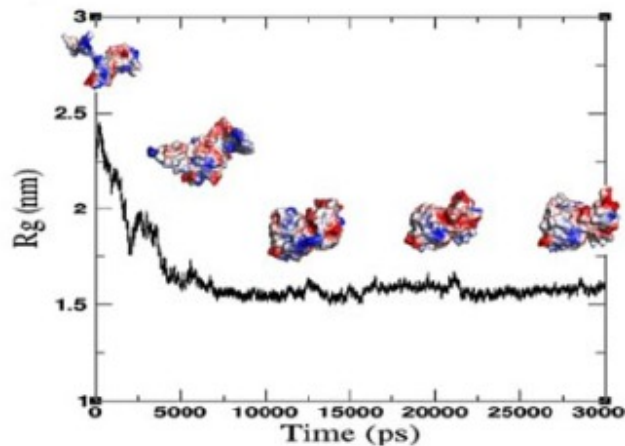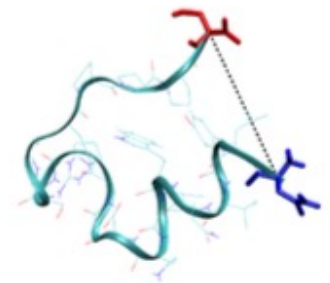# Other observables

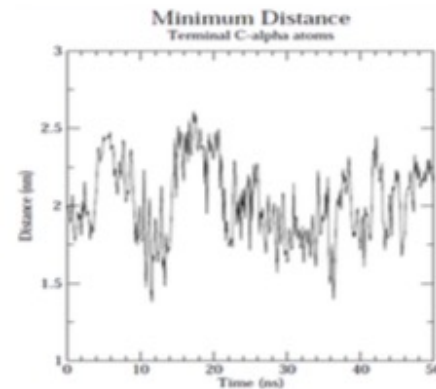

$$RMSD(t) = \left[\frac{1}{M} \sum_{i=1}^{N} m_i |\mathbf{r}_i(t) - \mathbf{r}_i^{\mathrm{ref}}|^2 \right]^{1/2}$$

$$RMSF_i = \left[\frac{1}{T} \sum_{t_j=1}^{T} |\mathbf{r}_i(t_j) - \mathbf{r}_i^{\mathrm{ref}}|^2 \right]^{1/2}$$

$$R_g = \left(\frac{\sum_i |\mathbf{r}_i|^2 m_i}{\sum_i m_i} \right)^2$$

To get information about contacts in the protein one can plot the **distances between two atoms** or the **minimum distance between two groups of atoms**

3D Structures of Biological Macromolecules
Exercise 1: Structural Comparison of Proteins

Jürgen Sühnel
jsuehnel@imb-jena.de

**Leibniz Institute for Age Research, Fritz Lipmann Institute,
Jena Centre for Bioinformatics
Jena / Germany**

Leibniz Institute for Age Research
Fritz Lipmann Institute (FLI)

PLoS one

Structural Relationships in the Lysozyme Superfamily: Significant Evidence for Glycoside Hydrolase Signature Motifs

Alexandre Wohlkönig[1], Joëlle Huet[2], Yvan Looze[2], René Wintjens[2,3]*

1 Structural Biology Brussels and Molecular and Cellular Interactions, VIB, Brussels, Belgium, 2 Laboratoire de Chimie Générale, Institut de Pharmacie, Université Libre de Bruxelles, Brussels, Belgium, 3 Interdisciplinary Research Institute, USR 3078 CNRS, Villeneuve d'Ascq, France

Abstract

Background: Chitin is a polysaccharide that forms the hard, outer shell of arthropods and the cell walls of fungi and some algae. Peptidoglycan is a polymer of sugars and amino acids constituting the cell walls of most bacteria. Enzymes that are able to hydrolyze these cell membrane polymers generally play important roles for protecting plants and animals against infection with insects and pathogens. A particular group of such glycoside hydrolase enzymes share some common features in their three-dimensional structure and in their molecular mechanism, forming the lysozyme superfamily.

Results: Besides having a similar fold, all known catalytic domains of glycoside hydrolase proteins of lysozyme superfamily (families and subfamilies GH19, GH22, GH23, GH24 and GH46) share in common two structural elements: the central helix of the all-α domain, which invariably contains the catalytic glutamate residue acting as general-acid catalyst, and a β-hairpin pointed towards the substrate binding cleft. The invariant β-hairpin structure is interestingly found to display the highest amino acid conservation in aligned sequences of a given family, thereby allowing to define signature motifs for each GH family. Most of such signature motifs are found to have promising performances for searching sequence databases. Our structural analysis further indicates that the GH motifs participate in enzymatic catalysis essentially by containing the catalytic water positioning residue of inverting mechanism.

Conclusions: The seven families and subfamilies of the lysozyme superfamily all have in common a β-hairpin structure which displays a family-specific sequence motif. These GH β-hairpin motifs contain potentially important residues for the catalytic activity, thereby suggesting the participation of the GH motif to catalysis and also revealing a common catalytic scheme utilized by enzymes of the lysozyme superfamily.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Rene.Wintjens@ulb.ac.be

HOMEWORK

Use Superpose to **determine the distance matrix** beteween representative structures of the lysozyme superfamily. Use the following PDB Ids:

GH19, GH22c, GH22i, GH23, GH24v, GH24l and GH46 were 3cql, 1iee, 2dqa, 153l, 2lzm, 1am7 and 1qgi, respectively.

Next monday we shall organize a journal club to discuss the papers by Volkand by Bae2004

Figure 1. Clustering tree of the 32 structures of lysozyme superfamily. Each enzyme is labelled by its protein code. GH families are indicated. A blue asterisk (*) is set to indicate an unreliable internal node according the jackknife test of Lanyon [70]. The scale below the tree indicated a length of 4.0 in the modified DaliLite scoring scheme (see method).

doi:10.1371/journal.pone.0015388.g001

Root Mean Square Deviation

- The RMSD is a measure to quantify structural similarity
- Requires 2 superimposed structures (designated here as "a" & "b")
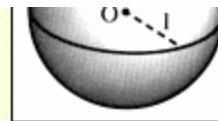- N = number of atoms being compared

$$RMSD = \sqrt{\frac{\Sigma\ (x_{ai} - x_{bi})^2 + (y_{ai} - y_{bi})^2 + (z_{ai} - z_{bi})^2}{\sqrt{N}}}$$

Two steps:

1. Identification of a set of related atom pairs
2. Superposition with minimum RMSD value

Superposition of two molecules

The problem of superpositioning two (or more) molecules onto each other is a common task and one would assume it is also very simple. In fact, there exist several techniques to solve this problem, but there are a few details to consider which can make coding of the problem subtle. In the following I will discuss a FORTRAN program which superimposes two pdb files using the elegant quaternion method as described by S.K.Kearsley [1].

Procedure

The superposition problem can be split up into two parts: A rotation around the geometric center of the molecule to bring it into the proper orientation, and a translation that superimposes the centers of the probe and target molecule(s).

The geometric center of the molecule can be found readily by averaging the x,y, and z coordinates of the n atoms k, respectively.

$$\langle x \rangle_k = \frac{1}{n} \sum_{k=1}^{n} \vec{x}_k$$

This center is not exactly the center of mass (we did not care what kind the atoms are, i.e., we did not calculate the moment). But a center defined this way will be the origin of the principal axes setting up the distance ellipsoid (similar to e.g., an anisotropic thermal parameter ellipsoid).
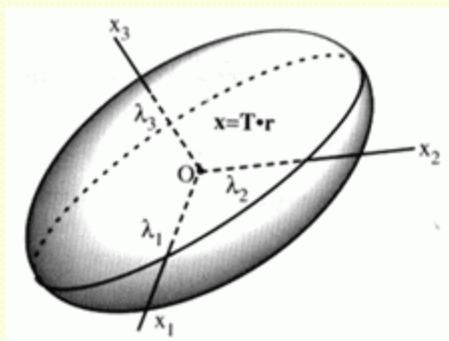
The (orthogonal) principal axes [2] can be obtained quite easily by orthogonalization of the matrix A (i.e., the matrix filled by the sum of the matrices of the metric tensor x) :

$$A_{ij} = \sum_{k=1}^{n} x_{(ij)k} \quad \text{where} \quad x_{(ij)} = \begin{vmatrix} xx & xy & xz \\ yx & yy & yz \\ zx & zy & zz \end{vmatrix}$$

by a Jacobi transformation. The Jacobi transformation (or orthogonalization) is an iterative application of rotations to a matrix until all the off-diagonal values are zero at machine precision [3]. If the matrix A is symmetric (which A of course is) then a diagonal matrix D exists so that

$$D = P^{-1}AP \quad \text{where D has the eigenvalues in its diagonal}: \quad D = \begin{vmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{vmatrix}$$

and P the normalized (orthogonal) eigenvectors of the principal axes in its columns. This orthogonalization is in fact equivalent to the solution of the eigenvalue problem [4]

$$A \cdot v = \lambda \cdot v$$

The feature of A is that it is real and symmetric means that a solution with real eigenvectors and eigenvalues must exist. This is good, but it is also the source of the *mmm* symmetry of the ellipsoid, a nasty property as we'll see shortly.

Once we have determined the center of each molecule, we can shift them to a common origin, usually (0,0,0). We are now faced with the problem to find the rotation that gives the best superposition of the molecules. What defines 'best' superposition? One might be tempted to just rotate the principal axes to an overlap. This could in principle work, and the rotation matrix to overlap the axes can be easily calculated by a Gauss-Jordan elimination [3] of three 3x3 linear systems. This does give the correct answer in cases where the molecules are pretty much pre-aligned, and the molecules can also have a different number of atoms. But as a result of the symmetry of the principal axes ellipsoid we have lost information about the orientation of the axes: we need to try 6 rotations (3 pairs of 180 deg rotated vectors) to find the actual solution. To determine the best one, we are back to a calculation of the pairwise distance r.m.s. (root mean square) between atoms which we hoped to avoid in the first place.

So do we have to solve this as a least squares problem of r.m.s. distances with all the problems associated (see, e.g. [1] for a discussion and further reading)? No, not really. The above idea using the principal axes rotation was quite elegant, but it needs to be improved. We need to find a way to resolve the degeneracy of the principal axes transformation. This can actually be done with an extension of the 3x3 problem of the metric tensor matrix to a 4x4 matrix which is constructed from a **non-cyclic** permutation of the pairwise **differences** between atom positions. The combination of elements follows the 4 constructors of the *Quaternion group H*, a small, non-cyclic subgroup of *GL2(C)* [5].

We solve our linear quaternion eigenvalue system again numerically by Jacobi transformation of the matrix Q

$$Q = \sum_{k=1}^{n} q_{(ij)k} \quad \text{, where again} \quad q_{(ij)} = \begin{vmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \\ q_{41} & q_{42} & q_{43} & q_{44} \end{vmatrix}$$

How the elements q are formed can be found in [1]. The smallest eigenvalue is the s.r.s.(sum of residuals squared) for the best rotation and the associated eigenvector contains element from which the 3x3 rotation matrix t for the best superposition is constructed. The largest EV contains the worst rotation, the associated eigenvalue the worst s.r.s. The Quaternion method is a very elegant way to solve the rotation problem and the r.m.s.d. least squares minimization in one shot!

The application of the rotation and the back-translation of the molecule are trivial final steps.

[1] S.K.Kearsley, *On the orthogonal transformation used for structural comparisons*, Acta Cryst. A45, 208 (1989)
[2] D.A.Danielson, *Vectors and Tensors in Engineering and Physics*, Addison-Wesley (1992), D.A.Sands, *Vectors and Tensors in Crystallography*, Dover Publishing (1982)
[3] W.H.Press, S.A.Teukolsky, W.T.Vettering and B.P. Flannery, *Numerical Recipes*, 2nd edition, Cambridge University Press (1992)
[4] S. Lipschutz, *Linear Algebra*, 2nd edition, Schaum Outline Series, McGraw-Hill (1991)
[5] M.Artin, *Algebra*, Prentice-Hall (1991)
[6] J.Hart, Quaternion demonstrator (SGI only), Stanford University

Definition

While the complex numbers are obtained by adding the element i to the real numbers which satisfies $i^2 = -1$, the quaternions are obtained by adding the elements i, j and k to the real numbers which satisfy the following relations.

$$i^2 = j^2 = k^2 = ijk = -1$$

If the multiplication is assumed to be associative (as indeed it is), the following relations follow directly:

$$\begin{aligned}
ij &= k, & ji &= -k, \\
jk &= i, & kj &= -i, \\
ki &= j, & ik &= -j.
\end{aligned}$$

(these are derived in detail below). Every quaternion is a real linear combination of the **basis quaternions** 1, i, j, and k, i.e. every quaternion is uniquely expressible in the form $a + bi + cj + dk$ where a, b, c, and d are real numbers. In other words, as a vector space over the real numbers, the set **H** of all quaternions has dimension 4, whereas the complex number plane has dimension 2. Addition of quaternions is accomplished by adding corresponding coefficients, as with the complex numbers. By linearity, multiplication of quaternions is completely determined by the multiplication table above for the basis quaternions. Under this multiplication, the basis quaternions, with their negatives, form the quaternion group of order 8, Q_8. The scalar part of the quaternion is a while the remainder is the vector part. Thus a **vector** in the context of quaternions has zero for scalar part.

resulting equations can be organized as an eigenvalue problem where $x_m = (x' - x)$, $x_p = (x' + x)$ with similar definitions for y_m, y_p, z_m and z_p.

$$
\begin{pmatrix}
\sum (x_m^2 + y_m^2 + z_m^2) & \sum (y_p z_m - y_m z_p) & \sum (x_m z_p - x_p z_m) & \sum (x_p y_m - x_m y_p) \\
\sum (y_p z_m - y_m z_p) & \sum (y_p^2 + z_p^2 + x_m^2) & \sum (x_m y_m - x_p y_p) & \sum (x_m z_m - x_p z_p) \\
\sum (x_m z_p - x_p z_m) & \sum (x_m y_m - x_p y_p) & \sum (x_p^2 + z_p^2 + y_m^2) & \sum (y_m z_m - y_p z_p) \\
\sum (x_p y_m - x_m y_p) & \sum (x_m z_m - x_p z_p) & \sum (y_m z_m - y_p z_p) & \sum (x_p^2 + y_p^2 + z_m^2)
\end{pmatrix}
\begin{pmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{pmatrix}
= \lambda
\begin{pmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{pmatrix}.
$$

Diagonalizing this symmetric matrix will give four orthogonal unit quaternions. The eigenvalues give the value of the residual for the rotation produced by application of the corresponding eigenvector. The r.m.s. deviation is given by $(\lambda/n)^{1/2}$ where n is the number of atoms compared. Smallest and largest eigenvalues give rotations that minimize and maximize the sum of the distances between all corresponding atoms.

Comparing Protein Structures
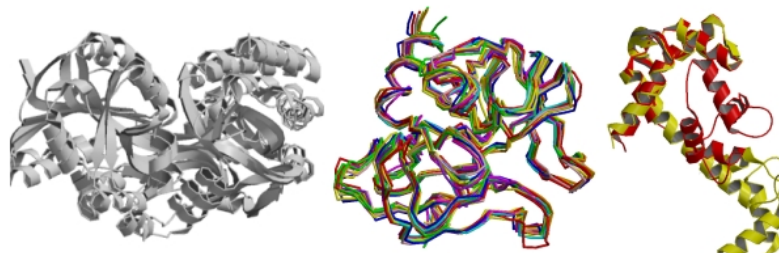
SuperPose Version 1.0

SuperPose is a protein superposition server. SuperPose calculates protein superpositions using a modified quaternion approach. From a superposition of two or more structures, SuperPose generates sequence alignments, structure alignments, PDB coordinates, RMSD statistics, Difference Distance Plots, and interactive images of the superimposed structures. The SuperPose web server supports the submission of either PDB-formatted files or PDB accession numbers.

Please cite the following: Rajarshi Maiti, Gary H. Van Domselaar, Haiyan Zhang, and David S. Wishart "SuperPose: a simple server for sophisticated structural superposition" Nucleic Acids Res. 2004 July 1; 32 (Web Server issue): W590W594. Click here for PDF.

If your PDB file contains multiple copies of a structure(ie. NMR files) you only need to enter one file or accession number. For additional information on how to run SuperPose, click [HELP]

---PDB Entry A---

Select the first PDB file [   ] [Durchsuchen...]

OR Enter a PDB accession number [   ]

---PDB Entry B (Optional)---

Select the 2nd PDB file [   ] [Durchsuchen...]

OR Enter a PDB accession number [   ]

Note: the uploaded file(s) must be in PDB format in order for this form to work.
Superimposing more than 2 files? Click here
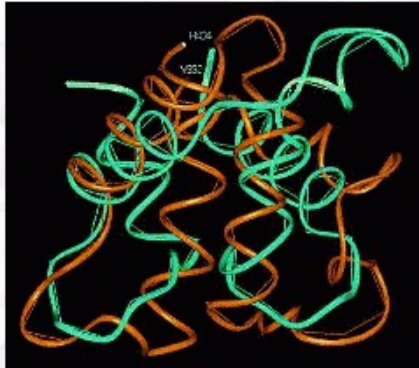[Submit]

Comparing Protein Structures

DaliLite | Pairwise comparison of protein structures

DaliLite is a program for pairwise structure comparison. Compare your structure(first structure) to a reference structure(second structure).

First Structure		Second Structure	
PDB entry code: [ ]	Chain ID: [ ]	PDB entry code: [ ]	Chain ID: [ ]
or upload a file in PDB format (.pdb,.ent,.dat,.brk)		or upload a file in PDB format (.pdb,.ent,.dat,.brk)	
[ ] Browse...		[ ] Browse...	
		Run DaliLite	Reset

Leibniz Institute for Age Research
Fritz Lipmann Institute (FLI)

http://www.ebi.ac.uk/DaliLite/

Comparing Protein Structures

JCB

Structural similarity between Acetylcholinesterase and Calmodulin found using CE (Tsigelny et al, *Prot Sci*, 2000, 9:180)

Databases and Tools for 3-D Protein Structure Comparison and Alignment

Using the Combinatorial Extension (CE) Method

?
More Info

Select from the following options by clicking the links on the right

FIND   Find structural alignments by selecting from **ALL** or **REPRESENTATIVES** from the PDB.

CALCULATE   Calculate structural alignment for **TWO CHAINS** either from the PDB or uploaded by the user. Calculate structural neighbors for one protein **UPLOADED BY THE USER AGAINST THE PDB.**
Calculate **MULTIPLE STRUCTURE ALIGNMENT.**

DOWNLOAD   Download the **SOFTWARE** and **DATABASES** used here for local use.

RESULTS   Review results from our own alignment experiments: **SUBDOMAINS [pdf]** | **PROTEIN KINASES [pdf]** | **ESTERASES, LIPASES [pdf]**

Leibniz Institute for Age Research
Fritz Lipmann Institute (FLI)

http://cl.sdsc.edu/

Beginning with an input PDB file or set of files, SuperPose first extracts the sequences of all chains in the file(s). Each sequence pair is then aligned using a Needleman–Wunsch pairwise alignment algorithm. If the pairwise sequence identity falls below the default threshold (25%), SuperPose determines the secondary structure using VADAR (volume, area, dihedral angle reporter) and performs a secondary structure alignment using a modified Needleman–Wunsch algorithm.

After the sequence or secondary structure alignment is complete, SuperPose then generates a difference distance (DD) matrix between aligned alpha carbon atoms. A difference distance matrix can be generated by first calculating the distances between all pairs of C atoms in one molecule to generate an initial distance matrix. A second pairwise distance matrix is generated for the second molecule and, for equivalent/aligned Calpha atoms, the two matrices are subtracted from one another, yielding the DD matrix. From the DD matrix it is possible to quantitatively assess the structural similarity/dissimilarity between two structures. In fact, the difference distance method is particularly good at detecting domain or hinge motions in proteins. SuperPose analyzes the DD matrices and identifies the largest contiguous domain between the two molecules that exhibits <2.0 Å difference.

From the information derived from the sequence alignment and DD comparison, the program then makes a decision regarding which regions should be superimposed and which atoms should be counted in calculating the RMSD. This information is then fed into the quaternion superposition algorithm and the RMSD calculation subroutine. The quaternion superposition program is written in C and is based on both Kearsley's method and the PDBSUP Fortran program developed by Rupp and Parkin. Quaternions were developed by W. Hamilton (the mathematician/physicist) in 1843 as a convenient way to parameterize rotations in a simple algebraic fashion. Because algebraic expressions are more rapidly calculable than trigonometric expressions using computers, the quaternion approach is exceedingly fast.

SuperPose can calculate both pairwise and multiple structure superpositions [using standard hierarchical methods and can generate a variety of RMSD values for alpha carbons, backbone atoms, heavy atoms and all atoms (average and pairwise). When identical sequences are compared, SuperPose also generates 'per residue' RMSD tables and plots to allow users to identify, assess and view individual residue displacements.

http://wishart.biology.ualberta.ca/SuperPose/

JCB

Identical/same sequence but different structure

Calmodulin: **1A29 vs. 1CLL**
                                    (open and closed form)

Similar structure but slightly different sequence length

Thioredoxin: **3TRX vs. 2TRX_a**

Similar structure but extremely different sequence

Thioredoxin/Glutaredoxin: **3TRX vs. 3GRX_1**

Calmodulin

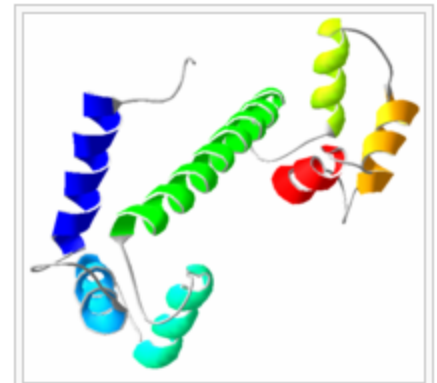*Your continued donations keep Wikipedia running!*

Calmodulin

From Wikipedia, the free encyclopedia

Calmodulin (CaM) is a Ca^{2+}-binding protein that is a key component of the Ca^{2+} second-messenger system and is involved in controlling many of the biochemical processes of cells.

Calmodulin is a small, acidic protein approximately 148 amino acids long (16706 Dalton) and, as such, is a favorite for testing protein simulation software. It contains four **EF-hand** "motifs" or domains, each of which binds a Ca^{2+} ion. It typically binds 0, 2, or 4 calcium ions, and binds and regulates different proteins in each state. There are over a hundred proteins known to bind calmodulin. It is highly conserved across all eukaryotes, and its expression is essential for biological cells to progress through mitosis.

Calmodulin-stimulated protein phosphatase (EC 3.1.3.16) and calmodulin-dependent kinases are the major calmodulin-binding proteins in the brain.

Calmodulin 3D structure

WIKIPEDIA
The Free Encyclopedia

Leibniz Institute for Age Research
Fritz Lipmann Institute (FLI)

Thioredoxin

InterPro

Glutaredoxins [1 , 2 , 3], also known as thioltransferases, are small proteins of approximately one hundred amino-acid residues. Glutaredoxin functions as an electron carrier in the glutathione-dependent synthesis of deoxyribonucleotides by the enzyme ribonucleotide reductase. Like thioredoxin, which functions in a similar way, glutaredoxin possesses an active center disulphide bond. It exists in either a reduced or an oxidized form where the two cysteine residues are linked in an intramolecular disulphide bond.

Glutaredoxin has been sequenced in a variety of species. On the basis of extensive sequence similarity, it has been proposed [4] that vaccinia protein O2L is most probably a glutaredoxin. Finally, it must be noted that phage T4 thioredoxin seems also to be evolutionary related.

Measures for Numerical Data

The choice of distances is important for applications, and the best choice is often achieved via combination of experience, skill, knowledge, and luck. Here we list some commonly used distances.

$$d_{euc}(\mathbf{x}, \mathbf{y}) = \left[\sum_{j=1}^{d} (x_j - y_j)^2 \right]^{\frac{1}{2}} = \left[(\mathbf{x} - \mathbf{y})(\mathbf{x} - \mathbf{y})^T \right]^{\frac{1}{2}} \qquad \textbf{Euclidean Distance}$$

$$d_{man}(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^{d} |x_j - y_j|. \qquad \textbf{Manhattan Distance}$$

$$d_{max}(\mathbf{x}, \mathbf{y}) = \max_{1 \leq k \leq d} |x_j - y_j|. \qquad \textbf{Maximum Distance}$$

$$d_{min}(\mathbf{x}, \mathbf{y}) = \left(\sum_{j=1}^{d} |x_j - y_j|^r \right)^{\frac{1}{r}}, \quad r \geq 1. \qquad \textbf{Minkowski Distance}$$

$$d_{mah}(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})\Sigma^{-1}(\mathbf{x} - \mathbf{y})^T}, \qquad \textbf{Mahalanobis Distance}$$

$$d_{ave}(\mathbf{x}, \mathbf{y}) = \left(\frac{1}{d} \sum_{j=1}^{d} (x_j - y_j)^2 \right)^{\frac{1}{2}}. \qquad \textbf{Average Distance}$$

From a distance matrix we can infer/predict
A topology in the space of protein structures…

THAT'S ALL FOLKS!