BIOINFORMATICS

Vol. 00 no. 00 2008
Pages 1–2

MSX-3D : A tool to validate 3D protein models using mass spectrometry

Michaël Heymann [1], David Paramelle [2], Gilles Subra [2], Eric Forest [3], Jean Martinez [2], Christophe Geourjon [1] and Gilbert Deléage [1*]

[1]Institut de Biologie et Chimie des Protéines (IBCP) UMR5086 CNRS, Université Lyon 1, IFR128 BioSciences Lyon-Gerland, 7 passage du Vercors 69367 Lyon cedex 07, France
[2]Institut des Biomolécules Max Mousseron (IBMM), UMR5247 CNRS, Universités Montpellier 1 et 2, 15 avenue Charles Flahault, 34000 Montpellier, France
[3]Protein Mass Spectrometry Laboratory, Institut de Biologie Structurale, UMR5075 CEA, CNRS, UJF, 41 rue Jules Horowitz, 38027 Grenoble Cedex 1, France

Received on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXX

ABSTRACT

Motivation: The technique of chemical crosslinking followed by mass spectrometry has proven to bring valuable information about the protein structure and interactions between proteic subunits. It is an effective and efficient way to experimentally investigate some aspects of a protein structure when NMR and X-ray crystallography data are lacking.

Results: We introduce MSX-3D, a tool specifically geared to validate protein models using mass spectrometry. In addition to classical peptides identifications, it allows an interactive 3D visualization of the distance constraints derived from a crosslinking experiment.

Availability: Freely available at http://proteomics-pbil.ibcp.fr

Contact: g.deleage@ibcp.fr

1 INTRODUCTION

With the ever growing size of sequence databases and the difficult task of experimentally elucidating the corresponding protein structure using classical techniques like NMR or X-ray crystallography, it was necessary to develop *in silico* methods in order to be able to predict the protein structure. But many heuristics and other approximations must be introduced to obtain theoretical models in a reasonable amount of time. Thus, when no close homolog protein structure is known, a protein model can be very unreliable. Molecular modeling can be especially tricky, too, for distantly related target and template protein sharing the same fold. That's why it is important to experimentally validate these models.

In this context, chemical crosslinking followed by mass spectrometry was found to be a technique of choice, both for its analytical speed and sensitivity (Back *et al*., 2003). Thus, different software were developed to facilitate the study of crosslinked components (Schilling *et al*., 2003; Gao *et al*., 2006; Wefing *et al*., 2006). However, these software appear to be mostly designed for mass spectrometrists, with a lot of fine-tuning possibilities, or optimized for a certain type of mass spectrometer. Moreover, all the tools available so far only work with a FASTA entry and are therefore unable to benefit from possibly existing structural data (e.g. after an *in silico* modeling experiment). That's why we developed MSX-3D, a tool designed for structural biologists, having in mind the ease of use and specifically focusing on the validation of theoretical protein models.

2 FUNCTIONALITY

MSX-3D is a command-line tool written in C++ and wrapped into a web interface freely available via the proteomics portal at the PBIL (http://proteomics-pbil.ibcp.fr). The user can upload a PDB file containing all the theoretical models. A model can even consist of more than one peptidic chain, thus opening the door for the study of protein complexes. After setting some experimental parameters (peak list, protease used for the digestion) and defining an unlimited and completely customizable list of possible modifications (crosslinker used, N-ter acetylation, etc.), MSX-3D will compute all the theoretical masses obtainable from this sample and compare theses values to the experimental masses initially given. Results will appear in a HTML page. For each hypothetical "hit", in addition to the simple peak identification, a complete overview of the distance constraints deduced from the crosslinks is given for each model (see Figure **1.b**). Moreover, an interactive three dimensional view of the crosslinked protein is presented thanks to a Jmol applet (Figures **1.c** and **1.d**).

3 CASE STUDY

It was shown that the cytochrome cd1 nitrite reductase from *Thiosphera pantotropha* refolds upon reduction (Williams *et al*., 1997). Almost complete structures of the oxidized and reduced forms can be found in the PDB database under the codes *1aoq* and *1aof*, respectively. We used these two structures as templates to model the nitrite reductase from *Hydrogenobaculum sp.* Y04AAS1[1]

*to whom correspondence should be addressed

[1] UniProt entry name: A7WHC2_9AQUI

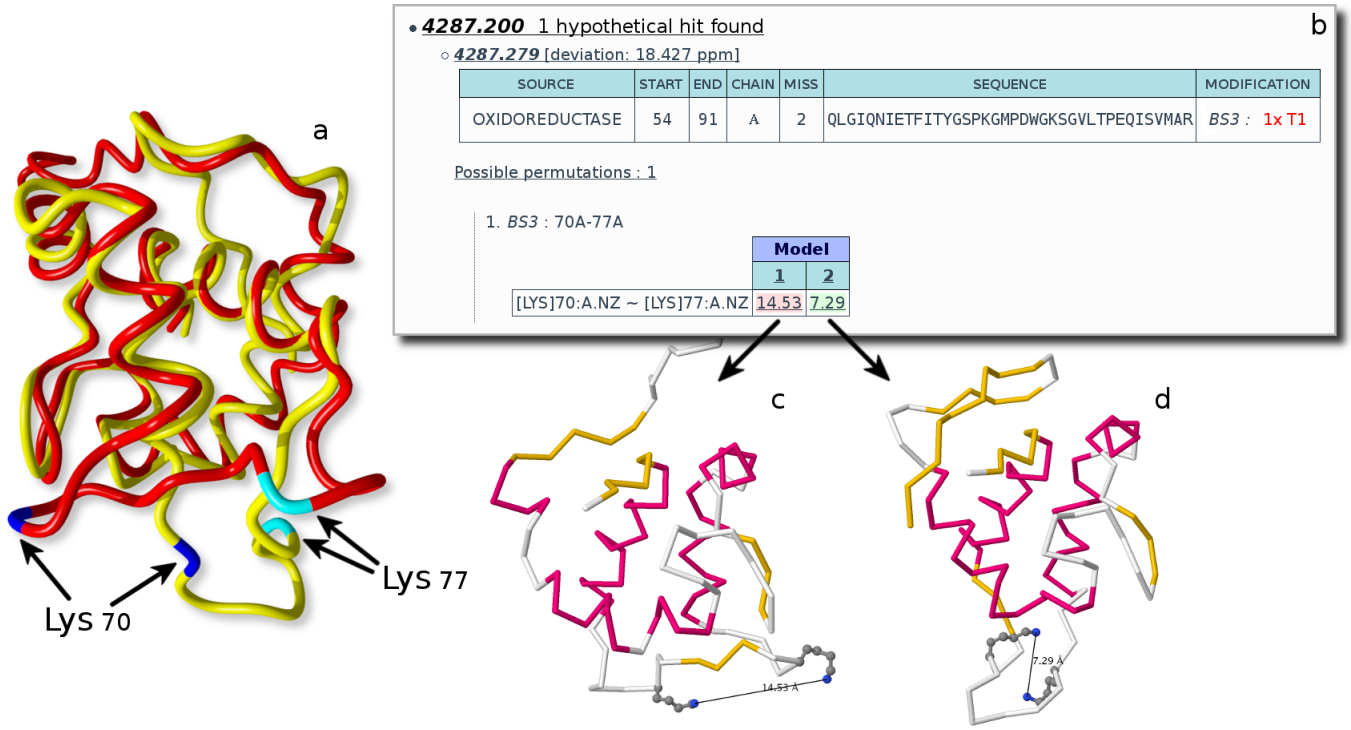

Fig. 1. (**a**) Structural alignment of the *c* domain of the two generated models. (**b**) An extract of MSX-3D output. Distances and model indexes are clickable items in order to communicate with a Jmol applet displaying an interactive 3D view emphasizing the selected crosslink, as shown in (**c**) and (**d**).

with Geno3D (Combet *et al.*, 2002). Figure **1.a** shows a structural alignment of the *c* domain for both models (reduced form in red, oxidized form in yellow); the rest of the protein, a rigid β-propeller domain, is not represented. Note the two lysines in the non-superimposed region (in dark blue and cyan, respectively). MSX-3D shows us that an experimental mass of 4287.200Da might correspond to a discriminative crosslink: Nζ of Lys 70 and Lys 77 are distant from 14.53Å in model 1 (reduced form) whereas they are distant from only 7.29Å in model 2 (oxidized form). The crosslinker we used, namely the bis(sulfosuccinimidyl) suberate (BS^3), which has an 11.4Å spacer arm length, would be unable to bind Lys 70 and Lys 77 of model 1. We can therefore affirm that the latter is an inacurate model in our case.

4 CONCLUSION

We introduced MSX-3D, a tool that easily and efficiently allows the structural biologist to investigate a protein structure, with the help of chemical crosslinking followed by mass spectrometry. It immediately emphasizes an inaccurate model or a possible ambiguity while attempting to identify the different mass peaks. It is a helpfull tool to discriminate accurate models froms inaccurate ones.

FUNDING

This work is supported by the Rhône-Alpes Region (Programme Emergence).

REFERENCES

Jaap W Back, Luitzen de Jong, Anton O Muijsers and Chris G de Koster (2003) Chemical cross-linking and mass spectrometry for protein structural modeling, *J Mol Biol*, **331**, 303-313.

Christophe Combet, Martin Jambon, Gilbert Deléage and Christophe Geourjon (2002) Geno3D: automatic comparative molecular modelling of protein, *Bioinformatics*, **18**, 213-214.

Qiuxia Gao, Song Xue, Catalin E Doneanu, Scott A Shaffer, David R Goodlett and Sidney D Nelson (2006) Pro-CrossLink. Software tool for protein cross-linking and mass spectrometry, *Anal Chem*, **78**, 2145-2149.

Birgit Schilling, Richard H Row, Bradford W Gibson, Xin Guo and Malin M Young (2003) MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides, *J Am Soc Mass Spectrom*, **14**, 834-850.

Stephan Wefing, Volker Schnaible and Daniel Hoffmann (2006) SearchXLinks. A program for the identification of disulfide bonds in proteins from mass spectra, *Anal Chem*, **78**, 1235-1241.

Pamela A. Williams, Vilmos Fülöp, Elspeth F. Garman, Neil F. Saunders, Stuart J. Ferguson and Janos Hajdu (1997) Haem-ligand switching during catalysis in crystals of a nitrogen-cycle enzyme, *Nature*, **389**, 406-412.