JCB

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

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

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

Comparing Protein Structures

JCB

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].

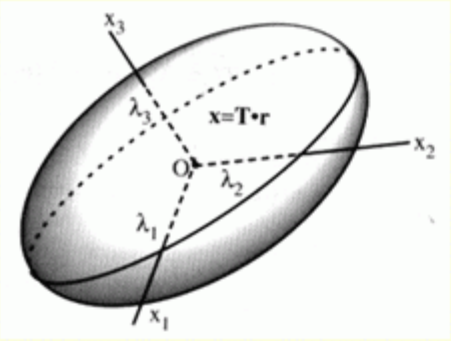
Procedure

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) :

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

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

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

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.

Comparing Protein Structures

JCB

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

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

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.

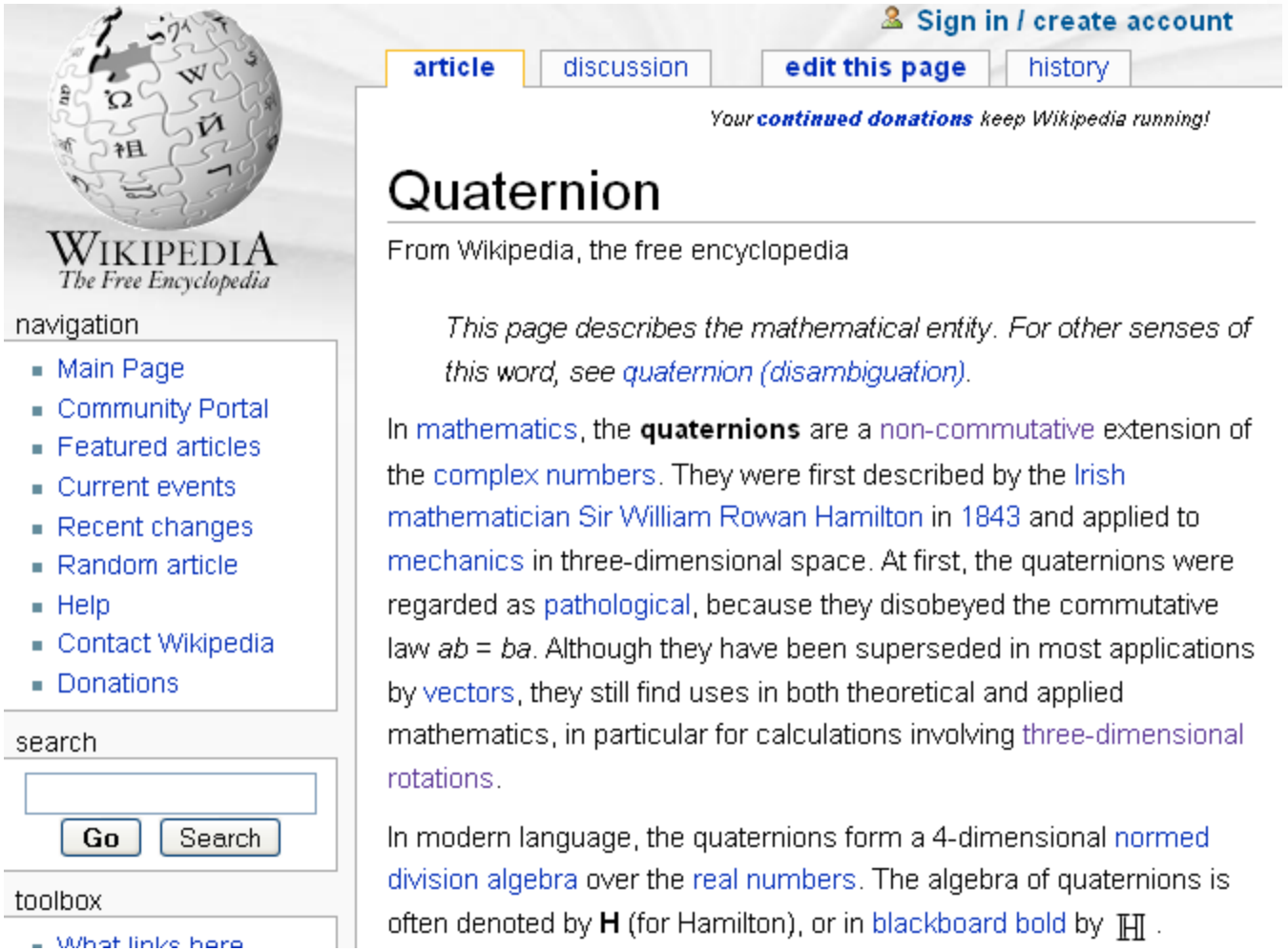
---

[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

---

Quaternions

JCB



Sign in / create account

article | discussion | edit this page | history

Your **continued donations** keep Wikipedia running!

WIKIPEDIA
The Free Encyclopedia

navigation

- Main Page
- Community Portal
- Featured articles
- Current events
- Recent changes
- Random article
- Help
- Contact Wikipedia
- Donations

search

Go | Search

toolbox

- What links here

Quaternion

From Wikipedia, the free encyclopedia

This page describes the mathematical entity. For other senses of this word, see quaternion (disambiguation).

In mathematics, the **quaternions** are a non-commutative extension of the complex numbers. They were first described by the Irish mathematician Sir William Rowan Hamilton in 1843 and applied to mechanics in three-dimensional space. At first, the quaternions were regarded as pathological, because they disobeyed the commutative law $ab = ba$. Although they have been superseded in most applications by vectors, they still find uses in both theoretical and applied mathematics, in particular for calculations involving three-dimensional rotations.

In modern language, the quaternions form a 4-dimensional normed division algebra over the real numbers. The algebra of quaternions is often denoted by **H** (for Hamilton), or in blackboard bold by $\mathbb{H}$.

Quaternions

JCB

Definition

[edit]

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.

Comparing Protein Structures

Acta Cryst. (1989). A**45**, 208–210

On the orthogonal transformation used for structural comparisons. By SIMON K. KEARSLEY, *Department of Chemistry, Yale University, 225 Prospect Street, New Haven, CT* 06511, *USA*

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.

$$\begin{pmatrix} \sum (x_m^2 + y_m^2 + z_m^2) & \sum (y_p z_m - y_m z_p) & \sum (x_m z_p - x_p z_m) \\ \sum (y_p z_m - y_m z_p) & \sum (y_p^2 + z_p^2 + x_m^2) & \sum (x_m y_m - x_p y_p) \\ \sum (x_m z_p - x_p z_m) & \sum (x_m y_m - x_p y_p) & \sum (x_p^2 + z_p^2 + y_m^2) \\ \sum (x_p y_m - x_m y_p) & \sum (x_m z_m - x_p z_p) & \sum (y_m z_m - y_p z_p) \end{pmatrix}$$

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

Comparing Protein Structures

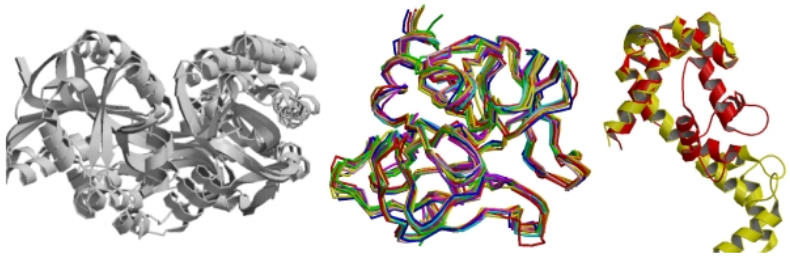
JCB

Acta Cryst. (1989). A**45**, 208–210

On the orthogonal transformation used for structural comparisons. By SIMON K. KEARSLEY, *Department of Chemistry, Yale University, 225 Prospect Street, New Haven, CT* 06511, *USA*

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)

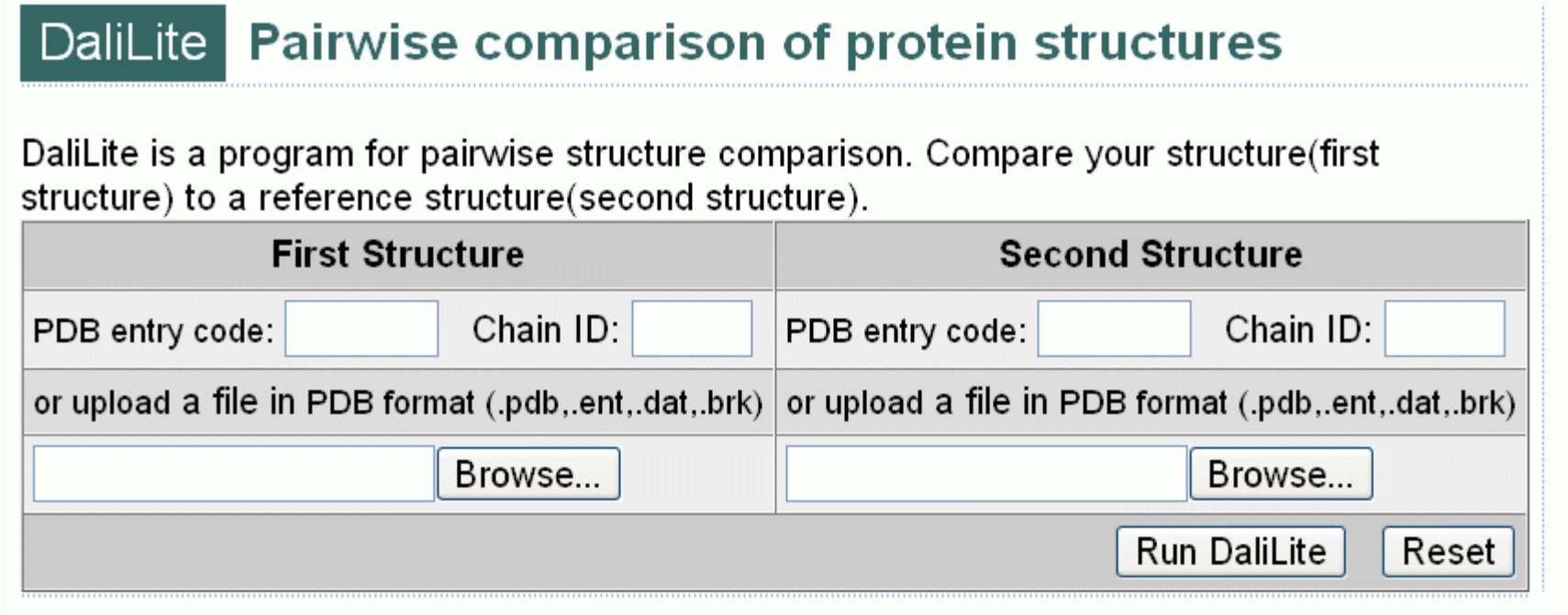
Select the 2nd PDB file Durchsuchen...

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

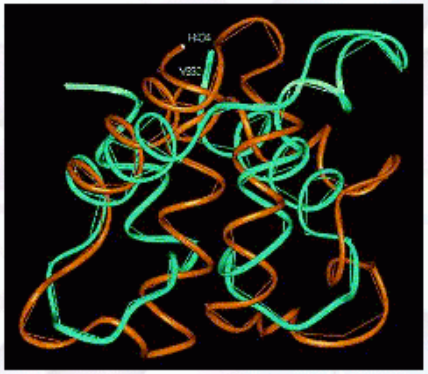
Submit

Comparing Protein Structures



Comparing Protein Structures

JCB



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

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

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

Using the Combinatorial Extension (CE) Method

More Info

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]

Comparing Protein Structures – SuperPose Server

JCB

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.

SuperPose Server: Examples

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

JCB

WIKIPEDIA
The Free Encyclopedia

navigation

- Main Page
- Community Portal
- Featured articles
- Current events
- Recent changes
- Random article
- Help
- Contact Wikipedia
- Donations

search

article | discussion | edit this page | history

Sign in / create account

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.

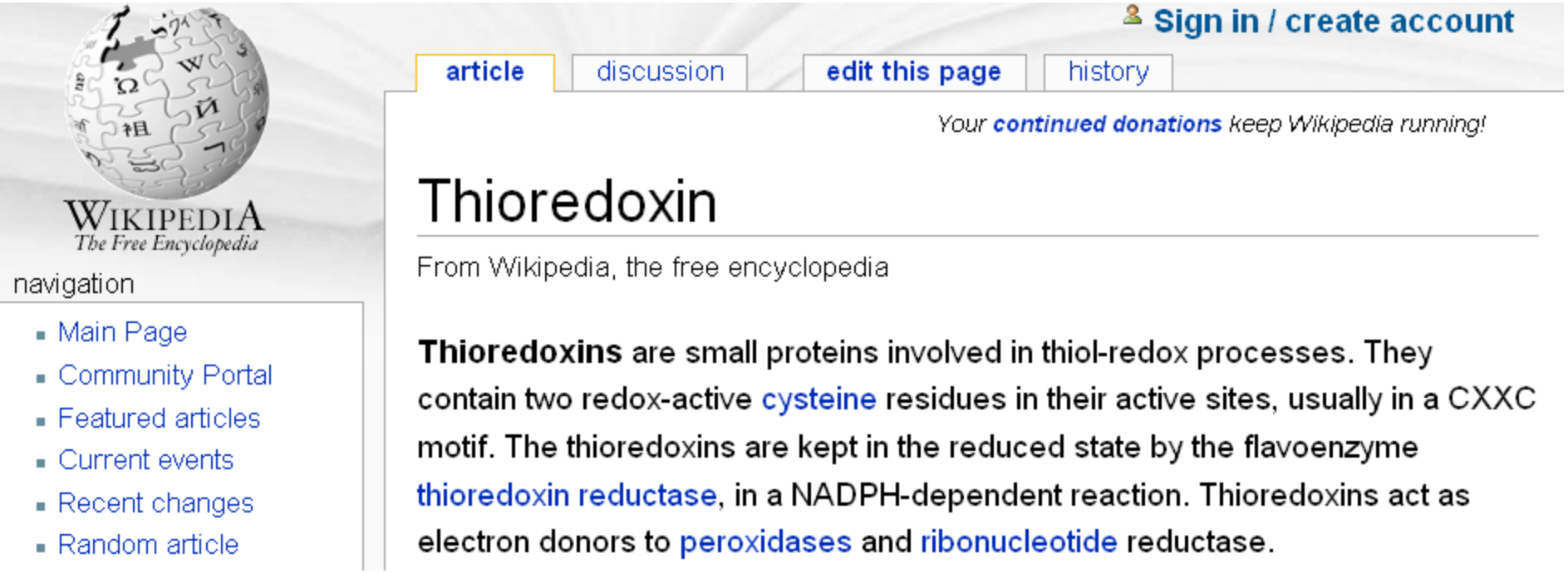
Calmodulin is a small, acidic protein approximately 148 amino acids long (16706 Dalton) and, as such, is a favorite for testing protein simulation software. It contains four **EF-hand** "motifs" or domains, each of which binds a Ca^{2+} ion. It typically binds 0, 2, or 4 calcium ions, and binds and regulates 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

Thioredoxin

JCB



Glutaredoxin

JCB

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.