Journal of Biomolecular NMR, **24:** 113–121, 2002.
KLUWER/ESCOM

SPINS: Standardized ProteIn NMR Storage. A data dictionary and object-oriented relational database for archiving protein NMR spectra

Michael C. Baran[a], Hunter N.B. Moseley[a], Gurmukh Sahota[a] & Gaetano T. Montelione[a,b,*]
[a]*Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, NJ 08854, U.S.A.;* [b]*Department of Biochemistry and Molecular Biology, Robert Wood Johnson Medical School, Piscataway, NJ 08854, U.S.A.*

Received 22 May 2002; Accepted 12 August 2002

Key words: integrated analysis process, NMR spectra archive, Oracle database, structural genomics

Abstract

Modern protein NMR spectroscopy laboratories have a rapidly growing need for an easily queried local archival system of raw experimental NMR datasets. SPINS (Standardized ProteIn Nmr Storage) is an object-oriented relational database that provides facilities for high-volume NMR data archival, organization of analyses, and dissemination of results to the public domain by automatic preparation of the header files required for submission of data to the BioMagResBank (BMRB). The current version of SPINS coordinates the process from data collection to BMRB deposition of raw NMR data by standardizing and integrating the storage and retrieval of these data in a local laboratory file system. Additional facilities include a data mining query tool, graphical database administration tools, and a NMRStar v2.1.1 file generator. SPINS also includes a user-friendly internet-based graphical user interface, which is optionally integrated with Varian VNMR NMR data collection software. This paper provides an overview of the data model underlying the SPINS database system, a description of its implementation in Oracle, and an outline of future plans for the SPINS project.

Abbreviations: BMRB – BioMagResBank database; FID – free induction decay; GUI – graphical user interface; JDBC – Java Database Connectivity; Perl DBI – Perl Database Interface; SPINS – Standardized Protein NMR Storage; SPINSds – SPINS directory structure; XML – eXtensible Markup Language.

Introduction

Biomolecular NMR research groups require efficient and simple access to archival NMR data, both for routine storage purposes and for the development and testing of novel computational methods for data analysis. Current common methods of archiving raw NMR data [usually in the form of time domain free-induction decay (FID) data] in use in most biomolecular NMR laboratories are often inefficient, out-dated, and error-prone, leading to frequent loss of valuable and expensive data. Archives on tape and optical media are difficult to track and recover, sometimes lack adequate organizational and querying facilities, and have limited longevity. On the other hand, disk space is now inexpensive enough to consider using mirrored live disk archives with regular tape backups. However, whether using archival media or live disk archives, the many different data sets generated in modern biomolecular NMR laboratories creates organizational problems that need to be addressed by an appropriate database structure. The growing demands on data organization and formatting in submitting NMR data and structures to public databases like the BioMagResBank (Seavey et al., 1991) and the Protein DataBank (Berman et al., 2000) also require simple methods of harvesting NMR data and moving this information from the NMR laboratory into appropriate archival formats. This is particularly challenging for the several pilot projects in structural proteomics (see for example

*To whom correspondence should be addressed. E-mail: guy@cabm.rutgers.edu

Heinemann, 2000; Montelione et al. 2000; Terwilliger, 2000; Yokoyama et al., 2000) which are being encouraged to submit into the public domain many more data items than has been traditionally expected from a conventional structural biology project. If properly organized and archived, these data will be invaluable to the NMR community in efforts to develop new data collection and analysis technologies.

Considering these needs of our laboratory and the community, we have developed SPINS (**S**tandardized **P**rote**I**n **N**mr **S**torage), an object-oriented relational database and data model that provide facilities for high-volume NMR data archival, data organization, and dissemination of raw NMR data (FIDs) to the public domain by automatic preparation of the header files needed for simple submission to the BioMagResBank (BMRB). SPINS is in routine use in our laboratory and is now available for use by other protein NMR laboratories. The current version provides means for archiving raw time-domain NMR data onto local disk drives together with all the information needed to define these data, and facilitates deposition of these data into the public domain. This paper provides an overview of the data model underlying the SPINS database system, a description of its implementation in Oracle, and an outline of future plans for the SPINS project.

The SPINS data model

We designed the SPINS data model to capture all information necessary to completely re-create an NMR experiment. The current schema of the SPINS data model is presented in Figure 1, and a complete summary of the SPINS Data Dictionary is available both as Supplementary Material to this manuscript (Supplemental Table 1) and from our laboratory web site (www-nmr.cabm.rutgers.edu). The Experiment Table (Table 1) is the central table of the SPINS schema. Its contents, together with information in tables to which it is linked and the NMR pulse sequence itself which is stored with the FID file, completely describe a protein NMR experiment.

Certain fields of the Experiment Table are keys relating records in the Experiment Table to records in other tables. These relationships are represented by the SPINS Entity Relationship Diagram (Figure 1). For example, each experiment includes a Pulse Sequence field entry in the Experiment Table representing the name of the NMR pulse sequence used to collect a particular NMR data experiment. It is important to make the distinction that while the Experiment Table holds the name of the pulse sequence used in the experiment, data fields describing the pulse sequence itself are stored in the Pulse Sequence Table. Since SPINS is a relational database, the PULSE_SEQUENCE field of the Experiment Table is simply a pointer to the corresponding Pulse Sequence Table. Therefore, by tracking a pulse sequence in the Experiment Table, all information in the Pulse Sequence Table associated with the tracked pulse sequence is also stored. Relationships of this nature (Figure 1) define the SPINS data model.

One of the important goals of our data model is to allow SPINS to automatically generate header files and contents of BMRB NMRStar files directly from the SPINS database. The BMRB (Seavey et al., 1991) 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 NMRStar formatted files, we have designed the internal SPINS data model to be as consistent as possible with the NMRStar v2.1.1 data model currently used by the BMRB. Although some data items have different names, the syntactical differences between the SPINS and NMRStar v2.1.1 data model are limited to those data items supported by SPINS and not yet supported by NMRStar. In particular, the SPINS description of the NMR_Sample, and the data tables and items describing the molecular structure are essentially identical to the NMR_Star v2.1.1. data model. Efforts are made to maintain this compatibility and consistency as these related NMRStar and SPINS data models evolve. Future releases of SPINS will expand upon the current SPINS data dictionary to include controlled vocabulary for all key parameters relevant to data collection and each step of the structure determination process. As the SPINS project evolves, its schema is guided by the well-characterized NMRStar dictionary and by the evolving universal data format under development by the CCPN project (http://www.bio.cam.ac.uk/nmr/ccp/datamodel/datamodel.html).

Overview of system functionality

An optimal solution for the implementation of the SPINS data model is a database that is well supported, scalable, fast, reliable, supports the use of

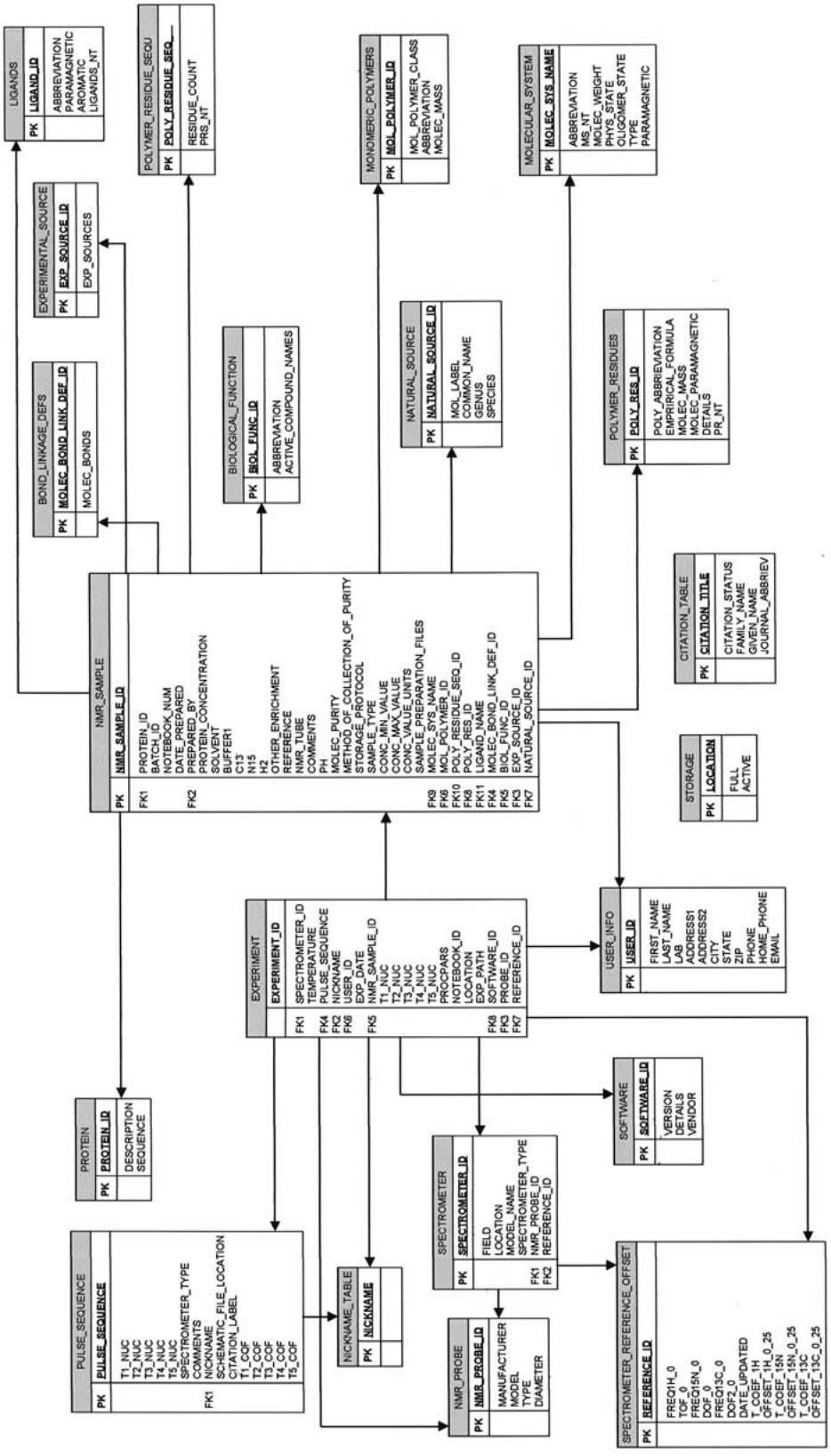

Figure 1. SPINS Entity Relationship Diagram. A schematic view of the SPINS tables and their relationships. Certain columns serve to link tables together. These relationships are designated next to the column name. PK designates a unique primary key, while FK designates the column as a foreign key link to another table. Details of the SPINS data model are presented in Supplemental Tables.

Table 1. A partial view of the SPINS data dictionary showing data fields of an Experiment Table. The SPINS data model was designed to be fully compatible with the public BioMagResBank (BMRB) data model. Where corresponding data items exist, the mapping of the SPINS model to the NMR-Star format is shown. Fields which are links to other tables appear in bold

Database field	NMR-star (BMRB) field	Description
EXPERIMENT_ID	Experiment_ID	Unique identifier of experiment
SPECTROMETER_ID	NMR_spectrometer_ID	Link to Spectrometer Table defining spectrometer used to collect data
TEMPERATURE	Variable_value(In sample_conditions)	Temp at which experiment was conducted
PULSE_SEQUENCE	Pulse_sequence_file_name	Link to NMR Pulse Sequence Table defining pulse sequence used
NICKNAME		Experiment nickname
USER_ID		Link to User Table defining the person(s) who did NMR experiment
EXP_DATE		Date experiment was performed
NMR_SAMPLE_ID	Sample_ID	Link to Sample Table defining sample
T1_NUC	Detected_nucleus	Nuclei labeled in t1 dimension
T2_NUC	Detected_nucleus	Nuclei labeled in t2 dimension
T3_NUC	Detected_nucleus	Nuclei labeled in t3 dimension
T4_NUC	Detected_nucleus	Nuclei labeled in t4 dimension
T5_NUC	Detected_nucleus	Nuclei labeled in t5 dimension
PROCPARS		All Varian *procpar* variables
NOTEBOOK_ID		Laboratory notebook i.d.
LOCATION		(base) Disk location
EXP_PATH		Disk location of experimental data
SOFTWARE_ID	Software_ID	Link to Software Table describing software used for data collection
PROBE_ID	NMR_probe_ID	Link to NMR Probe Table
REFERENCE_ID	NMR_spectral_processing_ID	Link to Spectrum Reference Data Table

complex data types, and runs on Linux. The Oracle 9i database meets all of these criteria. First, Oracle allows the implementation of the relational model, as well as the creation of complex object types necessary to represent certain data fields. Second, Oracle provides a proven scalable solution to warehouse the vast number of data sets to be generated within a large protein NMR laboratory or across a multi-institutional structural proteomics effort like the Northeast Structural Genomics Consortium (NESG) (www.nesg.org). Third, as an industry standard product, Oracle is well supported and provides multiple options for software development. Finally, the preservation of these raw NMR data is absolutely critical, and Oracle provides proven backup and recovery tools to ensure data integrity. Oracle also provides extensive XML support, including tools for parsing, transforming, querying and validating XML data. It is possible for SPINS to be ported with minimal changes into any SQL relational database which supports Java Database Connectivity (JDBC). However, an Oracle implementation is recommended for the current release of SPINS.

Figure 2 shows the implementation of SPINS in a standard three-tier architecture. The interface is completely web based, but designed to securely operate in an intranet. Static HTML pages are housed on an Apache web server. These pages link to Java servlets located on a Tomcat Java Server. Implementation of the interface as modules of server side Java code (servlets) allows for easy expansion of future releases by simply adding new modules to the existing code. The Java servlets communicate directly with the database through Java Database Connectivity (JDBC) or by executing Perl scripts which connect through the Perl Database Interface Module (DBI). An important feature of SPINS is that when data from an experiment are submitted, SPINS moves the data files into its own dynamically created directory system called

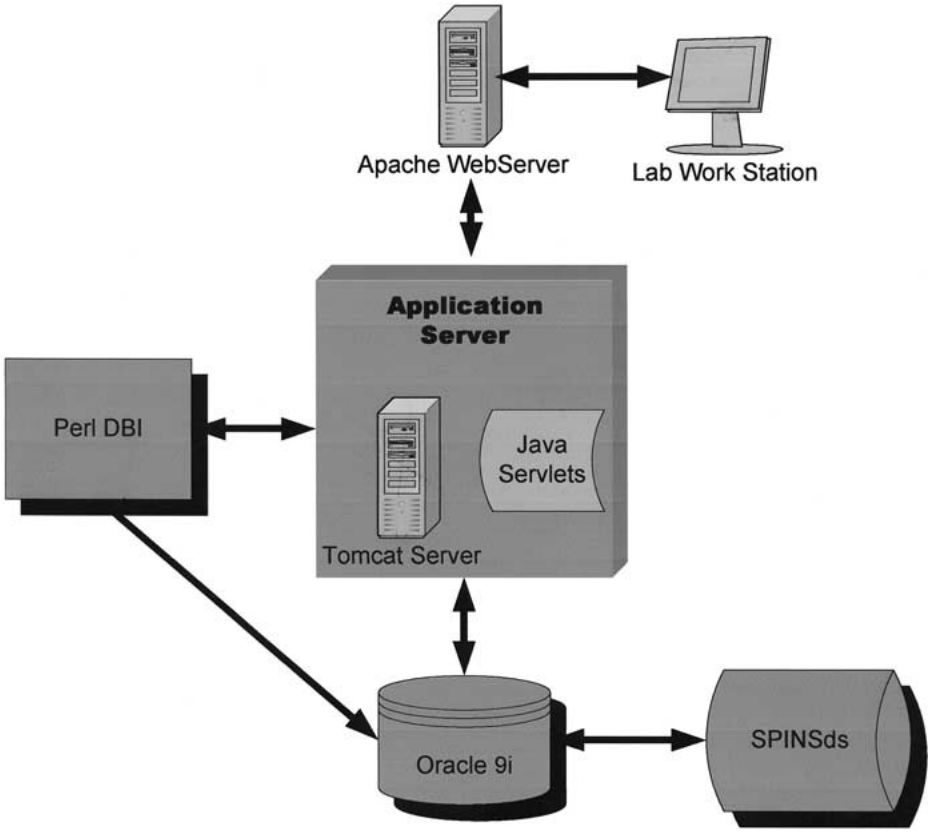

Figure 2. SPINS Three-Tier System Architecture. This figure shows the relationships between the major components implementing SPINS: Oracle 9i database, Tomcat Application Server and Apache Web server. We developed SPINS under Oracle 9i v9.0.1 enterprise edition database running on Red Hat Linux 7.1 using Java servlet technology in conjunction with Perl 5.005, and Tcl/TK v7.6. Java Database Connectivity (JDBC) drivers and the Perl Database Interface (DBI) drivers provide the direct interface between the Java servlets on the Tomcat server and the Oracle database. Limiting the Tomcat server access to the local intranet provides significant additional security.

SPINSds (SPINS directory structure). The database stores pointers to the location of the data files, but the actual files are stored within SPINSds. Files stored in the SPINSds include the FID file, the parameter file (e.g. the procpar file) specifying all parameters of data collection (i.e. data sizes, acquisition modes, etc.), the pulse sequence file describing the exact details of how the experiment was collected, and any additional data collection files (e.g., decoupler modulation shape files and files describing phase cycling) required to reproduce the experiment. By storing the actual pulse sequence used in the data collection and related data collection files with the FIDs, the database captures all the details of the data collection process without actually defining a controlled vocabulary for the complex process of NMR pulse sequence implementation. Pulse sequence parameters which are named differently in different pulse sequence implementations, such as parameters defining the center of the spectrum in each dimension, are mapped to standard parameter names in the Pulse Sequence Table. Additional pulse sequence parameter names, such as NOESY mixing time, will also be mapped in a similar fashion to standard internal names within SPINS in future versions of the data dictionary.

SPINS also features a user-friendly graphical user interface (GUI). A representative data view from this GUI is shown in Figure 3. A menu based navigation system is always present on the left hand side of the interface. The menu system provides links to the many SPINS features including SPINS Tools, Table Data entry forms, and online Help documentation. A separate data entry form in the interface represents each table in the SPINS data model (Figure 1). Menu options allow the user to carry out the following tasks: archive experiments, permit access to the SPINS Tools, which include the data mining tool and BMRB file generator, access forms for all tables in the SPINS schema,

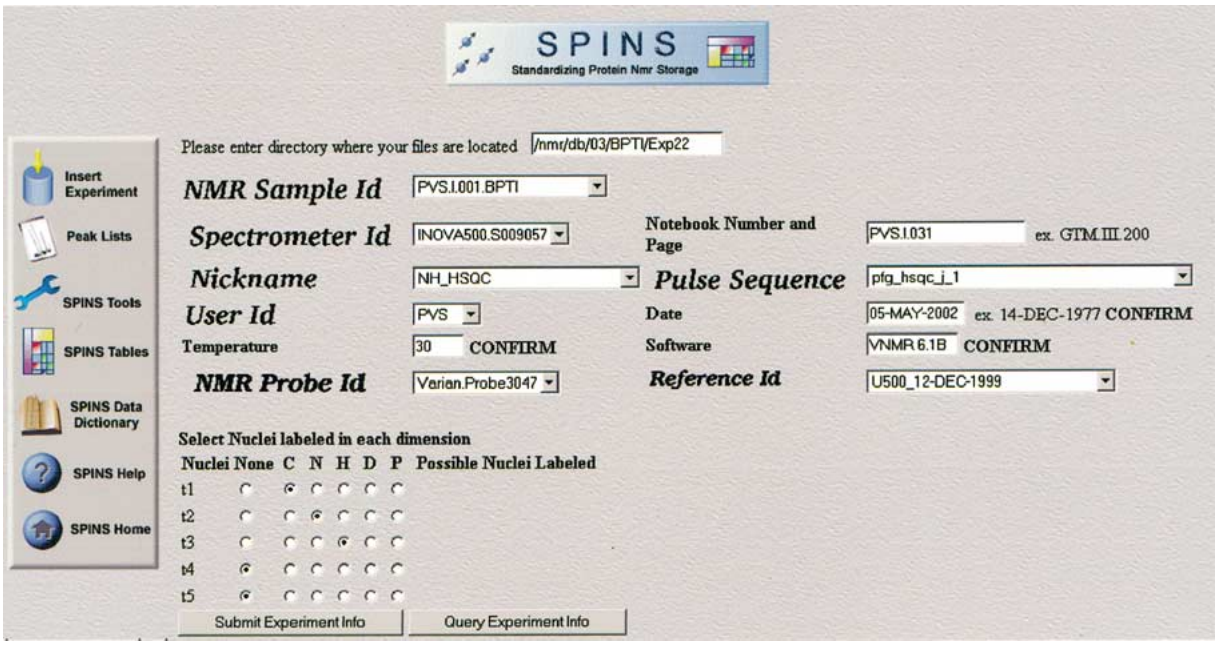

Figure 3. View of SPINS Graphical User Interface (GUI). Experiment Table data entry form, and example of the SPINS GUI. From the left menu bar, one can navigate the SPINS user environment. Options include links to Insert Experiments, use of SPINS Tools (including the Query Tool and BMRB file Generator), On-Line Help pages, the SPINS Data Dictionary, and a variety of SPINS administration tools. This figure is a screen shot of an Experiment insertion form. Most of the fields are pre-filled by the AutoFill tool and all that remains is for the user to check the entry, fill in any missing data, and click the submit button. Links are provided from this Experiment Table to other SPINS data tables (e.g., NMR Sample Id is a link to the corresponding NMR Sample Table).

and use the online help system. Each form provides insert/update/delete and extensive query functionality. For example, if querying the Pulse Sequence Table, the user can enter 'hnco' in the Pulse Sequence Name text entry field, and all pulse sequences containing the string hnco will be returned.

Archiving NMR data into SPINS is a simple process. SPINS supports two modes of data entry; (i) insertion of experimental NMR data previously stored on disk or (ii) insertion of NMR data directly through vendor-provided NMR data collection software (e.g., the Varian VNMR data collection software). This second method of data entry is not yet supported for Bruker NMR systems.

To insert an experiment previously saved to a hard disk the user simply launches the SPINS interface, opens the 'insert from disk' tool and selects the directory containing the FID file via a graphical pop up menu. The SPINS software then parses the Parameter File (e.g., Varian procpar file) and pre-fills many of the fields in the Experiment data entry form. The user is then asked to complete a minimal number of additional data fields manually, and the data are then submitted to the database by a click of a button.

SPINS also provides a Varian VNMR Integration Toolkit, which allows data to be entered into the SPINS database directly from Varian VNMR software. The VNMR Toolkit allows users to directly submit their experiments from the spectrometer host computer. It is built around the *makeText* macro, a Tcl/TK graphical utility within VNMR for generating a standardized text file readable by the SPINS software. To insert an experiment through VNMR the user first creates a text file using the *makeText* macro. Next, the experiment is saved onto disk using the macro *svfspins*, which saves the experiment to the present working directory as well as into a temporary directory where it remains until inserted into the database by the user. The user next launches the SPINS graphical user interface from the VNMR menu. Finally, the user selects the experiment to insert from a temporary directory area and verifies the correctness of the now completely pre-filled experiment data entry form before pressing the 'submit button' to copy the FID and associated files into the SPINSds file system.

Validations

To ensure data integrity, SPINS performs certain information and format validations. The following validations are run on all records submitted to the database using either data entry method described above:

Record uniqueness constraints

SPINS standardizes the naming of experiments as well as samples. Experiment id's are formed by combining the user's initials, notebook number, notebook page number, protein name, and pulse sequence (e.g., MCB.I.003_BPTI_pfg_hnco_se1.fid). The resulting Experiment id provides a unique label for the FID. A similar approach is used to uniquely assign the Sample id. The software verifies that such Experiment and Sample id's are unique before the data are accepted into the database.

Validate spectrometer reference date vs. date of experiment

Associated with each spectrometer are unique parameter values that are used in referencing the multidimensional NMR spectra (Monleon et al., 2002; M. Bayro et al., in preparation). These spectrometer-specific values may change from time to time due to adjustments or upgrades of spectrometer configurations. If a user tries to insert an experiment with a date earlier than the last update of the currently defined spectrometer-specific spectral referencing parameters stored in the Spectrometer Table, the interface voids the insertion and issues a warning. This warning can be overridden, but the user should make note that the spectrometer-specific referencing values stored with the experiment may be inaccurate.

Validate nuclei detected in experiment parameter set vs. pulse sequence definition

Each Pulse Sequence is defined to potentially use one of multiple nuclei types in each dimension (e.g., some experiments may be run with either ^{13}C or ^{1}H frequency labeling in the indirect t_1 dimension). If the user attempts to archive an experimental data set into the SPINS database with parameters inconsistent with these definitions associated with the corresponding pulse sequence, a warning is indicated.

Auto Fill tool validates fields vs. database

SPINS performs a validation to ensure that values read from a parameter file (e.g., a procpar file) or text file are consistent with the SPINS data dictionary as well as valid entries in the SPINS database. For example, a validation is performed to ensure the pulse sequence indicated by the user has a corresponding Pulse Sequence table entry in the SPINS database. Similarly, an experimental FID data set cannot be entered into the database unless there exists a NMR Sample Table entry describing the corresponding NMR sample. These kinds of validations are critical as they enforce the integrity of the relational model by preventing the user from inserting a data record without first inserting the associated records necessary to define these data; e.g., an FID data set cannot be inserted before the NMR sample and pulse sequence used to collect these data are defined.

SPINS tools

We developed several additional software tools to aid the user in archiving into and querying across the SPINS database. The Query/Data mining tool is used to search the database for archived experiments. This tool allows users to search the database based on one or a combination of constraints. For example, the user can search for all experiments done by user Scott using pulse sequences with the nickname hnco. Furthermore, the query tool allows searching by data collection parameter (e.g., procpar) names and values, allowing the user to, for example, search for all NOESY spectra recorded with the parameter 'mix' > 100 ms on samples at pH < 5. Along with the metadata describing the experiment, the tool also provides the user with the location (path) of the corresponding experimental files within the SPINSds.

As described above, the SPINSds is the logical hierarchy which SPINS creates to store experiments in the underlying file system. As an administrative tool, SPINS includes a SPINSds viewer. The viewer is an expandable tree representation of the SPINSds. Using the tool users can visually browse the SPINSds architecture to examine how files have been organized by SPINS and/or look for specific files that have been archived in the database.

SPINS can be used to generate header and content files for the BioMagResBank using its BMRB File Generator, which creates an NMR Star v2.1.1 file

populated with data held in SPINS that correspond to information supported by the BMRB archive. Using this feature, we have recently submitted from SPINS to the BMRB a group of six FID data sets corresponding to 3D triple resonance spectra recorded for bovine pancreatic trypsin inhibitor (BMRB entry 5307) suitable for automated analysis of backbone resonance assignments with our software packages AutoProc (M. Bayro et al., in preparation) and AutoAssign (Moseley et al., 2001). Oracle is fully compatible with XML formats for input and output. Future implementations of this tool will be based on XML data transformations into NMRStar v3.0 format, as well as other formats which we may choose to support.

SPINS also provides a 'disk space checker', which indicates a warning if the user attempts to submit an experiment and there is insufficient disk space to accommodate the data set. In this case, the user is given the choice to select a new storage location from the SPINS Storage Table, which defines the database storage disks. Finally, as mentioned above, SPINS also provides an AutoFill tool which pre-fills user forms to make the archival process as simple and user-friendly as possible. The AutoFill tool functions by parsing the data collection (e.g., Varian procpar file) and/or text files associated with the FID data set.

Documentation

Extensive online Help documentation is available through the SPINS menu navigation system. The help documentation is organized through a table of contents and features a series of screen shots to virtually walk the user through the use of the interface. A PDF version of the documentation is also included in the SPINS v1.0 distribution as well as a Java-based installation GUI.

Future directions

SPINS v1.0 and its associated data dictionary (Supplementary Table 1) represent the first phase of a multi-phase process integration project. SPINS v1.0 provides organization, archiving, and simple submission to the BMRB of time domain FID files and all the information needed to describe and reproduce these data. This version of SPINS is freely available to academic users. It provides a valuable tool to the biological NMR community for local data organization within a laboratory and for simplifying the process of archiving FID data into the public domain. Currently, our laboratory version of the SPINS database includes 432 FID data sets, requiring 19 Gbytes of disk space (i.e., ~ 44 Mbytes per FID). Relatively few raw NMR data sets (FIDs) are currently available in the public domain and routine archiving of such data using tools like SPINS will have significant scientific value.

We plan for SPINS to evolve into a central agent which integrates the entire process of NMR-based protein structure determination. As the protein spectroscopist progresses through the resonance assignment and structure determination process, SPINS will serve as the central archive, logging important information critical for documenting and reproducing each step of the NMR data analysis, and generating intermediate files in appropriate formats for the supported specific software applications. Our ultimate vision is for SPINS to take the form of a centralized automating agent, integrating all the NMR automation software from our lab [e.g., AutoProc (M. Bayro et al., in preparation), AutoPeak (Monleon et al., 2002), AutoAssign (Zimmerman et al., 1997; Moseley et al., 2001), and AutoStructure (Huang, 2001)] as well as other academic and public domain software [Sparky (Goddard et al., 2000), NMRPipe (Delaglio et al., 1995), DYANA (Güntert et al., 1997), CNS (Brunger et al., 1998), etc.], into a single GUI-based data analysis tool, putting the entire resonance assignment and structure determination process at the user's fingertips. The SPINS data dictionary is also designed to be consistent with the evolving structural genomics project databases, such as the SPINE database (Bertone et al., 2001) for tracking the protein expression and sample production processes of a large scale structural genomics project. Finally, SPINS will be capable of auto-submission of the associated intermediate and final data files generated in the process of NMR resonance assignments and structure analysis to the public domain BioMagResBank (Seavey et al., 1991) and Protein Data Bank (Berman et al., 2000) in a fully validated format.

Acknowledgements

We thank J. Aramini, J. Everett, D. Monleon, R. Paranji, C. Rios, T. Szyperski, and G.V.T. Swapna, along with J. Markley and E. Ulrich of the BioMagResBank for helpful discussions, suggestions, and advise. This work was supported by grants from the

Protein Structure Initiative of National Institutes of Health (P50 GM62413) and The New Jersey Commission on Science and Technology (2042-007-13).

References

Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shndyalov I. and Bourne, P. (2000) *Nucl. Acids Res.*, **28**, 235–242.

Bertone, P., Kluger, Y., Lan, Y., Zheng, D., Christendat, D., Yee, A., Edwards, A.M., Arrowsmith, C.H., Montelione, G.T. and Gerstein, M. (2001) *Nucl. Acids Res.*, **29**, 2884–2898.

Brunger, A.T., Adams, P.D., Clore, G.M., Delano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) *Acta Cryst.*, **D54**, 905–921.

Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.

Goddard, T.D. and Kneller, D.G. (2000) *SPARKY 3*, University of California, San Francisco, CA.

Güntert, P., Mumenthaler, C. and Wüthrich, K. (1997) *J. Mol. Biol.*, **273**, 283–298

Heinemann, U. (2000) *Nat. Struct. Biol.*, **7**, 940–942.

Huang, Y.J. (2001) *Automated Determination of Protein Structures from NMR Data by Iterative Analysis of Self-Consistent Contact Patterns*, Ph.D. Thesis, Rutgers, The State University of New Jersey, Piscataway, NJ.

Monleón, D., Colson, K., Moseley, H.N.B., Anklin, C., Oswald, R., Szyperski, T.A. and Montelione, G.T. (2002) *J. Struct. Funct. Genomics*, **2**, 93–101.

Montelione, G.T., Zheng, D., Huang, Y.J., Gunsalus, K.C. and Szyperski, T. (2000) *Nat. Struct. Biol.*, **7**, 982–985.

Moseley, H.N.B., Monleon, D. and Montelione, G.T. (2001) *Meth. Enzymol.*, **339**, 91–108.

Seavey, B.R., Farr, E.A., Westler, W.M. and Markley, J. (1991) *J. Biomol. NMR*, **1**, 217–236.

Terwilliger, T.C. (2000) *Nat. Struct. Biol.*, **7**, 935–939.

Yokoyama, S., Hirota, H., Kigawa, T., Yabuki, T., Shirouzu, M., Terada, T., Ito, Y., Matsuo, Y., Kuroda, Y., Nishimura, Y., Kyogoku, Y., Miki, K., Masui, R. and Kuramitsu, S. (2000) *Nat. Struct. Biol.*, **7**, 943–945.

Zimmerman, D.E., Kulikowski, C.A., Huang, Y., Feng, W., Tashiro, M., Shimotakahara, S., Chien, S.Y., Powers, R. and Montelione, G.T. (1997) *J. Mol. Biol.*, **269**, 592–610.