The NMR Exchange Format – Review and Recommendations

Pedro Romero, Eldon L. Ulrich, and John L. Markley

BMRB, Biochemistry Department, University of Wisconsin-Madison

Madison WI 53706, USA

Author contact information:

PR: promero@wisc.edu

ELU: elu@bmrb.wisc.edu

JLM: jmarkley@wisc.edu

Introduction

The NMR Exchange Format (NEF) (Gutmanas et al. 2015) is an NMR community effort aimed at creating a format for data related to three-dimensional structures to be deposited in the PDB archive. The data format is to be read/write compatible with relevant software packages and with data formats used by the public databases (PDB and BMRB).

The development of community-based, open data exchange formats is a valuable endeavor. Nevertheless, any such effort should be scrutinized in view of available data formats used by that same community, if only to avoid the proliferation of essentially duplicate or, even worse, incompatible ontologies, which, rather than facilitating the interoperation of database and software resources, cause complications and ambiguities that must be resolved. A format intended to be used to deposit data at the public databases must be capable of handling the deposition of all data in the categories required by that database (for NMR: chemical shifts and structure constraints as mandated by the PDB).

This report is aimed at evaluating the current version of the NEF, as presented in its public design documents, from the point of view of its compatibility with the data format in use by the members of the Worldwide Protein Data Bank (wwPDB) for the deposition of structures determined by NMR spectroscopy and associated experimental data. These formats are NMR-STAR, which was developed by the Biological Magnetic Resonance

Bank (BMRB) in collaboration with many in the NMR community (participants in structural genomics consortia worldwide and individual NMR community members) and with PDBx/mmCIF. The documents analyzed here were those on the NEF Github page (https://github.com/NMRExchangeFormat/NEF/) as of 1 September 2015, which consisted of a "proposal" document and commented examples. This task has been somewhat complicated by the adoption by the NEF of tag labels that differ from those used by the public databases, and the lack of a formal NEF definition document.

NEF Goals and Specification Documents

NEF Goals

1. **Provide a format for NMR restraint data that can be used for deposition to the wwPDB and BMRB and can be validated against atomic coordinates**

Comments

- For deposition to the resources mentioned, it makes sense to use a molecular system hierarchy description that matches NMR-STAR and PDBx/mmCIF with useful extensions rather than the older and superseded PDB format, so that validation software can be consistent with NMR-STAR and PDBx/mmCIF. As implied by its draft specifications, NEF appears to more closely match the older PDB format, which risks causing mapping and translation problems when used for deposition.
- NEF does not appear to consider assigned chemical shift data as restraints, and does not provide links to the data used to derive the assigned chemical shifts as provided for other kinds of restraints.
- Given the use of wildcard atom name definitions, validation of the chemical shift data against standard chemical shift statistics or back calculated chemical shift values could be difficult, unless the naming conventions are strictly enforced and a conversion code has been developed.
- Because overlapping resonances can be assigned unambiguously, but may still be a cause for ambiguity in structural constraints, the two need to be defined independently.
- It is imperative that the format can be converted to NMR-STAR so that data relevant to BMRB can be incorporated into its archive.
- Because the NEF format does not address hybrid structure determinations, which are becoming common, it may become restrictive and obsolete before even hitting the ground. NMR-STAR already accommodates several types of hybrid data (SAXS, FRET, MS, EMR).

- More generally, NEF is limited to NMR data for structure determination, and it seems to be restricted to single chain structures, which is a drawback, especially for deposition purposes: By focusing on structure-relevant data, it risks leaving out data that should be deposited to BMRB.

2. Provide a format that supports data common to that exported by a variety of software applications, but also provides provisions for capturing software specific data items

Comments

The issue of a common data format arises, as described in (Gutmanas et al. 2015), because of difficulties in the deposition of NMR restraint data, and this is a valid concern with NMR-STAR. We at BMRB have incorporated a large number of tags (~50) proposed by the NEF team that deal with this issue into the NMR-STAR data model and data dictionary. This is, indeed, an example of community-driven improvements aimed at dealing with perceived shortcomings of the current data models.

As to the issue of capturing software-specific data items, a number of different approaches might be envisioned as additions to the data model:

- In some cases software-specific data items could be handled by including enumerations specific to individual software applications for general tags.
- Another mechanism would be to define the software application and list the parameter names and their values.
- In cases where a more complex schema is required to clearly define the information and its interrelationships, a set of software-application-specific data structures would need to be defined. This structure would include tables, individual tag items, and methods for creating the proper relationships.

The NEF specification avoids the complications inherent to these approaches by simply allowing for each software developer to define their own software-specific namespaces, which means they are *not* part of the data format general specification, but simply “passed on” during data transfer. This solution, also available in NMR-STAR, allows for the addition of such important and unique but unspecified data to any database entry, without the need for enlarging the data model. Indeed, this solution has several advantages:

- No need to define namespaces and tags within the format specification to accommodate every item relevant to all current software applications.
- Information relevant to each software program would be kept in the most appropriate manner, as designed by the software developers themselves.

- Changes in software data formats would not have an effect on the BMRB or PDBx data models: the software developers would make sure older and newer data remain compatible with their software, and would implement any modifications to this effect in their own parsers, without affecting database records and/or other software data.
- As the NMR field and software applications advance, the data model specification can be updated and extended, as appropriate, on the basis of the information deposited using the above approaches.

We agree with the NEF developers that this is a correct and acceptable way of dealing with software-specific data. Moreover, because this approach can be handled within NMR-STAR, it can be compatible with the current BMRB archive.

NEF Proposal Document

The proposal document starts with a section on general issues that describe how different information and data are to be expressed in NEF as a subset of STAR. The main theme in this description is the flexibility allowed to the user. Some format decisions stray from the current PDBx/mmCIF standard, returning to "PDB-like" definitions that reintroduce inherent problems, which were resolved in NMR-STAR and PDBx/mmCIF. Some examples are listed below:

1. Residue and atom identifying strings are defined as case-sensitive, but only atom and standard residues are restricted to be all uppercase. Lower case is allowed for other codes, for which the use of upper-case is "recommended." This lack of a standardization, will necessitate the development of specialized parsers.
2. Sequence_code is defined to be a string to accommodate *any* so-called 'alt_codes' as suffixes (example: '127B'). No definition is provided for the meaning of alt_codes (the example seems to show an insertion code), and this appears to be left to the users to define. In NMR-STAR/PDBx/mmCIF, sequence_codes are integers and the sequence starts with '1' and increments sequentially; any defined 'alt code' or 'insertion code' has its own tag, so that the sequence codes are always sequential lists of integers. The use of program-specific numberings in a deposition will necessitate the definition of and inclusion of a deposition-specific namespace (an example of the extra meta data mentioned above).
3. Wildcards are used to represent atom sets. Stereochemical suffixes/wildcards are given a different meaning than in NMR-STAR. Both issues represent potential problems for format mapping, both with NMR-STAR and PDBx/mmCIF. By encoding chemical shift ambiguity in the atom nomenclature and not providing for additional chemical shift ambiguity annotation, chemical shift assignments that cannot be annotated properly will be left out of the chemical shift tables exported by the software and will not be

deposited in the public databases. Reporting of ambiguity in assignments for non-geminal atoms, including protein backbone atoms, does not appear possible with the current NEF specification. In NMR-STAR these intraresidue, interresidue, and inter-molecular ambiguities are represented, respectively, by ambiguity codes '4' (as currently used in 12,110 reports), '5' (3,911 reports), and '6' (1,020 reports).

4. "Field values and data types" states that, "Unlike NMR-STAR, the dollar sign ('$') is *not* used as a prefix to indicate a saveframe identifier." This is not an NMR_STAR convention, but a STAR one, and the change would make NEF non-STAR compliant. However, this idiosyncrasy of NEF should not cause problems if implemented correctly in software parsers.
5. The proposed sequence linking values (e.g., 'start', 'end', 'middle', 'cyclic', 'break', 'single', 'dummy') seem redundant, as bonding information makes them unnecessary.
6. As defined in the NEF proposal document, multiple peak lists can share a master shift list. More information is needed to make sense of this definition, as it is unclear how it is to be done. Is the idea that the master list will represent shift ranges and averages instead of unique values?

NEF Draft Example Files

The examples point to the need for several useful tags, in particular ones defining restraints that have been incorporated into the NEF but are missing from the NMR-STAR dictionary. These findings underscore the importance of community feedback as a mechanism for the improvement of current NMR information formats. In response, the BMRB has incorporated these into its dictionary (see Table 1 for a list of new tags incorporated to NMR-STAR). In other cases, the attempt to map tags between NEF and NMR-STAR uncovered ambiguities (noted in Table 1); some of these may simply reflect incomplete format specifications in NEF draft 0.8. It will be necessary to resolve these ambiguities to a one-to-one correspondence for the NEF to be usable.

Mapping of NEF to NMR-STAR

We have carried out a preliminary mapping of the available NEF specification to NMR-STAR (Table 1). This has led to the adoption of NEF-defined information into the NMR-STAR specification. We were unable to perform a more comprehensive mapping, owing to the lack of a more comprehensive NEF data dictionary.

Issues Identified

1. The current NEF specification allows for chemical shift data that cannot not be validated with existing software tools (e.g., chemical shifts reported in Hz without the specific field strength defined). It will be

necessary to convert atom nomenclature to the IUPAC conventions used by the wwPDB to validate chemical shift data.

2. Residue and atom numbering is an important and delicate issue that needs to be resolved. NMR-STAR handles up to four different counting schemas, conforming to PDBx/mmCIF sequential approach, while retaining PDB and author-specific numbering. This already imposes the use of extra qualifying tags for ambiguity, insertions and other numbering complications. It is recommended that the NEF, and, by extension, software packages, conform to this approach so as to avoid the addition of an extra layer of complication when mapping the data for deposition.
3. Some differences between NEF and NMR-STAR prevent proper mapping between formats:
    a. The current NEF cannot distinguish 1H chemical shifts from 2H chemical shifts, as can be done in NMR-STAR.
    b. The current NEF does not have a specification for defining isotope labeling. This is important for calibrating chemical shift values.
    c. The current NEF does not allow for the definition of chemical shift offsets caused by experimental design (TOCSY offsets for example)
    d. There is currently no provision for specifying the chemical shift referencing used. Is the intent that the user would enter this information at the time of deposition?
    e. The NEF includes a tag for chemical shift units in the consensus chemical shift table. Will the value to the tag always be 'ppm'? If the chemical shift data in the consensus chemical shift save frame are reported in Hz or data points, then tags for the spectrometer frequency for each nucleus (^{1}H, ^{13}C, ^{15}N, etc.) and possibly other details (sweep width and explicit carrier frequencies) will be needed to convert the chemical shifts to ppm for deposition and validation.
    f. The use of the term 'ordinal' implies that there is a specific order to the items in a set. This is not the indicated intent of many such named fields in NEF. An alternative term might be 'index' or 'count' that simply indicates a count or list of the items in the set (rows or lines), but not a specific order.
    g. The descriptions in the meta data save frame for software are unclear. In the loop where software, scripts, and parameters are all modeled, the tag for the software parameter is software-specific, 'cyana…' although there is already a tag that specifies the software name in the table row. Also, is it intended that the software name, the script name, and the script itself must be repeated for every parameter defined or are additional tags available to extend the loop when there are additional

parameters? If the latter is the case, is there a need for a tag that defines the parameter name in addition to the tag that defines the parameter value?

h. It is not clear why there are data in the 'software history loop' different from those in the 'nef_program_script' loop.

Future role of the NEF in the Worldwide Protein Data Bank (wwPDB)

The goal of the NEF is to provide a system for exchanging information between software packages and to provide a data format for deposition in the PDB and BMRB. The focus was to be on restraints, which currently are difficult to capture from different software packages. As such, it is assumed that other data items will continue to be represented in the NMR-STAR / PDBx formats currently used in the archives. There would appear to be two ways of handling the problem: one way is to develop a data format and dictionary for NMR restraints that is read-write compatible with NMR-STAR / PDBx, and the other is to develop tags and definitions to be added to an expanded NMR-STAR / PDBx.

If NEF is to be a separate deposition format, three main issues need to be resolved for it to become a viable exchange format that complies with its stated goals:

1. The format needs to be brought up to date with the current wwPDB standard, that is, PDBx/mmCIF. This includes being up to date with new types of NMR experiments and their data (e.g., hybrid structure determination experiments). This will facilitate deposition to both PDB and BMRB.
2. Specific and unambiguous definitions need to be developed, in which format flexibility is constrained to work within currently accepted parameters, thus avoiding mapping and translation problems.
3. The NEF must be fully compatible with the information handled by the public repositories, allowing for unambiguous mapping to the formats used by the public databases (i.e., NMR-STAR for BMRB, PDBx/mmCIF for PDB). Any current incompatibilities, such as the ones pointed out herein, must be worked out in the format's final definition.

Arguments can be advanced in favor of the second approach of making the NEF a subset of NMR-STAR. The data formats in current use by NMR-STAR represent the widely accepted ontology for NMR experimental data, including chemical shifts and many structural restraints. Furthermore, the NMR-STAR dictionary tag names are based on the PDBx/mmCIF standard defined by the PDB. NMR-STAR has evolved to facilitate the exchange of meta and quantitative data that meet the requirements of accuracy, clarity, coherence, completeness, and reproducibility. Meeting these criteria often requires complexity, and it would be a mistake to go backward to

a less stringent format. Thus, the NMR community—users, resources and software developers—may be better served by analyzing the current ontologies and correcting and extending them as required to support the needs of software developers. A possible –and recommended– approach would be to use PDBx/NMR-STAR tags for those data items where mapping is unambiguous. Tags that cannot be mapped unambiguously should be assessed for need and added, as appropriate, to NMR-STAR and/or PDBx. All tags in the NEF, whether NMR-STAR/PDBx based or newly defined, would be assembled by prepending to them the prefix "_NEF". This way, the exchange format would be guaranteed to consist of a proper subset of the current data models used by PDB and BMRB, and conversion would be reduced to removing the _NEF prefix and checking for the existence of the remaining tag in NMR-STAR/mmCIF. Any extra tags (i.e., those belonging to software-specific namespaces) would be saved and accessed "as is" without modification.

Whichever path is chosen by the NEF team, we believe that the BMRB and other branches of the wwPDB need to be involved in helping guide the future course of this worthy endeavor. We welcome further communication on these issues.

Reference

1. Gutmanas A et al. (2015) NMR Exchange Format: a unified and open standard for representation of NMR restraint data Nat Struct Mol Biol 22:433-434 doi:10.1038/nsmb.3041

Table 1

NEF Example File Analysis – NMR-STAR Correspondence

The following table shows a commented mapping analysis of NEF example files: The suggested NMR-STAR equivalent is shown for different NEF tags, along with the recommended NMR-STAR/PDBx match and appropriate comments. Tags have been grouped within save frames or loops to facilitate context visualization. Additions to NMR-STAR are indicated in the comments column.

NEF tags	NMR-STAR equivalent	Recommended NMR-STAR/PDBx match	Comments
save_XXX			
_nef_nmr_meta_data.sf_category	_Entry.Sf_category	_Entry.Sf_category	
_nef_nmr_meta_data.sf_framecode	_Entry.Sf_framecode	_Entry.Sf_framecode	
_nef_nmr_meta_data.format_name	_Entry.Format_name	_Entry.Format_name	
_nef_nmr_meta_data.format_version	_Entry.NMR_STAR_version	_Entry.NMR_STAR_version	
_nef_nmr_meta_data.program_name	_Entry.Creation_software_label	_Entry.Generated_software_label	Added to NMR-STAR
	_Entry.Creation_software_ID	_Entry.Generated_software_ID	Added to NMR-STAR
_nef_nmr_meta_data.program_version	_Entry.Creation_software_version	_Entry.Generated_software_version	Added to NMR-STAR
_nef_nmr_meta_data.creation_date	_Entry.Creation_date	_Entry.Generated_date	Added to NMR-STAR
_nef_nmr_meta_data.uuid	_Entry.UUID	_Entry.UUID	Added to NMR-STAR
_nef_nmr_meta_data.coordinate_file_name	_Conformer_family_coord_set.File_name	_Entry.Conformer_family_coord_set_file_name	Added to NMR-STAR
loop_			
_nef_related_entries.database_name	_Related_entries.Database_name	_Related_entries.Database_name	
_nef_related_entries.database_accession_code	_Related_entries.Database_accession_code	_Related_entries.Database_accession_code	
loop_			NMR-STAR and PDBx have models for capturing information about the software and methods used to create the experimental data found in a file. Further discussions are needed to understand the intent of this loop and to determine if a data structure matching this loop is needed in NMR-STAR.
_nef_program_script.program_name	_Conformer_family_software.Software_label		
	_Conformer_family_software.ID		
_nef_program_script.script_name	_Conformer_family_software.Method_label		
	_Conformer_family_software.Method_ID		
_nef_program_script.script	_Method_file.Text		
_nef_program_script.cyana_parameter_1			
loop_			As this information could become quite extensive, it is recommended to create a new save frame to be located at the end of the file. The NEF example indicates that this loop could contain information for many programs and scripts starting with the NMR experiment pulse programs and extending to at least the structure calculations.
_nef_run_history.run_ordinal			
_nef_run_history.program_name			
_nef_run_history.program_version			
_nef_run_history.script_name			
_nef_run_history.script			
save_			
save_cyana_additional_data_1			Domain defined save frames and tags are supported by NMR-STAR.
_cyana_additional_data.sf_category			
_cyana_additional_data.sf_framecode			
_cyana_additional_data.special_version			
save_			

NEF tags	NMR-STAR equivalent	Recommended NMR-STAR/PDBx match	Comments
save_nef_molecular_system			Equivalent information is in the NMR-STAR Assembly save frame
_nef_molecular_system.sf_category	_Assembly.Sf_category	_Assembly.Sf_category	
_nef_molecular_system.sf_framecode	_Assembly.Sf_framecode	_Assembly.Sf_framecode	
loop_			Currently, NEF does not appear to provide for the complete description (atoms and bonds) for a ligand or non-standard residue not present in existing PDB or BMRB libraries.
_nef_sequence.chain_code	_Chem_comp_assembly.Asym_ID	_Chem_comp_assembly.Entity_assembly_ID or _Chem_comp_assembly.Asym_ID (_PDBX_poly_seq_scheme.Asym_ID)	
_nef_sequence.sequence_code	_Chem_comp_assembly.Seq_ID	_Chem_comp_assembly.Comp_index_ID or _Chem_comp_assembly.Seq_ID (_PDBX_poly_seq_scheme.Seq_ID) _PDBX_poly_seq_scheme.PDB_ins_code	The NEF 'sequence_code' appears to allow for insertion codes. In mapping these insertion codes to NMR-STAR/PDBx the sequences may need to be renumbered to be mapped properly to 'Seq_ID'.
_nef_sequence.residue_type	_Chem_comp_assembly.Comp_ID	_Chem_comp_assembly.Comp_ID (_PDBX_poly_seq_scheme.Comp_ID)	
_nef_sequence.linking		_Chem_comp_assembly.Linking	Add this to NMR-STAR
_nef_sequence.residue_variant		_Chem_comp_assembly.Comp_variant_ID (_PDBX_poly_seq_scheme.Variant_ID)	Add this to NMR-STAR PDBX tag would need to be added to the pdbx dictionary?
loop_			Recommend including a loop for defining deleted atom when cross links are formed, as the leaving atoms may be ambiguous (stereospecific).
_nef_covalent_links.chain_code_1	_Bond.Entity_assembly_ID_1	_Bond.Entity_assembly_ID_1 or _Bond.Asym_ID_1	Review PDBx/mmCIF for analogous tags.
_nef_covalent_links.sequence_code_1	_Bond.Comp_index_ID_1	_Bond.Comp_index_ID_1 or _Bond.Seq_ID_1	
_nef_covalent_links.residue_type_1	_Bond.Comp_ID_1	_Bond.Comp_ID_1	
_nef_covalent_links.atom_name_1	_Bond.Atom_ID_1	_Bond.Atom_ID_1	
_nef_covalent_links.chain_code_2	_Bond.Entity_assembly_ID_2	_Bond.Entity_assembly_ID_2 or _Bond.Asym_ID_2	
_nef_covalent_links.sequence_code_2	_Bond.Comp_index_ID_2	_Bond.Comp_index_ID_2 or _Bond.Seq_ID_2	
_nef_covalent_links.residue_type_2	_Bond.Comp_ID_2	_Bond.Comp_ID_2	
_nef_covalent_links.atom_name_2	_Bond.Atom_ID_2	_Bond.Atom_ID_2	
		_Bond.Atom_stereo_config_1	Recommended addition to NEF
		_Bond.Atom_stereo_config_2	Recommended addition to NEF
		_Bond.Type	Recommended addition to NEF
		_Bond.Value_order	Recommended addition to NEF to distinguish single from double, etc.
save_nef_chemical_shift_list_1			Chemical shift referencing is not included in NEF, at this time, and could be important for validation.
_nef_chemical_shift_list.sf_category	_Assigned_chem_shift_list.Sf_category	_Assigned_chem_shift_list.Sf_category	
_nef_chemical_shift_list.sf_framecode	_Assigned_chem_shift_list.Sf_framecode	_Assigned_chem_shift_list.Sf_framecode	
_nef_chemical_shift_list.atom_chem_shift_units	_Assigned_chem_shift_list.Chem_shift_units	_Assigned_chem_shift_list.Chem_shift_units	Added to NMR-STAR (additional information required if the value is not ‘ppm’)

NEF tags	NMR-STAR equivalent	Recommended NMR-STAR/PDBx match	Comments
loop_			
_nef_chemical_shift.chain_code	_Atom_chem_shift.PDB_strand_ID	_Atom_chem_shift.Entity_assembly_ID or _Atom_chem_shift.PDB_strand_ID	
_nef_chemical_shift.sequence_code	_Atom_chem_shift.PDB_residue_no	_Atom_chem_shift.Comp_index_ID or _Atom_chem_shift.Seq_ID	
_nef_chemical_shift.residue_type	_Atom_chem_shift.PDB_residue_name	_Atom_chem_shift.Comp_ID	
_nef_chemical_shift.atom_name	_Atom_chem_shift.PDB_atom_name	_Atom_chem_shift.Atom_ID	
_nef_chemical_shift.value	_Atom_chem_shift.Val	_Atom_chem_shift.Val	
_nef_chemical_shift.value_uncertainty	_Atom_chem_shift.Val_err	_Atom_chem_shift.Val_err	
		_Atom_chem_shift.Atom_type	Recommended addition to NEF to distinguish a C-alpha from a calcium atom
		_Atom_chem_shift.Atom_isotope_number	Recommended addition to NEF to distinguish 1H from 2H
save_			
save_nef_distance_restraint_list_L1			
_nef_distance_restraint_list.sf_category	_Gen_dist_constraint_list.Sf_category	_Gen_dist_constraint_list.Sf_category	
_nef_distance_restraint_list.sf_framecode	_Gen_dist_constraint_list.Sf_framecode	_Gen_dist_constraint_list.Sf_framecode	
_nef_distance_restraint_list.potential_type	_Gen_dist_constraint_list.Potential_type	_Gen_dist_constraint_list.Potential_type	Added to NMR-STAR
_nef_distance_restraint_list.restraint_origin	_Gen_dist_constraint_list.Constraint_type	_Gen_dist_constraint_list.Constraint_type	
loop_			
_nef_distance_restraint.ordinal	_Gen_dist_constraint.Row_ID	_Gen_dist_constraint.Row_ID	Added to NMR-STAR
_nef_distance_restraint.restraint_id	_Gen_dist_constraint.ID	_Gen_dist_constraint.ID	
_nef_distance_restraint.restraint_combination_id	_Gen_dist_constraint.Set_ID	_Gen_dist_constraint.Set_ID	Added to NMR-STAR
_nef_distance_restraint.chain_code_1	_Gen_dist_constraint.PDB_strand_ID_1	_Gen_dist_constraint.Entity_assembly_ID_1	
_nef_distance_restraint.sequence_code_1	_Gen_dist_constraint.PDB_residue_no_1	_Gen_dist_constraint.Comp_index_ID_1 or _Gen_dist_constraint.Seq_ID_1	
_nef_distance_restraint.residue_type_1	_Gen_dist_constraint.PDB_residue_name_1	_Gen_dist_constraint.Comp_ID_1	
_nef_distance_restraint.atom_name_1	_Gen_dist_constraint.PDB_atom_name_1	_Gen_dist_constraint.Atom_ID_1	
_nef_distance_restraint.chain_code_2	_Gen_dist_constraint.PDB_strand_ID_2	_Gen_dist_constraint.Entity_assembly_ID_2	
_nef_distance_restraint.sequence_code_2	_Gen_dist_constraint.PDB_residue_no_2	_Gen_dist_constraint.Comp_index_ID_2 or _Gen_dist_constraint.Seq_ID_2	
_nef_distance_restraint.residue_type_2	_Gen_dist_constraint.PDB_residue_name_2	_Gen_dist_constraint.Comp_ID_2	
_nef_distance_restraint.atom_name_2	_Gen_dist_constraint.PDB_atom_name_2	_Gen_dist_constraint.Atom_ID_2	
_nef_distance_restraint.weight	_Gen_dist_constraint.Weight	_Gen_dist_constraint.Weight	Added to NMR-STAR
_nef_distance_restraint.target_value	_Gen_dist_constraint.Distance_val	_Gen_dist_constraint.Distance_val	If not a good match, add target value to NMR-STAR
_nef_distance_restraint.target_value_uncertainty	_Gen_dist_constraint.Distance_val_err	_Gen_dist_constraint.Distance_val_err	Added to NMR-STAR
_nef_distance_restraint.lower_linear_limit	_Gen_dist_constraint.Lower_linear_limit	_Gen_dist_constraint.Lower_linear_limit	Added to NMR-STAR
_nef_distance_restraint.lower_limit	_Gen_dist_constraint.Distance_lower_bound_val	_Gen_dist_constraint.Distance_lower_bound_val	
_nef_distance_restraint.upper_limit	_Gen_dist_constraint.Distance_upper_bound_val	_Gen_dist_constraint.Distance_upper_bound_val	
_nef_distance_restraint.upper_linear_limit	_Gen_dist_constraint.Upper_linear_limit	_Gen_dist_constraint.Upper_linear_limit	Added to NMR-STAR
		_Gen_dist_constraint.Member_logic_code	Recommend adding explicit logic code to NEF
save_			

NEF tags	NMR-STAR equivalent	Recommended NMR-STAR/PDBx match	Comments
save_nef_dihedral_restraint_list_L2			
_nef_dihedral_restraint_list.sf_category	_Torsion_angle_constraint_list.Sf_category	_Torsion_angle_constraint_list.Sf_category	
_nef_dihedral_restraint_list.sf_framecode	_Torsion_angle_constraint_list.Sf_framecode	_Torsion_angle_constraint_list.Sf_framecode	
_nef_dihedral_restraint_list.potential_type	_Torsion_angle_constraint_list.Potential_type	_Torsion_angle_constraint_list.Potential_type	Added to NMR-STAR
_nef_dihedral_restraint_list.restraint_origin	_Torsion_angle_constraint_list.Constraint_origin	_Torsion_angle_constraint_list.Constraint_origin	Added to NMR-STAR - Recommend the enumeration for the tag be analogous to the values used for distance constraints (i.e., chemical shifts, coupling constants) and not a software program like TALOS as in the NEF example
loop_			
_nef_dihedral_restraint.ordinal	_Torsion_angle_constraint.Row_ID	_Torsion_angle_constraint.Row_ID	Added to NMR-STAR
_nef_dihedral_restraint.restraint_id	_Torsion_angle_constraint.ID	_Torsion_angle_constraint.ID	
_nef_dihedral_restraint.restraint_combination_id	_Torsion_angle_constraint.Set_ID	_Torsion_angle_constraint.Set_ID	
_nef_dihedral_restraint.chain_code_1	_Torsion_angle_constraint.PDB_strand_ID_1	_Torsion_angle_constraint.Entity_assembly_ID_1 or _Torsion_angle_constraint.Asym_ID_1	
_nef_dihedral_restraint.sequence_code_1	_Torsion_angle_constraint.PDB_residue_no_1	_Torsion_angle_constraint.Comp_index_ID_1 or _Torsion_angle_constraint.Seq_ID_1	
_nef_dihedral_restraint.residue_type_1	_Torsion_angle_constraint.PDB_residue_name_1	_Torsion_angle_constraint.Comp_ID_1	
_nef_dihedral_restraint.atom_name_1	_Torsion_angle_constraint.PDB_atom_name_1	_Torsion_angle_constraint.Atom_ID_1	
_nef_dihedral_restraint.chain_code_2	_Torsion_angle_constraint.PDB_strand_ID_2	_Torsion_angle_constraint.Entity_assembly_ID_2	
_nef_dihedral_restraint.sequence_code_2	_Torsion_angle_constraint.PDB_residue_no_2	_Torsion_angle_constraint.Comp_index_ID_2 _Torsion_angle_constraint.Seq_ID_2	
_nef_dihedral_restraint.residue_type_2	_Torsion_angle_constraint.PDB_residue_name_2	_Torsion_angle_constraint.Comp_ID_2	
_nef_dihedral_restraint.atom_name_2	_Torsion_angle_constraint.PDB_atom_name_2	_Torsion_angle_constraint.Atom_ID_2	
_nef_dihedral_restraint.chain_code_3	_Torsion_angle_constraint.PDB_strand_ID_3	_Torsion_angle_constraint.Entity_assembly_ID_3	
_nef_dihedral_restraint.sequence_code_3	_Torsion_angle_constraint.PDB_residue_no_3	_Torsion_angle_constraint.Comp_index_ID_3 _Torsion_angle_constraint.Seq_ID_3	
_nef_dihedral_restraint.residue_type_3	_Torsion_angle_constraint.PDB_residue_name_3	_Torsion_angle_constraint.Comp_ID_3	
_nef_dihedral_restraint.atom_name_3	_Torsion_angle_constraint.PDB_atom_name_3	_Torsion_angle_constraint.Atom_ID_3	
_nef_dihedral_restraint.chain_code_4	_Torsion_angle_constraint.PDB_strand_ID_4	_Torsion_angle_constraint.Entity_assembly_ID_4	
_nef_dihedral_restraint.sequence_code_4	_Torsion_angle_constraint.PDB_residue_no_4	_Torsion_angle_constraint.Comp_index_ID_4 _Torsion_angle_constraint.Seq_ID_4	
_nef_dihedral_restraint.residue_type_4	_Torsion_angle_constraint.PDB_residue_name_4	_Torsion_angle_constraint.Comp_ID_4	
_nef_dihedral_restraint.atom_name_4	_Torsion_angle_constraint.PDB_atom_name_4	_Torsion_angle_constraint.Atom_ID_4	
_nef_dihedral_restraint.weight	_Torsion_angle_constraint.Weight	_Torsion_angle_constraint.Weight	Added to NMR-STAR
_nef_dihedral_restraint.target_value	_Torsion_angle_constraint.Angle_target_value	_Torsion_angle_constraint.Angle_target_value	Added to NMR-STAR
_nef_dihedral_restraint.target_value_uncertainty	_Torsion_angle_constraint.Angle_target_value_err	_Torsion_angle_constraint.Angle_target_value_err	Added to NMR-STAR
_nef_dihedral_restraint.lower_linear_limit	_Torsion_angle_constraint.Angle_lower_linear_limit	_Torsion_angle_constraint.Angle_lower_linear_limit	Added to NMR-STAR
_nef_dihedral_restraint.lower_limit	_Torsion_angle_constraint.Angle_lower_bound_val	_Torsion_angle_constraint.Angle_lower_bound_val	
_nef_dihedral_restraint.upper_limit	_Torsion_angle_constraint.Angle_upper_bound_val	_Torsion_angle_constraint.Angle_upper_bound_val	
_nef_dihedral_restraint.upper_linear_limit	_Torsion_angle_constraint.Angle_upper_linear_limit	_Torsion_angle_constraint.Angle_upper_linear_limit	Added to NMR-STAR
_nef_dihedral_restraint.name	_Torsion_angle_constraint.Torsion_angle_name	_Torsion_angle_constraint.Torsion_angle_name	
save_			

NEF tags	NMR-STAR equivalent	Recommended NMR-STAR/PDBx match	Comments
save_nef_rdc_restraint_list_1			
_nef_rdc_restraint_list.sf_category	_RDC_constraint_list.Sf_category	_RDC_constraint_list.Sf_category	
_nef_rdc_restraint_list.sf_framecode	_RDC_constraint_list.Sf_framecode	_RDC_constraint_list.Sf_framecode	
_nef_rdc_restraint_list.potential_type	_RDC_constraint_list.Potential_type	_RDC_constraint_list.Potential_type	Added to NMR-STAR
_nef_rdc_restraint_list.restraint_origin	_RDC_constraint_list.Constraint_origin	_RDC_constraint_list.Constraint_origin	Added to NMR-STAR
_nef_rdc_restraint_list.tensor_magnitude	_RDC_constraint_list.Tensor_magnitude	_RDC_constraint_list.Tensor_magnitude	Added to NMR-STAR
_nef_rdc_restraint_list.tensor_rhombicity	_RDC_constraint_list.Tensor_rhombicity	_RDC_constraint_list.Tensor_rhombicity	Added to NMR-STAR
_nef_rdc_restraint_list.tensor_chain_code	_RDC_constraint_list.Tensor_PDB_strand_ID	_RDC_constraint_list.Tensor_entity_assembly_ID or _RDC_constraint_list.Tensor_asym_ID	Added to NMR-STAR
_nef_rdc_restraint_list.tensor_sequence_code	_RDC_constraint_list.Tensor_PDB_residue_no	_RDC_constraint_list.Tensor_Comp_index_ID or _RDC_constraint_list.Tensor_Seq_ID	Added to NMR-STAR
_nef_rdc_restraint_list.tensor_residue_type	_RDC_constraint_list.Tensor_PDB_residue_name	_RDC_constraint_list.Tensor_Comp_ID	Added to NMR-STAR Review additional tags that exist in NMR-STAR for incorporation into NEF
loop_			
_nef_rdc_restraint.ordinal	_RDC_constraint.Row_ID	_RDC_constraint.Row_ID	Added to NMR-STAR
_nef_rdc_restraint.restraint_id	_RDC_constraint.ID	_RDC_constraint.ID	
_nef_rdc_restraint.restraint_combination_id	_RDC_constraint.Set_ID	_RDC_constraint.Set_ID	Added to NMR-STAR
_nef_rdc_restraint.chain_code_1	_RDC_constraint.PDB_strand_ID_1	_RDC_constraint.Entity_assembly_ID_1 or _RDC_constraint.Asym_ID_1	
_nef_rdc_restraint.sequence_code_1	_RDC_constraint.PDB_residue_no_1	_RDC_constraint.Comp_index_ID_1 or _RDC_constraint.Seq_ID_1	
_nef_rdc_restraint.residue_type_1	_RDC_constraint.PDB_residue_name_1	_RDC_constraint.Comp_ID_1	
_nef_rdc_restraint.atom_name_1	_RDC_constraint.PDB_atom_name_1	_RDC_constraint.Atom_ID_1	
_nef_rdc_restraint.chain_code_2	_RDC_constraint.PDB_strand_ID_2	_RDC_constraint.Entity_assembly_ID_2	
_nef_rdc_restraint.sequence_code_2	_RDC_constraint.PDB_residue_no_2	_RDC_constraint.Comp_index_ID_2 _RDC_constraint.Seq_ID_2	
_nef_rdc_restraint.residue_type_2	_RDC_constraint.PDB_residue_name_2	_RDC_constraint.Comp_ID_2	
_nef_rdc_restraint.atom_name_2	_RDC_constraint.PDB_atom_name_2	_RDC_constraint.Atom_ID_2	
_nef_rdc_restraint.weight	_RDC_constraint.Weight	_RDC_constraint.Weight	Added to NMR-STAR
_nef_rdc_restraint.target_value	_RDC_constraint.RDC_val	_RDC_constraint.RDC_val	
_nef_rdc_restraint.target_value_uncertainty	_RDC_constraint.RDC_val_err	_RDC_constraint.RDC_val_err	
_nef_rdc_restraint.lower_linear_limit	_RDC_constraint.RDC_lower_linear_limit	_RDC_constraint.RDC_lower_linear_limit	Added to NMR-STAR
_nef_rdc_restraint.lower_limit	_RDC_constraint.RDC_lower_bound	_RDC_constraint.RDC_lower_bound	
_nef_rdc_restraint.upper_limit	_RDC_constraint.RDC_upper_bound	_RDC_constraint.RDC_upper_bound	
_nef_rdc_restraint.upper_linear_limit	_RDC_constraint.RDC_upper_linear_limit	_RDC_constraint.RDC_upper_linear_limit	Added to NMR-STAR
_nef_rdc_restraint.scale	_RDC_constraint.RDC_val_scale_factor	_RDC_constraint.RDC_val_scale_factor	
_nef_rdc_restraint.distance_dependent	_RDC_constraint.RDC_bond_length	_RDC_constraint.RDC_bond_length	This may not be the correct match. If not, a tag similar to the NEF tag will be added to NMR-STAR
save_			

NEF tags	NMR-STAR equivalent	Recommended NMR-STAR/PDBx match	Comments
save_nef_nmr_spectrum_cnoesy1			
_nef_nmr_spectrum.sf_category	_Spectral_peak_list.Sf_category	_Spectral_peak_list.Sf_category	
_nef_nmr_spectrum.sf_framecode	_Spectral_peak_list.Sf_framecode	_Spectral_peak_list.Sf_framecode	
_nef_nmr_spectrum.num_dimensions	_Spectral_peak_list.Number_of_spectral_dimensions	_Spectral_peak_list.Number_of_spectral_dimensions	
_nef_nmr_spectrum.chemical_shift_list			Pointer to assigned chemical shift data save frame - in NMR-STAR this info is in the assignment loop in the peak list save frame
_nef_nmr_spectrum.experiment_classification	_Spectral_peak_list.Experiment_classification		This information is captured in the NMR-STAR NMR_spec_ext save frame This tag could be added to NMR-STAR
_nef_nmr_spectrum.experiment_type	_Spectral_peak_list.Experiment_type		This information is captured in the NMR-STAR NMR_spec_ext save frame This tag could be added to NMR-STAR
loop_			
_nef_spectrum_dimension.dimension_id	_Spectral_dim.ID	_Spectral_dim.ID	
_nef_spectrum_dimension.axis_unit	_Spectral_dim.Sweep_width_units	_Spectral_dim.Sweep_width_units	
_nef_spectrum_dimension.axis_code	_Spectral_dim.Spectral_region	_Spectral_dim.Spectral_region	
_nef_spectrum_dimension.spectrometer_frequency	_Spectral_dim.Spectrometer_frequency	_Spectral_dim.Spectrometer_frequency	Added to NMR-STAR
_nef_spectrum_dimension.spectral_width	_Spectral_dim.Sweep_width	_Spectral_dim.Sweep_width	
_nef_spectrum_dimension.value_first_point	_Spectral_dim.Value_first_point	_Spectral_dim.Value_first_point	Added to NMR-STAR
_nef_spectrum_dimension.folding	_Spectral_dim.Under_sampling_type	_Spectral_dim.Under_sampling_type	
_nef_spectrum_dimension.absolute_peak_positions	_Spectral_dim.Absolute_peak_positions	_Spectral_dim.Absolute_peak_positions	Added to NMR-STAR
_nef_spectrum_dimension.is_acquisition	_Spectral_dim.Aquisition_dimension	_Spectral_dim.Aquisition_dimension	Added to NMR-STAR
loop_			In NMR-STAR the information in this loop is captured with tags that are found in a loop similar to the one above in the NEF example. This loop is needed if information transfer into a dimension can come from more than one dimension.
_nef_spectrum_dimension_transfer.dimension_1			
_nef_spectrum_dimension_transfer.dimension_2	_Spectral_dim.Encoded_source_dimension_ID		
_nef_spectrum_dimension_transfer.transfer_type	_Spectral_dim.Encoding_code		
_nef_spectrum_dimension_transfer.is_indirect	_Spectral_dim.Encoding_indirect		Added to NMR-STAR

NEF tags	NMR-STAR equivalent	Recommended NMR-STAR/PDBx match	Comments
loop_			In NMR-STAR the info in this NEF loop is captured in four tables. This allows for more flexibility in describing the data and maintains the tables with fixed widths relative to the number of dimensions and the method used to measure the peak intensities. In NMR-STAR the peak characteristics are separated from the assignments so that information does not have to be repeated. If the community requests, the NMR-STAR peak tables can be remodeled as described in NEF.
_nef_peak.ordinal			
_nef_peak.peak_id	_Peak_general_char.Peak_ID		
_nef_peak.volume	_Peak_general_char.Intensity_val		
_nef_peak.volume_uncertainty	_Peak_general_char.Intensity_val_err		
_nef_peak.height			
_nef_peak.height_uncertainty			Recommended that NEF include tags for other peak characteristics as found in NMR-STAR (line-width, phase, etc.).
_nef_peak.position_1	_Peak_char.Spectral_dim_ID _Peak_char.Chem_shift_val		These three tags are used in NMR-STAR to record the info found in the NEF tags for chemical shift and uncertainty. In NEF 'position_1' is assumed to map to dimension 1 in the spectrum. In NMR-STAR this assumption is not made and the peak data is explicitly assigned to a spectral dimension.
_nef_peak.position_uncertainty_1	_Peak_char.Chem_shift_val_err		
_nef_peak.position_2			NEF lacks the tags to map these rows to specific rows in the chemical shift assignment table referenced by the tag at the top of this save frame
_nef_peak.position_uncertainty_2			
_nef_peak.position_3			
_nef_peak.position_uncertainty_3			
	_Assigned_peak_chem_shift.Spectral_dim_ID		This info is found in a separate category in NMR-STAR
_nef_peak.chain_code_1	_Assigned_peak_chem_shift.Entity_assembly_ID		
_nef_peak.sequence_code_1	_Assigned_peak_chem_shift.Comp_index_ID		Will need to add equivalent 'PDB' type tags or use 'Auth' tags for assignment hierarchy.
_nef_peak.residue_type_1	_Assigned_peak_chem_shift.Comp_ID		
_nef_peak.atom_name_1	_Assigned_peak_chem_shift.Atom_ID		
_nef_peak.chain_code_2			
_nef_peak.sequence_code_2			
_nef_peak.residue_type_2			
_nef_peak.atom_name_2			
_nef_peak.chain_code_3			
_nef_peak.sequence_code_3			
_nef_peak.residue_type_3			
_nef_peak.atom_name_3			
save_			
save_nef_peak_restraint_links			This save frame can be added to NMR-STAR. The exact implementation requires further discussion.
_nef_link_table.sf_category	_Constraint_expt_data_link_list.Sf_category	_Constraint_expt_data_link_list.Sf_category	Added to NMR-STAR
_nef_link_table.sf_framecode	_Constraint_expt_data_link_list.Sf_framecode	_Constraint_expt_data_link_list.Sf_framecode	Added to NMR-STAR

NEF tags	NMR-STAR equivalent	Recommended NMR-STAR/PDBx match	Comments
loop_			Chemical shifts are now used as restraints, but this system does not allow restraints to be linked to the chemical shift table. The chemical shift table does not have a unique 'shift_ID' that can be pointed to from a foreign key.
	_Constraint_peak_link.Spectral_peak_list_Sf_framecode	_Constraint_peak_link.Spectral_peak_list_Sf_framecode	Added to NMR-STAR
_nef_peak_restraint_link.nmr_spectrum_id	_Constraint_peak_link.Peak_char_spectral_peak_list_ID	_Constraint_peak_link.Peak_char_spectral_peak_list_ID	Added to NMR-STAR
_nef_peak_restraint_link.peak_id	_Constraint_peak_link.Peak_char_peak_ID	_Constraint_peak_link.Peak_char_peak_ID	Added to NMR-STAR
	_Constraint_peak_link.Constraint_Sf_framecode	_Constraint_peak_link.Constraint_Sf_framecode	Added to NMR-STAR
	_Constraint_peak_link.Constraint_Sf_category	_Constraint_peak_link.Constraint_Sf_category	Added to NMR-STAR
_nef_peak_restraint_link.restraint_list_id	_Constraint_peak_link.Constraint_list_ID	_Constraint_peak_link.Constraint_list_ID	Added to NMR-STAR
_nef_peak_restraint_link.restraint_id	_Constraint_peak_link.Constraint_ID	_Constraint_peak_link.Constraint_ID	Added to NMR-STAR
save_			