xVis: a web server for the schematic visualization and interpretation of crosslink-derived spatial restraints

Maximilian Grimm[1], Tomasz Zimniak[1], Abdullah Kahraman[2] and Franz Herzog[1,*]

[1]Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität München, Munich 81377, Germany and [2]Institute of Molecular Life Sciences, University of Zurich, Zurich 8057, Switzerland

ABSTRACT

The identification of crosslinks by mass spectrometry has recently been established as an integral part of the hybrid structural analysis of protein complexes and networks. The crosslinking analysis determines distance restraints between two covalently linked amino acids which are typically summarized in a table format that precludes the immediate and comprehensive interpretation of the topological data. *xVis* displays crosslinks in clear schematic representations in form of a circular, bar or network diagram. The interactive graphs indicate the linkage sites and identification scores, depict the spatial proximity of structurally and functionally annotated protein regions and the evolutionary conservation of amino acids and facilitate clustering of proteins into subcomplexes according to the crosslink density. Furthermore, *xVis* offers two options for the qualitative assessment of the crosslink identifications by filtering crosslinks according to identification scores or false discovery rates and by displaying the corresponding fragment ion spectrum of each crosslink for the manual validation of the mass spectrometric data. Our web server provides an easy-to-use tool for the fast topological and functional interpretation of distance information on protein complex architectures and for the evaluation of crosslink fragment ion spectra. *xVis* is available under a Creative Commons Attribution-ShareAlike 4.0 International license at http://xvis.genzentrum.lmu.de/.

INTRODUCTION

Chemical labeling of functional groups has a long-standing history in the structural analysis of proteins and ribonucleic acids (1). Initially, the tagging with different chemistries such as hydrogen/deuterium exchange or hydroxyl radical labeling detected solvent accessible residues, also known as footprinting. In contrast, bifunctional reagents do not only label surface exposed residues but also covalently link residues and thus reveal their proximity within the three-dimensional (3D) structure of biomolecules and their complexes.

In recent years, advances in mass spectrometric instrumentation that facilitate the detection of modified peptides at high mass accuracy, resolution and sensitivity resulted in the development of different workflows for the identification of crosslinked peptides in complex protein samples (2,3). Chemical crosslinking combined with mass spectrometry predominantly involves reagents carrying N-hydroxy-succinimidyl esters which are reactive against primary amines of lysine side chains and protein N-termini. Chemistries targeting acidic amino acids have been successfully introduced (4), however, have not yet led to routine applications. Similar spatial information is obtained by combining mass spectrometry with the site-specific incorporation of photoactivatable amino acids which give rise to radicals highly reactive against various side chains (5).

The reported protocols for the identification of crosslinked residues differ by the physicochemical properties of the spacer arm (6) and by the dedicated analysis software. In particular, recent workflows implemented isotope-coded and collision-induced-dissociation (CID) cleavable spacer arms. Isotope-tagging introduces an isotopic mass shift which detects crosslinks as isotopic pairs and aids in identifying the fragment ions derived from the two crosslinked peptides by the search engine *xQuest* (7). CID cleavable crosslinkers generate reporter ions and give rise to the linear peptides modified with the remaining crosslinker masses which facilitates the identification of the crosslink sites (8).

Regardless of the workflow, the structural information determined by chemical and photoactivatable (9–11) crosslinking is a distance restraint which is based on the length of the reagent. We are using disuccinimidyl suberate as a crosslinker which has a spacer arm length of 11.4 Å. In our study on a network of modular protein phosphatase 2A (PP2A) complexes, crosslinks were evaluated by measuring the distances between the two bridged lysines on available X-ray structures and homology models (12). The median Euclidian Cα-Cα distances of 287 intra-protein and

---

*To whom correspondence should be addressed. Tel: +49 89 2180 76755; Fax: +49 89 2180 76999; Email: herzog@genzentrum.lmu.de

70 inter-protein crosslinks were 15.4 and 19.6 Å, respectively. The recently released *Xlink Analyzer* software enables the automatic analysis and visualization of crosslinks in the context of the 3D structures (13). In comparison to Euclidian distances the simulation of distance restraints on the solvent accessible protein surface has been shown to be superior in constraining computational molecular modeling and docking experiments (14,15). Inter-protein crosslinks indicate direct protein–protein contacts and will thus significantly support systems biology efforts in predicting and reconstructing protein interactions of biological networks (16).

The potential of crosslinks in integrating structural data of different sources has been demonstrated by several groups. In particular, the heterogeneity of the sample preparations and the intrinsic flexibility of macromolecular protein complexes have in many cases prevented their structural elucidation by established high-resolution methods. These limitations led to the design of hybrid approaches that apply crosslink-derived distance restraints and computational modeling to integrate crystal structures and models of individual subunits with electron microscopic (EM) densities and subunit localization (17–19). Depending on the input data the integrative structural analysis results in models at different levels of detail. In recent studies we have obtained a hybrid model of RNA polymerase I at near atomic resolution (20) whereas the availability of only sparse high-resolution structural data on the INO80 complex resulted in a topological model of less detail (21).

The budding yeast INO80 complex is composed of 15 subunits and belongs to the INO80/SWR1 family chromatin remodelers. Its important role in transcription and genome maintenance is the exchange of histone H2A.Z/H2B dimers with free H2A/H2B. To elucidate the mechanism of histone exchange we analyzed the architecture of the INO80 complex interacting with its nucleosome substrate by combining EM 3D densities and subunit localization with chemical crosslinking and mass spectrometry. This hybrid structural analysis revealed distinct structural modules of an embryo-shaped complex that embraces the nucleosome like a flexible clamp. Chemical crosslinking together with subunit localization and biochemical assays identified individual subcomplexes and assigned them to structural models (Figure 1A). Furthermore, the crosslink-derived spatial restraints were used to draw a proximity map of structurally and functionally annotated domains which provided detailed insights into the mode of nucleosome binding and histone H2A variant exchange.

We thus developed *xVis*, a software solution for the fast and schematic visualization and interpretation of distance information on protein complexes and networks. *xVis* is able to display crosslink-derived distances obtained by different workflows in form of a circular, bar or network plot. The import of structurally and functionally annotated domains as well as the evolutionary conservation rates of amino acids is crucial for the initial structural and functional interpretation prior to the more laborious generation of 3D models. Furthermore, the program offers two options for the evaluation of crosslink identifications. The filtering of crosslinks according to the identification score or the false discovery rate (FDR) allows their selective representation below or above a certain cut-off value. For crosslink datasets identified by the search engine *xQuest*, we offer the possibility to connect to a *xQuest* server for the manual inspection of the crosslink fragment ion spectra. We also implemented sorting algorithms that arrange and group subunits according to crosslink densities. Strikingly, all subunits of the INO80 - nucleosome complex are correctly assigned to the previously identified subcomplexes solely based on hierarchical clustering or the Markov Cluster (MCL) Algorithm using crosslink-derived restraints.

MATERIALS AND METHODS

xVis setup

xVis is a server-client-based software solution which uses PHP to fetch protein information from the InterPro (http://www.ebi.ac.uk/interpro/) and UniProt (http://www.uniprot.org/) databases. The protein lengths and secondary structures from UniProt are accessed with dbfetch (22) provided by EMBL-EBI using UniProt identifiers. The annotations are downloaded from InterPro with MartWizard (23). The diagrams are generated on the client by a JavaScript library named D3 (Data-Driven Documents) (24). D3 is used to generate scalable vector graphics (SVG) and to animate the arrangements of proteins in the network diagram with force layout. To get a similar design on each browser we used Bootstrap and Jasny Bootstrap.

xVis can be installed on a server with PHP by coping the source code into the web directory of the server. We also provide a portable version containing the XAMPP server for Windows. The local installation of *xVis* is important for its connection to the *xQuest* server. The URL of the *xQuest* server has to be entered on the 'Settings' page.

Input data

The input data for the visualization of crosslink datasets comprises the coordinates of crosslinks including protein descriptors and positions of the linked residues, protein lengths information, quality scores for the mass spectrometric identification, annotation of protein domains and *xQuest* references which provide the possibility to link each crosslink with the corresponding fragment ion spectrum.

The input data are uploaded as comma separated files in the 'Input Parameters' form. In addition to the main input file, two accessory files help to customize the representations. *xVis* identifies the input values by the column headings, hence, the order of columns is not important whereas the headings have to precisely match the names given in the sample input files which can be retrieved from the 'Input Parameters' and 'Downloads' page.

The main 'Crosslink Data' file is derived from the *xQuest* output file, however, several crosslink datasets obtained by other workflows can be visualized if delivered in the appropriate input format. The 'Crosslink Data' file has to provide the protein names and the position of the linkage sites and optionally, may contain the identification score, the FDR and the *xQuest* reference of the fragment ion spectra. The *xQuest* reference of each crosslink is generated from values of different columns, hence requiring a complete *xQuest* output file as 'Crosslink Data' input file.
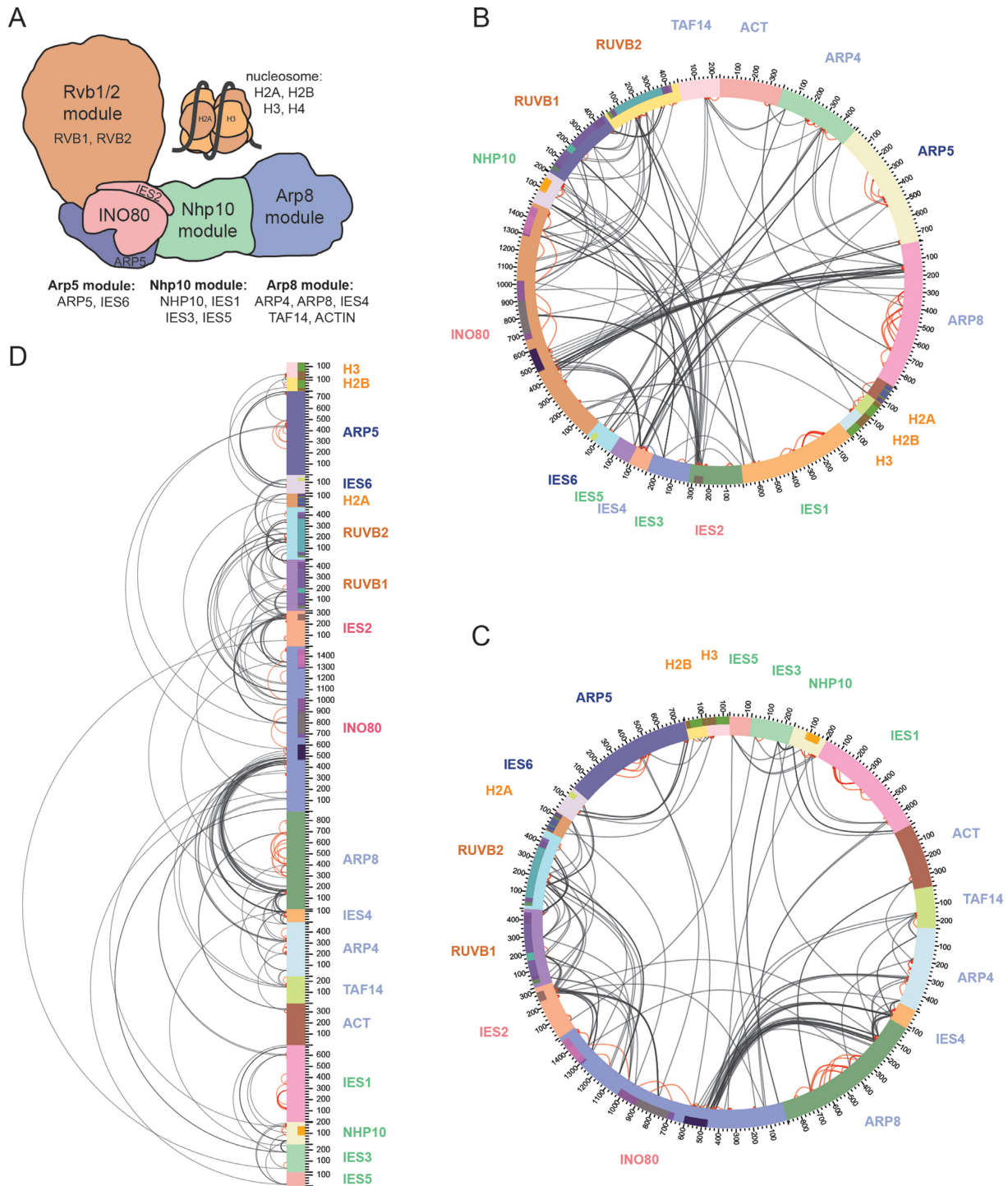
Figure 1. Topological analysis of the INO80 chromatin remodeler in complex with its nucleosome substrate. (**A**) Identification of subcomplexes and assignment to different structural modules using EM, chemical crosslinking and biochemical assays. (**B**) Crosslinked INO80 subunits are visualized and alphabetically sorted in a circular representation using *xVis*. The protein names are colored according to the modules in (A). (**C**) Hierarchical clustering of INO80 subunits based on crosslinks displayed in a circular diagram by *xVis*. (**D**) Hierarchical clustering of INO80 subunits based on crosslinks shown in a bar diagram by *xVis*. Bars represent InterPro domains (Figure 3A) (inter-protein crosslinks in black, intra-protein crosslinks in red).
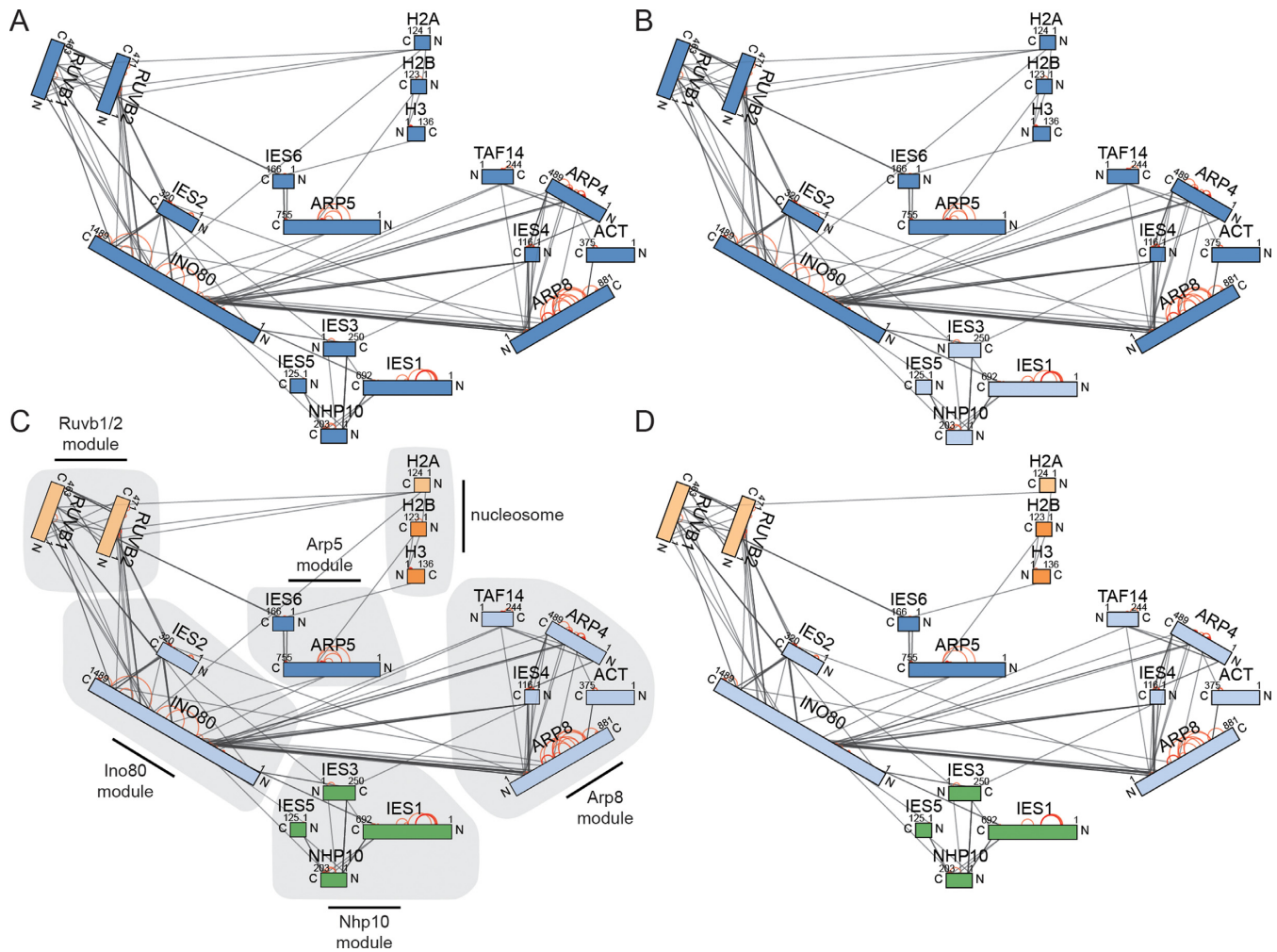
Figure 2. Topological analysis of the INO80 - nucleosome complex by applying the Markov Cluster (MCL) Algorithm in the network diagram. (**A**) Subunits non-clustered (all proteins blue colored). (**B**) Subunits clustered with the MCL parameters expansion=2 and inflation=1 separates the Nhp10 subcomplex (light blue). (**C**) Subunits clustered with expansion=1 and inflation=3 identifies several INO80 subcomplexes. (**D**) Proteins clustered like in (C) with crosslinks filtered by the identification score (cut-off 27).

The 'Protein Length' file has to use the same protein names as the 'Crosslink Data' file and specifies the user-defined protein lengths as amino acid numbers. Similarly, the 'Annotations' file links user-defined structural and functional annotations and their start and end positions with the protein names.

The proteins in the diagrams are labeled with the names given in the 'Crosslink Data' file. We recommend using the first part of the FASTA header (db|UniqueIdentifier|EntryName) as protein descriptor which facilitates automatic import of protein lengths and annotations through the primary accession number (UniqueIdentifier). The entry name or a user-defined name replacing the entry name in the descriptor is displayed in the representations. The exclusive application of user-defined names does not support the automatic import of protein lengths and annotation data and thus necessitates the upload of the respective input files.

Workflow

The typical workflow starts by selecting the input files and the plot type in the 'Input Parameters' form. Optional features of the schematic representations have to be activated at this step: (i) select automatic import of annotated secondary structures from UniProt; (ii) select automatic import of annotated domains, families, active sides, binding sites, conserved sites, repeats and post-translational modifications from InterPro; (iii) to introduce user-defined protein lengths deselect the UniProt protein lengths in order to access the respective upload mask; (iv) to group proteins based on the crosslink density choose the clustering type; (v) to filter the crosslink dataset define the column heading of the filter value; (vi) upload the Consurf (25) output files to display the evolutionary conservation of amino acids.

The 'Plot' button generates the respective crosslink representation in the same browser window. For each plot type the mouse over event on the crosslink line indicates the crosslink structure, the protein names, the absolute linkage positions and the selected score values. Mouse over on an-
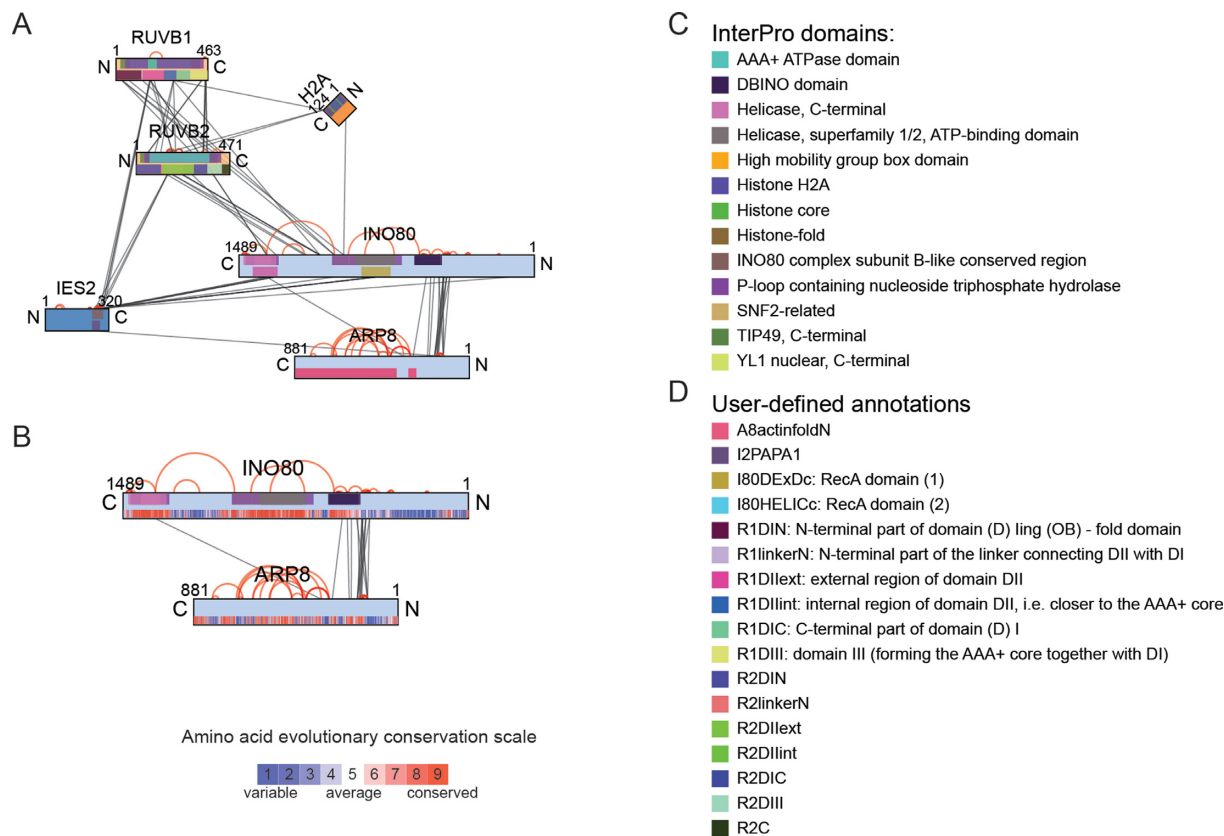
Figure 3. Customized annotations in *xVis*. (**A**) Selective representation of the Rvb1/2 and Ino80/Ies2 modules with the subunit Arp8 and histone H2A in a network diagram showing annotations from InterPro (top bars in the protein representation) and user-defined domains (bottom bars). (**B**) Evolutionary conservation of amino acid positions in the proteins Ino80 and Arp8 (bottom bars) obtained by the ConSurf webserver. (**C**) InterPro annotation legend used in (A) and (B). (**D**) User-defined domains displayed in (A).

notation fields displays the respective protein information with the sequence boundaries.

In the active plot the parameters of the plot type, sort type and filter value can be adjusted in the 'Change Input Parameters' menu. The 'Save and Restore Sessions' menu facilitates saving of all 'Input Parameters', 'Shapes', 'InterPro Annotations', 'Filter' and 'Markov Clustering Parameters' settings and of the protein positions in the network plot. A session starts with generating a plot in the 'Input Parameters' form. During the session you may switch plot types and change various parameters several times. Saving a session stores the parameters of the active plot and the protein positions of the network diagram in a file. Loading a session opens the latest diagram and restores all settings for further manipulation.

Circular and bar plots. The circular and bar plots (Figure 1B–D) offer the same visualization features. The proteins are represented as arcs or bars which are scaled to the protein lengths and crosslinks are drawn as circular lines. Protein secondary structures and annotations imported from public databases are plotted into the protein bars and arcs next to the sequence scale (Figure 1C and D) whereas user-defined domains are indicated in the opposite half of the objects (Figure 3A).

For these plot types the crosslinked proteins are arranged either alphabetically (Figure 1B), alphabetically in groups

or hierarchically clustered (Figure 1C). Protein complexes are indicated by by the hierarchical arrangement which orders the proteins according to the number of crosslinks. For this purpose we built a dendrogram by a bottom up algorithm that starts at the proteins with the most inter-protein crosslinks and uses the average number of restraints to calculate the new cluster. Strikingly, hierarchical clustering, even in the absence of supporting structural and biochemical data, is sufficient to accurately assign INO80 subunits to its subcomplexes (Figure 1A and C).

Several menus shown at the left-hand side of the diagram facilitate the customization and interpretation of the crosslink graph. The 'Secondary Structures' menu allows for selectively displaying the different secondary structures and their colors. The 'InterPro Annotations' menu provides the possibility to choose the annotations which are represented in the diagram. The quality of crosslink identifications is indicated by score values. The *xQuest* software calculates identification scores (Id-Score) and FDRs, however, also scores computed by different crosslink analysis software can be selected in the 'Input Parameters' form as filter value. Using the slider in the 'Filter' menu selectively shows crosslinks below or above the chosen threshold. This menu further enables the selective display of inter- and intra-protein crosslinks in different colors and marking individual proteins shows the selection of crosslinks that in-
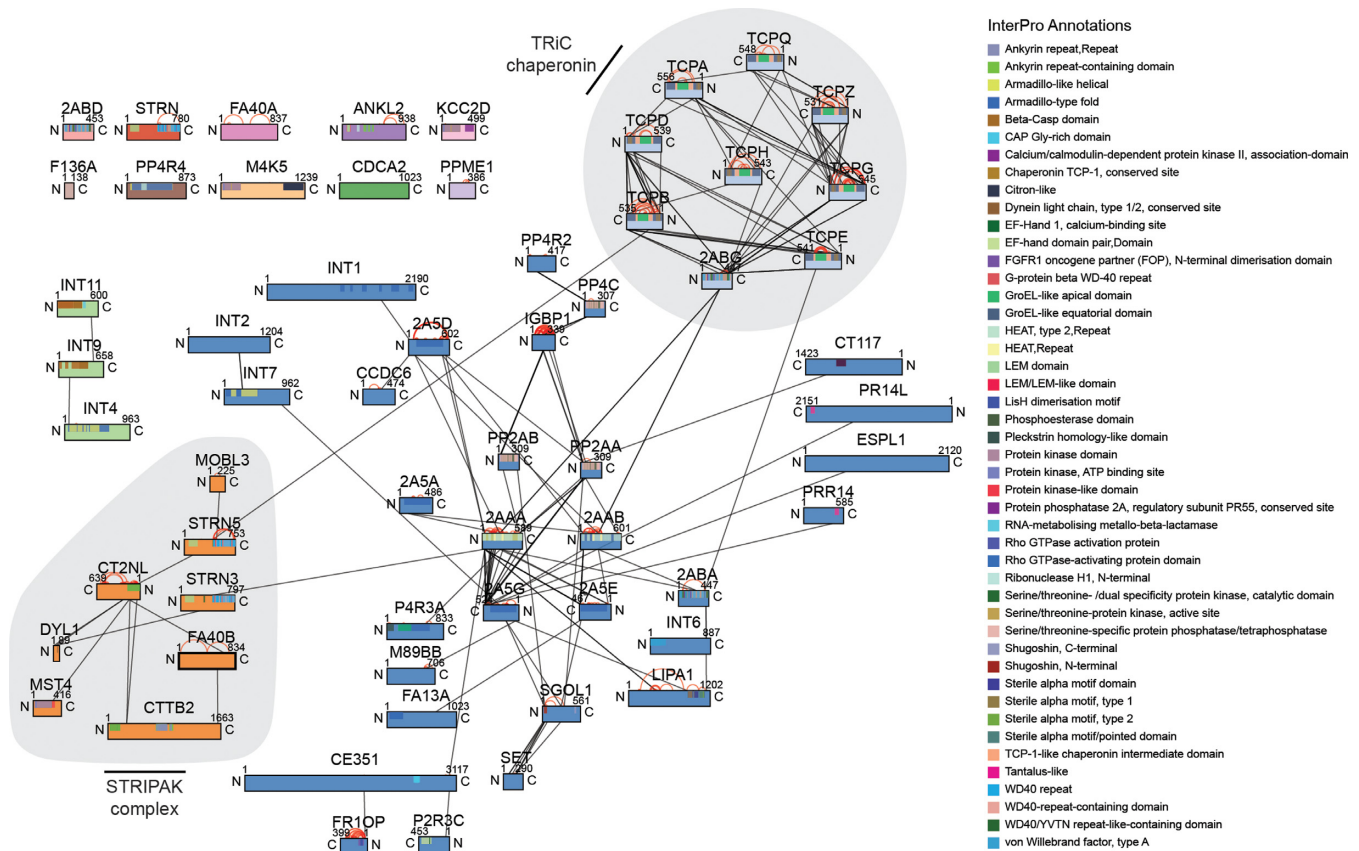
Figure 4. Topology of a network of modular PP2A complexes displayed in a network diagram. Proteins were grouped by the MCL algorithm applying the parameters expansion=2 and inflation=2 (STRIPAK, B'''/striatin-interacting phosphatase and kinase, STRIPAK, complex; TCP1 ring complex, TRiC).

tersect with these proteins. Once the layout of the circular or bar diagram is optimized the graphs may be exported as SVG.

Network diagram. The network is generated by a force-directed layout. In contrast to the circular and bar plots, selecting proteins in the 'Filter' menu' of the network diagram displays a selection of the subunits and their crosslinks. The sorting type of the network diagram has to be specified in the 'Input Parameters' or 'Change Input Parameters' form and similarly as for the circular plot, facilitates the grouping of proteins that are bridged by crosslinks. The MCL Algorithm (26) identifies clusters of proteins with high crosslink densities and separates them from low density areas. The 'Markov Cluster Parameters' menu helps to set the density restraints that define a cluster. The expansion and inflation values determine the cluster boundaries whereas the threshold excludes groups of proteins with numbers below this value from clustering.

Proteins of a cluster are indicated by the same color. In the network diagram without sorting the 15 subunits of the INO80 complex and the histones are same-colored (Figure 2A). The MCL Algorithm applied with parameters for wider cluster boundaries identifies the Nhp10 module, however, cannot discriminate between the remaining INO80 subunits and the histones (Figure 2B). Adjusting the parameters to narrow cluster boundaries perfectly resolves all

INO80 subcomplexes (Figure 2C) and places the histones H2B and H3 in a separate cluster. Similar to the hierarchical clustering in the circular and bar plot, histone H2A is assigned to the Rvb1/2 module indicating spatial proximity within the INO80 - nucleosome complex.

Evaluation and interpretation. A slider in the 'Filter' menu of the individual plots facilitates the selective display of crosslinks according to identification scores (Figure 2D) or FDRs.

To further assess the quality of crosslink identifications *xVis* offers the possibility to evaluate the fragment ion spectra. Once the local installation of *xVis* is connected to a *xQuest* server double-clicking on a crosslink line in a diagram opens the corresponding fragment ion spectrum in a separate tab. This allows the interactive manual inspection of the spectral quality of individual crosslink identifications in the dataset.

The analysis of the INO80 complex with *xVis* revealed that clustering of proteins according to the crosslink density identified subcomplexes that correspond to structural modules of the remodeler (Figure 2C).

In addition to the subunit architecture, plotting annotated protein domains and motifs visualizes their topology and provides further structural and functional insights. The crosslinks of the H2A core structures to the Snf2 ATPase domain of INO80 and next to the AAA$^+$ ATPases of Rvb1

and Rvb2 indicate that the H2A/H2B dimer is in close proximity to these ATPases and suggested an ATP-dependent mechanical activity for histone exchange (Figure 3A, C and D) (21).

As amino acids of functionally important sites, like binding interfaces or catalytic sites, tend to be evolutionary conserved, *xVis* facilitates the import of evolutionary conservation rates estimated by the Consurf web server (25). The Consurf conservation scores are plotted into the protein objects following the scale from variable (1) to conserved (9) amino acids (Figure 3B).

DISCUSSION

Here, we present the web server *xVis* for the fast and interactive visualization and interpretation of spatial restraints identified by different mass spectrometric workflows. The demand for tools for the 2D visualization of crosslinks is underscored by the software *xiNet* which was published during preparation of this manuscript (27) (http://crosslinkviewer.org/). *xiNet* represents crosslink data as node-link diagrams at residue resolution using an elaborate web-design for the manipulation of the graphs and for the visualization of annotated protein regions.

xVis offers, in addition to the network diagram, a circular or bar plot which in many cases is the preferred representation of spatial restraints on oligomeric complexes. Our software further indicates the evolutionary conservation of amino acids which might aid in pinpointing important residues that establish a protein binding interface. As an essential feature for the evaluation of the crosslink data *xVis* provides a link to the analysis software *xQuest* and displays the fragment ion spectrum for the qualitative assessment of each crosslink identification. Importantly, our web server uses clustering algorithms to identify proximal and distal subunits in a complex which has the potential to resolve the subcomplexes of the INO80 chromatin remodeler and complexes of a PP2A network (Figures 2C and 4).

We demonstrate on the INO80 chromatin remodeler in complex with its nucleosome substrate that the schematic visualization and interpretation of crosslinks by the *xVis* web server provides detailed mechanistic insights that agree with the previously reported findings (21). In particular in the absence of high-resolution structural data, the clustering of subunits according to the crosslink density identified distinct subcomplexes and revealed their architecture (Figure 2C). The visualization of annotated protein domains using *xVis* indicated that the INO80 partially embraces its nucleosome substrate and suggested how histone H2A variant exchange may be catalyzed. To demonstrate the potential of *xVis* for the interpretation of large crosslink datasets we visualized 239 inter-protein and 732 intra-protein restraints acquired on a network of PP2A complexes (12) and grouped the proteins using the MCL Algorithm. In this case, the discrimination between complexes separates the TRiC chaperonin and STRIPAK from PP2A core complexes and their adaptor and substrate proteins (Figure 4).

Spatial restraints acquired by crosslinking and mass spectrometry have been successfully implemented by different hybrid structural biology approaches. In a few studies, structural biologists have developed complete bioinformatic workflows providing 3D hybrid models of macromolecular protein complexes at different levels of detail (17,28). The schematic representations by *xVis* and other web servers visualize the subunit topology by positioning protein bars or nodes according to intermolecular crosslinks. Replacing these protein objects in future implementations by simplified subunit shapes derived from intramolecular restraints may delineate the 2D topology of protein complexes similar to 3D coarse-grained representations.

Remarkably, the circular and network diagrams obtained by *xVis* indicate the spatial proximity of annotated protein regions. The structural information at the level of protein domains or motifs facilitates the prediction of structural and functional models and the design of experiments to test their relevance *in vitro* and *in vivo*.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Fabris,D. and Yu,E.T. (2010) Elucidating the higher-order structure of biopolymers by structural probing and mass spectrometry: MS3D. *J. Mass Spectrom.*, **45**, 841–860.
2. Leitner,A., Walzthoeni,T., Kahraman,A., Herzog,F., Rinner,O., Beck,M. and Aebersold,R. (2010) Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol. Cell. Proteomics*, **9**, 1634–1649.
3. Sinz,A. (2006) Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom. Rev.*, **25**, 663–682.
4. Leitner,A., Joachimiak,L.A., Unverdorben,P., Walzthoeni,T., Frydman,J., Forster,F. and Aebersold,R. (2014) Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 9455–9460.
5. Forne,I., Ludwigsen,J., Imhof,A., Becker,P.B. and Mueller-Planitz,F. (2012) Probing the conformation of the ISWI ATPase domain with genetically encoded photoreactive crosslinkers and mass spectrometry. *Mol. Cell. Proteomics*, **11**, doi:10.1074/mcp.M111.012088.
6. Petrotchenko,E.V. and Borchers,C.H. (2010) Crosslinking combined with mass spectrometry for structural proteomics. *Mass Spectrom. Rev.*, **29**, 862–876.
7. Walzthoeni,T., Claassen,M., Leitner,A., Herzog,F., Bohn,S., Forster,F., Beck,M. and Aebersold,R. (2012) False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nat. Methods*, **9**, 901–903.
8. Kao,A., Chiu,C.L., Vellucci,D., Yang,Y., Patel,V.R., Guan,S., Randall,A., Baldi,P., Rychnovsky,S.D. and Huang,L. (2011) Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes. *Mol. Cell. Proteomics*, **10**, doi:10.1074/mcp.M110.002212.
9. Ai,H.W., Shen,W., Sagi,A., Chen,P.R. and Schultz,P.G. (2011) Probing protein-protein interactions with a genetically encoded photo-crosslinking amino acid. *Chembiochem*, **12**, 1854–1857.
10. Kramer,K., Sachsenberg,T., Beckmann,B.M., Qamar,S., Boon,K.L., Hentze,M.W., Kohlbacher,O. and Urlaub,H. (2014) Photo-cross-linking and high-resolution mass spectrometry for

assignment of RNA-binding sites in RNA-binding proteins. *Nat. Methods*, **11**, 1064–1070.

11. Krauth,F., Ihling,C.H., Ruttinger,H.H. and Sinz,A. (2009) Heterobifunctional isotope-labeled amine-reactive photo-cross-linker for structural investigation of proteins by matrix-assisted laser desorption/ionization tandem time-of-flight and electrospray ionization LTQ-Orbitrap mass spectrometry. *Rapid Commun. Mass Spectrom.*, **23**, 2811–2818.

12. Herzog,F., Kahraman,A., Boehringer,D., Mak,R., Bracher,A., Walzthoeni,T., Leitner,A., Beck,M., Hartl,F.U., Ban,N. *et al.* (2012) Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. *Science*, **337**, 1348–1352.

13. Kosinski,J., von Appen,A., Ori,A., Karius,K., Muller,C.W. and Beck,M. (2015) Xlink Analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J.Struct. Biol.*, **189**, 177–183.

14. Kahraman,A., Herzog,F., Leitner,A., Rosenberger,G., Aebersold,R. and Malmstrom,L. (2013) Cross-link guided molecular modeling with ROSETTA. *PLoS One*, **8**, e73411.

15. Potluri,S., Khan,A.A., Kuzminykh,A., Bujnicki,J.M., Friedman,A.M. and Bailey-Kellogg,C. (2004) Geometric analysis of cross-linkability for protein fold discrimination. *Pac. Symp. Biocomput.*, **2004**, 447–458.

16. Wass,M.N., Fuentes,G., Pons,C., Pazos,F. and Valencia,A. (2011) Towards the prediction of protein interaction partners using physical docking. *Mol. Syst. biol.*, **7**, 1–8.

17. Lasker,K., Forster,F., Bohn,S., Walzthoeni,T., Villa,E., Unverdorben,P., Beck,F., Aebersold,R., Sali,A. and Baumeister,W. (2012) Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 1380–1387.

18. Bui,K.H., von Appen,A., DiGuilio,A.L., Ori,A., Sparks,L., Mackmull,M.T., Bock,T., Hagen,W., Andres-Pons,A., Glavy,J.S. *et al.* (2013) Integrated structural analysis of the human nuclear pore complex scaffold. *Cell*, **155**, 1233–1243.

19. Plaschka,C., Lariviere,L., Wenzeck,L., Seizl,M., Hemann,M., Tegunov,D., Petrotchenko,E.V., Borchers,C.H., Baumeister,W., Herzog,F. *et al.* (2015) Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature*, **518**, 376–380.

20. Jennebach,S., Herzog,F., Aebersold,R. and Cramer,P. (2012) Crosslinking-MS analysis reveals RNA polymerase I domain architecture and basis of rRNA cleavage. *Nucleic Acids Res.*, **40**, 5591–5601.

21. Tosi,A., Haas,C., Herzog,F., Gilmozzi,A., Berninghausen,O., Ungewickell,C., Gerhold,C.B., Lakomek,K., Aebersold,R., Beckmann,R. *et al.* (2013) Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex. *Cell*, **154**, 1207–1219.

22. McWilliam,H., Valentin,F., Goujon,M., Li,W., Narayanasamy,M., Martin,J., Miyar,T. and Lopez,R. (2009) Web services at the European Bioinformatics Institute-2009. *Nucleic Acids Res.*, **37**, W6–W10.

23. Zhang,J., Haider,S., Baran,J., Cros,A., Guberman,J.M., Hsu,J., Liang,Y., Yao,L. and Kasprzyk,A. (2011) BioMart: a data federation framework for large collaborative projects. *Database*, **2011**, bar038.

24. Bostock,M., Ogievetsky,V. and Heer,J. (2011) D(3): Data-Driven Documents. *IEEE Trans. Vis. Ccomput. Graph.*, **17**, 2301–2309.

25. Glaser,F., Pupko,T., Paz,I., Bell,R.E., Bechor-Shental,D., Martz,E. and Ben-Tal,N. (2003) ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*, **19**, 163–164.

26. van Dongen,S. and Abreu-Goodger,C. (2012) Using MCL to extract clusters from networks. *Methods Mol. Biol.*, **804**, 281–295.

27. Combe,C.W., Fischer,L. and Rappsilber,J. (2015) xiNET: cross-link network maps with residue resolution. *Mol. Cell. Proteomics*, **14**, 1137–1147.

28. Alber,F., Dokudovskaya,S., Veenhoff,L.M., Zhang,W., Kipper,J., Devos,D., Suprapto,A., Karni-Schmidt,O., Williams,R., Chait,B.T. *et al.* (2007) Determining the architectures of macromolecular assemblies. *Nature*, **450**, 683–694.