

TopoLink: A software to validate structural models using chemical crosslinking constraints

Leandro Martinez (✉ leandromartinez98@gmail.com)

Martinez's Molecular Modeling Lab, Institute of Chemistry - University of Campinas

Allan Ferrari

Institute of Chemistry - University of Campinas

Fabio Cesar Gozzo

Institute of Chemistry - University of Campinas

Method Article

Keywords: chemical crosslinking, mass spectrometry, protein modeling, rosetta

Posted Date: March 25th, 2017

DOI: <https://doi.org/10.1038/protex.2017.035>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

TOPOLINK is a package to compute topological distances between atoms on the surface of proteins and validate structural models from cross-linking experimental data. The package is available at: "<http://leandro.iqm.unicamp.br/topolink>":<http://leandro.iqm.unicamp.br/topolink> Chemical cross-linking mass spectrometry is powerful experimental method for the evaluation of interatomic distances in biomolecular structures. These experiments provided upper bounds for the distances between residues, by the identification of pairs of atoms in the surface of protein which are within the linker reach. The evaluation of structural models using chemical cross-linking depends on the determination of the topological, or surface-accessible, distances of the reactive residues in the model, and its comparison with experimental observations. TopoLink is a package designed to compute topological distances and evaluate structural models, structured with the experimental result in mind. The user provides a description of the experiments performed, the types of cross-links that can be formed in each experiment, and an account of the experimental result, and TopoLink determines how consistent a structural model is consistent with that experimental results. For doing so, TopoLink computes all possible cross-links of the structure and their consistency with experimental observations. It also provides means for the visualization of the topological distances on the protein structure.

Procedure

All details on the usage of the software are available at:

"<http://leandro.iqm.unicamp.br/topolink>":<http://leandro.iqm.unicamp.br/topolink> The TopoLink package must be run with: `topolink inputfile.inp > topolink.log` where `inputfile.inp` is the parameter input file and `topolink.log` is the output log file of the run. **The basic TopoLink input file contains**: 1. The name of the PDB file of the model to be evaluated. 2. The type of links to be computed. 3. The specification of the linker used and experimental observations. 1. **Name of model file and output of linkers computed**. The name of the input file is provided with: `pdbfile model.pdb` and this definition can be overwritten by providing the name of the file in the command line, to facilitate the execution of TopoLink for multiple models with the same input file, as described above. TopoLink can output coordinates for the topological paths obtained, by using `printlinks yes linkdir ./links` where `./links` is the directory where the PDB files of the links will be written (the directory must exist). Note that there can be many links, so multiple files will be created. The files created are simple PDB files which can be open in any structure visualization software together with the model PDB file to visualize the topological distance. For the execution of TopoLink in multiple models, it is recommended not to write link files, using `printlinks no`. 2. **Types of links to be computed** The next important parameter of the TopoLink input file is the definition of which links are to be searched and computed. There are three options: `compute observed` `compute all` `compute reactive` When using `compute observed`, only the links that were observed experimentally (see below) will be computed. If `compute all` is used instead, all possible crosslinks will be computed. That means that, given the definition of the linker used, TopoLink will search for consistent topological distances for every pair of residues that could, by the chemical nature of the linker and the residues involved, be attached by the linker. Finally, `compute reactive` tells TopoLink to consider that only residues that were experimentally observed to react (by participating in

observed crosslinks or dead-ends) are reactive. Then, TopoLink will search for topological distances consistent with the linker used only between these pairs of "observed-reactive" residues. Additionally, the user may optionally choose to compute only the crosslinks between different chains of the PDB file, for instance to compute inter-proteins crosslinks in a complex. To do so, just add the interchain option to the input file. All intra-protein links will be skipped. 3. ****Specification of linker types and experimental observations**** The linker types and observations are specified in the TopoLink input file using an experiment-based structure, as described in ["http://leandro.iqm.unicamp.br/topolink/usage.shtml"](http://leandro.iqm.unicamp.br/topolink/usage.shtml):<http://leandro.iqm.unicamp.br/topolink/usage.shtml> Link types, reactive residues, and experimental observations (either crosslinks or deadends) are provided by the user with appropriate keywords.

Acknowledgements

We thank Fapesp, projects 2010/16947-9, 2013/05475-7 and 2013/08293-7 for financial support.

Figures

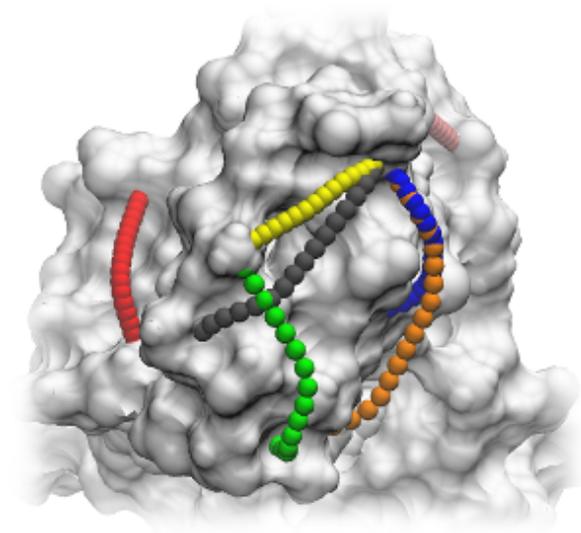


Figure 1

Topological links observed on the protein surface.