Journal of Magnetic Resonance 210 (2011) 177–183

Contents lists available at ScienceDirect

Journal of Magnetic Resonance

journal homepage: www.elsevier.com/locate/jmr

Singular spectrum analysis for an automated solvent artifact removal and baseline correction of 1D NMR spectra

Silvia De Sanctis [a], Wilhelm M. Malloni [a], Werner Kremer [a], Ana M. Tomé [b], Elmar W. Lang [a], Klaus-Peter. Neidig [c], Hans Robert Kalbitzer [a,*]

[a] Institute of Biophysics and Physical Biochemistry, University of Regensburg, D-93040 Regensburg, Germany
[b] Department of Electrical Engineering, Telecommunications and Informatics, University of Aveiro, P-3100 Aveiro, Portugal
[c] BioSpin GmbH, Software Department, Silberstreifen 4, D-76287 Rheinstetten, Germany

ARTICLE INFO

Article history:
Received 18 January 2011
Revised 27 February 2011
Available online 6 March 2011

Keywords:
AUREMOL-SSA/ALS
Singular spectrum analysis
Solvent suppression
Baseline correction
Oversampled digitally filtered data
Group delay
Finite response filter

ABSTRACT

NMR spectroscopy in biology and medicine is generally performed in aqueous solutions, thus in ^{1}H NMR spectroscopy, the dominant signal often stems from the partly suppressed solvent and can be many orders of magnitude larger than the resonances of interest. Strong solvent signals lead to a disappearance of weak resonances of interest close to the solvent artifact and to base plane variations all over the spectrum. The AUREMOL-SSA/ALS approach for automated solvent artifact removal and baseline correction has been originally developed for multi-dimensional NMR spectroscopy. Here, we describe the necessary adaptations for an automated application to one-dimensional NMR spectra. Its core algorithm is still based on singular spectrum analysis (SSA) applied on time domain signals (FIDs) and it is still combined with an automated baseline correction (ALS) in the frequency domain. However, both steps (SSA and ALS) have been modified in order to achieve optimal results when dealing with one-dimensional spectra. The performance of the method has been tested on one-dimensional synthetic and experimental spectra including the back-calculated spectrum of HPr protein and an experimental spectrum of a human urine sample. The latter has been recorded with the typically used NOESY-type 1D pulse sequence including water pre-saturation. Furthermore, the fully automated AUREMOL-SSA/ALS procedure includes the managing of oversampled, digitally filtered and zero-filled data and the correction of the frequency domain phase shift caused by the group delay time shift from the digital finite response filtering.

1. Introduction

Since its discovery, nuclear magnetic resonance spectroscopy has become an essential tool for numerous applications in biomedicine, biophysics and biochemistry. Multi-dimensional NMR spectroscopy of isotope-enriched proteins and nucleic acids clearly represents a powerful technique to solve complicated structural problems. However, one-dimensional spectroscopy of unlabelled samples has still important applications. Examples are functional studies of bio-macromolecules that are not available in isotope-enriched form, including interaction studies with small ligands in drug design or folding/unfolding studies of proteins. Another example may be the high-throughput NMR analysis of biofluids such as urine [1], blood plasma [2], bile, cerebrospinal fluids [3] and cervical mucus [4].

The quantitative one-dimensional ^{1}H NMR spectroscopic analysis of body fluids as well as macromolecules such as proteins is typically hampered by the presence of a dominant resonance arising from the solvent signal. Suppressing this solvent artifact, either by experimental methods (water suppression and/or tailored excitation) or by computational post-processing methods, is thus a key issue in proton NMR (for a review see [5]). Ideally, a suppression procedure should not (1) alter the intensities of solute resonances, (2) introduce baseline artifacts, or (3) obscure significant signals hidden underneath the intense solvent signal. With modern NMR spectrometers but not with the majority of older NMR spectrometers still in use, the two first points can be achieved experimentally rather well when the concentration of the solute is not too small, the third point represents a problem especially when using tailored excitation methods where a significant region around the water resonance frequency is blanked out.

Many post-processing suppression methods have been proposed in literature to solve the latter problem (listed e.g. by

Abbreviations: SSA, singular spectrum analysis; ALS, automated linear spline; FIR, finite impulse response filter.

* Corresponding author. Fax: +49 941 943 2479.
E-mail address: Hhans-robert.kalbitzer@biologie.uni-regensburg.de (H.R. Kalbitzer).

1090-7807/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2011.03.001

[5,6]). However, the existing filtering techniques developed so far adequately cannot recover the shapes and intensities of the solute resonances under the solvent peak and/or cannot be used automatically without human intervention. Malloni et al. [6] used singular spectrum analysis (SSA) combined with automated linear spline (ALS) baseline correction [6] for an automated removal of the solvent artifact and of baseline distortions in multi-dimensional NMR spectra. AUREMOL-SSA/ALS automatically identifies the solvent artifact signal and removes it while keeping those hidden resonances of interest that appear under the water signal. Moreover, an automated baseline correction is performed by a linear spline interpolation [7]. The crucial point here is an automated recognition of signals and baseline points independently of the type of the spectrum and the kind of sample. The published implementation of AUREMOL-SSA/ ALS was only developed for higher-dimensional NMR data [6] but did not give satisfactory results for one-dimensional NMR spectra. In particular, it is necessary to take into account that one-dimensional spectra of biomolecules are usually very crowded and may contain resonances with widely differing line widths. In addition, the digital resolution of one-dimensional NMR spectra is typically much higher than in multi-dimensional spectra and thus leads to the necessity of extracting much more components from the time domain signal in order to achieve an accurate separation of the solvent from the rest of the spectrum.

In this paper we will introduce the necessary adaptations on AUREMOL-SSA/ALS for the use in one-dimensional spectroscopy and show the applicability of the method and its limits in two typical situations: NMR spectroscopy of proteins and biofluids.

2. Materials and methods

2.1. Back-calculated dataset

A synthetic one-dimensional spectrum has been calculated with the AUREMOL routine RELAX-*JT*2 [8] starting from the three-dimensional structure of a mutant of histidine-containing phosphocarrier protein (HPr(H15A)) [9] from *Staphylococcus aureus* and using the corresponding experimental chemical shifts. RELAX-JT2 simulates multiplet structures as well as line widths. A one-dimensional 600 MHz NOESY spectrum with a mixing time of 0.15 s, a relaxation delay of 1.5 s, a spectral width of 12.65 ppm, and 2048 complex time domain data points was simulated. The resulting time domain data was filtered by exponential multiplication with a line broadening of 3 Hz. The water artifact was produced by measuring a one-dimensional spectrum of 90% H_2O/10% D_2O with solvent pre-saturation at 600 MHz, having the same acquisition parameters as those used for the HPr spectrum simulation. Before the Fourier transformation the water artifact signal was added to the synthetic time domain signal of the protein scaled in such a way that the maximum of the water signal was about 500 times stronger than a typical protein resonance.

2.2. Experimental data

The urine spectrum was recorded with a Bruker Avance-600 spectrometer operating at a ^{1}H frequency of 600.13 MHz. It was acquired using oversampling and digital filtering (Bruker DQD mode). A NOESY-type 1D pulse sequence was used for the sample, including a selective pre-saturation of the solvent resonance and a spoiling *z*-gradient pulse applied during the mixing time. For obtaining standardized conditions 133 mM sodium phosphate buffer, pH 7.4, 5% D_2O, and 0.1 mM DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) was added. The spectrum was recorded at 298 K using a mixing time of 10 ms and a relaxation delay of 5 s. The spectral width was 20.0 ppm and 128 K complex time domain points were sampled.

2.3. Software

The NMR data were acquired with the program TOPSPIN 2.0 (Bruker, Karlsruhe). All routines developed have been integrated in the AUREMOL software package and can be downloaded from www.auremol.de.

3. Theory and implementation

3.1. Singular spectrum analysis

Only aspects essential for an application of SSA [10] to one-dimensional spectra will be discussed here, since the general method has been described earlier for multi-dimensional spectra by Malloni et al. [6]. The general procedure is summarized in a flow chart (Fig. 1). The method starts in the time domain and if the data are recorded by oversampling and finite response (FIR) filtering they have to be preprocessed adequately. After subtracting the mean of the signal, SSA is applied. The obtained time domain data are multiplied with an appropriate filter function and Fourier transformed. A phase correction has to be applied in accordance to the group delay. The last step is the baseline correction performed by automated linear spline (ALS): the baseline points are identified automatically. The interpolated baseline regions are subtracted from the frequency domain data.

3.2. Preparation of data for SSA

Meaningful application of SSA for separating and removing the strongest signal (the solvent signal) requires data with periodic structure as it is represented by a complex FID signal. Modern spectrometers use oversampling of the data followed by digital frequency filtering by finite impulse response filters and a reduction of the stored spectral range [11]. FIR filtering leads to a delayed response where the first time points are corrupted. For practical reasons, these data do not contain useful spectral information and can be considered as a delayed data acquisition characterized by the group delay $\Delta_g = A_g\Delta t$ with A_g a factor and Δt the dwell time. One of the manufacturers (Bruker Biospin) stores these data points as part of the FID. The resulting time domain signal thus is increasing from zero before the damped "normal" FID $s(n\Delta t)$ begins. For application of SSA (and other time domain applications), the FID of length N has to be left shifted by N_g data points with

$$N_g = \mathrm{floor}(2A_g) \tag{1}$$

since the inclusion of this data leads to spectral artifacts. Therefore, before applying SSA, the data points belonging to the group delay are removed and stored for a subsequent reconstruction. They are excluded while the SSA identifies and nullifies the solvent component but they are re-introduced before the Fourier transformation. The time shift due to the removal of these points causes a phase shift in the frequency data. A correspondent phase correction in the frequency domain is automatically applied before correcting the baseline.

3.3. Implementation of SSA

The reduced, centered FID of length $N_r = N - N_g$ is embedded into a feature space of dimension M that corresponds to the number of desired components extracted, whereby a fixed shift of $\Delta n_t = 1$ is used to generate the trajectory matrix $\mathbf{X}$ of dimension $M \times Q$ with $Q = N_r - M + 1$. Thus $\mathbf{X}$ contains as many time lagged

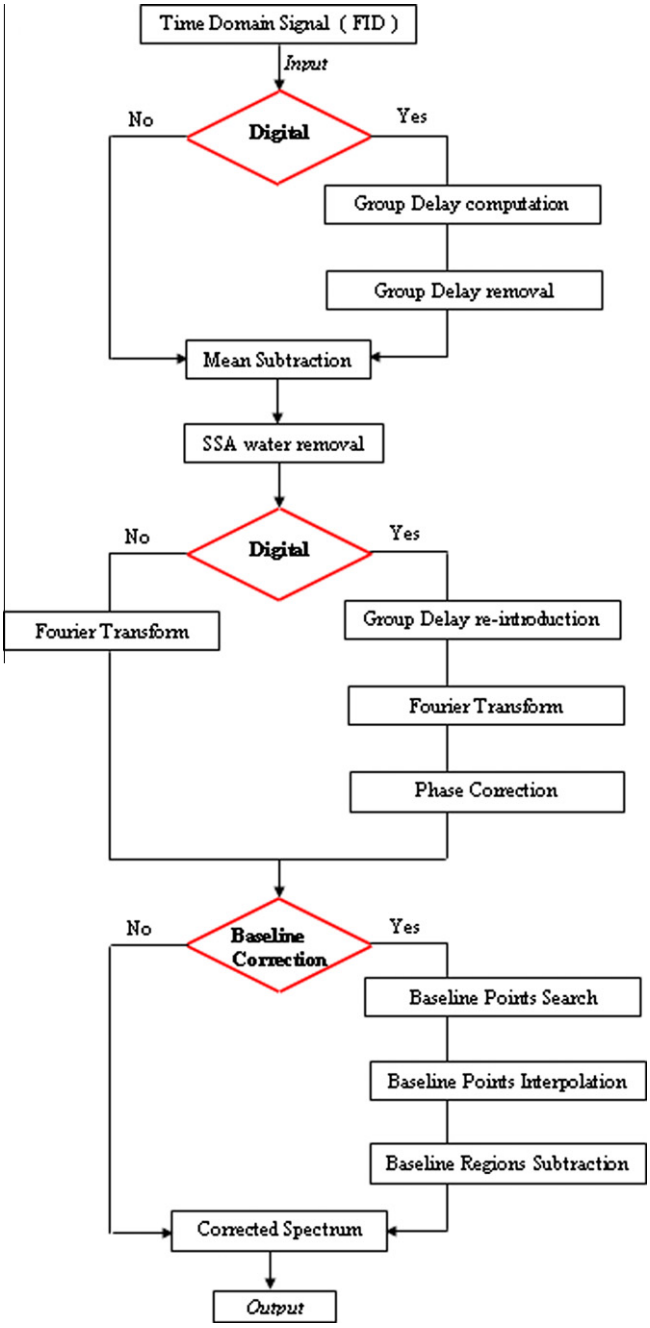

Fig. 1. Flow chart of AUREMOL-SSA/ALS.

copies of the complex FID $s(n\Delta t)$ as the number of components, shifted by one data point for every row of the data matrix. Due to the higher resolution of an usual one-dimensional spectrum, an embedding of $M = 20$ (used in default as described in [6]) is no more suitable thus it is increased to 40, whereas the rest of the algorithm is almost unchanged (see Section 4). Centering of the FID is done by calculating the mean of s as

$$\langle s \rangle = \frac{1}{N_r} \sum_{n=0}^{N_r-1} s(n\Delta t) \tag{2}$$

and subtracting it from s. Finally, the zero mean elements of **X** are computed as

$$x_{ij} = s(n\Delta t) \quad \text{with } n = M - i + j - 1 \tag{3}$$

The correlation matrix **C** is defined as

$$\mathbf{C} = \frac{1}{Q}\mathbf{X}\mathbf{X}^{\mathbf{t}} \tag{4}$$

with $\mathbf{X^t}$ the Hermitian transpose of **X**. The Hermitian matrix **C** can be decomposed by

$$\mathbf{C} = \mathbf{U}\mathbf{D}\mathbf{U}^t \tag{5}$$

with the unitary matrix **U** of eigenvectors $\mathbf{u_i}$ and the diagonal matrix **D**. Diagonalization of **C** is done by standard techniques according to [12,13]. **X** can be written as

$$\mathbf{X} = \mathbf{U}\mathbf{U}^{\mathbf{t}}\mathbf{X} = \sum_{i=1}^{M} \mathbf{u}_i \mathbf{u}_i^t \mathbf{X} = \sum_{i=1}^{M} \mathbf{X}_i \tag{6}$$

with the $(M \times 1)$ and $(1 \times M)$ matrices $\mathbf{u}_i$ and $\mathbf{u}_i^t$ containing the ith column of **U** and the ith row of $\mathbf{U}^t$, respectively. For solvent suppression the $\mathbf{u}_i$ and $\mathbf{u}_i^t$ corresponding to the largest eigenvalue of **C** (highest covariance) is set to zero resulting in a matrix $\mathbf{X}_{red}$ of reduced rank.

Generally, the elements along each descending diagonal of $\mathbf{X}_{red}$ will not be identical as they are in the original trajectory matrix **X**. This can be cured, however, by replacing the entries in each diagonal by their average, obtaining again a Toeplitz matrix $\mathbf{X}'_{red}$ [14,15]. As last step, an FID is reconstructed reverting the embedding from the matrix $\mathbf{X}'_{red}$ and the previously stored group delay points are appended to the reconstructed time domain signal. Since the group delay data points start invariably with a stretch of zeros, for avoiding truncation artifacts, the mean of the last quarter of the total data points of the FID obtained from $\mathbf{X}'_{red}$ has to be subtracted before appending the stored group delay data points. Since the group delay itself is usually not an integer multiple of the dwell time Δt, a first order phase correction φ can be calculated directly by $\varphi = 2\pi(2A_g - N_g)$ and has later to be applied after the final Fourier transformation of the data.

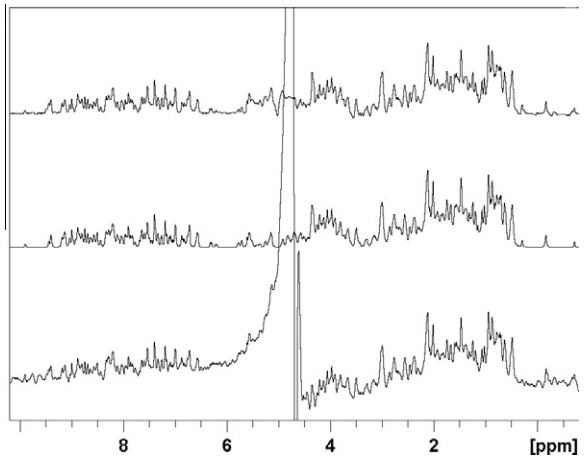

Fig. 2. AUREMOL-SSA/ALS applied to a 1D spectrum of a protein. The one-dimensional spectrum of HPr (H15A) from *Staphylococcus aureus* has been calculated with the AUREMOL routine RELAX-*JT*2. An experimental one-dimensional solvent FID measured in 90 % H_2O 10 % D_2O at 600 MHz with solvent presaturation was added to the protein time domain data simulated with the same parameters as used for the experimental data. The water artifact is approximately 500 times stronger than a typical protein resonance. Comparison of the complete simulated protein spectrum (middle trace) with the spectrum containing the water artifact (bottom signal) and the spectrum after application of AUREMOL-SSA/ALS (top trace).

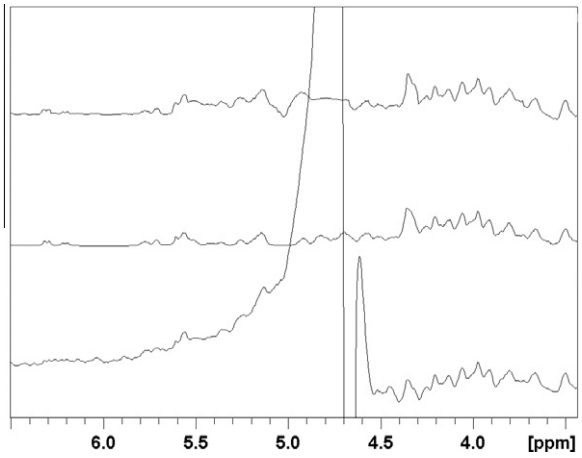

Fig. 3. Resonances close to the water line recovered by AUREMOL-SSA/ALS. Zoom of the spectrum shown in Fig. 2 with the protein spectrum including the water resonance (bottom signal), the original spectrum without any water signal added (middle trace) and after application of SSA/ALS (top spectrum).

3.4. Baseline correction

After removing the strong solvent signal if present, the baseline usually needs to be further corrected in the frequency domain. For higher-dimensional spectra the automated linear spline (ALS) [6,7] has shown to be efficient. The automated recognition of baseline points and signal points is crucial for the proper function of ALS. A method similar to that one proposed by Guentert and Wuethrich [16] but with a variable window size W is used in ALS. The size W of the window examined around a data point k must be clearly larger than the expected line width of a protein resonance peak for the recognition of baseline points. This parameter has to be automatically determined from an analysis of the line widths occurring in the spectrum and its correct choice is especially important in the complicated spectra of biological samples with large variations in line widths. In the multi-dimensional case [6] a histogram of all line widths occurring in the spectrum is built and W is fixed to twice the most frequently occurring line width. Tests on one-dimensional spectra of several biomaterials with a large variation of occurring line width values showed that this definition of W does not work *satisfactorily* since resonance lines with large line widths are not recognized correctly. It turned out that a window size W set to the double of the maximal occurring value yielded optimal results.

4. Results and discussion

4.1. Solvent suppression by SSA/ALS

SSA and ALS were applied to a synthetic one-dimensional spectrum of HPr protein and to an experimental one-dimensional

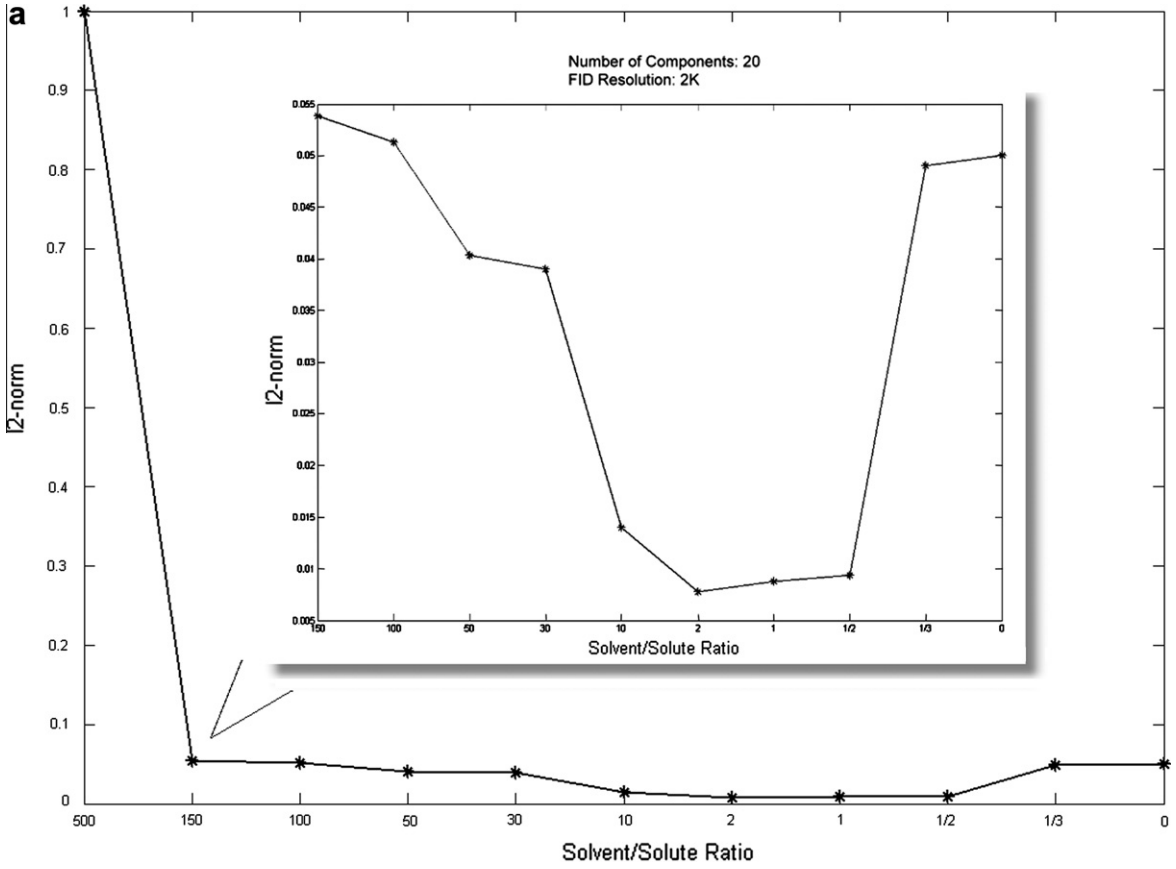

Fig. 4. Quantitative analysis of the performance of SSA. As in Figs. 2 and 3 a one-dimensional spectrum of HPr from *S. aureus* was simulated, the obtained FID was combined with an FID of an experimental water artifact signal with various relative intensities. As a measure for the performance of AUREMOL-SSA/ALS the l2-norm was calculated for the original simulated HPr spectrum and an HPr spectrum where a water signal was added and that was processed with AUREMOL-SSA/ALS. The l2 norm of a spectrum where the solvent signal was 500-times stronger than the most intense protein signal was arbitrarily set to 1. (a) Dependence of the l2 norm on the relative intensity of the solvent signal for a spectrum of HPr with 2 K complex data points. (b) Dependence of the performance of SSA on the number of components used. The l2 norm was calculated for a 2 K and 16 K FID with a ratio of solvent to protein signal intensity of 2.

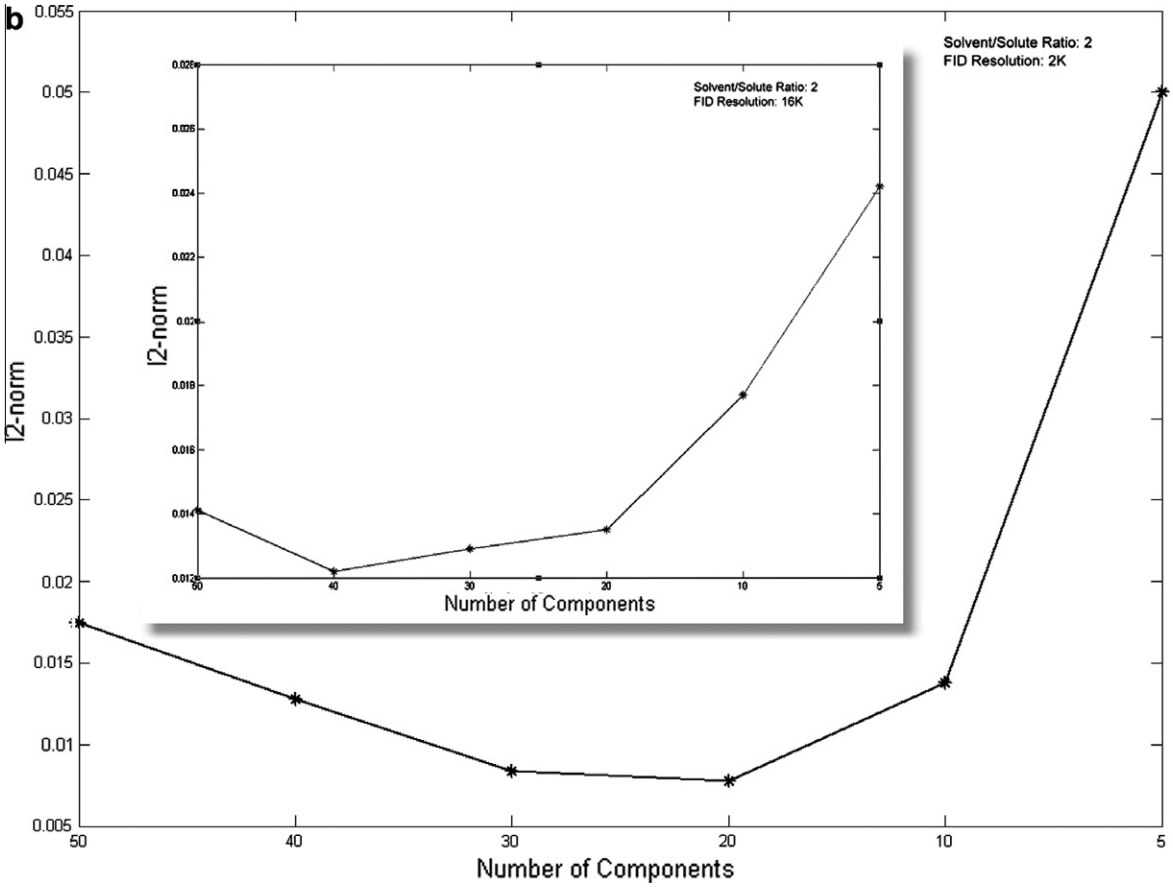

Fig. 4 (*continued*)

spectrum of human urine. The synthetic spectrum allows the quantitative investigation of the performance of SSA since the original protein spectrum before adding the solvent signal can be compared with the result after application of SSA/ALS. Especially, the back-calculation of the one-dimensional time domain signal of HPr protein allows the identification of the peaks of interest that are located in the water artifact region before adding the solvent signal. Since the back-calculated data are generated without any digital filtering, the algorithm automatically recognized that no group delay data points have to be taken into account. After the addition of the partly saturated water signal to the time domain signal of the protein, the SSA removal procedure was applied. As demonstrated later, for 1D-spectra consisting of 16 K ore more complex time domain data points, 40 components would be extracted by SSA. Since the simulated 1D spectrum of HPr has not a high digital resolution (2 K), 20 components showed to be sufficient and a window size W of 92 Hz was automatically calculated for ALS. The central trace of Fig. 2 shows the simulated HPr spectrum after exponential filtering and Fourier transformation of the time domain data. The bottom spectrum was obtained by addition of a very strong solvent time domain to the simulated FID using an identical domain filtering followed by a Fourier transformation. The strong water signal obscures the protein resonances lying in the center of the spectrum and introduces strong baseline distortions with anti-phase contributions and also additional truncation artifacts especially visible in the high field and the low field regions of the spectrum. After application of the 1D version of AUREMOL (Fig. 2 top trace), the water signal is almost reduced to zero, the baseline is almost perfect, and most of the truncation artifacts have been removed. The intensities of the protein signals are not influenced by the procedure, a property very important for quantitative evaluations of NMR spectra.

A zoom of the central area of the spectrum is shown in Fig. 3. The HPr resonances in the range between 4.4 ppm and 5.5 ppm are severely compromised by the residual water signal and can hardly be evaluated. The situation changes when SSA/ALS is applied (Fig. 3, top trace): apart from a small region between 4.7 ppm and 5.0 ppm all HPr resonances are recovered with accurate intensities. Thus, the algorithm shows a good performance by significantly reducing the distortions all over the spectrum and in the critical solvent region, but apparently the original water signal was too intense to allow a complete recovery of the peaks in the most compromised central region (see top trace of Fig. 3).

The qualitative assessment of the performance of AUREMOL-SSA/ALS by inspection of the spectra (Figs. 2 and 3) has been supported by a quantitative analysis. As starting point, the l2-norm (Euclidian norm) between the original simulated HPr spectrum and the same spectrum with the addition of a solvent signal 5000 times stronger than a typical amide protein resonance or 500 times stronger than the strongest protein signals (the superposed signals in methyl region) has been calculated after application of AUREMOL-SSA/ALS. The l2-norm of that spectrum has arbitrarily been set to 1 and the values obtained with other solvent/signal amplitude ratios were scaled correspondingly (Fig. 4). As demonstrated in Fig. 4a the use of SSA even on a very distorted spectrum (where the solvent signal is 150 times stronger than the most intense protein resonance) can definitively improve the spectral analysis. As to be expected, when the relative intensity of the

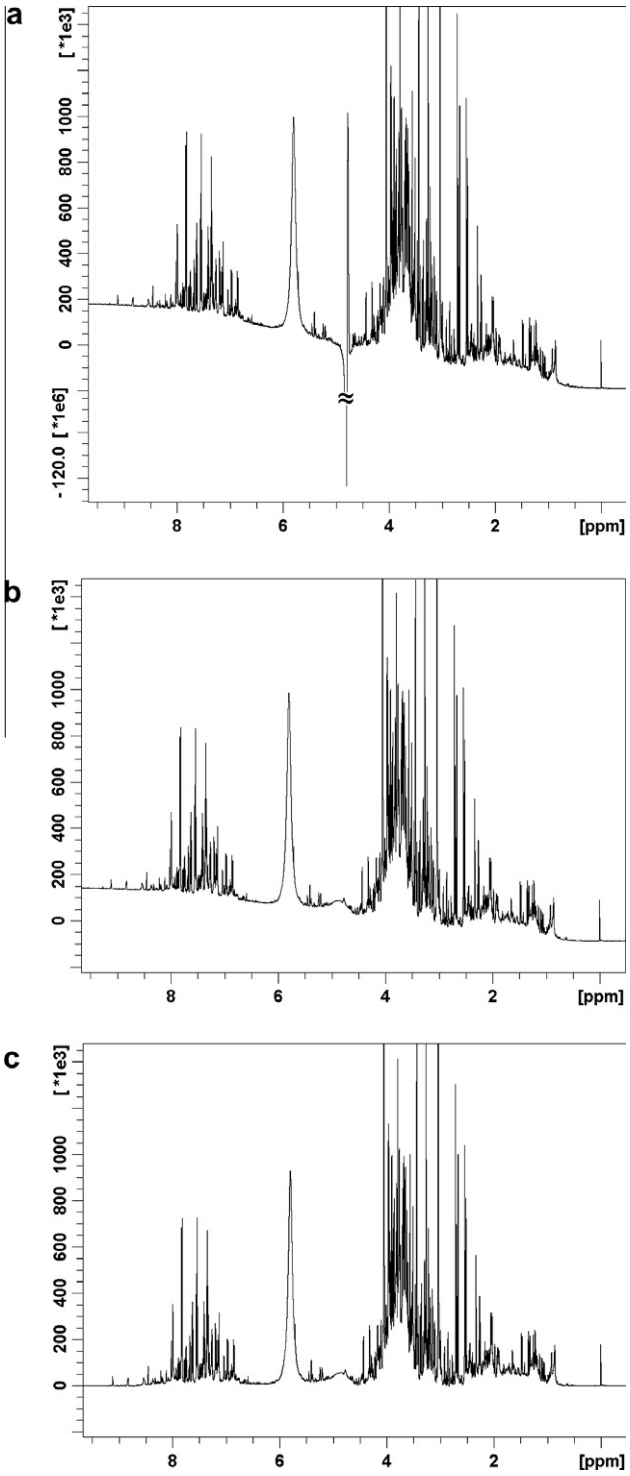

Fig. 5. AUREMOL-SSA/ALS applied on experimental data with strong baseline distortions. The one-dimensional urine spectrum is recorded at 600 MHz. Solvent removal and baseline correction by means of SSA and ALS. (a) Complete experimental spectrum of the urine, (b) spectrum after application of SSA, and (c) the spectrum after application of ALS.

solvent resonance is reduced, the performance of the algorithm as measured by the l2-norm is improved reaching an optimum when the water artifact amplitude is about twice that one of the strongest protein resonance (see the region zoomed out from Fig. 4a). Further reduction of the solvent signal mixed to the protein spectra leads to a slow increase of the l2-norm. This behavior directly follows from the method itself: SSA removes the component of the FID having the largest variance. When the intensity of the solvent signal becomes of the order of the intensity of the protein resonances, the water signal has not the largest variance and thus is not recognized properly by SSA. When its intensity is smaller than roughly the half of the most intense protein resonance, the algorithm removes just this component of the protein signal. Thus, a meaningful application of SSA is clearly related to the relative amplitude of the water artifact thus NMR spectra obtained by excitation sculpting are not suitable for SSA processing. AUREMOL gives out a warning that SSA should not be applied to data where the water signal is not the dominant signal.

In Fig. 4a an FID with low digital resolution (2 K complex data) was used and only 20 components were extracted for the SSA processing. However, as Fig. 4b shows for spectra with higher digital resolution more than 20 components have to be used for the analysis. Starting with FIDs with an optimal solvent to protein intensity ratio of 2, the dependence of the l2-norm on number of components used in the SSA analysis was followed for a 2 K and a 16 K FID. In accordance with Malloni et al. [6], optimal values are obtained for the low resolution FID with 20 components. Such a low resolution FID is usually recorded in multi-dimensional spectra. For the high resolution data the optimal removal of the water signal is reached using 40 components (see the region zoomed out in Fig. 4b). In AUREMOL the number of components used is automatically adapted to the size of the data handled.

The experimental urine spectrum shown in Fig. 5 is characterized by a large number of well resolved, strongly overlapping resonance lines. The water signal is still very strong (note that the negative parts of the signal were cut in Fig. 5a) but its line width is so small that only small distortions in the central part are visible. The actual spectrum was mainly selected because it also showed a strong baseline distortion that could possibly compromise SSA or ALS. The urine spectrum was also recorded with oversampling and digitally filtering and it is thus a test for the applicability of SSA/ALS to this type of experimental data. As described above, the algorithm automatically calculates the time domain data points belonging to the group delay in accordance with the acquisition parameters (72 time domain data points). They were removed before applying the SSA procedure and appended to the signal only before the final Fourier transformation. As shown above a decomposition by SSA in 20 components is not optimal for 1D spectra with high digital resolution. In all cases tested where SSA was applied to one-dimensional spectra (with a data size of 16 K and up to 128 K), empirically a separation of 40 components (as in the case reported here) was sufficient and mandatory to avoid the disappearance or the reduction of signals of interest that are close to the solvent resonance.

4.2. Automated linear spline (ALS) baseline correction

The automated baseline correction by ALS is an integral component of the solvent suppression technique by SSA since this latter attenuates baseline distortions caused by the solvent signal itself but cannot remove them completely. ALS can, of course, be used independently of SSA. A typical case where ALS has to be applied after solvent suppression by SSA would be on the urine spectrum (Fig. 5). Here, SSA as described in this application removes the strongest signal (that is the water signal) from the spectrum but has only a small effect on the baseline (see Fig. 5b). A sinusoidal baseline distortion caused by data clipping was additionally introduced in the spectrum. As Fig. 5c shows AUREMOL-ALS perfectly removes the baseline distortion. With the calculation method for the determination of the window size W used in multi-dimensional NMR a too small size of 6 Hz would result, however

the window size calculated from the largest effective line width is 120 Hz as it is required for a proper baseline correction.

5. Conclusion

We show in this paper that singular spectral analysis (SSA) combined with automated linear spline (ALS) provides a tool to automatically remove the solvent signal from NMR spectra of aqueous solutions and to correct the baseline distortions in one-dimensional NMR spectra. The method has been tested successfully for higher-dimensional NMR spectra earlier [6]. We demonstrate that with the modifications given here it can also be applied to one-dimensional spectra that usually show a much larger degree of signal overlap and large variations of line widths compared to the typical two- or three-dimensional spectra of proteins. Moreover, it has been re-arranged in a way to deal with time domain signals having a higher digital resolution where an increased number of components need to be extracted. In principle, it could be useful to nullify more than one component before reconstructing the signal, an idea that will be worked out in future. We have demonstrated that when a dominant solvent signal is no more present, ALS alone leads to an excellent baseline correction. A perfect baseline is especially mandatory for the quantification of the signals since even small baseline variations lead to large errors in the calculation of concentrations of the compounds in the sample. The solvent peak could hide many solute resonances of interest close to it or provide a bias to the determination of the peak integral. In protein NMR spectroscopy, application of SSA may recover the structurally important H^{α}-resonances that are otherwise only visible in spectra recorded in D_2O.

It is clear that using modern NMR spectrometers an excellent water suppression and almost perfect baselines can be achieved with selective excitation techniques such as WET [17], WATERGATE [18,19], and excitation sculpting [20] or by applying more complicated selective pre-saturation sequences such as PURGE [21] instead of the simple NOESY-type pre-saturation sequence used here. Selective excitation methods have always the disadvantage, that a rather large spectral range around the water signal is attenuated or not visible at all that may contain valuable information. In addition, the quality of the water suppression also depends on the concentration of solute under consideration. At millimolar concentrations these methods can attenuate the water signal to such a degree that it is not stronger than a typical resonance of the solute. In contrast at micromolar concentrations often relevant in biology these methods fail and the application of SSA/ALS would lead to a significant improvement of the spectral quality. Since the proposed algorithm is completely automated and also includes the handling of digitally filtered (oversampled) data, Fourier transformation, and phase correction in accordance with the group delay without any user intervention, it can be used just routinely on these data.

Acknowledgments

This work was supported by the Bavarian Science Foundation (ForNeuroCell), the European Union (SPINE-2), and the Fonds of the Chemical Industry (FCI).

References

[1] J.C.C. Brown, G.A. Mills, P.J. Sadler, V. Walker, ^{1}H NMR studies of urine from premature and sick babies, Magn. Reson. Med. 11 (1989) 193–201.
[2] J.K. Nicholson, M.P. O'Flynn, P.J. Sadler, A.F. MacLeod, S.M. Juul, P.H. Soenksen, Proton-nuclear-magnetic-resonance studies of serum, plasma and urine from fasting normal and diabetic subject, Biochem. J. 217 (1984) 365–375.
[3] O.A.C. Petroff, R.K. Yu, T. Ogino, High-resolution proton magnetic resonance analysis of human cerebrospinal fluid, J. Neurochem. 47 (1986) 1270–1276.
[4] U. Sahrbacher, A. Pehlke-Rimpf, G. Rohr, W. Eggert-Kruse, H.R. Kalbitzer, High resolution proton magnetic resonance spectroscopy of human cervical mucus, J. Pharm. Biomed. Anal. 28 (2002) 827–840.
[5] W. Gronwald, H.R. Kalbitzer, Automated structure determination of proteins by NMR spectroscopy, Prog. NMR Spectr. 44 (2004) 33–96.
[6] W.M. Malloni, S. De Sanctis, A.M. Tomé, E.W. Lang, C.E. Munte, K.P. Neidig, H.R. Kalbitzer, Automated solvent artifact removal and base plane correction from multidimensional NMR protein spectra by AUREMOL-SSA, J. Biomol. NMR 47 (2) (2010) 101–111.
[7] R. Saffrich, W. Beneicke, K.P. Neidig, H.R. Kalbitzer, Baseline correction in *n*-dimensional NMR spectra by sectionally linear interpolation, J. Magn. Reson. 101 (B) (1993) 304–308.
[8] A. Ried, W. Gronwald, J.M. Trenner, K. Brunner, K.P. Neidig, H.R. Kalbitzer, Improved simulation of NOESY spectra by RELAX-JT2 including effects of J-coupling, transverse relaxation and chemical shift anisotropy, J. Biomol. NMR 30 (2) (2004) 121–131.
[9] T. Maurer, S. Meier, N. Kachel, C.E. Munte, S. Hasenbein, B. Koch, W. Hengstenberg, H.R. Kalbitzer, High-resolution structure of the histidine-containing phosphocarrier protein (HPr) from *Staphylococcus aureus* and characterization of its interaction with the bifunctional HPr kinase/phosphorylase, J. Bacteriol. 186 (2004) 5906–5918.
[10] M. Ghil, M.R. Allen, M.D. Dettinger, K. Ide, Advanced spectral methods for climatic time series, Rev. Geophys. 40 (2002) 3.1–3.41.
[11] D. Moskau, Application of real time digital filters in NMR spectroscopy, Concepts Magn. Reson. 15 (2002) 164–176.
[12] K.I. Diamantaras, S.Y. Kung, Principal Component Neural Networks: Theory and Application, John Wiley, NY, 1996.
[13] I.T. Joliffe, Principal Component Analysis, Springer, Berlin, 2002.
[14] N. Golyandina, V. Nekrutkin, A. Zhigljavsky, Analysis of Time Series Structure: SSA and Related Techniques, Chapman and HALL/CRC, 2001.
[15] A.R. Teixeira, A.M. Tomé, M. Boehm, C.G. Puntonet, E.W. Lang, How to apply non-linear subspace techniques to univariate biomedical time series, IEEE Trans. Instrum. Meas. 58 (8) (2009) 2433–2443.
[16] P. Guentert, K. Wuethrich, FLATT-A new procedure for high-quality baseline correction of multidimensional NMR spectra, J. Magn. Reson. 96 (1992) 403–407.
[17] S.H. Smallcombe, S.L. Patt, P.A. Keifer, WET solvent suppression and its applications to LC-NMR and high-resolution NMR spectroscopy, J. Magn. Reson. Ser. A 117 (1995) 295–303.
[18] M. Piotto, V. Saudek, V. Sklenar, Gradient-tailored excitation for single-quantum NMR-spectroscopy of aqueous-solutions, J. Biomol. NMR 2 (1992) 661–665.
[19] V. Sklenar, M. Piotto, R. Leppik, V. Saudek, Gradient-tailored water suppression for H-1-N-15 HSQC experiments optimized to retain full sensitivity, J. Magn. Reson. A 102 (1993) 241–245.
[20] T.L. Hwang, A.J. Shaka, Water suppression that works—excitation sculpting using arbitrary wave-forms and pulsed-field gradients, J. Magn. Reson. A 112 (1995) 275–279.
[21] A.J. Simpson, S.A. Brown, Purge NMR: effective and easy solvent suppression, J. Magn. Reson. 175 (2005) 340–346.