Journal of Biomolecular NMR, **20:** 49–60, 2001.
KLUWER/ESCOM

MUNIN: A new approach to multi-dimensional NMR spectra interpretation

Vladislav Yu. Orekhov[a,*], Ilghiz V. Ibraghimov[b] & Martin Billeter[c]
[a]*Swedish NMR Centre at Göteborg University, Box 465, S-405 30 Göteborg, Sweden*
[b]*Saarbrücken University, Mathematical Department, D-66041 Saarbrücken, Germany*
[c]*Göteborg University, Biochemistry & Biophysics, Box 462, S-405 30 Göteborg, Sweden*

Received 30 November 2000; Accepted 5 March 2001

Key words: canonical decomposition, ^{1}H-^{15}N-NOESY-HSQC, non-uniform sampling, PARAFAC, parallel factor analysis, signal processing, three-way decomposition

Abstract

A new method, MUNIN (Multi-dimensional NMR spectra interpretation), is introduced for the automated interpretation of three-dimensional NMR spectra. It is based on a mathematical concept referred to as *three-way decomposition.* An NMR spectrum is decomposed into a sum of *components*, with each component corresponding to one or a group of peaks. Each component is defined as the direct product of three one-dimensional *shapes*. A consequence is reduction in dimensionality of the spectral data used in further analysis. The decomposition may be applied to frequency-domain or time-domain data, or to a mixture of these. Features of MUNIN include good resolution in crowded regions and the absence of assumptions about line shapes. Uniform sampling of time-domain data, a prerequisite for discrete Fourier transform, is not required. This opens an avenue for the processing of NMR data that do not follow oscillating behaviour, e.g. from relaxation measurements. The application of MUNIN is illustrated for a ^{1}H-^{15}N-NOESY-HSQC, where each component is defined as the set of all NOE peaks formed by a given amide group. As a result, the extraction of structural information simply consists of one-dimensional peak picking of the shape along the NOE-axis obtained for each amide group.

Abbreviations: MUNIN, multi-dimensional NMR spectra interpretation; DFT, discrete Fourier transform; PARAFAC, parallel factor analysis; NOESY, nuclear Overhauser enhanced spectroscopy; ^{1}H-^{15}N-HSQC, ^{15}N-based heteronuclear single quantum spectroscopy; 3D ^{1}H-^{15}N-NOESY-HSQC, three-dimensional combination of NOESY and ^{1}H-^{15}N-HSQC; E.COSY, exclusive correlated spectroscopy.

Introduction

MUNIN is an implementation of the concept that the interpretation of spectroscopic data is equivalent to achieving a separation of the spectral signals, and to cope with spectral noise. In terms of NMR spectroscopy, this concept translates into the decomposition of an experimental input consisting of raw time-domain data into components that describe single peaks or groups of related peaks. Separation of peaks in an NMR spectrum relies on their identification or, in other words, on the estimation of frequencies, amplitudes and other features. Conventionally, this goal is approached in two steps. First, discrete Fourier transform (DFT) is applied to the raw time-domain NMR data. Next, interactive or automated peak picking procedures are used to identify the individual signals. DFT, although being a well-established, fast and robust transformation, has several drawbacks when applied to multi-dimensional NMR spectroscopy. Input for DFT should be sampled at regular time intervals, which precludes optimal data sampling with respect to maximising both sensitivity and resolution for a given duration of an NMR experiment. Obviously, the

*To whom correspondence should be addressed. E-mail: orov@nmr.se

step from time-domain data to a (frequency-domain) spectrum, using DFT or other types of transforms, is most powerful when frequency differences are the major criteria for peak separation. However, other signal properties such as line shapes may not profit from Fourier transform. For signals that do not follow an oscillating behaviour, e.g. that are modulated mainly by relaxation decays or by J-coupling, DFT is almost useless.

MUNIN uses the mathematical concept of *three-way decomposition* (Carroll and Chang, 1970; Harshman, 1970) and a simple model for NMR spectra that is based on generally accepted assumptions to achieve high resolution and good sensitivity while avoiding artefacts. Though not widely known in NMR spectroscopy, the mathematical approach used in MUNIN has been applied for some time in fields like chemometrics or psychometrics (Harshman and Lundy, 1984). Three-way decomposition is also referred to as *parallel factor analysis* (PARAFAC; for a comprehensive description see Bro, 1997) or *canonical decomposition* (Carroll and Pruzansky, 1984). As a multi-linear model, three-way decomposition provides unambiguous results only for three- (as presented below) or higher-dimensional data sets. MUNIN separates the signals in all dimensions *simultaneously*. Most other methods such as DFT, wavelet transforms (Weaver et al., 1992; Barache et al., 1997), Bayesian and maximum likelihood techniques (Bretthorst, 1990; Rouh et al., 1994; Chylla and Markley, 1995; Kotyk et al., 1995; Ochs et al., 1999), maximum entropy methods (Hoch and Stern, 1996; Schmieder et al., 1997), linear prediction (Koehl, 1999) or the recently introduced filter diagonalisation (Hu et al., 2000) process in practice only one- or two-dimensional data sets at any given time. This complicates a self-consistent analysis of multi-dimensional spectra. Thus, MUNIN can be regarded as a high-dimensional complement to traditional spectral processing techniques, where dimensions are treated one at a time. It can also be thought of as an efficient method for reducing the dimensionality of multidimensional data sets. Namely, as a result of MUNIN processing, one obtains sets of one-dimensional objects, which can be more easily interpreted than a full 3D spectrum.

Deconvolution methods, which were used to resolve crowded regions of two-dimensional spectra, differ from the presently described three-way decomposition by their use of a priori known lineshapes and chemical shifts (Denk et al., 1986; Koradi et al., 1998). To our knowledge, three-way decomposition was never used for the analysis of NMR data, though applications of related ideas were reported (Abergel and Delsuc, 1993; Stilbs et al., 1996; Windig and Antalek, 1999). Here we present the new method, MUNIN, and two examples of its application to a 3D ^{1}H-^{15}N-NOESY-HSQC spectrum of the 14 kDa protein azurin (Karlsson et al., 1989; Van de Kamp et al., 1992). The first example illustrates computational feasibility and efficiency in resolving spectral components in the most crowded region of the spectrum. In the second example, calculations are performed on a subset of the NMR raw data that corresponds to non-uniform sampling in time along the ^{15}N dimension. It is shown that the same structural information can be obtained using only 20% of the original experimental data.

Materials and methods

Theory

Line shapes in all dimensions together with intensities can describe the signals, i.e. the peaks, of most multi-dimensional NMR spectra. In mathematical notation, NMR signals can thus be defined as direct products of one-dimensional entities. A spectrum becomes the sum over all signals, to which an error term may be added. Note that the term 'spectrum' is used here to describe time-domain NMR data, Fourier transformed frequency-domain data or a mixture of the two with Fourier transform applied only to selected dimensions. We restrict the following discussion to the three-dimensional case, but extension to higher dimensions is straightforward. The following equation can be thought of as a model to fit the experimental spectrum:

$$S_{i,j,k} = \sum_{m=1}^{R} A_m^m \cdot F1_i^m \cdot F2_j^m \cdot F3_k^m \quad + \quad e_{i,j,k} \qquad (1)$$

Here, the three-dimensional experimental data matrix S with size (I, J, K) is written as intensities for each point, $S_{i,j,k}$, with $i = 1,2\ldots I$; $j = 1,2\ldots J$; $k = 1,2\ldots K$. We will refer to the R terms in the sum as *components*; these may, but need not, correspond to cross peaks. The one-dimensional functions $F1^m$, $F2^m$ and $F3^m$ will be called *shapes,* which may, but need not, correspond to line shapes of peaks. The term *shape* is introduced here in relation with spectral line shapes; several synonyms can be found in the literature, e.g. loads, modes, etc. All shapes are

normalised, and A is a diagonal matrix containing the intensities of the R components. The function $e_{i,j,k}$ contains noise and artefacts, or more generally residual errors due to incomplete fitting of the experimental data by the components. In the absence of a priori constraints on the signals one has $R\cdot(I+J+K-2)$ unknowns corresponding to the normalised shapes with intensities. The system consists of $I \cdot J \cdot K$ measurements and is therefore over-defined for typical values of I, J, K and R, which means that one can formulate a least-square problem to obtain estimates of the unknowns.

A component may correspond to a cross peak in a Fourier transformed multi-dimensional spectrum. In this case the shapes are line shapes of the peak in all dimensions. However, it is important to emphasise that the components of Equation 1 do not always have a one-to-one correspondence to peaks. Thus, in an E.COSY spectrum several components are needed for the description of a single cross peak. In other cases, as in the examples discussed below, one component may represent several cross peaks. Note also that no assumption about the form of the shapes is made in Equation 1. Furthermore, if Equation 1 holds for shapes in frequency-domain, it will as a consequence of the linearity of the Fourier transform also hold for corresponding shapes in time-domain.

An important aspect of parallel factor analysis is the uniqueness of the solutions. In general, uniqueness can only be obtained for problems involving three or more dimensions (Kruskal, 1977, 1989; Sidiropoulos and Bro, 2000). For a specific discussion of MUNIN applied to 3D NMR spectra we define that the shapes of two components for a given dimension are different when their dot (or inner) product has an absolute value (significantly) smaller than 1. Note that shapes can be different even if their frequency maxima are identical, provided their line shapes differ (e.g. line widths or multiplet structure). For practical NMR applications (although not very strict in mathematical terms) the following three situations have to be distinguished, when considering the uniqueness of two components: (i) if the shapes are different in all three dimensions, the solution is unique; (ii) if for two dimensions the corresponding shapes coincide, the two components can be substituted by a single component without changing the residual of the fit. In practical calculations one always obtains a single component for this case. Thus, all NOE cross peaks to a given amide group in a ^{1}H-^{15}N-NOESY-HSQC have the same line shapes in two dimensions ($^{1}H_N$, ^{15}N), and they are consequently described by a single component; (iii) if exactly one shape is identical, we refer to the two components as being *mixed*. This situation corresponds to the two-dimensional problem discussed in the following.

For the two-dimensional case (only) Equation 1 can be rewritten in a matrix form:

$$\mathbf{X} = \mathbf{F1} \cdot \mathbf{A} \cdot \mathbf{F2}^T \tag{2}$$

where $\mathbf{X}$ is the reconstructed 2D spectrum of size $I \times J$ obtained by summation of the R components, $\mathbf{F1}$ and $\mathbf{F2}$ are shape matrices with dimensions $I \times R$ and $J \times R$, and $\mathbf{A}$ is the diagonal $R \times R$ matrix containing the amplitudes of the components. As is shown below, the model defined by Equation 2 does not result in a unique solution due to the so-called rotational ambiguity of the solution in two dimensions. For a given set of components we can rewrite Equation 2:

$$\mathbf{X} = \mathbf{F1} \cdot \mathbf{A} \cdot \mathbf{U}^{-1} \cdot \mathbf{U} \cdot \mathbf{F2}^T \tag{3}$$

where $\mathbf{U}$ is an arbitrary non-singular $R \times R$ matrix and $\mathbf{U}^{-1}$ is its inverse matrix. By defining new shape matrices $\tilde{\mathbf{F1}}$ and $\tilde{\mathbf{F2}}$

$$\tilde{\mathbf{F1}} = \mathbf{F1} \cdot \mathbf{A} \cdot \mathbf{U}^{-1} \text{ and } \tilde{\mathbf{F2}}^T = \mathbf{U} \cdot \mathbf{F2}^T \tag{4}$$

and, after normalization of these and moving the normalisation factors into a newly constructed diagonal matrix $\hat{\mathbf{A}}$, one obtains:

$$\mathbf{X} = \hat{\mathbf{F1}} \cdot \hat{\mathbf{A}} \cdot \hat{\mathbf{F2}}^T \tag{5}$$

Equation 5 describes the same reconstructed spectrum as Equation 2 using new components defined by the shapes $\hat{\mathbf{F1}}$ and $\hat{\mathbf{F2}}$ and intensities $\hat{\mathbf{A}}$. For the case $R = 2$, i.e. for the case of two mixed components, the matrix $\mathbf{U}$ can, without loss of generality, be written as:

$$\mathbf{U} = \begin{bmatrix} \cos(\alpha) & \sin(\alpha) \\ \cos(\beta) & \sin(\beta) \end{bmatrix} \tag{6}$$

Arbitrary linear combinations of the two components are defined by the two angles α and β. Equation 6 can easily be generalised for mixing of more than two components, e.g. by constructing $\mathbf{U}$ such that its rows represent different unit vectors written in polar co-ordinates of the corresponding dimensionality.

The ambiguity of the model given by Equation 2 can be eliminated, if additional constraints on the shapes in one or both dimensions are imposed. Orthogonality of the shapes is postulated when the solution is computed using singular value decomposition. An alternative method was proposed (Stilbs et al., 1996) for the analysis of 2D data sets obtained in

NMR diffusion experiments. Uniqueness of the resulting components was achieved by assuming known functional dependences (exponential) for the shapes in one of the dimensions, with the second dimension corresponding to unknown 1D spectra of compounds in a mixture. Usage of non-negativity constraints on the shapes was reported in the field of NMR imaging (Ochs et al., 1999). Symmetry constraints have also been used (Abergel and Delsuc, 1993; Steinbock et al., 1997). These or other constraints can be used to resolve the problem of *mixing* of the components, after or even during the MUNIN calculations on 3D NMR data. If the number of mixed components is small, e.g. two or three, manual procedures also work well (Stoyanova et al., 1995).

NMR spectroscopy

A gradient sensitivity-enhanced 3D ^{1}H-^{15}N-NOESY-HSQC spectrum (Zhang et al., 1994) was recorded for a 1 mM sample of reduced Azurin in potassium-phosphate buffer (T = 15 °C, pH 5). The spectrum was obtained on a 500 MHz Varian UNITY *Inova* spectrometer with 800, 160 and 44 complex points in the $^{1}H_N$, $^{1}H_{NOE}$ and ^{15}N dimensions, respectively. The spectral widths for these three dimensions were 8000, 8000, and 1810 Hz, respectively. A mixing time of 75 ms was used in the NOESY part of the experiment. The total recording time of the experiment was 47 h.

Fourier transforms and construction of 3D working data arrays

All 88 two-dimensional $^{1}H_N$-$^{1}H_{NOE}$ planes forming a time domain interferogram in the ^{15}N dimension were Fourier transformed, after application of square sine-bell weighting functions in each dimension, one by one using the VNMR software (version 6.1b, Varian Associates, Inc.). To reduce the amount of data and the number of components in the MUNIN calculations, rectangular regions from 8.11 to 8.36 ppm and from −0.5 to 9.23 ppm in the $^{1}H_N$ and $^{1}H_{NOE}$ directions, respectively, were extracted from each plane. The position of the rectangle was chosen to cover the most crowded region of the spectrum along the amide proton dimension corresponding to 22–27 amide signals (the exact number cannot be determined due to signal overlap, presence of weak peaks as well as tails from signals located outside the region). In the NOE dimension only empty flanking regions were removed. The set of the rectangle matrices extracted from all the $^{1}H_N$-$^{1}H_{NOE}$ planes constitutes a three-dimensional array with dimensions 33, 311 and 88 points in the $^{1}H_N$, $^{1}H_{NOE}$, and ^{15}N dimensions, respectively. Note that only the $^{1}H_N$ and $^{1}H_{NOE}$ dimensions were Fourier transformed, while the ^{15}N dimension remained in the time domain. For illustration purposes only (spectrum reconstruction and projections of $^{1}H_N$-^{15}N planes), 1D Fourier transforms in the ^{15}N dimension were performed on shapes obtained by the MUNIN calculations.

For calculations on an irregularly sampled data set, 70 of the 88 $^{1}H_N$-$^{1}H_{NOE}$ planes were selected randomly and removed. The remaining 18 planes (3, 9, 10, 16, 18, 23, 26, 45, 47, 51, 55, 56, 57, 61, 62, 66, 69, 73; odd and even numbers correspond to the real and imaginary planes, respectively), representing about 20% of the data, were collected in a new input data set of size 33 × 311 × 18 points, in the following referred to as the reduced data set. For this set Fourier transform is not feasible in the nitrogen dimension before or after MUNIN.

MUNIN calculations, resolving mixed components

MUNIN calculations were performed using home-made software, which comprised the following steps: (i) data preparation as described in the previous section; (ii) three-way decomposition (Ibraghimov, 1999); (iii) de-mixing of the resulting components when necessary; and (iv) reconstructions of projections of the three-dimensional spectrum for illustrations. Three-way decomposition was performed according to a known protocol (Bro, 1997; Hopke et al., 1998), i.e. minimisation of the mean square of residuals in the model given by Equation 1 for a specified number of components (see the Appendix). First, using a method referred to as 'Tucker3' (Tucker, 1966; Kroonenberg and Leeuw, 1980; Andersson and Bro, 1998), the input three-dimensional data arrays were compressed to 33 × 32 × 32 and 32 × 32 × 18 3D matrices for the full and reduced data sets, respectively. This compression yielded a dramatic speed-up of the subsequent calculations. The model given by Equation 1 for the compressed arrays was solved by alternative least-squares iterations (Harshman and Lundy, 1984). The resulting components were used, after un-compression, for the initiation of alternative least-squares iterations of the original data. Convergence was achieved in 15–20 min CPU time (SGI Octane, R10000 250 MHz processor) for 27 components on the full and reduced data sets.

Possible mixing of components was checked by calculating pairwise dot products of the resulting shapes (written as vectors). Two components were

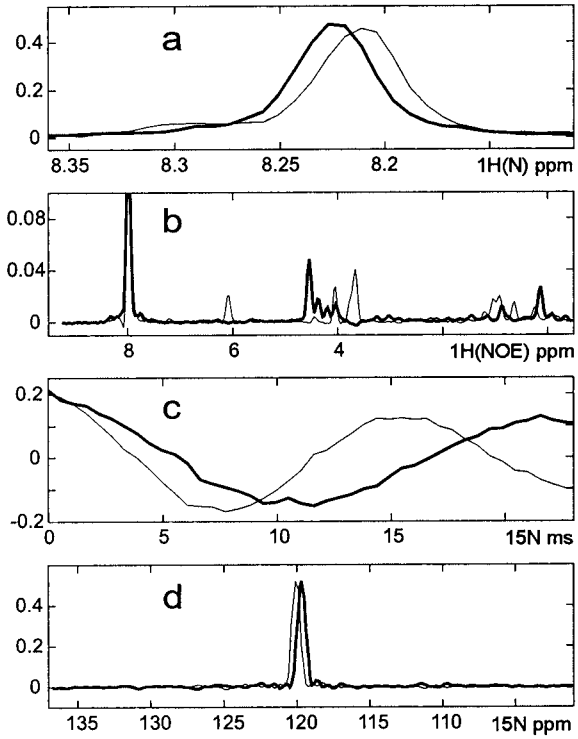

Figure 1. Normalised one-dimensional shapes of 2 (out of 27) MUNIN components referred to as component 1 (heavy lines) and component 2 (thin lines). Shapes are along the dimensions of (a) $^{1}H_{N}$ (amide proton), (b) $^{1}H_{NOE}$, and (c) ^{15}N. (d) The same shapes as in (c), but after Fourier transform. Arbitrary units are used for the vertical axis.

considered mixed if the absolute value of the dot product between their shapes in any of the three dimensions was higher than 0.95. If at least one of the two mixed components was mixed with a third component, all three components were collected in a group of mixed components. More components could be added to a group in the same way. Mixing of pairs of components was resolved (de-mixed) by manual adjustments of the angles α and β of Equation 6. No attempts were made to de-mix groups containing more than two components.

Results and discussion

Decomposition of a ^{1}H-^{15}N 3D NOESY-HSQC spectrum

A spectral region of a ^{1}H-^{15}N 3D NOESY-HSQC spectrum, as described in Methods, was decomposed into 27 components using the MUNIN procedure. In this type of spectrum all NOESY cross peaks stemming from a given amide proton are collected in a single component (since the shapes in both the ^{15}N and $^{1}H_{N}$ dimensions are identical). Figure 1 presents 2 of the 27 components, in the following referred to as *component 1* (heavy lines) and *component 2* (thin lines). Panels a–c of Figure 1 show the direct output of MUNIN, namely the one-dimensional shapes of the two components. The shapes in Figure 1a correspond to the line shapes of amide protons in the directly detected dimension. Each shape shown in Figure 1b, displaying the $^{1}H_{NOE}$ dimension, contains all the NOEs and the diagonal peak formed by the corresponding amide proton defined in Figure 1a. Finally, the cosine modulations in Figure 1c represent the shapes in the ^{15}N dimension. They correspond to time domain signals, since no Fourier transform was performed in this dimension prior to the MUNIN calculation. For presentation purposes, Figure 1d shows the ^{15}N shapes in the frequency domain after one-dimensional DFT of the shapes shown in Figure 1c.

All structural information of the NOESY-HSQC spectrum is found in the NOE shapes of Figure 1b, making a detailed analysis of the other dimensions unnecessary. With respect to the ^{15}N dimension data can be left in the time domain, avoiding the need for adjustment of phases or weighting functions; one may even skip decoupling in this dimension. As soon as components are resolved one also gets resolved diagonal signals. These signals, although they do not carry direct structural information, can be very useful for the calibration of the intensities of the NOE cross peaks, which may be affected by the different efficiencies of the HSQC step of the experiment. Resolution of components occurs simultaneously in all three dimensions. Therefore, the good characterisation and the high signal-to-noise ratios of the $^{1}H_{N}$ and ^{15}N shapes (Figure 1, a and c) can also be expected in the NOE dimension. From qualitative visual analysis we conclude that the signal-to-noise ratios of the NOE peaks in Figure 1c are at least as high as those obtained in the spectrum processed using regular DFT.

^{1}H-^{15}N projections from the ^{1}H-^{15}N 3D NOESY-HSQC and one of the $^{1}H_{N}$-$^{1}H_{NOE}$ NOESY planes from this spectrum are shown in Figures 2 and 3. The upper plots (Figure 2a and 3a) represent reconstructions using all 27 components resulting from the MUNIN calculations. Below, corresponding plots were made using the 3D spectrum obtained after regular DFT (Figures 2b and 3b). Clearly, the spectrum reconstructed from the MUNIN components and the Fourier transformed spectrum look very similar, which proves that the MUNIN calculations converged to a

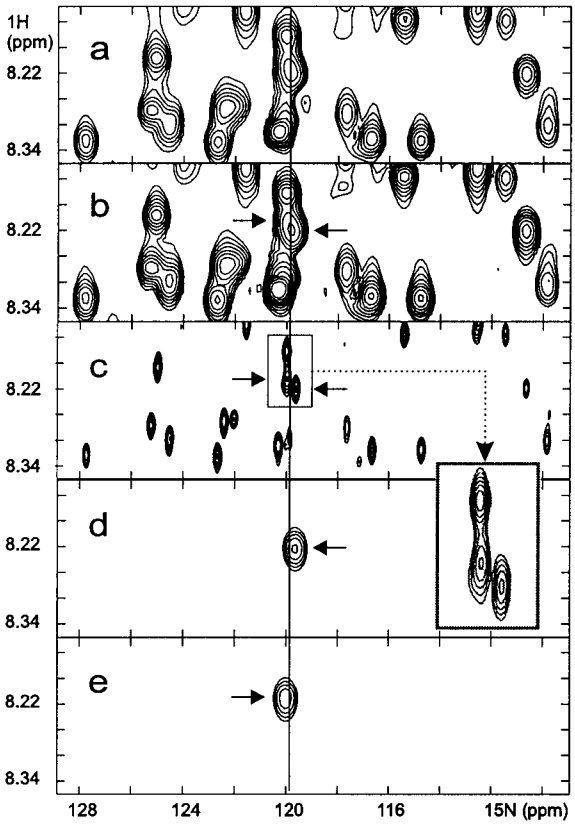

Figure 2. HSQC projections from the $^{1}H_{N}$-^{15}N 3D NOESY-HSQC spectrum. (a) Projection of the reconstructed 3D spectrum using all 27 MUNIN components. (b) Projection of the spectrum obtained with conventional Fourier transform. (c) Corresponding region from a $^{1}H_{N}$-^{15}N 2D HSQC spectrum recorded with four times longer acquisition time in the ^{15}N dimension than in the 3D spectrum. The relevant part of the spectrum is framed and magnified in the inset. (d, e) Projections of the reconstructed spectra using only component 1 or 2, respectively. The positions of the cross peaks corresponding to components 1 and 2 are indicated by arrows. The vertical line indicates the position in the ^{15}N dimension of the plane shown in Figure 3.

good solution. The difference matrix (not shown) between the spectra of Figure 3, a and b, yields only one or two contour levels at the positions of the strong diagonal signals. No regularities were observed in this difference spectrum, even when plotted at lower contour levels. One may conclude that the MUNIN procedure does not distort relative intensities of the signals, which may occur e.g. in maximum entropy reconstructions (Hoch and Stern, 1996). The slightly larger discrepancies between the HSQC projections (Figure 2, a and b) most probably indicate small differences in the baselines of the reconstructed and the Fourier transformed spectra. Minor baseline drifts can be expected, e.g. because of the distortion in the first point in time domain, and they can sum up to visible amplitudes in projections.

Though convergence to a global minimum is not guaranteed for the iterative approach used in MUNIN (Henrion, 2000), we always found good solutions in our calculations. Two other calculations for the same spectral region were performed (not shown) with the spectrum being Fourier transformed prior to the MUNIN calculations in either only one dimension ($^{1}H_{N}$) or in all three dimensions. Though the points in the input 3D arrays were weighted differently in these calculations due to application of weighting functions prior to DFT, the resulting sets of components were the same in all three cases. This indicates a good robustness of the MUNIN approach.

Resolving overlapped peaks

An obvious advantage of MUNIN is its ability to resolve overlapped signals, provided that their shapes are different. Confidence in the presence of two signals corresponding to components 1 and 2 is gained by the clean lines for each component in Figures 1a and 1c, and more importantly by the absence of 'cross-talk' in Figure 1b, i.e. most NOEs show up for only one component: with only one NMR signal giving rise to both components 1 and 2, the NOEs would have the same amide group as origin, and their intensities would be distributed among both components. For an independent test of resolution, the region of a $^{1}H_{N}$-^{15}N 2D HSQC corresponding to the projections of Figures 2a and 2b is shown in Figure 2c. The acquisition time in the ^{15}N dimension for this 2D spectrum was four times longer than for the 3D spectrum. At the positions marked by arrows and zoomed in the inset, two cross peaks are clearly distinguishable in the 2D spectrum, but not in the projections from the 3D spectrum. However, in the MUNIN calculations, these signals resulted as the individual components 1 and 2, i.e. they were completely resolved. Reconstruction (and subsequent projection) can obviously be restricted to selected components only. Reconstructions using only component 1 or component 2 are shown in Figures 2d and 2e. Similarly, the $^{1}H_{N}$-$^{1}H_{NOE}$ planes of Figures 3c and 3d were obtained by reconstruction using only component 1 or component 2, respectively. The substantial overlap of these two (and other) components in Figures 3a and 3b is completely resolved. It should be pointed out that the same structural information as presented in a 'traditional' way in Figure 3 is more

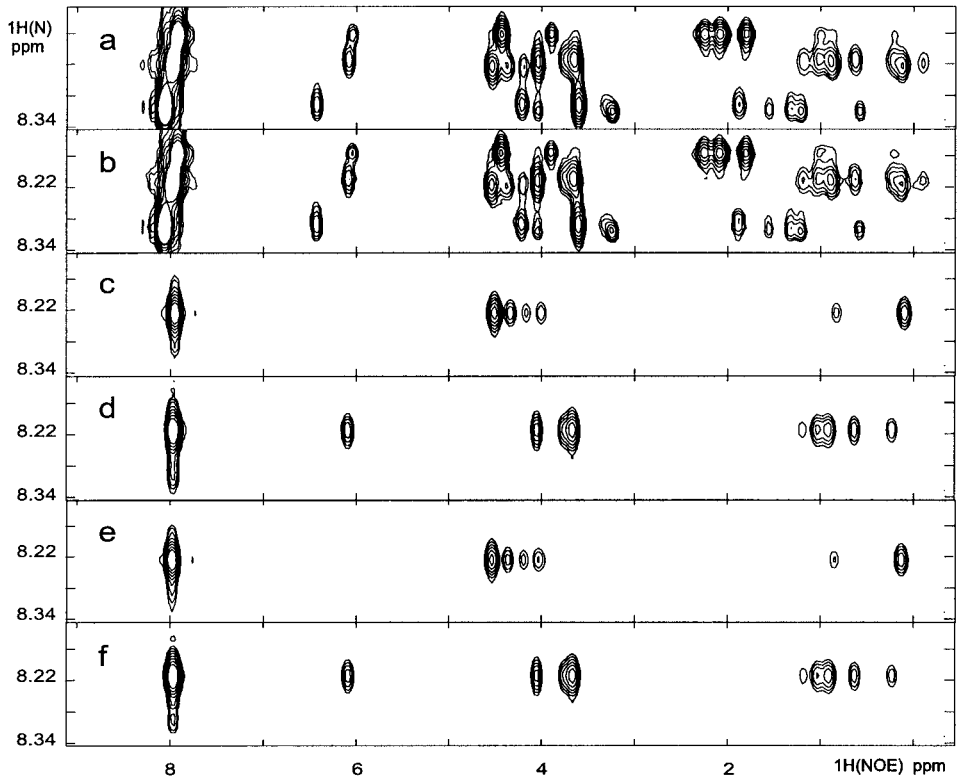

Figure 3. $^{1}H_{N}$-$^{1}H_{NOE}$ plane from the ^{1}H-^{15}N 3D NOESY-HSQC spectrum. The position of the plane in the ^{15}N dimension is indicated in Figure 2. (a) Reconstruction using all 27 MUNIN components. (b) Spectrum obtained by conventional Fourier transform. (c, d) Reconstructions using component 1 or 2, respectively. (e, f) Reconstructions using component 1′ or 2′, respectively, obtained in calculations using only 20% of the original experimental data.

easily obtained from the 1D $^{1}H_{NOE}$ shapes shown in Figure 1b.

Mixed components

As was discussed in the theoretical part, the solution of the MUNIN procedure is not unique if two or more components have very similar shapes in one of the dimensions. In Figure 4 we present a typical example of mixing observed between component 1 and a new component named component 3. Here, as in most cases, the mixing is caused by the degeneracy of the $^{1}H_{N}$ shapes. The most obvious manifestation of the component mixing seen in Figure 4 is the presence of more than one peak in the ^{15}N shapes (Figure 4f) and negative signals in the $^{1}H_{NOE}$ direction (Figure 4b). We observe mixing if the dot product between the normalized shapes is higher than ca. 0.95. The total set of 27 MUNIN components can be divided into 15 groups. The first eight groups contain individual non-mixed components. Three groups consist of pairs of components degenerate in amide proton shapes, and one group with two components is degenerate in nitrogen shapes. There is one group containing three components, and two groups contain four components, all with similarity in the $^{1}H_{N}$ dimension. It never occurs that two components have exactly the same pattern of peaks in the $^{1}H_{NOE}$ direction.

To simplify the spectral analysis it can be desirable to resolve the ambiguity within the groups or, in other words, to find pure, non-mixed components. To solve this problem one needs to introduce further constraints, based on a priori information available for a particular spectrum. The two components shown in Figure 4 were resolved by ensuring non-negativity of the $^{1}H_{NOE}$ shapes and/or the presence of a single peak in the ^{15}N shapes. Proper linear combinations (Equations 2–6) were obtained by manual adjustment of the angles α and β defined in Equation 6. The pure, non-mixed shapes are shown in Figure 4, c, e and g. It should be mentioned that each of the two angles α and β in Equation 6 affects only one of the two components when looking at a single dimension.

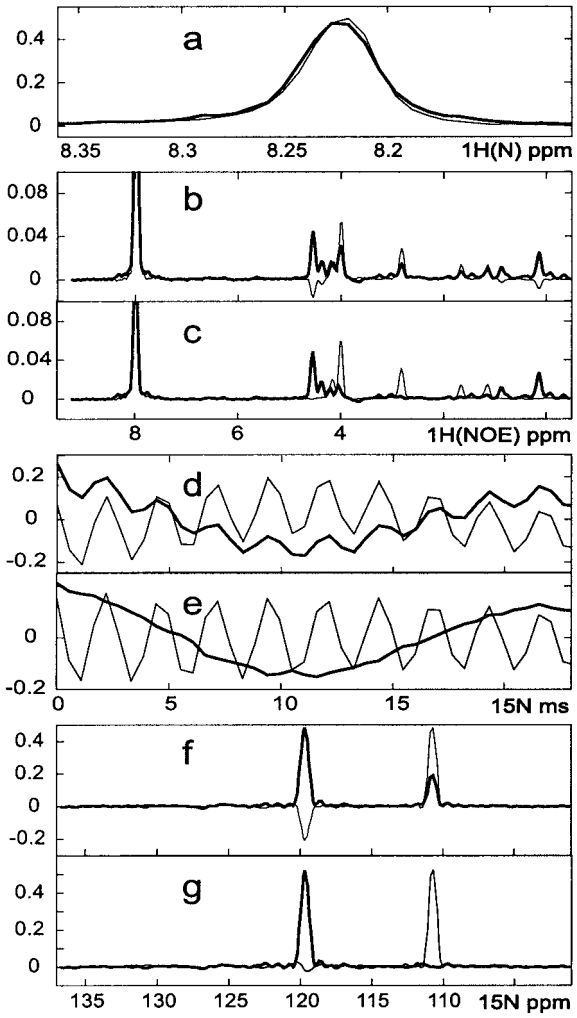

Figure 4. Mixing between the components 1 (heavy lines) and 3 (thin lines). (a) Normalised amide proton shapes, which 'overlap' with a dot product of 0.97. (b, d) $^{1}H_{NOE}$, and ^{15}N shapes, respectively; raw MUNIN output. (c, e) $^{1}H_{NOE}$ and ^{15}N shapes, respectively, of components 1 and 3 after manual de-mixing. (f, g) Fourier transformed ^{15}N shapes before and after de-mixing, respectively. Arbitrary units are used for the vertical axis.

Thus, if the constraints are defined only for one of the dimensions (e.g. ^{15}N), the angles become independent and can be tuned separately. The reduction of the cross-talk between the shapes after adjustment of the angles α and β is clearly distinguishable also for the time domain nitrogen shapes (Figures 4d and 4e). Although in many cases the automation of the de-mixing procedure is straightforward, we leave the consideration of the problem for a later publication, where the MUNIN processing of a complete 3D spectrum will be presented.

Strictly speaking, the raw result of our calculation can be thought of as a decomposition of the investigated spectral region into seven (mixed groups) plus eight (non-mixed components) independent subspectra. While individual components from a mixed group may not provide an adequate picture of spectral features, the sum of all components from this group does, and can thus be used for reconstruction of subspectra and their analysis. Even without de-mixing this decomposition brings significant simplification to the analysis, since mixing of the components does not imply that they necessarily overlap in the reconstructed 3D spectrum. For example, the components 1 and 3 shown in Figure 4 are mixed due to the degeneracy of the amide proton shapes (Figure 4a), but they are perfectly resolved in the reconstructed spectrum, because of the difference in nitrogen chemical shifts.

Reduced data set, non-uniform sampling

The main motivation for doing the ^{1}H-^{15}N 3D NOESY-HSQC experiment instead of a 2D NOESY is to achieve better signal resolution by spreading the NOE peaks from different amides along the ^{15}N dimension. Since structural information is coded only in the proton dimensions one could pay less attention to the nitrogen signals. A nice feature of the MUNIN decomposition is that the shapes can be considered totally independent from each other. This means that for the 'dummy' ^{15}N dimensions in the ^{1}H-^{15}N 3D NOESY-HSQC spectrum, one does not need a sampling procedure appropriate for phase sensitive Fourier transform. Instead, the sampled points could be unevenly distributed in the time domain to achieve optimal sensitivity and separation of the components. The resulting shapes corresponding to non-regular sampling in the time domain could be disregarded in the spectrum analysis. Alternatively, if one needs the positions of the nitrogen signals, these one-dimensional time series can be used for spectral estimations using other appropriate methods, e.g. maximum entropy.

Results of calculations on unevenly sampled data in the ^{15}N dimension are presented in Figure 3. Out of the 88 available NOESY planes, 18 were randomly selected and used for these calculations. Similar to the MUNIN calculations performed on the complete data set, 27 components were obtained. Two components, which are referenced below as $1'$ and $2'$, were identified to correspond to the components 1 and 2 defined earlier. $^{1}H_{N}$-$^{1}H_{NOE}$ planes from the 3D reconstruction using only component $1'$ or component $2'$ are shown in Figure 3, e and f, respectively, and can be com-

pared with their counterparts obtained from the full data set (Figure 3, c and d). It can be clearly seen that the reconstructions for the components obtained using the full and the reduced data sets are very similar and consequently carry the same structural information. Close inspection of the $^1H_{NOE}$ shapes (not shown) revealed slightly higher noise levels in the shapes of components $1'$ and $2'$ than in those of components 1 and 2. However, this can be expected since the anticipated acquisition time of the simulated experiment producing the reduced data set is almost five times smaller than the time of the original 3D experiment. It should also be noted that the number of components in this calculation is close to the theoretical limit imposed by the uniqueness condition for the three-way decomposition (Kruskal, 1977). Again, the similarity of the components 1 and $1'$, and 2 and $2'$, respectively, demonstrates the robustness of the MUNIN procedure and its tolerance with respect to spectral noise. It should be emphasised that the cross peaks corresponding to the components $1'$ and $2'$ were already not resolved when using the full data set processed by conventional Fourier transform (Figure 3b). A further dramatic reduction in resolution has to be expected after DFT if the acquisition time in the ^{15}N dimension would be reduced by a factor of five.

Factors influencing MUNIN analyses

Several factors can influence the outcome of a MUNIN analysis, examples being the number of components chosen for the analysis or spectral noise and artefacts. A systematic study of these has to include the analysis of many spectra including synthetic data with fully known contributions; this is beyond the scope of the present paper. The following remarks reflect our current experience. Noise does normally not give rise to separate components, since it does not follow the assumption made in Equation 1. If noise components do appear, then they normally do so with small amplitudes. Spectral artefacts with higher intensities than the signals pose more of a problem because the set of components may first try to describe those. The choice of the number of components to be used is not very critical. A too small choice will be readily detectable since several signals will be combined into single components, resulting in unexpected line shapes, e.g. two or more maxima in the shapes along the ^{15}N or ^1H_N dimensions. In our experience, a choice that exceeds the number of real signals by up to 20% has hardly any consequences. Finally, the processing of complete 3D data sets is for the ^1H-^{15}N NOESY-HSQC and many other types of spectra most effectively done by individually decomposing several spectral fragments followed by a collection of all components.

Conclusions

In this paper we introduce a new method, MUNIN, for spectral processing of 3D NMR data. It relies on a mathematical technique called three-way decomposition. The two examples presented above for a ^1H-^{15}N 3D NOESY-HSQC illustrate the computational feasibility of the method and the power of MUNIN to decompose signals in crowded spectra while avoiding artefacts or distortions to the line shapes. The set of components that MUNIN yields can be thought of as an efficient and natural method of compression (>100 times) of spectroscopic data, which can be very helpful for further spectral analysis. Quantification (e.g. peak picking and peak integration) of the resulting one-dimensional $^1H_{NOE}$ shapes is much easier than in the original multidimensional spectrum, which should significantly contribute to advance the automation in spectral analysis. Reduction by MUNIN of the dimensionality of a multidimensional spectrum down to one-dimensional shapes opens wider possibilities for other, non-Fourier based methods of spectral estimation, which can be otherwise hindered by computational limitations. Since Fourier transform is not required in all dimensions, non-uniform sampling schemes in time domain become practical in multidimensional NMR spectroscopy. Thus, one can be more flexible in designing experimental procedures optimised to achieve better sensitivity and resolution. The MUNIN decomposition can be applied to data sets other than three-dimensional spectra, e.g. to a set of 2D HSQC-type spectra in relaxation or J-coupling analysis. In this case the shapes along the third dimension of the 3D data array formed by the set of 2D spectra correspond to relaxation decays or J-modulations.

Acknowledgements

NMR experiments were performed at the Swedish NMR Centre at Göteborg University. The work was supported by a research grant from NFR (K-AA/KU 12071-302) and Lars Hiertas Minne foundation. We are grateful to B.G. Karlsson and J. Leckner for providing us with a uniformly ^{15}N-labelled sample of azurin and to M. Kubista for fruitful discussions.

References

Abergel, D. and Delsuc, M.A. (1993) *THEOCHEM: J. Mol. Struct.*, **105**, 65–70.

Andersson, C.A. and Bro, R. (1998) *Chemometrics Intell. Lab. Syst.*, **42**, 93–103.

Barache, D., Antoine, J.P. and Dereppe, J.M. (1997) *J. Magn. Reson.*, **128**, 1–11.

Bretthorst, G.L. (1990) *J. Magn. Reson.*, **88**, 571–595.

Bro, R. (1997) *Chemometrics Intell. Lab. Syst.*, **38**, 149–171.

Carroll, J.D. and Chang, J. (1970) *Psychometrica*, **35**, 283–319.

Carroll, J.D. and Pruzansky, S. (1984) In *Research Methods for Multimode Data Analysis* (Eds, Law, H.G., Snyder, C.W., Hattie, J.A. and McDonald, R.P.), Praeger, New York, NY, pp. 372–402.

Chylla, R.A. and Markley, J.L. (1995) *J. Biomol. NMR*, **5**, 245–258.

Denk, W., Baumann, R. and Wagner, G. (1986) *J. Magn. Reson.*, **67**, 617–636.

Harshman, R.A. (1970) *UCLA Working Papers in Phonetics*, **16**, 1–84.

Harshman, R.A. and Lundy, M.E. (1984) In *Research Methods for Multimode Data Analysis* (Eds, Law, H.G., Snyder, C.W., Hattie, J.A. and McDonald, R.P.), Praeger, New York, NY, pp. 122–215.

Henrion, R. (2000) *J. Chemometr.*, **14**, 261–274.

Hoch, J.C. and Stern, A.S. (1996) In *Encyclopedia of Nuclear Magnetic Resonance* (Eds, Grant, D.M. and Harris, R.K.), John Wiley, London, pp. 2980–2988.

Hopke, P.K., Paatero, P., Jia, H., Ross, R.T. and Harshman, R.A. (1998) *Chemometrics Intell. Lab. Syst.*, **43**, 25–42.

Hu, H.T., De Angelis, A.A., Mandelshtam, V.A. and Shaka, A.J. (2000) *J. Magn. Reson.*, **144**, 357–366.

Ibraghimov, I.V. (1999) In *Matrix Methods and Algorithms* (Ed., Tyrtyshnikov, E.E.), Inst. Comput. Mathematics RAS, Moscow, pp. 194–202.

Karlsson, B.G., Pascher, T., Nordling, M., Arvidsson, R.H.A. and Lundberg, L.G. (1989) *FEBS Lett.*, **246**, 211–217.

Koehl, P. (1999) *Prog. NMR Spectrosc.*, **34**, 257–299.

Koradi, R., Billeter, M., Engeli, M., Güntert, P. and Wüthrich, K. (1998) *J. Magn. Reson.*, **135**, 288–297.

Kotyk, J.J., Hoffman, N.G., Hutton, W.C., Bretthorst, G.L. and Ackerman, J.J.H. (1995) *J. Magn. Reson.*, **A116**, 1–9.

Kroonenberg, P.M. and Leeuw, J. (1980) *Psychometrica*, **45**, 69–97.

Kruskal, J.B. (1977) *Linear Algebra Appl.*, **18**, 95–138.

Kruskal, J.B. (1989) In *Multiway Data Analysis* (Eds., Coppi, R. and Bolasco, S.), North-Holland Elsevier Science Publishers, Amsterdam.

Ochs, M.F., Stoyanova, R.S., Arias-Mendoza, F. and Brown, T.R. (1999) *J. Magn. Reson.*, **137**, 161–176.

Rouh, A., Louisjoseph, A. and Lallemand, J.Y. (1994) *J. Biomol. NMR*, **4**, 505–518.

Schmieder, P., Stern, A.S., Wagner, G. and Hoch, J.C. (1997) *J. Magn. Reson.*, **125**, 332–339.

Sidiropoulos, N.D. and Bro, R. (2000) *J. Chemometr.*, **14**, 229–239.

Steinbock, O., Neumann, B., Cage, B., Saltiel, J., Muller, S.C. and Dalal, N.S. (1997) *Anal. Chem.*, **69**, 3708–3713.

Stilbs, P., Paulsen, K. and Griffiths, P.C. (1996) *J. Phys. Chem.*, **100**, 8180–8189.

Stoyanova, R., Kuesel, A.C. and Brown, T.R. (1995) *J. Magn. Reson.*, **A115**, 265–269.

Tucker, L.R. (1966) *Psychometrica*, **31**, 279–311.

Van de Kamp, M., Canters, G.W., Wijmenga, S.S., Lommen, A., Hilbers, C.W., Nar, H., Messerschmidt, A. and Huber, R. (1992) *Biochemistry*, **31**, 10194–10207.

Weaver, J.B., Xu, Y.S., Healy, D.M. and Driscoll, J.R. (1992) *Magn. Reson. Med.*, **24**, 275–287.

Windig, W. and Antalek, B. (1999) *Chemometrics Intell. Lab. Syst.*, **46**, 207–219.

Zhang, O.W., Kay, L.E., Olivier, J.P. and Forman-Kay, J.D. (1994) *J. Biomol. NMR*, **4**, 845–858.

Appendix: Least-squares minimization in MUNIN

Equation 1 can be rewritten as a minimization problem whose solution describes the individual shapes and amplitudes of the R components:

$$\min_{F1,F2,F3,A} \sum_{i,j,k} \left(S_{i,j,k} - \sum_{m=1}^{R} A_m^m \cdot F1_i^m \cdot F2_j^m \cdot F3_k^m \right)^2 \qquad \text{(A1)}$$

This appendix presents the PARAFAC approach, the choice of an initial approximation and the 'Tucker3' compression.

(1) PARAFAC (Harshman and Lundy, 1984), an iterative algorithm for the solution of Equation A1, uses the fact that if two shape matrices are fixed, the third one can be determined by solving a quadratic least-squares problem. Thus, in every PARAFAC iteration two shapes are taken from the previous iteration (or from an initial approximation) and considered fixed, and the third shape is computed. Fixing for example $F1$ and $F2$, we can update $F3$ and A by finding the minimum of

$$\min_{Z} \|B - CZ\|_F^2 \qquad \text{(A2)}$$

where the matrix B (dimensions $IJ \times K$) is obtained by joining the first two dimensions of S, the matrix C ($IJ \times R$) is defined as $C_{i,j}^m = F1_i^m F2_j^m$, and $Z = AF3^T$. (The Frobenius norm $\|X\|_F$ of an arbitrary matrix X is the square root of the sum of squares of all elements of X.) The update of $F3$ can be calculated by normalization of the rows in Z, and the updated matrix $A$is then constructed from the normalization factors. The minimum in Equation A2 is found by equating to zero all partial derivatives with respect to Z, yielding a system of linear equations:

$$C^T CZ = C^T B \qquad \text{(A3)}$$

Solving Equation A3 requires the solution of K systems of linear equations with the same matrix $C^T C$. Singular value decomposition (SVD) is used to resolve possible singularities in $C^T C$ during the iterations. In subsequent iterations the procedure is repeated to update the other two shape matrices. The fact that in every iteration the original functional given

by Equation A1 is minimized (with some variables fixed) ensures monotonous convergence of the iteration process. In total one needs for three iterations, updating all three dimensions, about $3IJKR + 3(I + J + K)R^3 + 15R^3$ arithmetic operations.

(2) Initial approximations for the first PARAFAC iteration can be obtained as follows (Ibraghimov, 1999). First an orthonormal matrix $F1$ is constructed from the R largest left singular vectors of the matrix S' ($I \times JK$), which is obtained from S by concatenation of the second and third dimensions. $F2$ and $F3$ can then be calculated by solving the quadratic least-squares problem that is derived from Equation A1 using the orthonormality of $F1$ and known properties of matrix norms:

$$\min_{F2,F3,A} \quad \sum_m \sum_{j,k} \left(\left(\sum_i S_{i,j,k} \cdot F1_i^m \right) - A_m^m \cdot F2_j^m \cdot F3_k^m \right)^2 + \sum_{i,j,k} S_{i,j,k}^2 - \sum_m \sum_{j,k} \left(\sum_i S_{i,j,k} \cdot F1_i^m \right)^2 \tag{A4}$$

The last two terms are constants and can be disregarded. With $D(m)_j^k = \sum_i S_{i,j,k} \cdot F1_i^m$ ($J \times K$), Equation A4 splits into R independent minimization problems:

$$\min_{F2^m, F3^m, A_m^m} \sum_{j,k} \left(D(m)_j^k - A_m^m \cdot F2_j^m \cdot F3_k^m \right)^2 \tag{A5}$$

Using SVD one obtains A_m^m, $F2^m$ and $F3^m$ as the largest singular value and the corresponding left and right singular vectors of the matrix D(m), respectively. Note that this method for the generation of an initial approximation requires that $R \leq \max(I, J, K)$.

(3) The 'Tucker3' step is a useful modification of the PARAFAC algorithm when R is much smaller than at least one of the numbers I, J or K (Tucker, 1966; Kroonenberg and Leeuw, 1980). Consider the following substitutions:

$$F1 = U1G1, \; F2 = U2G2, \; F3 = U3G3 \tag{A6}$$

where $U1$, $U2$ and $U3$ are matrices with orthonormal columns and dimensions $I \times R1$, $J \times R2$ and $K \times R3$, respectively. The matrices $G1$, $G2$ and $G3$ have dimensions $R1 \times R$, $R2 \times R$ and $R3 \times R$, respectively. $R1$, $R2$ and $R3$ are selected so that $\min(R, I) \leq R1 \leq I$, $\min(R, J) \leq R2 \leq J$, $\min(R, K) \leq R3 \leq K$. The original problem of Equation A1 can then be substituted by two usually simpler problems:

$$Q_{i',j',k'} = \sum_{m=1}^{R} A_m^m \cdot G1_{i'}^m \cdot G2_{j'}^m \cdot G3_{k'}^m + e2_{i',j',k'} \tag{A7}$$

$$S_{i,j,k} = \sum_{i',j',k'} Q_{i',j',k'} \cdot U1_i^{i'} \cdot U2_j^{j'} \cdot U3_k^{k'} + e1_{i,j,k} \tag{A8}$$

with $i' = 1 \ldots R1$, $j' = 1 \ldots R2$ and $k' = 1 \ldots R3$. In a first step Tucker3 compression is performed, i.e. the three-dimensional core matrix Q and the orthonormal factor matrices $U1$, $U2$ and $U3$ are estimated by solving the least-squares problem of Equation A8. Next, the PARAFAC algorithm is applied to solve the problem defined by Equation A7. The gain due to compression lies in the reduced size of Q compared to the original data S. If all residuals $e1$ in Equation A8 vanish, then the shapes $F1$, $F2$ and $F3$ defined by Equation A6 correspond exactly to those of the original minimization problem of Equation A1. Otherwise they represent a good initial approximation from which PARAFAC often converges after only a few iterations. Since the matrices $U1$ and $U2$ are orthonormal and using known properties of matrix norms the problem of Equation A8 can be rewritten as follows:

$$\min_{U3,Q} \sum_{i',j'} \sum_k \left(\left(\sum_{i,j} S_{i,j,k} \cdot U1_i^{i'} \cdot U2_j^{j'} \right) - \left(\sum_{k'} Q_{i',j',k'} \cdot U3_k^{k'} \right) \right)^2 \tag{A9}$$

and with $P_{i',j'}^k = \sum_{i,j} S_{i,j,k} \cdot U1_i^{i'} \cdot U2_j^{j'} (R1R2 \times K)$ and $W_{i',j'}^{k'} = Q_{i',j',k'} (R1R2 \times R3)$ as:

$$\min_{U3,W} \| P - WU3 \|_F^2 \tag{A10}$$

This least-squares problem can be solved by SVD of the matrix P, followed by construction of $U3$ from the right singular vectors corresponding to the $R3$ largest singular values. Q is obtained after the last iteration from $W = P \cdot U3^T$. Initial guesses for the matrices $U1$ and $U2$ are obtained from the largest left singular vectors of the matrix S', which is obtained from S by joining two dimensions. S should be permutated for

the construction of $U2$ to make J the leading dimension. The problem defined by Equation A8 converges much faster than PARAFAC on the original data due to the orthonormality of the factors $U1$, $U2$ and $U3$. The 'Tucker3' compression according to Equation A8 is not the time-limiting step of the MUNIN procedure.