**ABSTRACT**

Summary: PRODECOMP (projection decomposition) is an implementation of a multi-way decomposition algorithm for the analysis of two-dimensional projections of high-dimensional nuclear magnetic resonance spectra. The newest version, PRODECOMPv3, features a dramatic speedup, more reliable decompositions, a substantial reduction in memory demands, a new graphical user interface and integration into third-party software. These improvements extend the applicability of decompositions to novel types of NMR data on proteins, yielding backbone and side-chain assignments as well as structural information, and therewith enabling complete characterizations of proteins.

Availability: Program, short manual and an example calculation are freely available at www2.chem.gu.se/ccbpm/nmr/prodecomp.html

Contact: doroteya@chem.gu.se; martin.billeter@chem.gu.se

1 INTRODUCTION

Projection techniques in solution NMR (nuclear magnetic resonance) applied to proteins have the potential to reduce experiment times by several orders of magnitude, making the approach highly feasible for structure genomics or the functional investigation of intrinsically unstructured proteins. However, the interpretation of projection spectra presents some additional challenges. Multi-way analysis is a mathematical tool for characterizing higher order data with applications in many fields (Bro, 2006). Simultaneous decomposition of a set of NMR projections yields reliable solutions even with signal-to-noise ratios of the individual projections close to 1 (Malmodin and Billeter, 2005, 2006; Staykova et al., 2008). Implemented as a general tool for projection decomposition called PRODECOMP, this approach has a wide range of applications in NMR. Projections of high-dimensional NMR spectra of two proteins, ubiquitin (8.5 kDa) and azurin (14 kDa), were used to demonstrate its capabilities of decomposing NMR data involving heteronuclear coupling, TOCSY and NOESY magnetization transfers (Levitt, 2001). With the help of accompanying programs for processing the PRODECOMP output, automated resonance assignments for backbone and side chain atoms, and distance restraints for structure determinations were obtained. The algorithm is also incorporated into generally available packages such as CCPN,1 and interfaced to other software (TopSpin, Bruker Biospin GmbH).

2 MULTI-WAY DECOMPOSITIONS

Two-dimensional (2D) projections \( P_m(\omega_1, \omega_N) \) (projections enumerated by \( m \)) of a N-D spectrum \( S(\omega_1, \ldots, \omega_N) \) can be modeled as follows (Malmodin and Billeter, 2005, 2006):

\[
P_m(\omega_1, \omega_N) = \sum_k \left[ (F_k^1 \ast \ldots \ast F_k^{N-1})(\omega) \otimes F_k^N(\omega_N) \right] + \varepsilon \quad (1)
\]

where the symbols ‘∗’ and ‘⊗’ represent convolution operations and direct products. While the frequency parameter \( \omega_N \) is the same in the full spectrum \( S \) and the projections \( P_m \), the \( \omega \)-axis in \( P_m \) is a projection of the space spanned by the frequency axes \( \omega_1, \ldots, \omega_{N-1} \) in \( S \) (\( m \) refers to different linear combinations of frequencies on the left side which are reflected by different selections of \( F_k^1 \) on the right side). Terms in the sum enumerated by \( k \) are called components, and each component is characterized by 1D vectors \( F_k^1 \ldots F_k^N \), which are referred to as shapes. The latter characterize the resonances of all \( N \) nuclei present in the projections and thus provide the desired output; they can also be used to reconstruct the full-dimensional spectrum \( S \). The term \( \varepsilon \) is necessary to describe the difference due to noise or artifacts between the experimental input projections \( P_m \) on the left side of (1) and the model on the right side. Decomposition involves determining all shapes \( F_k^i \) by minimizing \( \varepsilon \).

3 RESULTS

The original implementation of multi-way decomposition according to (1), based on the FNLNS algorithm (Bro and De Jong, 1997), demonstrated the feasibility of the approach when applied to NMR projection data (PRODECOMPv1; Malmodin and Billeter, 2005, 2006). Code modifications led to better convergence rates and more reliable components in crowded regions (PRODECOMPv2), which allowed obtaining complete backbone resonance assignments of ubiquitin using a combined set of projections from two 5D experiments yielding 9D components (Staykova et al., 2008). This application revealed limitations caused by huge memory requirements with large number of components or with increased spectral size. The ensuing improvements (PRODECOMPv3) not only removed the memory barrier, but also yielded a significant speedup in calculation time.

3.1 Algorithm improvements

A detailed analysis of the original algorithm revealed two major bottlenecks: scaling discrepancies between input data and trial shapes, and over-dimensioned temporary matrices.
Several modifications were implemented in the current version of PRODECOMP:

- Normalizing of all input data substantially improved the convergence and removed ambiguities in the resulting shapes, revealing several previously missing components.
- Large sparse matrices in the code were replaced by element-wise tracing, multiplication and buildup of corresponding data.
- Among other improvements is also a definition of input and output formats coupled with conversion routines.

While the original version missed one of the six components in the test calculations of Figure 1 (white bars), scaling of input data resulted in observation of all components (gray bars). (Note that decomposition failures result in missing components rather than false positives.) Better memory management led to two important improvements: (a) a massive speedup in runtime (black bars) and (b) the possibility to decompose projections with higher resolution, e.g. increasing the number of points along the $\omega_N$-axis from 121 to 256 or even 512. The modified decomposition algorithm, implemented in PRODECOMPv3, was successfully tested on projected spectra based on heteronuclear coupling, TOCSY and NOESY experiments. Note that the increase in resolution is crucial for the latter two types of experiments.

### 3.2 Graphical user interface

The graphical user interface (GUI) of Figure 2 was designed to facilitate the input of projections and the choice of run-time parameters: selection of an interval along the $\omega_N$ dimension, number of expected components in this interval, a regularization factor and the number of iterations to be performed (Staykova et al., 2008). Projections in TopSpin format or from other sources via CEP format (top right panel: selected projections from a TOCSY-type experiment). Decomposition output is stored in XML/CCPN and ASCII format and visualized. The example in the lower right panel shows six components with strongly overlapping signals in $\omega_N$. In this experiment, each component corresponds to one protein residue; one component is highlighted by black lines (see www2.chem.gu.se/bcbp/nmr/prodecomp.html for individual plots of all components).

### 4 IMPLEMENTATION

PRODECOMPv3 is written in Python as a freely available software library. The decomposition function is being integrated within the CCPN model (proposed standard for data exchange between different programs and databases in macromolecular NMR) and TopSpin (instrument software by Bruker Biospin GmbH for recording and processing of experiments). PRODECOMPv3 has been tested on Windows, different Linux environments and Mac OS using the same code.

### ACKNOWLEDGEMENTS

We would like to thank Wolfgang Bermel from Bruker Biospin GmbH for implementing the NMR experiments.

**Funding:** EU (LSHG-CT-2005-018988).

**Conflict of Interest:** none declared.

### REFERENCES