SPINS: A laboratory information management system for organizing and archiving intermediate and final results from NMR protein structure determinations

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · MARCH 2006

Impact Factor: 2.63 · DOI: 10.1002/prot.20840 · Source: PubMed

7 AUTHORS, INCLUDING:

James M Aramini
City University of New York - Advanced Scie…
99 PUBLICATIONS **2,594** CITATIONS
SEE PROFILE

Daniel Monleón
Fundación de Investigación del Hospital Clí…
82 PUBLICATIONS **1,280** CITATIONS
SEE PROFILE

Giovanni T Monteleone
University of Rome Tor Vergata
254 PUBLICATIONS **7,594** CITATIONS
SEE PROFILE

SPINS: A Laboratory Information Management System for Organizing and Archiving Intermediate and Final Results from NMR Protein Structure Determinations

Michael C. Baran,[1] Hunter N. B. Moseley,[1] James M. Aramini,[1] Marvin J. Bayro,[1] Daniel Monleon,[1] Jessica Y. Locke,[1] and Gaetano T. Montelione[1,2,*]

[1]*Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, Rutgers University, and Northeast Structural Genomics Consortium, Piscataway, New Jersey*

[2]*Department of Biochemistry and Molecular Biology, Robert Wood Johnson Medical School, Piscataway, New Jersey*

ABSTRACT **Recent technological advances and experimental techniques have contributed to an increasing number and size of NMR datasets. In order to scale up productivity, laboratory information management systems for handling these extensive data need to be designed and implemented. The SPINS (Standardized ProteIn Nmr Storage) Laboratory Information Management System (LIMS) addresses these needs by providing an interface for archival of complete protein NMR structure determinations, together with functionality for depositing these data to the public BioMagResBank (BMRB). The software tracks intermediate files during each step of an NMR structure-determination process, including: data collection, data processing, resonance assignments, resonance assignment validation, structure calculation, and structure validation. The underlying SPINS data dictionary allows for the integration of various third party NMR data processing and analysis software, enabling users to launch programs they are accustomed to using for each step of the structure determination process directly out of the SPINS user interface. Proteins 2006;62:843–851.** © 2006 Wiley-Liss, Inc.

Key words: NMR spectra archive; structural genomics; integrated analysis process

INTRODUCTION

The average NMR structure determination produces roughly 30 to 50 gigabytes of data, generated by a wide range of NMR data analysis software. The process of moving data from one application to another, and archiving of intermediate and final results, presents an enormous organizational challenge. The BioMagResBank[1] (BMRB) and Protein Data Bank[2] (PDB) provide international repositories for the final results of the assignment and structure determination process. These databases are crucial for distributing structural data to the rest of the world. However, in order for the information to flow efficiently from the structural biology pipelines into the BMRB and PDB, adequate data management systems need to be implemented within structural biology laboratories.

Protein NMR spectroscopists depend on a number of software packages to facilitate the analysis of data. For this reason, the computational challenge of solving a protein structure by NMR presents a formidable technical challenge. While a number of software packages have been developed for the analysis of NMR data, a comprehensive solution for the complete automated analysis of NMR data from raw free-induction decay (FID) to structure has yet to be described. Users are forced to choose between a number of different software programs each specializing in a certain step of the structural determination process. As a result, a dramatic learning curve has emerged in which a true expert must be proficient with a number of different pieces of software in order to do his or her job. Furthermore, valuable time is often wasted on trivial tasks such as preparing the output of one program to be usable for the next. Also, inter- and, in some cases, even intralaboratory data exchange becomes extremely difficult when people are using a number of different formats offered by the various pieces of software available. To add to this complexity, with data passing between so many sources, organization quickly becomes a problem. Precious data is often lost due to disorganization. This disorganization can curb the development of future technologies by leading to irreproducible results.

Many of these problems have been addressed by the initial version of the SPINS[3] (Standardized ProteIn Nmr Storage) database which provides a Laboratory Informa-

The Supplementary Material referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat

Abbreviations: FID, free induction decay; GUI, graphical user interface; LIMS, laboratory information management system; NOESY, nuclear Overhauser effect spectroscopy; SPINS, Standardized Protein NMR Storage; SPINSds, SPINS directory structure; XML, eXtensible Markup Language.

Grant sponsor: Protein Structure Initiative of National Institutes of Health; Grant number: P50 GM62413; Grant sponsor: The New Jersey Commission on Science and Technology; Grant number: 99-2042-007-13.

*Correspondence to: Gaetano T. Montelione, CABM-Rutgers University, 679 Hoes Lane, Piscataway, NJ 08854. E-mail: guy@cabm.rutgers.edu

Received 1 June 2005; Revised 20 September 2005; Accepted 21 September 2005

Published online 4 January 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20840

tion Management System (LIMS) for organizing time domain NMR data. Other projects that provide data management of NMR data include the Sesame Project,[4] which has been developed for organizing the massive amounts of data generated by high throughput structure determination, and the CCPN project,[5–7] which has developed a detailed data model to describe the NMR structure determination process to be used as an application programming interface (API) to unify the development and integration of future NMR software. While both of these projects address the need to organize large-scale data production, they lack a mature methodology for deposition of these data into the public domain.

The SPINS software is a LIMS aimed at organizing NMR structure determination projects and seamlessly exporting that information to the BMRB. The software tracks intermediate files during each step of the structure determination process including: data collection, data processing, resonance assignments, and structure calculation. We have also prototyped the process of integrating various third-party NMR data analysis software with the SPINS database application. This integrated suite of NMR data analysis software provides a robust solution for the determination of protein structures from NMR data.

MATERIALS AND METHODS

Data Model

The original SPINS data model was designed to capture all the information necessary to completely re-create an NMR experiment.[3] As described in Figure 1, the SPINS data model has been expanded to further track all the information necessary to completely reproduce the structure determination process from FIDs to coordinates. A complete summary of the SPINS data model is available both as Supplementary Material to this manuscript (Supplemental Table I) and from our laboratory website (http://www-nmr.cabm.rutgers.edu).

Previously in SPINS,[3] the *Experiment* table had served as the central table of the database. This relationship is sufficient when describing data processing and peak picking which exist on a one-to-one basis with an experiment. However, resonance assignments and structure calculations are based on the analysis of multiple experiments. It is therefore necessary to introduce a higher level of logic to describe the overall process. The concept of a *Project* allows for the inclusion of multiple experiments and thus the complete description of resonance assignments and structure calculations.

One of the important goals of our data model is to allow SPINS to automatically generate complete BMRB NMR-STAR 3.0 files directly from the SPINS database. The BMRB[1] is the international repository of protein NMR data, and it is critical that any NMR database project coordinate closely with the data model in use by this community archive. To ensure smooth integration of SPINS data into NMR-STAR formatted files, we have designed the internal SPINS data model to be as consistent as possible with the new NMR-STAR v3.0 data model currently used by the BMRB.[1] While minor differences do exist between the two data models, there is a unique and complete mapping of the SPINS data model to NMR-STAR format. Efforts are made to maintain this compatibility and consistency as these related NMR-STAR and SPINS data models evolve.

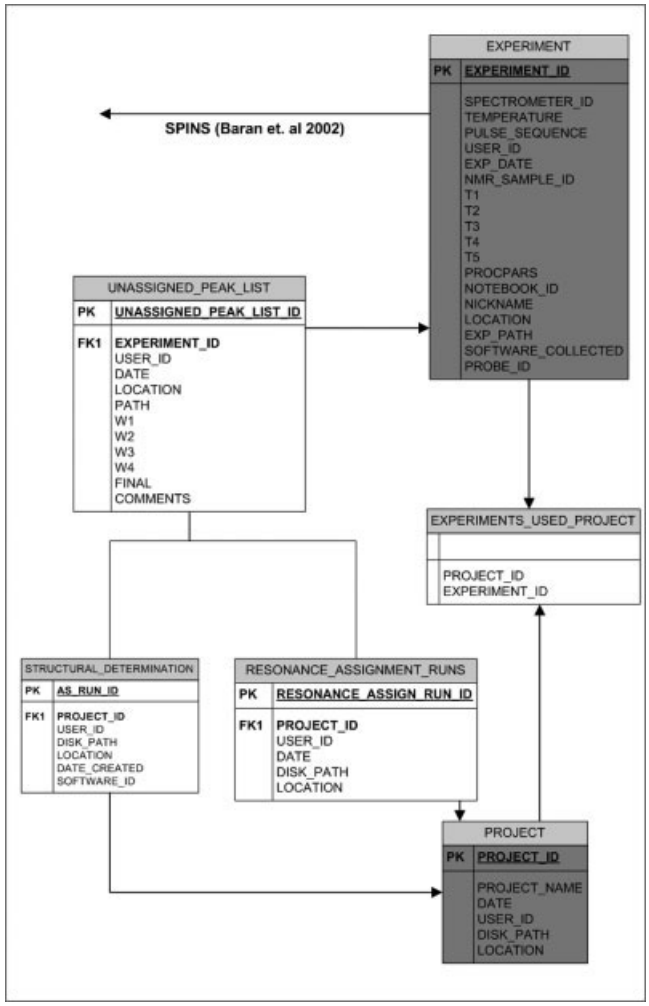

Fig. 1. SPINS Entity Relationship Diagram. A schematic view of the SPINS tables and their relationships. The *Experiment* and *Project* tables are shaded darker because they represent central tables of the data model. The *Experiment* table contains metadata describing the experiment as well as links to other tables which describe the experiment. Spectral peak lists are linked to experiments through the *Unassigned_Peak_list* table. Resonance assignments and structures are described by multiple experiments. Because this creates a one-to-many relationship, the *Experiment* and *Project* tables are linked by the *Experiments_used_Project* table. The *Project* table contains basic data concerning the project, its primary key is used to link the *Resonance_Assignment_Runs* and *Strucutral_Determinations* tables with the *Project* table. Details of the SPINS data model are presented in the Supplemental Material tables.

Overview of System Functionality

The SPINS database is installed on an Oracle 9i database running on a Linux platform. After installation of the SPINS software it is first necessary to populate the database with NMR experiments as previously described.[3] This includes populating the auxiliary tables necessary to

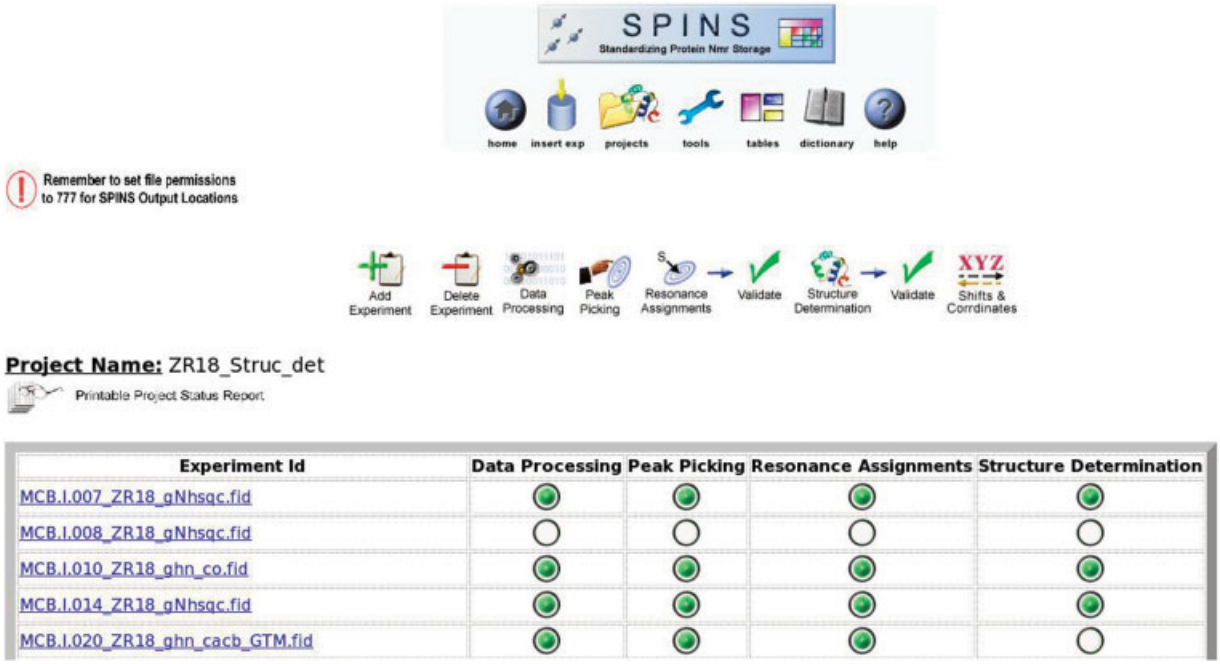

Fig. 2. SPINS Graphical User Interface. This figure shows the SPINS Project Management view from the interface. At the top of the page are two navigation menu bars. The top bar is the standard SPINS menu, providing links to the home page, experiment insertion tool, SPINS Tools, SPINS Tables, the Data Dictionary, and online help. Under the SPINS menu bar is the Project management menu bar. From this bar users can add or remove experiments from the project as well as launch tools for data processing/referencing, peak picking, automated resonance assignments, automated structure calculation, and validation of the complete process.

describe an experiment, such as sample and spectrometer specific data. Once experimental data has been inserted into SPINS one can create a *Project*, and associate experiments, resonance assignment runs, and structure calculation runs with that *Project*.

The Project Management page is the main interface for managing a NMR structure determination project (Fig. 2). The page provides a summary of all experiments currently inserted into the SPINS database for that project, as well as functionality to retrieve and insert additional data. Next to each experiment are columns corresponding to each step of the structure determination process. (Processing, Peak Picking, etc.) When data from one of these steps has been archived, a green circle will light up for that experiment under the corresponding field. For example, once data processing is complete and the corresponding processing scripts have been archived, a green light will light up next to that experiment under the data processing column. This provides a simple graphical way to track the progress of *Project*, as well as an easy way to retrieve data from the database.

Integrated Data Analysis Platform

In addition to its applications for NMR data archival, we have prototyped SPINS to run as an integrated NMR data analysis platform. To achieve this functionality SPINS integrates a number of third-party pieces of software corresponding to each step described above and presents them in a transparent manner so that they appear as a single application to the user, as summarized in Figure 3. SPINS is then be capable of launching that software from its interface, and archiving the input and output files generated.

In our laboratory, the SPINS software has been configured with the following programs: (i) the SPINS[3] database is used for storage and organization of raw FIDs. (ii) AGNUS,[25] a NMRPipe[9] script generating program utilized by SPINS to reference and process one-dimensional (1D), two-dimensional (2D), and three-dimensional (3D) NMR spectra in a near automated fashion. Processing results are then presented to the user for inspection and interactive phasing, if necessary, using NMRDraw[9] spectral visualization software. (iii) SPARKY[10] spectral visualization software for interactive spectral data analysis and peak picking. (iv) The automated NMR resonance assignment software AutoAssign[11–14] for determining backbone resonance assignments. (v) Assignment Validation Software (AVS)[8] suite, used to validate the quality of the assignments. (vi) The automated structure determination software AutoStructure[15–17] for iterative NOESY assignment and 3D structure genearation using the software DYANA,[18] NIH-XPLOR,[19,20] and/or CNS.[21] (vi) Protein Structure Validation Software suite (PSVS) and AutoQF[22] software, used to validate the structure quality.[23] These programs can be run out of the SPINS interface, and intermediate data files generated by these software are archived in SPINS throughout the analysis process.

The SPINS software provides a user-friendly approach toward using the software packages described above without having to worry about the numerous I/O complexities

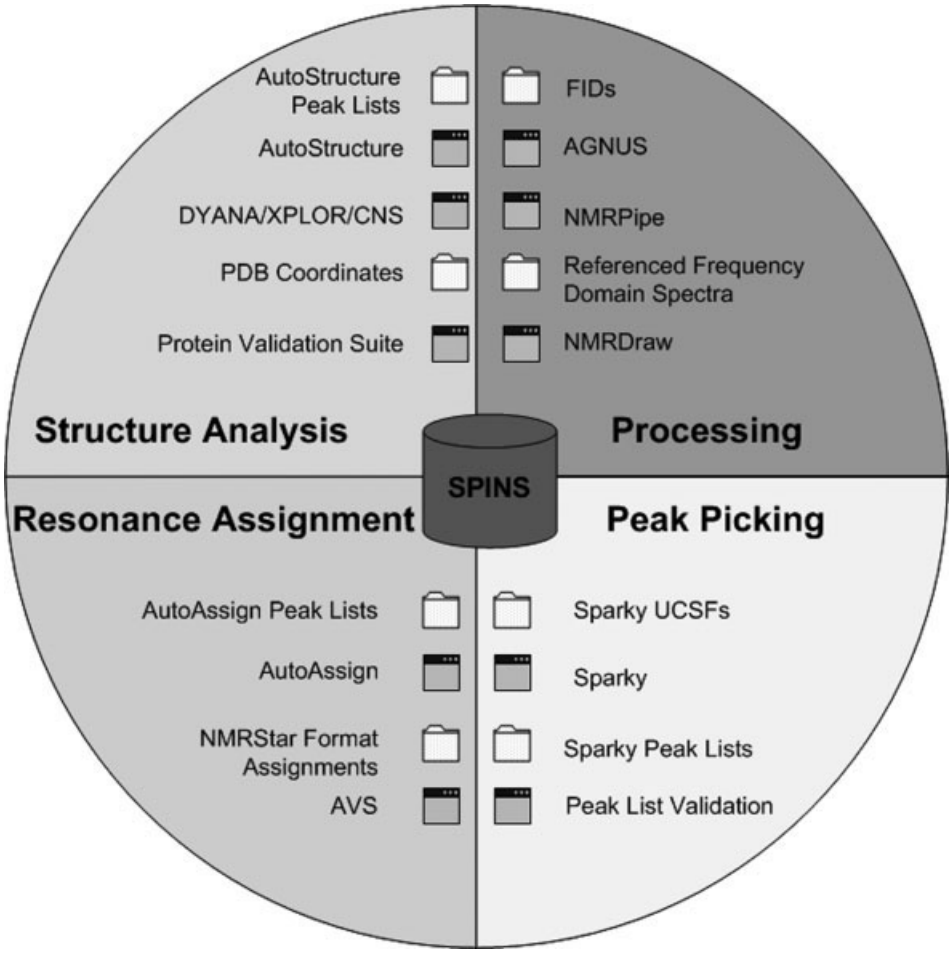

Fig. 3. SPINS Software Data Flow. This figure depicts the flow of data through the SPINS software from raw FIDs to backbone assignments. (**i**) The raw FIDs are housed in the SPINS database. (**ii**) AGNUS queries the SPINS database for autoreferencing and processing of experimental data. (**iii**) SPARKY is used for manual peak picking and peak list editing. (**iv**) AutoAssign software used for automated backbone resonance assignments. (**v**) Assignments Validation Software (AVS) suite is used to validate the reliability of backbone assignments. (**vi**) AutoStructure is run for iterative NOESY assignment and structure calculation. (**vii**) AutoQF and the Protein Structure Validation Suite are used to evaluate structure quality. In this process, side-chain resonance assignments are analyzed manually using the SPARKY interactive spectral analysis software.

associated with data analysis using multiple software packages. Furthermore, the process is completely reproducible and warehoused by the underlying SPINS database. The resulting intermediate and final files, along with the time domain raw NMR data, can be automatically exported into a NMR-STAR 3.0[24] file.

Project Archival Tool

In addition to raw experimental data, spectroscopists will typically generate processing scripts, peak lists, resonance assignments, and atomic coordinates during a protein structure determination. While SPINS does provide tools for manual entry of this data it can be very time consuming to archive a complete data set using the manual entry tools. To facilitate archival of NMR structure determination data, SPINS also provides a Project Archival Tool. This tool is designed to archive all files used to generate the protein structure from one easy-to-use automated interface [see Supplementary Material Fig. 1(A)]. Users are first asked to name their project and identify the directory where their NMR experiments are located. SPINS will then parse through that directory identifying all NMR experiments and display back to the user a list of experiments found, and ask for some supplementary information regarding those experiments, such as the sample id and spectrometer id [see Supplementary Material Fig. 1(B)]. After filling in the required information and pressing submit, SPINS will automatically identify and archive processing scripts and peak lists from the NMR experiment directories. User supplied resonance assignments and structure calculation data can also be archived with these tools. Upon completion of archival, all archived data can then be accessed using the Project Management interface described above.

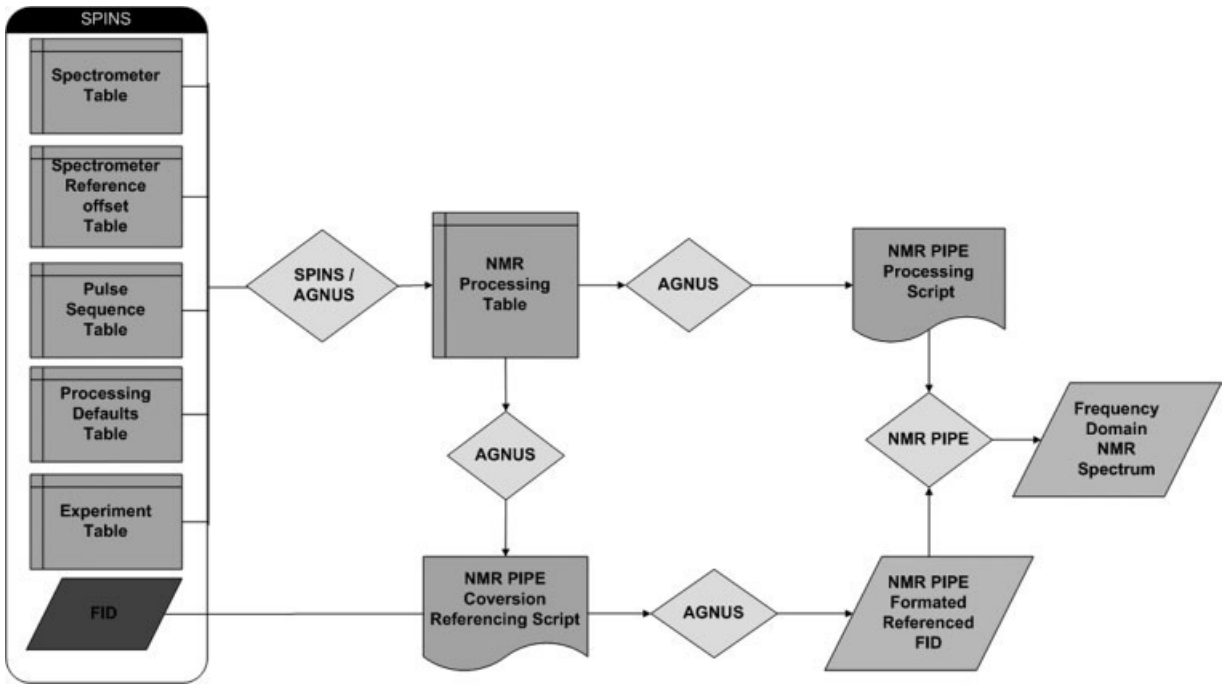

Fig. 4. AGNUS Data Flow. A number of the parameters required for the referencing and processing of NMR data are relatively sample independent. These parameters are captured by SPINS in the *Spectrometer*, *Spectrometer Reference Offset*, *Pulse Sequence*, *Processing Defaults*, *Processing*, and *Experiment* tables. These data are then used by AGNUS and SPINS to generate the NMR Processing Table in memory. All the information required to process and reference an experiment is stored within this data structure. The NMR Processing Table is then used by AGNUS to generate a NMR PIPE Conversion Referencing Script as well as a NMR PIPE Processing Script. The NMR PIPE Conversion Referencing Script is then used by AGNUS to convert the FID into an NMR PIPE formatted FID. This NMR PIPE formatted FID along with the previously generated NMR PIPE Processing Script is finally used as input into NMR PIPE to generate the final output of Frequency Domain NMR Spectra.

BMRB Deposition

The new BMRB NMR-STAR 3.0[24] format provides a complete description of the NMR structure determination process. Due to its size and complexity, manual preparation of a complete NMR-STAR 3.0 file is a very tedious process. Automatic generation of a complete NMR-STAR 3.0 file requires a local NMR LIMS system capable of archiving and organizing the wealth of data generated during an NMR structure determination. Using the previously described Project Archival Tool ensures that a complete representation of the structure determination process has been archived, thus making the SPINS LIMS an ideal candidate to provide automatic generation of complete NMR-STAR 3.0 files. SPINS provides an easy to use tool for automatically generating complete BMRB NMR-STAR 3.0 files. Users select the project for which they would like to generate a BMRB deposition, complete a few supplemental data fields and within seconds a complete NMR-STAR 3.0 file is generated.

In addition to generating BMRB entries, SPINS provides an integrated tool for uploading raw time-domain NMR data to the BMRB ftp server. After generating and validating the NMR-STAR file users can use SPINS to upload FIDs from the project as well the SPINS-generated NMR-STAR file to the BMRB. This process is all done automatically by having SPINS connect to the BMRB ftp server and transfer the data. Upon successful completion the user and the BMRB are contacted by e-mail.

NMR Data Processing Software Integration

AGNUS,[25] described in Figure 4, is software for referencing NMR spectra and generating macros for apodization and Fourier transformation using macro-based spectral processing software. It currently is implemented to provide data conversion and processing scripts for the software package NMRPipe.[9] AGNUS takes as input Varian NMR or Bruker BioSpin format FIDs, together with libraries of spectrometer and pulse-sequence specific table files. The software references the data in the direct and indirect dimensions using spectrometer specific parameters stored in the *Spectrometer* table and creates scripts suitable for subsequent processing with NMRPipe[9] using default and/or user-provided processing parameters. The ^{1}H frequency offset value corresponding to 0.0 ppm at 25°C, and its temperature dependence, is measured on the methyl resonance of a standard sample of aqueous (DSS) 2,2-dimethyl-2-silapentane-5-sulfonic acid (10 m*M*, 99.9% D_2O). Offset values and their temperature dependence for additional nuclei are then calculated using IUPAC-recommended chemical shift referencing protocols and ratios of gyromagnetic ratios[26]; for example, $\gamma_{N/H}$ = 0.101329118 and $\gamma_{C/H}$ = 0.251449530. A detailed protocol for determining these frequency offset values and their temperature dependence is presented on our laboratory web site at http://www-nmr.cabm.rutgers.edu/labdocuments/nmrprotocols/offset_referencing.html. The AGNUS

software runs as a stand-alone command line application or a tightly integrated component of the SPINS software.

The AGNUS software can be launched directly from the SPINS Project Management Page. After selecting Data Processing from Project Management Menu bar, users are asked to select an experiment from the current project to process. The user is first asked to confirm some parameters that SPINS has extracted from the database such as the spectrometer, temperature, output location, etc. The user is next asked to confirm the AGNUS calculated conversion parameters as well as number of processing parameters such as window function, zero-filling, and phasing parameters in each spectral dimension [Fig. 5(A)]. As the user works through the GUI forms, sample-independent processing parameters are pre-filled by the AGNUS software with suggested values corresponding to the selected experiment. These parameters are expert defined suggestions stored in the SPINS *Processing_Defaults* table, which have been shown to work in the past. The user is given the ability to adjust these parameters as required. Phasing is handled interactively by launching the NMRDraw[9] software directly out of the web browser–based SPINS interface. Upon completion the output is written to the user defined output location and the processing scripts are archived in the SPINS database.

Spectral Visualization Software Integration

SPARKY[10] is a commonly used, spectral visualization software program. In our laboratory, after data processing with AGNUS and NmrPipe, the referenced formatted frequency domain spectra (.pipe or.ft2 file) are converted to the SPARKY format (.ucsf) and peak picked. Using the Project Management page of the SPINS interface the user can launch SPARKY and archive the resulting peak list in the SPINS database. Peak lists are stored internally in XML format. Data format conversions are automatically handled by applying an XSLT transformation to the data. XML technology simplifies data exchange and provides flexibility for future support of various software programs. Currently SPINS supports peak list conversion from XML to the SPARKY,[10] AutoAssign,[11–14] and AutoStructure[15,17] peak list formats. However, support for other peak list formats could easily be adapted via XSLT style sheets.

Backbone Resonance Assignment Software Integration

AutoAssign[11–14] is software for the automated analysis of backbone resonance assignments using triple resonance NMR experiments. SPINS automatically generates all AutoAssign input and then launches and executes the program to make largely automated backbone resonance assignment a routine process [Fig. 5(B)]. In order for AutoAssign to run optimally, high-quality input data is required. The input peak lists must be properly referenced, and registered with respect to one another to minimize interspectral differences in chemical shift. The manual preparation of this data can be demanding. However, based on data extracted from the SPINS database as well as statistics provided by peak list analysis, SPINS can automatically generate properly formatted AutoAssign peak lists, the AutoAssign control file, as well as an assignment constraints file. Prior to executing the AutoAssign algorithm, the user is asked to validate the auto generated input data. The results are then analyzed using the Assignment Validation Suite (AVS)[8] software and saved as Connectivity Map (Cmap) image files. The Cmap describes all inter- and intraconnectivities and can be customized to include sequential NOE, chemical shift index (CSI),[27] amide hydrogen exchange, and scalar coupling data. The current version of the AutoAssign program provides only Cβ, Hβ, and some methyl side-chain resonance assignments; the balance of side-chain assignments are determined by interactive manual analysis using SPARKY.

Assignment Validation Software Integration

SPINS applies the Assignment Validation Suite (AVS) algorithms to highlight the problematic areas of the resonance assignment.[8] These results are then saved as a Connectivity Map (Cmap) image. SPINS runs the AVS[8] assignment validation algorithm and compares the assigned chemical shifts to the BMRB average chemical shifts to identify possible incorrectly assigned residues. Finally, the user is presented with the final Cmap image which graphically illustrates the entire assignment process [Fig. 5(B)].

Structure Calculation Software Integration

AutoStructure[15,17] is software for the iterative automated assignment of NOESY cross peaks throughout cycles of structure calculation using the structure calculation software DYANA,[18] nih-XPLOR,[19,20] and/or CNS.[21] SPINS automatically generates all AutoStructure input and launches its graphical user interface, giving users full access to the AutoStructure software [Fig. 5(C)]. This input includes properly formatted peak lists and the control file used to execute the AutoStructure software. When satisfied with their results users can then archive the AutoStructure input and output files in the SPINS database.

NMR R Factors

Using the AutoStructure software allows one to also execute the AutoQF[22] software for a quantitative NMR structure quality score. AutoQF calculates an F-measure quality factor derived from Recall and Precision scores commonly used in data retrieval systems. Recall measures the percentage of peaks in the NOESY peak lists consistent with the resonance assignments that are also consistent with the average interproton distances of the query structures. Precision measures the percentage of close distance proton pairs in the query structures whose back-calculated NOE interactions are detected in the NMR data. Both Recall and Precision quantify how well the 3D model structures agree with resonance assignment and NOESY cross peak data.

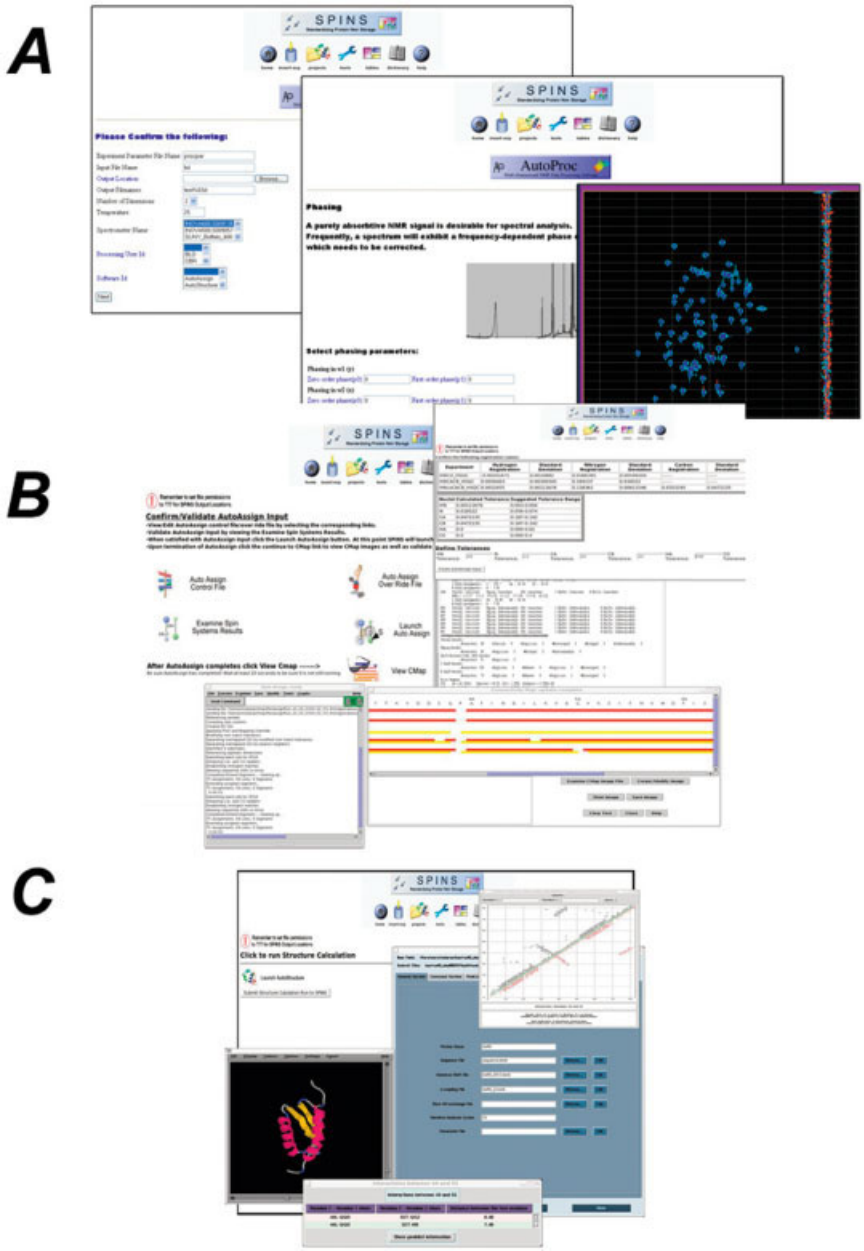

Fig. 5. SPINS NMR Data Analysis Integration. (**A**) AGNUS features an easy-to-use web-based interface. The software can be run as a stand-alone version, or as an integrated component of the SPINS database software. The user navigates through a series of forms containing default processing parameters which have been shown to work in the past for a selected experiment. (**B**) SPINS can also be used to generate the input and execute AutoAssign. This includes converting peak lists to AutoAssign format, registering the peak lists, generation of the AutoAssign control file, and generation of an AutoAssign assignment constraints file. After executing AutoAssign from the SPINS software the users is presented with a Connectivity Map (Cmap) summarizing the results of the automated assignment run. The Cmap highlights areas of sequence degeneracy where the AutoAssign algorithm may have had trouble. Furthermore, the *validate_assignments* program of the Assignment Validation Software suite is executed to highlight possible misassigned residues. (**C**) After completing all resonance assignments the input for AutoStructure can be generated by SPINS. The AutoStructure GUI is then launched directly out of the SPINS interface. Through the interface the autogenerated input can be modified by the user and structure calculations can be sent in batch to a multiprocessor CPU cluster.

Structure Validation Software Integration

In addition to AutoQF, our laboratory has also developed the Protein Structure Validation Software suite which can be used for both NMR and X-ray structures.[28] This software integrates a number of other software packages that evaluate structural quality based on a number of parameters including fold and packing quality, deviations of bond lengths and bond angles from standard values, backbone

and side-chain dihedral angle distributions, hydrogen-bond geometry, and close contacts between atoms. This approach allows users to quantitatively evaluate structure quality best on a consensus of quality scores. The current implementation of SPINS is configured to also launch the structure validation suite for analysis of protein structure quality.

Additional SPINS Tools

As previously described, SPINS includes an NMR-STAR v3.0 file generator. In the future, we plan on importing this file into a stand-alone version of NMR-ADIT for final manual editing prior to deposition. Furthermore, in addition to the tools previously described,[3] SPINS provides tools for peak list viewing in multiple formats, a JCoupling Calculator which automatically calculates $^{3}J(H^{N}\text{-}H^{\alpha})$ scalar coupling data from an assigned HNHA[29] peak list, and an AutoPaper tool which generates semi-complete Structure Notes formatted for rapid publication in *Proteins: Structure, Function, Bioinformatics*.

Documentation and Availability of Software

A 77-page PDF version of the documentation is included in the SPINS distribution which is also accessible through the software. The SPINS LIMS software described here requires the installation and configuration of an underlying Oracle database, Sun Java 1.4, and the Jakarta Tomcat java application server. SPINS has been tested on the Linux RedHat 7.3 operating system and requires a Mozilla 1.4 or higher web browser and java plugin. It is recommended that the computer network is Network File System (NFS) enabled.

RESULTS AND DISCUSSION

The SPINS database has been online within our laboratory since 2002. It currently contains approximately 3000 NMR experiments and serves as the main archive for our laboratory's NMR data, as well as HSQC screening data for hundreds of NESG protein targets. While the initial version of SPINS was designed to archive and manage raw NMR data, the expanded data model and software described in this article aims to capture the complete NMR protein structure determination process as well as disseminate that data to the BMRB. There are currently seven complete data sets archived in SPINS using the expanded SPINS data model (pdb: 1pqx, 1xhj, 1sqr, 1xhs, 1n91, 1ny4, 1yez). These data sets contain FIDs, processing scripts, spectral peak lists, and intermediate and final resonance assignment and structure calculation analysis data. The SPINS software has been used in the automatic generation of BMRB deposition files and/or the transfer of time domain data sets for 12 proteins (bmrbid: 4968, 5359, 5307, 5596, 5656, 5691, 5844, 6173, 5596, 6448, 6533, 6505).

In addition to capturing the data produced during a structure determination process, we have prototyped integrating various third-party NMR data analysis software with the SPINS application. The aim is to reduce the user learning curve by removing the complexities of data format conversion and program execution from the user. Recently, the CCPN[6,7] project has demonstrated the feasibility and value of this strategy by integrating a suite of NMR data analysis software under the CCPN data model. The CCPN project provides an extensive data model for LIMS implementation. It is not, however, an integrated LIMS system. Although the SPINS and CCPN data models have evolved differently, they share common data items as they are both BMRB NMR-STAR 3.0-compatible. In particular, final and intermediate NMR data files generated with CCPN-compliant software can be archived within the SPINS LIMS system, and used in programs launched from SPINS.

To validate our approach, the Northeast Structure Genomics Consortium (http://www.nesg.org) target SeR8 (80 residues PDBID: 1xhj) was analyzed using the SPINS software as described in this article. All experimental, intermediate, and final results were stored in the SPINS database. The time domain fids were processed through the SPINS/AGNUS interface using the NMRPipe[9] software. Using these methods, all NMR spectra were processed on the fly, directly as they came off the spectrometer. By utilizing the AGNUS software as well as the parallel processing capabilities of the NMRPipe[9] software, time spent on spectral processing was greatly reduced. As a result, complete backbone resonance assignments were made within hours of completing the final data acquisition. Side-chain analysis was done manually and structure calculations were run using the AutoStructure 1.1.2,[16,17] interfaced with XPLOR-nih-2.9.7.[19,20] Last, the BMRB NMR-STAR 3.0 deposition was generated automatically using the SPINS software (Accession 6355).

CONCLUSION

The fusion of informatics, biophysics, and biology will be key in the progression of modern science. The current era of science is generating large amounts of data that are proving difficult to manage by traditional methods. The SPINS Laboratory Information Management System addresses these problems in protein NMR data analysis by providing an archive of the complete assignment and structure determination process. It provides a path to archive intermediate and final files, as well as raw FID data, in the public domain BMRB database. We have also demonstrated the integration of third-party data analysis software with the SPINS database application. While other data models, such as the BMRB and CCPN data models, support a more detailed description of the resonance assignments and structure calculations, SPINS aims to archive the input/output data used by other software to analyze resonance assignments and 3D structures. In the future, we plan to move some of the process engineering/integration advancements learned from the SPINS project into graphical interfaces for in-house and third-party data analysis software. This will allow streamlining the current SPINS LIMS into a lightweight stand-alone client application to be used primarily for its most unique functions of data management, local data archival,

and automated preparation of public domain data depositions.

ACKNOWLEDGMENTS

We thank J. Everett, P. Rossi, T. Szyperski, G. V. T. Swapna, Y. P. Huang, along with J. Markley and E. Ulrich of the BioMagResBank for helpful discussions, suggestions, and advise.

REFERENCES

1. Seavey BR, Farr EA, Westler WM, Markley JL. A relational database for sequence-specific protein NMR data. J Biomol NMR 1991;1:217–236.
2. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
3. Baran MC, Moseley HN, Sahota G, Montelione GT. SPINS: standardized protein NMR storage. A data dictionary and object-oriented relational database for archiving protein NMR spectra. J Biomol NMR 2002;24:113–121.
4. Zolnai Z, Lee PT, Li J, Chapman MR, Newman CS, Phillips GN, Jr., Rayment I, Ulrich EL, Volkman BF, Markley JL. Project management system for structural and functional proteomics: Sesame. J Struct Funct Genomics 2003;4:11–23.
5. Fogh R, Ionides J, Ulrich E, Boucher W, Vranken W, Linge JP, Habeck M, Rieping W, Bhat TN, Westbrook J, Henrick K, Gilliland G, Berman H, Thornton J, Nilges M, Markley J, Laue E. The CCPN project: an interim report on a data model for the NMR community. Nat Struct Biol 2002;9:416–418.
6. Fogh RH, Boucher W, Vranken WF, Pajon A, Stevens TJ, Bhat TN, Westbrook J, Ionides JM, Laue ED. A framework for scientific data modeling and automated software development. Bioinformatics 2005;21:1678–1684.
7. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED. The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 2005;59:687–696.
8. Moseley HN, Sahota G, Montelione GT. Assignment validation software suite for the evaluation and presentation of protein resonance assignment data. J Biomol NMR 2004;28:341–355.
9. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 1995;6:277–293.
10. Goddard TD, Kneller DG. SPARKY 3. San Francisco: University of California; 2000.
11. Zimmerman DE, Montelione GT. Automated analysis of nuclear magnetic resonance assignments for proteins. Curr Opin Struct Biol 1995;5:664–673.
12. Zimmerman DE, Kulikowski CA, Huang Y, Feng W, Tashiro M, Shimotakahara S, Chien C, Powers R, Montelione GT. Automated analysis of protein NMR assignments using methods from artificial intelligence. J Mol Biol 1997;269:592–610.
13. Moseley HN, Montelione GT. Automated analysis of NMR assignments and structures for proteins. Curr Opin Struct Biol 1999;9:635–642.
14. Moseley HN, Monleon D, Montelione GT. Automatic determination of protein backbone resonance assignments from triple resonance nuclear magnetic resonance data. Methods Enzymol 2001;339:91–108.
15. Huang YJ. Automated determination of protein structures from NMR data by iterative analysis of self-consistent contact patterns. New Brunswick, NJ: Rutgers University; 2001.
16. Huang YJ, Swapna GV, Rajan PK, Ke H, Xia B, Shukla K, Inouye M, Montelione GT. Solution NMR structure of ribosome-binding factor A (RbfA), a cold-shock adaptation protein from *Escherichia coli*. J Mol Biol 2003;327:521–536.
17. Huang YJ, Moseley HN, Baran MC, Arrowsmith C, Powers R, Tejero R, Szyperski T, Montelione GT. An integrated platform for automated analysis of protein NMR structures. Methods Enzymol 2005;394:111–141.
18. Güntert P, Mumenthaler C, Wüthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 1997;273:283–298.
19. Brünger AT. X-PLOR, Version 3.1: a system for X-ray crystallography and NMR. New Haven: Yale University Press; 1992. xvii, 382 p.
20. Schwieters CD, Kuszewski JJ, Tjandra N, Clore MG. The Xplor-NIH NMR molecular structure determination package. J Magn Reson 2003;160:65–73.
21. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54:905–921.
22. Huang YJ, Powers R, Montelione GT. Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. J Am Chem Soc 2005;127:1665–1674.
23. Bhattacharya A, Tejero R, Montelione GT. Protein structure validation software (PSVS) suite, and its applications in evaluating protein structures generated by structural genomics consortia. In preparation.
24. Hall S. The STAR File: a new format for electronic transfer and archiving. Chem Inf Comput Sci 1991;31:326–333.
25. Monleon D, Colson K, Moseley HN, Anklin C, Oswald R, Szyperski T, Montelione GT. Rapid analysis of protein backbone resonance assignments using cryogenic probes, a distributed Linux-based computing architecture, and an integrated set of spectral analysis tools. J Struct Funct Genomics 2002;2:93–101.
26. Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD. 1H, 13C and 15N chemical shift referencing in biomolecular NMR. J Biomol NMR 1995;6:135–140.
27. Wishart DS, Sykes BD. The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. J Biomol NMR 1994;4:171–180.
28. Baran MC, Huang YJ, Moseley HN, Montelione GT. Automated analysis of protein NMR assignments and structures. Chem Rev 2004;104:3541–3556.
29. Vuister GW, Bax A. Quantitative J Correlation: a new approach for measuring homonuclear three-bond J (HN-Ha) coupling constants in 15N-enriched proteins. J Am Chem Soc 1993;115:7772–7777.