

@Tome v3 - Documentation

version : 09/08/2019

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

Table des matières

@Tome explained in a few words	3
@Tome flowchart:.....	4
Stage 1 : How to submit a query to @Tome ?.....	5
Stage 2 : Structural alignment (S.A.) research.....	6
Stage 3 : Structural alignment analyses & modeling.....	7
Stage 4 : @Tome rank calculation.....	8
Stage 5 : Ligand research & profile of contacts.....	9
Stage 6 : Complexes modeling.....	10
GUI : Database storage and graphic user interface.....	11
GUI : Available Processing Modules.....	13
Modules for structural alignment :.....	13
Module : Edit / adjust alignment.....	13
Module : Modify selection of Atome.....	13
Module : Hide or remove selected alignments.....	13
Module : Modify the clustering of 3D cores.....	13
Module : Add your PDB file.....	13
Modules for Modeller:.....	14
Module : Modeling Selected Alignment (<i>Modeller license is mandatory</i>).....	14
Module : Multi Template Modeling (<i>Modeller license is mandatory</i>).....	14
Module : Automatic selection & modeling (<i>Modeller license mandatory</i>).....	14
Modules for protein/ligand complexes :.....	15
Module : Modify ligands selection.....	15
Module : Make complexes by crossed comparative docking.....	15
Module : Profile of receptor-ligand contacts.....	15
Module : Recalculate models of complexes.....	15
Modules for virtual screening :.....	16
Module : Modify complexes selection.....	16
Module : Run virtual screening.....	16
Modules for extract & download :.....	17
Module : Alignment with Pat.....	17
Module : Structural Superimposition.....	17
Module : Binding Site Information.....	17
Module : Download models or ligands.....	17
Module : @Tome values in CSV format.....	17
Stage 7 : Comparative Virtual Screening.....	18
Technical overview.....	21
References of embedded tools.....	22

@Tome explained in a few words ...

@Tome (@utomatic Threading Optimization Modeling and Evaluation) is a web pipeline dedicated to protein structure modeling and small ligand docking based on comparative analyses.

@Tome allows fold recognition, template selection, structural alignment editing, structure comparisons, 3D-model building and evaluation. In case of protein-ligand complexes modeling, the binding site is automatically predicted and the ligand position is calculated by comparative docking or virtual screening. In our pipeline, softwares are efficiently interconnected to accelerate and automate the modeling process. Each step of the process includes several tools in order to obtain an accurate consensus score for evaluation. The main steps are summarized below.

@Tome can be used for modeling your molecule but also in studies of large set of proteins such as protein family or genome. Some resulting models are available on this platform. Furthermore, @Tome offers virtual screening tools on pre-calculated family of receptors such as nuclear receptor or protein-kinase.

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

@Tome looks for structural alignments of a user primary sequence using dedicated softwares (HHSearch, Fugue, Psi Blast, Sp3) and predict molecular structures with the help of software such as Tito, Scwrl or Modeller. @Tome provides a graphical and intuitive presentation of the alignments results. Results are evaluated with an important set of 1D, 2D & 3D evaluation scores extracted from softwares such as Tito, TCoffee, Verify3D, 3D-Jury and 3D clustering methods.

3D structure modeling ...

For each structural alignment, three types of model are calculated :

- 1/ Partial 3D common core: the Tito software keeps conserved amino acid residues and proposes a common core closed to the experimental template (same experimental coordinates).
- 2/ Complete 3D common core: the Scwrl software adds missing side chains on 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 ...

Receptor/ligand complexes can be modeled 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,...

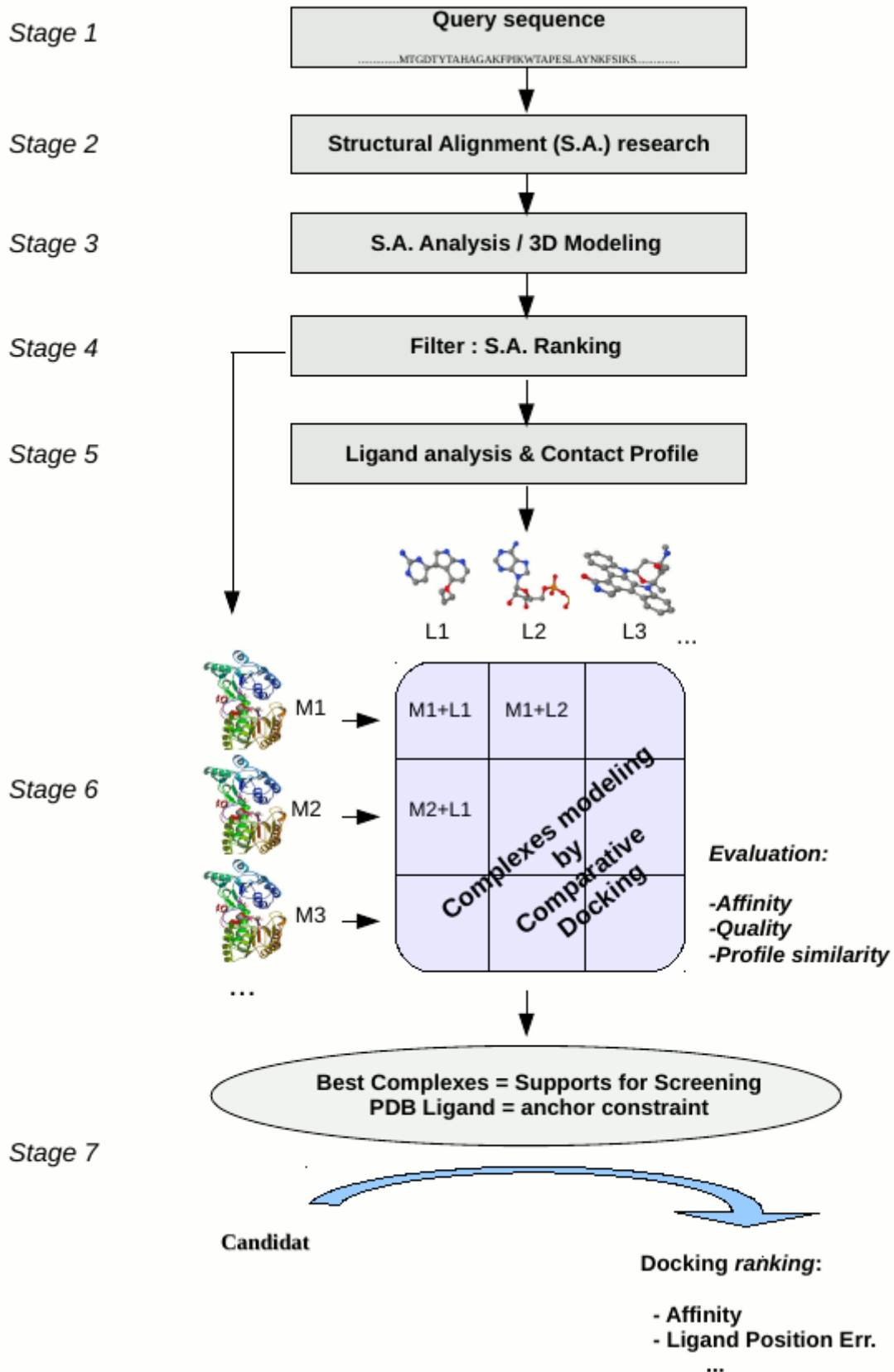
A profile of receptor/ligand contacts is calculated from the homologous crystallographic templates. At the end, each modeled complexes can be evaluated by calculating a similarity to this profile (Pat software).

Virtual screening

@Tome proposes to calculate virtual screening on modeled or crystallographic structure targets. In addition, the similar crystallographic ligands can be used during the screening process, as a shape restraints. Classical affinities score (pKd) are calculated using scoring functions. A additional score to evaluate the position of the ligand (Ligand Position Error (LPE)) was developed from a SVM machine learning. All these evaluation scores help us to select the best result.

@Tome flowchart:

This flowchart shows a complete process. All stages are described below in the text.



Stage 1 : How to submit a query to @Tome ?

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

The submission page is in the "My Modeling" menu.

This page allows the user submitting the primary sequence, title and email address (screen shot below).

Many tools are selected by default but the user can change options parameters and tools by clicking on the corresponding buttons. All the options are described below.

@TOME v3 Platform (Apr 2019)

Submit your sequence

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

@TOME permits also to create 3D model of receptor/ligand complexes by comparative docking of ligand between the experimental protein and model. At the end, ligand of modeled complex can be used to guide the positioning of new drugs during a virtual screening steps.

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

@TOME is a meta-server that uses many tools developed by the international scientific academic community. Most of these programs are for non-commercial use only. The service is provided on a "as is" basis without warranties of any kind.

The complete description of the pipeline @TOME is in this [Document](#)

Title :

Primary Sequence: [One letter code] eg:
RSKKNLSLALS LTADQMVSALLDAEPPILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRVP
GFVDLTLHDQVHLLC A WLEILMI GLVWRSM EHPGKLLFAPNLLDRNQGKCEGMVEIFDMLLAT
SSRFRMMNLQGEFVCLKSI ILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLIHLMA
[See example result here \(read-only\)](#)

Email :

Search for Homologous Sequences & Comparative Modeling

Ligands Selection, Complexes Modeling & Screening

Interface with Experimental Methods

Stage 2 : Structural alignment (S.A.) research.

The query primary sequence is submitted to selected structural alignment softwares. The following tools are embedded:

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 results. Homologous sequences are found into the PDB and the CBE databases (Contact Between Entities). CBE is a home made database that contains all complexes from the PDB and 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.

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 the 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 analyses & modeling

All structural alignments (S.A.) found are evaluated with a set of 1D, 2D and 3D scores. The tab below describes the sequential processes and the information extracted for each S.A.:

Features extracted from the process for each S.A.	Tools
Sequence identity	@Tome
3D structure of template (PDB files and chains)	@Tome
Classification & information : <i>cath, scop, uniprot, smart, prosite, Pfam, EC-Number csa, procognate, experimental method, ...</i>	@Tome
Secondary structure prediction	Psea
Sequence alignment score	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
Binding site prediction	@Tome
Ligand information <i>(small molecule / peptide / Nucleic Acid / Sugar / Ion)</i> <i>Name, Weight, contact distance,</i> <i>binding site conservation</i>	@Tome

Process calculated on all the SA together	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 results. 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:

$$Score(i) = \frac{\sum_i^T \left(\left(\frac{V_i - V_{min}}{V_{max} - V_{min}} \cdot 100 \right) \cdot N_t - 1 \right)}{\sum_i^T N_t}$$

where:

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

N_t = Total number of tools

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

Stage 5 : Ligand research & profile of contacts

Ligand research: @Tome calculates and evaluates 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 (default), peptides, sugars, nucleic acids or ions according to the study. Ligands with an atom at 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 (ϕ , ψ , ω , χ_1 , χ_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:

- Different 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 $> 50\%$

Stage 6 : Complexes 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). Best model(s) and ligand(s) are selected automatically. @Tome transfers 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).

Cross comparative docking is also available. A ligand can be inserted in several model(s) increasing variability of binding contacts and possibilities.

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).

GUI : Database storage and graphic user interface

Stages 1 to 5 are made in background on the @Tome grid computing. Scores & information are stored in a local database (in XML format). After calculation, @Tome provides access to all these scores on a graphical user interface.

The user interface consists of two parts. An main form that allows one to display all the database information of the study and permits to select S.A., models or ligands. After selection, the user can run calculation modules to perform many type of analyses ("Select Module" menu at the bottom of the page) . The results of these calculations are included in the database and dynamically appear in the main form.

Interactive main form:

The screenshot displays the @TOME V3 (Apr 2019) interface. On the left, there are several menu sections: 'Global output mode' (Atome Result, Contacts with Ligands), 'Show alignment' (Atome selection), 'Column output' (Atome Rank Value, Alignment, 3D Common Core, Structural Clustering, Modeller Result, Complexes Modelling, Screening, Templates Information, Sequence & Result Tab), and 'Server load'. The main area shows a table of results for a query sequence. The table has columns for 'Atome Classification', 'Score', 'Align', '3DCore', 'Cluster', 'PDB', and 'Complex'. The results are sorted by score, with the highest score being 92.06 for 'Blasp_CBE'. The table also includes a 'Sequence Alignment' section at the bottom, showing the query sequence and its alignment with various protein sequences, with residues highlighted in different colors to indicate specific properties.

Select information in interactive main form: In the left frame, two types of menu are available. The first one, called "Global output mode" permits to display S.A. 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, TCOffee Identity Score, 3D Clustering, Ligands Contact, hydrophobic amino-acid, disulfides bonds.

Column Output: For each S.A. entries, all the results can be displayed. The following table resumes available information in the menu / column:

Column & values available	Explanations
Score @Tome Rank: Tool score:	Consensus rank calculated by @Tome Raw score of structural alignment tools : Psiblast, Fugure, HHsearch or SP3
Alignment Identity : TCoffee Score:	Sequence identity between query and template (%) Sequence alignment score evaluated by T-Coffee Software.
3D Core Tito Score : Qmean : Verify3D:	Tito sequence-structure compatibility evaluation score Qmean score of Tito 3D model Verify3D Score of Tito 3D model
3D Clustering Cluster Num: 3D-Jury Struct. Score:	Cluster number 3D-Jury (MaxSub) Score :most frequent sub-structure found within a model set.
Modeling 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
Information Uniprot: Cath class: Scop class: Smart class: Prosites class: Pfam class: EC Number: Experiment Method: Taxonomy:	Uniprot classification and link Cath classification and link Scop classification and link Smart classification and link Prosites classification and link Pfam classification and link Enzyme Commission 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)
Complexes Modeling Show Ligands : Complex Affinity : Binding Site identity :	Show and select ligand associated to the structure Show the best affinity found for the complexe. Identity rate of main binding site (S1)
Sequence PDB Template : Complex Tab :	Display template primary sequence Show all the calculated complex in a same tab

In each menu, "Hide column" permits to hide column in the main form. The header of each column is a yellow button that change the display order.

GUI : Available Processing Modules

After selection of one or more S.A. (or ligands) the user can run calculation modules to perform many type of analyses on the selected entries. All the modules are accessible from the menu in the bottom of form.

To facilitate its use while maintaining a flexible access to most functionalities and complex pipelining, @Tome is run in basic mode by default but an expert mode can be activated on the result pages. This gives access to additional menus.

Modules are explained here:

Modules for structural alignment :

Module : Edit / adjust alignment

This module permits to improve the alignment of selected S.A. It proposes a manual re-alignment (Vito v2) or to use automatic softwares such as ClustalW or Muscle. Structural re-alignment is also possible with the Matt software .

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 : Modify selection of Atome

@Tome selects automatically the 20 best S.A. according calculated rank. This selection can be modified in order to hide redundancy or remove bad selection.

Module : Hide or remove selected alignments

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

Module : Modify the clustering of 3D cores

In stage 3 of the processing, @Tome clusterize all the Tito models with similar structure. This 3D clustering is done by the software MaxCluster with the default options. This module allows to restart the calculation with different options.

Each cluster is sorted by color. The main cluster is in green. The centroid model of cluster appears in bold characters.

Module : Add your PDB file

This module permits to add a new S.A. in the database. This local pdb file is used as a template for the new model. Th sequence alignment is made by the software Muscle. This new entry can be used like the other SA.

Modules for Modeller:

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 classify models (rank).

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 taken into consideration: distance, H bond and disulfide bond.

Module : Multi Template Modeling (*Modeller license is mandatory*)

Modeller can use more than one template in order to create better models. The modeled structure are added as a new SA entry 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 to "Modeling Selected Alignment" module. The code of the Modeller model is a 4 letters 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 taken into consideration: distance, H bond and disulfide bond.

Module : Automatic selection & modeling (*Modeller license mandatory*)

In automatic mode, Modeller is used to make mono-template models from the 20 best S.A. selected by @Tome. 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.

Modules for protein/ligand complexes :

Module : Modify ligands selection

This module allows user selecting ligand(s) from PDB template. Many filters are available: type of ligand, weight, receptor, binding site.... The selected ligands will be used for complexes modeling.
@Tome selects automatically ligands ("small molecules" type) of the main binding site (S1).

Module : Make complexes by crossed comparative docking

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

Module : Profile of receptor-ligand contacts

This module shows the current profile of ligand/receptor interactions previously calculated on homologous crystallographic templates (PDB). 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/ligands selected by the user.

Module : Recalculate models of complexes

This module makes it possible to recalculate the protein / ligand complexes in automatic mode. Some information is requested from the user such as the identity threshold or the type of ligand. The previous complexes are erased.

Modules for virtual screening :

Module : Modify complexes selection

The user can select here the best complexes after comparative docking. This complexes will be used as a target to guide the virtual screening. Selected complexes can be downloaded. A set of complexes is automatically selected by @Tome but can be modify here.

Module : Run virtual screening

Virtual screening is made on a database of suitable homologous complexes supports (models or structures) previously calculated in @Tome by comparative docking. The candidate ligand (mol2 file) is docked into the selected binding site taking into consideration the orientation of the crystallographic ligands found in the templates (shape restraint anchoring).

Exploration of different docking positions is made by the software Plants that calculates multiple conformations by rotation of dihedral angles or ligand translation. The side chains of residues of the binding sites are not taken into account in the calculation.

For each modeled complex many descriptors are calculated:

- A theoretical affinity (pKd) is calculated using several scoring functions (MedusaScore, X-score, DSX).
- Similarity with profile of contacts ligand/receptor in homologous crystallographic complexes (PDB).
- Theoretical ligand position error (LPE) is calculated by multiple regression (SVM) with many descriptors.

Modules for extract & download :

Module : Alignment with Pat

This module allows one to download 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 : Structural Superimposition

3D structures can be superimposed and downloaded from 3D core Tito result, Tito/Scwrl, Modeller or PDB template. Different formats are available: NMR type bundle, single PDB file with one chain by model or separate PDB file.

For superimposition step, @Tome uses the Matt software (long time consuming) or MaxSub/Maxcluster software.

Module : Binding Site Information

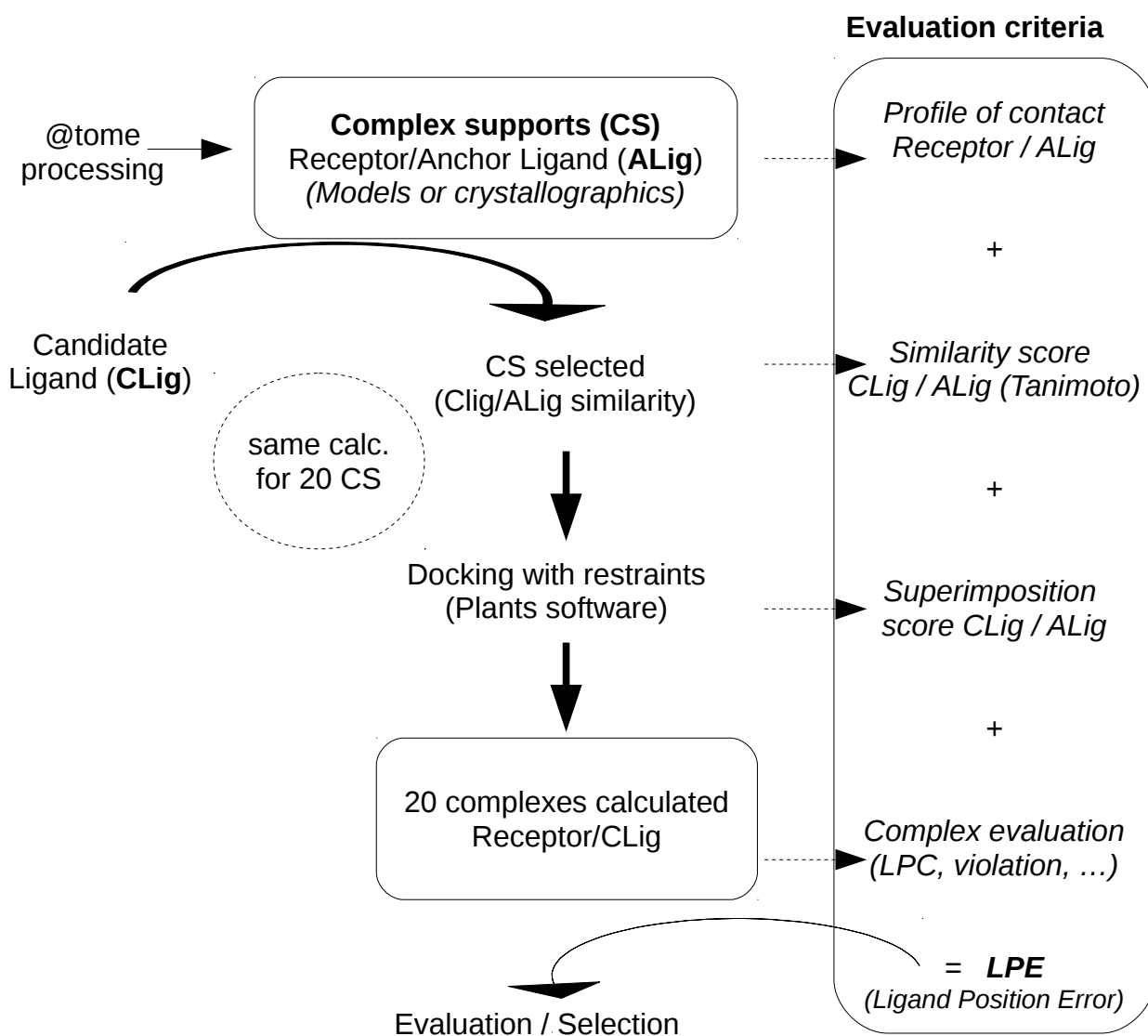
Module : Download models or ligands

Module : @Tome values in CSV format

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 shape restraint) 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 “complexes selection” module permits to choose target for the virtual screening (called complex support). The selected crystallographic ligands are used as anchors. Complex supports for screening can be selected according to many criteria: receptor model quality, PSim, 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 virtual 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.

Complexes evaluation :

Evaluation of new complexes is made with the same descriptors used in the complex modeling module (see above) : calculated affinity (pKa), similarity to profile (PSim), 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 Torch software) built on appropriate complexes found in the PDB. The setup of this supervised algorithm is explained here:

Structures used for training and test databases of SVM are extracted from representative part of the PDB including 95 crystallographic complexes from different families in CATH classification. Selected structures are re-modeled with various templates, various level of sequence identity and a large set of ligands found in the same family of complex (crossed comparative docking).

The 7235 models of complexes obtained are more or less close to crystallographic structures and their ligands are used as shape restraint in a docking recalculation of crystallographic ligand.

For each docking, a large set of descriptors (table below) is calculated and related to RMSD between the modeled and the crystallographic position of ligand.

49 PDB entries (3850 vectors) are used to train a supervised SVM machine in a multi-variable linear regression mode. 46 PDB entries (3385 vector) are used for validation. After optimization, 5 descriptors are selected for SVM learning: Tanimoto similarity score, superimposition between candidate and anchor ligand, similarity to crystallographic profile of interaction (PSim), complementarity function score (LPC) and the size of candidate ligand (number of atoms).

SVM is used with the following settings: Gaussian kernel in regression mode, high precision required to stop learning (-e = 0.001) and sparse data mode. At the end, a theoretical RMSD called “Ligand Position Error or LPE” is calculated with the same descriptors extracted from new cases. LPE permits to easily select docking close to crystallographic homologous complexes.

The Pearson coefficient of correlation (R) is equal to 0,81 on test base and the average error between LPE and real RMSD is equal to 1,26 Angstrom.

Clustering docking orientation :

Ligand position clustering (LigPosition)

For each candidate ligand, the new modeled complexes are superimposed in order to detect 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 best LPE score.

Best docking selection

The LPE is calculated for each position and permits to select the most probable cluster of orientation of ligand. In the selected cluster, the best affinity (pKa) 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 :

Values (in csv file)	
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: Douquet D. and Labesse G (2001), *Easier threading through web-based comparisons and cross-validations*, *Bioinformatics*, 17(8), 752-3.
- @Tome-V2: *a new pipeline for comparative modeling of protein-ligand complexes*. Pons & Labesse - *Nucleic Acids Research, Web Server Issue- 2009* - doi: 10.1093/nar/gkp368
- PDB: 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: Altschul, et al. *Nucleic Acids Res.* 25(17): 33100-3402 (1997) Schaffer, et al. *Nucleic Acids Res.* 29(14): 2994-3005 (2001)
- BLAST : 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 : Soding, J. *Protein homology detection by HMM-HMM comparison. Bioinformatics (2004), bioinformatics.*
- FUGUE V2 : 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: Hongyi Zhou and Yaoqi Zhou. *PROTEINS: Structure, Function, and Bioinformatics* 58:321–328 (2005)
- PSIPRED : Jones DT. (1999) *Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol.* 292: 195-202.
- T-COFFEE : : T-Coffee: *A novel method for fast and accurate multiple sequence alignment. J Mol Biol* , 2000 Sep 8;302(1):205-17.
- PSEA:: Labesse, Colloc'h, Pothier & Moron, 1997, *Bioinformatics*
- JALVIEW : Clamp, M., Cuff, J., Searle, S. M. and Barton, G. J. (2004), "The Jalview Java Alignment Editor", *Bioinformatics*, 20, 426-7.
- CLUSTALW : 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: 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: Jerome Gracy and Laurent Chiche (2005), *Nucleic Acids Research Web Server issue*, 33, W65-71.
- TITO: G. Labesse and J.-P. Moron, *Bioinformatics*, 14, 206-350
- SCWRL: 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 : Eisenberg D, Lüthy R, Bowie JU. *VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol.* 1997;277:396-404.

- MODELLER : N. Eswar, D. Eramian, B. Webb, M. Shen, A. Sali. *Protein Structure Modeling With MODELLER*. in press, 2006.
- JMOL: Web Site: www.jmol.org
- MAXCLUSTER: 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.
- PROFIT 3.1 : Fitting was performed using the McLachlan algorithm (McLachlan, A.D., 1982 "Rapid Comparison of Protein Structures", *Acta Cryst A*38, 871-873) as implemented in the program ProFit (Martin, A.C.R. and Porter, C.T., <http://www.bioinf.org.uk/software/profit/>)
- MAXSUB: 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: 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: M. Menke, B. Berger, L. Cowen, "Matt: Local Flexibility Aids Protein Multiple Structure Alignment", 2007
- KINDOCK: 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: 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 an Empirical Binding Free Energy Function *J. Computational Chemistry*, 19: 1639-1662.
- UniProtKB / Swiss-Prot: Leinonen R., Diez F.G., Binns D., Fleischmann W., Lopez R., Apweiler R.- UniProt Archive, *Bioinformatics* 20:3236-3237(2004).
- CATH: 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 : 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: 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: 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: 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 :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)
- MedusaScore: 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)

- *LPC*: Sobolev V., Sorokine A., Prilusky J., Abola E.E., Edelman M., (1999) Automated analysis of interatomic contacts in proteins. *Bioinformatics*, 15, 327-332
- *PLANTS*: Korb, O.; Stützle, T.; Exner, T. E. "PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design" *Lecture Notes in Computer Science* 4150, 247-258 (2006).
- *FCONV*: Neudert G , Klebe G. fconv: Format conversion, manipulation and feature computation of molecular data. *Bioinformatics*. 2011 Apr 1;27(7):1021-2. doi: 10.1093/bioinformatics/btr055. Epub 2011 Feb 18.
- *XSCORE*: Wang, R.; Lai, L.; Wang, S. Further Development and Validation of Empirical Scoring Functions for Structure-Based Binding Affinity Prediction. *J. Comput.-Aided Mol. Des.* 2002, 16, 11-26.
- *DSX* :G Neudert G Klebe DSX: A Knowledge-Based Scoring Function for the Assessment of Protein–Ligand Complexes - *J. Chem. Inf. Model.*, 2011, 51 (10), pp 2731–2745
- *PDBBind* :Wang, R.; Fang, X.; Lu, Y.; Wang, S. "The PDBbind Database: Collection of Binding Affinities for Protein-Ligand Complexes with Known Three-Dimensional Structures", *J. Med. Chem.*, 2004; 47(12); 2977-2980
- *NJPLOT*: Perrière, G. and Gouy, M. (1996) WWW-Query: An on-line retrieval system for biological sequence banks. *Biochimie*, 78, 364-369.
- *BindingDB* : iu,T., Lin,Y., Wen,X., Jorriksen,R.N. and Gilson,M.K. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities *Nucleic Acids Research* 35:D198-D201 (2007).
- *VITO* : Catherinot,V. and Labesse,G. (2004) ViTO: tool for refinement of protein sequence structure alignments. *Bioinformatics*, 20, 3694-3696.
- *Open BABEL*: NO'Boyle, M Banck,CA James, C Morley,T Vandermeersch and GR Hutchison; *Open Babel: An open chemical toolbox*, *Journal of Cheminformatics*20113:33, 10.1186/1758-2946-3-33, 2011
- *QMEAN*: Benkert P1, Tosatto SC, Schomburg D. A comprehensive scoring function for model quality assessment.. *Proteins*. 2008 Apr;71(1):261-77.
- *SVMtorch*: Collobert and Bengio, *SVMtorch: Support Vector Machines for Large-Scale Regression Problems*, *Journal of Machine Learning Research*, vol 1, pages 143-160, 2001.