Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Magnetic Resonance 176 (2005) 47–53
JMR Journal of Magnetic Resonance
www.elsevier.com/locate/jmr

Signal identification in NMR spectra with coupled evolution periods

Daniel Malmodin, Martin Billeter *

Biophysics Group, Department of Chemistry, Göteborg University, Sweden

Received 23 March 2005; revised 17 May 2005
Available online 21 June 2005

Abstract

Novel multidimensional NMR experiments rely on modified time-domain sampling schemes to provide significant savings of experimental time. Several approaches are based on the coupling of evolution times resulting in a reduction of the dimensionality of the recorded spectra, and a concomitant saving of experimental time. We present a consistent and general tool, called EVOCOUP, for the analysis of these reduced dimensionality spectra. The approach is flexible in the sense that the input can consist of various forms of reduced dimensionality spectra, that any piece of information can be removed (provided enough information is left), e.g., signals undetectable due to poor signal-to-noise or covered by artifacts, and that it can be applied to spectra involving any number of nuclei. The use of a general optimization procedure and an appropriate target function provides for a robust approach with well-defined results and ensures optimal use of redundant information normally present in the input. Spectral overlap in the directly detected dimension is resolved in a fully automated manner, avoiding the assessment of signal quality and its use in combinatorial trials. The positions of all peaks in a corresponding full-dimensional spectrum are obtained without need for reconstruction of this spectrum. In a systematic analysis of a complete spectrum recorded for the 14 kDa protein azurin and involving five different nuclei, only four spin systems were missed and no false spins systems were detected.
© 2005 Elsevier Inc. All rights reserved.

Keywords: EVOCOUP; GFT; Projection–reconstruction; Reduced dimensionality; Singular value decomposition

1. Introduction

Fast multidimensional NMR spectroscopy has recently received considerable attention with the presentation of several novel methods (for reviews, see for example [1–3]). For example, Hadamard spectroscopy employs complex irradiation schemes to select specific frequencies and produces a subspectrum in a fraction of the time needed for the entire spectrum [4]. Single scan spectroscopy, where resonance information from different spins is spatially encoded, may in the future offer even larger time savings [5]. Other approaches allow non-uniform data sampling along the indirectly detected dimensions that can be optimized with respect to experiment time, resolution, and/or sensitivity. The resulting sparse data sets, which cannot be subjected to discrete Fourier transform, may be transformed directly with the help of maximum entropy reconstruction [6,7] or submitted to three-way decomposition to reconstruct uniformly sampled data [8–10].

Another type of saving of NMR instrument time is achieved by the coupling of different evolution times in a multidimensional experiment. Obviously, any elimination of an indirectly detected dimension caused by the coupling of the evolution time associated with one nucleus to that of another nucleus significantly shortens the time required for the experiment. A price to pay is that several chemical shifts are now encoded on a single frequency axis, and consequently the resulting spectral data contains peaks with positions defined by linear combinations of several chemical shifts rather than single chemical shifts. This concept, which relies itself on the earlier proposed "Accordion" spectroscopy [11], has been proposed more

* Corresponding author. Fax: +46 31 773 3910.
E-mail address: martin.billeter@chem.gu.se (M. Billeter).

1090-7807/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2005.05.020

than 10 years ago [12,13], and it has been realized in various types of experiments [14,15]. More recently, two novel approaches based on reduced dimensionality made the interest in this concept rise again. In GFT spectroscopy [16–18], scans of the pulse sequence are repeated with different combinations of 0 and 90° phase shifts (originally suggested in [19]) at the start of the second and subsequent evolution periods. Proper processing, including in particular multiplication of the spectral data with a "G-matrix," yields a set of spectra with peaks whose shifts along the reduced dimension are linear combinations of the chemical shifts of the nuclei involved. For example, the indirect dimension in spectrum one contains the shifts of a first nucleus, ω_1, in spectrum two and three it contains the sum and difference of shifts of two nuclei, $\omega_1 + \omega_2$ and $\omega_1 - \omega_2$, respectively, and so on. In a similar fashion, the so-called projection–reconstruction method [20–22] considers projections of two or more dimensions onto an axis oriented along user defined angles with respect to the original chemical shift axes, which again establishes a linear relation between the shifts measured on this new axis and the chemical shifts of the nuclei involved. As the name indicates, this approach emphasizes reconstruction of the full-dimensional spectrum from the recorded projection spectra.

When interpreting a set of spectra obtained by coupled evolution experiments one faces some additional problems compared to the analysis of a conventional, full-dimensional spectrum. For each spin system (defined by the nuclei coupled through the evolution periods in the experiment), one needs to identify a set of peaks with typically exactly one peak stemming from each spectrum. Exceptions include, for example, spin systems with glycine in experiments involving the chemical shift of α-protons. Together, these peaks should form a certain pattern according to the linear combinations of chemical shifts defined by the coupling of evolution periods in the experiment. Because of the spreading of signal intensities to the various members of such a peak set, signal-to-noise may become a critical issue. In this case, the identification of the peak patterns may be hampered by the difficulty in detecting some peaks. A similar situation arises if artifacts hide peaks, or with other types of data corruption. Overlap of chemical shifts, in particular in the directly detected dimension, will cause additional problems due to the mixing of sets of peaks belonging to two or more spin systems. Any approach designed for resolving these problems should rely on the fact that the input data contain a significant degree of redundancy. This in turn may result in ambiguous answers with combinatorial approaches that do not make strict use of a well-defined penalty function. The following approach for processing data from spectra recorded by coupling evolution periods, termed EVOCOUP, is designed to avoid or overcome most of the above mentioned problems.

2. Theory

The relation between observed peak shifts in spectra recorded in an experiment with coupled evolution periods and the true chemical shifts of the nuclei involved in the experiment can be expressed as a system of linear equations. Written in matrix form, the vector $\boldsymbol{\omega}$, which contains n unknown chemical shifts, needs to be determined from the vector $\mathbf{p}$ with m peak coordinates observed in m spectra of the experiment and a $(m*n)$-matrix $\mathbf{A}$ describing the linear combinations specific to a given experiment

$$\mathbf{A} \cdot \boldsymbol{\omega} \approx \mathbf{p}. \tag{1}$$

This general notation is valid for various types of experimental descriptions. Thus, assuming a 4D experiment with coupling of the three evolution periods, the total set of spectra in a (4, 2)D GFT experiment [16] would be described by the following matrix $\mathbf{A}$

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & -1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{pmatrix}. \tag{2}$$

Kupče and Freeman [21] have introduced a nomenclature, which describes coupling schemes by relations expressing evolution times by a common time parameter t and projection angles $\alpha, \beta, \ldots$ For the example of a 4D spectrum with evolution times t_1, t_2, and t_3, these relations would look as follows:

$$\begin{aligned} t_1 &= t \cos\alpha \cos\beta, \\ t_2 &= t \sin\alpha \cos\beta, \\ t_3 &= t \sin\beta. \end{aligned} \tag{3}$$

The corresponding matrix $\mathbf{A}$ takes the form

$$\begin{pmatrix} \cos\alpha \cdot \cos\beta & \sin\alpha \cdot \cos\beta & \sin\beta \\ \cos\alpha \cdot \cos\beta & -\sin\alpha \cdot \cos\beta & \sin\beta \\ \cos\alpha \cdot \cos\beta & \sin\alpha \cdot \cos\beta & -\sin\beta \\ \cos\alpha \cdot \cos\beta & -\sin\alpha \cdot \cos\beta & -\sin\beta \end{pmatrix}. \tag{4}$$

The number m of equations, i.e., spectra, must be at least as large as the number n of unknown chemical shifts. Typically, m is strictly larger than n and one has an over-determined system of equations. We assume that the matrix $\mathbf{A}$ has rank n, i.e., there are n linearly independent equations. The matrix $\mathbf{A}$ cannot be inverted, and Eq. (1) can in general not be strictly fulfilled (therefore the "≈" symbol in this equation). The optimal solution of Eq. (1) is the set of chemical

shifts $\boldsymbol{\omega}'$ that, when multiplied with the matrix **A**, best approximates (in a least squares sense) the experimental data **p**. This solution can be found in a general way, treating all equations simultaneously and without attributing special weight to any equation, using reduced singular value decomposition (see, for example [23]). With the above assumptions, the (m*n)-matrix **A** can be represented by a matrix **U** of size (m*n) (that is derived from an orthogonal matrix by truncation), an orthogonal (n*n)-matrix **V**, and a diagonal (n*n)-matrix **D** with strictly non-zero elements on the entire diagonal

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}. \tag{5}$$

Note that due to lack of orthogonality in **U**, the product $\mathbf{U}\mathbf{U}^{\mathrm{T}}$ is not equal to a unit matrix, preventing the determination of an inverse matrix of **A**. By introducing a *pseudoinverse* $\mathbf{A}'$ as follows

$$\mathbf{A}' = \mathbf{V}\mathbf{D}^{-1}\mathbf{U}^{\mathrm{T}}, \tag{6}$$

one can calculate a chemical shift vector

$$\boldsymbol{\omega}' = \mathbf{A}' \cdot \mathbf{p}. \tag{7}$$

Inserting this vector with the resulting chemical shifts into Eq. (1), one obtains a vector $\mathbf{p}'$ with approximate values of the experimental observations in **p**

$$\mathbf{p}' = \mathbf{A} \cdot \boldsymbol{\omega}'. \tag{8}$$

It can be shown that $\mathbf{p}'$ best approximates the experimentally observed shifts in **p** [23], and thus that $\boldsymbol{\omega}'$ represents the optimal solution of Eq. (1).

The length of the difference vector $\|\mathbf{p} - \mathbf{p}'\|$ is a good internal measure of the reliability of the resulting chemical shifts. In the analysis of a set of spectra obtained with coupled evolution periods one frequently encounters situations with consistent peak positions from most spectra, but with one or a few peak positions that do not properly match the pattern. This could be due to the use of an artifact peak, but also because of overlap along the directly detected dimension. In the latter case, it will be easy to detect peak combinations were different spin systems contribute each with more than one peak. The most challenging situations occur when all but one peak originate from the same spin system and only one peak comes from a different spin system. Therefore, the size of the maximal component of the difference vector, *maxdiff*, is used as internal test rather than the length of the entire vector

$$maxdiff = \max_i(p_i - p'_i) \quad i = 1, 2 \ldots m. \tag{9}$$

It should, however, be noted that in all applications presented in the figures and tables below, both measures *maxdiff* and $\|\mathbf{p} - \mathbf{p}'\|$ would properly discriminate between correct and wrong peak combinations.

3. Results and discussion

The goal in the calculations with EVOCOUP is to obtain optimal estimates of all chemical shifts for as many spin systems as possible while avoiding false positives. In particular, we explore (a) the influence of overlap in the directly detected HN dimension (up to five chemical shifts of different HNs are found in intervals of 0.02 ppm width for our test protein, azurin, with 128 residues), (b) the flexibility with respect to the input by using data sets that may be considered incomplete (e.g., from a GFT point of view), and (c) the robustness of the procedure by combining data sets that were not strictly recorded under identical conditions. The reliability and robustness of these shift calculations was tested on a GFT spectrum of the type (5,2)D HACACONHN, where a 5D spectrum involving the nuclei Hα, Cα, CO, N, and HN is reduced to a set of 2D spectra [16]. The following 13 spectra from this experiment were analyzed (spectra are identified by the linear combination of chemical shifts in the indirectly detected dimension): ω_{N}, $\omega_{\mathrm{N}} \pm \omega_{\mathrm{CO}} \pm \omega_{\mathrm{C}\alpha}$, and $\omega_{\mathrm{N}} \pm \omega_{\mathrm{CO}} \pm \omega_{\mathrm{C}\alpha} \pm \omega_{\mathrm{H}\alpha}$. The first spectrum in this series was recorded as a conventional ^{15}N HSQC at least one year earlier than the other 12 spectra, which were recorded as a single data set. The spectra characterized by the simple combinations $\omega_{\mathrm{N}} \pm \omega_{\mathrm{CO}}$, which are typically recorded as a separate data set, were not considered. Note that this particular choice of spectra and the difference in conditions during acquisition was made to explore points (b) and (c) above, and is clearly not meant as a recommendation. Fig. 1 shows, for a narrow chemical shift interval in the directly detected dimension around $\omega_{\mathrm{HN}} =$ 10 ppm, these 13 spectra (labeled basic, first and third order spectra in the GFT nomenclature).

A first test is based on the spectral slices shown in Fig. 1, which include the HN frequencies for Thr 52 with $\omega_{\mathrm{HN}} = 10.01$ ppm and Thr 113 with $\omega_{\mathrm{HN}} = 10.04$ ppm. While the difference of 0.03 ppm is in principle sufficient to distinguish the two spin systems, we assume complete overlap of the two ω_{HN} frequencies for the present test. Peak positions providing entries for the vector **p** in Eq. (1) were obtained by peak picking with the fully automated program AUTOPSY [24]. The resulting peak lists were used directly, i.e., without any manual interaction. Due to the assumed overlap along ω_{HN}, the peak lists provide an ambiguity of two peak candidates from each spectrum, and a total of $2^{13} = 8192$ different peak combinations from all 13 spectra for the two spin systems. With only two spin systems present it is still feasible to calculate $\boldsymbol{\omega}'$ according to Eq. (7) for all combinations and subsequently use *maxdiff* as defined in Eq. (9) to discriminate between correct and wrong combinations. Fig. 2 summarizes these 8192 calculations after their ordering according to increasing values of *maxdiff* by reporting the 20 combinations with lowest values for *maxdiff* (sol-

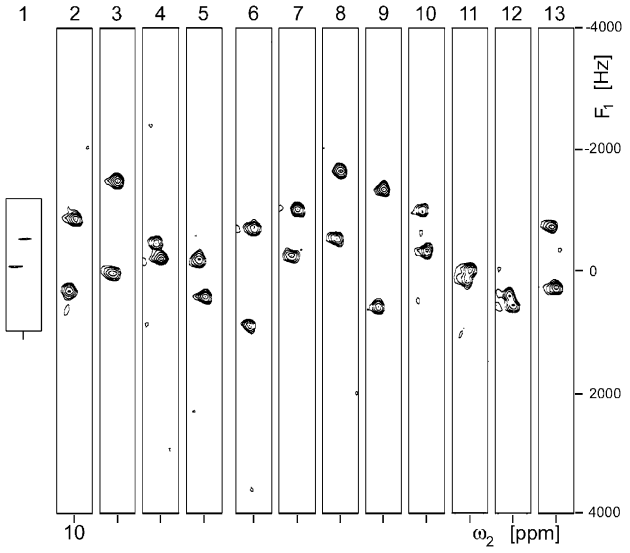

Fig. 1. ^{15}N HSQC spectrum and 12 spectra from a (5,2)D HACACONHN GFT experiment [16] recorded for the 128 amino acid long blue copper protein azurin from *Pseudomonas aeruginosa* [28]. Narrow strips around $\omega_2 = 10$ ppm along the directly detected HN dimension are shown with peaks for Thr 52 and Thr 113. The ^{15}N HSQC spectrum was recorded in a conventional way and at an earlier date; thus the smaller spectral width in ω_1. The GFT experiment was performed at 600 MHz for a 1 mM solution of reduced azurin labeled with ^{15}N and ^{13}C. Fifty-three and 733 complex points covering sweep widths of 8000 and 10000 Hz were recorded along the reduced and the HN dimension, respectively. Translated to a ppm scale for ^{15}N nuclei, the 8000 Hz in the mixed frequency dimension cover an interval from 54.1 to 185.7 ppm. The shift combinations along the indirect dimension for spectra 2–13 are as follows: $\omega_N + \omega_{CO} + \omega_{C\alpha}$, $\omega_N + \omega_{CO} - \omega_{C\alpha}$, $\omega_N - \omega_{CO} - \omega_{C\alpha}$, $\omega_N - \omega_{CO} + \omega_{C\alpha}$, $\omega_N + \omega_{CO} + \omega_{C\alpha} + \omega_{H\alpha}$, $\omega_N + \omega_{CO} + \omega_{C\alpha} - \omega_{H\alpha}$, $\omega_N + \omega_{CO} - \omega_{C\alpha} - \omega_{H\alpha}$, $\omega_N + \omega_{CO} - \omega_{C\alpha} + \omega_{H\alpha}$, $\omega_N - \omega_{CO} - \omega_{C\alpha} - \omega_{H\alpha}$, $\omega_N - \omega_{CO} - \omega_{C\alpha} + \omega_{H\alpha}$, $\omega_N - \omega_{CO} + \omega_{C\alpha} + \omega_{H\alpha}$, and $\omega_N - \omega_{CO} + \omega_{C\alpha} - \omega_{H\alpha}$. The enumeration of the spectra indicated at the top is used in the text.

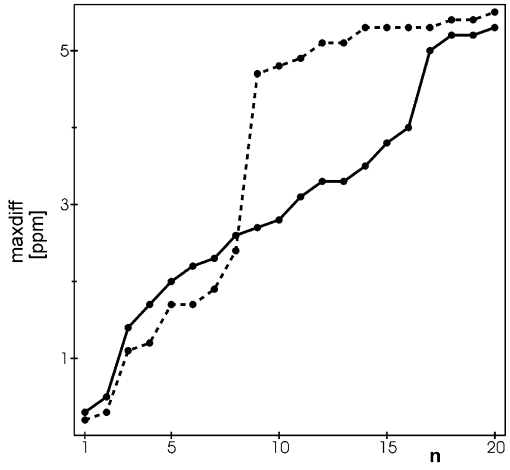

Fig. 2. Size of the maximal component *maxdiff* (see Eq. (9)) of the difference vector $\|\mathbf{p} - \mathbf{p}'\|$ for various combinations of peaks from Fig. 1 picked in all 13 spectra (solid line) or only in spectra 6–13 (dashed line). The 20 combinations with lowest values for *maxdiff* are enumerated (symbol *n*) in increasing order on the horizontal axis.

id line in Fig. 2). In a second set of calculations, only the eight spectra involving all four nuclei (i.e., with shifts according to $\omega_N \pm \omega_{CO} \pm \omega_{C\alpha} \pm \omega_{H\alpha}$; spectra 6–13 in Fig. 1) were used, providing $2^8 = 256$ different combinations (dashed line in Fig. 2). For both input sets, the first two combinations correspond to the correct assignments of Thr 52 and Thr 113. A significant increase for *maxdiff* is observed between the correct and the wrong combinations, both when using 13 or only 8 spectra, demonstrating that *maxdiff* is a reliable internal measure for resolving overlap situations. Situations with one wrong peak, typically exhibiting the next lowest values of *maxdiff* besides the correct combinations, are further examined in Table 1.

Five different calculations for Thr 52 of azurin are summarized in Table 1. The column with the first calculation reports the result from a fully correct input for this residue using all 13 spectra. The internal consistency criterion, *maxdiff*, is small with a value of 0.3 ppm. The resulting chemical shifts correspond closely to data obtained under comparable conditions [25]. In the next two columns, the input has been reduced by omitting the entries from spectrum 12 alone or from both spectra 12 and 9 of Fig. 1 (e.g., assuming unavailability due to overlap or data corruption). The internal test criterion provides the same values for *maxdiff*, and the deviations of the resulting chemical shifts from those of the first calculations are negligible (Table 1). Calculations 4 and 5 report results when a false peak position is entered, namely the peak of Thr 113 instead of that of Thr 52 from the spectrum 12 or the spectrum 9, respectively. Note that while the peak shifts for the two residues differ by about 30 ppm in spectrum 9, they

Table 1
Calculations of chemical shifts for Thr 52 from spectra of a (5,2)D HACACONHN GFT experiment on azurin

Calculation	1	2	3	4	5
Number of spectra used[a]	13	12	11	13	13
Incorrect entries[b]	none	none	none	12	9
maxdiff[c]	0.3	0.3	0.3	2.0	19.7
$\omega_N - \omega'_N$(ppm)[d]	—	0.02	0.05	0.22	2.42
$\omega_{CO} - \omega'_{CO}$(ppm)[d]	—	0.01	0.00	0.10	1.06
$\omega_{C\alpha} - \omega'_{C\alpha}$(ppm)[d]	—	0.01	0.00	0.10	1.06
$\omega_{H\alpha} - \omega'_{H\alpha}$(ppm)[d]	—	0.00	0.01	0.04	0.40

[a] Peak coordinates from 11, 12 or 13 of the GFT spectra shown in Fig. 1 were used. In calculation 2, no input was used from spectrum 12, and in calculation 3, input from spectra 9 and 12 was eliminated.

[b] The spectra strips in Fig. 1 show peaks for Thr 52 and Thr 113. Incorrect entries for the calculations for Thr 52 in this table refer to the use of peak coordinates for Thr 113. The number of the spectrum, from which an incorrect entry was taken, is given.

[c] This entity serves as internal check on the correctness of the results; see Eq. (9). The values for $\|\mathbf{p} - \mathbf{p}'\|$ for the five different calculations would be 0.6, 0.6, 0.5, 2.5, and 24.7.

[d] Individual chemical shift deviations with respect to the first calculations. The results from the first calculation coincide very closely with published chemical shifts obtained under comparable conditions [25].

almost coincide in spectrum 12 with a difference of only 3 ppm (Fig. 1). Nonetheless, the internal criterion based on *maxdiff* provides for the incorrect input from spectrum 12 a value that is about seven times higher than for the correct calculations. With incorrect input from spectrum 9, the difference increases to a factor of about 70. The errors in the resulting chemical shifts remain rather small when the wrong input from spectrum 12 is used, e.g., 0.04 ppm for Hα, but become significant for false input from spectrum 9, e.g., 0.4 ppm for Hα (Table 1). This clear detection of erroneous input by the internal criterion, even when the resulting shifts become only slightly incorrect, demonstrates the robustness and reliability of the approach. Thus, the use of 13 spectra from this (5,2)D GFT experiment provides a large degree of redundancy; the over-determination with only eight spectra is often sufficient to resolve overlap situations.

EVOCOUP calculations have also been performed systematically for the entire azurin protein. The input consisted of peak lists obtained directly from the program AUTOPSY [24] for all 13 spectra of Fig. 1. Due to limited sensitivity, these peak lists do not contain all expected peaks, e.g., no peaks could be observed by AUTOPSY in the spectra with mixed frequencies for HN of Asn38 with $\omega_{HN} = 11.36$ ppm. In this systematic analysis, the HN dimension was scanned with intervals of 0.02 ppm width in steps of 0.01 ppm, and all peaks observed in a given interval were considered simultaneously. As mentioned earlier, substantial overlap is observed along the HN dimension, requiring the simultaneous processing of up to five peaks from each projection, i.e., sorting out the five correct combinations out of a total $5^{13} \approx 10^9$ combinations. This prohibits exhaustive systematic calculations as for the examples above. A simple alternative approach was chosen that starts with a subset of equations from Eq. (1) and, provided a solution is found, adds more equations. Calculations were started with peak combinations from five equations, and if the maximal component *maxdiff* of the vector $\|\mathbf{p} - \mathbf{p}'\|$ was smaller than a cutoff of 0.5 ppm, additional equations were added exploring all combinations until this cutoff was exceeded or all 13 equations were used. If this procedure yielded a consistent set of peaks from all 13 spectra, this was considered an assignment, and the peaks used were scratched from the list of available peaks before continuing the search. This procedure was repeated until no more consistent combinations of 13 peaks were possible. Next, the requirement on the number of peaks was relaxed to 12 and so on down to a minimal number of required peaks of 8. Note that this simple method does not allow the use of a peak in two assignments, which may prohibit the detection of some frequencies such as those from both Hα atoms in glycines. With this simple procedure, all but four of the 123 azurin backbone spin systems with amide groups could be assigned. For one of the four missing spin systems ($\omega_{HN} = 11.36$), no peaks were detected by AUTOPSY except in the ^{15}N HSQC due to very poor signal-to-noise (the intensity of this peak in the ^{15}N HSQC is about five times smaller when compared to other peaks in this spectrum). The other three cases occurred in heavily overlapped regions around $\omega_{HN} = 8.70$ and $\omega_{HN} = 8.95$ ppm. Peaks of the missed spin systems overlap in the indirectly detected dimension with peaks from other spin systems with nearly identical ω_{HN}. For some of the peaks with overlap in both dimensions, the input peak lists contained only one entry, and when this had been used by one spin system it was not available to any further spin systems. The three spin systems did thus not have eight unique peaks left and were missed. For five of the 11 glycines, both $\omega_{H\alpha}$ frequencies could be detected. Four spin systems not belonging to backbone moieties also appeared; they all are caused by peaks assigned to side chains of asparagines or glutamines.

4. Conclusions

A consistent and general approach is described for the calculation of chemical shifts from spectra with coupled evolution periods. The result is always well-defined as the set of chemical shifts that can best approximate the linear combinations of these shifts observed in the experiments, i.e., the set of chemical shifts that minimizes the difference vector $\mathbf{p} - \mathbf{p}'$. Instead of the length of the vector, its largest component, *maxdiff*, was chosen to serve as internal test for the reliability of the result. The chemical shift calculations are robust in the sense that when a wrong peak is selected in one of the spectra, *maxdiff* will increase faster with the error in the peak position than the errors in the resulting chemical shifts. In the application to a 128 residues long protein, the approach yields no false results (detection of side chain related spin systems is not strictly wrong). It misses only in three cases, always due to extensive overlap in both dimensions. Overlap caused by projections can always be resolved by recording a properly chosen new projection [26]. As an alternative, we work on a modified search scheme that allows input peaks to be used for several spin systems while avoiding the need to search through all peak combinations (e.g., all 10^9 combinations in the case of five overlapping HN frequencies). In a fourth case, no peaks were provided from the peak picker as input. Picking peaks in projections, i.e., prior to searching for peak patterns in all projections, represents a weakness of the present approach that is shared with other methods based on the same order of events [27]. The reason is the reduced signal-to-noise in individual projections compared to a joint detection of all peaks stemming from a given spin systems. Reconstruc-

tion of the full-dimensional spectrum represents one attempt to avoid the loss in sensitivity related to peak picking in individual projections [22]. Because the approach is based on systematic searches, thus avoiding complicated combinatorial searches, it is simple to use. Input peak lists from various schemes of evolution coupling can be considered, and this flexibility extends also to the fact that any spectrum can easily be ignored, e.g., in the case of poor signal-to-noise, overlap with artifacts or other data corruption. This represents an advantage over methods that rely explicitly on doublet symmetry relations defined in their pattern models [27]. Obviously, the input data must be sufficiently complete to uniquely determine the results. Beyond this, any redundancy is used in an unbiased manner to resolve ambiguities, e.g., caused by overlap. The output corresponds to a conventional peak list with precise peak positions that would result from peak picking in a corresponding full-dimensional spectrum, making for many applications a reconstruction unnecessary. Several methods for the analysis of reduced dimensionality data are currently emerging, examples being spectrum reconstruction [20] and *PatternPicker* [27]. A systematic comparison among EVOCOUP and these methods is beyond the scope of this first presentation of EVOCOUP because of the so far limited number of applications presented for most methods, but also due to different types of output, such as reconstructed spectra versus peak lists. Nonetheless, the diversity of concepts among all these approaches is likely to result in major improvements through a synthesis of different ideas.

Acknowledgments

This project was supported by the Swedish Research Council (VR Grant 621-2003-4048) and the Swedish NMR Centre. We thank V.Yu. Orekhov for help with the spectroscopy, and B.G. Karlsson for generously making a doubly labeled azurin sample available.

References

[1] R. Freeman, E. Kupče, New methods for fast multidimensional NMR, J. Biomol. NMR 27 (2003) 101–113.
[2] R. Freeman, E. Kupče, Distant echoes of the accordion: reduced dimensionality, GFT-NMR, and projection–reconstruction of multidimensional spectra, Concepts Magn. Reson. 23A (2004) 63–75.
[3] D. Malmodin, M. Billeter, High-throughput analysis of protein NMR spectra, Prog. Nucl. Magn. Reson. Spectrosc. 46 (2005) 109–129.
[4] E. Kupče, T. Nishida, R. Freeman, Hadamard NMR spectroscopy, Prog. Nucl. Magn. Reson. Spectrosc. 42 (2003) 95–122.
[5] L. Frydman, T. Scherf, A. Lupulescu, The acquisition of multidimensional NMR spectra within a single scan, J. Am. Chem. Soc. 125 (2003) 11385–11396.
[6] A.S. Stern, K.-B. Li, J.C. Hoch, Modern spectrum analysis in multidimensional NMR spectroscopy: comparison of Linear-Prediction extrapolation and Maximum-Entropy reconstruction, J. Am. Chem. Soc. 124 (2002) 1982–1993.
[7] D. Rovnyak, D.P. Frueh, M. Sastry, Z.-Y.J. Sun, A.S. Stern, J.C. Hoch, G. Wagner, Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction, J. Magn. Reson. 170 (2004) 15–21.
[8] V.Yu. Orekhov, I. Ibraghimov, M. Billeter, Optimizing resolution in multidimensional NMR by three-way decomposition, J. Biomol. NMR 27 (2003) 165–173.
[9] M. Billeter, V.Yu. Orekhov, Three-way decomposition and nuclear magnetic resonance, in: Sloot, Abramson, Bogdanov, Dongarra, Zomaya, Gorbachev (Eds.), Computational Science-ICCS2003, Springer Verlag, Berlin, 2003, pp. 15–24.
[10] V. Tugarinov, W.-Y. Choy, V.Yu. Orekhov, L.E. Kay, Solution NMR-derived global fold of a monomeric 82-kDa enzyme, Proc. Natl. Acad. Sci. USA 102 (2005) 622–627.
[11] G. Bodenhausen, R.R. Ernst, Direct determination of rate constants of slow dynamic processes by two-dimensional ''Accordion'' spectroscopy in nuclear magnetic resonance, J. Am. Chem. Soc. 104 (1982) 1304–1309.
[12] T. Szyperski, G. Wider, J.H. Bushweller, K. Wüthrich, 3D ^{13}C-^{15}N-heteronuclear two-spin coherence spectroscopy for polypeptide backbone assignments in ^{13}C-^{15}N-double-labeled proteins, J. Biomol. NMR 3 (1993) 127–132.
[13] T. Szyperski, G. Wider, J.H. Bushweller, K. Wüthrich, Reduced dimensionality in triple-resonance NMR experiments, J. Am. Chem. Soc. 115 (1993) 9307–9308.
[14] K. Ding, A. Gronenborn, Novel 2D triple-resonance NMR experiments for sequential resonance assignments of proteins, J. Magn. Reson. 156 (2002) 262–268.
[15] W. Kozminski, I. Zhukov, Multiple quadrature detection in reduced dimensionality experiments, J. Biomol. NMR 26 (2003) 157–166.
[16] S. Kim, T. Szyperski, GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information, J. Am. Chem. Soc. 125 (2003) 1385–1393.
[17] S. Kim, T. Szyperski, GFT NMR experiments for polypeptide backbone and $^{13}C^{\beta}$ chemical shift assignment, J. Biomol. NMR 28 (2004) 117–130.
[18] H.S. Atreya, T. Szyperski, G-matrix Fourier transform NMR spectroscopy for complete protein resonance assignment, Proc. Natl. Acad. Sci. USA 101 (2004) 9642–9647.
[19] B. Brutscher, N. Morelle, F. Cordier, D. Marion, Determination of an initial set of NOE-derived distance constraints for the structure determination of $^{15}N/^{13}C$-labeled proteins, J. Magn. Reson. B 109 (1995) 238–242.
[20] E. Kupče, R. Freeman, Projection–Reconstruction of three-dimensional NMR spectra, J. Am. Chem. Soc. 125 (2003) 13958–13959.
[21] E. Kupče, R. Freeman, Projection–Reconstruction technique for speeding up multidimensional NMR spectroscopy, J. Am. Chem. Soc. 126 (2004) 6429–6440.
[22] E. Kupče, R. Freeman, Fast multidimensional NMR: radial sampling of evolution space, J. Magn. Reson. 173 (2005) 317–321.
[23] D.C. Lay, Linear Algebra, Addison-Wesley, New York, 2003.
[24] R. Koradi, M. Billeter, M. Engeli, P. Güntert, K. Wüthrich, Automated peak picking and peak integration in macromolecular NMR spectra using AUTOPSY, J. Magn. Reson. 135 (1998) 288–297.
[25] J. Leckner, Folding and Structure of Azurin—the Influence of a Metal, Chalmers University of Technology, Göteborg, Sweden, 2001.
[26] E. Kupče, R. Freeman, Resolving ambiguities in two-dimensional NMR spectra: the 'TILT' experiment, J. Magn. Reson. 172 (2005) 329–332.

[27] H.N.B. Moseley, N. Riaz, J.M. Aramini, T. Szyperski, G.T. Montelione, A generalized approach to automated NMR peak list editing: application to reduced dimensionality triple resonance spectra, J. Magn. Reson. 170 (2004) 263–277.

[28] B.G. Karlsson, T. Pascher, M. Nordling, R.H. Arvidsson, L.G. Lundberg, Expression of the blue copper protein azurin from *Pseudomonas aeruginosa* in *Escherichia coli*, FEBS Lett. 246 (1989) 211–217.