

Straightforward and complete deposition of NMR data to the PDBe

Christopher J. Penkett · Glen van Ginkel · Sameer Velankar · Jawahar Swaminathan · Eldon L. Ulrich · Steve Mading · Tim J. Stevens · Rasmus H. Fogh · Aleksandras Gutmanas · Gerard J. Kleywegt · Kim Henrick · Wim F. Vranken

Received: 4 July 2010 / Accepted: 19 July 2010 / Published online: 3 August 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract We present a suite of software for the complete and easy deposition of NMR data to the PDB and BMRB. This suite uses the CCPN framework and introduces a freely downloadable, graphical desktop application called CcpNmr Entry Completion Interface (ECI) for the secure editing of experimental information and associated datasets through the lifetime of an NMR project. CCPN projects can be created within the CcpNmr Analysis software or by importing existing NMR data files using the CcpNmr FormatConverter. After further data entry and checking with the ECI, the project can then be rapidly deposited to the PDBe using AutoDep, or exported as a complete deposition NMR-STAR file. In full CCPN projects created with ECI, it is straightforward to select chemical shift lists, restraint data sets, structural ensembles and all relevant associated experimental collection details, which all are or

will become mandatory when depositing to the PDB. Instructions and download information for the ECI are available from the PDBe web site at <http://www.ebi.ac.uk/pdbe/nmr/deposition/eci.html>.

Keywords Database deposition · CCPN · wwPDB · Structure calculation · Structure validation · NMR-STAR

Introduction

Public databases that archive scientific data hold a crucial record of experimental information, especially in relation to associated publications. In the field of structural biology, the Protein Data Bank (PDB; Berman et al. 2000) has been storing three-dimensional structural data of mainly proteins, DNA and RNA since 1971. The predominant experimental techniques to determine these structures are X-ray crystallography and Nuclear Magnetic Resonance (NMR) spectroscopy. In recent years, the worldwide PDB (wwPDB; Berman et al. 2007), the organisation that manages the PDB, has begun to require the mandatory deposition of an increasing amount of experimental data, associated parameters and meta-data. For structures determined by X-ray crystallography, the deposition of structure factors has been mandatory since 2008. In NMR, where data management is typically more complex because of the variety of data that can be obtained and the general lack of consistent data formats, deposition of the restraints used to calculate the structure has been mandatory since 2008 as well (Markley et al. 2008). The deposition of chemical shifts and associated referencing information will become mandatory during 2010. Furthermore, it is not inconceivable that additional types of data and information will become mandatory for deposition at the PDB in the future.

Electronic supplementary material The online version of this article (doi:10.1007/s10858-010-9439-3) contains supplementary material, which is available to authorized users.

C. J. Penkett · G. van Ginkel · S. Velankar · J. Swaminathan · A. Gutmanas · G. J. Kleywegt · K. Henrick · W. F. Vranken (✉)
Protein Data Bank in Europe, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK
e-mail: wim@ebi.ac.uk

E. L. Ulrich · S. Mading
BioMagResBank, Department of Biochemistry,
University of Wisconsin Madison, 433 Babcock Dr.,
Madison, WI 53706, USA

T. J. Stevens · R. H. Fogh
Department of Biochemistry, University of Cambridge,
80 Tennis Court Road, Cambridge CB2 1GA, UK

Whilst obtaining this additional NMR data is of increasing value to a varied group of researchers; for example, in chemical shift-based structure calculation (Cavalli et al. 2007; Wishart et al. 2008; Shen et al. 2008), structure recalculation efforts (Nederveen et al. 2005), and large scale data analyses (Vranken 2007; Vranken and Rieping 2009), it has also made the deposition of NMR data more complicated. For deposition of macromolecular NMR data to the PDB, there are currently two web-based options: the first is AutoDep (Sen et al. 2007), which is hosted by the Protein Data Bank in Europe (PDBe; Velankar et al. 2010) group at the European Bioinformatics Institute (EBI). This allows users to submit coordinates, NMR-derived structure restraints and other NMR data items such as chemical shifts or peak lists along with associated information such as authors, citation references, molecule/sequence data, sample and experimental information. The second option is ADIT-NMR at the BioMagResBank (BMRB, University of Wisconsin and PDBj-BMRB, Osaka, Japan; Ulrich et al. 2008), which allows submitters to deposit similar information, but with a different web-based input tool. The main disadvantage of web submission is that it can be slow and cumbersome to fill in the forms and to locate and upload all the necessary data and files.

To reduce and simplify the extra work required by data depositors, we have developed a complete and easy-to-use set of tools for depositing NMR data for biological macromolecules. These desktop and web-based software tools use the framework provided by the Collaborative Computational Project for NMR (CCPN; Fogh et al. 2002, 2005, 2010) to gather all the mandatory data together in a single CCPN project where all data is stored in an interlinked and consistent way. These CCPN projects can then be deposited using AutoDep at PDBe (Velankar et al. 2010) or exported as full NMR-STAR files ready to be uploaded to ADIT-NMR at BMRB or PDBj-BMRB (Ulrich et al. 2008) (Fig. 1). In particular, we introduce CcpNmr ECI (Entry Completion Interface) for editing all the deposition information safely and securely on the user's desktop. For CCPN projects made using CcpNmr (Vranken et al. 2005) or Extend-NMR (<http://www.extend-nmr.eu/>) software, it becomes straightforward to select chemical shift lists, restraint sets, structural ensembles, and peak lists at the click of a button. The CcpNmr FormatConverter and associated deposition file import tool can be used to create CCPN projects from existing NMR data files. In addition, all associated meta-data required for AutoDep and ADIT-NMR are easily added to the same CCPN project.

These developments ensure higher quality of the deposited data because the depositor either uses the internally consistent CCPN framework while analysing data, or is interactively involved in resolving issues with ambiguous data when creating a CCPN project for deposition only from

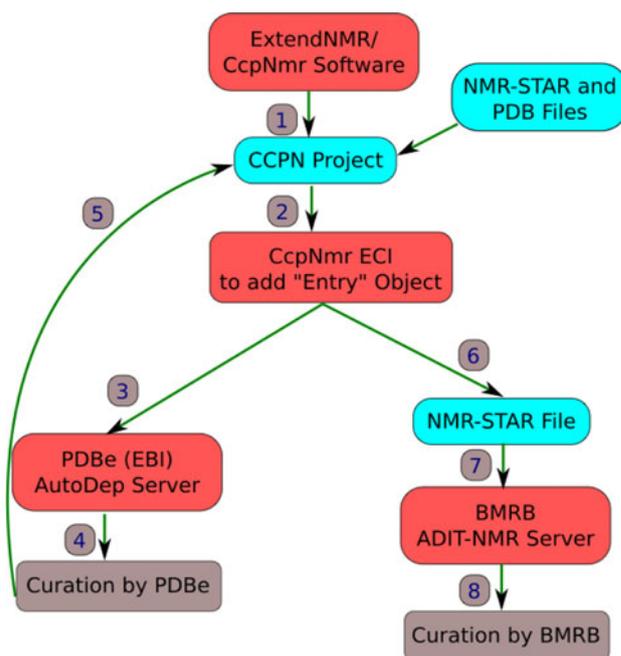


Fig. 1 The CCPN system for depositing NMR data with the PDB. (1) In order to use the deposition system, a CCPN project, created with CcpNmr software, is required. (2) Use CcpNmr ECI to add an “Entry” object. Also, edit and select the data that is to be submitted for public use. (3) For depositions with macromolecular structures, first upload the CCPN project to the CCPN-AutoDep server at the PDBe. (4) Once this submission is complete, the depositor will receive a PDB ID code and PDBe annotators will curate the structural data. (5) The curated data is incorporated into the submitted CCPN project. (6) The updated CCPN project will be exported as an NMR-STAR file and sent with all the other data to the BMRB. (7) This NMR-STAR file will be uploaded to the ADIT-NMR server at BMRB; at this point, the depositor will be supplied with a link for them to complete the submission at BMRB. (8) On completion of this submission, annotators at BMRB will curate the NMR data

existing NMR data formats. They also greatly reduce the need for annotator intervention during deposition, because the depositor ensures the consistency of the incoming data pre-deposition. This is important for the future of the wwPDB, as annotation staff numbers remain constant while the number of structure depositions increase and the deposition of additional experimental data becomes mandatory. The tools for deposition of NMR structures and related data described here thus help to maintain the quality of the PDB archive and uphold current response times to depositors.

Methods

All the code is written in Python and uses the CCPN API (Application Programming Interface) libraries to read CCPN XML (eXtensible Markup Language) project files into Python objects (Fogh et al. 2010). Reading and writing of external data files (for example, coordinates from PDB

files, or NMR restraints files from various formats) are performed using FormatConverter libraries (Vranken et al. 2005). Reading of NMR-STAR files is also done using FormatConverter libraries, which have been extended as part of this work, to import all data found in NMR-STAR 3.1 files and the header information from PDB files. For export of NMR-STAR files from completed CCPN projects, we have developed new code that uses a Python-based dictionary (Ccpn_To_NmrStar.py) and associated parser (NmrStarExport.py). For each NMR-STAR data value (tag), the dictionary specifies a CCPN data value (attribute) or in more complex cases a Python subroutine that will provide the relevant data. The dictionary controls looping over CCPN data objects where there are multiple values (e.g., chemical shifts) to be written out in an NMR-STAR data values table, and can handle further complications in the export process; for example, when data for a single NMR-STAR data category (saveframe) must be extracted from several different types of CCPN objects. It is possible to define mappings between specific CCPN framework releases and NMR-STAR versions in the NMR-STAR export framework; it can thus maintain point-to-point compatibility between previously mapped CCPN releases and NMR-STAR versions, and this setup is essential for compatibility testing between new CCPN and NMR-STAR versions before they are released.

CcpNmr ECI is written in Python Tk, using libraries similar to those used for CcpNmr Analysis (Vranken et al. 2005), so that it can be used on the user's desktop for extra security and convenience. It can be run either as a stand-alone application, from CcpNmr Analysis or from the Extend-NMR GUI (graphical user interface). It is tab-based (Fig. 2) and has separate data sections in each tab for editing and adding the deposition information. The organisation of the tabs aims to be intuitive, and is set up so that data items can be connected to each other; the "Experiments" tab in particular allows the user to connect each experiment to chemical shift references, associated chemical shift lists, samples used, experimental conditions, and information about the NMR spectrometer and probe.

Two main Python scripts handle CCPN projects as part of the AutoDep web interface. The first script (Ccpn2AutoDep.py) converts the CCPN data into AutoDep XML files. It also automatically exports structures in PDB format and distance restraints in CNS format when the CCPN project is uploaded. After the data has been deposited and curated, a second script (Autodep2Ccpn.py) reads the final AutoDep XML file and the curated PDB file to identify any new data that was added or information that was updated during deposition or curation, which can then be updated in the original CCPN project.

The deposition pipeline described here has been tested on 102 real projects received at the PDBe over the last five

years, including projects with complicated molecular systems (Table S1), and covering a wide variety of NMR data (Table S2). Note that the FormatConverter covers a much wider range of software (<http://www.ebi.ac.uk/pdbe/nmr/software/formatConverterIOTable.html>).

CcpNmr software is released in two versions, both of which include ECI; they can be downloaded from: http://www.ebi.ac.uk/pdbe/nmr/deposition/eci.getting_started.downloads.html. The releases have different advantages from a deposition point of view: the full release includes CcpNmr Analysis, providing greatly enhanced visualisation and analysis capabilities, and is fully supported on 32 and 64-bit Linux, Mac OSX Intel/PPC and Windows. The Format Converter-only release is essentially platform-independent as it only requires the widely available Python and Tcl/Tk packages.

Results

Figure 1 shows an overview of the deposition system described here. The individual components are described below. Tutorials and detailed help for each component can be found on the PDBe web site (<http://www.ebi.ac.uk/pdbe/nmr/deposition/>).

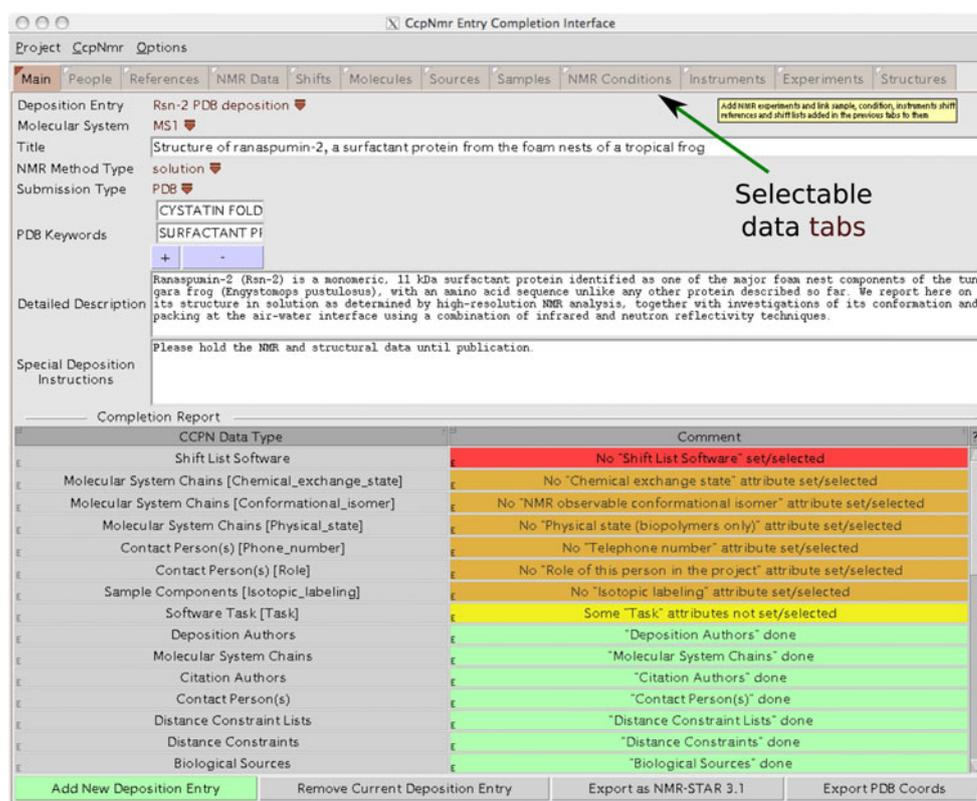
Creating CCPN projects

The investigator typically begins with a CCPN project created whilst working with the suite of CCPN-framework integrated software. Currently, the main tools available are CcpNmr FormatConverter and CcpNmr Analysis (Vranken et al. 2005). CcpNmr Analysis is a spectrum visualisation, resonance assignment and NMR data analysis application. For users who have custom pipelines to calculate NMR structures, FormatConverter allows for the import of most types of NMR data from the most common other NMR assignment, peak picking and structure calculation programs. It is available both as a desktop tool and in a web-based version (see: <http://www.ebi.ac.uk/pdbe/nmr/software/formatConverterUsage.html>). Users of CcpNmr Analysis, or the CcpNmr Extend-NMR software, will already have all their data in a CCPN project. Whatever the starting point, the end result is a CCPN project that contains derived NMR data such as chemical shifts, restraint sets and structure ensembles (see <http://www.ebi.ac.uk/pdbe/nmr/deposition/overview.html>) (Fig. 1).

CcpNmr ECI

The Entry Completion Interface (ECI) is designed to supplement the data in an existing CCPN project so that it contains almost all information required for quick web-based deposition. In the ECI, it is first necessary to create a

Fig. 2 CcpNMR ECI. The tabs correspond to the types of meta-data that can be added and CCPN data that can be selected for deposition. The *bottom frame* shows a report on the completeness of the deposition, colour coded to distinguish between sections that are still not done (*red*), already completed (*green*) or partially completed (*orange* and *yellow*). Clicking on any incomplete section (red, orange or yellow) will activate the relevant tab to add the missing information. When hovering the mouse pointer over the titles and widgets, a tool tip presenting some brief help text will pop up



new "Entry" record in the "Main" tab; this "Entry" can track and store all the deposition data in the CCPN project (Fig. 2). The "NMR Data" tab allows selection of the NMR data underlying the deposition and any associated publications (Fig. 3). In the "Structures" tab, the depositor can select the final structural ensemble and restraints used to calculate this ensemble. The other ECI tabs allow adding and editing of different kinds of meta-data, such as contact authors and associated information like e-mail addresses, entry authors, publications, software used, molecular system information, biological and experimental sources of bio-polymers, sample and isotope labelling details, NMR experiment details and conditions, and spectrometer and probe information. In the "Main" tab, which holds information such as PDB title and keywords, there is also a frame showing how complete the mandatory deposition data is. Data items are colour coded to show whether or not the information is available and complete. Main sections (corresponding to CCPN objects like "Person" or "Citation") that are mandatory and still need to be completed are shown in red, whilst those that have been filled-out are displayed in green. Orange colour is used for empty mandatory subsections whilst partially completed ones are shown in yellow.

For users wishing to start with data from previous submissions to the wwPDB, it is possible to import PDB header information and NMR-STAR v 3.1 files into ECI. This data can then be edited and modified to suit the new

submission, with very similar projects requiring little user input. Furthermore, it is possible to use the compatible CcpNmr DataShifter (which is available as part of the CCPN releases) to copy data from other CCPN projects quickly into the new project. It is worth noting that archived NMR data (Ulrich et al. 2008) and "cleaned up" restraints (Doreleijers et al. 2009), which are available from the BMRB as NMR-STAR files, can be imported into CCPN and analysed further using CcpNmr and Extend-NMR software.

For validation purposes, there is an option in the "Shifts" tab to do a quick check of chemical shift values. Chemical shifts are compared to the standard distribution of all shifts for that particular atom, as determined by the BMRB. Outlier shifts are highlighted: a shift row in the table is represented in pink if the deviation from the mean is more than two standard deviations and in red if it is more than three standard deviations (Fig. 4). More help and links to pages about each tab can be found on the ECI help page (<http://www.ebi.ac.uk/pdbe/nmr/deposition/eci.html>).

AutoDep

Once a CCPN project is ready for deposition, the user has to archive and compress (tar and gzip) the CCPN directory structure into one file before upload to AutoDep. The CCPN project can be uploaded to AutoDep on the PDB web site (<http://www.ebi.ac.uk/pdbe-xdep/autodep/AutoDep?>

Fig. 3 Selecting NMR data in ECI using a mouse click. Many different types of data can be present in a CCPN project such as chemical shifts, relaxation data, peak lists, kinetic experiments and derived parameters. The type of data can be selected in the top frame and individual data sets can be selected in the bottom frame

The screenshot shows the 'CcpNmr Entry Completion Interface' with the 'NMR Data' tab selected. The top table lists data types and their counts. The bottom table shows selected shift lists, and a dropdown menu is open for the 'ShiftList : 303K shifts : 2' row, showing options: AZARA, CNS, CcpNmr Analysis, Topspin, and <None>. A green arrow points to the 'CcpNmr Analysis' option.

Choose how the shift list was made with the pulldown menu

Fig. 4 The “Shifts” tab of the ECI. The individual chemical shifts are listed for each chemical shift list selected for deposition. For validation purposes, the chemical shifts are compared to the standard distribution of all shifts for that particular atom, and outlier shifts are *highlighted*: a shift row in the table is represented in *pink* if the deviation from the mean is more than two standard deviations and in *red* if it is more than three standard deviations

The screenshot shows the 'CcpNmr Entry Completion Interface' with the 'Shifts' tab selected. The top table lists chemical shift references. The bottom table shows unusual chemical shifts for 'Shift List: 303K shifts : 2'. Outlier shifts are highlighted in pink or red.

#	Class	Isotope	Experiments	Mol. Name	Atom	Value	Unit	Ref Type	Indirect Shift Ratio	Sample Geometry	Location	Axis
1	Internal	¹ H	10	H ₂ O	H	4.643	ppm	direct				
2	Internal	¹³ C	8	H ₂ O	H	0	ppm	indirect	0.251			
3	Internal	¹⁵ N	7	H ₂ O	H	0	ppm	indirect	0.101			
4	Internal	¹ H	16	H ₂ O	H	4.702	ppm	direct				
5	Internal	¹³ C	4	H ₂ O	H	0	ppm	indirect	0.251			
6	Internal	¹⁵ N	13	H ₂ O	H	0	ppm	indirect	0.101			

#	Isotope	Resonance	PPM	SD	BMRB Mean	Random Coil
703	¹ H	MS1-A 50GluHa	5.684	7.956e-03	4.420	4.420
1365	¹ H	MS1-A 73LysHdb	1.017	2.045e-03	1.680	1.680
1361	¹ H	MS1-A 73LysHga	0.576	2.362e-03	1.370	1.420
880	¹ H	MS1-A 70LysHda	0.842	1.294e-03	1.606	1.680
1434	¹³ C	MS1-A 70LysCe	40.291	8.360e-03	41.967	42.350
5	¹ H	MS1-A 72ArgH	9.612	0.020	8.251	8.390
1044	¹³ C	MS1-A 32IleCg1	31.739	0.030	27.780	27.460
701	¹ H	MS1-A 51AspHbb	3.447	0.017	2.726	2.980
1364	¹ H	MS1-A 73LysHda	0.926	1.327e-03	1.606	1.680
1410	¹ H	MS1-A 18AsnHd2a	6.135	0	7.356	7.590
879	¹ H	MS1-A 70LysHdb	0.899	0.018	1.606	1.680
1433	¹ H	MS1-A 70LysHga	1.357	3.045e-03	2.919	

[section=basedOnCCPN](#)). A password and Autodep accession number allow the user to access and modify the submission from this point on. In the Autodep web pages, any fields completed in ECI will be auto-completed, so that as more data was added in ECI, fewer fields will require editing in AutoDep. For projects marked as “all green” in the completeness report in ECI, there will be very little user input required in AutoDep (see Fig. 5). Any experimental data items that can be deposited (for example, a restraints list) will be automatically extracted from the CCPN project (if selected in the ECI) so that the depositor will not be required to upload those data files. If all mandatory data sets are present in the CCPN project, then the file upload page in AutoDep will be marked as green when the project is first uploaded and will be automatically skipped. It is possible to go back to this page and add more files if they were not imported as part of the CCPN project. On completion of the AutoDep submission, annotators at the PDBE will curate the structural data. Any information added by the depositor in AutoDep or edited during curation will then be added and/or modified in the original CCPN project, which will be stored at the PDBE as part of the originally deposited data and will remain available for future reference (Fig. 1). Detailed instructions on Autodep deposition for NMR structures are available online at <http://www.ebi.ac.uk/pdbe/nmr/deposition/autodep.html>.

ADIT-NMR

The curated CCPN project is also exported to NMR-STAR format and this file is forwarded to the BMRB, together with the original and curated data, after the coordinate annotation is finished (typically within two days after the AutoDep deposition). The BMRB will then initiate a new ADIT-NMR deposition for the NMR data only and e-mail the depositor a web link where the BMRB data submission procedure can be completed. In our experience, for data sets that are “all green” in ECI (*i.e.* all data necessary for deposition are available), more than 90% of the fields in ADIT-NMR will be populated, and little user input will be required to finish the deposition. At this point, annotators at the BMRB will curate the NMR data in the submission (for more details about completing ADIT-NMR pages, see: <http://www.ebi.ac.uk/pdbe/nmr/deposition/adit-nmr.html>).

Discussion and conclusions

The basic philosophy behind the web tools for depositing NMR data (AutoDep and ADIT) is very different: AutoDep is designed to provide context-dependent forms based on the experimental method and refinement software used, and stores data temporarily in an internal XML format that can

later be transformed into proper archive formats such as PDB, mmCIF (Bourne et al. 1997), or NMR-STAR. However, ADIT (hosted at RCSB, Rutgers University) and ADIT-NMR store the data from their sessions in mmCIF or NMR-STAR files. In spite of these differences, both AutoDep and ADIT(-NMR) suffer from a need to have the user to separately save each completed page. This can make the deposition process slow and time-consuming and is also error prone if the user fails to notice, for example, typographical errors and then changes the web page to a new deposition section. A desktop-based solution like the one presented here is much more user-friendly. It is easier to navigate, will flag missing or incorrect data and has the advantage that all the meta-data that has been entered is stored locally and remains available. This solution does require installation of a software package, but since the software used (Python and Tk) is platform independent and the CCPN installation scripts are now well developed and tested, this does not present a major obstacle.

For NMR data storage in the PDB, the primary archive format is NMR-STAR (Ulrich et al. 2008). It is a text-based format, similar to STAR and CIF used for crystallographic data, and uses identifiers in save frames and tables to uniquely tag data items that can then be referenced elsewhere in the same file. Because of its text-based and human-readable nature, it is a good format for long-term archival of data. For software development, it is more important that data can be transformed directly and unambiguously between files and data structures in memory, and that the data consistency and validity can be assured. This is where the UML-based CCPN data model excels (Fogh et al. 2010). CCPN projects consist of many XML files, which are less intelligible to humans, but are read and written directly by the subroutine libraries that come with the CCPN implementation. Once data is in memory, the subroutine libraries (available in Python, Java and C) ensure data access and maintain data integrity, while application programs perform the actual calculations on chemical shifts, peak lists, NMR restraints, atomic coordinates and other available data.

Although it is also possible to write out an NMR-STAR file from ECI and send it to the BMRB directly, we strongly encourage that, for structure-based depositions, the whole CCPN project is uploaded first into AutoDep at the PDBE, especially if CcpNmr Analysis or Extend-NMR software were used (Fig. 1). A CCPN project contains a more complete record of the information gathered during the spectral analysis (for example, incomplete assignments or peaks that were observed but not used) and structure calculation process (e.g., the restraint lists that were used for the first iteration of a structure calculation); data that would be lost if only the final data were archived. This ability of the CCPN data model to describe all aspects of

Fig. 5 AutoDep at the PDB. The first page requests the details of the contact person for the data deposition. There is a navigation side bar on the left to move between different pages. A *green circle* indicates completed pages, whilst incomplete pages are shown with a *red arrow*. On each page, a *green star* shows complete fields whilst incomplete fields are indicated with a *red star*. Fields that are not mandatory have an “N/A” button (“not applicable”) next to them, so that non-relevant fields can be rapidly completed. There is also a “Global N/A” button that auto-completes all the non-mandatory fields on the page. If everything is set to green in ECI, then very little information is left to enter in AutoDep. Each input field links to a popup window containing brief help text

the process of macromolecular structure determination using NMR, combined with complete and faithful inter-conversion with the archive/deposition NMR-STAR file format as described in this paper, makes CCPN projects an ideal medium for NMR groups to deposit their NMR data with the PDB, as well as allowing for longer term, internal archival of data in a compact and consistent format. Finally, there is the benefit of a simpler and faster deposition process both for the users and the data curators. The user only needs to make one CCPN project that can be used for all aspects of deposition and particularly the collation of all mandatory data and associated information in a simple, secure and organised fashion. As increasing amounts of NMR data gradually become mandatory for journal authors to obtain PDB and BMRB accession codes, authors will find that this unified approach saves them large amounts of time when depositing NMR models and data, and annotators will not have to spend time dealing with unnecessary data consistency issues.

There are a wide variety of NMR software programs that can now create or use CCPN projects. One good example of this is the iCing server (currently in beta form at <http://nmr.cmbi.ru.nl/icing/>) that allows NMR spectroscopists to validate their own CCPN projects and identify potential problems with respect to structural ensemble geometry (WHATIF; Hoofst et al. 1996 and PROCHECK-NMR; Laskowski et al. 1996), NOE violations (Doreleijers et al. 2005) and chemical shift data (SHIFTX; Neal et al. 2003).

In conclusion, a simple, secure and complete deposition system is presented for NMR depositors, that allows local editing and storage of all information related to an NMR structure determination project. Since this system is based on a framework that consistently stores all types of NMR data, it is also easily amenable to include new mandatory data types for deposition in the future.

Acknowledgments The authors thank Brian Smith, Yanan Fu, Vitaliy Gorbatyuk, Marie Phelan and Nicole Cheung for testing the tools with their CCPN NMR projects made using Analysis. We also acknowledge Wayne Boucher for programming help and Chris Spronk for help with the documentation pages. This project was funded by the UK Biotechnology and Biological Sciences Research Council (BBSRC) grant BBE0075111, with equipment support from The Wellcome Trust grant 088944. BMRB is supported by grant LM05799 from the US National Library of Medicine.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Berman HM, Westbrook JD, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242

- Berman H, Henrick K, Nakamura H, Markley JL (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res* 35:D301–D303
- Bourne PE, Berman HM, McMahon B, Watenpaugh KD, Westbrook JD, Fitzgerald PM (1997) Macromolecular crystallographic information file. *Methods Enzymol* 277:571–590
- Cavalli A, Salvatella X, Dobson CM, Vendruscolo M (2007) Protein structure determination from NMR chemical shifts. *Proc Natl Acad Sci USA* 104:9615–9620
- Doreleijers JF, Nederveen AJ, Vranken W, Lin J, Bonvin AM, Kaptein R, Markley JL, Ulrich EL (2005) BioMagResBank databases DOCR and FRED containing converted and filtered sets of experimental NMR restraints and coordinates from over 500 protein PDB structures. *J Biomol NMR* 32:1–12
- Doreleijers JF, Vranken WF, Schulte C, Lin J, Wedell JR, Penkett CJ, Vuister GW, Vriend G, Markley JL, Ulrich EL (2009) The NMR restraints grid at BMRB for 5, 266 protein and nucleic acid PDB entries. *J Biomol NMR* 45:389–396
- 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 (2002) The CCPN project: an interim report on a data model for the NMR community. *Nat Struct Biol* 9:416–418
- Fogh RH, Boucher W, Vranken WF, Pajon A, Stevens TJ, Bhat TN, Westbrook J, Ionides JM, Laue ED (2005) A framework for scientific data modeling and automated software development. *Bioinformatics* 21:1678–1684
- Fogh RH, Boucher W, Ionides JMC, Vranken WF, Stevens TJ, Laue ED (2010) MEMOPS: data modelling and automatic code generation. *J Integr Bioinform* 7:123. doi:10.2390/biecoll-jib-2010-123
- Hooft RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. *Nature* 381:272
- Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8:477–486
- Markley JL, Ulrich EL, Berman HM, Henrick K, Nakamura H, Akutsu H (2008) BioMagResBank (BMRB) as a partner in the Worldwide Protein Data Bank (wwPDB): new policies affecting biomolecular NMR depositions. *J Biomol NMR* 40:153–155
- Neal S, Nip AM, Zhang H, Wishart DS (2003) Rapid and accurate calculation of protein ^1H , ^{13}C and ^{15}N chemical shifts. *J Biomol NMR* 26:215–240
- Nederveen AJ, Doreleijers JF, Vranken W, Miller Z, Spronk CA, Nabuurs SB, Güntert P, Livny M, Markley JL, Nilges M, Ulrich EL, Kaptein R, Bonvin AM (2005) RECOORD: a recalculated coordinate database of 500 + proteins from the PDB using restraints from the BioMagResBank. *Proteins* 59:662–672
- Sen S, Van Ginkel G, Kapopoulou A, Sahni G, Swaminathan GJ, Newman RH, Velankar S, Henrick K (2007) Autodep 4.1: a web-based deposition and archival system. *Acta Cryst A* 63:s141
- Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu G, Eletsky A, Wu Y, Singarapu KK, Lemak A, Ignatchenko A, Arrowsmith CH, Szyperski T, Montelione GT, Baker D, Bax A (2008) Consistent blind protein structure generation from NMR chemical shift data. *Proc Natl Acad Sci USA* 105:4685–4690
- Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Mazziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Kent Wenger R, Yao H, Markley JL (2008) BioMagResBank. *Nucleic Acids Res* 36:D402–D408
- Velankar S, Best C, Beuth B, Boutselakis CH, Cobley N, Sousa Da Silva AW, Dimitropoulos D, Golovin A, Hirshberg M, John M, Krissinel EB, Newman R, Oldfield T, Pajon A, Penkett CJ, Pineda-Castillo J, Sahni G, Sen S, Slowley R, Suarez-Uruena A, Swaminathan J, van Ginkel G, Vranken WF, Henrick K, Kleywegt GJ (2010) PDBe: Protein Data Bank in Europe. *Nucleic Acids Res* 38:D308–D317
- Vranken W (2007) A global analysis of NMR distance constraints from the PDB. *J Biomol NMR* 39:303–314
- Vranken WF, Rieping W (2009) Relationship between chemical shift value and accessible surface area for all amino acid atoms. *BMC Struct Biol* 9:20
- Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59:687–696
- Wishart DS, Arndt D, Berjanskii M, Tang P, Zhou J, Lin G (2008) CS23D: a web server for rapid protein structure generation using NMR chemical shifts and sequence data. *Nucleic Acids Res* 36:W496–W502