

EMBO Practical Course

23.06.2016, Heidelberg

Prof. Rebecca C. Wade
Antonia Stank, Stefan Richter

Full Tutorial PDF can be downloaded her:
wiki.mcm.h-its.org/doku.php?id=courses:public:public

ProSAT⁺ webserver

<http://prosat.h-its.org/>

Analyze protein sequence annotations in context of the three dimensional protein structure.

Visual Molecular Dynamics (VMD)

<http://wiki.mcm.h-its.org/doku.php?id=courses:public:vmd>

Learn different functionalities of the molecular visualization program for displaying, animating, and analyzing large biomolecular systems using 3-D graphics.

webPIPSA

<http://pips-a.hits.org/>

Learn how to use the PIPSA server to compare different interaction fields (electrostatics) for different proteins.

SYCAMORE/ qPIPSA

<http://sycamore.eml.org/>

Estimate kinetic parameters based on protein structural information with qPIPSA for the first step in the glycolysis.

LigDig

<http://mcm.h-its.org/ligdig/>

Explore the new LigDig webserver, a web application for investigating ligand-protein interactions.

MolSurfer

<http://projects.villa-bosch.de/dbase/molsurfer/>

Learn how to use MolSurfer, a graphical tool that links a 2D projection of a macromolecular interface to a 3D view of the macromolecular structures.

ProSAT⁺ webserver Tutorial

Analyze protein sequence annotations in context of the three dimensional protein structure.

ProSAT⁺ allows a quick visualization of a 3D protein structure with highlighted functional regions together with annotations of the functional effects of point mutations. Sequence annotations from UniProt are displayed, but user annotations, provided in a URL to the viewer can also be visualized. Using this approach, one can publish structure sequence mapping annotations using an URL.

In this example you will learn how to use ProSAT⁺ and you will use it to

- display different sequence features on the protein structure
- search for a motif within the protein structure
- export the current visualization state as a picture (e.g. jpg)
- generate a URL that contains user specified annotations for a protein

You will work on hemoglobin in this tutorial, but you can use also a protein that you are currently working on or you are interested in. Most of the following steps are general and can be applied for any protein. In the other following tutorials you will also work with the heat shock protein 90 (HSP90) (webPIPSA tutorial) and Glucokinase/Isomerase (SYCAMORE tutorial). You can also check these proteins using ProSAT⁺ with the PDB ids 1uy8, 3f9m and 1iat respectively.

1. Go to <http://prosat.h-its.org/>. On the start page you can see that you can search for a 3D protein structure by PDB id, UniProt accession number or a protein/DNA sequence.
2. Search for the PDB id 1j7y and click the 'Check PDB' button. After a few seconds you can see the result page with the visualized protein structure on the right and different informations about the protein on the left.
3. Find the sequence scroll bar on the left and scroll through the protein sequence. Move your mouse over one of the colored squares shown above the amino acids and get some information about a known sequence annotation for this specific amino acid. Click on one of the check-boxes and see what happens. In case there is no structural data available for an amino acid the check-box is disabled (e.g. in 1j7y position number 1).
4. Click on the 'Show feature panel' button or on the ProSAT⁺ logo on the upper left of the website. The feature panel on the left will open and different categories of sequence annotations are listed (corresponding to the different colors of the squares in the sequence scroll bar). You can show all annotations for one category with the help of the check-box, or you open the drop down list for a specific category and select only one annotation.
Imagine you wanted to look up a hemoglobin variant called 'M-Boston/M-Osaka' in the alpha subunit. Select chain A in the feature panel and in the 'variant' section you can find the one of interest. The additional information that this variant leads to a decrease in oxygen-binding affinity can be verified through the link to SwissVar. By selecting this annotation, the corresponding histidine at position 59 is visualized and labeled in the 3D structure. The amino acid is near to the bound heme (HEM) ligand of the corresponding subunit. In the category 'binding', you find two annotated residues and by selecting both, you can see that not only does the histidine at position 59 bind the heme iron at the distal position but also the histidine at position 88 is also annotated as relevant for binding and binds the heme iron at the proximal position. This multiple residue annotation can also be seen in the sequence scroll bar.
5. Close the feature panel again and use the motif search module on the lower left part of the page

to search for the 'HVD' motif within the protein structure. You get a list of matches for each chain and if you select the match the sequence scroll bar automatically jumps to that sequence position and the residues are highlighted. You can also search for a single amino acid, for example, tryptophan to see if it occurs in the structure.

6. You can use some predefined visualizations by using the 'Configure Jmol' menu button at the top and select one of the listed visualizations. Imagine you prepared a visualization that shows some labeled sequence annotations and the protein in a meaningful representation and now you want to save this as a picture. For this, you can use the export commands in the 'Configure Jmol' menu, e.g. 'Download current image (jpg)'.
7. Now you want to send the information about the 'M-Boston/M-Osaka' variant and its influence on the decrease in oxygen-binding affinity to a collaboration partner. To avoid any misunderstandings you want to send a ProSAT⁺ URL that automatically visualizes the 3D structure and highlights the relevant annotation.

First, ensure you still have the M-Boston/M-Osaka' variant selected and then open the URL getter via the 'Generate URL' menu at the top. Use the 'Add Selected Features' to automatically put the selected annotation(s) into the table. You can edit the entry in the list by clicking on any value, for example, you could change the annotation name to specify that position 59 is also known as metal ion-binding site. If you finished editing click on the 'Get URL' button and copy the generated link. Now, you can send this link to your collaboration partner and he will see the protein structure with your defined annotations.

Visual Molecular Dynamics (VMD) Tutorial

Learn different functionalities of the molecular visualization program for displaying, animating, and analyzing large biomolecular systems using 3-D graphics.

In this tutorial you will learn how to use VMD to:

- Exploring the RCSB Protein Data Bank (PDB)
- Protein and ligand visualization and selection of structure parts
- Create Graphics representations
- Inspecting Protein Structure: The Ramachandran Plot (via Extensions menu)
- Coloring Chains and Hydrogen patterns
- Alignment of proteins
- Visualization of electrostatic potentials

The tutorial can be started via the following link:

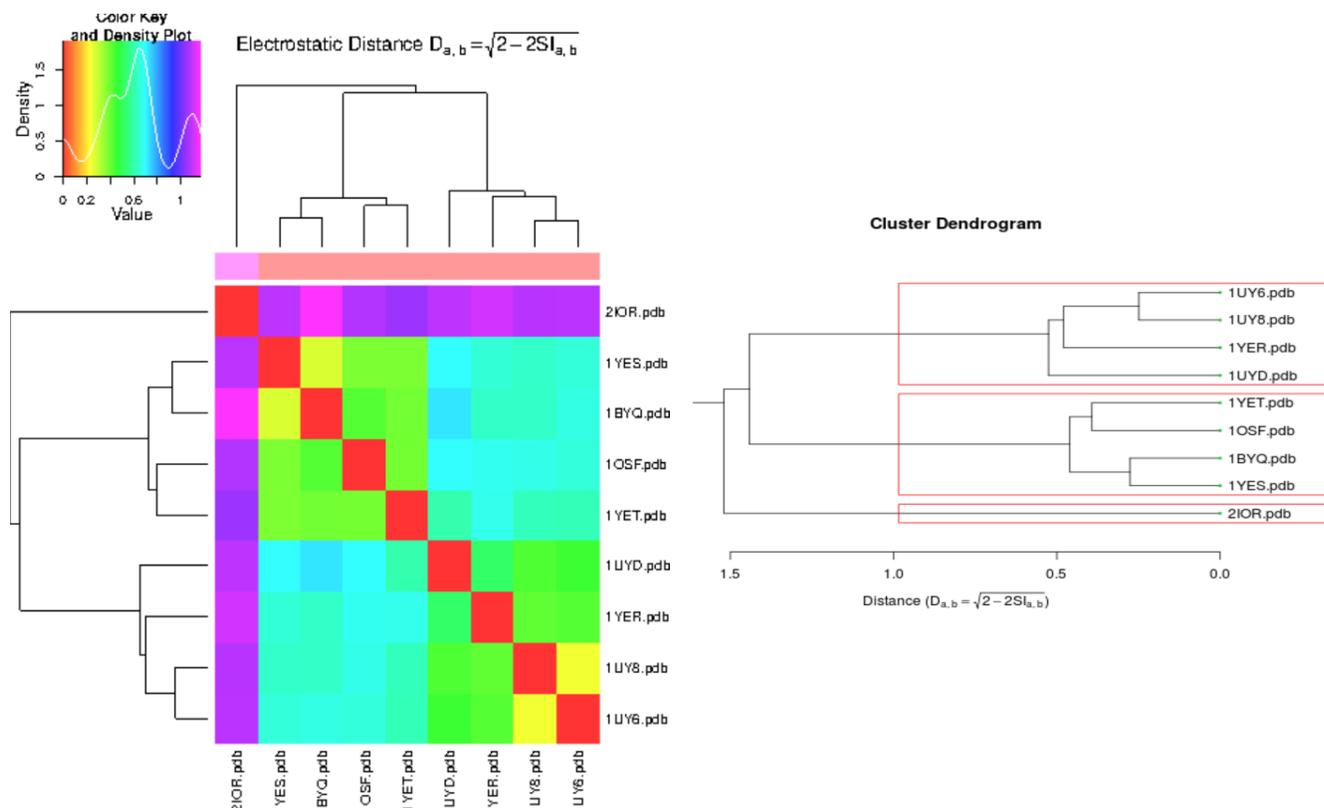
<http://wiki.mcm.h-its.org/doku.php?id=courses:public:vmd>

webPIPSA Tutorial

Using the PIPSA server to compare different interaction fields (electrostatics) for different proteins.

In this example we will focus on the chaperone HSP90 molecule. It stabilizes proteins including those that are required for tumor growth and that is why there is an interest to find HSP90 inhibitors. This protein has a binding site that is surrounded by helices. Some of these helices can be distorted. Here, we will compare the electrostatics surrounding the binding sites of nine different HSP90 structures and check if there are differences.

We have already performed a PIPSA analysis with these nine structures and grouped them into three clusters. You can see the cluster results below:



Question: Do you think the number of clusters is reasonable? What could be the reason for the bigger distance between some HSP90 proteins?

This previous analysis was performed on the whole (global) superimposed HSP90 structures. Now we want to perform a local analysis around the binding site with the distorted helix as center and see how the result changes.

1. Go to the PIPSA webserver <http://pipsa.h-its.org/>
2. Start a PIPSA analysis with PDB structures as input. Select and import the nine local PDB files from http://wiki.mcm.h-its.org/lib/exe/fetch.php?media=courses:public:webpipsa_pdb.zip. These pdb files contain the already superimposed structures from the previous analysis. Give your email and use APBS (this goes faster). You do not need to superimpose the structures

again.

To set the center type a name (e.g. center) and the following coordinates:

x: 16.305 y: 7.275 z: 24.916

As radius you can use 10 and/or 15 Å (you can define a center two times with a different radius)

3. When the analysis is complete (this can take several minutes), note the different clusters of structures. Use the tabs at the top to switch between the different analysis.

Question: Which differences can you observe?

4. Look at the structures using the integrated applet. Focus on the distorted helices. What is your conclusion?

Question: What is the difference between the global and the local comparison runs?

SYCAMORE Tutorial

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.

Modelling the first step of glycolysis

1st 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 Km 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 Km parameter of the human hexokinase IV using available protein structure for that enzyme and available kinetic data for other enzymes.
4. Go to the Sycamore website, <http://sycamore.eml.org/>, and click the button “Enter SYCAMORE”
5. Use the last option by clicking on “parameter estimation”
6. Register at the Parameter Estimation page by clicking on New Registration or enter your user name and password and press Enter

Registration

We request registration because parameter calculations 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 calculations. In any case, you will be notified by email on the result of the calculation(s), therefore an email address is required. Without email registration, parameter calculations will not be performed.

Additionally, the workflow for parameter calculations makes use of the program MODELLER, which requires an End-User Software License Agreement. For academic users, the use of the program is free. You can get a key [here](#). (For now, you can get a key from [Stefan Richter](#) or [Andreas Weidemann](#)).

Furthermore, SYCAMORE offers the possibility to save your models as projects and (in near future) to view and keep track of all calculated parameters.

In order to avoid repeated registrations, we offer the facility to store your email address on the server and login by using your username and password. We appreciate when you also provide your name and institution but this is not mandatory. However, this helps us in the evaluation who is using SYCAMORE for parameter calculations.

REGISTRATION

first name	last name
institution	city
country	MODELLER key
maxMustermann	maxMusterman@home.
user name *	email *
password *	confirm password *
(* required)	

register

7. In this example you are querying for the liver-specific glucokinase (hexokinase IV or D) from *Homo sapiens*.
 1. Insert the UniProt identifier of that amino acid sequence: P35557.

2. Click on the Start search button. You will then see a screen like this:

webPIPSA
Protein Interaction Property Similarity Analysis

Heidelberg Institute for Theoretical Studies
 HITS

Work in progress

This page will be reloaded automatically in a few seconds. Please wait.

The time required for a PIPSA analysis depends on several factors:

- the number of structures you submitted for analysis.
- the number of regions you have selected.
- the number of jobs in the request queue (submitted by other users)

Typically a PIPSA analysis takes from 10 min to 2 hours. A qPIPSA analysis requires for the download phase approx. 5-20 min, whereas the analysis requires 30 to 180 minutes.

If you start a [new search](#), the current search will only be accessible via the link in the email you receive.

Your request is being processed. There is 1 job in the request queue.

Here is an overview of the actions done and the advancement of the information retrieval (latest at top):

- Parsing of Swiss-Prot entry P0A4E1 using <http://www.expasy.org/uniprot/P0A4E1.txt> completed.
- Parsing of Swiss-Prot entry O31392 using <http://www.expasy.org/uniprot/O31392.txt> completed.
- Parsing of Swiss-Prot entry A120M6 using <http://www.expasy.org/uniprot/A120M6.txt> completed.
- Parsing of Swiss-Prot entry P54495 using <http://www.expasy.org/uniprot/P54495.txt> completed.
- Parsing of Swiss-Prot entry Q53Y25 using <http://www.expasy.org/uniprot/Q53Y25.txt> completed.
- Parsing of Swiss-Prot entry A0SZU9 using <http://www.expasy.org/uniprot/A0SZU9.txt> completed.
- Parsing of Swiss-Prot entry Q4E4E1 using <http://www.expasy.org/uniprot/Q4E4E1.txt> completed.
- Parsing of BRENDA entry 2.7.1.2 completed.

8. 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

1. Structural Information
2. Region selection
3. Kinetic data

Structural information

Models

Found 20 protein structures. 20 come from PDB.

#	Seq.: Max Cov. Max Iden. Min RMS with 1GLK [?]	Chains (Seq. Cov./ Seq. Iden./ Align. Score)	Details [?]
#1	98% 0% 100000.0 (see logfile)	<input type="checkbox"/> A P35557 from 1 to 465 (98% / /)	Accession: 1GLK Resolution: NOT APPLICABLE Method: THEORETICAL MODEL EC references: - 2.7.1.1 Heteroatoms: - GLUCOSE [GLC] ▶ Missing Residues [click to expand](7)
#2	97% 0% 100000.0 (see logfile)	<input type="checkbox"/> A P35557 from 12 to 465 (97% / /)	Accession: 3QIC Method: X-RAY DIFFRACTION R-factor: 0.213 Organism: HOMO SAPIENS Crystallization conditions: 25% (w/v) PEG3350, 0.1M HEPES PH 7.3, 0.2M AMMONIUM SULFATE, VAPOR DIFFUSION, SITTING DROP, TEMPERATURE 293K EC references: - 2.7.1.2 Heteroatoms: - chain A: ALPHA-D-GLUCOSE [GLC] - chain A: PHOSPHOSERINE [SEP] - chain A: GLYCEROL [GOL] ▶ Missing Residues [click to expand](17)
	96% 0% 100000.0 (see logfile)	<input type="checkbox"/> A P35557 from 13 to 466 (96% / /)	Accession: 1V4S Method: X-RAY DIFFRACTION R-factor: 0.232 Organism: HOMO SAPIENS Crystallization conditions: PEG1500, HEPES, PH 8.8, VAPOR DIFFUSION, HANGING

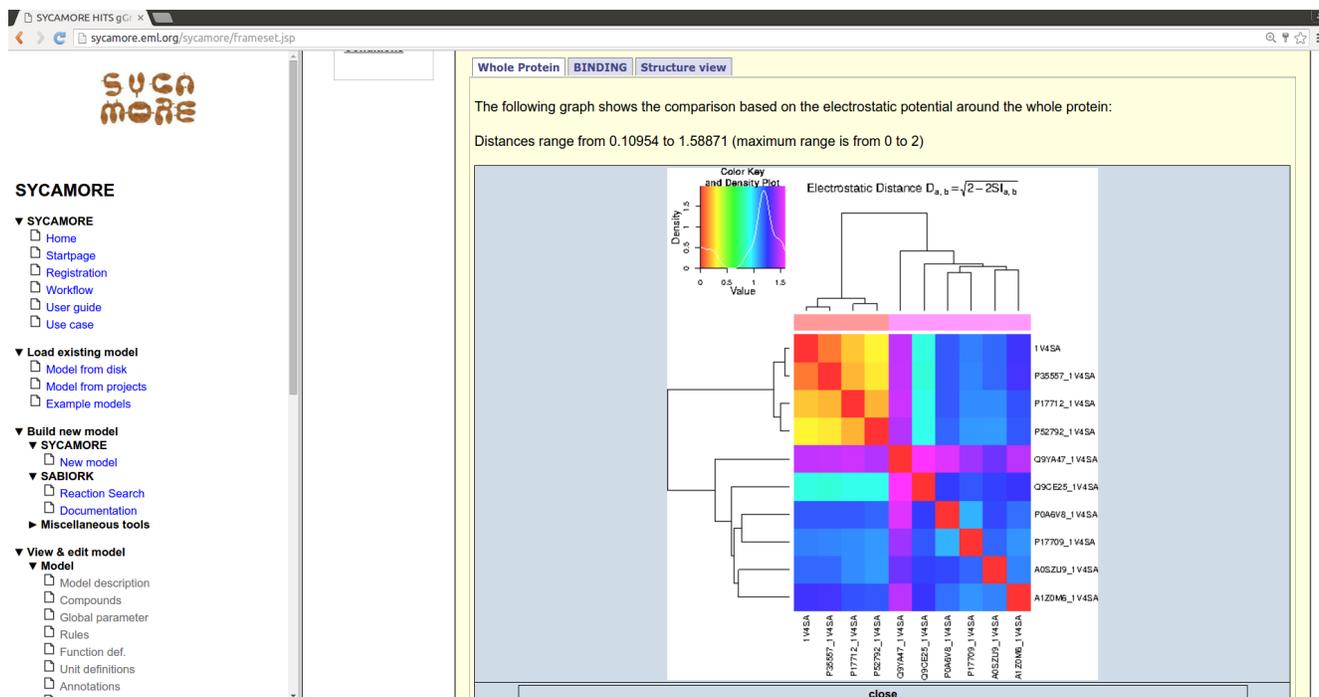
9. SYCAMORE found 20 protein structures with relevance to the human hexokinase IV. All come from the PDB, none from ModBase or SwissModel.
10. Choose 1V4S as a PDB template structure for subsequent structural modelling based on sequence identity with the UniProt query and completeness of X-ray structure.
11. Open the Region Selection submenu. Here, annotations from the UniProt entry P35557 are reported. Select Region BINDING Substrate in order to also perform a comparison restricted to that region.
12. 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 all entries from BRENDA and leave all entries from SABIO-RK “on”.
13. 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.
14. The calculation starts. The web site will refresh automatically and you will be notified by email when the calculation is finished. (this can take several minutes)

The screenshot shows the SYCAMORE webPIPASA interface. The browser address bar indicates the URL is `sycamore.em.lorg/sycamore/frameSet.jsp`. The page features the SYCAMORE logo on the left and the Heidelberg Institute for Theoretical Studies logo on the right. The main content area is titled "webPIPASA Protein Interaction Property Similarity Analysis" and displays a "Work in progress" status with a yellow background. A message states: "This page will be reloaded automatically in a few seconds. Please wait." Below this, it explains that the time required for a PIPASA analysis depends on several factors: the number of structures submitted, the number of regions selected, and the number of jobs in the request queue. It notes that a typical PIPASA analysis takes 10 minutes to 2 hours, while a qPIPASA analysis requires a 5-20 minute download phase and 30 to 180 minutes of analysis. A message indicates: "Your request is being processed. There is 1 job in the request queue." A list of completed modellings is provided:

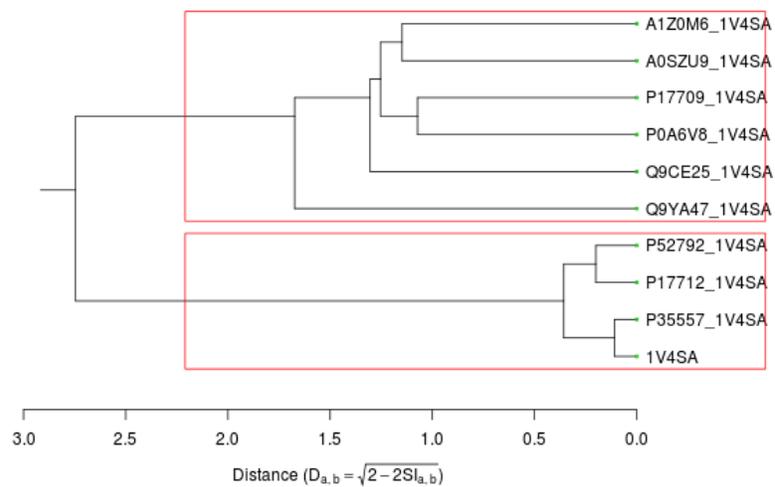
- The modelling of P52792 on 1V4SA is complete.
- The modelling of A0SZU9 on 1V4SA is complete.
- The modelling of P17712 on 1V4SA is complete.
- The modelling of P17709 on 1V4SA is complete.
- The modelling of P0A6V8 on 1V4SA is complete.
- The modelling of A1Z0M6 on 1V4SA is complete.
- The modelling of P35557 on 1V4SA is complete.
- Started modelling of all chosen sequences on the pdb structures (modeller, [2]).

15. 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.

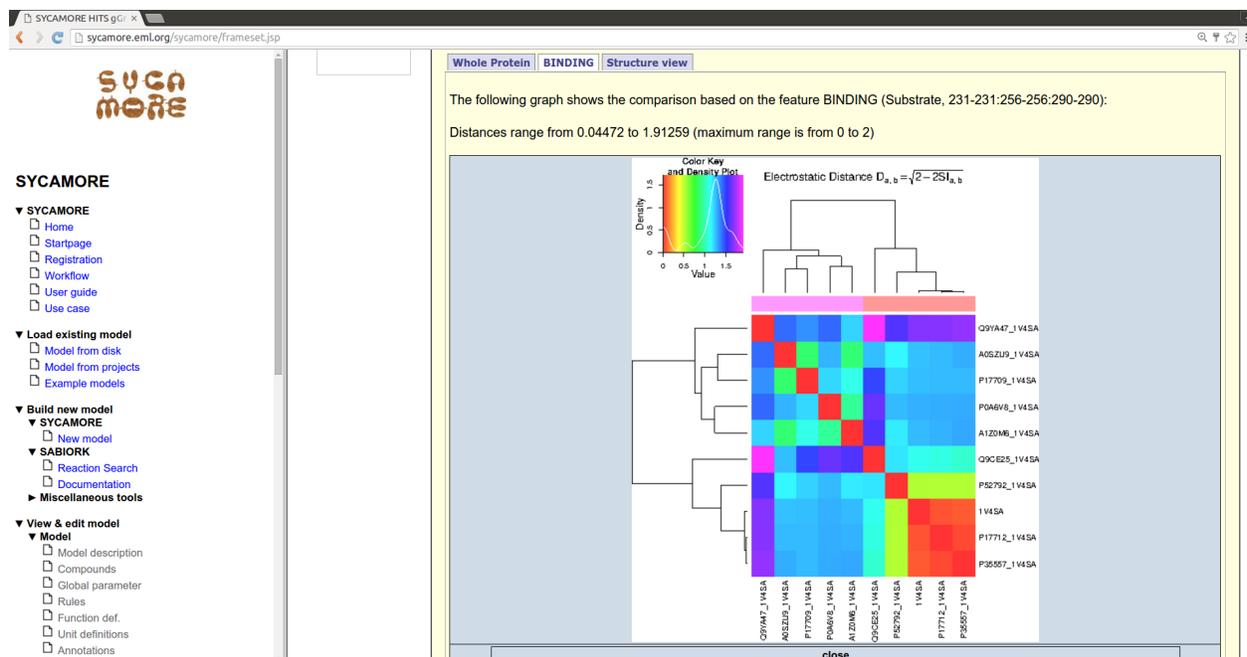
1. The first dendrogram is for the electrostatic potential of the entire protein.



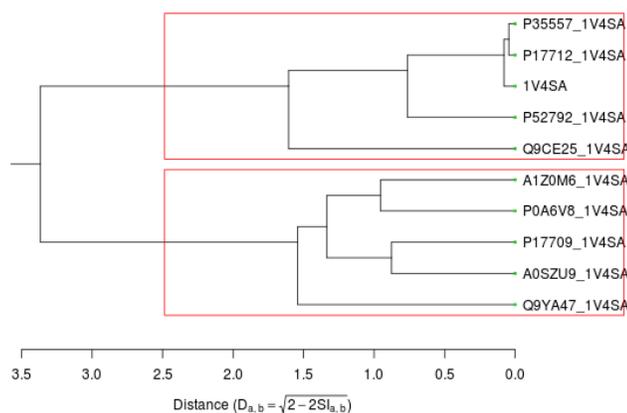
Cluster Dendrogram



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



Cluster Dendrogram



The electrostatic is conserved for enzymes from human (P35557), rat (P17712), and mouse (P52792). The enzymes from the other organisms exhibit electrostatic potentials distant from the others. This puts your query sequence P35557 very close to the hexokinase from rat (P17712), and remote from, for example, yeast (P17709). Thus, kinetic parameters from rat could be used if you were to model the hexokinase IV from *Homo sapiens*.

This also agrees with experimentally available K_m values D-glucose from rat and from human. Those are much more similar between rat and human than to yeast, which is drastically lower.

The modelling of the second step of glycolysis Glucose-6-phosphate isomerase (G6PI) catalyzes the reaction glucose-6-phosphate \rightarrow fructose-6-phosphate. This would be a similar procedure as before by using the UniProt identifier for the enzyme glucose-6-phosphate isomerase (P06744).

The actual simulation of the final model can be done according to the example explained here:

<http://sycamore.eml.org/sycamore/documentation/usecase.jsp>

LigDig Tutorial

The LigDig webservice provides seven different modules. Here, you will get an impression how to use LigDig and for which purposes it might help you.

1. Go to the LigDig website: <http://mcm.h-its.org/ligdig/> and click on 'Example Cases' in the right panel.
2. Select an example case which is most interesting for you:
 - How and where can information be found for the compound fructose 1,6-bisphosphate?
 - Which biological macromolecular structures in complex with fructose 1,6-bisphosphate are available?
 - Which structural and functional families are these complexes in?
 - In which functional role does the compound bind?
 - Which proteins does fructose 1,6-bisphosphate bind to?
 - What other proteins do these compounds bind to?
 - How and where can information be found for the compound fructose 1,6-bisphosphate and a set of other compounds?
 - How can I prepare the structure of a protein-ligand complex for further analysis or calculations (by protonating it and adding missing atoms)?
 - How can I compare the similarity of ligand binding sites (ligand based)?
 - How can I compare the similarity of ligand binding sites (binding site residue based)?
3. Go through the example and test the respective module(s) in LigDig.

MolSurfer Tutorial

Learn how to use MolSurfer, a graphical tool that links a 2D projection of a macromolecular interface to a 3D view of the macromolecular structures.

The Tutorial can be started via the following link:

<http://projects.villa-bosch.de/dbase/mol-surfer/tutor/instructions.html>