

# @TOME v2.2 - Documentation

version : 09/11/2015

Authors and References

**@TOME server is available here:** <http://atome.cbs.cnrs.fr>

**Authors:** Jean-Luc Pons & Gilles Labesse

Centre de Biochimie Structurale / CBS - CNRS UMR5048 - INSERM U1024 – Montpellier  
University - 29 rue de Navacelles 34090 MONTPELLIER France

**Reference:**

@TOME-2: a new pipeline for comparative modeling of protein-ligand complexes. Pons & Labesse  
- **Nucleic Acids Research**, Web Server Issue- **2009** - doi: 10.1093/nar/gkp368 [Link](#)

**Contact Email :** [atome@cbs.cnrs.fr](mailto:atome@cbs.cnrs.fr)

## Table des matières

@TOME explained in a few words .....	4
@TOME V2 flowchart: .....	5
Stage 1 : How to submit a query to @TOME ?.....	6
Stage 2 : Structural Alignment (S.A.) research.....	7
Stage 3 : Structural Alignment (S.A.) analyses & modeling.....	8
Stage 4 : @TOME Rank calculation.....	9
Stage 5 : Ligand research & Contact profile.....	10
GUI : Database storage and graphic user interface.....	11
GUI : Running Processing Modules.....	13
Module : Perform Alignment.....	13
Module : change 3D cluster.....	13
Module : Change @tome Selection.....	13
Module : Modeling Selected Alignment (Modeller license is mandatory).....	13
Module : Multi Template Model (Modeller license is mandatory).....	13
Module : Automatic Modeling (Modeller license mandatory).....	14
Module : Show or Recalculate Profile of Contacts.....	14
Module : Ligand Selection Tool.....	14
Module : Transfer Selected Ligand(s) to Selected Model(s).....	14
Module : Automatic Complexes Modeling.....	14
Module : Binding Site Prediction.....	15
Module : Select templates for the screening & Module : Run Virtual Screening.....	15
Module : Export Alignment with Pat.....	15
Module : Hide or Remove Selected Entries.....	15
Module : Structural Superimposition.....	15
Module : Add User PDB file as a New Entry.....	15
Module : Search Ligands from Homologous Structures.....	15
Stage 6 : Complex Modeling.....	16
Stage 7 : Comparative Virtual Screening.....	17
Technical overview.....	20
References of embedded tools.....	21

## @TOME explained in a few words ...

@TOME v.2.2 is the new version of the meta-server **@tome v1**. This web server is a pipeline dedicated to model 3D structures of receptor/ligand complexes based on homologous protein-ligand complexes of the protein data-bank (PDB). Each step of the process includes several tools to obtain an accurate consensus score. The main steps are summarized here:

### *Looking for homologous structural alignments in PDB data-bank ...*

From the amino acid sequence submitted by the user, @TOME V2.2 searches homologous sequences using dedicated softwares (**HHSearch, Fugue, Psi Blast, Sp3**) and predict molecular structures (**Tito, Scwrl**). @TOME allows finding the best sequence alignment and the best 3D experimental structures to use as template. Results are evaluated with an important set of 1D, 2D & 3D evaluation tools such as **Tito, TCoffee, Verify3D, 3D-Jury** and **3D clustering** methods.

### *3D structure modeling ...*

For the comparative modeling study, three types of model are calculated by three different softwares:

- 1/ The **Tito** software (Tool for Incremental Threading Optimization) keeps conserved amino acid residues and proposes a common core closed to the experimental template (same experimental coordinates).
- 2/ The **Scwrl** software calculates missing side chains into the Tito common core.
- 3/ Finally, the **Modeller** software offers a recalculated and sophisticated complete model.

All models are evaluated and ranked by many tools such as **Qmean, Dope, Errat, Dfire, ...**

### *Protein / ligand complexes modeling ...*

@TOME offers a module to model receptor/ligand complexes by homology. @TOME detects all the ligands stored into the template files (PDB) and proposes to move them into the model. The transfer of ligands is allowed by the superimposition of the experimental structure and the model (**comparative docking method**). Each modeled complex is evaluated with several scoring functions such as **MedusaScore, Xscore or Plants,...** A profile of receptor/ligand contacts is calculated from the best crystallographic templates. Each new complexes can be evaluated by calculating a similarity to this profile (**Pat** software).

### *Virtual screening ....*

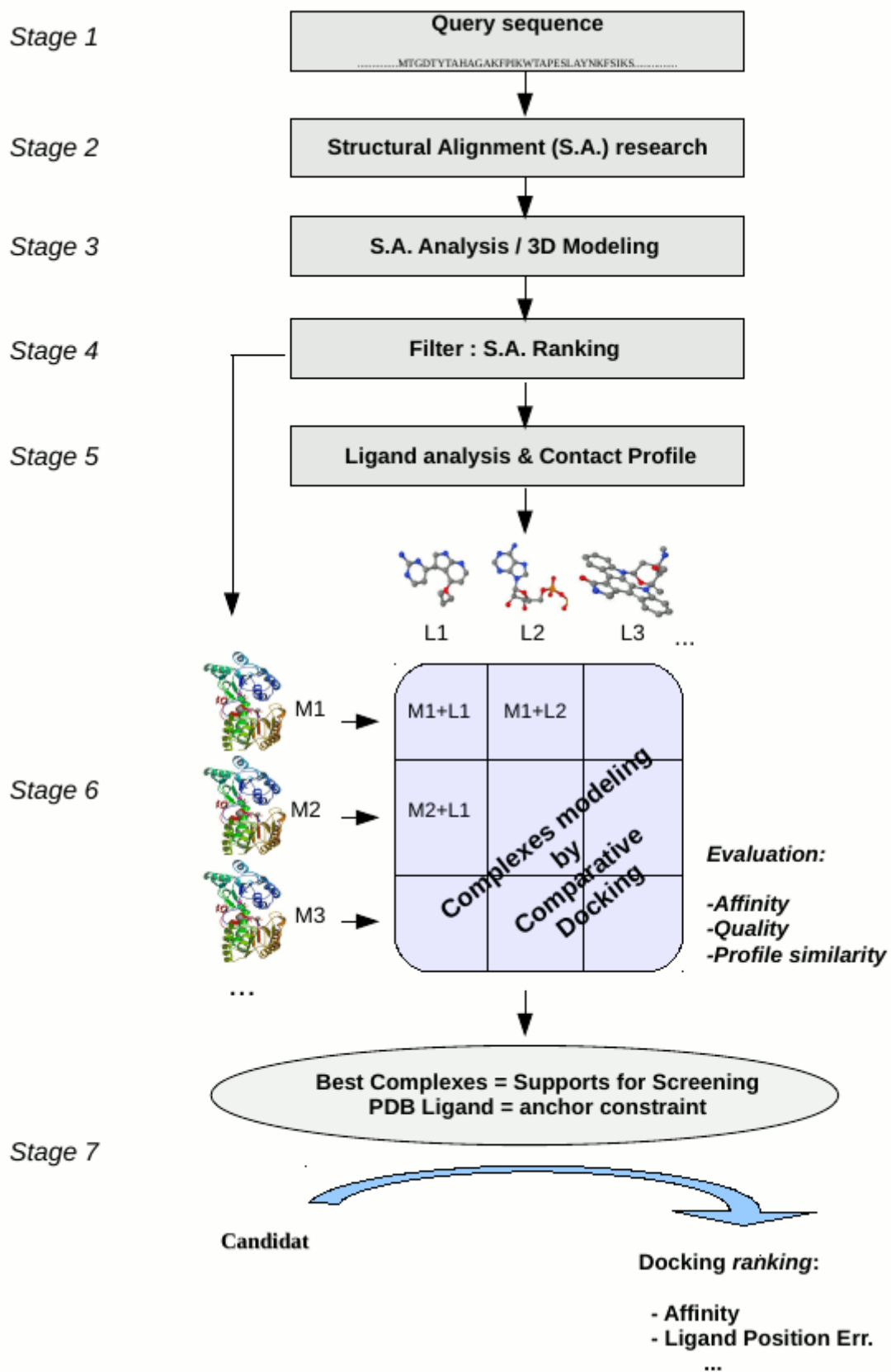
@TOME proposes a virtual screening module using the modeled receptor (or crystallographic structure if available). Each previously modeled complex (in step "Protein/ligand complexes modeling) is evaluated and the best complexes can be used to guide the positioning of new drugs. Affinities (dissociation constants) are calculated using scoring functions. A novel score (Ligand Position Error / **SVM method**) permits to evaluate similarity between docking results and many crystallographic features such as profile of receptor/ligand interactions, orientation of the ligand or similarity with PDB ligands. This score help us to select the best results.

### *And ...*

Some other modules such as "Molecular replacement" or "NMR" are available.

# @TOME V2 flowchart:

This flowchart shows a complete study by @TOME. All stages are described below in the text.



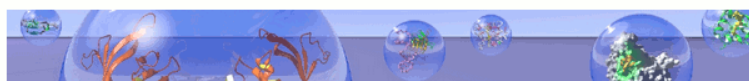
# Stage 1 : How to submit a query to @TOME ?

The home page of @TOME V2 is here : <http://atome.cbs.cnrs.fr>

This home page allows the user submitting the primary sequence, title and email address (screen shot below). Some tools are not selected by default for the process but it is possible to change options [The user can change the parameters and choose tools not selected by default]. Some modules such as "NMR", "molecular replacement" or "screening" can be selected. Options are described below in the text.



**@TOME**  
NEW V. 2.2



@TOME: a pipeline for comparative modeling of protein-ligand complexes.

This meta-server, named @TOME 2, allows one to submit an amino acid sequence to embedded softwares dedicated to homologous sequences search, structural predictions and fold recognition. @TOME V2 facilitates the recognition of the better 3D template and the best automatic alignment according an important set of evaluation tools results.

@TOME 2 permits also to create 3D model of receptor/ligand complexes by comparative docking of ligand between the experimental protein and model.

Best models of complexes can be used to serve to guide the positioning of new drugs during a screening steps

***Warning: Access to @TOME2 server is restricted to academic use only !***

*This website is free and open to all academic users and there is no login requirement.*

***This website is optimized for Firefox.***

[Documentation](#)

[Embedded tools](#)

[HHsearch](#)  
[Fugue, PS3](#)  
[Pisblast](#)

[Pspired, P-Sea](#)

[Tito, Scwrl](#)  
[Modeller](#)

[QMean](#)  
[Verify3D](#)  
[Eval23D](#)

[T-Coffee](#)  
[Clustalw](#)  
[Muscle](#)  
[Jalview](#)

[MaxCluster](#)  
[Profit, Matt](#)

[Plants](#)  
[MedusaScore](#)  
[XScore](#)  
[DSX](#)

[Pat](#)  
[JMol](#)  
[Babel](#)  
[Fconv](#)  
...

**Title :**

**Primary Sequence:** [One letter code] eg:  
MPSHRNSNLKFCTVCASNNRSMESHKVLQEAGYNVSSYGTGSAVRLPGLSIDKPNVYSFGT  
PYNDIYNDLLSQSADRYKSNGLLQMLDRNRRLKKAPEKWQESTKVFDFVFTCEERCDFDAVCE  
DLMNRGGKLNKIVHVINVDIKDDDENAKIGSKAILEADMLNDKIEQCEKDDIPFEDCIMDI  
LTEWQSSHSQLPSLYAPSY

**Email :**

---

Search for Homologous Sequences & Comparative Modeling

Ligands Selection & Complexes Prediction by Comparative Docking

Modules

---

Ref: Pons & Labesse - Nucleic Acids Research, Web Server Issue- 2009

## Stage 2 : Structural Alignment (S.A.) research.

@TOME submits the query primary sequence to structural alignment softwares previously selected. The following tools are embedded in @TOME:

Tools	DBase	Selected	Output
Psiblast	PDB	default	20 best results
Psiblast	CBE	optional	user choice
HHSearch	PDB	default	20 best results
Fugue	Homstrad	default	10 best results
SP3	SP3 DB	optional	10 best results

Each tool uses a different algorithm (see below). The multiplicity of tools allows the user retrieving more analyses and thus having more confidence in the common results of the different algorithms. Homologous sequences are detected into the **PDB**. The **CBE** database (Contact Between Entities) which contains complexes from the PDB, permits to easily find all the ligands of the homologous structures. Many features have been previously calculated in CBE in order to increase the speed of @TOME process. CBE is available at <http://atome.cbs.cnrs.fr/AT2B/cbe.htm>.

Embedded tools, algorithms and options are explained here:

- **Psi-Blast** on PDB : a sequence-profile comparison software  
Default parameters : 8 iterations, E-Value threshold = 1E-3.
- **Psi-Blast** on CBE : detects additional ligands not found by Psi-Blast on PDB  
Default parameters : 8 iterations, E-Value threshold = 1E-3.  
Low limit of identity between query and template : selected by user  
Maximum number of additional S.A. found : selected by user
- **HHsearch** : a profile-profile comparison software  
Parameters for Psiblast: 8 iterations, E-Value threshold=1E-3,  
Database = PDB, PsiPred is running.
- **Fugue** : a sequence-structure comparison software / default options of Fugue.
- **SP3** (Optional) : a profile-profile + structure comparison software / default options of SP3.

## Stage 3 : Structural Alignment (S.A.) analyses & modeling

All S.A. found are evaluated with a set of 1D, 2D and 3D tools.

The tab below describes the process for each S.A.:

Sequential processing for each S.A.	Tools
Sequence identity	@TOME
Extraction of related PDB files and chains	@TOME
S.A. information : <i>cath, scop, uniprot, smart, prosite, Pfam, EC-Number csa , procognate, experimental method, ....</i>	@TOME
Secondary structure prediction	Psea
Alignment score calculation	Tcoffee
Common core modeling (Threading)	Tito
Side chain calculation	Scwrl
3D evaluation	Verify3D, Qmean
Detection of disulfide bonds	Pat
Ligand detection (< 8Å to protein)	@TOME
Ligand information <i>(small molecule / peptide / Nucleic Acid / Sugar / Ion) Name, Weight, contact distance, binding site conservation</i>	@TOME

Process on all the set of S.A.	Tools
3D clustering threshold = 0.8Å / sequence independent Hierarchical Clustering algorithm cluster name = C1, C2,...	Maxcluster / MaxSub 3DJury
Research common features from templates (extract restraints for Modeller)	Pat / FCT
Profile of protein-ligand contacts	Pat

All scores are used in the process of S.A. selection (consensus method).

All these indicators can be displayed on the graphic user interface of @TOME in order to facilitate the selection of the best alignments. In the default configuration, only the twenty best sequences alignments are displayed on the graphic user interface.

## Stage 4 : @TOME Rank calculation

For each S.A. an additional **@TOME rank** score is calculated including all result scores previously defined. Each value is normalized within a range from 0 to 100 (0 = lowest value, 100 = highest value).

**@tome rank is an average of the following tool results:**

- Identity (%) (Percentage of identity \*(length of template sequence / length of query sequence))
- 3D-Coffee sequence alignment score \*(length of template sequence / length of query sequence)
- 3DJury rank (coeff 1/2)
- Tito Score divided by the number of residues in the sequence
- Verify3D & QMean evaluation scores of Scwrl models (Tito Common core + calculated side chains).

**@tome score is calculated as:**

$$Score(i) = \frac{\sum_i \left( \left( \frac{Vi - Vmin}{Vmax - Vmin} \cdot 100 \right) \cdot Nt - 1 \right)}{\sum_i Nt}$$

where:

Vt i = Score value of tool (t) for S.A. i

Nt = Total number of tools

@tome rank score is displayed in the user graphic interface for each entry.



## Stage 5 : Ligand research & Contact profile

**Ligand research:** @TOME permits to calculate and evaluate protein/ligand complexes. To this end, it looks for all the ligands in the homologous templates of the PDB. The user can select several types of ligands: small molecules, peptides, sugars, nucleic acids or ions according to the study. All the nearest ligands (at least one atom less than 4 Angstroms from the protein) are taken into consideration and classified according to the distance contacts (at 4, 6, 8 Angstroms).

**Binding site detection:** All ligands found in homologous templates are clustered by superimposition of crystallographic complexes. This step permits to detect and rank all binding sites (S1, S2). The main binding site is called S1 (the largest number of ligands), the second S2 and so on.

**Co-Procognate process:** Procognate ligands are not systematically detected in the cognate DBase. Tanimoto similarity score is used to extend this concept to all ligands. Ligand with  $\text{tanimoto} > 0.5$  have same biological state.

**Receptor/ligand contact profile:** the 20 best homologous crystallographic complexes are taken into consideration to calculate the profile of protein/ligand interactions. Complexes are superimposed in order to extract features such as contact distances, angles ( $\phi$ ,  $\psi$ ,  $\omega$ ,  $\chi_1$ ,  $\chi_2$ ), atom types, ... All this calculation is made by the **Pat** software. At the end, a profile of receptor/ligand interactions for the family of receptors is calculated.

Similarity to this profile will be used as a quality criteria for model evaluation (Profile Similarity = PSim).

The 20 best crystallographic complexes are selected according to following features:

- All types of ligands (without redundancy) except Ions
- Receptors are selected in the main 3D common core cluster
- Only the main binding site (S1) is selected and binding site conservation has to be  $> 50\%$

# GUI : Database storage and graphic user interface

Stages 1 to 5 are made in background on the @TOME grid computing. All the previously defined information are stored in a local database (in XML format). After calculation, @TOME provides access to all the information on a graphical user interface.

The user interface consists of two parts. An interactive main form allows the user to view all the database information of @TOME and permits him to select S.A., models or ligands. After selection, the user can use some modules form (several options) to perform many type of process analyses. The results of these calculations are included in the database and dynamically appear in the main form.

## Interactive main form:

Atome Color : MutatedHelixStrandConserved

Query sequence : >PF00001\_5HT1B : (2014-07-29)  
MEEQGIQCAPPFPATSQTGVPLANLSHNCADDYIQDSIALPWKVLVALLALITLATTLSNAFVIATYRTRKLTHTPNYLIASLAVTDLLVSLVMPISMIVTGTGRWTLGQVDFWLLSSDITCASIMHLCVIALDRYWAITDAVDYSAKRTPKRA

Atome Classification :

(20 SA)	Score	Align	3DCore	Cluster	PDB	Complex	Info
31 HSearch	97.44	80%	-135.81	-C1-	4IAR	2	5HT1B_HUMAN (first) Aq[9-398]
1 PsiBlast_PDB	96.29	73%	-135.57	-C1-	4IAR	2	5HT1B_HUMAN (first) Aq[9-398]
36 HSearch	77.67	33%	-118.74	-C1-	3UON	1	ACM3_HUMAN (first) Aq[23-457]
10 PsiBlast_PDB	76.06	34%	-157.26	-C1-	2Y02	12	ADRB1_MELGA Aq[10-300]
4 PsiBlast_PDB	75.06	34%	-142.81	-C1-	2YCW	6	ADRB1_MELGA Aq[10-299]
15 PsiBlast_PDB	74.94	34%	-154.56	-C1-	3ZPR	9	ADRB1_MELGA Aq[10-299]
14 PsiBlast_PDB	74.84	34%	-155.38	-C1-	3ZPQ	10	ADRB1_MELGA Aq[10-300]
13 PsiBlast_PDB	74.77	34%	-152.92	-C1-	4AMJ	11	ADRB1_MELGA Aq[10-300]
11 HSearch	74.54	34%	-156.15	-C1-	2Y03	8	ADRB1_MELGA Aq[10-300]
9 PsiBlast_PDB	74.52	34%	-156.43	-C1-	2Y01	12	ADRB1_MELGA Aq[10-300]
21 Fugue	74.31	30%	-122.30	-C1-	2RH1	10	ADRB2_HUMAN (first) Aq[36-477]
8 PsiBlast_PDB	74.13	34%	-154.96	-C1-	2Y00	12	ADRB1_MELGA Aq[10-300]
12 PsiBlast_PDB	74.11	34%	-156.43	-C1-	2Y04	12	ADRB1_MELGA Aq[10-300]
7 PsiBlast_PDB	73.96	34%	-143.62	-C1-	2YCY	6	ADRB1_MELGA Aq[10-299]
34 HSearch	73.76	32%	-119.43	-C1-	2RH1	10	ADRB2_HUMAN (first) Aq[36-477]
39 HSearch	73.29	26%	-121.21	-C1-	4DKL	5	LYS_BPT4 (first) Aq[14-453]
32 HSearch	73.22	39%	-150.37	-C1-	4AMJ	11	ADRB1_MELGA Aq[7-300]
37 HSearch	72.57	25%	-119.12	-C1-	4E1Y	25	AA2AR_HUMAN (first) Aq[29-428]
35 HSearch	72.03	33%	-119.19	-C1-	3PBL	2	DRD3_HUMAN (first) Aq[41-472]

Server load: [Progress Bar] (9 %)

After selection chose an action: [Select Module] [Continue] [Result Summary]

**Select information in interactive main form:** In the left frame, two types of selection menu are available. The first one, called "Global output mode" permits to display S.A. entries according to several criteria (sort by tools, by cluster and so on). The second group of menu called "Column output" permits to select which information can be viewed in the different columns.

**Sequence color type :** the user can choose to color according to : Secondary Structure, Toffee Identity Score, 3D Clustering, Ligands Contact, hydrophobic amino-acid, disulfides bonds.

**Column Output:** For each S.A. entry, several values are available. The following table defines all the menu entries and associated values:

<b>Score</b> @tome Rank: Tool score:	calculated @TOME rank Raw score of SA search tools : Psiblast, Fugure, HHsearch or SP3
<b>Alignment</b> Identity : TCoffee Score:	Sequence identity (%) Sequence alignment score evaluated with T-Coffee Software.
<b>Structural Core</b> Tito Score : Qmean : Verify3D:	Tito sequence-structure compatibility evaluation score Qmean score of Tito 3D model Verify3D Score of Tito 3D model
<b>3D Clustering</b> Cluster Num: 3D-Jury Struct. Score:	Cluster number 3D-Jury (MaxSub) Score :most frequent sub-structure found within a model set.
<b>Model</b> Model Rank : QMean Score : Dfire Score : Dope Score : Errat Score : Modeller Score : Verify3D Score :	Rank of model (consensus of all scores below) Qmean score of modeller model DFire score of modeller model Dope score of modeller model Errat score of modeller model Raw Modeller calculated energy of the model Verify3D Evaluation Score of the modeller model
<b>Infoclass</b> Uniprot: Cath class: Scop class: Smart class: Prosite class: Pfam class: EC Number: Experiment Method: Taxonomy:	Uniprot classification and link Cath classification and link Scop classification and link Smart classification and link Prosite classification and link Pfam classification and link EC Number and link NMR or XRay structure (NMR constraints and structure factor are available here) Distribution of Pfam domain in classification HEBAV (Human-Eukaryote-Bacteria-Archaea-Viruses)
<b>Complexes Modeling</b> Show Ligands : Complex Affinity : Binding Site identity :	Show and select ligand close to the current main chain Show the best affinity for all calculated complexes. Identity rate of main binding site (S1)
<b>Sequence</b> PDB Template : Complex Tab :	Display template primary sequence Show all the results of a calculated complex

For each column, "Hide column" entry permits to hide relative column in the main form. The header of each column is a yellow button that allows the user sorting the column.

## GUI : Running Processing Modules

@TOME permits to run different process on the S.A. entries. The principle is to select one or more S.A. (or ligands) and to run a module processing. All the modules are accessible with the menu in the bottom of GUI form. Each module is explained here:

### Module : Perform Alignment

The "Perform Alignment" module permits to improve the alignment of selected S.A. It proposes a manual re-alignment (**Jalview**, **Vito**) or to use automatic softwares such as **T-Coffee**, **ClustalW** or **Muscle**. Structural alignment is also possible with **Matt** software and **Vito**.

Each new alignment can be selected and included into the general form in the group called "Multi Alignment Result". Thus, it may be used in @TOME as others S.A.

### Module : change 3D cluster

@TOME calculates structural clusters automatically on Tito common core models (see before). However, it is possible to recalculate clusters optimizing some options such as the choice of model or the threshold.

The clustering module of @TOME uses the software **MaxCluster**. Each cluster is sorted by color. The centroid model of cluster appears in bold characters.

### Module : Change @tome Selection

@TOME selects automatically the 20 best S.A. according @TOME rank. You can change this selection in order to hide redundancy or extract some bad selected SA.

### Module : Modeling Selected Alignment (*Modeller license is mandatory*)

@TOME proposes to calculate more sophisticated models using the **Modeller V9** software. The evaluation of Modeller models is done through the internal Modeller energy score and the **Verify3d**, **Qmean**, **Errat**, **Dope** and **Dfire** softwares. A consensus score permits to rank models. New models are available and can be displayed in the "Model" column. This permits to the user to easily select models with better results.

*Optional* : FCT = Frequent Contact Tree (**Pat** software)

Add restraints to modeling. Restraints are extracted from the best crystallographic templates. Three types of restraints are take into consideration: distance, H bond and disulfide bond.

### Module : Multi Template Model (*Modeller license is mandatory*)

**Modeller** can also use multiple S.A. to create better models. This feature was embedded in @TOME. In this case, only one model will be calculated from all the selected S.A. The new multi-template model will be added as a new entry in the @TOME database. The new entries will be displayed in the main form into the "Multi Template Modeling" group. It may be used like any other S.A..

The evaluation of Modeller models is identical than "Modeling Selected Alignment" module. The code of the Modeller model is a 4 letter code which starts by "M" letter.

*Optional* : FCT = Frequent Contact Tree (**Pat** software)

Add restraints to modeling. Restraints are extracted from the best crystallographic templates. Three types of restraints are take into consideration: distance, H bond and disulfide bond.

### **Module : Automatic Modeling** (*Modeller license mandatory*)

*This module can be selected in the @tome home page (in query submission form).*

Here, @TOME uses **Modeller** software in order to make mono-template models from the 20 best S.A. selected automatically. In a second step, 4 multi-template models were computed by Modeller. For each model to construct, 4 templates have been selected according to the best scores from @TOME, **Scwrl**, **Tito** and **Qmean**.

At the end, mono and multi template models are ranked according to a consensus score calculated using several tools : **Qmean**, **Dope**, **Dfire**, **Errat** and Modeller score.

### **Module : Show or Recalculate Profile of Contacts**

This module shows the current profile of ligand/receptor interactions extracted from homologous complexes found in the PDB database. The profile is a set of constraints calculated by the PDBGEO & CT4D tools from the **PAT** software (show above). The restraints correspond to the most frequent atomic contacts (receptor/ligand) observed in the superimposed structures. It is possible here to recalculate profile with other template/S.A. & ligands selected by the user.

### **Module : Ligand Selection Tool**

This module allows user selecting ligand(s) from local database. Many filters are available: type of ligand, weight, receptor, binding site....

### **Module : Transfer Selected Ligand(s) to Selected Model(s)**

This **comparative docking** module allows modeling receptor/ligand complexes. It is explained in the "Complex Modeling" documentation below.

### **Module : Automatic Complexes Modeling**

@tome can select automatically the better S.A. and ligands for modeling complexes by comparative docking. Two run modes are available: direct or crossed comparative docking.

Direct comparative docking :

Insert ligand of template in the corresponding model

Process steps :

- 1/ Automatic selection of ligands & models (main binding site & main 3D cluster)
- 2/ Transfer of ligands from template to corresponding model by comparative docking
- 3/ Recalculation of non-conserved side chains after ligand transfer (binding site)
- 4/ Evaluation of complexes (theoretical pKa)
- 5/ Binding site prediction & return email

Crossed comparative docking :

Insert all selected ligand in all selected models.

Process steps:

- 1/ Automatic selection of the 5 best Scwrl models (from the main structural cluster)
- 2/ Automatic selection of ligands (main binding site & main structural cluster)
- 3/ Transfer of all selected ligands from template to all selected models by comparative docking
- 4/ Recalculation of non-conserved side chains after ligand transfer (binding site)
- 5/ Evaluation of complexes (theoretical pKa)
- 6/ Binding site prediction & return email

### **Module : Binding Site Prediction**

Here, the user has to select the best complexes generated by comparative docking in order to detect a consensus binding site.

### **Module : Select templates for the screening & Module : Run Virtual Screening**

These two modules are explained in the "virtual screening" documentation below.

### **Module : Export Alignment with Pat**

This module allows you saving (downloaded file) the alignment of selected sequences. The following formats are available: Fasta, Pir, Selex and MSF. The format conversion is made by the **Pat** software.

### **Module : Hide or Remove Selected Entries**

In order to eliminate irrelevant alignments and reduce the database, this module can hide (from general form) or remove (from database) the selected S.A..

### **Module : Structural Superimposition**

This module allows the user choosing the structures to be superimposed (Tito, Tito/Scwrl, Modeller or PDB template). It then proposes to extract result according to different formats (NMR type bundle, single PDB file with one chain by model or separate PDB file).

In this step, @TOME uses the **Matt** software (long time consuming) or **MaxSub/Maxcluster** software.

### **Module : Add User PDB file as a New Entry**

This module allows the user inserting a new structure in the database (local pdb). This structure may be used as a target for the transfer of ligand or any other processing modules.

### **Module : Search Ligands from Homologous Structures**

Process steps :

- 1/ @TOME selects 10 best alignments in order to make a profile for **Psiblast** research
- 2/ Psiblast (with profile) research homologous protein chains in the CBE Database
- 3/ Ligands in contact with homologous protein chains are extracted
- 4/ A multi-mol2 text file with extracted ligands is returned and an email is send

## Stage 6 : Complex Modeling

### *Comparative docking*

In the first steps of processing, @TOME generates common core models of submitted query (stage 3) and gets ligands from PDB templates (stage 5). Model(s) and ligand(s) can be selected independently by user or automatically.

This “complex modeling” module allows transferring the selected ligand(s) from experimental structure(s) to selected model(s) target. The position of ligand in the model is obtained by the superimposition of the experimental complex with the model (**comparative docking**)

Two run modes are available: direct or crossed comparative docking.

-Direct comparative docking : only ligand from the template of model is inserted into the model (complex close to the PDB template).

-Crossed comparative docking : all selected ligand(s) are inserted in all selected model(s) (increasing variability of binding contacts).

### *Complex evaluation*

For each new modeled complex, 3 descriptors are calculated to evaluate the quality of complex model :

1/ *The calculated binding affinity (pKa)* is an average of 4 scoring functions: **MedusaScore**, **Xscore**, **DSX** and **Plants**.

A large study on 95 crystallographic complexes of **PDBbind** shows that the average increases the precision and robustness of the prediction (see Table).

*Table shows the correlation of Pearson (R) between calculated and measured affinity:*

Score funct.	Pearson coef. (R)
MedusaScore	0.49
Xscore	0.39
Plants	0,44
DSX	0,53
@TOME Avg.	0,53

2 / *Similarity to the crystallographic templates (PSim)* : in the first steps of @TOME processing, a profile of contacts between the receptor and the ligand is calculated on the crystallographic templates (Pat software). For each dock, similarity to this profile can be calculated (PSim) in order to evaluate the docking.

3/ *Quality of complex (Q)* is a consensus score including several features such as internal energy of ligand (**AMMP force field** calculation including angle, dihedral, bonds, VdW), complementarity function between ligand and binding site (**LPC** software), quality of receptor structure (**Qmean** software), type of contacts (Number of violation, H-Bonds).

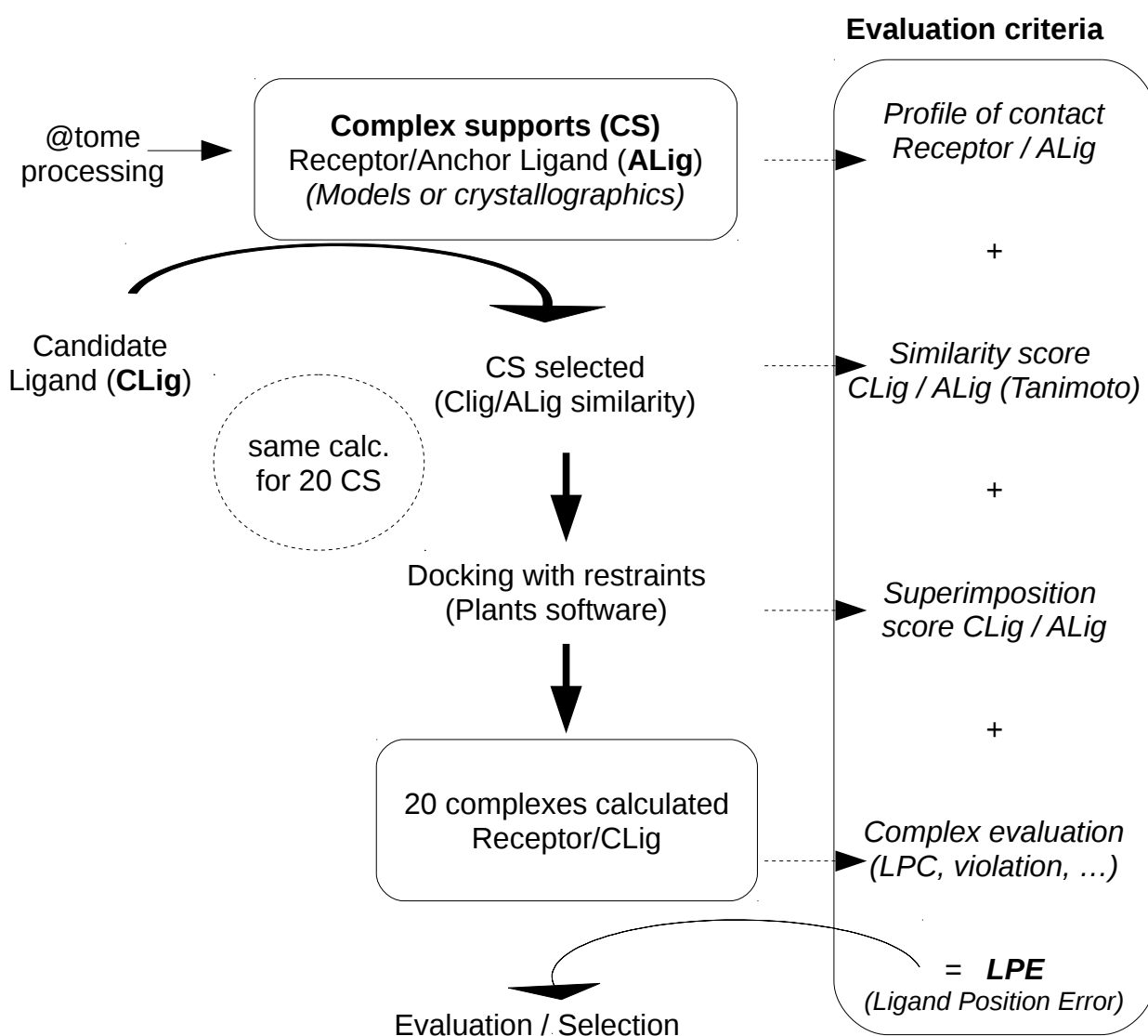
### *User interface*

All the modeled 3D structures of complex can be displayed on @tome thanks to Jmol 3D viewer.

## Stage 7 : Comparative Virtual Screening

The screening module of @TOME allows docking new ligand (candidate) taking into consideration the orientation of the crystallographic ligands (anchor) in the templates and their profile of contacts. These features are extracted from a database of suitable homologous complexes previously selected in the result of @TOME pipeline modeling (complex supports). The principle is described here.

*Flowchart of comparative screening processing & evaluation with LPE (Ligand Position Error)*





### ***Selection of the best modeled complexes to guide the docking (complex supports) :***

The “Select targets for screening” module permits to choose complex supports for the comparative screening. The selected crystallographic ligands are used as anchors. Complex supports for screening can be selected according to many criteria: receptor model quality, P<sub>Sim</sub>, complex model quality, predicted affinity, candidate/anchor **Tanimoto** similarity score between the candidate ligand and the anchor...

In the automatic mode, the closest ligands are first selected (Tanimoto score).

### ***Screening process :***

In the second step, the user has to upload a (multi) mol2 file with new candidate ligands and run the screening process. Exploration of different docking positions of candidate ligands is made by the **Plants** software which tests multiple conformations by rotation of dihedral angles or translation of the candidate ligand. A weight is added if the orientation of the new candidate ligand is close to a position of the ligand already known in the crystallographic complex (shape restraint).

Plants is used with the default settings. The weight given to the superposition with the ligand and anchor (shape constraint) is -3. Computational speed is equal to 2. The side chains of the binding site are not changed. The volume of the binding site is equal to the size of the ligand + 8Å.

### ***Clustering docking orientation :***

#### ***Ligand position clustering (LigPosition)***

For each candidate ligand, the new modeled complexes are superimposed in order to detect and rank the different orientations found in the docked process. These orientations are clustered (Fconv software (ref)). The most probable orientation corresponds to the cluster having the highest number of orientations.

### ***Complexes evaluation :***

Evaluation of new complexes is made with the same descriptors used in the complex modeling module (see above) : calculated affinity (pK<sub>a</sub>), similarity to profile (P<sub>Sim</sub>), complex quality (Q). Many other descriptors are extracted from the process (see tab below).

**A theoretical RMSD (Ligand Position Error or LPE in Angstrom)** is calculated using an SVM multi-variables regression method (**SVM<sub>Torch</sub>** software) built on appropriate complexes found in the PDB. After optimization, 5 descriptors are selected for SVM learning: Tanimoto similarity score, superimposition between candidate and anchor ligand, similarity to crystallographic profile of interaction (P<sub>Sim</sub>), complementarity function score (LPC) and the size of candidate ligand (number of atoms).

LPE permits to easily select docking close to crystallographic homologous complexes.

### ***Best docking selection***

The average of the LPE is calculated for each cluster of position and allows selecting the most probable orientation of ligand. In the selected cluster, the best affinity (pK<sub>a</sub>) determines the best docking selected by @tome.

### ***Docking values tab***

For each docking many values are extracted. All results can be displayed in the graphic HTML form or downloaded in a csv file. Values are described in the table below :

<b>Values (in csv file)</b>	
File;ModelID	User study name, S.A @tome . id
Template;Chain	Template PDB code & chain
AtomSA;	S.A. @tome score
QMean;	QMean score of receptor model
Identity	Sequence identity with template (%)
Anchor;Origin	Ligand PDB code & PDB id source
CDock	Comparative docking mode (= Raw)
AnchKd	Affinity calculated between receptor/anchor (pKa)
Candidate	User candidate name (in mol2 file)
Plants;PlantsRaw	Plants score (with and without weight)
PlantsKd	Plants pKa (calculated by linear regression on PDBbind)
MedusaRaw;MedusaSc;	Raw score & pKa calculated by linear regression (see ref)
XScore	XScore affinity score (pKa)
DSX_Raw;DSX_Score	Raw score & pKa calculated by linear regression on PDBbind
AtomeSc	@tome affinity score : (pKa = Avg. (Plants, Xscore, Medusa, DSX))
NbAtLig	Number of atoms in the candidate ligand (without hydrogen)
Tanimoto	Similarity between candidate ligand and anchor ligand
AnchorFit	Candidate/ligand superimposition score (Plants software)
LigandEnergy	Internal energy of ligand (AMMP force field)
LPC	LPC software score (receptor/ligand complementarity function)
AffinityMeasure;AffinityUnit	Measured affinity and unit (if available)
HB;Violation	Number of putative Hbond & violation in the receptor/ligand complex
PSim	Similarity to profile of receptor/ligand interaction in PDB templates
CpxQuality	Complex quality consensus score
LigPosition;NbPos	Orientation of the ligand (cluster number) & size of cluster
Reliab (= LPE)	Ligand Position Error (SVM multi-variables linear regression)

## **Technical overview**

@tome is a set of 3 multi core Linux servers in parallel configuration (load balancing & redundancy). A set of tera-byte disk of storage is available in secure RAID configuration mode. @tome pipeline is written in Perl 5 and each tool was installed locally on computer. This allows @tome being faster and more robust. The firsts stages (1 to 5) of processing are made on a background grid in parallel mode.

## References of embedded tools

- @tome V1:

Web Site: <http://abcis.cbs.cnrs.fr/>

Ref: Douguet D. and Labesse G (2001), Easier threading through web-based comparisons and cross-validations, *Bioinformatics*, 17(8), 752-3.

- PDB: Web Site: <http://www.rcsb.org/pdb/>

Ref: H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne: The Protein Data Bank. *Nucleic Acids Research*, 28 pp. 235-242 (2000).

- PSI-BLAST: Web Site:

Ref: Altschul, et al. *Nucleic Acids Res.* 25(17): 33100-3402 (1997) Schaffer, et al. *Nucleic Acids Res.* 29(14): 2994-3005 (2001)

- BLAST : Web Site: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>

Ref: Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." *J. Mol. Biol.* 215:403-410.

- HHSEARCH Web site HHPred: <http://toolkit.tuebingen.mpg.de/hhpred>

Ref: Soding, J. Protein homology detection by HMM-HMM comparison. *Bioinformatics* (2004), *bioinformatics*.

- FUGUE V2 Web site : <http://www-cryst.bioc.cam.ac.uk/fugue...> Version 2:

<http://tardis.nibio.go.jp/fugue/prf...>

Ref: Shi et al (2001). FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure- dependent gap penalties. *J. Mol. Biol.*, 310, 243-257.

- SP3: Web Site: <http://sparks.informatics.iupui.edu/sp3>

Ref: Hongyi Zhou and Yaoqi Zhou. *PROTEINS: Structure, Function, and Bioinformatics* 58:321–328 (2005)

- PSIPRED Web Site: <http://bioinf.cs.ucl.ac.uk/psipred/>

Ref: Jones DT. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292: 195-202.

- T-COFFEE : Web Site: <http://www.tcoffee.org/>

Ref: T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* , 2000 Sep 8;302(1):205-17.

- PSEA: Web Site: <http://abcis.cbs.cnrs.fr>

Ref: Labesse, Colloch, Pothier & Mornon, 1997, *Bioinformatics*

- JALVIEW: Web Site: <http://www.jalview.org>

Ref : Clamp, M., Cuff, J., Searle, S. M. and Barton, G. J. (2004), "The Jalview Java Alignment Editor", *Bioinformatics*, 20, 426-7.

- CLUSTALW : Web Site:

Ref: Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22(22):4673-4680.

- MUSCLE: Web Site: <http://phylogenomics.berkeley.edu/c...>

Ref: Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 2004, Vol. 32, No. 5 1792-1797

- PAT: Web site: <http://pat.cbs.cnrs.fr>

Ref :Jerome Gracy and Laurent Chiche (2005), Nucleic Acids Research Web Server issue, 33, W65-71.

- TITO: Web Site :[http://bioserv.cbs.cnrs.fr/HTML\\_BIO...](http://bioserv.cbs.cnrs.fr/HTML_BIO...)

Ref: G. Labesse and J.-P. Mornon, Bioinformatics, 14, 206-350

- SCWRL:

Ref: A. A. Canutescu, A. A. Shelenkov, and R. L. Dunbrack, Jr. A graph theory algorithm for protein side-chain prediction. Protein Science 12, 2001-2014 (2003).

- VERIFY3D Web Site: [http://nihserver.mbi.ucla.edu/Verify\\_3D/](http://nihserver.mbi.ucla.edu/Verify_3D/)

Ref: Eisenberg D, Lüthy R, Bowie JU. VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol. 1997;277:396-404.

- MODELLER : Web site : [http://salilab.org/modeller/about\\_m...](http://salilab.org/modeller/about_m...)

ref: N. Eswar, D. Eramian, B. Webb, M. Shen, A. Sali. Protein Structure Modeling With MODELLER. in press, 2006.

- JMOL: Web Site: [www.jmol.org](http://www.jmol.org)

- MAXCLUSTER: Web Site: <http://www.sbg.bio.ic.ac.uk/maxclu...>

Ref: Ortiz, A. R., Strauss, C. E., and Olmea, O. (2002). MAMMOTH (matching molecular models obtained from theory): an automated method for model comparison. Protein Sci, 11, 2606-21.

- MAXSUB:

Ref: Siew N, Elofsson A, Rychlewski L, Fischer D. MaxSub: an automated measure for the assessment of protein structure prediction quality. Bioinformatics. 2000 Sep;16(9):776-85.

- 3D-JURY:

Ref: Ginalski K, Elofsson A, Fischer D, Rychlewski L. 3D-Jury: a simple approach to improve protein structure predictions. Bioinformatics. 2003 May 22;19(8):1015-8.

- MATT: Web Site: <http://matt.csail.mit.edu>

Ref: M. Menke, B. Berger, L. Cowen, "Matt: Local Flexibility Aids Protein Multiple Structure Alignment", 2007

- KINDOCK:

Ref: Martin L, Catherinot V, Labesse G, kinDOCK: a tool for comparative docking of protein kinase ligands.Nucleic Acids Res. 2006 Jul 1;34(Web Server issue):W325-9.

- AUTODOCK:

Ref: Morris, G. M., Goodsell, D. S., Halliday, R.S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J. (1998), Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function J. Computational Chemistry, 19: 1639-1662.

- UniProtKB / Swiss-Prot: Web Site: <http://www.uniprot.org/uniprot/> Ref: Leinonen R., Diez F.G., Binns D., Fleischmann W., Lopez R., Apweiler R.- UniProt Archive, Bioinformatics 20:3236-3237(2004).

- CATH: Web Site: <http://www.cathdb.info/>

Ref : CATH: A Hierarchic Classification of Protein Domain Structures. Orengo,C.A., Michie,A.D., Jones,D.T., Swindells,M.B., Thornton,J.M. (1997). Structure, 5, 1093-1108

- SCOP: Web Site: <http://scop.mrc-lmb.cam.ac.uk/scop/>

Ref : Murzin A. G., Brenner S. E., Hubbard T., Chothia C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247, 536-540.

- SMART: Web Site: <http://smart.embl.de/>

Ref: SMART, a simple modular architecture research tool: Identification of signaling domains , Jörg Schultz, Frank Milpetz, Peer Bork and Chris P. Ponting (1998) Proc. Natl. Acad. Sci. USA 95, 5857-5864

- PROSITE: Web Site: <http://www.expasy.ch/prosite/>

Ref: Hulo N., Bairoch A., Bulliard V., Cerutti L., Cuče B., De Castro E., Lachaize C., Langendijk-Genevaux P.S., Sigrist C.J.A. The 20 years of PROSITE. Nucleic Acids Res. 2007 Nov 14.

- PFAM: Web Site: <http://pfam.sanger.ac.uk/>

Ref: The Pfam protein families database: R.D. Finn, J. Tate, J. Mistry, P.C. Coghill, J.S. Sammut, H.R. Hotz, G. Ceric, K. Forslund, S.R. Eddy, E.L. Sonnhammer and A. Bateman Nucleic Acids Research (2008) Database Issue 36:D281-D288

- MOLREP Web Site: <http://www.ccp4.ac.uk/html/molrep.html>

Ref: A. Vagin, A. Teplyakov, MOLREP: an automated program for molecular replacement., J. Appl. Cryst. (1997) 30, 1022-1025.

- X-RMSDist Web Site: <http://abcis.cbs.cnrs.fr> Autor: Stefano Trapani (CBS)

- Open BABEL Web Site:

- MedusaScore Web Site:

Ref: S. Yin, L. Biedermannova, J. Vondrasek, and N. V. Dokholyan, "MedusaScore: An accurate force-field based scoring function for virtual drug screening" Journal of Chemical Information and Modeling, 48: 1656-1662 (2008)