Journal of Biomolecular NMR, **24:** 191–201, 2002.
KLUWER/ESCOM

Three-way decomposition of a complete 3D ^{15}N-NOESY-HSQC

Aleksandras Gutmanas[a], Patrik Jarvoll[b], Vladislav Yu. Orekhov[b] & Martin Billeter[a,*]
[a]*Biochemistry and Biophysics, and* [b]*Swedish NMR Centre, Göteborg University, Box 465, 405 30 Göteborg, Sweden*

Received 8 May 2002; Accepted 4 October 2002

Key words: automated spectral analysis, azurin, demixing, MUNIN, ^{15}N-NOESY-HSQC, PARAFAC, peak-picking

Abstract

Three-way decomposition is applied for the structural analysis of a complete three-dimensional ^{15}N-NOESY-HSQC of the 128 residues long protein azurin. The procedure presented includes decomposition using the software MUNIN, providing an initial characterization of the complete spectrum by 355 components. This is followed by post-processing yielding a final list of 149 components, 123 of which characterize 1859 NOE peaks from backbone N-H groups. Components from three-way decomposition are defined as direct products of one-dimensional shapes along the three dimensions. Thus, a complete set of distance constraints from this spectrum is obtained by one-dimensional peak picking of the shapes along the NOE dimension. Correctness and completeness of this set of NOEs are tested for all backbone amide groups against both an independent peak picking algorithm and the three-dimensional crystal structure of azurin, and a coincidence of about 95% is observed. Automated 'demixing' of components that are 'mixed' in a complex manner due to overlap of the HN and/or ^{15}N frequencies is illustrated.

Abbreviations: FID, free induction decay; MUNIN, multi-dimensional NMR spectra interpretation; ^{15}N-NOESY-HSQC, three-dimensional combination of NOESY and ^{15}N-HSQC; 1D, one-dimensional; 3D, three-dimensional.

Introduction

One of the most exciting features of biomolecular NMR is its wide applicability, covering structural studies at atomic resolution to investigations of fold and domain orientation, various aspects of internal molecular dynamics or the detection and characterization of ligand binding to macromolecules. As a consequence, a wide variety of different experiments has been designed, requiring either a large number of different processing schemes or more general approaches. The importance of optimal data processing methods is a consequence of limitations on instrument time and hardware, the need for high-throughput schemes and the fact that many new targets such as membrane systems challenge the limits of today's capabilities of NMR.

*To whom correspondence should be addressed. E-mail: martin.billeter@bcbp.gu.se

Three-way decomposition (Caroll and Chang, 1970; Harshman, 1970) implemented in the software MUNIN has been introduced recently as a general method for the evaluation of NMR spectra (Orekhov et al., 2001). Analyses of very different types of NMR data demonstrated the general applicability of this approach besides the use on various 3D spectra after Fourier transform in all three dimensions. Thus, 3D spectra with one dimension left in time-domain opened an avenue to the use of non-uniformly sampled spectral data (Orekhov et al., 2001). Sets of 2D ^{15}N-HSQC data were used in two different contexts. First, highly overlapped signals in relaxation-modulated spectra could be separated due to the simultaneous considerations of differences in their exponential decays (Korzhnev et al., 2001). A further application concerns the efficient screening for binding of potential ligands to a target protein as in *SAR by NMR* (Hajduk et al., 1999), using again a set of 2D ^{15}N-HSQC, where one of the three dimensions

of the input data set for MUNIN corresponded to the enumeration of the spectra (Damberg et al., in press).

The basic model assumption of the three-way decomposition method is that line shapes along each dimension combined with amplitudes can describe NMR signals. Mathematically, a 3D NMR data set $\boldsymbol{S}$ can be expressed as the sum of direct products of 1D line shape vectors:

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

where $S_{i,j,k}$ is an experimental data point, and the indices *i, j* and *k* cover the whole spectrum $\boldsymbol{S}$. Each of the R terms in the sum is referred to as a *component*, with A^m being its amplitude. The factors $F1^m$, $F2^m$ and $F3^m$ are the normalized 1D shape vectors in the corresponding dimensions and are henceforth named *shapes*. Finally, $e_{i,j,k}$ is a point from the 3D residual of the decomposition. For a more detailed discussion of three-way decomposition of NMR spectra see Orekhov et al. (2001). The input for a calculation is the data set $\boldsymbol{S}$ and the number of expected components R, whereas the shapes $F1^m$, $F2^m$ and $F3^m$ and the amplitudes A^m represent the output. The goal of the decomposition is to optimally describe $\boldsymbol{S}$ by the R components, i.e., to minimize the norm of the residual matrix with elements $e_{i,j,k}$. The scalar R is typically chosen much smaller than the number of signals (peaks) in $\boldsymbol{S}$. Thus, the decomposition algorithm must collect several signals, choosing in particular those with high intensity, into single components. Since only signals with common features such as the same frequency in two dimensions can be described by a single component, noise is expected to largely remain in the residual.

The wide applicability of three-way decomposition is a consequence of the fact that Equation 1 does not imply any specific form of the shapes. Thus, shapes may describe frequency-domain spectral data, time-domain FIDs, exponential relaxation decays, two-valued functions with for example '1' for binding and '0' for non-binding and many other experimental results.

In the following we demonstrate the use of MUNIN to a complete 3D ^{15}N-NOESY-HSQC of azurin. The entire spectrum is analyzed, signals (NOEs) are identified and their correctness and completeness with respect to an independent peak-picking algorithm, AUTOPSY (Koradi et al., 1998), as well as to the 3D protein crystal structure (4AZU.pdb, Nar et al., 1991) are examined. The phenomenon of loss of uniqueness of the decomposition known as 'mixing' (Kruskal, 1977, 1989; Orekhov et al., 2001) and automated 'demixing' are demonstrated for complex situations.

Materials and methods

The spectrum

A 3D gradient sensitivity-enhanced ^{15}N-NOESY-HSQC spectrum (Zhang et al., 1994) of azurin (a blue copper protein, Karlsson et al., 1989) was recorded on a Varian Inova 600 MHz instrument. It contained 512 complex points in the direct (HN) dimension, and 200 and 40 in the first (H_{NOE}) and second (^{15}N) indirect dimensions, respectively. The spectrum was processed with the NMRPipe software (Delaglio et al., 1995). The two proton dimensions were Fourier transformed after zero filling to 1024 and 512 points, respectively, whereas the ^{15}N dimension was kept in time domain. Each of the resulting 80 ^{1}H-^{1}H planes contained 455×512 points after extraction of the relevant section in the HN dimension, corresponding to 12.0–5.5 ppm and 12.08–2.55 ppm along HN and H_{NOE}, respectively.

Mixing and demixing

Decomposition according to Equation 1 is unique, provided that the components differ in all three shapes. If however the shapes of any two or more components are identical along one dimension, the solution is no longer unique (Kruskal, 1977, 1989; Orekhov et al., 2001). Components possessing identical shapes in one dimension are termed *mixed* in that dimension, and the procedure to select the most appropriate of the many solutions is called *demixing*. It requires additional constraints on the shapes in at least one of the two non-identical dimensions. As a consequence of applying such constraints, the second non-identical dimension also changes. A detailed theoretical account of demixing is given in Orekhov et al. (2001). However, the solution presented there was limited to manual demixing of only two components.

Groups of potentially mixed components were identified by calculating pairwise scalar products of the shapes in all dimensions as described in Orekhov et al. (2001) with a cutoff value of 0.9. Note that groups may be defined based on mixing in several dimensions. Thus, component A may be mixed with

component B in one dimension and with component C in another dimension; such a case is presented and discussed in Results. As described in Orekhov et al. (2001), each demixed shape vector is a linear combination of the mixed ones, which implies that an appropriate transformation matrix must be optimized. Moreover, since the new shapes are independent, each line of this matrix may be optimized individually.

In the work presented here two criteria were used for automated demixing. If the components were mixed in the HN dimension, the respective ^{15}N shapes were combined in order to produce a single complex exponent for each of the new components. If on the other hand the components were mixed in the ^{15}N dimension, a combination of HN shapes was sought that each contained a single Gaussian peak. No criteria were applied to the NOE dimension, which was demixed as a by-product as stated above. Both demixing criteria were implemented in MATLAB (The MathWorks, Inc.), making use of the standard Levenberg-Marquardt algorithm (e.g., Press et al., 1992) for non-linear least squares minimization. Note that if two or more components have identical shapes in two dimensions, they can effectively be described by only one component, whose shape in the third dimension is the sum of the corresponding original shapes, weighted by their amplitudes. However, normally the similarity of the shapes is only partial, and therefore a special function to treat these cases was devised. It reconstructed a sub-spectrum using Equation 1 but summing up only the components in question. This sub-spectrum was then submitted to three-way decomposition using a reduced number of components.

Peak picking in the NOE shapes

One-dimensional peak picking on the NOE shapes was performed by an *ad hoc* home-built routine (Jarvoll, 2002), which is shortly summarized here. Each shape was treated independently. The noise was assumed to be normally distributed. A noise level was determined by removing the furthest outliers of the distribution that describe the signals. At each iteration the point contributing most to the skewness of the distribution was eliminated, and the iterations were repeated until the kurtosis approached the value expected for a normal distribution. Next, the shape was divided into noise and peak regions and each peak region was treated separately. A linear combination of a Gaussian and a Lorentzian line was fitted to each maximum with the contribution from each type of line shape being an optimization parameter. Peaks identified in this manner were subtracted from the MUNIN shape, and the fitting procedure was repeated until no intensity above the noise level remained. Then the original peak region was considered again and a common fit was calculated using all previously found peak positions. Peaks with less than three points above the noise level or with obviously insignificant amplitudes and line widths were eliminated. Namely, the amplitude had to be at least twice (adjustable parameter) the noise level and the line width had to be below 15 data points (0.4 ppm, adjustable parameter). After the fitting of peaks to the resulting peak regions, a mean peak width was calculated and peaks whose width was 1.75 times higher than the mean (adjustable parameter) were identified. Attempts were made to split each of these broad peaks by fitting two smaller peaks, whose parameters were derived from the original peak. The new splitted peaks were minimized together with the other peaks and were only accepted if the fit improved and both new peaks satisfied the above significance criteria.

For the resulting significant peaks, volumes were determined as percentages of the diagonal peak. Quality factors, Q, were calculated for each peak using the following expression.

$$Q = \frac{2 \cdot \sum_i p_i d_i}{\sum_i \left(p_i^2 + d_i^2\right)}. \qquad (2)$$

Here, the index i enumerates the N data points d_i describing the peak, and p_i are the corresponding fitted points. Because the fitting of the p_i involves four parameters, peaks consisting of $N = 3$ and $N = 4$ data points are given quality factors of $Q = 0.7$ and $Q = 0.8$, respectively. Equation 2 represents an *ad hoc* definition and is mainly intended to detect possibly wrong results of the peak picking that require further analysis. In the case of essentially non-negative data (if negative intensity is expected a somewhat more complex expression is needed) a quality factor of $Q = 1.0$ indicates a perfect fit, and $Q = 0.0$ is obtained for example when a peak is not fitted at all ($p_i = 0$ for all indices i). Systematic deviations between d_i and p_i, where all p_i are either twice or half as big as the corresponding d_i, would yield $Q = 0.8$. Thus, peaks with quality factors of less than 0.5 require further investigation.

Results

MUNIN runs

A ^{15}N-NOESY-HSQC was recorded and Fourier-transformed in both proton dimensions as described in Methods. It was then divided into 17 overlapping regions along the HN dimension. Fifteen central regions, covering the HN frequency range from 9.5 to 6.5 ppm, were each constructed to be 21 data points wide (7 data points correspond to 0.1 ppm). The two flanking regions, covering the low-field and the high-field part of the spectrum, were 181 and 71 data points wide, respectively. All regions overlap with their immediate neighbors by 7 data points; this value was chosen to be somewhat larger than the average signal linewidth. In the other two dimensions, all regions had the same size, namely 470 data points or 13.43 ppm along the NOE dimension and 80 time-domain points along the ^{15}N dimension. In order to avoid the strong noise present at the spectral edges, strips along the NOE dimension with a width of 0.29 ppm at the low-field end and 0.91 ppm at the high-field end were removed. Figure 1 shows regions 6 to 8 of the first HN-H_{NOE} plane, illustrating the overlap used between neighboring regions and its relation to peak widths.

The input to individual MUNIN runs consisted of 3D spectral data corresponding to one region plus the expected number of components that are necessary to describe all signals within the region. For the present spectrum, this number corresponds to the number of N-H groups with proton frequencies in the interval defined by the region limits. A high-resolution HSQC spectrum was used for the estimation of the number of components as follows. The size of a given region was increased in the HN dimension by five points on each side, and all HSQC peaks having their maximum in this enlarged region were counted. This number was increased by three to account for the water signal and other spectral artifacts. Due to incomplete assignments of the side chain resonances and their strong overlap, the expected number of components needed to be further increased for the regions covering the HN frequency interval from 7.71 to 6.61 ppm. An indication for the need to increase the number of expected components was the fact that the water line present in the spectrum was described by only one component whereas two to three components were required in other regions. This estimation yielded between 6 and 40 components for each region. The calculation for all

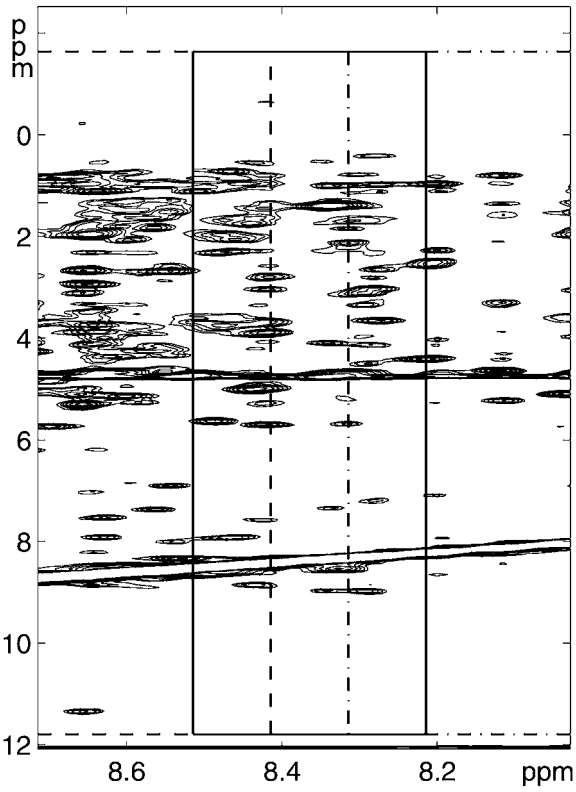

Figure 1. Spectral section from the first NOESY plane of the ^{15}N-NOESY-HSQC (with time-domain data along the ^{15}N dimension). Solid lines delineate a central region, dashed and dash-dotted lines overlapping regions on each side. Decomposition examples in the text and the following figures are taken from the central region.

17 regions required about 30 hours on a single R10000 CPU of an Origin2000 SGI computer.

Analysis of components

The results of the 17 MUNIN runs are summarized in Table 1. From the total of 355 components obtained, 206 could be removed for one of the following reasons. In each region one or several components are required to describe the strong intensity of the water signal. These are easily identified, as their shapes along the HN dimension are lines at constant intensity. Other components are duplicated because they are located in an overlap of two neighboring regions; the less complete description was removed. Typically, the number of components used for calculation is larger than what is necessary to characterize the true signals. The additional components are then used to describe strong noise or certain artifacts and can be identified by their low amplitude and their unusual shape (e.g., the complete absence of a Lorentzian-like line shape

Table 1. Overview of components resulting from the MUNIN analysis

Number of components		All 17 regions[a]	Example region[a]
Total		355	30
Removed[b]:	Redundant	118	9
	Water	47	2
	Other	41	5
Total removed		206	16
Relevant[c]:	Directly	53	2
	After demixing	96	12
Total used		149	14
Class[d]:	Backbone	123	14
	Side chain	23	0
	Negative	3	0

[a]The 17 regions cover the entire spectrum; the example region is the central region shown in Figure 1.
[b]Components were removed because (i) of their simultaneous presence in neighboring regions (redundant), (ii) they describe the water signal or (iii) they have small amplitudes and distorted shapes.
[c]Relevant components are all components that have not been removed for the reasons described in footnote 2 and thus describe true spectral signals. A majority of these components required demixing.
[d]All relevant components were classified into components describing (i) backbone or (ii) side chain N-H groups, or (iii) components with negative signals that most likely characterize folded signals from unassigned side chains.

in the HN dimension). A different situation that is encountered for a few component pairs is the near identity of two of the three shapes; such components can be combined into a single component as described in Methods.

The remaining 149 components could either be used directly, or they were mixed and thus needed to be demixed. Comparison with an assigned 2D ^{15}N-HSQC, after Fourier transform of the ^{15}N dimension, shows the following result of the MUNIN analysis of the ^{15}N-NOESY-HSQC. One hundred twenty-three components describe NOEs to all N-H groups of the backbone, 22 components are due to 11 assigned side chains of asparagine and glutamine, and one component corresponds to a tryptophane side chain. Three negative components could not be identified but likely represent folded side chain signals. In the following analysis, only the backbone components were used.

Example components

This section provides some insight into situations that were encountered when analyzing the output of MUNIN, i.e. the components. The example components are taken from the central region of the three shown in Figure 1. In total this region was decomposed using 30 components. Figures 2A and 2B, show the resulting shapes for six selected components as they appeared as output from MUNIN. Examples of easy cases (not mixed or a group of two mixed components) were reported previously (Orekhov et al., 2001). Figure 2A illustrates different situations that were observed in the HN dimension. The thick dash-dotted lines in Figures 2A and 2B, characterize a peak that overlaps with the neighboring region and is better considered in this other region, while the thin dash-dotted lines in Figures 2A and 2B, describe intensity due to water. Both components with dash-dotted lines were therefore removed at this stage.

The remaining four components all include a peak at 8.43 ppm in the HN dimension (Figure 2A) and therefore require demixing. The two dashed lines exhibit normal shapes in this dimension with one maximum and otherwise near zero intensity. The two solid lines on the other hand show in addition significant negative intensity to the left of the peak. In the ^{15}N dimension several features are observed (after Fourier transform; Figure 2B). The thick dashed line shows besides its most prominent positive peak two negative peaks at the frequencies of other components, which is typical for mixing. Similarly, the thin dashed line has a shoulder to the right of its maximum intensity and a weaker peak to the left (with very low intensity, which nonetheless disappears after demixing). Finally, the thick solid line exhibits exclusively negative intensity at the same frequency as the peak of the thin solid line. Due to the presence of similarity between the solid lines along both the HN and the ^{15}N dimension (note that different signs are irrelevant with respect to similarity), the demixing had to occur in two steps. Since both solid lines were obviously describing two HN frequencies, one at 8.43 ppm and the other to the left of the region, they required demixing. The automatic demixing was thus set up to first demix the two components with solid lines based on their similarity in ^{15}N, followed by elimination of the resulting component that describes the HN frequency outside the region. In a second step the other resulting component with a HN frequency of 8.43 ppm was subjected to automatic demixing

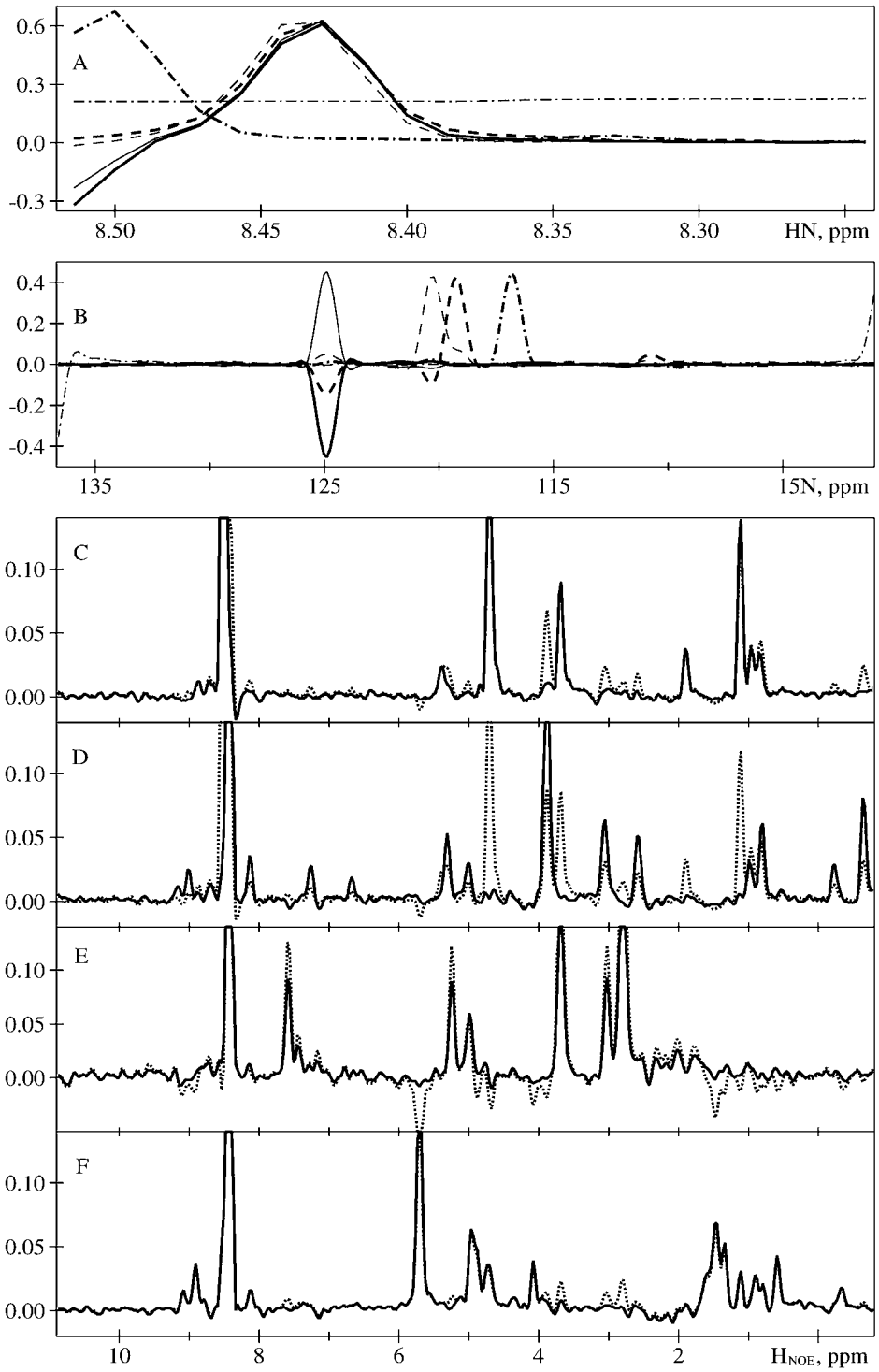

Figure 2. Examples of MUNIN output components from the central region of Figure 1 and automatic demixing results. Four of the six selected components describe a case of mixing due to similarity along the HN dimension (solid and dashed lines), with two components showing further similarity along the ^{15}N dimension (solid lines), while the two remaining components exemplify a case of overlap with the neighboring region (thick dash-dotted line) and describe the water line (thin dash-dotted line). (A) Shapes along the HN dimension. (B) Shapes along the ^{15}N dimension after Fourier transforms. (C–F) Dotted and solid lines show the MUNIN results before and after demixing for the four mixed components shown in (A) and (B) with thick solid lines (C), thin solid lines (D), thick dashed lines (E) and thin dashed lines (F).

with the two components represented by dashed lines based on their similarity in HN. In the first step, demixing was based on the expectation that a single Gaussian curve should be a good approximation of the shapes along the HN dimension, while in the second step the shapes along the ^{15}N dimension (not Fourier transformed!) were approximated by single complex exponential functions.

The effects of these demixing steps along the NOE dimension are shown in Figures 2C–F, with the shapes before and after the demixing procedure given by dotted and solid lines, respectively. The panels from top to bottom correspond to the components with thick and thin solid, and thick and thin dashed lines in Figures 2A and 2B, in this order. The major effect is the complete disappearance of NOE peaks from certain shapes. In particular, all negative peaks are eliminated. Other peaks change intensity, and in the top two panels the diagonal peaks become narrower, i.e., better defined. In the HN dimension all shapes adopt after demixing simple curves with one maximum each, and in the ^{15}N dimension (after Fourier transform) negative peaks and shoulders disappeared (not shown).

Reconstructions

The extent and correctness of the MUNIN decomposition of the spectrum is best illustrated by complete or partial reconstructions of the spectrum using all or selected MUNIN components according to Equation 1. For these reconstructions the equation is used to calculate a 3D spectrum (left side of the equation) from shapes $F1^m$, $F2^m$ and $F3^m$ and amplitudes A^m describing all or a selected subset of the components. The selected shapes are multiplied to form 3D components and subsequently added up according to the right side of the equation. Figure 3 shows the first ^{1}H-^{1}H plane for reconstructions or their differences to the original experimental data of the central region of Figure 1. Differences to the original data are given for the direct MUNIN results prior to (Figure 3A) and after demixing (Figure 3B), using in both cases all components. Obviously the MUNIN decomposition represents a very good approximation of the original spectrum, with only minor differences affecting mostly diagonal peaks. The effect of the classification of components into those components describing NOEs and those marked for removal (Table 1) is illustrated in Figures 3C, D. Figure 3C was calculated as the sum of all components representing NOEs to the central N-H groups of this region, i.e. the essential output of MUNIN for this region. Figure 3D is a reconstruction describing border components, the water line and other artifacts. The sum of these two reconstructions corresponds thus to the calculated spectrum that was subtracted from the original data in Figure 3B. Figure 3E provides traces in the first ^{1}H-^{1}H plane along the HN frequency indicated by dashed lines in Figures 3A–D. The thin solid line shows the original, experimental data, while the thick solid line shows the difference between this original data and the MUNIN decomposition; this line thus corresponds to a cross section along the dashed line of Figure 3A. It proves again the high similarity between the experimental data and the MUNIN decomposition. The dashed line shows the difference between the original data and the partial reconstruction using selected components as in Figure 3D (i.e., components marked for removal). The relatively strong presence of the dashed line at the diagonal and the corresponding diagonal intensity in Figure 3D result from five removed components (Table 1). These components were weaker than all the relevant ones and possessed distorted line shapes in at least one dimension. They describe features like satellite peaks and distortions along the ^{1}H-^{1}H diagonal, which are observed in the spectrum.

Peak picking in the NOE shapes

One-dimensional peak picking as described in Methods was applied to the shapes along the NOE dimension of all backbone components, which were first demixed if required. Peaks with less than three consecutive points above the noise level were removed, and particularly broad peaks were considered for splitting (see Materials and methods). Negative intensity in these shapes, if present, was not considered. The final amplitudes of fitted peaks had to be at least two times the noise level. Peak amplitudes were then scaled by setting the amplitude of the diagonal peak in each shape to unity. A quality factor was determined for each peak according to Equation 2. An illustration of this 1D peak picking is given for the NOE shape of His 46 (Figure 4), and all peaks obtained for this residue are listed in Table 2 (left side). The NOE shape of this component is plotted in Figure 4A, and the individual peaks identified by the peak picking are plotted in Figure 4B. The difference between these two plots is shown as a bold trace in Figure 4A. All obvious peaks are identified together with shoulders near 1.1, 1.6, 7.1 and 8.4 ppm. The latter is a shoulder of the diagonal peak and is thus, with a quality factor of

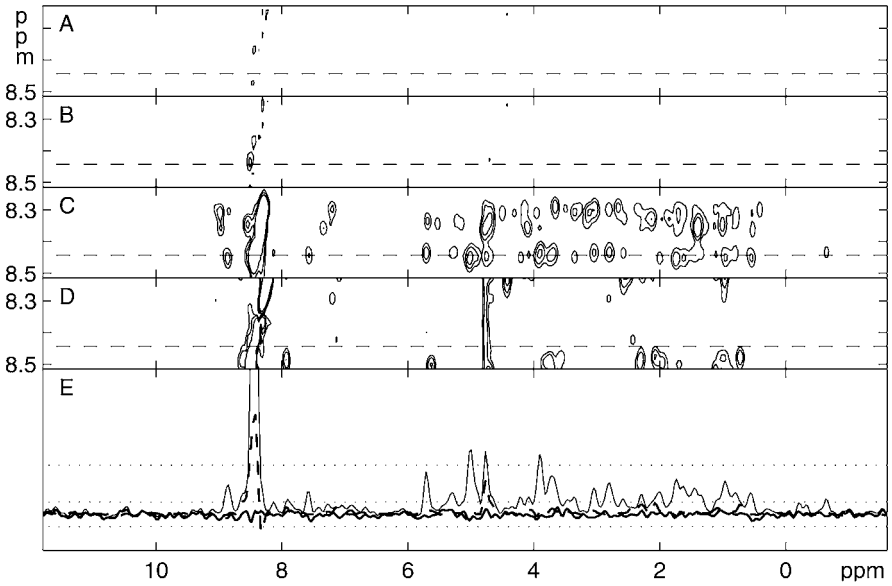

Figure 3. Extent and correctness of the MUNIN decomposition illustrated using the first ^{1}H-^{1}H plane of the example region from Figure 1. (A) Difference spectrum between the original experimental data and the spectrum reconstructed from all components representing the direct output of MUNIN. (B) Difference between the original data and the reconstruction from all MUNIN components after demixing. (C) Reconstruction from selected demixed components that describe the NOE peaks extracted from this region by MUNIN. (D) Reconstruction from selected demixed components that were removed because they represent border components, water lines or artifacts. (E) 1D cross sections with the original data (thin solid line), difference to the reconstruction using all components before demixing (thick solid line; corresponds to the dashed line in (A)), and difference to the reconstruction from selected demixed components (dashed line; corresponds to the dashed line in (D)). Dashed lines in (A)–(D) show the position of the cross sections in (E). In (A)–(D) solid lines give both positive and negative contour levels; the maximal and minimal heights of these contour lines are indicated by dotted lines in (E).

0.68 (Table 2), less reliable. For a test of correctness and completeness with respect to the 3D structure, the crystal structure of azurin with the PDB-code 4AZU (Nar et al., 1991) was used. Hydrogen atoms were added to this structure using a function of the program MOLMOL (Koradi et al., 1996). The right side of Table 2 provides for all hydrogen atoms within 5 Å of the HN of His 46 both the distances to this atom and the independently assigned chemical shift (these shifts were obtained manually using additional spectra). Shifts are matched between the two sides of the table if they differ by less than 0.08 ppm, accounting mainly for shift changes between different spectra. Figure 4C shows the corresponding region from the conventionally processed spectrum to help the following discussion of discrepancies between MUNIN results and the crystal structure. Besides the shoulder peak of the diagonal signal mentioned above, four more signals could not be matched to a short proton-proton distance in the structure (at 7.48, 7.15, 0.99 and 0.77 ppm). All four signals are, however, clearly present in the NOE shape (Figure 4A) as well as in the original spectrum (Figure 4C). Since for these amide groups no overlap occurs in the spectrum, it can be concluded that these resonances belong to unassigned atoms of the protein or some impurity. Two distances from the right side of Table 2 cannot be assigned to peaks from the left side (at 4.89 and 3.72 ppm). Both are larger than 4.5 Å and may therefore be invisible, possibly due to small structural differences of azurin in the crystal and in solution. Both signals are missing in the shape (Figure 4A) and in the original spectrum (Figure 4C). Peak picking of the NOE shapes from all 123 backbone components yielded a total of 1859 NOEs after removal of those with amplitudes less than 1% of the respective diagonal peak or with quality factors < 0.5 (see Equation 2). Their correctness is discussed below.

Discussion

The correctness and completeness of the analysis of this ^{15}N-NOESY-HSQC was further assessed by comparing the resulting peaks from all components to another peak picker and to the 3D structure of azurin

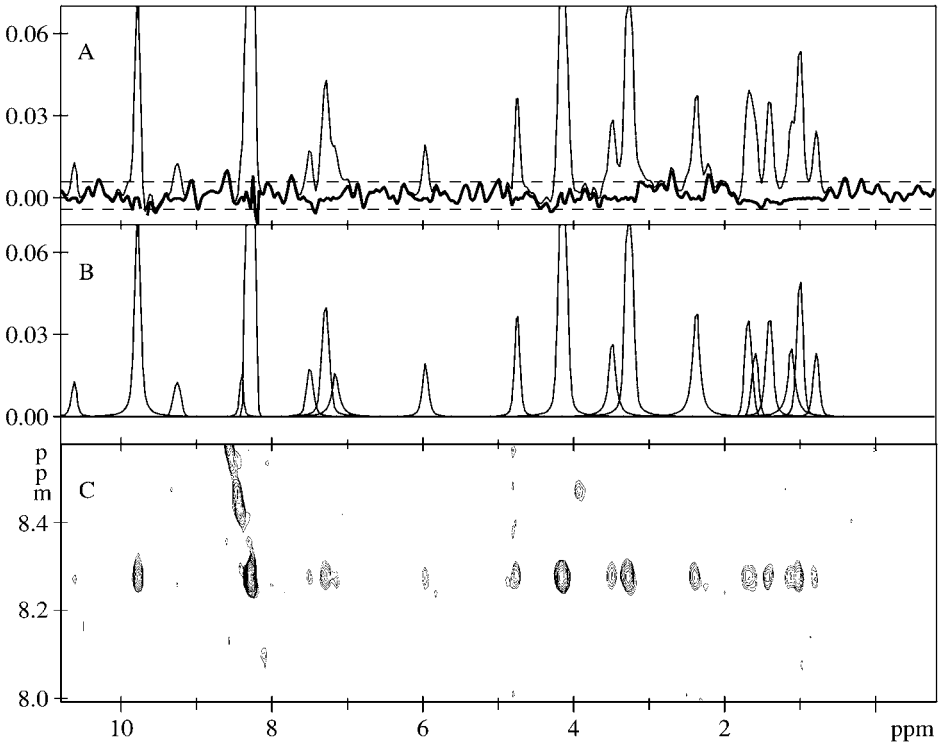

Figure 4. Peak picking of the NOE shape of the component describing the HN of His 46. (A) Shape resulting from MUNIN (thin line) and difference (thick line) of this shape to the peaks picked and shown in (B). (B) Peaks picked as mixed Gaussian–Lorentzian signals by the automatic procedure described in the text. (C) Corresponding region from the conventionally processed spectrum.

(Table 3). AUTOPSY (Koradi et al., 1998) is a sophisticated peak picker that can operate on (fully Fourier-transformed) 3D spectra. We have applied AUTOPSY in the context of another project to the same ^{15}N-NOESY-HSQC as used here (Malmodin et al., pers. commun.), and we use the results based on AUTOPSY run-time parameters defined in the other study. A further test is based on the 3D structure 4AZU.pdb (Nar et al., 1991), from which short proton-proton distances were extracted as described. The three peak lists involving backbone HNs used for the following comparison consisted of 1859 peaks for MUNIN combined with 1D peak picking, 2260 distances shorter than 5 Å from the 3D structure and 1686 peaks picked by AUTOPSY. The 1000 strongest peaks (largest amplitude or shortest distance) from one group (called group A in Table 3) were matched to all entries of the other group (called group B). For the MUNIN peaks this includes all peaks with volumes relative to the corresponding diagonal peak of more than 3.75%. For peaks derived from the crystal structure, the 1000 most likely are those due to distances shorter than 3.76 Å. Matches are peaks differing by less than 0.04, 0.08 and 1.0 ppm in the HN, NOE and ^{15}N dimensions, respectively. They exceed 935 peaks for all group combinations (Table 3). Visual inspection in the spectrum of all peaks reported by MUNIN but missing in the lists obtained by analysis of the crystal structure or from AUTOPSY or vice versa shows that in each case some intensity is present that may or may not represent a cross peak. Thus, all discrepancies arise due to doubtful cases and do not represent clear failures of any method. A similar degree of agreement as between MUNIN and the structure or AUTOPSY is also observed between the latter two lists (Table 3).

Conclusions

In conclusion, the present analysis of a complete ^{15}N-NOESY-HSQC demonstrates the applicability of MUNIN for the extraction of structural data from 3D NMR spectra. With a coincidence of about 95% when compared to peaks derived either from a crystal structure or from an independent peak-picking program, the correctness and completeness of the resulting NOEs

Table 2. Comparison for the backbone amide of His 46 between NOE peaks from MUNIN decomposition and NOEs predicted from the 3D structure[a]

MUNIN peaks[b]			Distances in structure[c]		
Quality[d]	Volume[d]	δ_M [ppm][d]	δ_S [ppm][e]	d [Å][e]	Atom[e]
0.70	2.06	10.61	10.62	4.36	47 Asn HN
1.00	14.60	9.77	9.77	2.63	87 Ile HN
0.80	2.02	9.24	9.23	4.96	88 Gly HN
0.68	2.75	8.39			
1.00	100.00	8.27	8.29	0	46 His HN
0.80	2.89	7.48			
0.98	6.61	7.28	7.29	2.97	46 His Hα
0.88	2.56	7.15			
1.00	3.12	5.95	5.98	4.73	45 Gly HN
			4.89	4.52	87 Ile Hα
1.00	6.39	4.74	4.75	3.92	86 Leu Hα
			4.74	4.85	47 Asn Hα
1.00	2.27	4.13	4.16	2.32	45 Gly Hα
			4.07	4.12	88 Gly Hα
			3.72	4.91	35 His Hβ
0.99	4.44	3.48	3.49	4.62	41 Lys Hα
1.00	16.00	3.25	3.30	2.76	46 His Hβ
			3.25	3.07	45 Gly Hα′
1.00	6.30	2.36	2.36	3.86	46 His Hβ′
0.94	5.71	1.67	1.68	4.04	87 Ile Hγ1
0.84	3.72	1.58	1.61	3.29	87 Ile Hγ1’
1.00	5.93	1.39	1.41	3.14	86 Leu Hγ
			1.31	4.38	41 Lys Hγ′
0.65	4.07	1.10	1.11	4.12	86 Leu $H_3\delta1$
0.97	8.18	0.99			
1.00	3.78	0.77			

[a]Peaks are matched, i.e., are listed on the same line, if their chemical shifts differ by less than 0.08 ppm.
[b]Peaks obtained by MUNIN decomposition and subsequent one-dimensional peak picking of the MUNIN shape along the NOE dimension corresponding to the HN of histidine 46.
[c]Distances from the HN of histidine 46 to hydrogen atoms that are closer than 5 Å in the crystal structure with PDB-code 4AZU. Hydrogen atoms were added to crystal structure using a routine of the program MOLMOL (Koradi et al., 1996).
[d]Quality factors, volumes and the chemical shifts δ_M of the peaks from the MUNIN analysis are a result of the one-dimensional peak picking of the MUNIN shape along the NOE dimension corresponding to the HN of His 46. Quality factors were determined with Equation 2. Volumes are in percent of the diagonal peak.
[e]Distances and atom identifications stem from the crystal structure, while the chemical shifts δ_S are taken from a manual sequence-specific assignment of azurin.

Table 3. Comparison between NOE peaks from MUNIN decomposition and either NOEs predicted from the crystal structure or obtained by peak picking of the 3D spectrum with program AUTOPSY[a]

Comparison[b]		Match[c]
Group A	Group B	
MUNIN	Structure	937
Structure	MUNIN	966
MUNIN	AUTOPSY	985
AUTOPSY	MUNIN	987
Structure	AUTOPSY	953
AUTOPSY	Structure	961

[a]Comparisons were between groups of observed or predicted NOE peaks. For each group, the 1000 most likely peaks (largest observed volumes or predicted from the shortest distances) are checked for a match in the full list of a second group.
[b]‘MUNIN’ represents the 1859 NOE peaks obtained by MUNIN decomposition and subsequent one-dimensional peak picking of all MUNIN shape along the NOE dimension. ‘Structure’ represents the 2260 proton-proton distances shorter than 5 Å. ‘AUTOPSY’ represents the 1686 NOE peaks from peak picking in the 3D spectrum with the program AUTOPSY (Koradi et al., 1998).
[c]Peaks from the 1000 most likely peaks of group A that match a peak from group B. Match indicates chemical shift difference of less than 0.04, 0.08 and 1.0 ppm in the HN, NOE and ^{15}N dimensions, respectively.

should be sufficient to warrant direct structure calculations with a robust method such as those implemented in the program packages CNS/ARIA (Brünger et al., 1998; Nilges et al., 1997) or DYANA (Güntert et al., 1997). Advantages of MUNIN include the direct classification of cross peaks into groups belonging to different N-H moieties prior to peak-picking and the absence of a need for Fourier transform in any dimension, opening an avenue to non-uniform sampling (Orekhov et al., 2001). Although the organization of various demixing steps in complex situations requires human intervention, the proper demixing steps are fully automated, providing reliable results for various criteria applied such as single complex exponents for FIDs or single maxima for frequency data. As with other types of NMR applications of MUNIN, e.g., to relaxation data (Korzhnev et al., 2001) or in drug discovery (Damberg et al., in press), the major strength of three-way decomposition stems from the simultaneous use of data from all three dimensions.

Acknowledgements

The azurin sample was a generous gift of B.G. Karlsson. Use of the facilities of the Swedish NMR Centre, help with recording the spectra by T. Papavoine and resonance assignments for the present conditions by I. Bezsonova are gratefully acknowledged. This research

is supported by NFR/VR grants K-AA/KU 12071-300 and K-AA/KU 12071-303.

References

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) *Acta Crystallogr.*, **D54**, 905–921.

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

Damberg, C.S., Orekhov, V.Y. and Billeter, M. (2002) *J. Med. Chem.*, in press.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.

Güntert, P., Mumenthaler, C. and Wüthrich, K. (1997) *J. Mol. Biol.*, **273**, 283–298.

Hajduk, P.J., Meadows, R.P. and Fesik, S.W. (1999) *Q. Rev. Biophys.*, **32**, 211–240.

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

Jarvoll, P. (2002) *Automatic Peak Picking*, Diploma Thesis, Göteborg University, Göteborg.

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

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

Koradi, R., Billeter, M. and Wüthrich, K. (1996) *J. Mol. Graph.*, **14**, 51–55, 29–32.

Korzhnev, D.M., Ibraghimov, I.V., Billeter, M. and Orekhov, V.Y. (2001) *J. Biomol. NMR*, **21**, 263–268.

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

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

Nar, H., Messerschmidt, A., Huber, R., van de Kamp, M. and Canters, G.W. (1991) *J. Mol. Biol.*, **221**, 765–772.

Nilges, M., Macias, M.J., O'Donoghue, S.I. and Oschkinat, H. (1997) *J. Mol. Biol.*, **269**, 408–422.

Orekhov, V.Y., Ibraghimov, I.V. and Billeter, M. (2001) *J. Biomol. NMR*, **20**, 49–60.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1992) *Numerical Recipes in C*, Cambridge University Press, Cambridge.

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