Conclusion of NEF survey and discussion

Three proposals were unanimously accepted:

- Q1 - Change 'residue_type' to 'residue_name
- Q3 - Remove _nef_chemical_shift_list.atom_chemical_shift_units column (always use 'ppm')
- Q4 - Add '_programspecific'_raw_data saveframe for input files and unstructured data

Questions Q2 (Change form 'ordinal' to 'index_id' for pro-forma line number columns) and Q6 (Add _nef_sequence.ordinal column to preserve line order for databases) both raised some discussion. It became clear that the name 'index_id' was seen as unnecessary complex, and that there was confusion about what was and was not an ordinal. The simplest solution seemed to be to use 'index' for all line-numbering columns (including _nef_sequence.ordinal), and to rename '_nef_run_history.run_ordinal to _nef_run_history.run_number, since this, on second thoughts, is not an ordinal, but a run serial number.

For question 7 (Residue variant codes) there was understandable scepticism about introducing a new set of topology descriptions that would then have to be maintained and expanded. This still leaves the shortcomings of the current situation, which the proposal was intended to address: The mmCif variant names (RCSB curated) have been rejected previously as too complicated. Furthermore, the set of curated variants is limited to the twenty regular amino acids, and there seems to be no realistic prospect of future expansion.

The following interpretation should address the concerns of both sides of the argument:

- NEF specifies residues and their variants in the nef_sequence loop using the combination of residue_name, linking, and residue_variant, as described in the NEF specification.
- The RCSB-curated descriptions for chemical compounds and residue variants are the reference for residue names, residue topologies and atom names.
- For residues with supported mmCif variant topologies (currently the twenty standard amino acids), the NEF standard specifies how NEF residue_name, linking, residue_variant triplets map to the mmCif variant codes, including which are the default variants used in NEF files. The mapping can be found in the specification, in the mmCIF_NEF_variant_mapping.txt file, which could be expanded to support additional curated variant topologies if and when they are introduced (after consultation and agreement only).
- For all other chemical compounds, the only supported topology specification corresponds to the Chemical Compound description.
- The supported variants for standard residues, including which is the default variant, is specified in the Residue_Variants.txt file. In summary, these are the mmCif-supported variants of the twenty standard amino acids (with a few additional variants for protonation states of backbone NH3 and COOH), as well as the 5'-terminal, 3'-terminal, intra-chain and free versions of the four standard DNA and four standard RNA residues. It should be noted

that the default variant of 5'-terminal nucleotides is the dephosphorylated one. A NEF-conforming program must be able to read a file containing these variants, interpreting each variant as the most appropriate topology supported internally. There is no obligation for programs to support (or export) all the variants. In discussion it was clear that some programs supported only a single variant of e.g. most N- and C- terminal amino acids, and would convert whatever variant was specified in the file to the supported variant. This behaviour is acceptable (and anyway unavoidable) in a conforming program. The only requirement is that on reading variants are mapped to the internal form that most closely matches the given information, and that on writing the exported variant codes match the topology used internally.

- NEF supports the specification of variants that do not have a matching mmCif topology specification. These are interpreted by adding and subtracting atoms relative to the parent compound, and by additional bonds in the nef_covalent_links loop, as described in the specification. The information available for such variants is unavoidably less complete, but it does allow a correct description of the atoms and bonding network for variants that would otherwise remain undescribed.
  There is no requirement for programs to be able to interpret non-standard variants, and programs are free to refuse to read files that contain them.
- The standard wildcard expressions for NMR-equivalent proton groups are derived from atom names and the general principles of the NEF specification. The Residue_Variants.txt file lists the wildcard expressions of the standard amino acid, DNA, and RNA that programs are used to refer to NMR-equivalent groups of nuclei (methylenes, methyls, isopropyls, rotating aromatic rings, ...).

Question 5 (Change non-stereo wildcards from X/Y (upper case) to x/y (lower case), and tighten wildcard rules) was also raised for discussion. The proposal originated from a point raised by the BMRB that the allowing people to choose their own casing for names would lead to confusion, and that a tighter specification was needed. This made it necessary that the casing of names should be a part of the standard. Since it had already been established that residue names should be mixed case and case sensitive, to support the eventual introduction of mixed case names for e.g. saccharide residues, an all-upper-case system was not possible. The best solution seemed to be to use the mmCif casing for atom and residue names, but to use lower case x and y to distinguish prochiral wildcards from atom names.