A User Manual

of

IDEAL

***I**ntrinsically **D**isordered proteins with **E**xtensive **A**nnotations and **L**iterature*

IDEAL, Intrinsically Disordered proteins with Extensive Annotations and Literature (http://www.ideal.force.cs.is.nagoya-u.ac.jp/IDEAL/), is a collection of knowledge on experimentally verified intrinsically disordered proteins. IDEAL contains manual annotations by curators on intrinsically disordered regions, interaction regions to other molecules, post-translational modification sites, references, and structural domain assignments.

IDEAL development team
Satoshi Fukuchi, Ph. D. Maebashi Institute of Technology
Motonori Ota, Ph. D. Nagoya University

Ideal-admin@force.cs.is.nagoya-u.ac.jp

1) *Top page*

User can find proteins in interest by the search tool [1]. User can select one of “Full text”, “Uniprot accession”, “protein name”, and “PDB id” categories. The search tool is also available from the “Search” button [3].

Another way to access IDEAL entries is to open the entry list by clicking “Browse” [2].

IDEAL also provides the BLAST search [4]. User can input an amino acid sequence to find the homologs in IDEAL.

All of the data in IDEAL is available in the XML format [5].

The logo, IDEAL [6], is the link to the top page. This header (blue bar) always locates at the top of any pages in IDEAL.

IDEAL

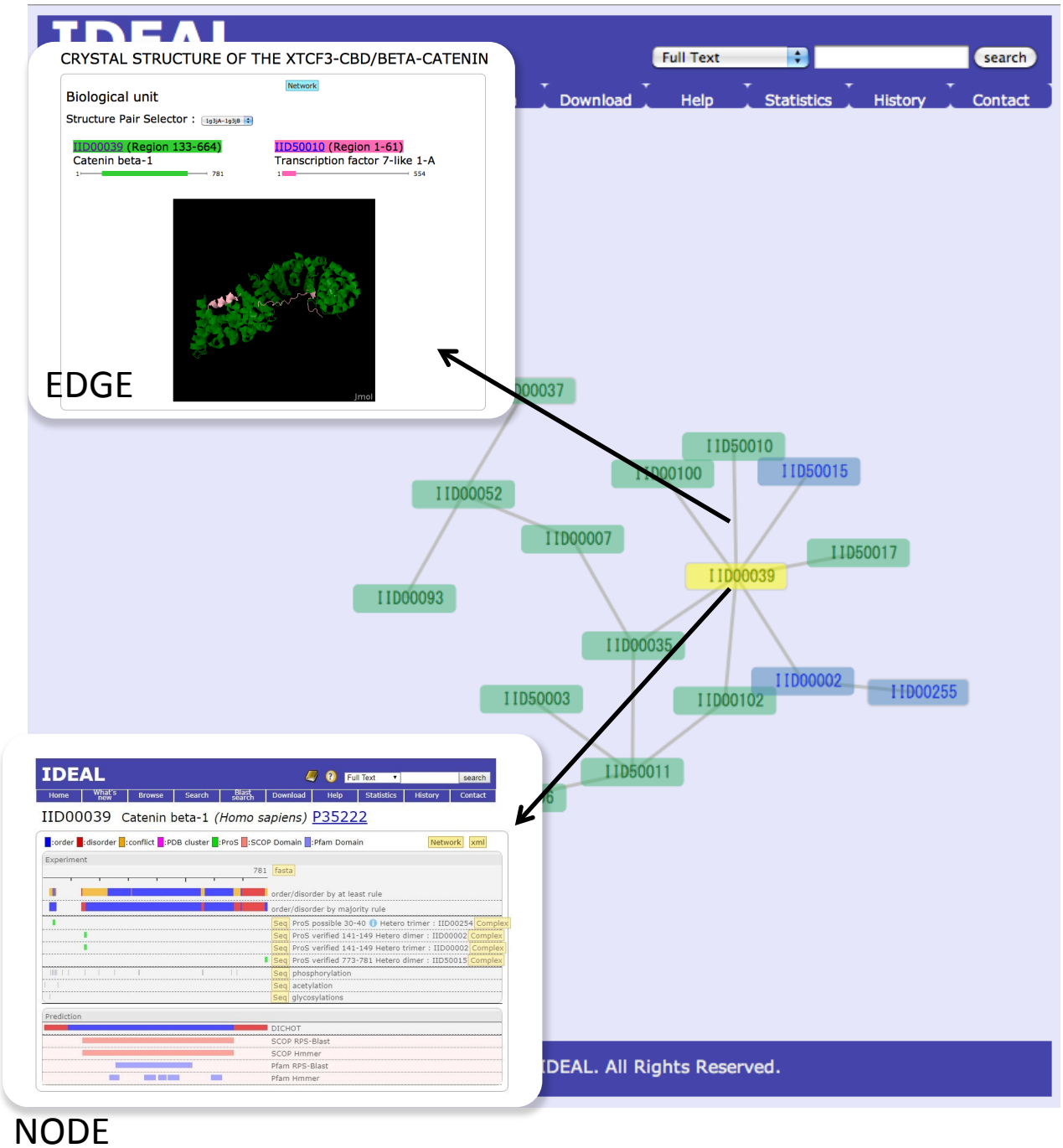
What is IDEAL ?
IDEAL provides a collection of knowledge on experimentally verified intrinsically disordered proteins (IDPs) more

	What'new	Find the new features of IDEAL
[2]	**Browse**	See the list of all proteins.
[3]	**Search**	Search by keyword, Uniprot accession and PDB ID.
[4]	**Blast search**	Find similar sequences in IDEAL.
[5]	**Download**	Get IDEAL in XML format.
	Help	Find the way to dig up IDEAL.
	Statistics	See the current status of IDEAL

System Requirements
In order to use this website safely and cormfortably, we recommend the use of the following browsers and versions:Windows Internet Explorer 9.0, Firefox 8.0, Safari 5.0, Google Chrome 16.0, Macintosh Safari 5.0, Firefox 8.0

2) *NODE and EDGE*

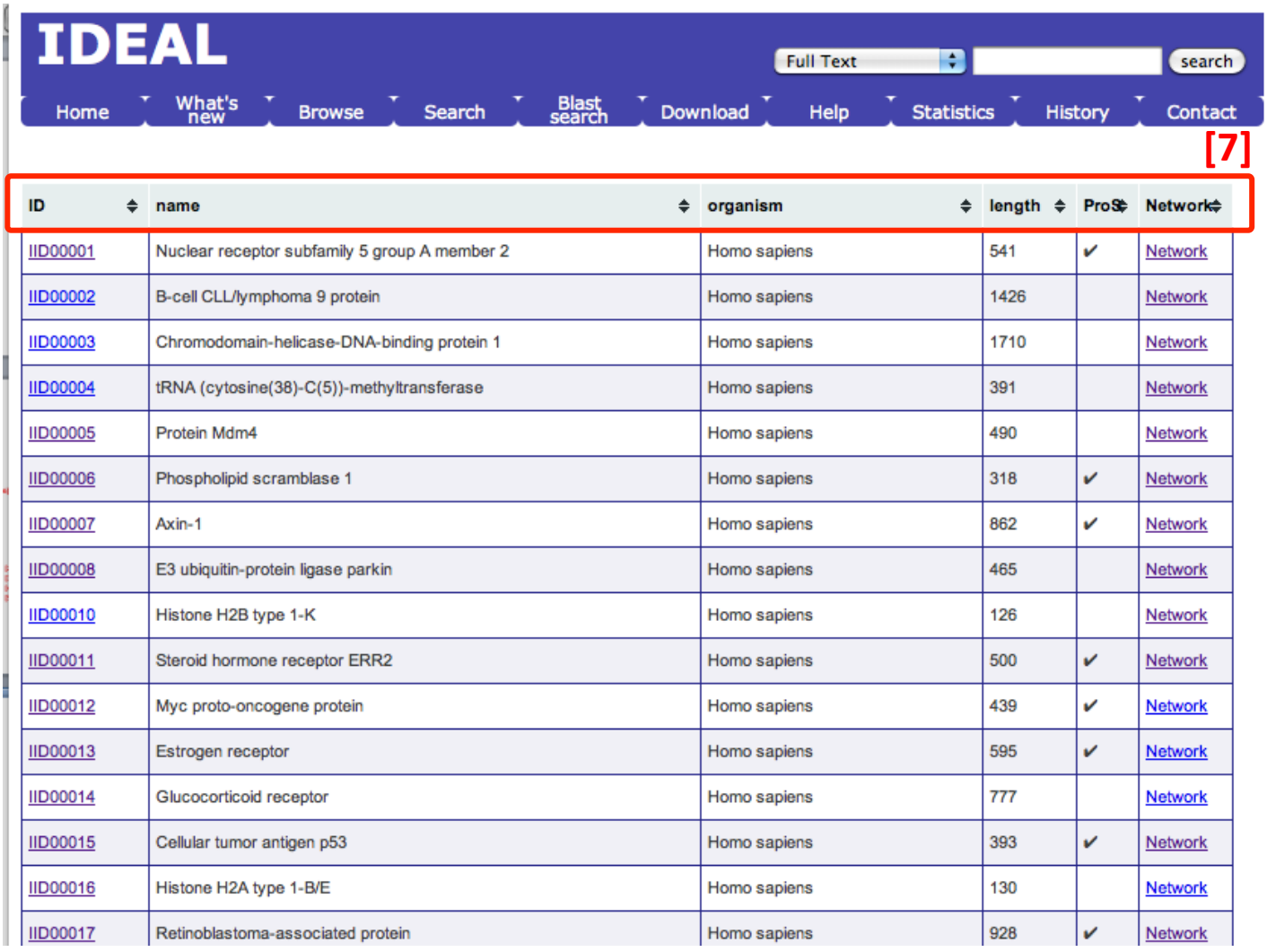
IDEAL refers to an entry (protein) as a NODE, and an interaction of two entries (PPI) as an EDGE, preparing NODE pages as well as EDGE pages. The former contains the detailed information for an IDP. The latter shows a structural complex of the entry and its binding partner. The NODE pages can be linked by the EDGE pages to compose PPI networks. An example of the PPI networks is shown below. The networks can be available from the entry list, each of NODE pages, or EDGE pages.



3) *The entry (NODE) list*

The list tabulates all of the entries in the descending order by the IID (default). IID is labeled on each protein, starting from IID0001 for human proteins, IID5001 for other eukaryotic proteins, and IID9001 for proteins of the remaining organisms including viruses. IID can be clickable to present each entry.

User can sort the list by cricking the items in the header [7]. The column ProS shows presence or absence of the protean segment, which is IDRs with ability of structural transformation. The Network links to the PPI network map, which contains the NODE entry.



ID	name	organism	length	ProS	Network
IID00001	Nuclear receptor subfamily 5 group A member 2	Homo sapiens	541	✔	Network
IID00002	B-cell CLL/lymphoma 9 protein	Homo sapiens	1426		Network
IID00003	Chromodomain-helicase-DNA-binding protein 1	Homo sapiens	1710		Network
IID00004	tRNA (cytosine(38)-C(5))-methyltransferase	Homo sapiens	391		Network
IID00005	Protein Mdm4	Homo sapiens	490		Network
IID00006	Phospholipid scramblase 1	Homo sapiens	318	✔	Network
IID00007	Axin-1	Homo sapiens	862	✔	Network
IID00008	E3 ubiquitin-protein ligase parkin	Homo sapiens	465		Network
IID00010	Histone H2B type 1-K	Homo sapiens	126		Network
IID00011	Steroid hormone receptor ERR2	Homo sapiens	500	✔	Network
IID00012	Myc proto-oncogene protein	Homo sapiens	439	✔	Network
IID00013	Estrogen receptor	Homo sapiens	595	✔	Network
IID00014	Glucocorticoid receptor	Homo sapiens	777		Network
IID00015	Cellular tumor antigen p53	Homo sapiens	393	✔	Network
IID00016	Histone H2A type 1-B/E	Homo sapiens	130		Network
IID00017	Retinoblastoma-associated protein	Homo sapiens	928	✔	Network

3) *NODE page*

Summary of the annotated regions

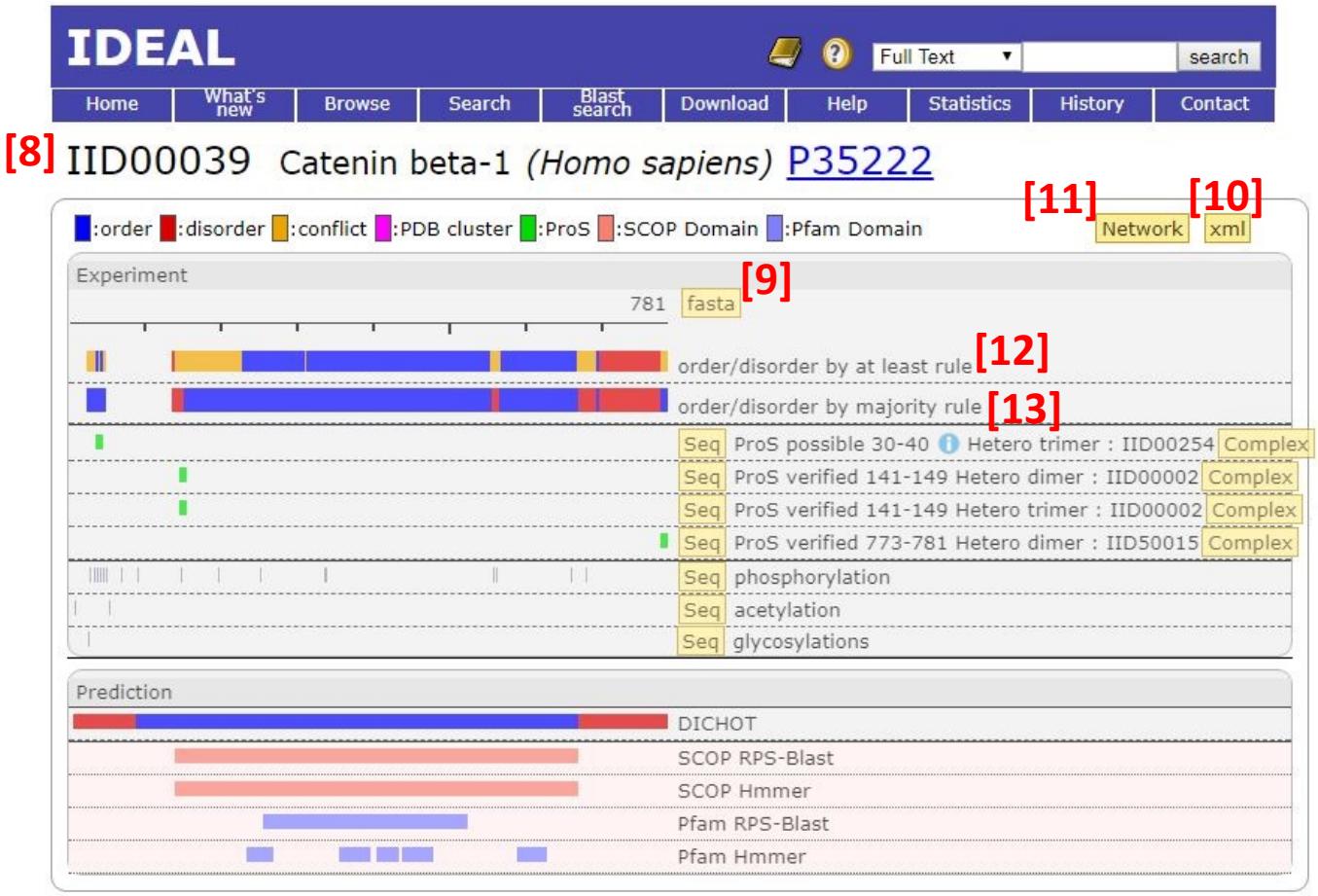
This is an example of a NODE page.

The identifier, IID, the protein name, the source organism, and the link to Uniprot are listed [8]. The annotated regions, functional regions, and domain assignments are presented color bars.

The amino acid sequence in the FASTA format and the information in the XML format are available at [9] and [10], respectively. The network map is linked by [11].

Two color bars [12] and [13] summarize the order/disorder annotations. The “at least rule” bar [13] shows the summary based on the “at least rule”. The "at least rule" assigns “order” (“disorder”) to a region if the region was annotated as ordered (disordered) at least once. When the annotation is inconsistent, the region is annotated as conflict.

The “majority rule” bar [12] shows the summary by the “majority rule”. The "majority rule" assigns “order” or “disorder” to a region according to the majority decision of all evidences.

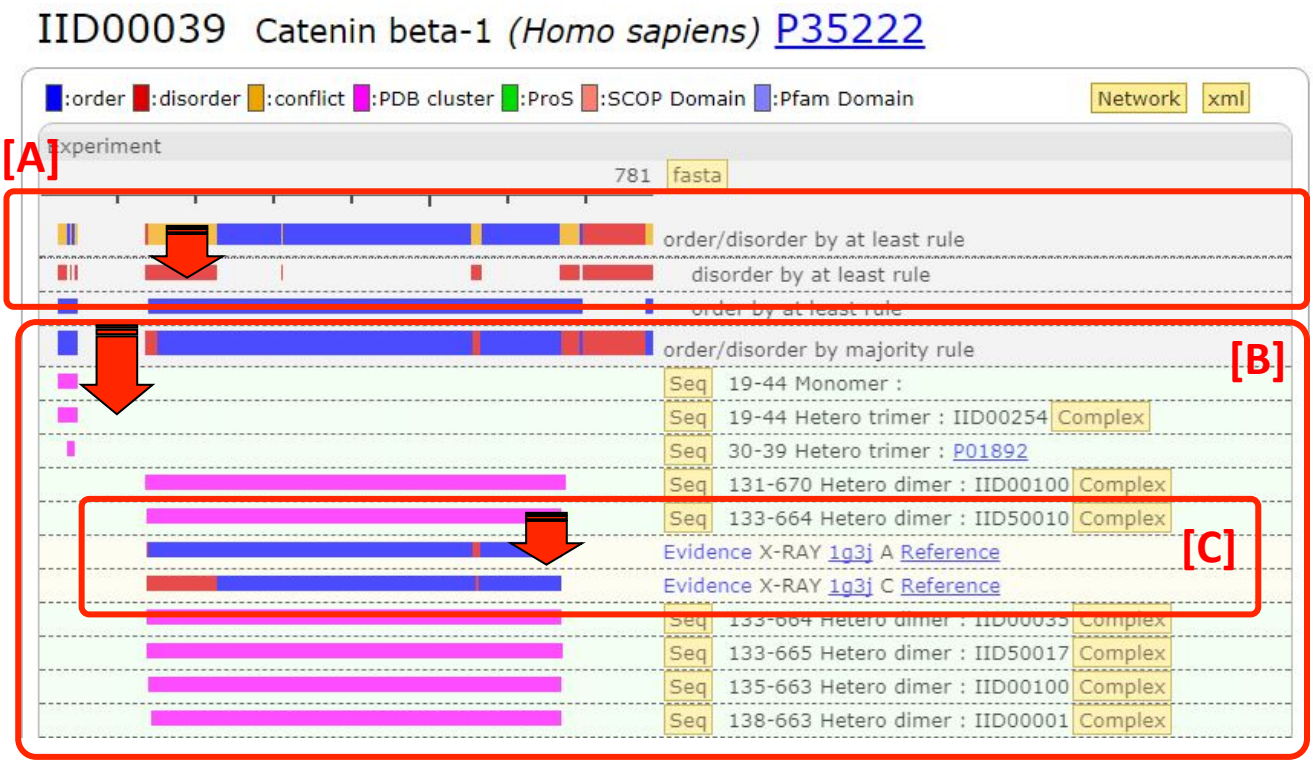


How to see the regions annotated.
Some of the bars in the chart are click-able to show the detailed information.

[A] shows the breakdown of the “at least rule”, which appears by clicking the “at least rule” bar. The break down of “at least rule” includes two bars. The first and the second bars correspond to the “at least” ordered regions, and the “at least” disordered regions, respectively.

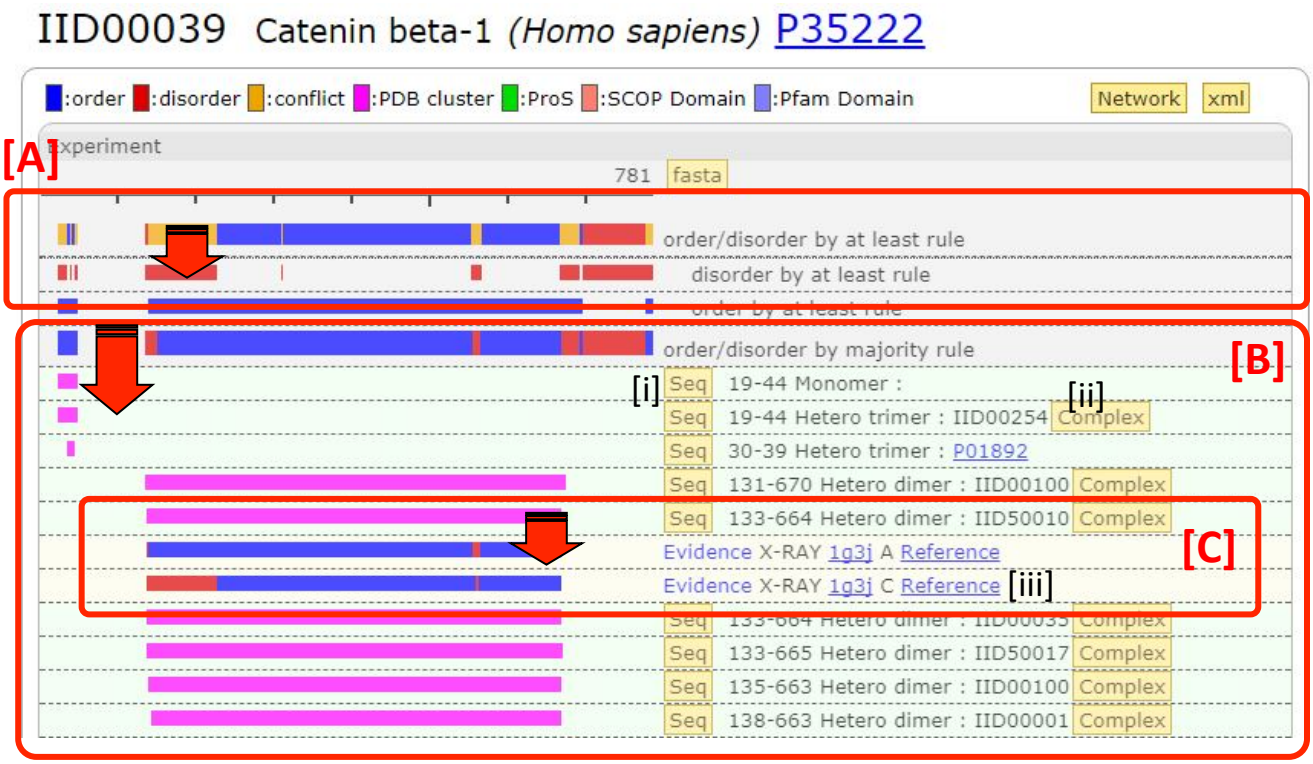
[B] shows the break down of the “majority rule”, where all of the annotated ordered/ disordered regions are presented. PDB entries in this field are clustered, and magenta bars are clickable to present clustered regions [C]. Clustering threshold is described below.

Clustering PDB
We constructed clusters of almost equivalent PDB entries, employing biological units. In the comparison of two complexes, they were firstly divided into subunits. When two subunits (a subunit pair) taken from each complex show more than 70% sequence identity, or their gap sites in the alignment are less than 7, the subunit pair is considered equivalent. Note that the latter condition is applied to compare short segments. When all subunits pairs in two complexes are equivalent, and the interacting-subunit pairs are the same, the complexes are considered equivalent, and should be clustered. Based on this rule, we conducted a single-linkage clustering, and obtained clusters of protein complexes. Monomers were also clustered in the same manner.



Details in the annotation.

[i] The Seq button presents the FASTA formatted sequence high-lighting the corresponding region. Structured/unstructured status, region stat/end, and oligomeric state follow. When binding partners exist, IID or uniprot accession is presented. The Complex button is a link to the EDGE page containing the protein complex of this protein and the partner [ii]. The magenta bar shows detailed annotation clustered [C]. Red and blue represent disordered and ordered regions. “Evidence” shows the experimental data for the annotation together with experimental method, PDB identifiers with chain ID, and “Reference” linking to PubMed [iii].

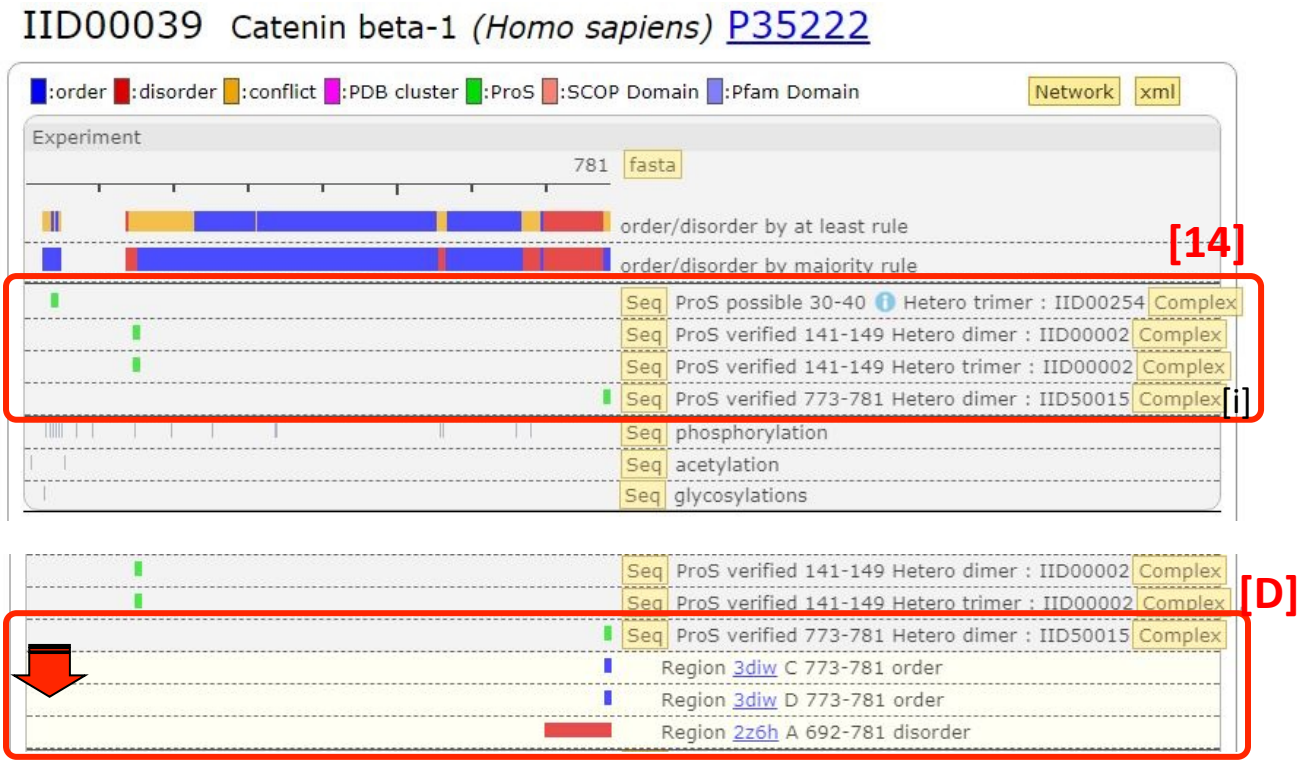


The protean segment (ProS)

The section [14] shows ***protean segments, ProS***. One of the reasons why IDPs have drawn much attention is attributed to the phenomenon so-called coupled folding and binding, where a short flexible segment can bind to its binding partner with forming a specific structure to act as a molecular recognition element. IDEAL explicitly annotates these regions as ***protean segments***.

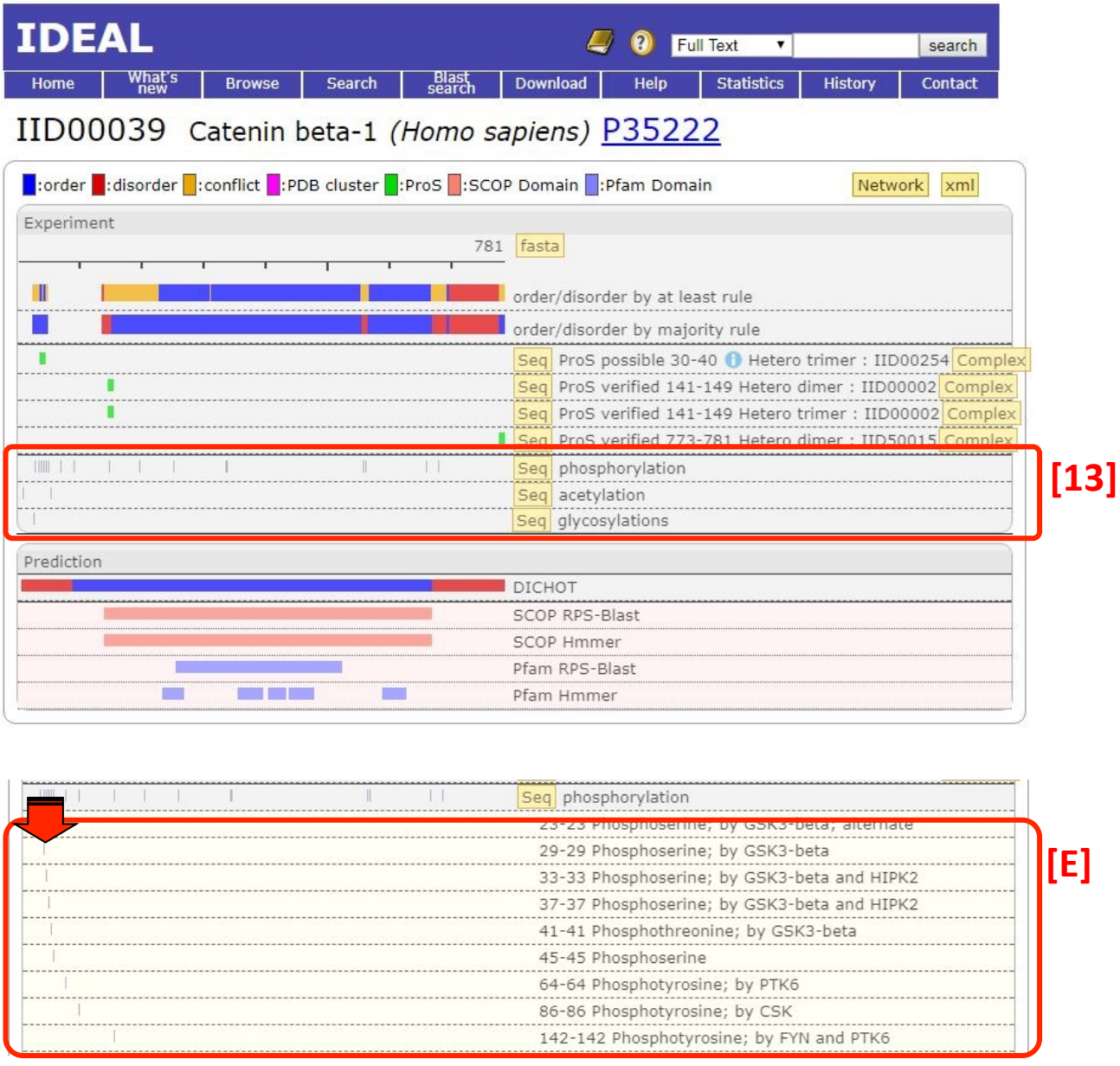
We defined three categories for ProS, verified ProS, possible ProS and predicted ProS. A verified ProS is defined as the sequence, which has both evidences of disordered in an isolated state and of ordered in a binding state with a partner molecule. A possible ProS is defined as the sequence, which only has an evidence of ordered in a binding state, but is thought to be a ProS from circumstantial evidences, for example disorder evidence in homologs, even though it has no evidence of disordered in an isolated state. A predicted ProS is a new category introduced from the IDEAL version 20/Nov/2017. A predicted ProS is defined as the sequence, which only has an evidence of ordered in a binding state, but is thought to be disordered in an isolated state by manual inspections and the results of several disorder-prediction tools (DICHOT, Mobi, P^2D^2 etc.). When the binding partner exists in IDEAL, a link to the EDGE page is presented [i].

The green bars in the ProS section can expand by a click to show the ordered and disordered regions accounting for the ProS [D]. In the case of “possible ProS” and “predicted ProS”, only a ordered region is presented. The clustering results are adopted in the ProS presentation. In this case, C and D chains of PDB, 3diw are clustered, being presented in a single green bar.



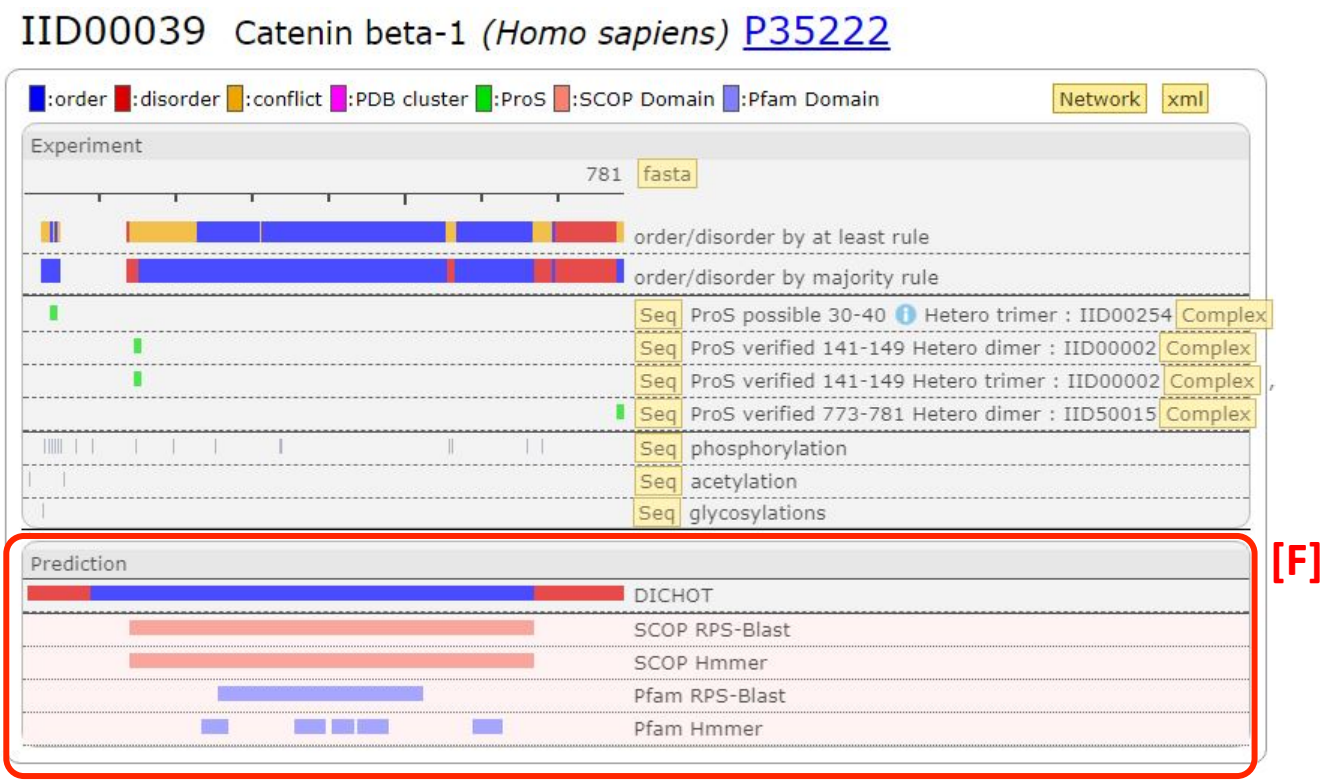
Miscellaneous information from UniProt.

Below the ProS section, miscellaneous information such as modification sites from Uniprot is summarized [13]. Each bar is crick-able to see the details [E].



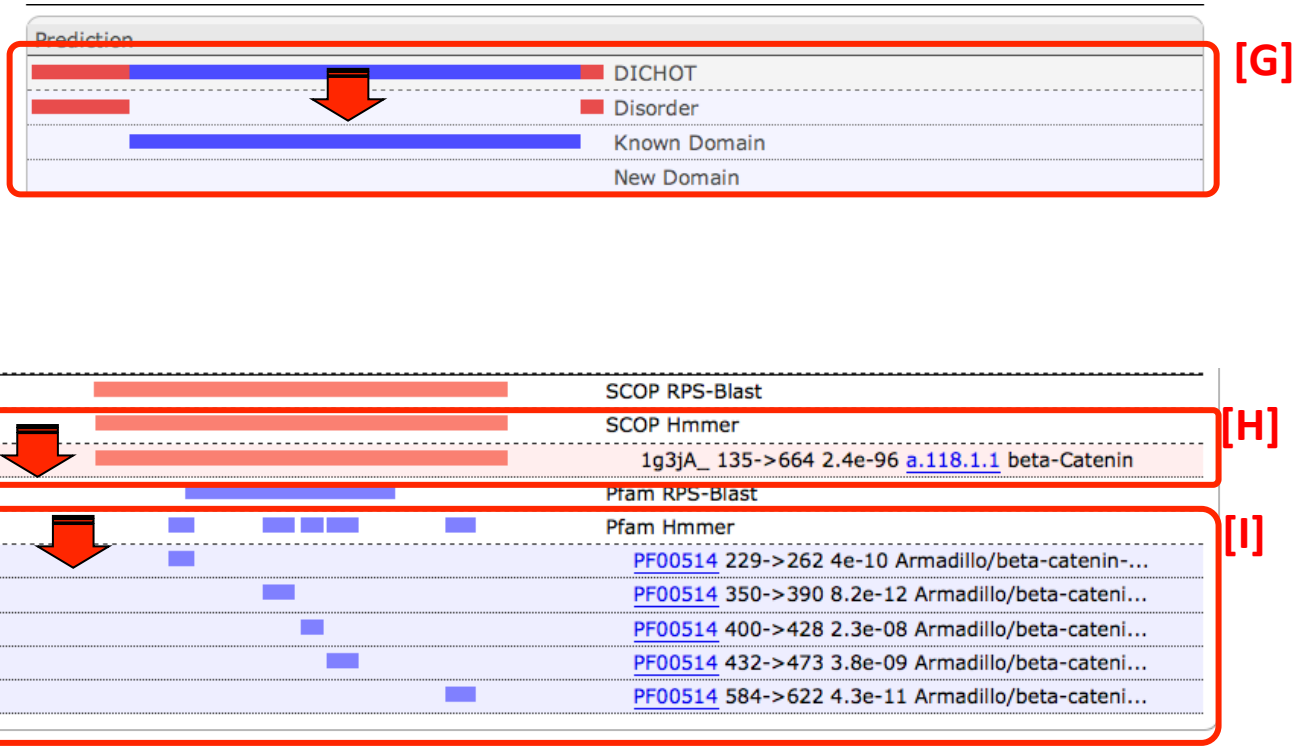
Prediction section.

IDEAL provide the Experiment section and the prediction section. The prediction section presents information based on experimental evidences, whereas the prediction section presents prediction results such as disorder/order prediction by DICHOT, SCOP domain assignments, and Pfam domain assignments [F].



Each bar is a crick-able to show the details. [G] shows DICHOT prediction result. Predicted ordered and disordered regions are presented by blue and red bars. Cryptic domain is domains, which are predicted to be structural domains but their 3D structures have not be known.

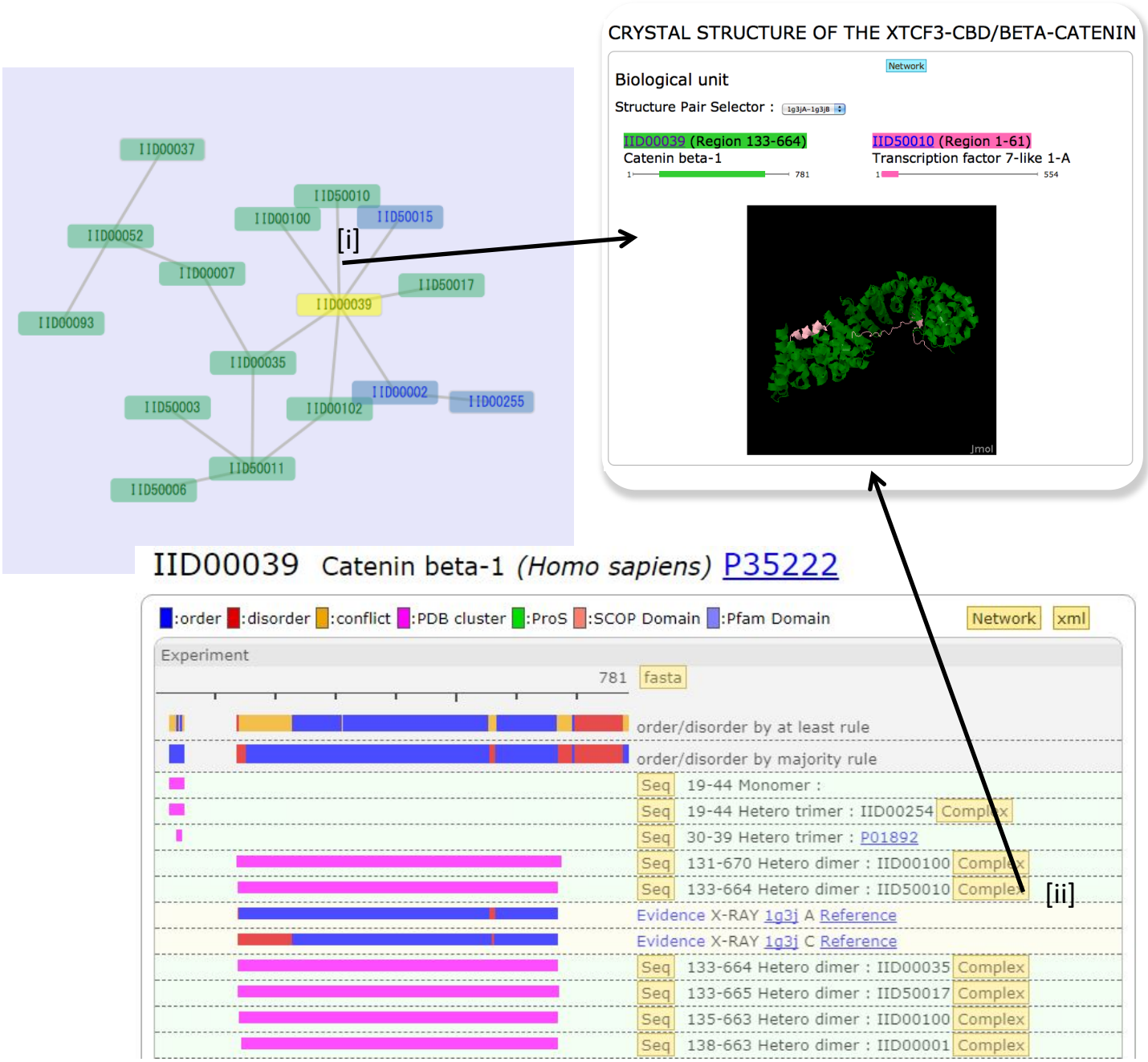
[H] and [I] present SCOP and Pfam prediction by HMM. In the Panel [H], the line shows the domain ID (PDB ID and domain number), the assigned region, the e-value, the SCOP concise classification strings with the link to SCOP, and the description of the domain. In the panel [I], each line shows the Pfam ID with the link to Pfam, the assigned region, the e-value, and the description of the domain.



3) *EDGE page*

How to access EDGE pages

You can access EDGE pages from a edge in a PPI map [i], or an arrow in a NODE page [ii]. Arrows linked to EDGE pages can be found by clicking the majority rule bar or in in the ProS section.



Details in the EDGE pages.

This is an example of the EDGE pages. Edge pages provide structural complexes of an IDEAL entry (NODE) and its binding partner. In this case, the complex of catenin beta-1 and transcription factor 7-like 1-A is shown. The structure is displayed by the J-mol applet [i], in which you can rotate, zoom, and other operations. The cartoon is colored in the same color shown in [J], where the corresponding regions are also presented. [ii] and [iii] are the links to each NODE entry. PDB entries are clustered (see the clustering PDB section), and Structure pair selector [iv] enable one to select a protein complex displayed. By clicking the Network button [v], you can find the PPI network, to which this protein complex belongs.