massXpert 2: a cross-platform software environment for polymer chemistry modelling and simulation/analysis of mass spectrometric data

Filippo Rusconi

Laboratoire de biophysique, Muséum national d’Histoire naturelle, CNRS UMR7196 - INSERM U565 - MNHN
USM0503; 57, rue Cuvier – Case postale 26 – F-75231 Paris Cedex 05 – France

Received on April 23, 2009; revised on July 3, 2009; accepted on July 30, 2009

Advance publication September 9, 2009

Associate Editor: Thomas Lengauer

ABSTRACT

Summary: Since the middle of the 90s, mass spectrometry has evolved into an almost indispensable tool in structural studies on an ever-growing variety of (bio-)polymers, of which proteins, sugars and nucleic acids are the most prominent. Since the first public release of massXpert, the advances of mass spectrometry have motivated continuous and thorough maintenance of that software, in the form of two full software rewrites, culminating with massXpert 2, which we describe in this report. We shall describe the profound changes in massXpert that were performed so as to keep up with the technical advances in mass spectrometry since a decade.

Availability: The massXpert 2 software is an open source and free software project hosted at http://www.massxpert.org.

Contact: rusconi@mnhn.fr

Supplementary information: Supplementary data are available at Bioinformatics online.

1 INTRODUCTION

Since the introduction of large biopolymer analysis with soft ionization techniques, mass spectrometry as a whole has become a powerful combination of instrumentation, techniques and methodologies. Advances in each of these fields as well as in sample preparation methodologies have widened the scope of mass spectrometry from its main application field—biopolymer structure characterization—to other scientific areas, like cell science. We have witnessed these advances and practised them to some extent: our software development had to adapt to the new requirements set forward either by our own experiments, our collaborators, or the advances of mass spectrometry since a decade.

2 IMPLEMENTATION

2.1 Methods

The development of massXpert 2 was performed using the well-known cross-platform open source free software Qt libraries from Nokia (C++ language; software freely available at http://www.qtsoftware.com).

massXpert 2 was developed as an integrated environment in which four functional modules inter-relate: XpertDef for polymer chemistry modelling; XpertCalc for programmable desktop calculator-based calculations; XpertEdit for sequence editing and for all the biochemical and mass spectrometric simulations; XpertMiner for \((m/z, z)\) pair list-based data mining.

2.2 Results

2.2.1 Evolving computing landscape massXpert 1 was written at a time when the Microsoft Windows computing platform was overwhelmingly dominant. This domination has reduced since, which motivated the first Linux-based rewrite into GNU polyxmass (Rusconi, 2006 and references therein, for comparisons of program features with other existing software). A second rewrite leveraged on the experience gained during the first one, adding new features and most importantly adding native cross-platform support: massXpert 2 executes natively on all the common computing platforms and, in particular, can execute side by side with the software that drives the mass spectrometer. Indeed, massXpert was designed from the ground up to be used as a decision-making aid that should be harnessed while performing mass data acquisitions.

2.2.2 Increasing diversity of analysed polymers massXpert 1 only dealt with proteins, and almost all its data were internally hard-coded. The ever-increasing variety of polymers being analysed by mass spectrometry prompted us to perform a full software redesign so as to let users define polymer chemistries \(ex\ nihilo\). The XpertDef module lets the user define any aspect of a polymer chemistry, from atoms through monomers and chemical modifications to cleaving agents and fragmentation patterns. massXpert 2 is shipped with definitions for proteins, saccharides and nucleic acids.

2.2.3 Electrospray ionization and multi-charged ions. When massXpert 1 was first developed, MALDI-TOF mass spectrometry was predominant, hence the software only dealt with mono-protonated species. Because electrospray ionization was later widely adopted, massXpert 2 had to provide a way to configure ionization levels for any simulation. This new feature was considered essential, as GNU polyxmass and massXpert 1 were of almost no use in an
specifying the involved monomers as a semicolon-separated list of
linking is performed by choosing one cross-linker from the list
indicating the number of incomplete cross-links. Monomer cross-
the selected sequence, the program alerts the user with a message
selected, which do not encompass all the entities cross-linked to
cross-linked oligomers. When one or more sequence regions are
small chain links superimposed onto their vignette. The editor now
sequence editor with cross-linked cysteinyl residues displayed with
easily modelled. Figure 1 shows the Kunitz inhibitor loaded in the
the Kunitz (1945) inhibitor—with its seven disulfide bonds—can be
number of cross-links to be set to a polymer sequence. For example,
new cross-link framework designed in
2.2.5 A completely new cross-link simulation framework
The new cross-link framework designed in massXpert 2 allows any
number of cross-links to be set to a polymer sequence. For example,
the Kunitz (1945) inhibitor—with its seven disulfide bonds—can be
easily modelled. Figure 1 shows the Kunitz inhibitor loaded in the
sequence editor with cross-linked cysteinyl residues displayed with
small chain links superimposed onto their vignette. The editor now
features a multiple-region selection mode that allows simulation of
cross-linked oligomers. When one or more sequence regions are
selected, which do not encompass all the entities cross-linked to
the selected sequence, the program alerts the user with a message
indicating the number of incomplete cross-links. Monomer cross-
linking is performed by choosing one cross-linker from the list
of available cross-linkers (in the current polymer definition) and
specifying the involved monomers as a semicolon-separated list of
monomer positions. Any cross-link might involve any number of
monomers. As an example, the very peculiar cross-links involved
in the formation of the fluorescent proteins’ chromophore are easily
simulated in massXpert 2. In this case, cross-linking the three cyan
fluorescent protein’s Thr, Trp and Gly residues taking their part in
the chromophore would involve specifying their positions as the
‘65:66:67’ list. Polymer sequence cleavage supports cross-links,
even with partial cleavages.

2.2.6 Refactored chemical modification framework
A lot of progress has been made in the characterization of biopolymer
chemical modifications and some modifications were found to be
much more frequent and relevant than initially thought (methylation
is a very important protein and DNA chemical modification). These advances prompted inclusion in massXpert 2 of a number of
compelling improvements in its chemical modification framework,
like being able to modify a given monomer in the sequence with
any number of same or different modifications. Note that provisions
set in the polymer chemistry definition ensure that the user does not
inadvertently perform unrealistic simulations.

3 CONCLUSION

massXpert 1 was developed during the late 90s. Ten years of
development and two rewrites later, massXpert 2 brings the
improvements required by the evolution of mass spectrometric
techniques. The most salient improvements are (i) the fact that
scientists can model new polymer chemistries, which are then
made available to the other modules in the software framework;
(ii) a powerful cross-link framework allows the simulation of
the most complicated situations; (iii) the sequence editor has
new capabilities such as multi-region selections; (iv) a redesigned
chemical modification framework ensures enough simulation
flexibility for the most complex cases; (v) pervasive handling of
multi-charged ions makes simulations more closely and reliably fit
mass spectrometric experimental data; (vi) the data-mining module
allows thorough comparisons of mass peak lists. Also of importance
is the fact that the software is cross-platform and that a detailed
(200 pp approximately) user manual in PDF format is available (see
Supplementary information).

Finally, while massXpert 1 was released as proprietary no-cost
software, the massXpert 2 software package is released under an
open source free software license (GPLv3) so as to encourage
collaborative development starting from current source code.

Conflict of Interest: none declared.

REFERENCES

data in proteomic projects by using massXpert. Bioinformatics, 18, 644-645.
In Mass Spectrometry Data Analysis in Proteomics. Humana Press, Totowa, New Jersey.
Strohalm,M. et al. (2008) mMass data miner: an open source alternative for mass
668-669.

Fig. 1. Sequence editor window, multi-region selection. The Kunitz inhibitor
is shown. One multi-region selection simulates a tryptic cross-linked entity
made of two peptides cross-linked by a disulfide bond. The computed masses
may optionally take the cross-links into account.

electrospray ionization context, which is characterized by multi-
charged ions.

2.2.4 (m/z,z) pair list-based data mining
The XpertMiner module features the ability to import mass data from outside
massXpert 2, typically from the mass spectrometer data acquisition
software, from spectrum display/analysis tools like mMass
(Strohalm et al., 2008) or from databases. The imported datasets
translate into (m/z,z) pair lists which are available for automated
comparison with similarly created lists from massXpert 2-based
simulations. Lists can be manipulated according to arbitrary
parameters (add/remove masses or formulas, for example), and
then used for peak-matching work with a configurable tolerance.
Handling (m/z,z) pairs instead of m/z single values makes the whole
data-mining process charge independent, which is essential with
arbitrarily charged ions. The XpertMiner module is operated using
simple drag-and-drop and copy/paste clipboard paradigms.

2.2.5 A completely new cross-link simulation framework
The new cross-link framework designed in massXpert 2 allows any
number of cross-links to be set to a polymer sequence. For example,
the Kunitz (1945) inhibitor—with its seven disulfide bonds—can be
easily modelled. Figure 1 shows the Kunitz inhibitor loaded in the
sequence editor with cross-linked cysteinyl residues displayed with
small chain links superimposed onto their vignette. The editor now
features a multiple-region selection mode that allows simulation of
cross-linked oligomers. When one or more sequence regions are
selected, which do not encompass all the entities cross-linked to
the selected sequence, the program alerts the user with a message
indicating the number of incomplete cross-links. Monomer cross-
linking is performed by choosing one cross-linker from the list
of available cross-linkers (in the current polymer definition) and
specifying the involved monomers as a semicolon-separated list of
monomer positions. Any cross-link might involve any number of
monomers. As an example, the very peculiar cross-links involved
in the formation of the fluorescent proteins’ chromophore are easily
simulated in massXpert 2. In this case, cross-linking the three cyan
fluorescent protein’s Thr, Trp and Gly residues taking their part in
the chromophore would involve specifying their positions as the
‘65:66:67’ list. Polymer sequence cleavage supports cross-links,
even with partial cleavages.

2.2.6 Refactored chemical modification framework
A lot of progress has been made in the characterization of biopolymer
chemical modifications and some modifications were found to be
much more frequent and relevant than initially thought (methylation
is a very important protein and DNA chemical modification). These advances prompted inclusion in massXpert 2 of a number of
compelling improvements in its chemical modification framework,
like being able to modify a given monomer in the sequence with
any number of same or different modifications. Note that provisions
set in the polymer chemistry definition ensure that the user does not
inadvertently perform unrealistic simulations.

3 CONCLUSION

massXpert 1 was developed during the late 90s. Ten years of
development and two rewrites later, massXpert 2 brings the
improvements required by the evolution of mass spectrometric
techniques. The most salient improvements are (i) the fact that
scientists can model new polymer chemistries, which are then
made available to the other modules in the software framework;
(ii) a powerful cross-link framework allows the simulation of
the most complicated situations; (iii) the sequence editor has
new capabilities such as multi-region selections; (iv) a redesigned
chemical modification framework ensures enough simulation
flexibility for the most complex cases; (v) pervasive handling of
multi-charged ions makes simulations more closely and reliably fit
mass spectrometric experimental data; (vi) the data-mining module
allows thorough comparisons of mass peak lists. Also of importance
is the fact that the software is cross-platform and that a detailed
(200 pp approximately) user manual in PDF format is available (see
Supplementary information).

Finally, while massXpert 1 was released as proprietary no-cost
software, the massXpert 2 software package is released under an
open source free software license (GPLv3) so as to encourage
collaborative development starting from current source code.

Conflict of Interest: none declared.

REFERENCES

data in proteomic projects by using massXpert. Bioinformatics, 18, 644-645.
In Mass Spectrometry Data Analysis in Proteomics. Humana Press, Totowa, New Jersey.
Strohalm,M. et al. (2008) mMass data miner: an open source alternative for mass
668-669.