

## Parameter estimation based on protein structural information with qPIPSA

For an introduction into the methodology and the underlying theory, the user is referred to the SYCAMORE documentation and the original scientific publications.

### 4.1. Modelling the first step of glycolysis

1<sup>st</sup> Step of Glycolysis: Glucose + ATP -> Glucose-6-phosphate + ADP

1. Modelling scenario:

1. You are interested in modelling this reaction for the liver-specific isozyme hexokinase IV in *Homo sapiens*.
2. You found kinetic data (for example the  $K_m$  value of the substrate glucose) for the glucokinase from yeast and rat but are not sure which to use in your mathematical model.
3. Perform an estimation of the  $K_m$  parameter of the human hexokinase IV using available protein structure for that enzyme and available kinetic data for other enzymes.
4. Register at the **Parameter Estimation** page by clicking on **New Registration** or enter your user name and password and press **Enter**

File Edit View Go Bookmarks Tools Help

http://sycamore.embl.org/sycamore/frameset.jsp

Firefox Help Firefox Support Plug-in FAQ

- Global parameter
  - Rules
  - Function def.
  - Unit definitions
  - Pathway map
- ▼ **Compartments**
  - All compartments
- ▼ **Reactions**
  - All reactions
- ▼ **Refine & analyze model**
  - Completeness
  - Sensitivity analysis
  - ▼ **Model simulation**
    - Copasi
    - General
  - ▼ **Parameter estimation**
    - Start**
    - User guide
- ▼ **Save model**
  - View XML code
  - Save on disk
  - Save as project
- **Resources**
  - Registration
  - User guide
  - Use case
  - Imprint
  - Home

### Registration

We request registration because parameter estimations may take between a couple of minutes and - in case of very complex calculations - several days. In addition, because CPU capacity is restricted, all requests are queued, thus it might be that it takes some time even for simple estimations. In any case, you will be notified by email on the result of the estimation(s), therefore an email address is required. Without email registration, parameter calculations will not be performed.

In order to access (load, view, edit, save) your personal models, you must register. In case that you are not registered yet, choose 'new registration'.

**REGISTERED USERS**

Name:    
(user name) (password)

[I can't remember user name / password](#)

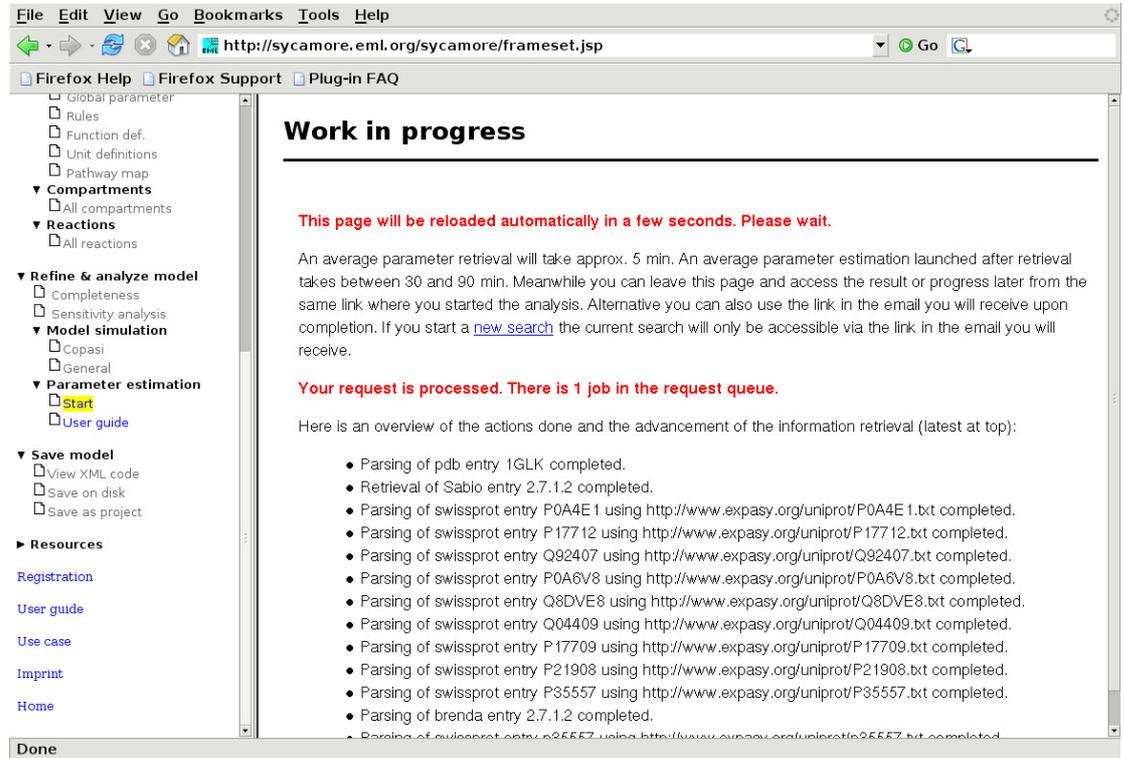
Done

5. In this example you are querying for the liver-specific glucokinase (hexokinase IV or D) from *Homo sapiens*.

1. Insert the SwissProt identifier of that amino acid sequence

- (P35557).
- If you do not know the SwissProt identifier of your sequence of interest, go to [www.expasy.org](http://www.expasy.org) and query for it.
  - Click on the **Start search** button.

You will then see a screen like this:



- The server is querying databases which contain protein structures (the PDB, Modbase and SwissModel) from X-ray analysis or theoretical models and BRENDA and SABIO-RK for existing kinetic information. Results are reported on

- Structural Information**
- Region selection**
- Kinetic data**

The screenshot shows the SYCAMORE web interface. The browser address bar displays <http://sycamore.eml.org/sycamore/frameset.jsp>. The left sidebar contains a navigation menu with categories like 'Global parameter', 'Rules', 'Function def.', 'Unit definitions', 'Pathway map', 'Compartment', 'Reactions', 'Refine & analyze model', 'Model simulation', 'Parameter estimation', 'Save model', and 'Resources'. The main content area is titled 'Structural information' and contains a section for 'Models'. A text box indicates 'Found 5 protein structures. 5 come from PDB.' Below this is a table with columns for '#', 'Seq. Max Cov. Max Iden.', 'Chains (Seq. Cov./ Seq. Iden./ Align. Score)', and 'Details'. Two rows are visible: #1 for PDB code 1GLK and #2 for PDB code 1V4S.

#	Seq. Max Cov. Max Iden.	Chains (Seq. Cov./ Seq. Iden./ Align. Score)	Details
#1	98% 100%	1GLK	Accession: 1GLK Resolution: NOT APPLICABLE Method: THEORETICAL MODEL Heteroatoms: • GLUCOSE [GLC]
#2	96% 99%	P35557 P35557 PDB from 14 to 461 ( 96% / 99% / 990 )	Accession: 1V4S Resolution: 2.30 Å Method: X-RAY DIFFRACTION R-factor: 0.232 Organism: HOMO SAPIENS Crystallization conditions: PEG 1500, HEPES, PH 6.6, VAPOR DIFFUSION, HANGING DROP, TEMPERATURE 293K EC references: • 2.7.1.1 Heteroatoms: • 2-AMINO-4-FLUORO-5-[(1-METHYL-1H-IMIDAZOL-2-YL) [MRK] • ALPHA-D-GLUCOSE [AGC] • SODIUM ION [NA]

5. SYCAMORE found 5 protein structures with relevance to the human hexokinase IV. 5 come from the PDB, none from ModBase or SwissModel.

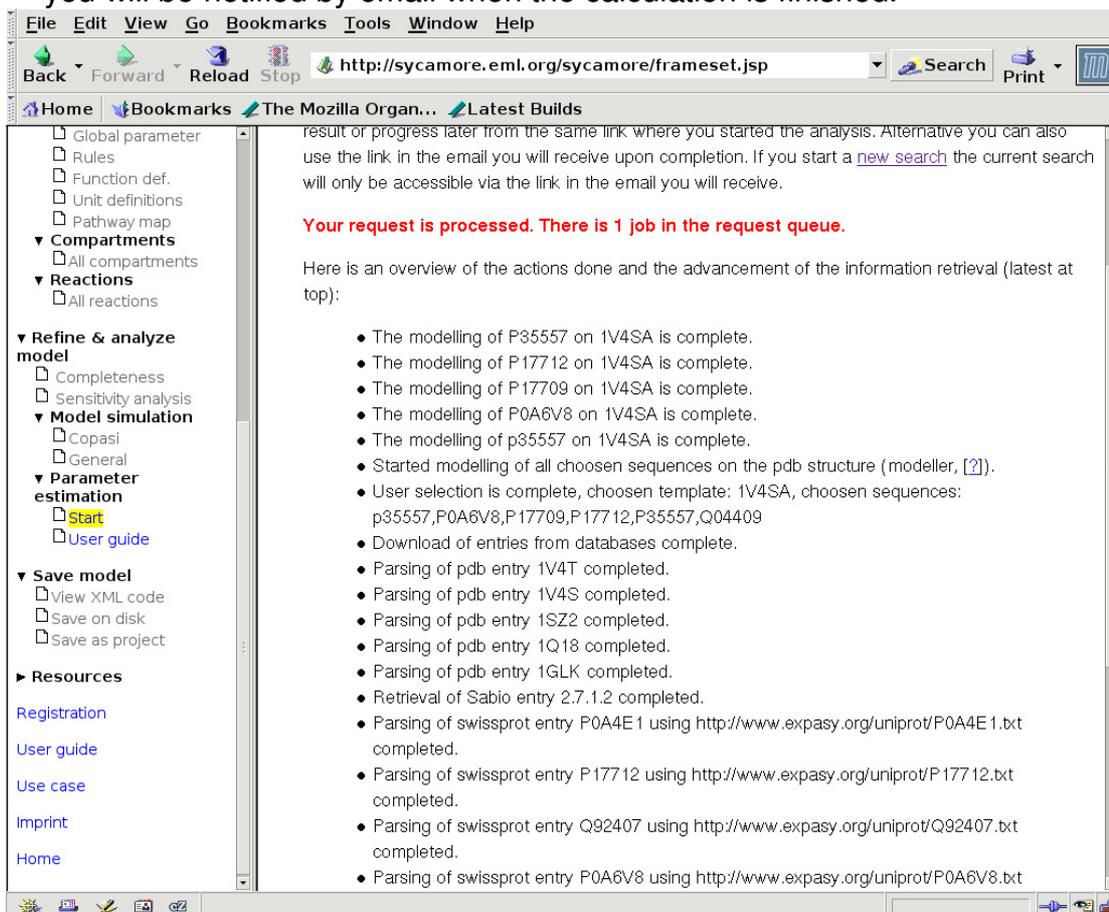
The retrieved structures are

1. 1GLK (theoretical model of human beta-cell glucokinase)
  2. 1V4S (*H. sapiens*; closed form)
  3. 1V4T (*H. sapiens*; super-open form)
  4. 1Q18 (*E. coli*)
  5. 1SZ2 (*E. coli*)
6. Choose **1V4S** as a PDB template structure for subsequent structural modelling based on sequence identity with SwissProt query and completeness of X-ray structure.
  7. Open the **Region Selection** submenu. Here, annotations from the SwissProt entry P35557 are reported. Select **Region Glucose-binding (Potential)** in order to also perform a comparison restricted to that region.
  8. Scroll down further to the **Kinetic data** section.
    1. Kinetic data are retrieved from BRENDA and SABIO-RK. By default, all entries are ticked as “selected”.
    2. **Deselect** several entries from BRENDA. Leave all entries from SABIO-RK “on”:
      1. “On” should be

1. *Homo sapiens*
2. *Saccharomyces cerevisiae*
3. *E. coli*
4. *Rattus norvegicus*

9. Go to the top of the page and click on “**Calculate electrostatically similar proteins**” to start a comparison of the selected protein structures and their properties.

10. The calculation starts. The web site will refresh automatically and you will be notified by email when the calculation is finished.



11. The output is a tree-like diagram like the following which positions your query sequence (P35557) in relation to the other chosen organisms for which kinetic data are available.

1. The first dendogram is for the electrostatic potential of the entire protein.

File Edit View Go Bookmarks Tools Window Help

Back Forward Reload Stop http://sycamore.eml.org/sycamore/frameset.jsp Search Print

Home Bookmarks The Mozilla Organ... Latest Builds

Reaction Search  
Documentation

▼ View & edit model

▼ Model

- Model description
- Global parameter
- Rules
- Function def.
- Unit definitions
- Pathway map

▼ Compartments

- All compartments

▼ Reactions

- All reactions

▼ Refine & analyze model

- Completeness
- Sensitivity analysis

▼ Model simulation

- Copasi
- General

▼ Parameter estimation

- Start
- User guide

▼ Save model

- View XML code
- Save on disk
- Save as project

► Resources

Registration

User guide

Use case

Imprint

Home

Total Protein REGION Structure view

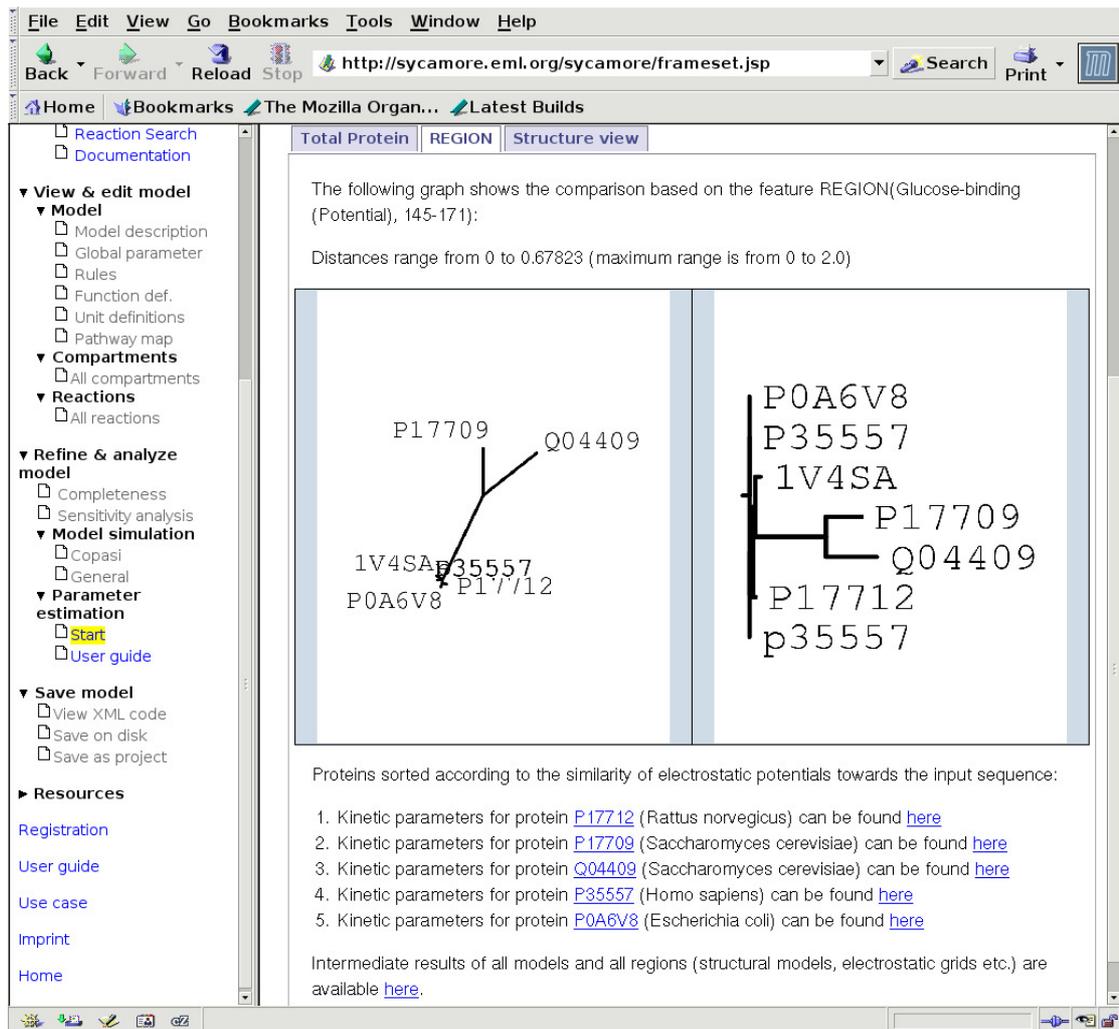
The following graph shows the comparison based on the total protein:  
Distances range from 0 to 0.78294 (maximum range is from 0 to 2.0)

Proteins sorted according to the similarity of electrostatic potentials towards the input sequence:

1. Kinetic parameters for protein [P35557](#) (Homo sapiens) can be found [here](#)
2. Kinetic parameters for protein [P0A6V8](#) (Escherichia coli) can be found [here](#)
3. Kinetic parameters for protein [P17712](#) (Rattus norvegicus) can be found [here](#)
4. Kinetic parameters for protein [Q04409](#) (Saccharomyces cerevisiae) can be found [here](#)
5. Kinetic parameters for protein [P17709](#) (Saccharomyces cerevisiae) can be found [here](#)

Intermediate results of all models and all regions (structural models, electrostatic grids etc.) are available [here](#).

2. The second dendrogram is for the selected Region “glucose-binding”



The electrostatic is conserved for enzymes from human (P35557), rat (P17712), and *E. coli* (P0A6V8). The enzymes from yeast (P17709 and Q04409) exhibit electrostatic potentials distant from the others.

This puts your query sequence P35557 very close to the hexokinase from rat (P17712), and remote from yeast (P17709, Q04409). Thus, kinetic parameters from rat could be used if you were to model the hexokinase IV from *Homo sapiens*.

This agrees with experimentally available  $K_m$  values for of 7.4 and 7.7 mM for D-glucose from rat; and 6 and 8.6 mM for D-glucose from human.

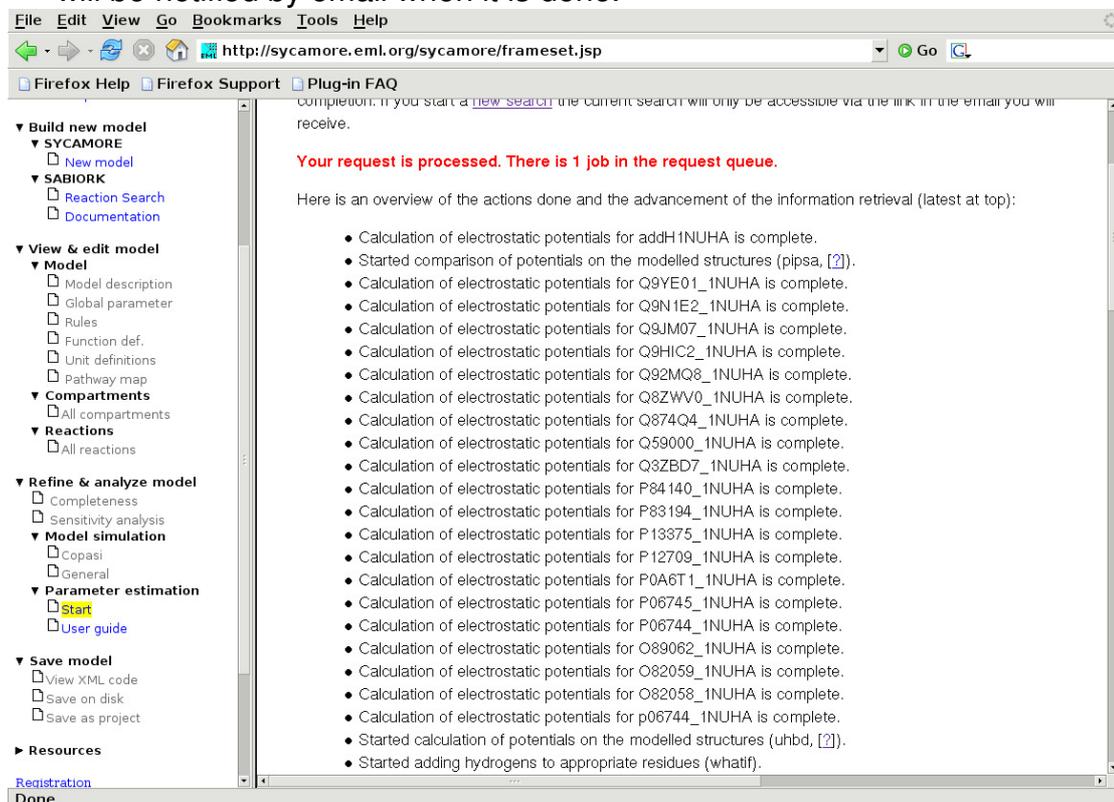
The glucokinase from *E. coli* exhibits a drastically lower  $K_m$  value of 0.78 mM for D-glucose and points to a different physiological role and enzymatic mechanism despite a conserved electrostatic potential near the glucose binding site.

The  $K_m$  value for D-glucose from yeast is 0.028 mM is far from the ones for rat and human.

## 4.2 Modelling the second step of glycolysis

Glucose-6-phosphate isomerase (G6PI) catalyzes the reaction  
glucose-6-phosphate → fructose-6-phosphate

1. You are querying for the  $K_m$  value of the substrate of the glucose-6-phosphate isomerase from *Homo sapiens*.
2. Insert the SwissProt identifier for that enzyme (P06744).
3. In **Structural Information**, 30 homologous protein structures are found in the PDB with the same EC number 5.3.1.9
  1. Select the entry “1NUH” as a structural template based on sequence identity with query sequence (both from *Homo sapiens*) and co-crystallization with a transition state analogue compound.
4. In **Region Selection**, select “ACT\_SITE” “By similarity” “From residue 389 to residue 389” as a centre for electrostatic potential comparison.
5. Select BRENDA and SABIO-RK entries for which  $K_m$  values for the substrate glucose-6-phosphate are available.
6. Click on “**Calculate electrostatically similar proteins**”
7. The calculation will start. The web site will refresh automatically and you will be notified by email when it is done.



The screenshot shows a web browser window with the URL <http://sycamore.eml.org/sycamore/frameset.jsp>. The left sidebar contains a navigation menu with categories like 'Build new model', 'View & edit model', 'Refine & analyze model', and 'Save model'. The main content area displays a message: 'Your request is processed. There is 1 job in the request queue.' Below this, it lists 30 completed calculations of electrostatic potentials for various protein structures, including P06744, P06744\_1NUHA, and others. The list ends with 'Started adding hydrogens to appropriate residues (whatif)'.

8. The results are two tree-like diagrams:

# 1. One for the electrostatic potential of the entire protein

The screenshot shows the SYCAMORE web interface. The browser address bar displays <http://sycamore.eml.org/sycamore/frameset.jsp>. The interface includes a navigation menu on the left with sections for 'Load existing model', 'Build new model', 'View & edit model', 'Refine & analyze model', and 'Save model'. The main content area has tabs for 'Total Protein', 'ACT\_SITE', and 'Structure view'. The 'Total Protein' tab is active, displaying a dendrogram comparing protein electrostatic potentials. The dendrogram shows a central node with branches leading to various protein IDs. The proteins are sorted by similarity to the input sequence. Below the dendrogram, a list of 12 proteins is provided with links to their kinetic parameters.

Total Protein | ACT\_SITE | Structure view

The following graph shows the comparison based on the total protein:  
Distances range from 0 to 1.20291 (maximum range is from 0 to 2.0)

Proteins sorted according to the similarity of electrostatic potentials towards the input sequence:

1. Kinetic parameters for protein [P06744](#) (Homo sapiens) can be found [here](#)
2. Kinetic parameters for protein [P06745](#) (Mus musculus) can be found [here](#)
3. Kinetic parameters for protein [Q9JM07](#) (Mus musculus) can be found [here](#)
4. Kinetic parameters for protein [Q9N1E2](#) (Oryctolagus cuniculus) can be found [here](#)
5. Kinetic parameters for protein [Q9YE01](#) (Aeropyrum pernix) can be found [here](#)
6. Kinetic parameters for protein [Q3ZBD7](#) (Bos taurus) can be found [here](#)
7. Kinetic parameters for protein [Q89062](#) (Mus musculus) can be found [here](#)
8. Kinetic parameters for protein [Q82058](#) (Spinacia oleracea) can be found [here](#)
9. Kinetic parameters for protein [P12709](#) (Saccharomyces cerevisiae) can be found [here](#)
10. Kinetic parameters for protein [Q874Q4](#) (Aspergillus niger) can be found [here](#)
11. Kinetic parameters for protein [Q9HIC2](#) (Thermoplasma acidophilum) can be found [here](#)
12. Kinetic parameters for protein [Q8ZWV0](#) (Pyrobaculum aerophilum) can be found [here](#)

# 2. One for the electrostatic potential around the chosen region (ACT\_SITE residue 389).

File Edit View Go Bookmarks Tools Help

http://sycamore.eml.org/sycamore/frameaset.jsp

Firefox Help Firefox Support Plug-in FAQ

▼ Load existing model

- Model from disk
- Model from projects
- Example models

▼ Build new model

- SYCAMORE
  - New model
- SABIORK
  - Reaction Search
  - Documentation

▼ View & edit model

- Model
  - Model description
  - Global parameter
  - Rules
  - Function def.
  - Unit definitions
  - Pathway map
- Compartments
  - All compartments
- Reactions
  - All reactions

▼ Refine & analyze model

- Completeness
- Sensitivity analysis
- Model simulation
  - Copasi
  - General
- Parameter estimation
  - Start
  - User guide

▼ Save model

- View XML code
- Save on disk
- Save as project

► Resources

- Registration
- User guide
- Use case
- Imprint
- Home

Done

interpretation of the output can be found in the [STAPAC help](#).

Please click on the tabs to get results based on regions of the protein (in case you have selected regions for analysis before). In case you print this document, all graphs will be shown.

Total Protein ACT\_SITE Structure view

The following graph shows the comparison based on the feature ACT\_SITE (By similarity, 389-389):

Distances range from 0 to 1.40428 (maximum range is from 0 to 2.0)

Proteins sorted according to the similarity of electrostatic potentials towards the input sequence:

1. Kinetic parameters for protein [P06744](#) (Homo sapiens) can be found [here](#)
2. Kinetic parameters for protein [P0A6T1](#) (Escherichia coli) can be found [here](#)
3. Kinetic parameters for protein [Q9JM07](#) (Mus musculus) can be found [here](#)
4. Kinetic parameters for protein [Q9HIC2](#) (Thermoplasma acidophilum) can be found [here](#)
5. Kinetic parameters for protein [Q92MQ8](#) (Rhizobium melliotti) can be found [here](#)
6. Kinetic parameters for protein [Q8ZWV0](#) (Pyrobaculum aerophilum) can be found [here](#)
7. Kinetic parameters for protein [Q874Q4](#) (Aspergillus niger) can be found [here](#)
8. Kinetic parameters for protein [O82059](#) (Spinacia oleracea) can be found [here](#)

## 9. Results

1. The electrostatic potential at the active site of the query sequence P06744 (Homo sapiens) is closest to the ones from mouse (P06745), bovine (Q3ZBD7) and rabbit (Q9N1E2).
2. Most distant are *M. janaschii* (Q59000), spinach (O82058), *B. stearothermophilus* (P13375), *P. furiosus* (P83194), *Th.litoralis* (P84140).
3. This is in agreement with the experimentally measured  $K_m$  values for glucose-6-phosphate:
  1. *Homo sapiens* 0.445 mM
  2. *Bos taurus* 0.45-0.58 mM
  3. *Mus musculus* 0.48 mM
  4. *Oryctolagus cuniculus* 0.6 mM
  5. *M. janaschii* 1.0 mM
  6. *P. furiosus* 1.99 mM
  7. *Th.litoralis* 11.7 mM
4. Thus, kinetic parameters from, for example bovine, mouse and rabbit can be used with some confidence in simulations of the human enzyme.



