Short Communication

Complex Portal - A Unifying Protein Complex Database

B. Meldal[1,*], C. Combe[2], H. Bye-A-Jee[1], J. Heimbach[3], M. Koch[1], J. Sullivan[3], Y. Yehudi[3], G. Micklem[3], H. Hermjakob[1], P. Porras[1], S. Orchard[1]

[1] European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambs CB10 1SD, United Kingdom
[2] Wellcome Trust Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, United Kingdom
[3] Cambridge Systems Biology Centre/ Department of Genetics, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, United Kingdom

[*] Correspondence: bmeldal@ebi.ac.uk

SUMMARY

The EBI Complex Portal is a manually curated, unifying resource of macromolecular complexes from model organisms. Each entry has a unique and stable identifier and links participating molecule to their unique reference database (UniProt, ChEBI and RNAcentral). Each complex is annotated with information about their stoichiometry, topology and structural assembly, function, complex-centric Gene Ontology terms and evidence codes. Complexes are extensively cross-referenced to ChEMBL, EMDB, Experimental Factor Ontology, Intenz, MatrixDB, the PDB and Reactome. Bespoke visualisation tools for the general topology and stoichiometry, crystal structures, molecular reactions and gene expression data are provided. All data is open-source and available in PSI-MI xml2.5 and xml3.0 standard formats, MI-JSON and tab-delineated ComplexTAB format.

KEYWORDS

Protein complex database; Protein-protein interactions; Manual curation

BODY

Understanding the composition and role of protein complexes is integral to understanding biological processes but identifying the biologically functional "unit" of proteins and other molecules from raw data such as Proteomics or Transcriptomics data is not trivial. The EBI Complex Portal [1] (www.ebi.ac.uk/intact/complex) provides a unifying, easy-to-search-and-retrieve resource for protein complexes and a custom-made visualisation tool, ComplexViewer [2] (Figure 1 and Figure 2).

The use of the Evidence Code Ontology allows us to indicate for which entries direct experimental evidence is available in molecular interaction databases such as IntAct (www.ebi.ac.uk/intact) or if the complex has been inferred based on homology or background scientific knowledge.

The current curation focus is on model organisms (human, mouse, yeast, *E. coli*, and recently *Arapidopsis thaliana*). While it is difficult to establish our coverage for human due to the presence of many variants of a single complex, for *Saccharomyces cerevisiae* we have now released approximately 80% of known complex with most of the remaining 20% already in the quality control pipeline. For 76% of published entries we have experimental evidence in a protein-protein-interaction database or from crystallisation or electron microscopy data.

Queries can contain single search terms, lists of identifiers or combinatorial searches using our bespoke complex query language (CQL). Standard identifiers such as UniProt, ChEBI, RNAcentral and GO IDs, protein, gene and complex names or synonyms as well as complex cross-references (e.g. PDB IDs) and species IDs and names can be used as query terms. Data is available via our web service or our ftp site in the new MI-JSON format, in the standard interaction formats [3], PSI-MI XML 2.5 & 3.0, and in a bespoke, tab-delimited format, ComplexTAB. Other formats can be made available on request.

Each entry has a unique identifier and contains information about the participating molecules (including small molecules and nucleic acids integral to the complex), their stoichiometry, topology and structural assembly. We provide a function summary and complex-centric Gene Ontology annotations for each complex. Additionally, complexes are extensively cross-referenced to ChEMBL, EMDB, Experimental Factor Ontology (for diseases), Intenz, MatrixDB, the PDB and Reactome. We provide bespoke visualisation tools for crystal structures, reactions and gene expression data.

ComplexViewer is a stand-alone JavaScript tool that displays the complex and its binding features dynamically. It is portable and can be included in other resources (e.g. www.humanmine.org).

User requests and input from the representative model organism communities are welcome.

CONFLICT OF INTEREST

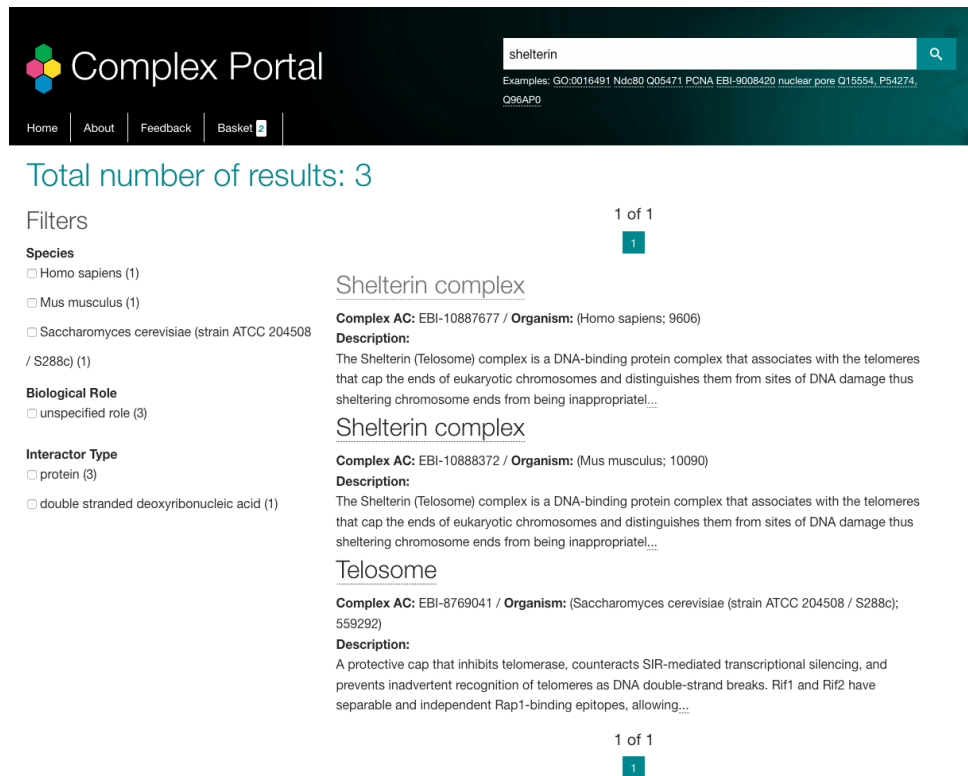The authors declare no conflict of interest.
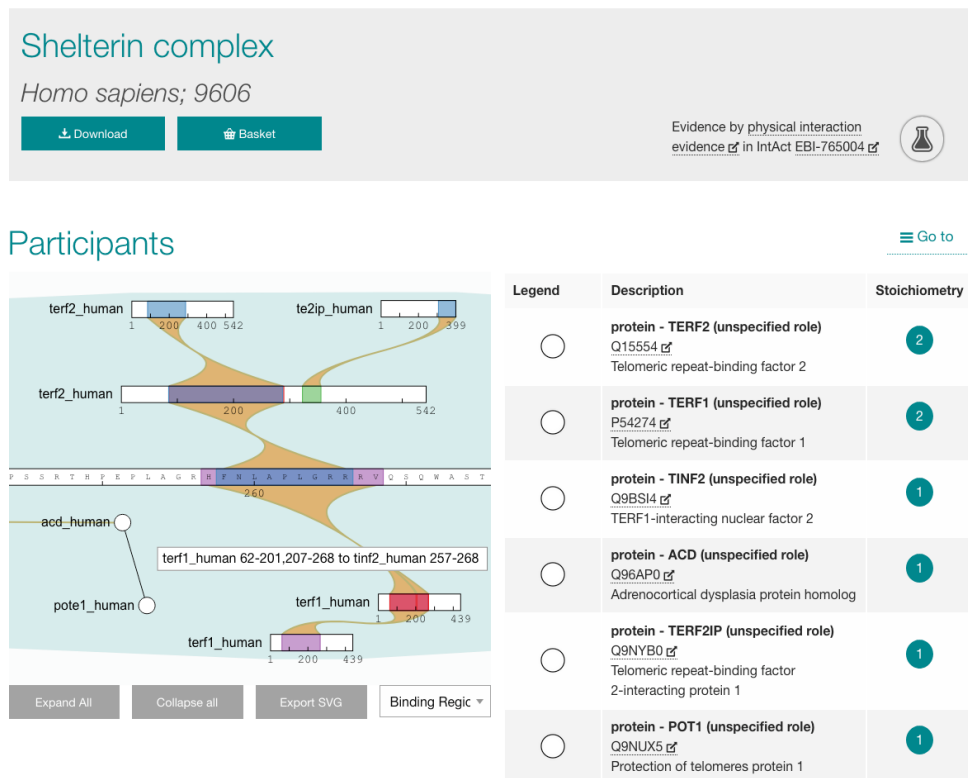
Figure 1: Search results for "shelterin".

Figure 2: visualisation of the human Shelterin complex.

REFERENCES

1. Meldal BHM, Forner-Martinez O, Costanzo MC, Dana J, Demeter J, Dumousseau M, et al. **The complex portal - an encyclopaedia of macromolecular complexes**. Nucleic Acids Research. 2015;43(D1):D479–D484. doi:10.1093/nar/gku975.

2. Combe CW, Sivade MD, Hermjakob H, Heimbach J, Meldal BHM, Micklem G, et al. **ComplexViewer: visualization of curated macromolecular complexes**. Bioinformatics. 2017;33(22):3673–3675. doi:10.1093/bioinformatics/btx497.

3. Kerrien S, Orchard S, Montecchi-Palazzi L, Aranda B, Quinn AF, Vinod N, et al. **Broadening the horizon – level 2.5 of the HUPO-PSI format for molecular interactions**. BMC Biology. 2007 Oct;5(1):44. doi:10.1186/1741-7007-5-44.