BISC: Binary SubComplexes in proteins database

Thomas Juettemann[1,2] and Dietlind L. Gerloff[2,*]

[1]School of Biology, University of Edinburgh, Edinburgh EH9 3JJ, UK and [2]Dept of Biomolecular Engineering, University of California, Santa Cruz, CA 95064, USA

ABSTRACT

Binary subcomplexes in proteins database (BISC) is a new protein–protein interaction (PPI) database linking up the two communities most active in their characterization: structural biology and functional genomics researchers. The BISC resource offers users (i) a structural perspective and related information about binary subcomplexes (i.e. physical direct interactions between proteins) that are either structurally characterized or modellable entries in the main functional genomics PPI databases BioGRID, IntAct and HPRD; (ii) selected web services to further investigate the validity of postulated PPI by inspection of their hypothetical modelled interfaces. Among other uses we envision that this resource can help identify possible false positive PPI in current database records. BISC is freely available at http://bisc.cse.ucsc.edu.

INTRODUCTION

Structural biological techniques have advanced in recent years and are able to resolve the quaternary structures of increasingly large, transient or permanent, protein complexes. Functional genomics experiments to investigate protein–protein interaction (PPI) have diversified and new technologies are emerging rapidly. Most laboratories submit their findings to literature and publicly available databases. While integration of such repositories is urgently needed and improving, different criteria for data selection and curation can be of interest to specialist users and a large number of different PPI databases will prevail [for a recent review, see Ref. (1)].

Here, we introduce a new database resource: BInary SubComplexes in proteins (BISC). BISC allows users to explore known and modellable subcomplex (SC) structures in current functional genomics PPI databases in user-friendly manner. BISC is organized in three main sections:

Characterized subcomplexes (BISCHom + BISCHet)

This section contains binary substructures extracted from crystallographically determined structures in the Protein Data Bank (PDB) (2) and the predicted biological assemblies based on their coordinates by PISA (3). Homodimeric and heterodimeric interfaces are separated for further analysis, since the latter are evolutionarily more constrained with respect to mutations, due to their symmetry. Users can search by keyword or sequence similarity (with either of two suspected interaction partners or both) (Figure 1A) to retrieve an informative BISC page for each SC of interest (Figure 1B). Alongside fully interactive, embedded Jmol (4) displays emphasising the interface, structural information and links are provided, e.g. SCOP (5) classification of both partners; interface size and residues [the buried surface area computed by NACCESS (6)] and energy scores by PISA (3). Importantly the SCs can serve as potential template structures for modelling protein complexes by homology (see below).

Modellable interactions in functional genomics databases (BISC-MI)

BISC-MI features a list of modellable interactions (MIs) in three of the most widely used databases of experimentally reported PPI (1): the BioGRID (7), IntAct (8) and the Human Protein Reference Database (HPRD) (9). Modellable PPI share sequence similarity between both partners and a template structure in BISCHom or BISCHet (Figure 1C). Users can generate protein structural models dynamically, by an automated procedure using the program MODELLER (10). Multiple sequence alignments (MSAs) for each partner family are generated beforehand because MSA-based modelling generally delivers better models than pairwise alignment procedures. A link to the output page (Figure 1D) is returned by email reply. It provides embedded interactive displays of the two protein family MSAs [using Jalview (11)] and the model as well as a link for downloading the atomic coordinates. Athough it is important to stress the speculative and unproven nature of any such model, it can often serve as a valuable starting point for further validation.

Data and tool linkouts for validative analyses using structural bioinformatics

The user may want to inspect the surface properties of the template and/or modelled SCs, e.g. electrostatic charge and
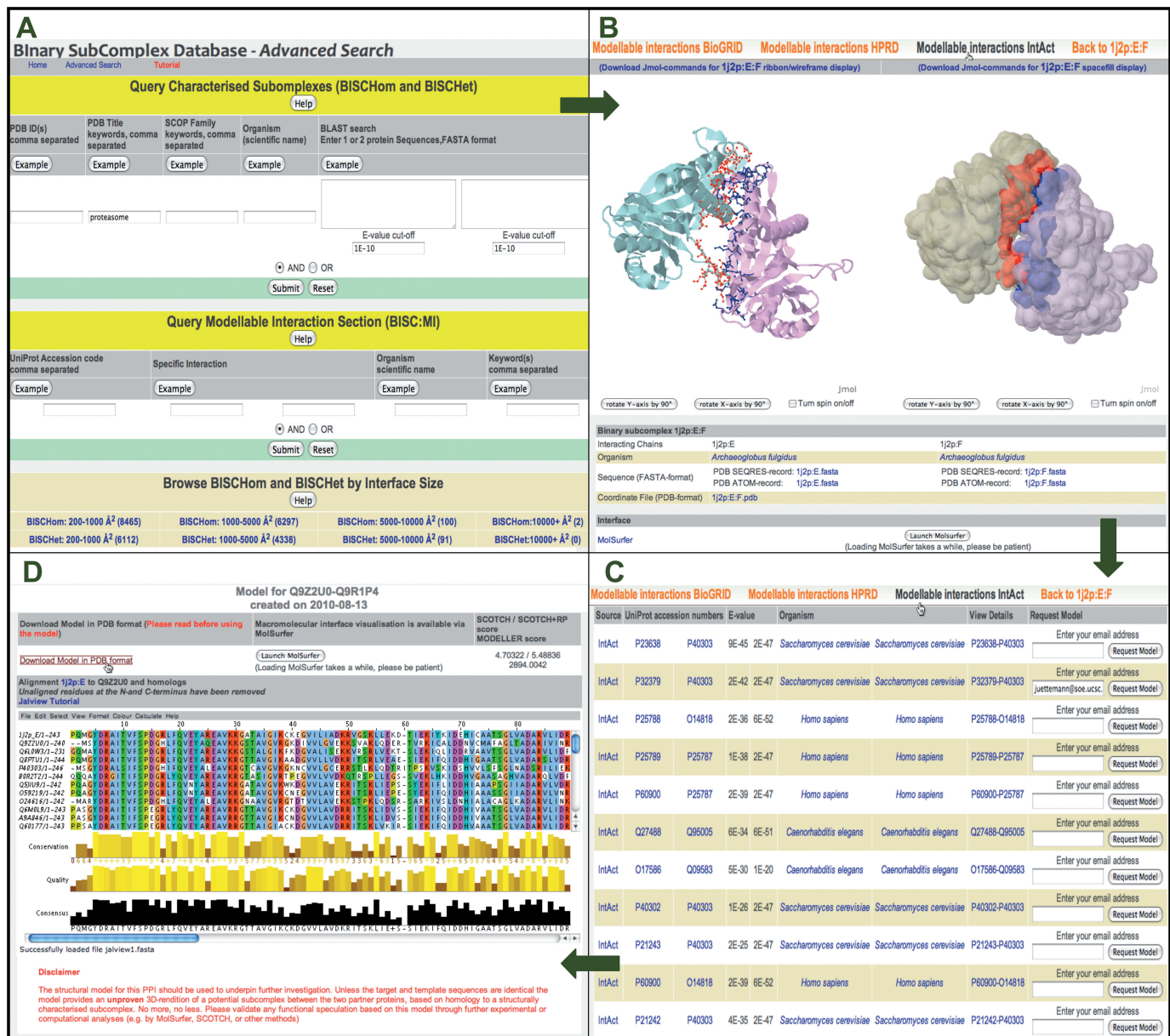

Figure 1. Screenshots of selected BISC output pages. (**A**) In this example, a user searched the Characterized SC Section with the keyword 'prote-asome'. (**B**) An excerpt of the output page for one of the binary SCs of the known proteasome α-ring structure from the archaeon *Archaeoglobus fulgidus* (PDB:1J2P) is shown to illustrate BISC page layout, Jmol (4) displays and options. Links at the top of the page indicate that MIs are available (in the BISC-MI Section) in all three functional genomics PPI databases, the BioGRID (7), HPRD (9) and IntAct (8). (**C**) Models of the hypothetical SCs relating to these PPI (in various species) can be requested by a web form. (**D**) An excerpt of the results page for one of these requests. Validative scores from MODELLER (10) and SCOTCH (13) are displayed at the top of the page, MolSurfer (12) can be launched to compute compatibility scores ECC and HCC. One of the family alignments for each interaction partner is visible in the interactive Jalview (11) display, scrolling down would reveal the other.

hydrophobicity. Validation programs can help evaluate the feasibility of a hypothetical PPI, e.g. by evaluating complementarity of the two interacting protein surfaces. However, some require specialized input files [e.g. the web/Java tool MolSurfer (12) and the valid-ation program SCOTCH (13)] or produce output too difficult to interpret for non-experts [e.g. PISA (3)]. Information from these programs are easily obtained from BISC pages.

Currently BISC is fully updated three times each year. Help documentation and a tutorial example are available at the BISC web site.

DATA SOURCES AND SELECTION

Figure 2 shows an overview schematic of how BISC content is selected. Below we provide further
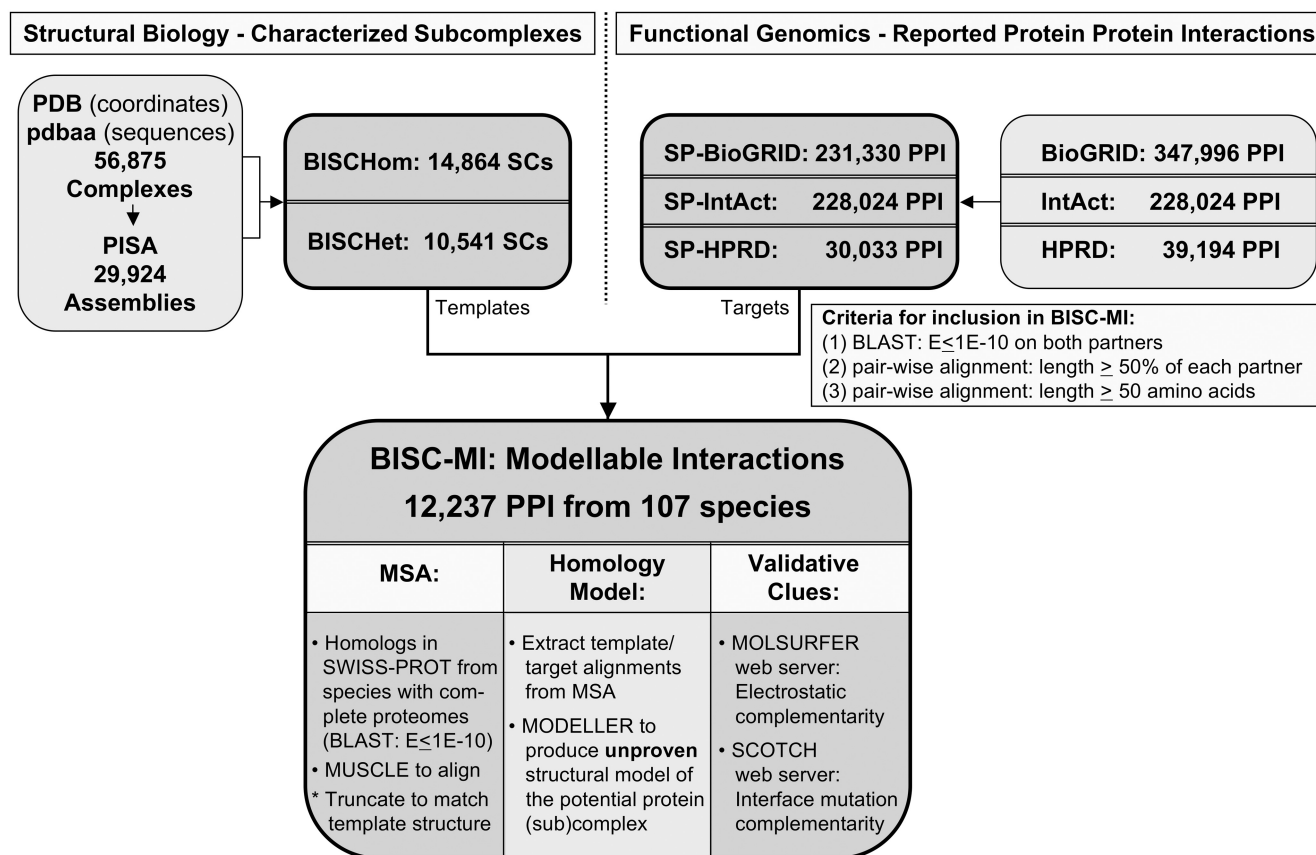
Structural Biology - Characterized Subcomplexes

Functional Genomics - Reported Protein Protein Interactions

PDB (coordinates)
pdbaa (sequences)
56,875
Complexes
↓
PISA
29,924
Assemblies

BISCHom: 14,864 SCs

BISCHet: 10,541 SCs

SP-BioGRID: 231,330 PPI

SP-IntAct: 228,024 PPI

SP-HPRD: 30,033 PPI

BioGRID: 347,996 PPI

IntAct: 228,024 PPI

HPRD: 39,194 PPI

Templates

Targets

Criteria for inclusion in BISC-MI:
(1) BLAST: $E \leq 1E-10$ on both partners
(2) pair-wise alignment: length $\geq$ 50% of each partner
(3) pair-wise alignment: length $\geq$ 50 amino acids

BISC-MI: Modellable Interactions

12,237 PPI from 107 species

MSA:	**Homology Model:**	**Validative Clues:**
• Homologs in SWISS-PROT from species with complete proteomes (BLAST: $E \leq 1E-10$) • MUSCLE to align * Truncate to match template structure	• Extract template/target alignments from MSA • MODELLER to produce **unproven** structural model of the potential protein (sub)complex	• MOLSURFER web server: Electrostatic complementarity • SCOTCH web server: Interface mutation complementarity

Figure 2. BISC design and content in schematic overview. The main data sources for the two BISC database sections described in the text [BISC(Hom + Het) and BISC-MI] are shown with the numbers of records that were processed and collated in the most recent update (October 2010; interface sizes >200 Å^2). PISA, protein interfaces, surfaces and assemblies Protein Interfaces Surfaces and Assemblies (3); SP, SWISS-PROT (15); BioGRID (7); IntAct (8); HPRD, Human Protein Interaction Database (9). For inclusion in BISC-MI section, all selection criteria must be met.

implementation details. For technically interested users additional method diagrams within the BISC online documentation offer further detail on depicting data extraction from PDB and PISA and the BISC-MI modelling pipeline.

Characterized subcomplexes (BISCHom + BISCHet)

BISCHom and BISCHet are generated by extracting binary SCs from crystallographic protein structures in the PDB (2) (Figure 2, top left). Generally we retain PPI between partner proteins from the same species. To capture all plausible binary interfaces (even those not present in the PDB co-ordinate file, for example PDB:1JL5 contains only one subunit of the homo-tetrameric protein complex), we run PISA (3) on the set and extract the top-ranked predicted stable assembly for each structure as an alternative source of SCs. Interfaces present in both PDB and PISA assemblies are identified by cross-matching their ATOM records because chain identifiers may differ. If the top-ranked PISA assembly contains more protein subunits than the PDB record it is carried forward. Next, multi-protein complexes are carried forward (separately for PDB and PISA-derived complexes) if several partner proteins are longer than 40 amino acids. Buried surface area is calculated by NACCESS (6) to identify direct interfaces (values listed

are half the total buried area on both subunits). All interacting subunit pairs are extracted originally. However, we only return SCs with buried surface area >200 A^2 in standard uses of BISC (searches and browsing), to prevent that non-expert users are misled by peripheral contacts. Homodimeric and heterodimeric SCs are then separated to produce BISCHom and BISCHet, respectively, after the sets are redundancy filtered to <95% pairwise identity using PISCES scripts (14) on the corresponding pdbaa sequence file (this is provided by the Dunbrack group for weekly PDB updates). To prevent that non-identical interfaces in homo-oligomeric structures (from PDB and PISA) are removed from BISCHom, all interfaces are checked for size differences before PISCES. A list of all PISCES-eliminated structures is retained and included in keyword and PDB-ID searches by default.

Modellable interactions (BISC-MI)

For BISC-MI, the data source are PPI records from the three functional genomics databases: the BioGRID (7), IntAct (8) and HPRD (9) (Figure 2, top right). Currently only PPI with SWISS-PROT (SP) accession codes among the list of protein identifiers are carried forward. A postulated PPI is listed in BISC-MI if (i) both partners elicit E-values $<10^{-10}$ to different chains

in the same SC in BISCHom or BISCHet (which would be used as a template structure for modelling); (ii) the BLAST local alignments with this template SC span at least 50% of each of its partner sequences; and (iii) are at least 50 amino acids long. MIs are reported in separate tables depending on their PPI-database source. Systematic species names are obtained from the National Center for Biotechnology Information (NCBI) via the TaxID annotation in their SP sequence records.

Model building

Modelling requests from BISC-MI are processed by an automated template-based modelling (homology modelling) protocol using the popular MODELLER program (10) (Figure 2). To ensure that the two required pairwise alignments (between the target and template proteins) are as accurate as possible, each is based on a MSA. Homologues are collected by BLASTing each target sequence against species with complete proteomes in SP (15) and including the most similar protein in each species with E-values $\leq 10^{-10}$. Both partner families are filtered so they include the same species set (this is so that future implementations can model all likely orthologues simultaneously). The two MSAs are generated, and the respective template sequences added, with MUSCLE (16) using the alignment and sequence-to-profile alignment options, respectively. Aligning full-length protein sequences with the often shorter template sequences automatically can produce unsatisfactory results with any method. To pre-empt sub-standard models being derived in such instances, we apply empirical criteria to either generate a warning (but still return a model) or abort the process. These criteria look for excessive gapping in the MSA and length differences; their specifics are documented at the BISC web site. Finally the requested co-ordinate model is produced by MODELLER through a standard multi-chain template protocol and returned by email. To stress the speculative and unproven nature of any such model our results pages carry a warning to reflect this. By offering MSA and model production as a service, rather than a pre-computed section within BISC, we account for the possibility that additional homologues become available in between regular BISC updates. However, all request results are stored. If a request is made for a model that was produced previously, the user can either produce a new model or gain immediate access to a previously produced MI results page.

Validative interface scores and methods

It is currently impossible to discern true PPI from artefactual PPI with certainty, computationally, even if a crystal structure is available. However, BISC provides interface scores and methodology that should be helpful in this regard, e.g. by flagging outliers. For PPI in the Characterized SC Section, BISC provides information through: direct contact in a crystal structure; inclusion (or not) in the PDB or PISAs top-ranked assembly or both; PISA-interface scores; a MolSurfer launch option to derive electrostatic correlation coefficients (ECCs) and

hydrophobicity correlation coefficients (HCC); see below. For a hypothetical PPI in the MI Section of BISC, validative clues come through: experiment (the basis for its record in a functional genomics database); MODELLER objective function score; SCOTCH and SCOTCH + RP scores; MolSurfer launch (ECC and HCC). Among these, the SCOTCH and MolSurfer web servers are less well-known but helpful (see also the usage example, below). MolSurfer (12) calculates the electrostatic surface potential of each partner and projects the result onto a planar interpolation of the interface between the two protein subunits forming a binary complex. This enables computing a correlation coefficient score (ECC). A HCC is also returned. By contrast the SCOTCH (13) web server uses a given SC structure to discern which amino acids are located at the interface and investigates statistical correlation between the mutations found at relevant alignment positions within the family. SCOTCH evaluates interface complementarity based on standardized amino acid properties of the residues that come in proximity of one another at the potential interface. Neither MolSurfer nor SCOTCH depend on precise atom positions for a rudimentary assessment of the feasibility of a binary complex. Our threshold value recommendations for their scores are based on simple calibration experiments and experience (Quan, X. and Gerloff, D.L., unpublished data); R. Guerois, personal communication).

CURRENT CONTENT (OCTOBER 2010 RELEASE)

Currently the core of BISCHom (SCs with interface sizes $>200\,\text{Å}^2$) contains 14 864 records relating to 11 636 unique sequences and 11 433 PDB records (homooligomeric complexes sometimes contain several interfaces on the same subunit; we record these separately); 1423 species (TaxIDs) are represented. BISCHet core contains 10 541 entries from 3060 PDB-records; 446 species are represented. The current BISC-MI fragments resulted from a screen of 231 330 records with SP accession codes in the BioGRID; 228 024 in IntAct; and 30 033 in HPRD. A total of 12 237 MI with interface sizes $>200\,\text{Å}^2$ were found from 107 species (BioGRID: 3518 MI; IntAct: 4853 MI; HPRD: 4949 MI).

Version numbers or download dates for all third-party software and data sets other than the PDB, used in the most recent update of BISC (October 2010) are listed below. PISA 01 May 2009; BioGRID 3.0.67; IntAct 28 Jul 2010; HPRD Release 9; BLAST 2.2.19; PISCES Dec 142005; NACCESS 2.1.1; PDB2PQR 1.6.0 (input files for MolSurfer); MUSCLE 3.7; MODELLER 9.6; MolSurfer web server http://projects.villa-bosch.de/dbase/molsurfer/submit-elec.html; SCOTCH web server http://biodev.extra.cea.fr/scotch/; Jmol 11.6.24; JalView 2.4.

USAGE EXAMPLE

BISC content can be exploited in systematic searches, for example for possible false positive PPI in literature reports

(Juettemann, T. and Gerloff, D.L., manuscript in preparation). In contrast, the benefits of BISC as an interactive resource are best illustrated through an example. In this hypothetical scenario a user is interested in PPI within the proteasome. For the sake of argument we will assume that the crystal structure of an eukaryotic proteasome is still not available at the time of investigation. [The first eukaryotic proteasome structure was deposited in the PDB in April 1998 (PDB:1RYP).] More detailed documentation of this example and additional figures relating to the first part can be found in the online tutorial at the BISC web site.

Using either the keyword or BLAST search option, the user found characterized SCs in BISC. These were derived from archaeal proteasome structures, e.g. from *Thermoplasma acidophilum*. [The first *T. acidophilum* proteasome structure was deposited in the PDB in June 1996 (PDB:1PMA).] Of the 28 protein subunits making up the core particle structure, only one heterodimeric SC and one homodimeric SC of structure 1PMA, were present in BISC. This is because the particle is highly sequence redundant. Its barrel shape is built of four heptameric rings with 14 identical subunits making up the inner rings and the 14 other subunits making up the outer rings.

Linking to the corresponding two SC pages in BISC (1PMA:I:J and 1PMA:Y:Z) revealed that several MIs were recorded in the functional genomics PPI databases. The BISC-MI lists made apparent that many MIs are eukaryotic, most are from budding yeast (*Saccharomyces cerevisiae*). Unexpectedly, the user encountered multiplicity compared with the archaeal structure—13 unique yeast interactions were modellable onto the template 1PMA:I:J and 6 onto 1PMA:Y:Z. The user noted that a plausible explanation for this observation is that, probably, eukaryotic proteasome particles are composed of chains that are similar in structure but different in sequence. While this fact would have been known by proteasome experts before the first eukaryote structure appeared, this scenario illustrates how browsing a resource like BISC can provide interesting clues to a novice.

Next, hypothetical structures were requested by the user for all *S. cerevisiae* PPIs using the BISC modelling pipeline. Links to the result pages were returned by email and reported various third-party scores for each model (Figure 3A; see references and BISC documentation for more specific information on each score). Additionally, the result web sites displayed the MSAs used to create each model using Jalview (Figure 3B) and the model using Jmol (not shown).

Given the similar degrees of sequence identity (Seq id) of all target sequences to the template, all models were expected to be of similar quality. However, when examining the SCOTCH + RP scores and electrostatic complementarity using MolSurfer (ECC), one notices that the last modelled PPI in the list (P40302:P32379, both are component proteins of the yeast proteasome according to SP) obtained exceptionally unfavourable scores compared with the others: SCOTCH 7.33; SCOTCH + RP 8.48; ECC 0.003 (Figure 3A). A look at the multiple sequence alignment created during the modelling process for target sequence P32379) points to a possible explanation. P32379 carries an insertion (sequence: SGEERLM), which was placed at positions 119–125. Upon visual inspection, the location of the insertion within the alignment seems adequate. However, MODELLER created novel electrostatically unfavourable atomic interactions in this hypothetical model to accommodate the loop, which impacted on the deviating scores (albeit not that provided by MODELLER itself).

It is known, meanwhile, that these two protein subunit form a direct physical PPI in the yeast proteasome particle. Thus it would have been incorrect to dismiss the postulated PPI solely due to the deviating scores. However, the example illustrates how exploring unusual or aberrant scores in BISC can uncover unusual features in PPIs such as the inserted loop. To follow up on such clues more thoroughly, expert users can download all files created during the modelling process, e.g. to refine models using a local MODELLER installation or to run additional validative methods.

DISCUSSION AND FUTURE DIRECTIONS

We will soon make it possible for BISC users to request models for all likely orthologues within the target families, not only the specific target PPI; this should enable additional research into the feasibility of a potential PPI. We welcome feedback regarding further existing methods that would be helpful for validating the approximate interfaces in our models, and additional PPI databases to include.

Previously published resources exist that feature similar content as the Characterized SC Section in BISC, e.g. PROTCOM (17), PISA (3) and some of the many databases focusing on known interfaces (18–23). However, to the best of our knowledge, BISC is currently unique in that it makes extensive use of PISA-predicted assemblies alongside those delivered in the downloadable co-ordinate files from the PDB. Thereby BISC provides additional binary SCs that other databases (e.g. PROTCOM) would miss. The recently developed IBIS resource (24) also used structurally characterized PPI as template structures and infers PPI between closely homologous partner pairs if such inference is supported by a set of verification criteria. Some other published databases and servers offer modelling-by-request services as we do within the Modellable Interaction Section in BISC, e.g. MODBASE (25) and SWISS-MODEL DB (26), although these resources do not specialize in modelling binary complexes and do not provide additional validative services besides modelling scores.

The most important practical distinction between BISC-MI and other databases is its emphasis on PPI suggested by functional genomics experiments. By offering structural bioinformatics follow-up in an easy-to-use resource BISC will help ensure that probable false-positive laboratory results are identified, reinvestigated and corrected. Conversely, existing bioinformatics methodology attempting to validate hypothetical complexes is often sufficient to provide informative clues, but insufficient in most cases to discriminate
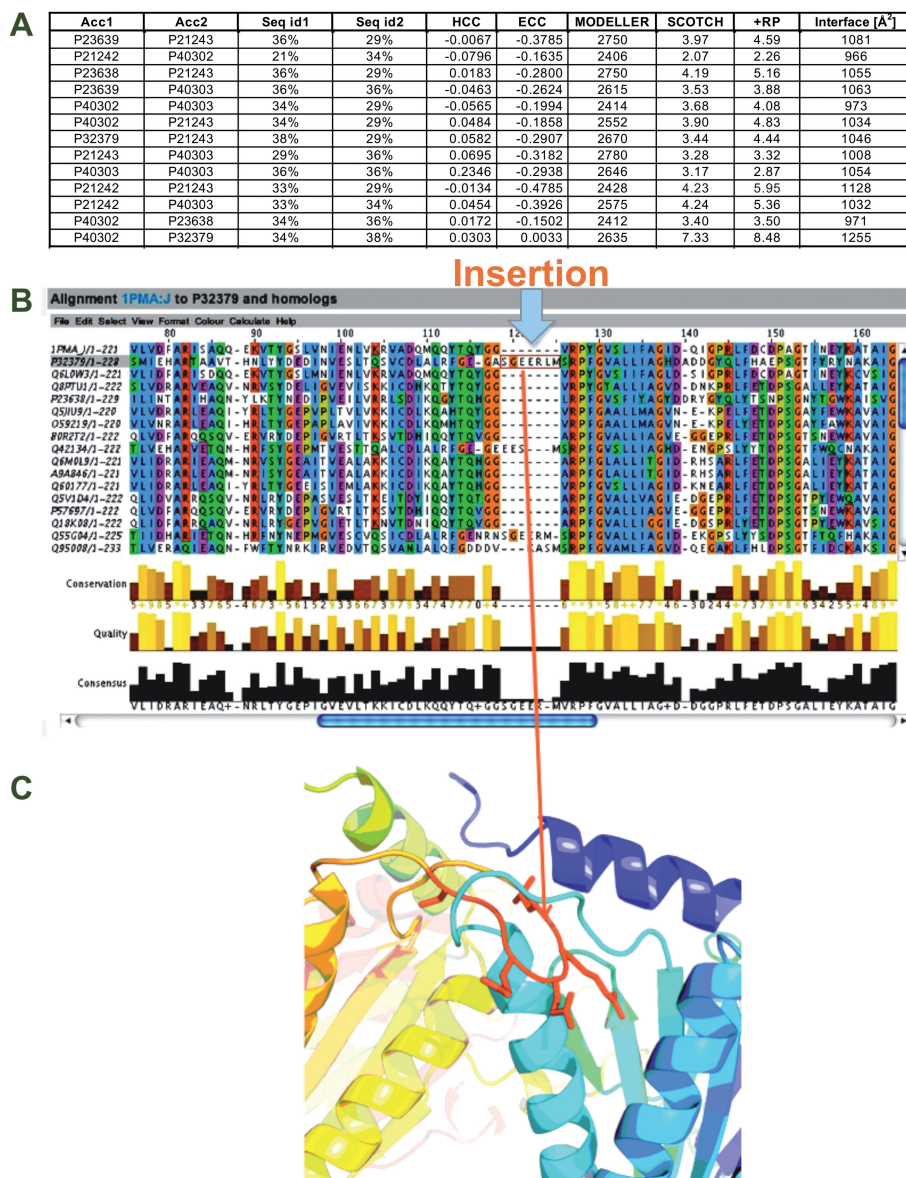
Figure 3. Usage scenario: proteasome. (**A**) Validative scores returned for each modellable hypothetical SC in the *S. cerevisiae* proteasome (see text); Acc, accession; sequence identity (with template 1PMA:I:J); HCC, MolSurfer hydrophobicity correlation coefficient (12); ECC, MolSurfer electrostatic correlation Seq id coefficient (12); MODELLER, objective score function (10); SCOTCH score (13); +RP, SCOTCH score with RP (13). (**B**) An excerpt of the results page returned by the BISC-MI modelling pipeline for hypothetical PPI P40302:P32379 highlighting the insertion in the multiple sequence alignment for P32379. (**C**) The position of the corresponding loop segment (red) in the MODELLER (10) model is shown in context of the surrounding protein structure.

between interacting and non-interacting hypothetical complexes on its own. Recently the database developers of the new genome-wide docking database (GWIDD) (27) have also linked up structural bioinformatics and functional PPI databases. Since GWIDD and BISC differ in their selection of data sources and structural bioinformatics techniques they are currently different in content and could be integrated in the future. Like everyone in the field, BISC will greatly benefit from further standardization and consistency in PPI-annotation and curation. This is driven forward by several professional organizations, such as the proteomics standards initiative (PSI) by the Human Proteome Organization (HUPO) (28) and/or the IMEx (29) and MIMIx standards (30).

AVAILABILITY

Access to BISC and the services it provides is free to anyone at http://bisc.cse.ucsc.edu, and does not require registration or login. Some of the optional requests deliver the link to a results page by email and thus require a valid email address. Results are readily viewable in any web browser with enabled Java. Rare minor differences are attributable to bugs in the browser

software (e.g. Mozilla Firefox on Mac OS X) and are easily resolved (e.g. by clicking onto the scroll bar to complete launching the molecular viewer Jmol). Users interested in downloading the entire database please contact the authors.

REFERENCES

1. Ooi,H.S., Schneider,G., Chan,Y.-L., Lim,T.-T., Eisenhaber,B. and Eisenhaber,F. (2010) Databases of Protein-Protein Interactions and Complexes. *Methods Mol. Biol.*, **609**, 145–159.
2. Berman,H., Henrick,K., Nakamura,H. and Markley,J.L. (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.*, **35**, D301–D303.
3. Krissinel,E. and Henrick,K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **372**, 774–797.
4. Jmol: an open-source Java viewer for chemical structures in 3D. http://www.jmol.org/ (1 February 2009, date last accessed).
5. Andreeva,A., Howorth,D., Chandonia,J.M., Brenner,S.E., Hubbard,T.J., Chothia,C. and Murzin,A.G. (2008) Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res.*, **36**, D419–D425.
6. Hubbard,S.J. and Thornton,J.M. (1993) *'NACCESS' Computer Program*. Department of Biochemistry and Molecular Biology, University College London.
7. Breitkreutz,B.J., Stark,C., Reguly,T., Boucher,L., Breitkreutz,A., Livstone,M., Oughtred,R., Lackner,D.H., Bähler,J., Wood,V. *et al*. (2008) The BioGRID Interaction Database: 2008 update. *Nucleic Acids Res.*, **36**, D637–D640.
8. Aranda,B., Achuthan,P., Alam-Faruque,Y., Armean,I., Bridge,A., Derow,C., Feuermann,M., Ghanbarian,A.T., Kerrien,S., Khadake,J. *et al*. (2010) The IntAct molecular interaction database in 2010. *Nucleic Acids Res.*, **38**, D525–D531.
9. Keshava Prasad,T.S., Goel,R., Kandasamy,K., Keerthikumar,S., Kumar,S., Mathivanan,S., Telikicherla,D., Raju,R., Shafreen,B., Venugopal,A. *et al*. (2009) Human Protein Reference Database—2009 update. *Nucleic Acids Res.*, **37**, D767–DD772.
10. Sali,A. and Blundell,T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol Biol.*, **3**, 779–815.
11. Waterhouse,A.M., Procter,J.B., Martin,D.M., Clamp,M. and Barton,G.J. (2009) Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.
12. Gabdoulline,R.R., Wade,R.C. and Walther,D. (2003) MolSurfer: A macromolecular interface navigator. *Nucleic Acids Res.*, **31**, 3349–3351.
13. Madaoui,H. and Guerois,R. (2008) Coevolution at protein complex interfaces can be detected by the complementarity trace with important impact for predictive docking. *Proc. Natl Acad. Sci. USA*, **105**, 7708–7713.
14. Wang,G. and Dunbrack,R.L. Jr (2005) PISCES: recent improvements to a PDB sequence culling server. *Nucleic Acids Res.*, **33**, W94–W98.
15. The UniProt Consortium. (2009) The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res.*, **37**, D169–D174.
16. Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
17. Kundrotas,P.J. and Alexov,E. (2007) PROTCOM: searchable database of protein complexes enhanced with domain-domain structures. *Nucleic Acids Res.*, **35**, D575–D579.
18. Davis,F.P. and Sali,A. (2005) PIBASE: a comprehensive database of structurally defined protein interfaces. *Bioinformatics*, **21**, 1901–1907.
19. Finn,R.D., Marshall,M. and Bateman,A. (2005) iPfam: visualization of protein–protein interactions in PDB at domain and amino acid resolutions. *Bioinformatics*, **21**, 410–412.
20. Winter,C., Henschel,A., Kim,W.K. and Schroeder,M. (2006) SCOPPI: a structural classification of protein-protein interfaces. *Nucleic Acids Res.*, **34**, D310–D314.
21. Jefferson,E.R., Walsh,T.P., Roberts,T.J. and Barton,G.J. (2007) SNAPPI-DB: a database and API of Structures, iNterfaces and Alignments for Protein-Protein Interactions. *Nucleic Acids Res.*, **35**, D580–D589.
22. Bordner,A.J. and Gorin,AA. (2008) Comprehensive inventory of protein complexes in the Protein Data Bank from consistent classification of interfaces. *BMC Bioinformatics*, **9**, 234.
23. Günther,S., von Eichborn,J., May,P. and Preissner,R. (2009) JAIL: a structure-based interface library for macromolecules. *Nucleic Acids Res.*, **37**, D338–D341.
24. Shoemaker,B.A., Zhang,D., Thangudu,R.R., Tyagi,M., Fong,J.H., Marchler-Bauer,A., Bryant,S.H., Madej,T. and Panchenko,A.R. (2010) Inferred Biomolecular Interaction Server (IBIS)—a web server to analyze and predict protein interacting partners and binding sites. *Nucleic Acids Res.*, **38**, D518–D524.
25. Pieper,U., Eswar,N., Webb,B.M., Eramian,D., Kelly,L., Barkan,D.T., Carter,H., Mankoo,P., Karchin,R., Marti-Renom,M.A. *et al*. (2006) MODBASE: a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.*, **34**, D291–D295.
26. Kiefer,F., Arnold,K., Kunzli,M., Bordoli,L. and Schwede,T. (2009) The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.*, **37**, D387–D392.
27. Kundrotas,P.J., Zhu,Z. and Vakser,I.A. (2010) GWIDD: Genome-wide protein docking database. *Nucleic Acids Res.*, **38**, D513–D517.
28. Orchard,S., Albar,J.P., Deutsch,E.W., Binz,P.A., Jones,A.R., Creasy,D. and Hermjakob,H. (2008) Annual Spring Meeting of the Proteomics Standards Initiative, 23–25 April 2008, Toledo, Spain. *Proteomics*, **8**, 4168–4172.
29. Orchard,S., Kerrien,S., Jones,P., Ceol,A., Chatraryamontri,A., Salwinski,L., Nerothin,J. and Hermjakob,H. (2007) Submit your interaction data the IMEx way. *Proteomics*, **7**, 28–34.
30. Orchard,S., Salwinski,L., Kerrien,S., Montecchi-Palazzi,L., Oesterheld,M., Stümpflen,V., Ceol,A., Chatraryamontri,A., Armstrong,J., Woollard,P. *et al*. (2007) The minimum information required for reporting a molecular interaction experiment (MIMIx). *Nat. Biotechnol.*, **25**, 894–898.