J Biomol NMR (2009) 43:197–210
DOI 10.1007/s10858-009-9304-4

ARTICLE

Chemical shift optimization in multidimensional NMR spectra by AUREMOL-SHIFTOPT

Kumaran Baskaran · Renate Kirchhöfer · Fritz Huber · Jochen Trenner · Konrad Brunner · Wolfram Gronwald · Klaus-Peter Neidig · Hans Robert Kalbitzer

Received: 8 September 2008 / Accepted: 18 December 2008 / Published online: 21 February 2009

Abstract A problem often encountered in multidimensional NMR-spectroscopy is that an existing chemical shift list of a protein has to be used to assign an experimental spectrum but does not fit sufficiently well for a safe assignment. A similar problem occurs when temperature or pressure series of *n*-dimensional spectra are to be evaluated automatically. We have developed two different algorithms, AUREMOL-SHIFTOPT1 and AUREMOL-SHIFTOPT2 that fulfill this task. In the present contribution their performance is analyzed employing a set of simulated and experimental two-dimensional and three-dimensional spectra obtained from three different proteins. A new *z*-score based on atom and amino acid specific chemical shift distributions is introduced to weight the chemical shift contributions in different dimensions properly.

K. Baskaran · R. Kirchhöfer · F. Huber · J. Trenner · K. Brunner · W. Gronwald · H. R. Kalbitzer (✉)
Department of Biophysics and Physical Biochemistry, University of Regensburg, Postfach, 93040 Regensburg, Federal Republic of Germany
e-mail: hans-robert.kalbitzer@biologie.uni-regensburg.de

K.-P. Neidig
Software Department, Bruker BioSpin GmbH, Silberstreifen 4, 76287 Rheinstetten, Federal Republic of Germany

Present Address:
R. Kirchhöfer · F. Huber
LipoFIT Analytic GmbH, Josef-Engert Str. 9, 93053 Regensburg, Federal Republic of Germany

Present Address:
W. Gronwald
Institute of Functional Genomics, University of Regensburg, Josef-Engert Str. 9, 93053 Regensburg, Federal Republic of Germany

Keywords Chemical shift · Peak assignment · Multidimensional NMR spectra · AUREMOL

Introduction

Nuclear Magnetic Resonance is an important tool for structure elucidation of biological macro molecules, and quite useful to study the dynamical behavior of molecules. Measuring the "true" chemical shifts accurately in experimental spectra is not straightforward in NMR because of severe overlap of resonance peaks and the presence of noise and artifacts. Here, a number of optimized peak picking routines were developed (Neidig et al. 1984; Glaser and Kalbitzer 1987). The inverse problem, the projection of known chemical shifts (assignments) to an experimental spectrum is also not trivial because of the same reasons: the peak maximum may be shifted by noise or by superposition with other peaks or artifacts. In addition, the digital resolution provides a general limit of accuracy. However, by far the most important problem is caused by chemical shift variations due to temperature shifts or small changes of the sample composition and the buffer conditions (e.g., pH and ionic strength). Here, the already existing chemical shift table (usually created from a large set of multidimensional spectra) does not correspond exactly to the spectrum under investigation. Another application would be TROSY-spectroscopy where the cross peaks are shifted by *J*/2. Since for structural determination information from whole sets of nD-spectra has to be combined, the variation in chemical shifts between the different spectra has to be taken into account. In principle, chemical shift recognition is part of automated procedures for assigning peaks in multidimensional spectra. Several

automated peak assignment procedures are reported in the literature (Catasti et al. 1990; Hare and Prestegard 1994; Zimmerman et al. 1997; Xu et al. 2001; Herrmann et al. 2002; Gronwald et al. 2002) using neural networks and other optimization techniques. But all these methods are aimed at structure elucidation and not giving much importance to the optimization of chemical shift of every atom. For the chemical shift optimization of an individual spectrum that is recorded at different conditions they are not useful.

In the present paper we propose two different algorithms to adapt a given chemical shift table optimally. They are compared with each other and their accuracy for different spectral types is assessed. The proposed algorithms can also be used in other fields such as metabolomics where an alignment of peaks in spectra of different mixtures improves a multivariate analysis.

Materials and methods

Simulation of NOESY data sets

Test data sets were created by spectral simulation of HPr (histidine containing phosphocarrier protein) from Staphylococcus aureus, a 88 residue phosphocarrier protein. The ^{1}H and ^{15}N NMR spectra are completely assigned (Maurer et al. 2004) and are deposited in the BioMag data base. The NMR structure is deposited on the PDB data base (PDB ID:1ka5). A 2D NOESY spectrum of HPr was simulated using RELAX (Görler and Kalbitzer 1997; Gronwald et al. 2000) module in AUREMOL (Gronwald and Kalbitzer 2004). Test data set was created using 466 proton chemical shifts of HPr protein with a mixing time of 250 ms, a relaxation delay of 1.75 and having 2,048 data points in both dimensions. The spectrometer frequency was set to 800.2 MHz. The simulation used an overall rotational correlation time of 3.9 ns, included internal mobility on the basis of the model-free approach with standard main chain and side chain order parameters, fast methyl rotations and slow ring flip motions. J-coupling and chemical shift anisotropy was not included in the simulation. The detection limit was set to 0.5 nm leading to 9,035 resonance peaks in the simulated NOESY spectrum.

3D carbon and nitrogen edited NOESY-HSQC spectra were simulated analogously. The chemical shift table of carbons contained 368 entries that of nitrogen 95 entries. The digital resolutions in $\delta 1$, $\delta 2$, and $\delta 3$ were 0.0987, 2.198, and 0.016 ppm/point for the ^{13}C edited spectra. The proton resonance frequency was 800.2 MHz. For the ^{15}N edited spectra, the digital resolutions in $\delta 1$, $\delta 2$, and $\delta 3$ were 0.0987, 2.055, and 0.011 ppm/point.

Gaussian noise was added to the simulated spectrum with a standard deviation scaled to the mean cross peak intensity of the spectrum <I>. Ten percent noise would correspond to a standard deviation σ of 0.1 <I>. The noise is created by randomly picking probability densities p(z) at the normalized intensity z from a Gaussian distribution function with a mean of zero. A second random-number generator was used to decide if z is accepted or not. The random numbers x are projected to the interval [0, p_{max}]. For $x \leq p(z)$ z is accepted, otherwise rejected. (Box and Muller 1958).

Experimental test spectra

Experimental ^{1}H, ^{15}N-HSQC spectra of HPr from *S. carnosus* were recorded at different pressures and temperatures as described earlier (Kalbitzer et al. 2000). The ^{1}H frequency was 750 MHz. The ^{1}H and ^{15}N spectral widths were 13.9474 and 50.012 ppm, respectively. The corresponding data size of the time domain data were 2,048 × 256 complex data points. The spectrum recorded at a temperature of 298 K and a pressure of 3 MPa was manually reassigned with the data published earlier (Görler et al. 1999a, b) and was used as a reference spectrum to assign the remaining series of the spectra.

In addition a three-dimensional HNCA spectrum of Saratin (Gronwald et al. 2008) was used. The spectral widths were 13.98, 41.11, and 41.41 ppm, in the ^{1}H, ^{15}N and ^{13}C dimensions, respectively. The corresponding size of the time domain data was 2,048 × 64 × 128, respectively. The proton resonance frequency was 600 MHz.

Theoretical considerations

Chemical shift optimization

Our aim is to find chemical shift values that optimally explain a given spectrum starting from an initial chemical shift table $\mathbf{S}_0$ for atoms (spins) or group of atoms (e.g., methyl protons) (j) $\{\delta(j) | j = 1,\ldots,J\}$. The chemical shifts $\delta(j)$ can be degenerated that is more than one atom in the protein may have the same chemical shift ($\delta(j) = \delta(k)$). In principle, an atom can have more than one chemical shift $\delta(j)$ value when it occurs in different states n (e.g., local conformations of the proteins), a fact that can be described by introduction of the corresponding superscript to $\delta^{n}(j)$. In addition, in experimental spectra where a spin is represented in more than one dimension (as in typical homonuclear spectra) the experimental chemical shifts of the same atom (spin) may be different because (1) errors with referencing did occur or because (2) differences in digital resolution dominate the peak positions. Error (1) can be reduced to ±1 data point by carefully referencing the spectra, error (2)

cannot be avoided but again it should be smaller/equal to one data point. The list $\mathbf{S}_0$ is usually incomplete in protein spectra since often for some atoms of a protein the resonance frequencies cannot be identified. At the end a final chemical shift table $\mathbf{S}_f$ is generated that optimally fulfills some optimization criteria often in the form of a target function.

The experimental *N*-dimensional spectrum contains cross peaks at positions $\boldsymbol{\delta}_i$,

$$\boldsymbol{\delta}_\mathbf{i} = \begin{pmatrix} \delta_1^i \\ \cdots \\ \delta_N^i \end{pmatrix} \quad (1)$$

where the possible combinations of the components of vector $\boldsymbol{\delta}_i$ depend on the actual type of experiment and the sample composition. In a classical multidimensional spectrum the allowed frequencies are a subset of all resonance frequencies $\delta(j)$. Assignment of a cross peak at $\boldsymbol{\delta}_i$ would mean the assignment of different $\delta(j) \in \mathbf{S}_0$ to all components δ_k^i $(k = 1,...,N)$ of $\boldsymbol{\delta}_i$. The chemical shift optimization can be then formulated as the search for a diagonal matrix $\mathbf{A}$ with

$$\mathbf{S}_f = (1 + \mathbf{A})\mathbf{S}_0 \quad (2)$$

and the initial and the final chemical shift tables $\mathbf{S}_0$ and $\mathbf{S}_f$ written as vectors. The simulation of the spectra with $\mathbf{S}_f$ should optimally explain the experimental spectrum under consideration. In an ideal, noise-free spectrum all experimental cross peaks at positions $\boldsymbol{\delta}_i^e$ should be explained by simulated cross peaks at positions $\boldsymbol{\delta}_m^s$.

Algorithms and general definitions

Starting with a chemical shift table $\mathbf{S_0}$ as input two different procedures were developed here that are combined in AUREMOL-SHIFTOPT. Both approaches contain as a key element the comparison between simulated and experimental spectra. They are tested in this paper on different types of spectra for showing their advantages and disadvantages. Both methods include a set of common procedures that already exist in AUREMOL, namely a peak picking procedure (Neidig et al. 1984), the calculation of peak probabilities to discriminate true resonances from artifacts and noise (Antz et al. 1995; Schulte et al. 1997), the assignment of peaks from the input shift table, the calculation of the assignment probability from the chemical shift deviation (and optional additional information), and the calculation of a corrected chemical shift table from the probability weighted cross peak coordinates (Fig. 1). The last three steps may be iterated until the target function is optimized.

In NOESY-type spectra the peak volume is an important additional source of information. Experimental peak

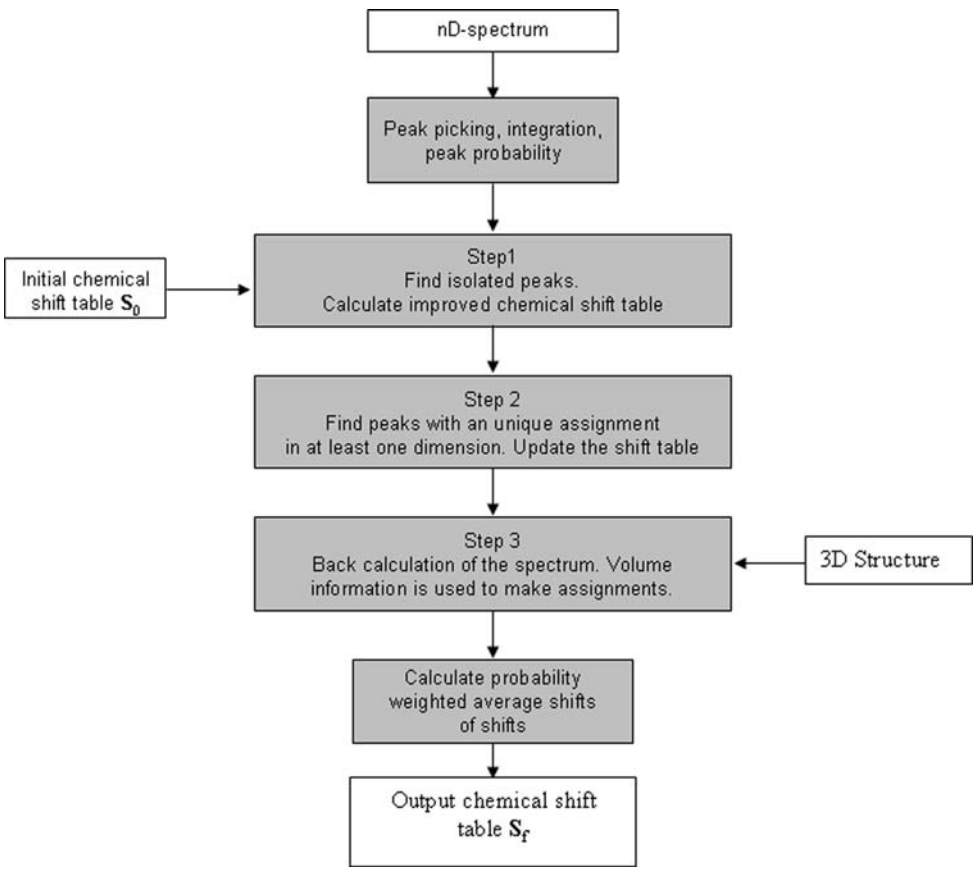

Fig. 1 The three-step chemical shift optimization of SHIFTOPT1

volumes are obtained by an iterative segmentation procedure (Geyer et al. 1995), the simulated peak volumes are calculated on the basis of the full relaxation matrix formalism (Görler and Kalbitzer 1997; Görler et al. 1999a, b; Ried et al. 2004). When an experimental structure is not yet available, the relaxation matrix is built from the most likely pair wise distances obtained from an unbiased structural data base (Nasser et al. manuscript in preparation). In other cases the spectrum is simulated by a procedure established in AUREMOL that allows to predict any *n*-dimensional spectrum from a chemical shift list and an internal description of the expected cross peak patterns on a semi quantitative basis (Brunner et al. manuscript in preparation).

Since it is the general assumption that a spectrum back calculated with $\mathbf{S}_0$ should be similar to the experimental spectra, the corresponding experimental cross peaks should not be too far away from the simulated cross peaks in terms of a distance metric in the chemical shift space.

In a multidimensional NMR spectrum with the dimension N the cross peak position is defined by the N-dimensional vector $\boldsymbol{\delta}_\mathbf{i}$. Since the digital resolution R_k in all dimensions k is usually not identical, the precision of the peak position is also different in the different dimensions. A lower error bound T_k^l is given

$$T_k^1 = \pm\left(R_k + \frac{L_k}{2}\right) \tag{3}$$

L_k is the expected line width in dimension k. The upper error bound T_k^u depends on the experimental conditions and defines the initial search range for a peak. The vectors

$$\mathbf{T}^l = \begin{pmatrix} T_1^l \\ \cdots \\ T_N^l \end{pmatrix} \tag{4}$$

$$\mathbf{T}^u = \begin{pmatrix} T_1^u \\ \cdots \\ T_N^u \end{pmatrix} \tag{5}$$

They are used to define the upper and lower search range in all N dimensions.

SHIFTOPT1

This first approach aims mainly on cases where experimental NOESY type spectra plus additional structural information are available for shift optimization. Employing the initial shift list and the structural information NOESY spectra are simulated. However, it is not limited to NOESY-spectra but can be applied to other types of spectra. The simulated spectra predict the frequency combinations $\boldsymbol{\delta}_\mathrm{m}$ where cross peaks s_m^s are to be expected and the approximate cross peak volume for a given structure. Peak positions and volumes in the simulated spectra can be directly compared with those in the experimental spectra obtained by peak picking and peak integration of the experimental cross peaks s_j^e. Since the experimental spectrum contains true signals as well as noise and artifact peaks the signals are classified by a Bayesian analysis (Antz et al. 1995; Schulte et al. 1997) and probabilities $\mathrm{P_j}$ are calculated that an experimental cross peak is a true signal. The experimental peaks inside the search areas are sorted according to their probability and all peaks with a low signal probability $\mathrm{P_j} < \mathrm{P_t}$ are removed from the peak list, the other peaks are included in the analysis but with consideration of $\mathrm{P_j}$. $\mathrm{P_t}$ is calculated in such a way that in the search area 10% more peaks are left as expected from the spectrum simulation. Another criterion for the general acceptance of cross peaks is the cross peak volume that should be in the correct range. In the NOESY-spectra cross peaks with intensities that are significantly larger than that of the smallest possible H–H distance of 0.18 nm can be omitted from the assignment procedure. When a reliable structure exists, a more detailed volume comparison can be done on the basis of an assignment hypothesis. Here, the full relaxation matrix formalism provides good estimates of the cross peak intensities to be expected.

The actual procedure used is a three step procedure (Fig. 1). Step 1 selects those cross peaks s_j^e that are isolated and uniquely assignable to a simulated peak s_m^s inside the error limits of $\boldsymbol{\delta}_\mathrm{m} \pm \mathbf{T}^\mathrm{u}$. The corresponding chemical shift values of $\mathbf{S}_1$ are set to new values. In step 2 those cross peaks are selected where at least in one dimension a unique chemical shift assignment exists. Here, as well as in the step 3 (no unique chemical shifts) volume information is used to solve ambiguities. Simultaneously with the assignment procedure the chemical shift tolerance is reduced for individual shifts to $\mathbf{T}^\mathrm{l}$.

The peak assignment and chemical shift refinements are performed in an iterative way. Initially, all experimental peaks are unassigned and are part of list of unassigned experimental peaks (*U*-list) that consists of the experimental peak positions $\boldsymbol{\delta}_\mathrm{i}^\mathrm{e}$, the volumes V_i^e and the probability values $\mathrm{P_i}$. In step 1 first the uniquely assignable peaks are identified and written to the *A*-list. Starting with an arbitrary peak $\mathrm{s}_\mathrm{i}^\mathrm{e}$ from the *A*-list with chemical shifts $\boldsymbol{\delta}_\mathrm{i}^\mathrm{e}$ the chemical shift table $\mathbf{S}$ is slowly refined by updating iteratively their chemical shifts $\delta(\mathrm{j})$

$$\delta(j) = \delta(j) + \xi_i\left(\delta_{i,k}^e - \delta(j)\right) \tag{6}$$

with

$$\xi_i = \frac{P_i}{a + P_i} \tag{7}$$

and $\delta_\mathrm{i,k}^\mathrm{e}$ the chemical shift coordinate assigned to $\delta(\mathrm{j})$. Using the updated chemical shift table $\mathbf{S}$ a new *A*-list is

created that may now contain different peaks and the procedure is repeated. ξ_i controls the influence of an individual experimental peak on the updated chemical shift table **S**. Especially it is ensured that the influence of artifact and noise peaks is limited by the inclusion of the peak probability values P_i in ξ_i, while the parameter a controls the general influence of an individual experimental peak. Initially the parameter a is set to 10 and is now increased by 1. After 5 cycles the tolerance is reduced to max ($\mathbf{T}^u/2$, $\mathbf{T}^l$) and the procedure is repeated N-times (typically $N = 15$). The final A-list now contains the uniquely assigned cross peaks. Simultaneously the chemical shift table **S** has been updated and consists of two subsets S_{nr} and S_r that contain the non-refined and the refined chemical shift values, respectively. In case of NOESY spectra the peaks of the final A-list are used to normalize a back calculated NOESY-spectrum for the next steps.

In step 2 again all cross peaks are scanned iteratively (usually 20 times) with the new chemical shift table **S** and the corresponding error bounds. Only cross peaks are accepted that fulfill the condition that at least in one dimension an unambiguous assignment is possible. In addition, for NOESY spectra the peak volume V must be inside the allowed range corresponding to the distance range between 0.18 and 5 nm. If more than one assignment is possible, the assignment of a cross peak s_i^e to a simulated peak s_m^s where the deviation of chemical shifts is smallest and the volume is closest to the expectation is taken. For making that decision a z-score Q is defined as

$$Q(s_m^s) = \frac{1}{b}\sqrt{\sum_{k=1}^{n} z_k^2(s_m^s) + z^2(s_m^s)} \tag{8}$$

With n the dimension in the N-dimensional spectrum considered. The parameters z_k are defined by

$$z_k = \frac{\delta_k(s_m^s) - \delta_k(s_i^e)}{\sigma_k} \tag{9}$$

with $\delta_k(s_m^s)$ and $\delta_k(s_i^e)$ the chemical shifts in dimension j of the experimental and the simulated peaks. The expected standard deviations σ_k are defined by

$$\sigma_k = T_k^l/2 \tag{10}$$

The parameter z is defined as a function of the normalized cross peak volumes

with

$$\sigma = \frac{|V_{max}^{-1/6} - V_{min}^{-1/6}|}{2} \tag{11}$$

and V_{max} and V_{min} the volumes corresponding to the smallest possible distance (0.18 nm) and the maximum detectable distance (0.5 nm), respectively. The constant b in Eq. 8 is given by

$$b = \sqrt{\frac{V_{min}^{-1/6}}{\sigma} + \sum_{k=1}^{n} \frac{T_k^2}{\sigma_k^2}} \tag{12}$$

with T_k the actual error range of the chemical shift. Only solutions with $Q \leq 0.5$ are considered. If such a solution exists the assignment with the lowest Q value is taken and the shift list is updated as in step 2.

In the last step also peaks with ambiguous assignments are taken and the solution with the lowest Q value is selected. After typically 20 cycles the final shift list is calculated from all assigned cross peaks as the average weighted with the peak probabilities P_i, i.e., when the component k of the chemical shift vectors $\boldsymbol{\delta}_i$ (i = 1,.., N) is assigned to a specific atom j then

$$\delta_{final}(j) = \frac{1}{\sum_{i=1}^{N} P_i} \sum_{i=1}^{N} P_i \delta_i(k) \tag{13}$$

with P_i the Bayesian peak probability. When an atom is represented in more than one dimension in a spectrum, e.g., in dimension k and p, than the P_i is replaced by P_i^* to

$$P_i^* = \frac{D_R(k) P_i}{D_R(k) + D_R(p)} \tag{14}$$

SHIFTOPT2

SHIFTOPT2 represents an alternative way to optimize a chemical shift table and is useful for spectra with a not too large number of cross peaks as HSQC or HNCA spectra. Again a model spectrum is generated from the given input chemical shift table. This model spectrum is compared with corresponding experimental spectrum. The general preparation of the data corresponds to SHIFTOPT1 where after peak picking and Bayesian analysis the most probable experimental peaks s_i^e in the search areas are selected as defined above. The number of simulated cross peaks is reduced to n_s cross peaks by removing all simulated cross peaks s_m^s where an experimental cross peak s_i^e does not exist with $\boldsymbol{\delta}_i^e$ inside the error limits of $\boldsymbol{\delta}_m^s \pm \mathbf{T}^u$.

Two probability matrices $\mathbf{Q}^e$ and $\mathbf{Q}^s$ are constructed for the experimental and simulated cross peaks with the elements Q_{im}^e and Q_{im}^s. Q_{im}^e represents a measure for the probability of an experimental peak s_i^e at position $\boldsymbol{\delta}_i^e$ to be assigned to a simulated peak s_m^s at position $\boldsymbol{\delta}_m^s$. Q_{im}^s represents a measure for the probability of a simulated peak s_m^s at position $\boldsymbol{\delta}_m^s$ to be assigned to the experimental peaks s_i^e at position $\boldsymbol{\delta}_i^e$. Q_{im}^e and Q_{jm}^s are given analogously to Eq. 8 as a function of a generalized variable $\mathbf{z}^{im}$. The components z_k^{im} of the vector $\mathbf{z}^{im}$ are defined by

$$z_k^{im} = \frac{\delta_i^e - \delta_m^s}{p_i \sigma_j} \tag{15}$$

The values $1/\sigma_j$ are atom and amino acid specific weighting factors with respect to the assignment of the simulated peak as defined earlier (Schumann et al. 2007). For atoms or molecules not contained in the data base the averages of the data base are taken with $\sigma(^1H)$ 1.55 ppm, $\sigma(^{15}N)$ 0.236 ppm, and $\sigma(^{13}C)$ 0.447 ppm (Schumann et al. 2007). The signal probability p_i is obtained from the Bayesian analysis of the experimental spectra. The elements Q_{im}^e are given as

$$Q_{im}^e = \frac{\exp\left(\frac{-(\mathbf{z}^{im})^2}{2}\right)}{\sum_r \exp\left(\frac{-(\mathbf{z}^{ir})^2}{2}\right)} \tag{16}$$

with summation over all peaks s_r^s in the search range. Correspondingly, Q_{im}^s is defined by

$$Q_{im}^s = \frac{\exp\left(\frac{-(\mathbf{z}^{im})^2}{2}\right)}{\sum_s \exp\left(\frac{-(\mathbf{z}^{sm})^2}{2}\right)} \tag{17}$$

with summation over all peaks s_s^e in the search range.

From the two matrices an averaged probability matrix $\mathbf{Q}$ is calculated by

$$\mathbf{Q} = \frac{\mathbf{Q}^e + \mathbf{Q}^s}{2} \tag{18}$$

In step 1 pairs of peaks s_i^e and s_m^s with $Q_{im} = 1$ are uniquely assigned and all matrix elements Q_{ik} ($k = 1,\ldots, n_s$; $k \neq m$) and Q_{lm} ($m = 1,\ldots, n_e$;$l \neq i$) are set to 0. For the remaining peaks new probabilities are calculated from the reduced set of peaks. The procedure is repeated until no further element with $Q_{im} = 1$ are found. In step 2 the maximum element Q_{im}^{max} of $\mathbf{Q}$ is identified and $\mathbf{Q}$ is renormalized by $\boldsymbol{Q} = \mathbf{Q}/Q_{im}^{max}$. Step 1 is repeated but only

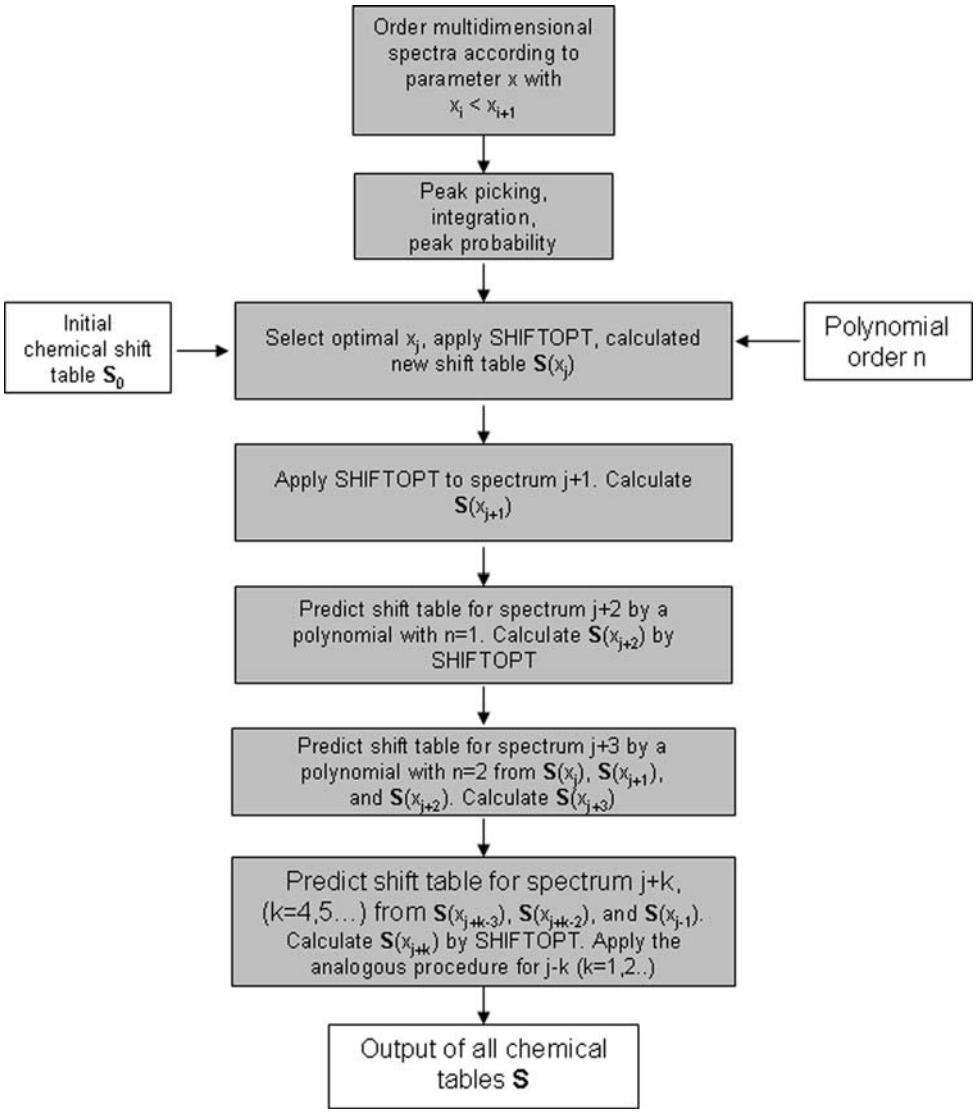

Fig. 2 Schematical view of the chemical shift optimization in a series of spectra. Note that a polynomial function of the order of 2 is used. In general $n + 1$ chemical shift lists are required for the prediction

peaks are taken as assigned where $Q_{im} = 1$ and $Q_{ik} < 1$ ($k = 1,\ldots, n_s$; $k \neq m$) and $Q_{lm} < 1$ ($m = 1,\ldots, n_e$;$l \neq i$) holds. This procedure is repeated until no new assignments are found.

Adaptation of assignments to a series of spectra

In many cases the chemical shifts in a series of *n*-dimensional NMR spectra can be represented as continuous functions of a parameter *x* (such as temperature, pressure, pH or ligand concentration). When more than two spectra are available the approximate positions of cross peaks in spectrum i + 1 can be predicted from the already assigned spectra by a polynomial of the order n set by the user.

In SHIFTOPT the following strategy is used (Fig. 2): First the spectra are ordered according to the parameter x in such a way that for all spectra i and i + 1 $x_i < x_{i+1}$ holds. In a next step that spectrum j is selected where the x(j) is closest to the conditions where the chemical shift table $\mathbf{S}_0$ fits to. The spectrum is assigned with SHIFTOPT and the obtained chemical shift table **Sj** is used to assign spectrum j + 1 or j − 1. If $x_{j+1} - x_j \leq x_j - x_{j-1}$ holds spectrum $k = j + 1$ is selected, otherwise spectrum j − 1. For the following we describe the algorithm for increasing values of k but it can (and generally has to) be applied also for decreasing values of *k*, $k < j$. The chemical shifts $\delta_m(s_i^s)$ of the simulated peaks in dimension m of spectrum $k = j + 2$ are then predicted by a polynomial of order 1 from the optimized $\delta_m(s_i^e)$ from spectrum j and j + 1. After optimization of the chemical shift list for x_{j+2}, new coefficients $a_{ml,}$ are calculated by

$$\delta_m\left(s_i^s, x_k\right) = \sum_{l=0}^{n} a_{ml} x_k^l \tag{19}$$

The coefficients a_{ml} can be calculated by rewriting Eq. 19 in matrix notation, using the Vandermonde matrix

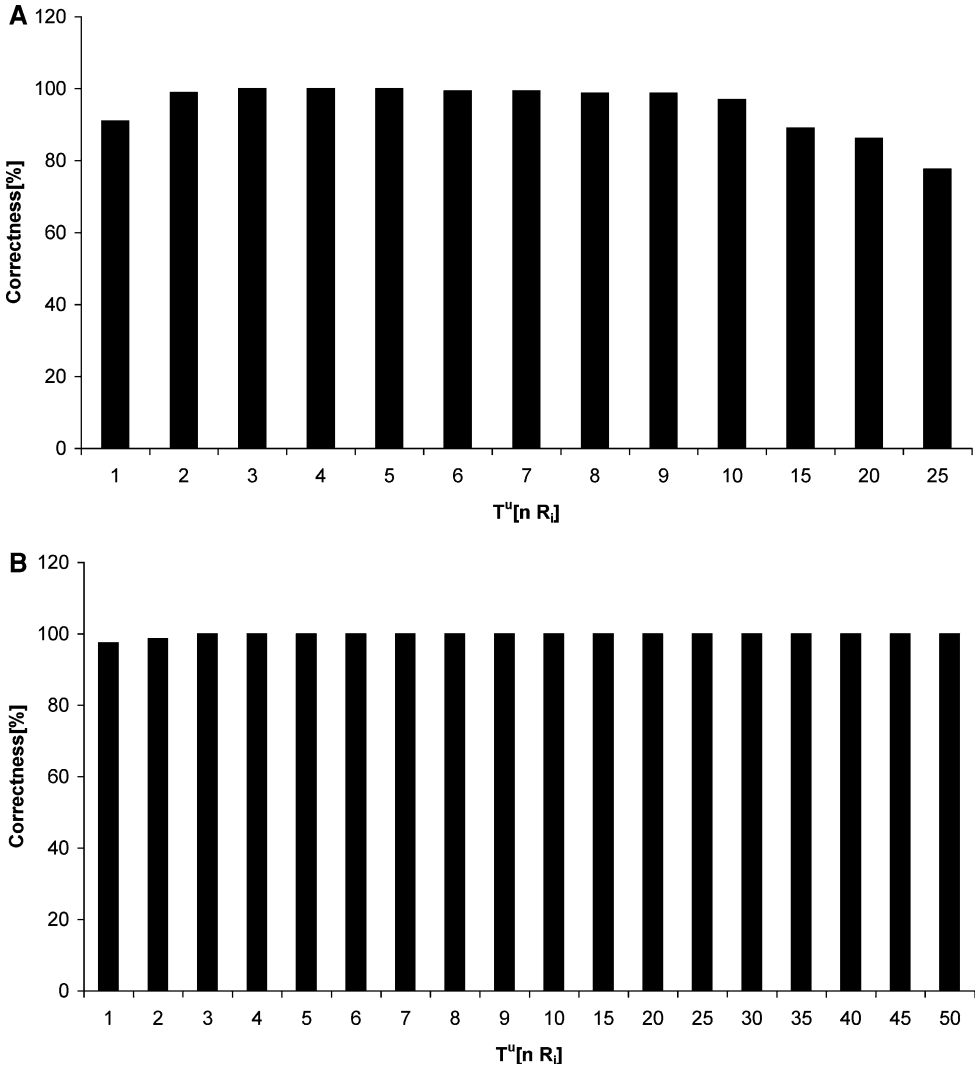

Fig. 3 Stability of the algorithms as a function of the search range T^u. **a** Application of SHIFTOPT1 to a simulated 2D NOESY-spectrum HPr from *S. aureus* with a ^{1}H digital resolution R_i in both dimensions of 0.0062 ppm and **b** application of SHIFTOPT2 to an experimental 2D-^{1}H, ^{15}N HSQC-spectrum from HPr from *S. carnosus* with a ^{1}H and a ^{15}N digital resolution R_i of 0.0068 and 0.195 ppm, respectively. The percentage of correct solutions with ($|\delta_i^{opt} - \delta_i^s| \leq R_i$) is plotted as a function of *n* with *n* a multiple of the search range $T^u = n \cdot R_2$. δ_i^{opt} and δ_i are the chemical shift values after optimization and the correct values before optimization, respectively

$$\begin{bmatrix} 1 & x_0 & x_0^2 & . & . & . & x_0^n \\ 1 & x_1 & x_1^2 & . & . & . & x_1^n \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 1 & x_n & x_n^2 & . & . & . & x_n^n \end{bmatrix} \begin{bmatrix} a_{m0} \\ a_{m1} \\ . \\ . \\ . \\ a_{mn} \end{bmatrix} = \begin{bmatrix} \delta_m\left(s_i^s, x_0\right) \\ \delta_m\left(s_i^s, x_1\right) \\ . \\ . \\ . \\ \delta_m\left(s_i^s, x_n\right) \end{bmatrix} \quad (20)$$

The coefficients a_{ml} can be obtained from the linear Eq. 20 using standard techniques for solving simultaneous equations (William et al. 1992).

Results

Stability of the search algorithms

The size of the search area $\mathbf{T}^u$ determines the number of possibilities to solve the optimization problem. A too small size will lead to incorrect solutions for resonances outside the search interval. A too large size increases the computational time and may also increase the ambiguities and thus the error probability. Therefore, we have systematically increased the size of the search range from a value corresponding to the digital resolutions R_i to very large values (Fig. 3). Two different spectra were used, a simulated 2D-NOESY spectrum containing 9,035 cross peaks and an experimental ^{1}H, ^{15}N-HSQC spectrum containing 79 cross peaks. The spectra were subjected to peak picking, integration and the Bayesian peak recognition routine of AUREMOL was applied. When we start with the correct chemical shift table, false chemical shift values are only obtained when the search range $\mathbf{T}^u$ is either too small or at very large values of $\mathbf{T}^u$. At very small search ranges the peak maximum may be shifted out of the search range since overlapping of peaks may shift the peak maxima. Very large search ranges increase the ambiguities and thus may lead to errors. This is especially important in the crowded NOESY-type spectra. Here, SHIFTOPT1 gets some incorrect result after the search range is increased to values larger than 0.03 ppm and becomes significantly less efficient at values larger than 0.06 ppm (48 Hz), whereas for the much less crowded HSQC spectrum the size of the search range virtually does not play any role.

However, when an ideal spectrum without noise and artifacts and infinitely small line widths was created by multiplying σ(H) with γ_H/γ_X. **a** Simulated 2D NOESY-spectrum with a ^{1}H digital resolution of 0.0062 ppm, application of SHIFTOPT1. The spectrum contains 9,035 cross peaks from the protein. **b** as (**a**) but after application of SHIFTOPT2. **c**, **d** Simulated 3D ^{15}N edited NOESY spectrum with a ^{1}H digital resolution of 0.005 and of 0.098 ppm in the direct and indirect dimension, a ^{15}N digital resolution of 0.764 ppm. Only data for SHIFTOPT1 are shown

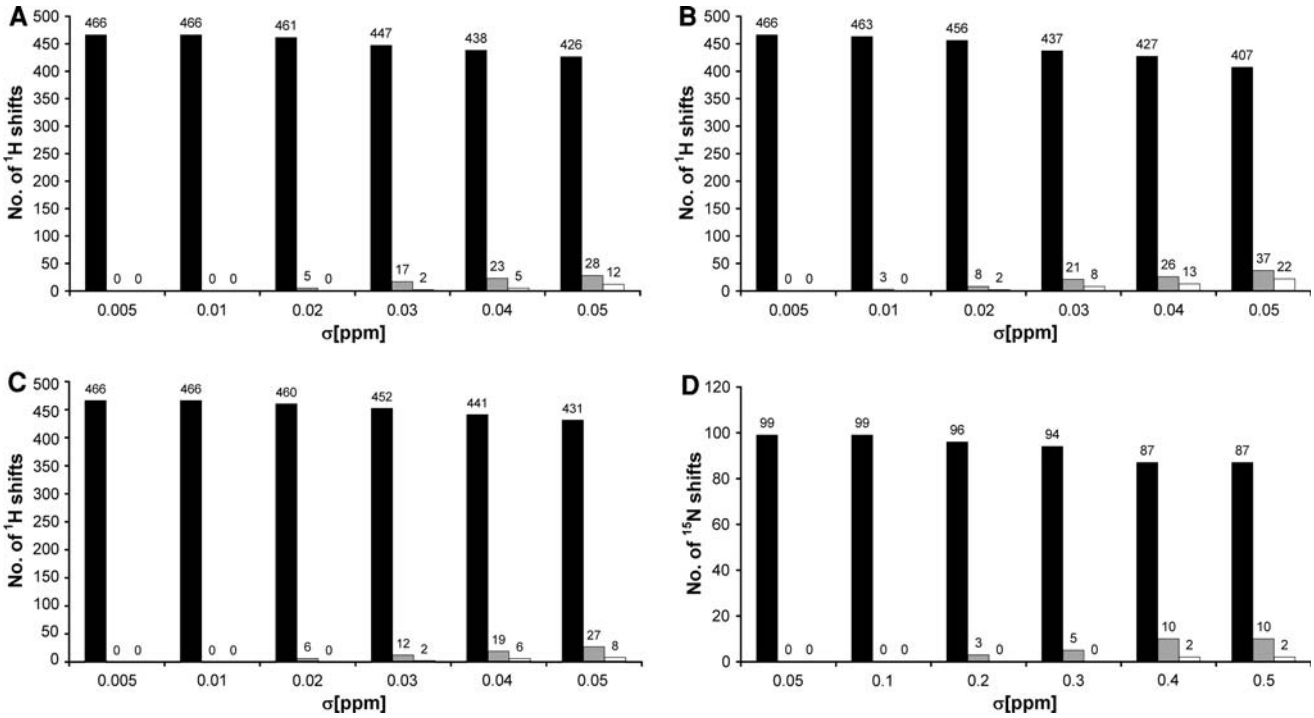

Fig. 4 Reliability of SHIFTOPT1 and SHIFTOPT2 for NOESY-type spectra in the absence of noise. The number of completely correctly predicted chemical shifts ($|\delta_i^{opt} - \delta_i^e| \leq R_i$) (black bars), of improved or unchanged chemical shifts ($|\delta_i^{opt} - \delta_i^e| \leq (|\delta_m^s - \delta_i^e|)$ (grey bars), and inadequately optimized chemical shifts ($|\delta_i^{opt} - \delta_i^e| > (|\delta_m^s - \delta_i^e|)$ (white bars) are plotted as a function of σ in the dimension k under consideration. δ_i^{opt} is the chemical shift after optimization. Note that for nuclei X others than ^{1}H the standard deviation was modified by

calculating a peak list directly from the chemical shifts both methods are very stable for all search ranges tested.

Performance in the absence of noise

In a next test the chemical shift table was perturbed by adding or subtracting a value $\Delta\delta_i$ to the chemical shifts δ_i. The values $\Delta\delta_i$ were randomly selected from a Gaussian distribution with a standard deviation σ and a mean of δ_i. For nuclei X others than ^{1}H the standard deviation was modified by multiplying σ(H) with γ_H/γ_X. Only values less than 3σ were taken. In this way, a chemical shift table $\mathbf{S}_0$ was produced that does not fit optimally to the experimentally observed chemical shifts δ_i^e. For small values of σ and thus for small chemical shift variations between the simulated shifts δ_m^s and the experimental shifts δ_i^e excellent results are obtained with SHIFTOPT1 for the 2D-NOESY (Fig. 4a) spectrum as well as for the 3D ^{15}N edited NOESY spectrum (Fig. 4c and d). Up to a σ of 0.01 ppm all chemical shifts values are correct after the application of SHIFTOPT1 in the 2D-NOESY spectrum, up to σ of 0.02 ppm all values are either improved or not modified. Only after σ has increased to 0.03 ppm, a few chemical shift values (2 out of 466) are worse than before. A similar picture is obtained for the 3D-NOESY spectrum. Up to a σ of 0.01 and 0.1 ppm for ^{1}H and ^{15}N all chemical shifts values are correct, up to a σ of 0.02 and 0.3 ppm all ^{1}H and ^{15}N chemical shifts values are either improved or unchanged. Only after σ has increased to 0.03 and 0.4 ppm, a few chemical shift values are slightly worse than before.

In its last step SHIFTOPT1 relies strongly on the calculation of cross peak volumes, whereas SHIFTOPT2 does not use this information. If this information is available and the cross peak volumes vary much as it is typical for NOESY-type spectra the performance of SHIFTOPT2 should be inferior to that of SHIFTOPT1. Indeed, at values of the standard deviation where SHIFTOPT1 works perfectly, SHIFTOPT2 already makes some errors (Fig. 4b).

A different type of data sets are represented by 2D ^{1}H, ^{15}N-HSQC spectra or 3D HNCA spectra that contain only a small number of cross peaks with a relatively small dynamic range of cross peak intensities. Here, the fast, direct algorithm of SHIFTOPT2 should perform well. Figure 5c and d show that this is indeed true. Up to a σ of 0.01 ppm all chemical shifts values are correct after the application of SHIFTOPT2 in the 2D-HSQC spectrum, up

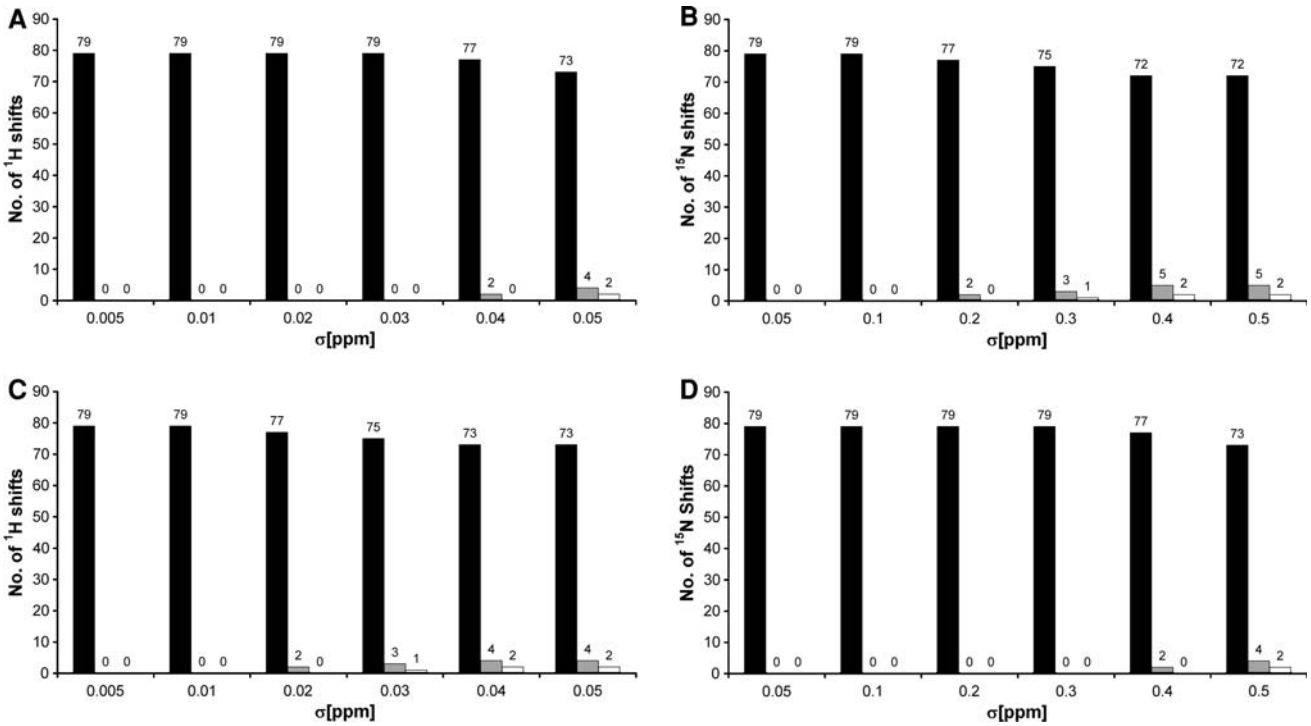

Fig. 5 Reliability of SHIFTOPT1 and SHIFTOPT2 for HSQC-type spectra in the absence of noise. The number of completely correctly predicted chemical shifts ($|\delta_i^{opt} - \delta_i^e| \leq R_i$) (black bars), of improved or unchanged chemical shifts ($|\delta_i^{opt} - \delta_i^e| \leq (|\delta_m^s - \delta_i^e|)$ (grey bars), and inadequately optimized chemical shifts ($|\delta_i^{opt} - \delta_i^e| > (|\delta_m^s - \delta_i^e|)$ (white bars) are plotted as a function σ in the dimension k under consideration. δ_i^{opt} is the chemical shift after optimization. Note that or nuclei X others than ^{1}H the standard deviation was modified by multiplying σ(H) with γ_H/γ_X. **a**, **b** Application of SHIFTOPT1 to a 2D ^{1}H, ^{15}N HSQC spectrum with a ^{1}H digital resolution of 0.0068 ppm and a ^{15}N digital resolution of 0.19 ppm, where all noise peaks were removed after peak picking. **c**, **d** As (**a**, **b**) but using SHIFTOPT2 on a 2D ^{1}H,^{15}N-HSQC-spectrum

to a σ of 0.02 ppm all ^{1}H chemical shift values are either improved or not modified. Only after σ has increased to 0.05 ppm, a few chemical shift values are worse than before. In the ^{15}N domain standard deviations of up to 0.4 ppm (that is up to a maximum error of $3\sigma = 1.2$ ppm) are accepted without error. A similar picture is obtained for the 3D-HNCA spectrum (Fig. 6). Up to a σ of 0.01, 0.1, and 0.1 ppm all ^{1}H, ^{15}N and ^{13}C chemical shifts values are correct after the application of SHIFTOPT2 in the 3D-HNCA spectrum, up to a σ of 0.03, 0.3, and 0.3 ppm all ^{1}H, ^{15}N and ^{13}C all values are either improved or not modified. Only after σ has increased to 0.04, 0.4, and 0.4 ppm for the ^{1}H, ^{15}N, and ^{13}C, respectively, a few chemical shift values are slightly worse than before. SHIFTOPT1 (Fig. 5a, b) does not perform as well as SHIFTOPT2 in a HSQC spectrum (Fig. 5c, d) giving wrong results for ^{1}H and ^{15}N shift deviations of $\sigma > 0.02$ ppm and > 0.2 ppm, respectively.

Performance in the presence of noise

Since noise and artifact peaks can lead to false assignments, Gaussian noise was added to the simulated 2D NOESY spectrum before peak picking. Additional cross peaks at random positions were added in the case of the experimental 2D HSQC spectrum. Since the performance of SHIFTOPT1 was superior for NOESY-type spectra to SHIFTOPT2 but inferior for HSQC-type spectra, SHIFTOPT1 was only tested for NOESY-type spectra (Fig. 7) and SHIFTOPT2 for HSQC-type spectra (Fig. 8).

Two different cases of practical importance were studied, a relatively small maximum chemical shift variation of 0.01 ppm (8 Hz at 800 MHz) and 0.02 ppm (18 Hz at 800 MHz). The number of noise peaks wrongly recognized as signals were increased by increasing the level of the standard deviation of the Gaussian noise continuously. The spectrum contained 9,035 true cross peaks. When 653 additional peaks where wrongly recognized all chemical shifts where improved or at least not changed. When 4,143 noise peaks were recognized 407 of the 466 chemical shifts where corrected perfectly, 54 improved or unchanged, and only 5 (1.1%) corrected in the wrong way. At a σ of 0.02 ppm still most of the peaks were improved (Fig. 7b).

The application of SHIFTOPT2 to an experimental ^{1}H, ^{15}N HSQC gives similar results: Up to 22 additional cross peaks are tolerated when 79 true signals are present and the chemical shift list contains errors up to 0.03 (^{1}H) and 0.3 ppm (^{15}N). When 48 additional cross peaks are added,

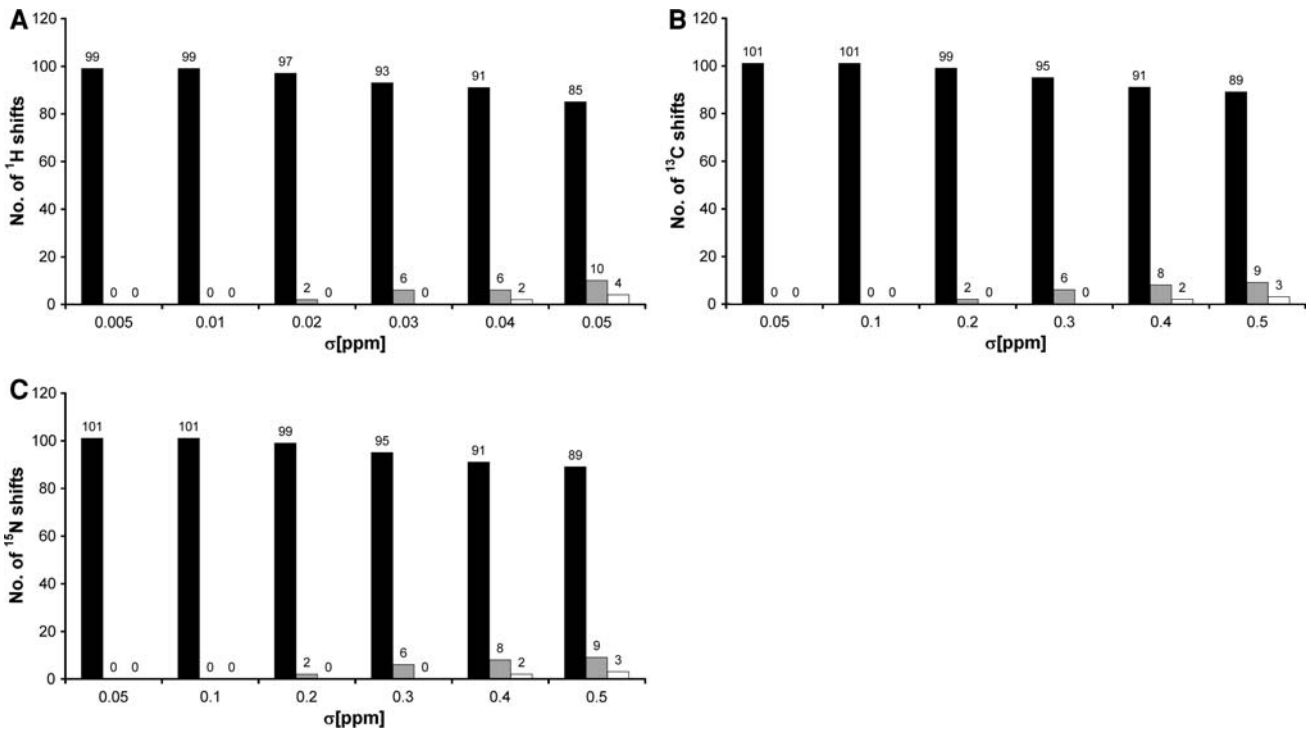

Fig. 6 Reliability of SHIFTOPT2 for a 3D HNCA spectrum in the absence of noise. The number of completely correctly predicted chemical shifts ($|\delta_i^{opt} - \delta_i^{e}| \leq R_i$) (black bars), of improved or unchanged chemical shifts ($|\delta_i^{opt} - \delta_i^{e}| \leq (|\delta_m^{s} - \delta_i^{e}|)$ (grey bars), and inadequately optimized chemical shifts ($|\delta_i^{opt} - \delta_i^{e}| > (|\delta_m^{s} - \delta_i^{e}|)$ (white bars) are plotted as a function σ in the dimension k under consideration. δ_i^{opt} is the chemical shift after optimization. Note that for nuclei X others than ^{1}H the standard deviation was modified by multiplying σ(H) with γ_H/γ_X. 3D HNCA with a ^{1}H digital resolution of 0.0068 ppm (**a**), ^{13}C digital resolution of 0.1617 ppm (**b**), and **c** a ^{15}N digital resolution of 0.321 ppm

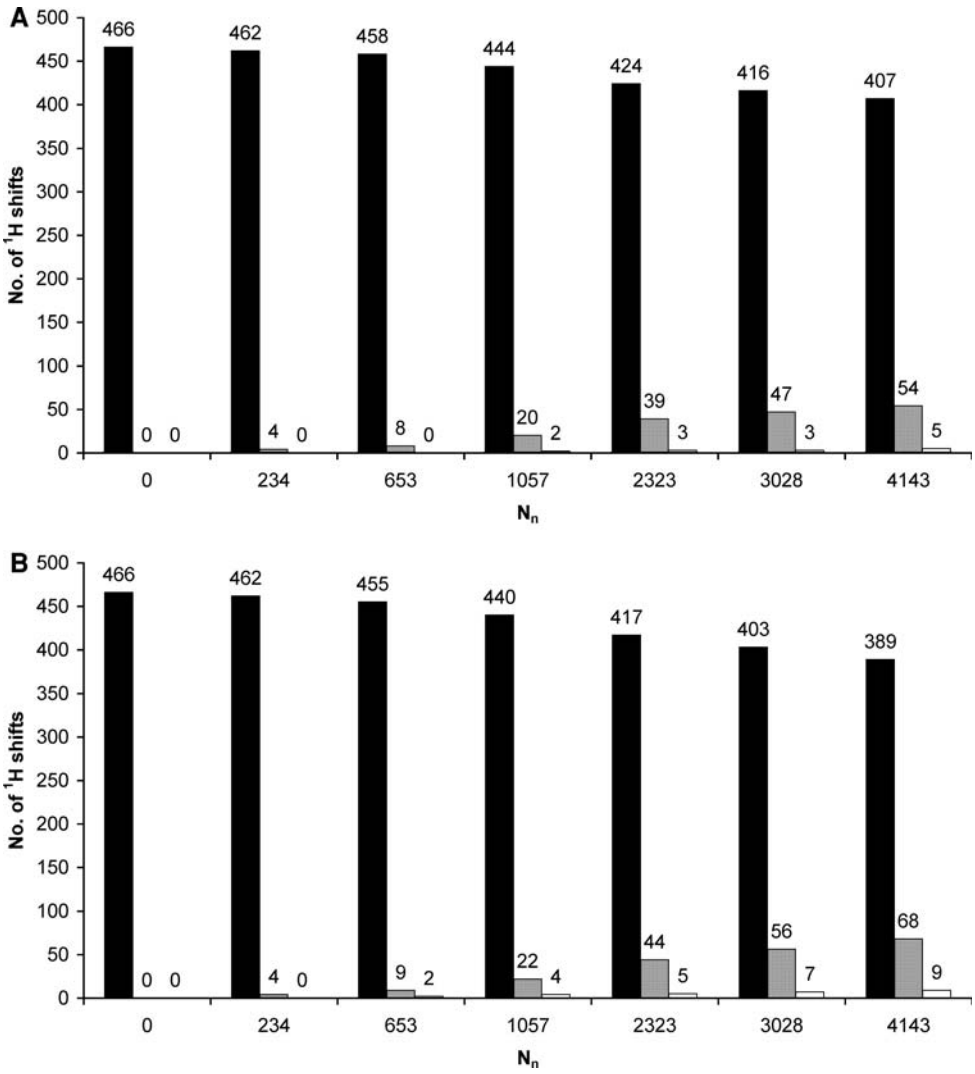

Fig. 7 Reliability of the shift optimization procedure SHIFTOPT1 as a function of noise level. The number of completely correctly predicted chemical shifts ($|\delta_i^{opt} - \delta_i^e| \leq R_i$) (black bars), of improved or unchanged chemical shifts ($|\delta_i^{opt} - \delta_i^e| \leq (|\delta_m^s - \delta_i^e|)$ (grey bars), and inadequately optimized chemical shifts ($|\delta_i^{opt} - \delta_i^e| > (|\delta_m^s - \delta_i^e|)$ (white bars) are plotted as a function of the number of noise peaks in the dimension k under consideration. The noise level was increased gradually, so that at the peak picking threshold N additional noise peaks were identified. δ_i^{opt} is the chemical shift after optimization. **a** Simulated 800 MHz 2D NOESY-spectrum with a ^{1}H digital resolution of 0.0062 ppm, application of SHIFTOPT1. The spectrum contains 9,035 valid protein cross peaks. Variations of chemical shifts with a standard deviation $\sigma = 0.01$ ppm. **b** Same as (**a**) but with a σ of 0.02 ppm

the majority of the chemical shifts is still improved, only 3 chemical shifts are modified in the wrong direction (Fig. 8a, b). At the higher maximum chemical shift deviation of 0.06 and 0.6 ppm respectively (Fig. 8c, d) a similar picture is obtained, the number of wrongly corrected chemical shifts is still unchanged.

Automated chemical shift assignment in a set of pressure dependent HSQC spectra

Figure 9a shows a part of a ^{1}H, ^{15}N HSQC spectrum of a HPr from *S. carnosus* measured at different pressures. The initial ^{1}H and ^{15}N chemical shift values at ambient pressure were taken from Görler et al. (1999a, b). In a first step SHIFTOPT2 was applied to a data set recorded at 3 MPa (Kalbitzer et al. 2000) and could assign all shift values correctly (Fig. 9b). This optimized chemical shift table was then applied to a second data set recorded at 50 MPa where again all chemical shifts were correctly found. In a next step, the chemical shifts expected at 100 MPa were predicted by a linear extrapolation and than optimized by applying SHIFOPT2. The chemical shifts were used for a second order prediction and the procedure was repeated as before. The chemical shifts of the 79 amide groups could be correctly obtained for all pressures. As a comparison, all chemical shifts were obtained by applying SHIFTOPT2 without prior prediction of the chemical shift development. Here, errors occurred for higher pressures, although most of the chemical shifts obtained were still correct (Fig. 9).

Discussion

Even when using the same sample chemical shifts of cross peaks vary slightly from spectrum to spectrum due to small variations of temperature (caused e.g., by broadband decoupling) and differences in digital resolution. In practice, chemical shift lists as published in the

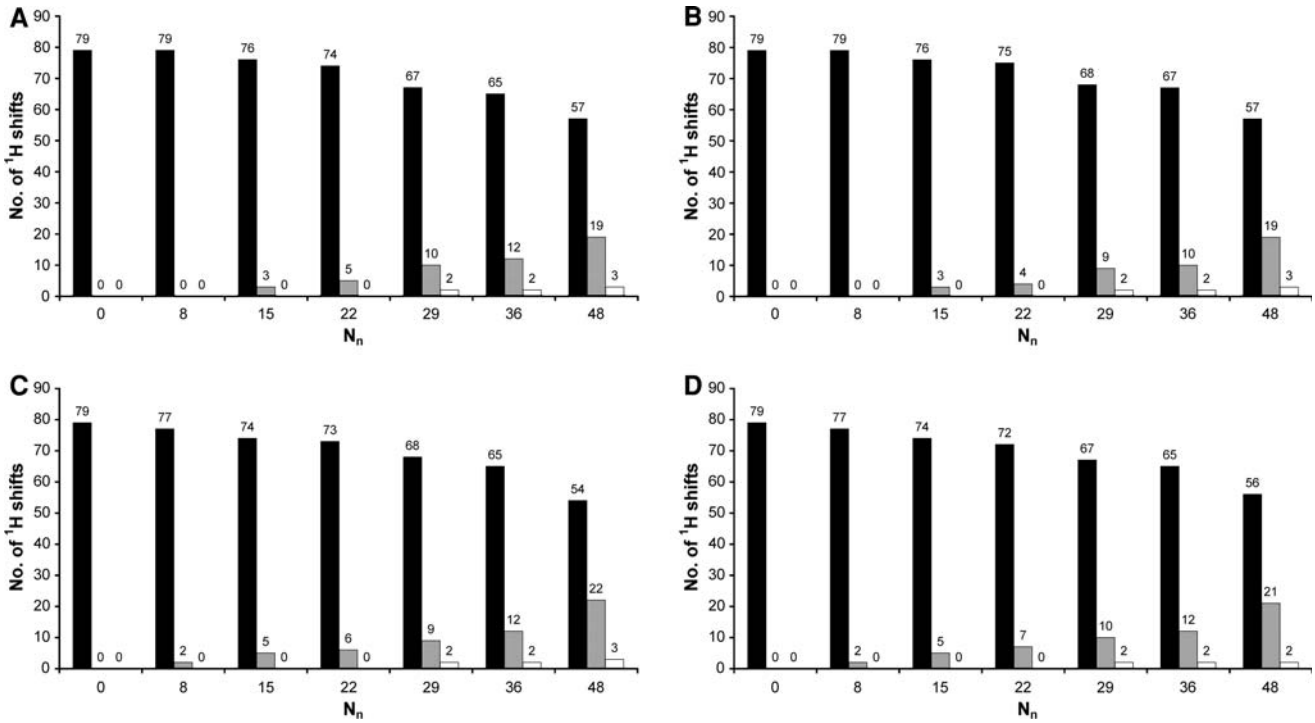

Fig. 8 Reliability of the shift optimization procedure SHIFTOPT2 as a function of noise level. The number of completely correctly predicted chemical shifts ($|\delta_i^{opt} - \delta_i^e| \leq R_i$) (black bars), of improved or unchanged chemical shifts ($|\delta_i^{opt} - \delta_i^e| \leq (|\delta_m^s - \delta_i^e|$) (grey bars), and inadequately optimized chemical shifts ($|\delta_i^{opt} - \delta_i^e| > (|\delta_m^s - \delta_i^e|$) (white bars) are plotted as a function of the number of N. The spectrum contains 79 valid protein cross peaks. δ_i^{opt} is the chemical shift after optimization. Note that for nuclei X others than ^{1}H the standard deviation was modified by multiplying σ(H) with γ_H/γ_X (**a**, **b**) Experimental 2D ^{1}H, ^{15}N HSQC spectrum, ^{1}H digital resolution of 0.0068 ppm and ^{15}N digital resolution of 0.019 ppm. Variations of chemical shifts with $\sigma = 0.01$ ppm in the direct dimension and 0.1 ppm in the indirect dimension. **c**, **d** Same as (**a**, **b**) but with σ-values of 0.02 ppm and of 0.2 ppm

BMRB data base are composed from different data sets measured under various experimental conditions (e.g., data recorded in D_2O and H_2O), and thus do usually not fit exactly to a given experimental spectrum. Here, a chemical shift optimization as implemented in SHIFTOPT1 and SHIFTOPT2 can help.

Limits of accuracy

Although a completely correct result of the chemical shift optimization is the ultimate goal, for most applications, an improvement of the chemical shift lists is still a satisfactory result. There are several factors that lead with high probability to insufficient results: (1) Incomplete spectra, where some chemical shifts are not represented at all do not contain the required information and thus cannot be used for calculating optimized shifts. When working on peak lists as SHIFTOPT1 and SHIFTOPT2 do, artefact peaks with sufficient intensity may be wrongly interpreted as true signals and may be used for the chemical shift calculation. This is especially dangerous in low redundancy spectra such as HSQC-spectra where they may be wrongly assigned because of the used distance metric.

Here, a limited search range is an important mean to avoid a misinterpretation, since an artefact inside in the search range cannot be recognized as such. However, the application of a powerful artefact recognition routine prior to the application of SHIFTOPT reduces the likelihood of such a wrong interpretation and the starting value is not modified. The Bayesian routine used in AUREMOL assigns signal probabilities to all peaks. It has been proved powerful to calculate the number of cross peaks K expected in the search areas, and to retain only the a*K peaks ($a = 1.1$) with the highest probabilities in the peak list. In our experience about 10% more peaks should be retained, that is a is set to 1.1. (2) Chemical shift degeneracy as it often occurs in 2D-HSQC spectra with the amide proton and nitrogen resonances identical cannot be handled by the actual routines, since there is no information available, to decide if the cross peaks are superposed or one cross peak is missing or shifted significantly. However, in NOESY-type spectra usually the redundancy is very high and normally at least one cross peak with different chemical shift combinations is available and can be used. (3) Very strong shifts in crowded spectra may lead to wrong decisions since the used metric favours small shift changes.

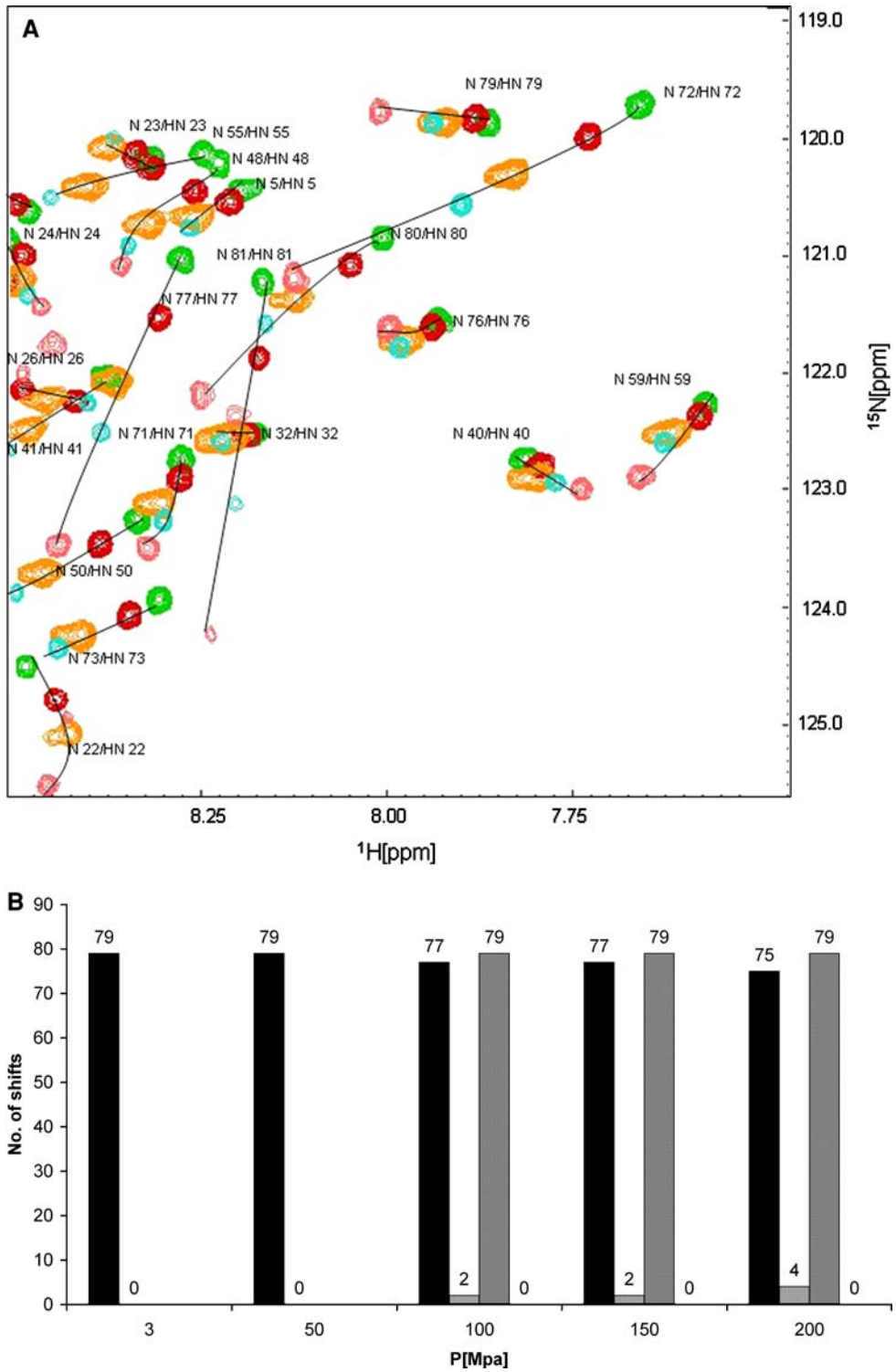

Fig. 9 Automated chemical shift recognition in a set of pressure dependent HSQC-spectra. **a** A set of ^{1}H, ^{15}N NMR spectra of ^{15}N enriched HPr from *S. carnosus* was recorded at 298 K and various pressures. (green) 3 MPa, (red) 50 MPa, (yellow) 100 MPa, (blue) 150 MPa, (pink) 200 MPa. Only part of the spectrum is shown. Solid lines connect residues automatically assigned using a polynomial of the order of 2. **b** The number of completely correctly optimized chemical shifts ($|\delta_i^{opt} - \delta_i^{e}| \leq R_i$) (black bars), of improved or unchanged chemical shifts ($|\delta_i^{opt} - \delta_i^{e}| \leq (|\delta_m^{s} - \delta_i^{e}|$) and (grey bars) are plotted as a function of the pressure. Using the predicted shifts from the chemical shift polynomial as input shift, the assignment getting better. The number of completely correctly optimized chemical shifts ($|\delta_i^{opt} - \delta_i^{e}| \leq R_i$) (black doted bars), improved or unchanged chemical shifts ($|\delta_i^{opt} - \delta_i^{e}| \leq (|\delta_m^{s} - \delta_i^{e}|$) (grey doted bars) using the polynomial is plotted in the figure. The spectra contain 79 valid protein cross peaks

Performance of the routines

In the absence of significant noise peaks in NOESY type spectra SHIFTOPT1 works perfectly up to a sigma of 0.01 ppm (Fig. 4a) that corresponds to maximum chemical shift changes of ±0.03 ppm. For higher chemical shift changes a few shifts are only partly corrected. At a maximum shift change of 0.09 ppm 2 (out of 462) shift values are not improved anymore, but are wrongly corrected by a shift change in the wrong direction. This is usually a pair wise ambiguity where the used metric favours the wrong assignment for two resonances with very close chemical shifts. In the 3D-NOESY-HSQC spectrum comparable results are obtained, up to maximum shifts of the order of the line width very good results are obtained. When larger shift variations are allowed, the used distance metric leads in a few cases to wrong decisions. In practice, it means that the subsequent steps still have to allow an appropriate error of the chemical shifts.

When more than one spectrum exists with continuous shift changes, the proposed prediction procedure leads to accurate results.

Acknowledgements This work was supported by the European Union (SPINE-2), the Fonds der Chemischen Industrie, the Deutsche Forschungsgemeinschaft (DFG), and the Bundesministerium für Bildung und Forschung (BMBF).

References

Antz C, Neidig K-P, Kalbitzer HR (1995) A general Bayesian method for an automated signal class recognition in 2D NMR spectra combined with a multivariate discriminant analysis. J Biomol NMR 5:287–296

Box GEP, Muller ME (1958) A note on the generation of random normal deviates. The Annals of Mathematical Statistics 29(2):610–611

Catasti P, Carrara E, Nicolini C (1990) Pepto: an expert system for automatic peak assignment of two-dimensional nuclear magnetic resonance spectra of proteins. J Comput Chem 11(7):805–818

Geyer M, Neidig K-P, Kalbitzer HR (1995) Automated peak integration in multidimensional NMR spectra by an optimized iterative segmentation procedure. J Magn Reson B 109:31–38

Glaser S, Kalbitzer HR (1987) Automated recognition and assessment of cross peaks in two-dimensional NMR spectra of macromolecules. J Magn Reson 74:450–463

Görler A, Kalbitzer HR (1997) RELAX: a flexible program for the analysis of NOESY-Spectra by back calculation based on the complete relaxation matrix formalism. J Magn Reson 124: 177–188

Görler A, Hengstenberg W, Kravanja M, Beneicke W, Maurer T, Kalbitzer HR (1999a) Solution structure of histidine containing phosphocarrier protein (HPr) from *Staphylococcus carnosus*. Appl Magn Reson 17:465–480

Görler A, Gronwald W, Neidig K-P, Kalbitzer HR (1999b) Computer assisted assignment of ^{13}C or ^{15}N edited 3D-NOESY-HSQC spectra using back calculated and experimental spectra. J Magn Reson 137:39–45

Gronwald W, Kalbitzer HR (2004) Automated structure determination of proteins by NMR spectroscopy. Prog NMR Spectrosc 44:33–96

Gronwald W, Kirchhöfer R, Görler A, Kremer W, Ganslmeier B, Neidig K-P, Kalbitzer HR (2000) RFAC, a programme for automated NMR-R-factor estimation. J Biomol NMR 17:137–151

Gronwald W, Moussa S, Elsner R, Jung A, Ganslmeier B, Trenner J, Kremer W, Neidig K-P, Kalbitzer HR (2002) Automated assignment of NOESY NMR spectra using a knowledge based method (KNOWNOE). J Biomol NMR 23:271–287

Gronwald W, Bombke J, Maurer T, Domogalla B, Huber F, Schumann F, Kremer W, Fink F, Rysiok T, Frech M, Kalbitzer HR (2008) Structure of the Leech protein Saratin and characterisation of its binding to collagen. J Mol Biol 381:913–927

Hare BJ, Prestegard JH (1994) Application of neural networks to automated assignment of NMR spectra of proteins. J Biomol NMR 4:35–46

Herrmann T, Güntert P, Wthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 319:209–227

Kalbitzer HR, Görler A, Li H, Dubovskii P, Hengstenberg W, Kowolik C, Yamada H, Akasaka K (2000) ^{15}N and ^{1}H NMR study of histidine containing protein (HPr) from *Staphylococcus carnosus* at high pressure. Prot Sci 9:693–703

Maurer T, Meier S, Kachel K, Munte CE, Hasenbein S, Koch B, Hengstenberg W, Kalbitzer HR (2004) High resolution structure of the histidine containing phosphocarrier protein (HPr) from *Staphylococcus aureus* and characterisation of its interaction with the bifunctional HPrKinase/phosphorylase. J Bacteriol 186:5906–5918

Neidig P, Bodenmüller H, Kalbitzer HR (1984) Computer aided evaluation of two-dimensional NMR spectra of proteins. Biochem Biophys Res Commun 125:1143–1150

Press W, Teukolsky S, Vetterling W, Flannery B (1992). §3.1 Polynomial interpolation and extrapolation. In: numerical recipes in C. The art of scientific computing, 3rd edn. Cambridge University Press, Cambridge

Ried A, Gronwald W, Trenner JM, Brunner K, Neidig K-P, Kalbitzer HR (2004) Improved simulation of NOESY spectra by RELAX-JT2 including effects of J-coupling, transverse relaxation, and chemical shift anisotropy. J Biomol NMR 30:121–131

Schulte AC, Görler A, Antz C, Neidig K-P, Kalbitzer HR (1997) Use of global symmetries in automated signal class recognition by a Bayesian method. J Magn Reson 129:165–172

Schumann FH, Riepl H, Maurer T, Gronwald W, Neidig K.-P., Kalbitzer HR (2007) Combined chemical shift changes and amino acid specific chemical shift mapping of protein–protein interactions. J Biomol NMR 39:275–289

Xu Y, Jablonsky MJ, Jackson PL, Braun W, Rama Krishna N (2001) Automatic assignment of NOESY cross peaks and determination of the protein structure of a new world scorpion using NOAH/DIAMOD. J Magn Reson 148:35–46

Zimmerman DE, Kulikowski CA, Huang Y, Feng W, Tashiro M, Shimotakahara S, Chien C-y, Powers R, Montelione GT (1997) Automated analysis of protein NMR assignments using methods from artificial intelligence. J Mol Biol 269:592–610