Deliverable D7.7

Project Title:	World-wide E-infrastructure for structural biology	
Project Acronym:	West-Life	
Grant agreement no.:	**675858**	
Deliverable title:	Quality analysis workflow for predicted complexes	
Lead Beneficiary:	EMBL	
Contractual delivery date:	Month 30	
Actual delivery date:	Month 30	
WP No.	7	
WP Title	Joint research	
WP leader:	Jose María Carazo	CSIC
Contributing partners:	EMBL-EBI	

Deliverable written by Nurul Nadzirin

Work Package 7, Deliverable 7.7: Quality analysis workflow for predicted complexes

Deliverable D7.7 1
Work Package 7, Deliverable 7.7: Quality analysis workflow for predicted complexes 2
Introduction and rationale of the quality analysis workflow 3
Web server 4
Landing page and assessment mode 4
Expert interface 6
Input and result 7
References 8

Introduction and rationale of the quality analysis workflow

Protein–protein interactions and protein assemblies play a crucial role in all cellular processes. Their disruption often leads to disease. Much of what we know about protein complexes is derived from data on the three-dimensional (3D) structures of these complexes determined by experimental methods. But the number of protein assemblies for which detailed structural information is available represents only a small fraction of the protein assemblies in the cell that can be detected by proteomics and other methods (Mishra, 2012).

Computational procedures capable of reliably generating structural models of multi-protein assemblies starting from the atomic coordinates of the individual components, the so-called ‘‘docking’’ methods, therefore play an important role in helping to bridge the gap. The Critical Assessment of Predicted Interactions (CAPRI) (http://www.ebi.ac.uk/msd-srv/capri/ and http://capri-docking.org), a community-wide initiative established in 2001, has been at the fore-front in this endeavor (Janin, et al., 2003).

In CAPRI, individual groups, that develop docking procedures, predict the 3D structure of a protein complex from the known structures of the components. The assessment of predicted models is carried out by comparing these with experimental structure (the Target). Since its inception, the CAPRI committee has organized six evaluation meetings. During these evaluation meetings, discussions have led to established standards for the parameters and criteria used to evaluate the quality of the predicted complexes. These parameters and criteria are explained in a previous WestLife report (refer to Milestone 7.6 https://zenodo.org/record/1035144).

The assessment of predicted complexes is carried out by an assessment team. The evaluation is based on a number of criteria, largely adopted by the CAPRI community. Now, this assessment program, CAPRI Analyse, is made available to users worldwide as a web server. Theoretically, if a user has experimental structure of macromolecular complex (target structure) available, and has modelled the constituent components (the unbound structures) to predict the structure of macromolecular complex using computation methods, the “CAPRI Analyse” server can provide evaluation of the quality of the predicted models based on CAPRI protocol. CAPRI analysis protocol has been developed over a period of more than a decade, and this is the first time users have direct access to the assessment program via a web server.

Web server

The site runs on an Apache 2.4 web server. The server-end was developed using the web framework Django (version 1.11), which was useful to streamline validation of user input based on Django models and handle URL mapping.

Landing page and assessment mode

The landing page for CAPRI Analyse can be accessed from:

https://www.ebi.ac.uk/pdbe/complex-pred/capri-validation/

Two assessment modes are allowed; namely the standard assessment, and the peptide assessment.

Since its inception, the focus of CAPRI has expanded significantly, to include complexes of proteins with other macromolecules, peptides, nucleic acids, and carbohydrates. Several CAPRI Rounds have also focused on the problems of predicting protein binding affinities and modeling the positions of water molecules in protein–protein interfaces. Users should use the Standard mode if their target is a standard protein-protein complex, whereas the Peptide mode should be used by users with a standard protein-peptide complex as a target. Other users are encouraged to use the standard mode, however users who are familiar with the parameters can choose to run both the standard and peptide assessment modes using the Expert interface. For example, water analysis can be turned on if needed, or structural alignment can be carried out using Multiprot instead of the default Sofist software program. More details on the Expert interface are provided in the next section.

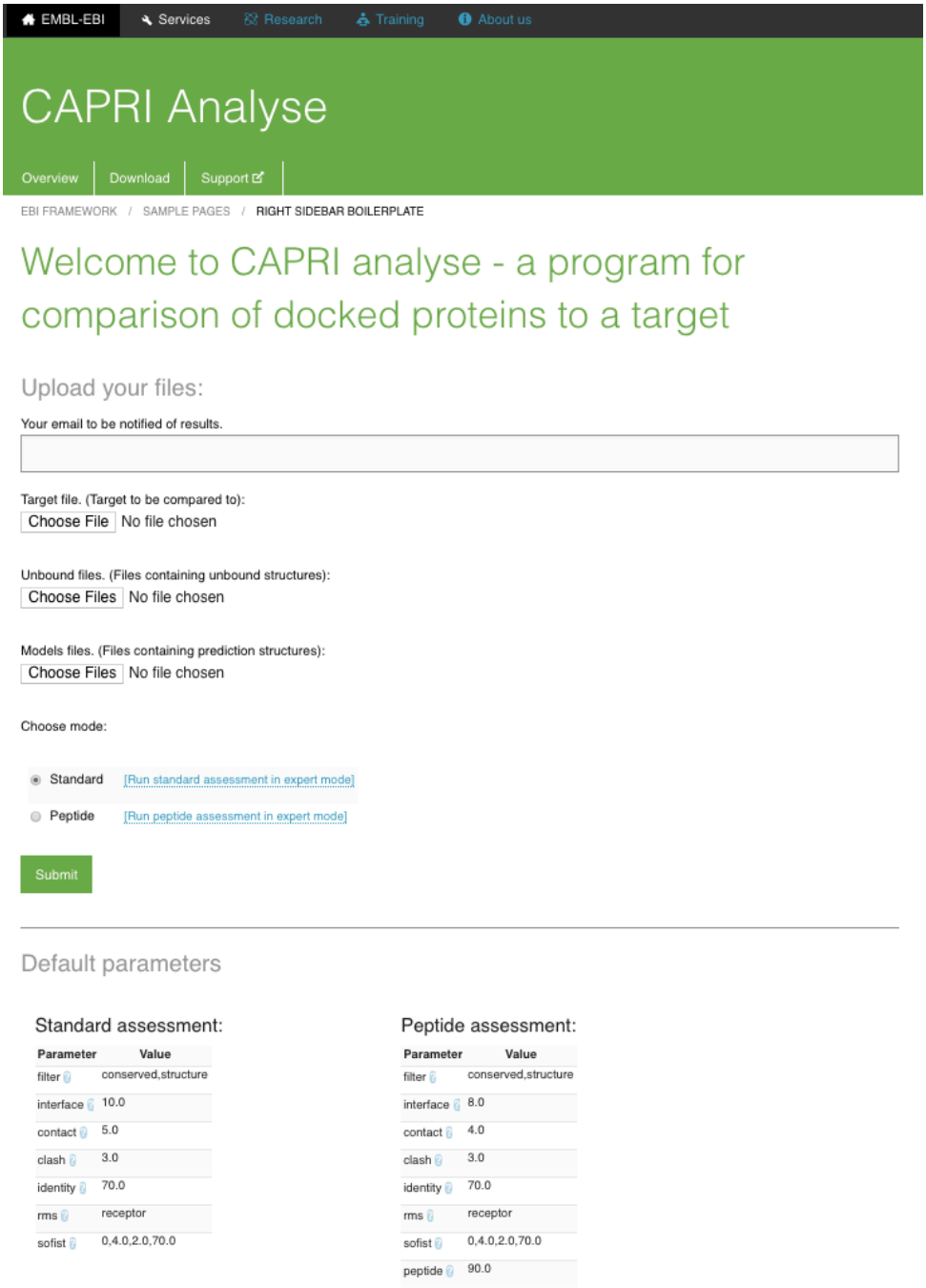

Figure 1: Landing page of the web server

Expert interface

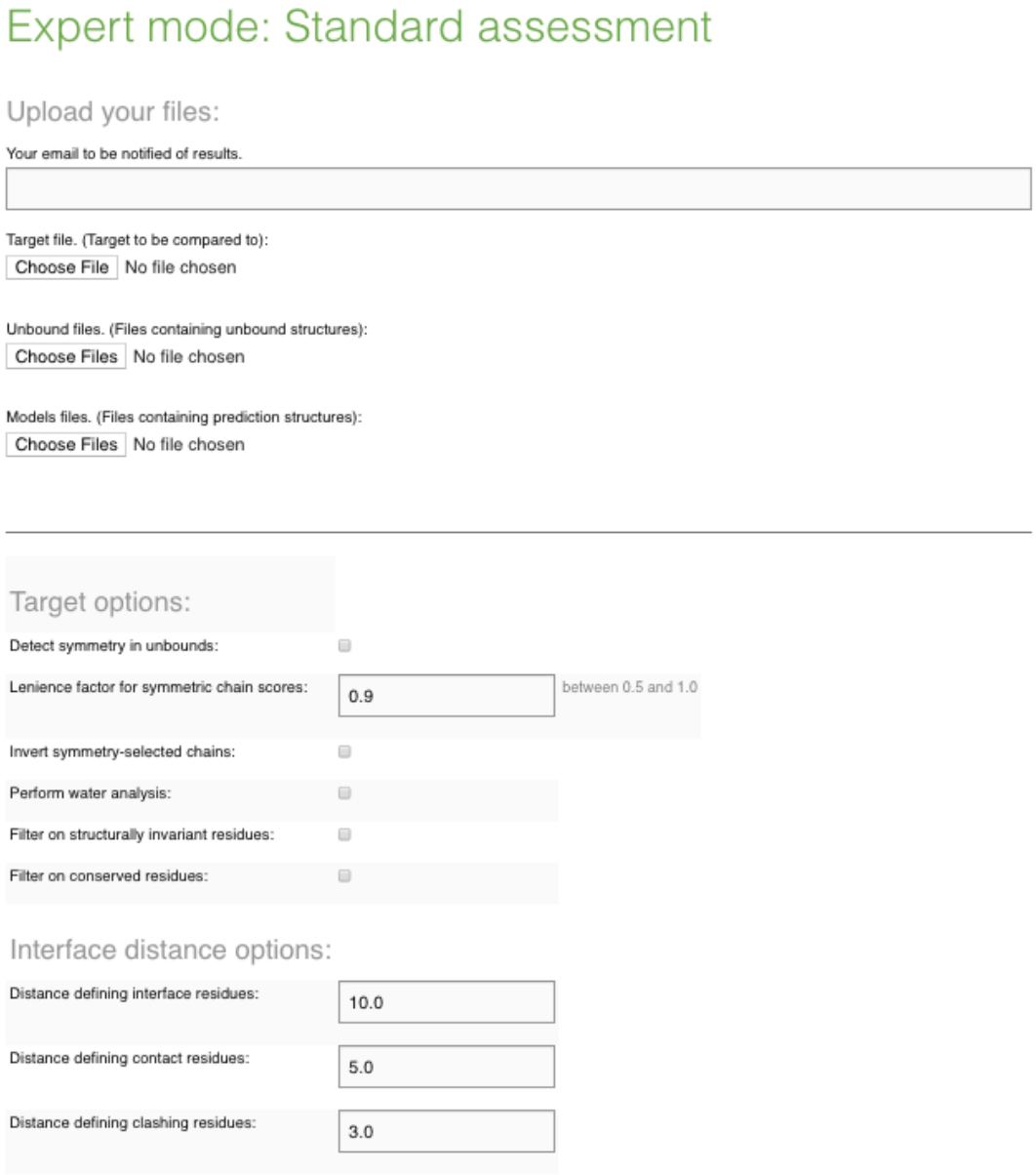

Figure II: Parts of the interface showing the expert mode for standard assessment

The expert interface allows users to change the parameters used to run the assessment pipeline. The parameters and expert levels are detailed below:

Parameter	**Type**	**Default value**	**Condition**	**Expert level (0 – standard, 1-expert)**
Protein-peptide assessment	Boolean	FALSE	-	0
Peptide Identity	Float	90	Percentage (0-100)	0
Lenience factor for symmetric chain scores	Float	0.9	Between 0.5-1.0	1
Detect symmetry	Boolean	FALSE	-	1
Invert symmetry-selected chains	Boolean	FALSE	-	1

Perform water analysis	Boolean	FALSE	-	1
Filter on structurally invariant residues	Boolean	TRUE	-	1
Filter on conserved residues	Boolean	TRUE	-	1
Distance defining interface residues	Float	10	5-20	1
Distance defining contact residues	Float	5	2-10	1
Distance defining clashing residues	Float	3	0-4	1
Create unbound and interface complexes	Boolean	FALSE	-	1
Ignore in clash threshold calculation	Boolean	FALSE	-	1
Set clash threshold for evaluated structures	Float	-1	-1 to 200	1
Maximum distance between consecutive residues	Float	20	?	1
Minimum required sequence identity and presence	Float	70	Percentage (0-100)	1
RMS fit, receptor or receptor interface	Choice (receptor or interface)	“receptor”	-	1
RMS fit on mass-weighted	Boolean	FALSE	-	1
Sofist clustering option	Integer	0	0 or 1	1
Sofist maximum RMS aligned structures	Float	4	0-4	1
Sofist segment pairs	Float	2	0-2	1
Sofist min fraction aligned SS elements	Float	70	Percentage (0-100)	1
How many models per file to evaluate at most	Integer	0	-	1
Allow automatic chain splitting	Boolean	TRUE	-	1
Use strict chain checking	Boolean	FALSE	-	1
Use Multiprot for structure alignment	Boolean	FALSE	-	1

Input and result

To analyse predicted complex models against the target structure, three type of file inputs are necessary:

- Target structure, i.e. the structure of the complex obtained using experimental methods
- Structures of the unbound components in the complex
- A file containing the predicted models.

Users also need to provide their email address so they can be notified when the program has completed, which typically takes less than 5 minutes for a standard protein-protein complex and two predicted models. The results are shown as a sortable table with a number of parameters, most interesting being the classification of predictions denoted as ‘high’, ‘medium’, or ‘acceptable’ accuracy. The reader is referred to the previous West-Life report for a detailed description of these

parameters and the corresponding thresholds used in classifying predictions.

References

Janin, J., *et al.* CAPRI: a Critical Assessment of PRedicted Interactions. *Proteins* 2003;52(1):2-9.
Mishra, S. Computational prediction of protein-protein complexes. *BMC Res Notes* 2012;5:495.