Journal of Magnetic Resonance 203 (2010) 129–137

Contents lists available at ScienceDirect

Journal of Magnetic Resonance

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

Benchmarking NMR experiments: A relational database of protein pulse sequences

Russell R.P. Senthamarai [a], Ilya Kuprov [c,*], Konstantin Pervushin [a,b,*]

[a] School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551, Singapore
[b] Biozentrum of University of Basel, Klinlegberg-Str. 70, CH-4056 Basel, Switzerland
[c] Oxford e-Research Centre, University of Oxford, 7 Keble Road, Oxford OX1 3QG, UK

ARTICLE INFO

Article history:
Received 23 September 2009
Revised 7 December 2009
Available online 23 December 2009

Keywords:
NMR
Relational database
Spin state space restriction
Spinach
Spin dynamics
Benchmarking of NMR experiments

ABSTRACT

Systematic benchmarking of multi-dimensional protein NMR experiments is a critical prerequisite for optimal allocation of NMR resources for structural analysis of challenging proteins, e.g. large proteins with limited solubility or proteins prone to aggregation. We propose a set of benchmarking parameters for essential protein NMR experiments organized into a lightweight (single XML file) relational database (RDB), which includes all the necessary auxiliaries (waveforms, decoupling sequences, calibration tables, setup algorithms and an RDB management system). The database is interfaced to the Spinach library (http://spindynamics.org), which enables accurate simulation and benchmarking of NMR experiments on large spin systems. A key feature is the ability to use a single user-specified spin system to simulate the majority of deposited solution state NMR experiments, thus providing the (hitherto unavailable) unified framework for pulse sequence evaluation. This development enables predicting relative sensitivity of deposited implementations of NMR experiments, thus providing a basis for comparison, optimization and, eventually, automation of NMR analysis. The benchmarking is demonstrated with two proteins, of 170 amino acids I domain of αXβ2 Integrin and 440 amino acids NS3 helicase.

1. Introduction

Nuclear magnetic resonance (NMR) spectroscopy is an important analytical tool, widely used for structural, kinetic and dynamic studies of biological molecules [1]. It is a remarkably informative technique; however, with large macromolecular systems NMR spectroscopy is complicated by the low sensitivity, rapid transverse relaxation and spectral overlap [2]. Though sensitivity improvements are continuing [3–5], molecular size is a fundamental limitation: transverse relaxation rates grow linearly with the effective volume [2]. The loss of sensitivity due to transverse relaxation depends on the details of the trajectory that the spin system describes under a given pulse sequence, and even similar sequences are known to differ markedly in their sensitivity to relaxation [6,7]. A theoretical tool providing a set of unified performance measures is essential for pulse sequence benchmarking.

Over the years, a number of computational approaches have evolved for spin dynamics simulation. While the analytical simulation tools [8–11], are helpful in understanding spin dynamics under most pulse sequences, they can only be used for very small systems and simple waveforms. Numerical simulation programs (ANTIOPE [12], GAMMA [13], SIMPSON [14], SIMPLTN [15], VNMRS [16], QSim [17], SPINEVOLUTION [18], BlochLib [19] and many others) offer a way to simulate and analyze whole NMR experiments with accuracy that is sufficient for benchmarking. However, most existing numerical tools are limited to about 5–8 spins by the "exponential scaling wall" as the dimensionality of the computational problem increases rapidly with the size of the system [20]. The only software package currently capable of operating with polynomial scaling is Spinach [21,22].

In this article, we report a relational XML database (NMR-RDB) of Bruker NMR sequences that includes the auxiliary data (waveforms, decoupling sequences, calibration tables, setup algorithms, etc.) and is interfaced to the algebraic core of the Spinach library [21,22]. The resulting software framework is able to simulate a variety of multi-channel, multi-dimensional NMR experiments (including phase cycles, simultaneous shaped pulses, composite pulse decoupling, etc.) on a large (200 + spins, 32 are used below) generic spin system representing a fragment of a protein. Full DD–CSA relaxation superoperator [23,24] including all cross-correlation terms is implemented. The result is a unified framework for benchmarking protein NMR experiments in solution, providing selection criteria for NMR experiments suitable for the protein at hand and eventually enabling automation of the experiment setup.

2. Methods and algorithms

The state space restriction approximation used by the Spinach library has been described elsewhere [22]. It is based on a fairly

* Corresponding authors.
E-mail addresses: ilya.kuprov@oerc.ox.ac.uk (I. Kuprov), kpervushin@ntu.edu.sg (K. Pervushin).

1090-7807/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2009.12.008

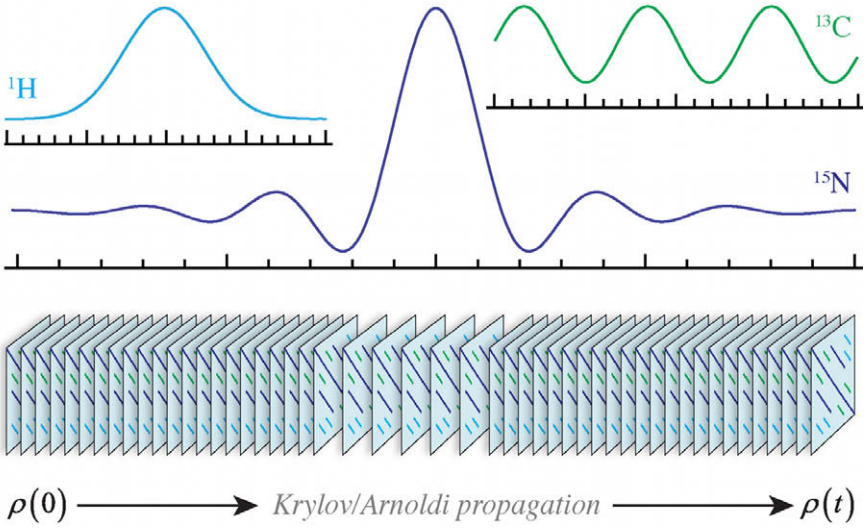

Fig. 1. Common time grid resampling: figure showing the varying time grids of the input waveforms and their sampling onto a common time grid.

simple observation that, in finite-time experiments, any large quantum system only ever occupies a tiny fraction of its Hilbert space. This is particularly true for liquid state NMR experiments and leads to dramatic reduction in matrix dimensions [22]. The current stable version of Spinach, including documentation, is available at http://spindynamics.org and in Supplementary information.

2.1. Time propagation

The time propagation of a quantum system in the Liouville space [25] amounts to a suitably time-ordered sequence of the following operations:

$$\hat{\rho}(t+\Delta t) = \exp\left(-i\hat{\hat{L}}(t)\Delta t\right)\hat{\rho}(T) \tag{1}$$

where $\hat{\rho}(t)$ is the density matrix (including, if necessary, spatial degrees of freedom), $\hat{\hat{L}}(t)$ is the Liouvillian superoperator (including relaxation and chemical kinetics contribution, if present) and Δt is kept sufficiently small. Importantly, efficient algorithms exist for direct calculation of $\hat{\rho}(t+\Delta t)$ from $\hat{\rho}(t)$ and sparse representation of $i\hat{\hat{L}}(t)\Delta t$, bypassing the calculation of the matrix exponential [21]. The resulting density operator trajectory is an ordered list of density operators which can be projected onto an observable operator.

In most cases the temporal sequence of NMR controls (rf-pulses, delays, etc.) can be compiled into a sequence of matrices and the simulation can be carried forward by repeated application of Eq. (1) followed by projection onto the observable operator of interest. A notable technicality arising in the implementation of such schemes can be called *common grid resampling* Fig. 1. While the sequence of events is strictly aligned in their position and duration, their internal time discretisation may vary (for example, a 50-point Gaussian waveform on ^{15}N could be longer than a simultaneous 200-point Q5 pulse on ^{1}H), necessitating the introduction of a common time grid, which should (locally) be at least as dense as the fastest event in the channel stack. The simultaneous shaped pulse function implemented in Spinach performs this resampling and uses the resulting Liouvillian stack to propagate the system in a sequence of local Krylov subspaces [26,27].

2.2. Relaxation treatment

Two implementations of relaxation superoperator are currently available – a local one, where relaxation superoperators are computed exactly in every spin cluster and then re-coupled [22], and global, where the relaxation term between every pair of states in the restricted basis is calculated explicitly. The former is described in detail in Supplementary information. The latter is based on the analysis of the direct product structure of individual states as described below.

The internal basis representation in Spinach is based on irreducible spherical tensor indexing. For every state, only the indices of its $\hat{T}_{l,m}$ components are stored:

$$\hat{\rho} = \hat{T}_{l_1 m_1} \otimes \hat{T}_{l_2 m_2} \otimes \ldots \otimes \hat{T}_{l_n m_n} \Longleftrightarrow \begin{pmatrix} l_1 & l_2 & \ldots & l_n \\ m_1 & m_2 & \ldots & m_n \end{pmatrix} \tag{2}$$

This provides a compact state description and permits the usage of a formidable body of analytical knowledge about the properties of such direct products.[1] In the case of Bloch–Redfield–Wangsness relaxation superoperator [28,29], it is known that $\hat{\hat{R}}$ has a very sparse structure – basis vectors differing in the state of more than three spins do not cross-relax under dipole and CSA interactions considered, and specific relaxation mechanisms have well-defined cross-relaxation paths. Thus, for each bilinear interaction Hamiltonian $\hat{H}_{ij}(t)$ connecting operators $\hat{T}_{l_i m_i}$ and $\hat{T}_{l_j m_j}$, a set of spin states may be identified with the members differing only in the operators $\hat{T}_{l_i m_i}$ and $\hat{T}_{l_j m_j}$. This enables compact tabulation of all relevant auto- and cross-relaxation rates for a generic pair of interacting spins computed as described in Supplementary information. This table is used to rapidly calculate the relaxation superoperator (currently in the isotropic tumbling approximation) in the restricted state space. For each individual pair of spins, dipole–dipole interactions and the (axial) CSAs of each spin are included, thus enabling the simulation of the TROSY effect. We further restrict computations of the cross-correlated cross-relaxation terms to DD and CSA interactions to directly *J*-coupled spins. This significantly reduces the computation time by limiting the number of unique off-diagonal terms in the relaxation matrix. The computation is completed via iteration through all available DD and CSA interactions.

In specific cases where relaxation to thermal equilibrium (rather than all-zero density matrix) is required for accurate experiment description, the full non-uniform Liouville–von Neumann equation incorporating the thermal equilibrium state $\hat{\rho}_0$ is rewritten as a uniform equation by concatenating $\hat{\rho}(t)$ and $\hat{\rho}_0$ into a single column vector [30].

[1] Many routines in Spinach operate semi-analytically to take advantage of this.

$$\frac{\partial}{\partial t}\begin{pmatrix}\hat{\rho}\\ \hat{\rho}_0\end{pmatrix}=\begin{pmatrix}-i\left(\hat{\hat{L}}+\hat{\hat{R}}\right)\hat{\rho} & \hat{\hat{R}}\hat{\rho}_0\\ 0 & 0\end{pmatrix}\begin{pmatrix}\hat{\rho}\\ \hat{\rho}_0\end{pmatrix} \tag{3}$$

and using the standard exponential propagator solution. This procedure doubles the matrix dimension, but it is still the most straightforward way of dealing with the non-uniform version of Liouville–von Neumann equation.

2.3. Magnetization transfer graphs and relaxation profiles

In order to visualize the magnetization flow through the experiment, a pulse sequence timeline is built Fig. S1, with the nodes at each time point corresponding to the largest coefficients $c_i(t)$ in the expansion of the density operator in a chosen orthonormal operator basis $\{\hat{B}_i\}$ (time stepping is chosen to coincide with the major pulse sequence events). The node size is set proportional to $|c_i(t)|$. The coherence flux s_{ijk} between nodes i and j at the sequence stage k is computed as

$$s_{ijk}=\left(\left[\hat{\hat{U}}_k\hat{B}_i\right]\hat{B}_j^{\dagger}\right) \tag{4}$$

where $\hat{\hat{U}}_k$ is the total propagator for the sequence stage k. The display threshold for the absolute value of s_{ijk} was set to 0.03, which was found to provide a good balance between pathway importance and timeline clutter.

The resulting timelines have proven to be useful for visual and automatic identification of productive magnetization transfer pathways connecting the initial and target density operators and non-productive pathways resulting into the loss of magnetization highlighted, respectively, in magenta and yellow in Fig. S1 (Supplementary information). This separation of the density operator into productive and non-productive parts at every stage in the pulse sequence enables construction of the coherence loss profile of NMR pulse sequences defined as an ordered list of the effective coherence loss rates r_k defined as

$$r_k=\frac{||\rho_k^{\text{prod}}||-||\rho_{k+1}^{\text{prod}}||}{\Delta t_k} \tag{5}$$

where Δt_k is the duration of the kth stage of the pulse sequence. With sufficiently detailed time quantization of the sequence, Eq. (5) approximates the time-dependent rate of useful coherence loss, $r(t)=d\|\rho^{\text{prod}}(t)\|/dt$, so that the efficiency profile of the experiment can be constructed.

2.4. Performance metrics

The experiments deposited in NMR-RDB are performance annotated using the following parameters as the numerical quantum bounds [31,32] in the absence ($b_{\max}$) and presence ($b^r_{\max}$) of relaxation:

$$b^r_{\max}=Tr(\hat{O}^{\dagger}\hat{\rho}_{\text{acq}}) \tag{6}$$

where $\hat{O}$ is the normalized observable operator and $\hat{\rho}_{\text{acq}}$ is the density operator at the beginning of the signal acquisition. To account for the quadrature detection scheme, these calculations can be performed twice for both quadrature components

The other useful parameters listed in Tables 1a and 1b) are derivatives from the two above and include R_{PP}, the apparent relaxation rate of the entire pulse sequence with the chemical shift evolution periods set to their respective minimal values, and R_{CS}, the apparent relaxation rate of the coherences evolving in the time-incremented chemical shift encoding periods. With the corresponding duration of the pulse program T, these rates can be used to predict the signal-to-noise ratio $S/N^{n\text{D}}$ in the nD spectra acquired with the corresponding pulse sequence starting from the observed sensitivity in the seeding 1D ^{1}H (see next section).

One of the central performance measures computed with NMR-RDB/Spinach is the signal-to-noise ratio $S/N^{n\text{D}}$ in nD spectra in relation to $S/N^{1\text{D}}$ in the reference 1D ^{1}H-detected spectrum measured with the same free induction decay (FID) acquisition parameters as the directly detected dimension of the nD spectrum. For simplicity we assume that all spectra are measured with one scan. If more scans are employed, e.g. $ns>1$, then the resulting S/N are scaled up with $\sqrt{ns}$. In the first step, $S/N^{2\text{D}}$ in two-dimensional orthogonal planes of the nD experiment are computed by time-domain sampling along the corresponding indirect dimension while keeping all other incremented delays to their minimum. Eq. (7) provides the reconstructed one-dimensional slice along the indirect dimension of the 2D plane normalized to the amplitude of the reference 1D ^{1}H spectrum,

$$S(\omega)=F_+\left[(f_{tr}(t)e^{-R_{\text{eff}}t}+n(t))f_{\text{apod}}(t)\right] \tag{7}$$

where F_+ is a discrete Fourier transform, R_{eff} is the effective relaxation rate defined above, $f_{\text{apod}}(t)$ is the apodization function and $n(t)$ is Gaussian noise with standard deviation typically set to 10% of $f(t)$ (see below), and t is the incremented time-delay in the indirectly sampled dimension. The normalized transfer function is then defined as

Table 1a
Simulated performance of 2D ^{1}H–^{15}N correlation experiments for uniformly ^{13}C,^{15}N-labeled protein (αXβ2 Integrin I domain) with τ_c = 20 ns.

Experiments[a]	Performance metrics[b]					
	$b_{\max}/b^r{}_{\max}$[c]	T[d] (ms)	R_{PP}[e] (s^{-1})	R_{CS}[f] (s^{-1})	$\Delta\nu_{1/2}$ ^{15}N[g] (Hz)	$S/N^{2\text{Dh}}$
HMQC	0.69/0.49	12.7	30.5	36.0	20.8	19.6
HSQC	0.67/0.46	13.2	28.5	26.0	18.4	22.5
HSQC-PEP-PFG	0.69/0.32	26.5	26.0	26.0	18.4	22.1
TROSY	0.38/0.23	17.8	28.0	9.5	16.2	20.3
TROSY-PFG	0.38/0.21	22.8	25.0	9.5	16.2	18.4
TROSY-D.N.	0.38/0.22	23.9	20.2	9.5	16.2	19.3

[a] Experiments are HMQC [43], HSQC [44], HSQC-PEP-PFG [45–47], TROSY [37], TROSY-PFG [37], TROSY-D.N. [39].
[b] In t_1 dimension 128 complex increments with spectral width sw(^{15}N) = 2980 Hz at B_0 = 700 MHz are simulated.
[c] Maximum transfer bounds computed in the absence and presence of relaxation.
[d] Time duration of the pulse sequence with chemical shift evolution periods set to the minimal values.
[e] Effective relaxation rate of the entire experiment with chemical shift evolution periods set to the minimal values.
[f] Effective relaxation rates for the coherences used to indirectly encode chemical shifts.
[g] Expected line-width in the indirectly detected dimension.
[h] Signal-to-noise ratio expected in 2D spectrum in the presence of relaxation with S/N = 10 is the seeding 1D ^{1}H spectrum.

Table 1b
Simulated performance of 2D ^{1}H–^{15}N correlation experiments for uniformly ^{2}H,^{13}C,^{15}N-labeled protein (NS3Helicase) with τ_c = 40 ns.

Experiments[a]	Performance metrics[b]					
	b_{max}/b^{r}_{max}[c]	T[d] (ms)	R_{PP}[e] (s^{-1})	R_{CS}[f] (s^{-1})	$\Delta\nu_{1/2}$ ^{15}N[g] (Hz)	S/N^{2Dh}
HMQC	0.69/0.52	12.7	22.0	33.0	18.0	21.6
HSQC	0.67/0.52	13.2	19.2	32.4	18.0	23.3
HSQC-PEP-PFG	0.67/0.37	26.5	21.2	32.4	18.0	24.1
TROSY	0.38/0.24	17.8	25.6	10.6	15.2	20.7
TROSY-PFG	0.38/0.22	22.8	22.4	10.6	15.2	19.5
TROSY-D.N.	0.38/0.24	23.9	19.0	10.6	15.2	20.7

[a] Experiments are HMQC [43], HSQC [44], HSQC-PEP-PFG [45–47], TROSY [37], TROSY-PFG [37], TROSY-D.N. [39].
[b] In t_1 dimension 128 complex increments with spectral width sw(^{15}N) = 2980 Hz at B_0 = 700 MHz are simulated.
[c] Maximum transfer bounds computed in the absence and presence of relaxation.
[d] Time duration of the pulse sequence with chemical shift evolution periods set to the minimal values.
[e] Effective relaxation rate of the entire experiment with chemical shift evolution periods set to the minimal values.
[f] Effective relaxation rates for the coherences used to indirectly encode chemical shifts.
[g] Expected line-width in the indirectly detected dimension.
[h] Signal-to-noise ratio expected in 2D spectrum in the presence of relaxation with S/N = 10 is the seeding 1D ^{1}H spectrum.

$$f_{tr}(t) = \left| \frac{Tr(\hat{O}^{\dagger}\hat{\rho}(t))}{\max[\|Tr(\hat{O}^{\dagger}\hat{\rho}(t))\|]} \right| \tag{8}$$

The simulations are performed for each component of the Cartesian orthogonal or echo-anti-echo quadrature sampling scheme thus resulting in the complex, phase-sensitive spectrum. To compute S/N^{2D}_{H-X} 10% of Gaussian noise $n(t)$ is added to the time domain signal prior to apodization and Fourier transform, which corresponds to the signal-to-noise ratio in the reference 1D spectrum of 10:1. The parameters S/N^{2D}_{H-X} and $\Delta\nu_{1/2}$ are directly estimated from the simulated spectrum as shown in Fig. S2 (Supplementary information). Due to the added random noise the simulations are repeated 200 times to report the mean and RMSD values.

The sensitivity in nD spectrum is computed according to Eq. (9)

$$S/N^{nD} = b^{r}_{max} \prod_{i=1}^{n-1} S/N^{2D}_{H-X_i} \tag{9}$$

where b^{r}_{max} is the maximum magnetization transfer bound in the presence of relaxation and the product is performed over the number of indirectly detected dimensions. For example, S/N^{3D} in the 3D HNCA experiment will be a product of S/N^{2D}_{H-N} and $S/N^{2D}_{H-C\alpha}$ spectral planes. Finally, we assume that the same steady state level of initial spin polarization is generated for all experiments, thus avoiding the need for explicit modeling of the return of the spin polarization back to its thermal equilibrium.

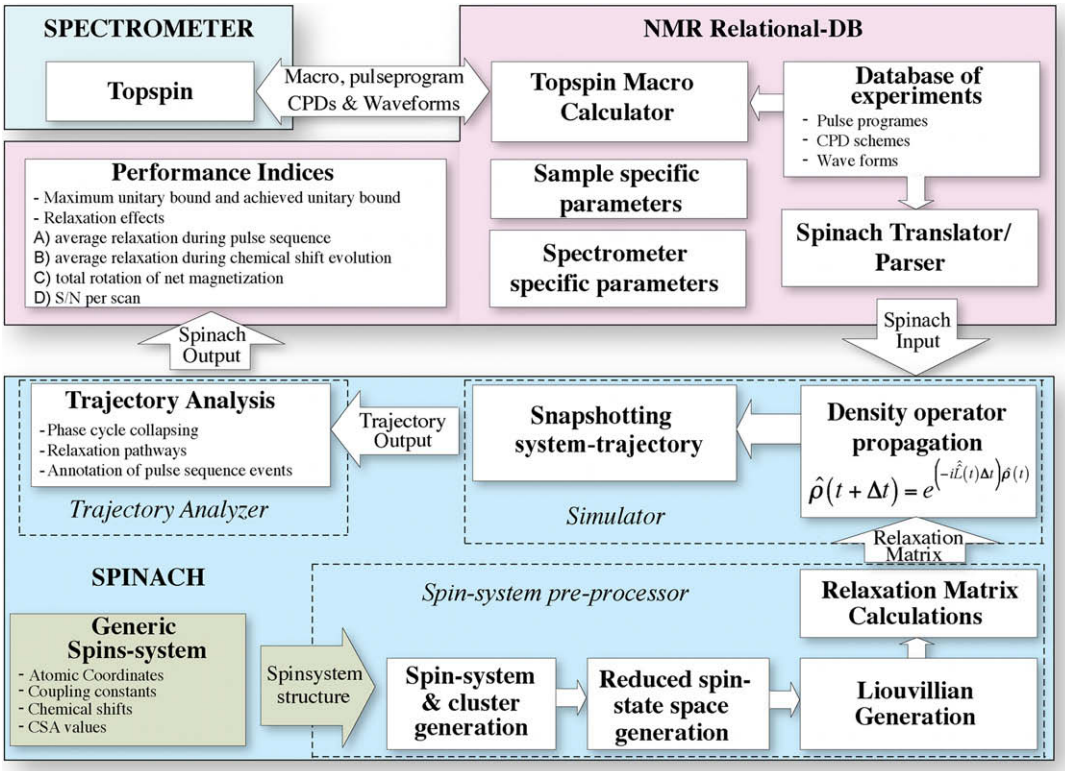

Fig. 2. Spinach-NMR-RDB: the flow of events involved in the setup and analysis of experiments deposited in the NMR-RDB. The arrows extend from the boxes, continuously, indicates, the procedure is automated and the arrows connected with the boxes indicates the need of manual intervention, despite the complete procedure could be automated with an external Matlab (www.mathworks.com) script. The description of individual modules and procedures are described in the Supplementary information.

2.5. Relational database and analysis of NMR experiments

The relational database is used to generate setup procedures for an NMR experiment or prepare input for Spinach. The block schematic is shown in Fig. 2.

Relational database of NMR experiments (RDB) is an XML formatted lightweight (stored as a single file) database grouping individual components of NMR experiments such as the pulse sequence, decoupling sequences, RF waveforms, calibration tables, sample-specific information and algorithms needed to set up and process NMR experiments using common attributes found in the data set. For example, all the files needed to setup a TROSY-HNCA experiment can be brought together under the corresponding experiment entry Fig. S3 (Supplementary information). Alternatively, these records can be annotated and grouped by attributes such as data type, contributing author, date of creation, URL reference, etc.

Relational database management system is implemented in the Lua application extension language (www.lua.org) as an extension script for the popular spectrum analysis program CARA (www.nmr.ch) and is itself stored in RDB making it self-contained. Since CARA is a single-file monolithic application (not requiring installation procedure) available for all common platforms, the RDB and management system are not constrained by the choice of platform. The RDB management system exports and imports all components of NMR experiments from and to spectrometer data structure of TopSpin (Bruker Biospin AG), online database (www.trosy.com) or standalone files. For the selected spectrometer and sample it generates macro performing experimental setup procedures at the spectrometer or semi-automatically translates the Bruker pulse programs and parameters to the Spinach representations for the numeric analysis. Due to access to the spectrometer calibration data, a semantic interpretation of the effects of RF

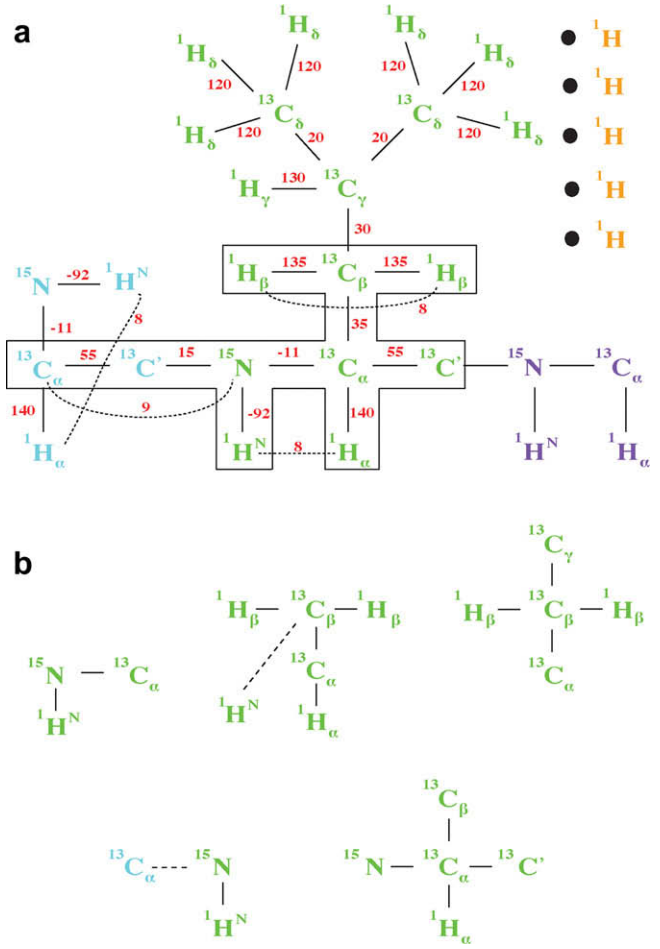

Fig. 3. Spin system: (a) Flattened 2D picture of the generic protein spin system used for sequence benchmarking. Atomic coordinates were taken from NS3Helicase (2JLQ.pdb). Typical coupling constants (in Hz) are shown in red. Dotted lines indicates couplings between indirectly connected spins (b) selected clusters for C_α spin constructed by inspection of the magnetization transfer across C_α spin as described in Section 2. Dotted line represents indirect connectivity between C_α and $^1H^N$. (For interpretation of the references in color in this figure legend, the reader is referred to the web version of this article.)

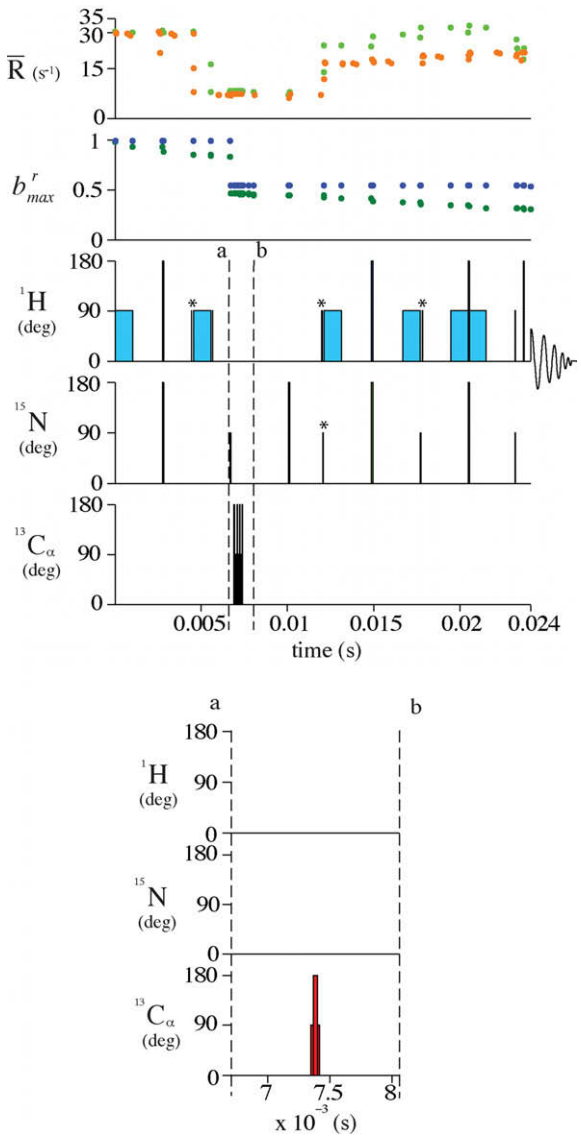

Fig. 4. An example analytics of the TROSY pulse sequences [37,39] simulated using generic spin system, shown in Fig. 3a. Chemical shift evolution, events between (a) and (b), is shown, simplified, at the bottom. The Chemical shift offsets are as follows: $\omega(^1H) = 4.7$ ppm, $\omega(^{15}N) = 119$ ppm and $\omega(^{13}C) = 110$ ppm. The rest of the parameters and the Spinach input script can be found in NMR-RDB online at www.trosy.com. Spectrometer frequency was set to 700 MHz. Chemical shift evolution is simulated for four t_1 – time incremental points. Pulse sequence events, amplitudes and relaxation rates were plotted versus time (milliseconds). Pulses on each channel are represented by bars with the height of the bar specifying the total angle of rotation and the width of the bar marks the pulse duration (some pulses are short enough to appear as thin lines). The norm of the density operator along the productive magnetization pathways, with (green) and without relaxation (blue) were plotted versus time. At the top panel, the relaxation rates, at the end of the corresponding events, are plotted in green and orange, corresponding to ST2-PT and CTE pulse sequence events. (For interpretation of the references in color in this figure legend, the reader is referred to the web version of this article.)

pulses in terms of induced rotations as well as direct analysis of the performance of the experiments is easily available.

Spinach is functionally divided into three modules, spin system pre-processor, simulator and trajectory analyzer.

Spin system generator receives the description of coupled spin system as the input and generates the spin system data structure. This module builds the restricted basis set, generates the Liouvillian and the relaxation superoperator. Detailed procedures and algorithms of these steps are described elsewhere [21,22,33].

Simulator receives translated pulse sequence instructions in the form of Spinach function calls and carries out those actions using the spin system structure obtained from the generator. At this level, the events such as delays and pulses can be quantized to trace evolution of the density operator with the desired time resolution. Snapshots of the density operator along its trajectory as well as auxiliary information such global time scale binding of the events, events durations, chemical shift evolution incremental cycles, phase cycles, quadrature schemes, etc. for the pulse sequence are stored for the subsequent analysis. In order to achieve this, a number of functions on top of the Spinach algebraic engine are implemented [Table S1] (Supplemented information). Together with the snapshotting of density operator, these functions enable monitoring the spin evolution/relaxation during the pulse sequence execution [Table S2] (Supplementary information).

Trajectory analyzer reads in the density operator trajectory and produces magnetization transfer graph upon collapsing the phase cycle. Pulse sequence events are tabulated and the norm of the density matrix and apparent relaxation rates along the productive pathways is outputted in conjunction with each pulse sequence events. In addition, a measure of deviation of the density operator from its relaxation-free trajectory is produced. In essence, this module semantically annotates the pulse sequence events and generates performance measures. The detailed description of algorithms used in the implementation of these modules has been given above.

3. Results and discussion

3.1. Construction of a generic spin system

The key development in the numerical tools described above is the ability to use a single generic spin system to adequately simu-

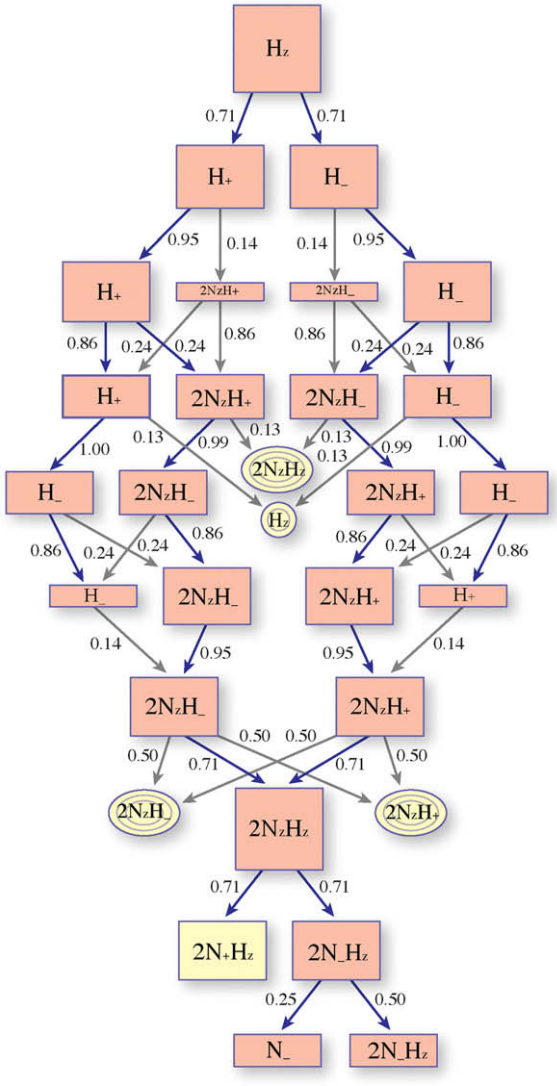

Fig. 5a. Magnetization flow graph: automatically generated flow of magnetization chart during the first INEPT block of TROSY-ST2-PT. Size of each box containing spin state is proportional to the magnetization transferred via that particular spin state during the corresponding event in the pulse sequence. Blue lines represent the strong magnetization flow channels. Productive spin states are shown in magenta and non-productive are shown in yellow. (For interpretation of the references in color in this figure legend, the reader is referred to the web version of this article.)

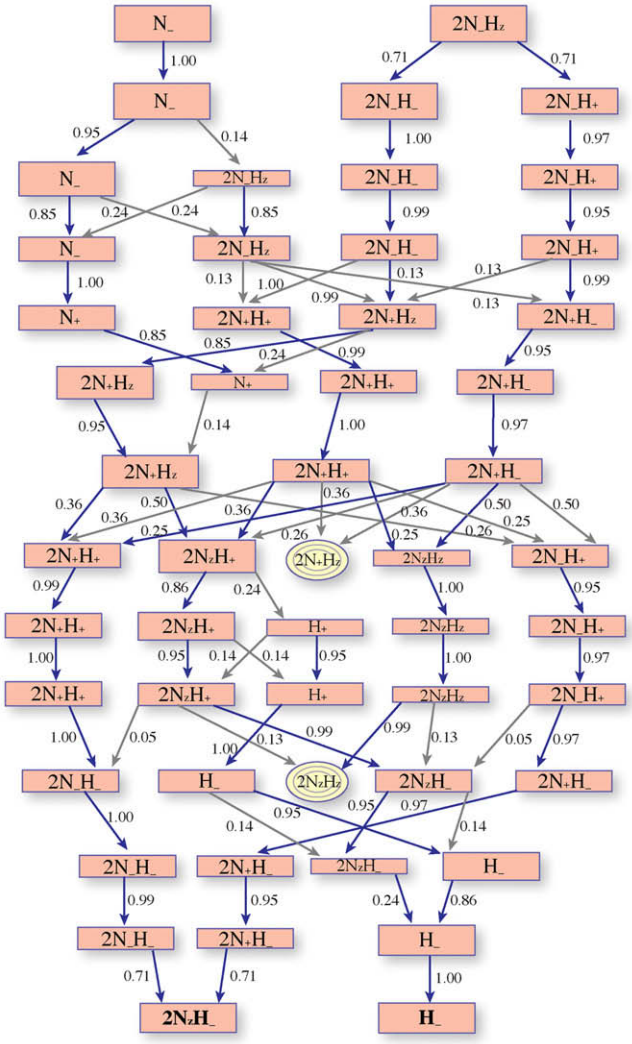

Fig. 5b. Magnetization flow graph: automatically generated flow of magnetization chart during the ST2-PT element of TROSY-ST2-PT. Same conventions are used, as in Fig. 5a.

late the majority of solution state NMR experiments deposited in RDB. The use of the single spin system relaxes the need to tailor specific spin system suitable for the experiment in analysis, thus providing a uniform framework for performance annotation across majority of known pulse sequences. We opted to use a fragment from NS3Helicase (PDB ID: 2JLQ.pdb) comprising 27 *J*-coupled spins from residues Gln-384 to Ser-386 as well as five spatially proximate protons belonging to other residues. Atomic coordinates and chemical shielding tensor values of each atom constitutes the generic spin system can be found in [Table S3] (Supplementary information). The positions of these protons were slightly adjusted to maintain the dipolar density at the backbone spins typically encountered in compact globular proteins. [Tables S4a and S4b] (Supplementary information) report the adjusted dipolar densities in the generated spin system and averaged ^{1}H–^{1}H dipolar densities found in NS3. For perdeuterated proteins, all aliphatic ^{1}H are replaced with ^{2}H isotopes. At the time of writing, a typical calculation takes about 1–2 h on an average workstation.

The values of coupling constants are adopted from NMRGuide 3.5 (Bruker Biospin AG). In the cases of conformation-dependent coupling constants [34], e.g. $^{3}J_{HNH\alpha}$, a single value corresponding to the middle of the range is set. Chemical shielding tensor values of C_α, C_β, carbonyl C, backbone N and amide protons were adopted from [35,36] and generic chemical shielding values were used for the rest of the spins. The chemical shifts of all spins were set to the averaged values reported for the restricted set of diamagnetic proteins in the Biological Magnetic Resonance Data Bank (http://www.bmrb.wisc.edu). Fig. 3a shows a flattened 2D representation of the generated spin system with the coupling constants shown in red.

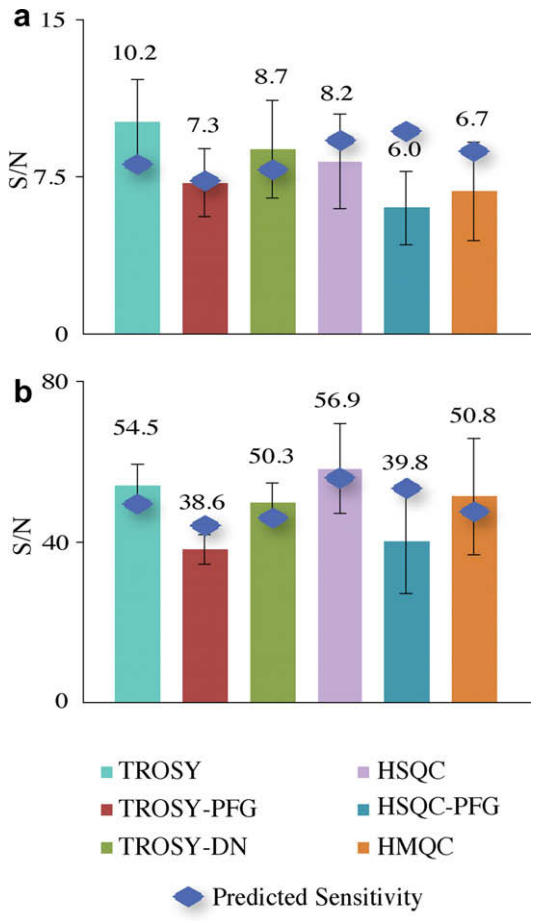

Fig. 6. Comparison of calculated and experimental *S*/*N* between a set of 2D ^{1}H,^{15}N-correlation experiments deposited in NMR-RDB measured with (a), ^{2}H,^{13}C,^{15}N-labeled 51 kDa NS3Helicase in complex with RNA and (b), ^{13}C,^{15}N-labeled 25 kDa αXβ2 Integrin I domain. A value above each bar represents average *S*/*N* of defined number of peaks, common in all the spectra. Acquisition parameters: (a) Number of scans = 8, number of complex time domain points sampled in the indirect dimension = 128 and spectral width (^{15}N) = 2838 Hz. (b) Number of scans = 4, number of complex time domain points sampled in the indirect dimension = 100 and spectral width (^{15}N) = 2980 Hz. The predicted S/N^{2D} are calculated with Eq. (9) using data of [Tables 1a and 1b] and S/N^{1D} = 8.0 and S/N^{1D} = 2.0 for αXβ2 Integrin I domain and NS3Helicase, respectively.

In order to construct the restricted spin state space, the coupled spins are grouped together to form clusters comprising spins entangled into highest spin orders excited in pulse sequences. Each cluster size is kept to minimum and for computational efficiency does not exceed 6. The main cluster set is automatically generated by the JNES algorithm [22] and is manually extended after inspection of propagation of highest order coherences encountered in triple resonance experiments. It is important to note that clusterization is a *basis set generation strategy* [22] – the spin system is always simulated in its entirety in the resulting connectivity-adapted basis set. Fig. 3b depicts the clusters of importance in magnetization transfer via C_α as a hub. However, it should be noted that in TROSY-HNCACB, in C_α/C_β chemical shift evolution period the coherence $(E/2 + H^{N}{}_{z})N_{z}C^{\alpha}{}_{z}C^{\beta} + H^{\beta}{}_{z}H^{\beta}{}_{z}C^{\gamma}{}_{z}$ is generated, requiring the inclusion of eight-spin orders into the basis. This is currently difficult (Spinach library is at the development stage), but a simple workaround was found involving replacing the two ^{1}H 180° decoupling pulses in the middle of C_α/C_β evolution period Fig. S4 (Supplementary information) with continuous composite pulse decoupling sequence. Alternatively, the $^{1}J_{C\beta H\beta}$ couplings can be temporarily disabled, thus preventing the build-up of high spin orders involving H_β.

3.2. Performance annotation of 2D [^{1}H,^{5}N]-correlation experiments

Two experimental schemes of TROSY experiments from NMR-RDB employing ST2-PT [37] and improved coherence transfer elements (CTE) [38,39] are simulated using NMR-RDB/Spinach framework with the generic spin system shown in Fig. 3a. Fig. 4 shows the density operator trajectory analysis including schematic representation of the pulse sequences (note that ST2-PT and improved CTE differ by pulses marked with asterisks). In order to monitor the norm of the density operator and the apparent relaxation rates as a function of time, we plotted the pulse program events as function of time. Therefore, we consistently use the vertical scale to semantically annotate the rotation of magnetization caused by RF pulses, wherein the height and the width represent the total angle of rotation and the event duration on the common time grid, respectively. The corresponding magnetization transfer graph for ST2-PT highlighting the productive and the non-productive magnetization transfer pathways is shown in Fig. 5a and b.

For both experiments the initial density operator represents the Boltzmann equilibrium magnetization on $^{1}H^{N}$ and ^{15}N spins as it is shown in the magnetization transfer graph of Fig. 5a. In the absence of relaxation only ½ of the norm of the density operator is retained after the ^{15}N transverse coherence $2N_{-}(E/2 + H_z)$ is excited by the first ^{15}N 90° pulse due to selection of the TROSY single transition operator by the phase cycle. The anti-TROSY transition is not transferred to the detectable magnetization and, thus, falls to a non-productive pathway highlighted in yellow in Fig. 5a. From the simulated relaxation profile it is clear that the main relaxation losses stem from the three coherence transfer INEPTS, so that the improved CTE [38,39] contributes to higher sensitivity by providing the reduced apparent relaxation rates. However, since the improved CTE ends with a ^{1}H 90° pulse, it requires additional delays for flipping water magnetization back to +*z* direction, which is usually combined with the echo-anti-echo PFG-based quadrature

encoding scheme resulting in overall longer pulse sequence. Since ST2-PT TROSY [37] does not require such extra delays, it effectively delivers similar signal-to-noise.

3.3. Sensitivity in nD experiments

In order to more systematically compare performance of widely used 2D $^{1}H,^{5}N$-correlation experiments, we calculated the performance measures and experimentally measured S/N^{2D} in series of 2D spectra using two test proteins, 25 kDa $^{13}C,^{15}N$-labeled αXβ2 Integrin I domain (1N3 V.PDB), and 51 kDa $^{2}H,^{13}C,^{15}N$-labeled NS3Helicase (2JLQ.PDB) from *Dengue* virus in a complex with 14 nucleotide single stranded RNA [40]. Fig. S5 (Supplementary information) shows that high quality spectra of these two proteins are obtained which are suitable for quantitative comparisons. The rotational correlation times of 20 and 40 ns for αX-I and NS3 proteins, respectively, were estimated using protein solvent accessible surface area as well as known 3D shapes [41].

Fig. 6 shows S/N^{2D} in series of 2D $^{1}H,^{15}N$-correlation spectra and compares it with predicted *S/N*. Overall, a good agreement between predicted and experimental *S/N* is observed with notable exception of HSQC-PEP-PFG experiment for both proteins [Tables S5a and S5b] (Supplementary information). We traced this underperformance of the latter experiment to saturation of water signal due to lack of water-flip-back [42] in our implementation. Although the aim of the performance annotation of RDB is to predict signal-to-noise in *n*D spectra from seeding lower dimensional spectra, here we chose to use the resolved 2D correlation spectra as the reference and back-predict S/N^{1D} in the corresponding seeding 1D spectra, $S/N^{1D} = 8.0$ and $S/N^{1D} = 2.0$ for αXβ2 Integrin I domain and NS3Helicase, respectively. These values closely correspond to the experimentally observed *S/N* in the seeding 1D ^{1}H spectra, albeit the direct comparison is hampered by the absence of well resolved individual protons in 1D spectrum.

We attempted to predict *S/N* in TROSY series of 2D and 3D experiments measured with both proteins using as input the corresponding rotational correlation times τ_c and S/N^{1D} in the seeding 1D ^{1}H spectra. The corresponding experimental parameters as well as calculated performance measures are listed in [Tables S6 and S7] (Supplementary information) for NS3Helicase and αXβ2 Integrin I domain, respectively. Comparison between predicted and experimental *S/N* in 2D TROSY and 3D HNCA and 3D HNCACB as well as in individual 2D [$^{1}H,^{13}C$] and [$^{1}H,^{15}N$] shows that the sensitivities can be reasonably predicted Fig. 7 thus demonstrating viability of the proposed benchmarking of NMR experiments.

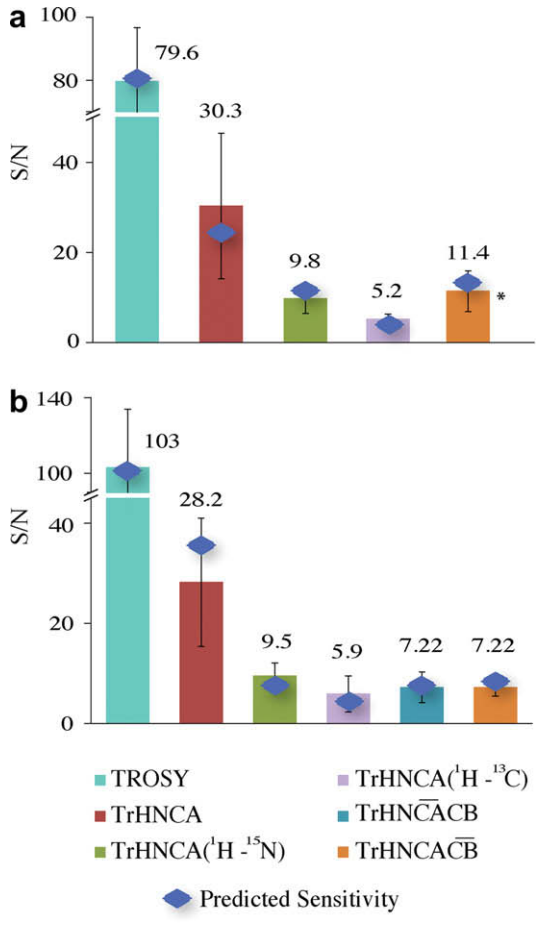

Fig. 7. Comparison of predicted and experimental *S/N* in 2D [$^{1}H,^{15}N$]-TROSY, 2D [$^{1}H,^{13}C$]- and [$^{1}H,^{15}N$]-planes of TROSY-HNCA, TROSY-HNCACAB, measured with (a) $^{2}H,^{13}C,^{15}N$-labeled 51 kDa NS3Helicase and (b) $^{13}C,^{15}N$-labeled 25 kDa αXβ2 Integrin I domain. A value above each bar represents the average *S/N* of those assigned peaks. The relevant experimental parameters are listed in [Tables S6 and S7] (Supplementary information). The predicted S/N2D are calculated with Eq. (9) using data of [Tables S6 and S7] and $S/N^{1D} = 12.5$ and $S/N^{1D} = 13.7$ for αXβ2 Integrin I domain and NS3Helicase, respectively. Note, that the used samples are different from those reported in Fig. 6 ($* T_{CACB} = 14.4$ ms; full magnetization transfer from C_α to C_β).

4. Conclusions

The relational database of NMR experiments and the Spinach library enable numerical modeling of NMR experiments on a large user-specified spin system representative of a protein. Directly comparable performance metrics can be computed for every sequence in the database, enabling straightforward comparison and performance benchmarking. The resulting software package is a complete tool for pulse sequence analysis, testing of new pulse sequences and optimization of the existing experiments. The system also serves as an educational tool for understanding basic principles of construction and performance of NMR experiments.

Acknowledgments

The authors thank Singapore Ministry of Education as well as Swiss National Science Foundation for financial support (MOE Tier-2 grant and SNF grant to Konstantin Pervushin). The authors thank Peter Hore, Matthew Krzystyniak, Hannah Hogben and Gareth Charnock for helpful discussions, and Dr. Zhao Xiao and Leo Wong for making spectra of αXβ2 Integrin and NS3Helicase available for us. This work was supported by EPSRC (EP/F065205/1, EP/H003789/1).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmr.2009.12.008.

References

[1] J. Cavanagh, W.J. Fairbrother, A.G. Palmer, N.J. Skelton, Protein NMR Spectroscopy: Principles and Practice, Academic Press, New York, 1996.
[2] A.E. Ferentz, G. Wagner, NMR spectroscopy: a multifaceted approach to macromolecular structure, Q. Rev. Biophys. 33 (2000) 29–65.
[3] D.P. Frueh, T. Ito, J.S. Li, G. Wagner, S.J. Glaser, N. Khaneja, Sensitivity enhancement in NMR of macromolecules by application of optimal control theory, J. Biomol. NMR 32 (2005) 23–30.
[4] T.E. Skinner, T.O. Reiss, B. Luy, N. Khaneja, S.J. Glaser, Application of optimal control theory to the design of broadband excitation pulses for high-resolution NMR, J. Magn. Reson. 163 (2003) 8–15.
[5] K. Bromek, D. Lee, R. Hauhart, M. Krych-Goldberg, J.P. Atkinson, P.N. Barlow, K. Pervushin, Polychromatic selective population inversion for TROSY experiments with large proteins, J. Am. Chem. Soc. 127 (2005) 405–411.

[6] N. Khaneja, T. Reiss, B. Luy, S.J. Glaser, Optimal control of spin dynamics in the presence of relaxation, J. Magn. Reson. 162 (2003) 311–319.
[7] N. Khaneja, B. Luy, S.J. Glaser, Boundary of quantum evolution under decoherence, Proc. Natl. Acad. Sci. USA 100 (2003) 13162–13166.
[8] J. Shriver, Concepts Magn. Reson. 4 (1992) 1–33.
[9] P. Guntert, N. Schaefer, G. Otting, K. Wuthrich, Poma – a complete Mathematica implementation of the NMR product-operator formalism, J. Magn. Reson. A 101 (1993) 103–105.
[10] K. Young, G.B. Matson, V. Govindaraju, A.A. Maudsley, Spectral simulations incorporating gradient coherence selection, J. Magn. Reson. 140 (1999) 146–152.
[11] A. Jerschow, MathNMR: spin and spatial tensor manipulations in Mathematica, J. Magn. Reson. 176 (2005) 7–14.
[12] F.S. Debouregas, J.S. Waugh, ANTIOPE, a program for computer experiments on spin dynamics, J. Magn. Reson. 96 (1992) 280–289.
[13] S.A. Smith, T.O. Levante, B.H. Meier, R.R. Ernst, Computer simulations in magnetic resonance. An object-oriented programming approach, J. Magn. Reson. A 106 (1994) 75–105.
[14] M. Bak, J.T. Rasmussen, N.C. Nielsen, SIMPSON: a general simulation program for solid-state NMR spectroscopy, J. Magn. Reson. 147 (2000) 296–330.
[15] T. Allman, A.D. Bain, J.R. Garbow, SIMPLTN, a program for the simulation of pulse NMR spectra, J. Magn. Reson. A 123 (1996) 26–31.
[16] P. Nicholas, D. Fushman, V. Ruchinsky, D. Cowburn, The virtual NMR spectrometer: a computer program for efficient simulation of NMR experiments involving pulsed field gradients, J. Magn. Reson. 145 (2000) 262–275.
[17] M. Helgstrand, P. Allard, QSim, a program for NMR simulations, J. Biomol. NMR 30 (2004) 71–80.
[18] M. Veshtort, R.G. Griffin, SPINEVOLUTION: a powerful tool for the simulation of solid and liquid state NMR experiments, J. Magn. Reson. 178 (2006) 248–282.
[19] W.B. Blanton, BlochLib: a fast NMR C++ toolkit, J. Magn. Reson. 162 (2003) 269–283.
[20] W. Kohn, Nobel Lecture: Electronic structure of matter; wave functions and density functionals, Rev. Mod. Phys. 71 (1999) 1253–1256.
[21] I. Kuprov, Polynomially scaling spin dynamics II: further state-space compression using Krylov subspace techniques and zero track elimination, J. Magn. Reson. 195 (2008) 45–51.
[22] I. Kuprov, N. Wagner-Rundell, P.J. Hore, Polynomially scaling spin dynamics simulation algorithm based on adaptive state-space restriction, J. Magn. Reson. 189 (2007) 241–250.
[23] P. Kumar, A. Kumar, Effect of remote cross correlations on transverse relaxation, J. Magn. Reson. A 119 (1996) 29–37.
[24] A. Kumar, P.K. Madhu, Cross-correlations in multispin relaxation, Concepts Magn. Reson. 8 (1996) 139–160.
[25] R.R. Ernst, G. Bodenhausen, A. Wokaun, The Principles of Nuclear Magnetic Resonance in One and Two Dimensions, Clarendon Press, Oxford, 1987.
[26] Z. Bai, Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems, Appl. Num. Math. 43 (2002) 9–44.
[27] R.W. Freund, Model reduction methods based on Krylov subspaces, Acta Numer. 12 (2003) 267–319.
[28] A.G. Redfield, On the theory of relaxation processes, IBM J. Res. Dev. 1 (1957) 19–31.
[29] M. Goldman, Formal theory of spin–lattice relaxation, J. Magn. Reson. 149 (2001) 160–187.
[30] D.J. Tannor, Introduction to Quantum Mechanics: A Time- Dependent Perspective, University Science Books, New York, 2006.
[31] S.J. Glaser, T. Schulte-Herbruggen, M. Sieveking, O. Schedletzky, N.C. Nielsen, O.W. Sorensen, C. Griesinger, Unitary control in quantum ensembles: maximizing signal intensity in coherent spectroscopy, Science 280 (1998) 421–424.
[32] K. Pervushin, B. Vogeli, T.N. Heinz, P.H. Hunenberger, Measuring H-1-H-1 and H-1-C-13 RDCs in methyl groups: example of pulse sequences with numerically optimized coherence transfer schemes, J. Magn. Reson. 172 (2005) 36–47.
[33] I. Kuprov, N. Wagner-Rundell, P.J. Hore, Bloch–Redfield–Wangsness theory engine implementation using symbolic processing software, J. Magn. Reson. 184 (2007) 196–206.
[34] V.F. Bystrov, Spin-spin coupling and the conformational states of peptide systems, Prog. NMR Spectr. 10 (1976) 41–82.
[35] H. Sun, L.K. Sanders, E. Oldfield, Carbon-13 NMR shielding in the twenty common amino acids: comparisons with experimental results in proteins, J. Am. Chem. Soc. 124 (2002) 5486–5495.
[36] K. Loth, P. Pelupessy, G. Bodenhausen, Chemical shift anisotropy tensors of carbonyl, nitrogen, and amide proton nuclei in proteins through cross-correlated relaxation in NMR spectroscopy, J. Am. Chem. Soc. 127 (2005) 6062–6068.
[37] K. Pervushin, G. Wider, K. Wuthrich, Single transition-to-single transition polarization transfer (ST2-PT) in [N-15, H-1]-TROSY, J. Biomol. NMR 12 (1998) 345–348.
[38] D.W. Yang, L.E. Kay, Improved (HN)-H-1-detected triple resonance TROSY-based experiments, J. Biomol. NMR 13 (1999) 3–10.
[39] D. Nietlispach, Suppression of anti-TROSY lines in a sensitivity enhanced gradient selection TROSY scheme, J. Biomol. NMR 31 (2005) 161–166.
[40] T. Xu, A. Sampath, A. Chao, D. Wen, M. Nanao, P. Chene, S.G. Vasudevan, J. Lescar, Structure of the Dengue virus helicase/nucleoside triphosphatase catalytic domain at a resolution of 2.4 A, J. Virol. 79 (2005) 10278–10288.
[41] V.V. Krishnan, M. Cosman, An empirical relationship between rotational correlation time and solvent accessible surface area, J. Biomol. NMR 12 (1998) 177–182.
[42] M. Piotto, V. Saudek, V. Sklenar, Gradient-tailored excitation for single-quantum NMR- spectroscopy of aqueous-solutions, J. Biomol. NMR 2 (1992) 661–665.
[43] L. Muller, Sensitivity enhanced detection of weak nuclei using heteronuclear multiple quantum coherence, J. Am. Chem. Soc. 101 (1979) 4481–4484.
[44] G. Bodenhausen, D.J. Ruben, Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy, Chem. Phys. Lett. 69 (1980) 185–189.
[45] J. Cavanagh, A.G. Palmer, P.E. Wright, M. Rance, Sensitivity improvement in proton-detected 2-dimensional heteronuclear relay spectroscopy, J. Magn. Reson. 91 (1991) 429–436.
[46] A.G. Palmer, J. Cavanagh, P.E. Wright, M. Rance, Sensitivity improvement in proton-detected 2-dimensional heteronuclear correlation NMR-spectroscopy, J. Magn. Reson. 93 (1991) 151–170.
[47] L.E. Kay, P. Keifer, T. Saarinen, Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity, J. Am. Chem. Soc. 114 (1992) 10663–10665.