Supporting information for

going beyond base-pairs: topology-based characterization of base-multiplets in RNA

Sohini Bhattacharya,1 Ayush Jhunjhunwala,1 Antarip Halder,1,a Dhananjay Bhattacharyya,2 and Abhijit Mitra1,*

1Center for Computational Natural Sciences and Bioinformatics (CCNSB), International Institute of
Information Technology (IIIT-H), Gachibowli, Hyderabad 500032, India.
2Computational Science Division, Saha Institute of Nuclear Physics (SINP), 1/AF, Bidhannagar,
Kolkata 700064, India.
*aPresent address: Solid State and Structural Chemistry Unit, Indian Institute of Science, Bangalore
560012, India.

Section S1: Justification for priority of alphabetical order over geometric family considerations

Section S2: Detailed description of nomenclature convention for cyclic-3 and cyclic-4 quartet topologies

Section S3: Priority rules for nomenclature of higher order structural elements having one or more branch points but no cyclic components

Section S4: Comparison between 3DNA and BPFind: Why we have chosen BPFind output to identify multiplets?

Section S5: Name of all PDB files analyzed
Section S1:
Justification for priority of alphabetical order over geometric family considerations

The structures of two base-pairs belonging to the same base-pair family may be comparable only if the base-pairs are also isosteric, based on the identities of the constituent bases. Two base pairs, which are formally isosteric in isolation, may also differ from each other in their interactions with their respective local environment, depending on their base identities. In general, the structural similarity between base pairs within a given geometric family, is limited to only their respective local strand orientations. Thus, for a cluster of mutually interacting nucleotides, as in a multiplet, the structure and function are more sensitive to the base identities than to the geometric families of the constituent base pairs.

For example, if we consider linear quartets belonging to SSC-WWT-SSC family, there are seven varieties of base combinations are observed- AAAG, AAUA, AUAA, CAAG, CGCG, GAAU, and UAAU. In this case the isostericity pattern is conserved (I1-I4-I1) in AAAG, CAAG, CGCG, GAAU and UAAU. For AAUA and AUUA, the isostericity pattern is I1-I1-I1 and for CGCG the isostericity pattern is I1-I2-I1. These quartets are structurally different and do not appear in homologous structural contexts. Similar scenario is observed in other topologies also.

Further, the isostericity considerations are not only important in terms of C1'-C1' distances, and the relative positioning of the bases based on their hydrogen bond donor acceptor interactions; tertiary and neighbor interactions also need to be taken into account. For example, as was shown long back (Šponer, Jiří, et al. "Unique tertiary and neighbor interactions determine conservation patterns of Cis Watson–Crick A/G base-pairs." Journal of Molecular Biology 330.5 (2003): 967-978.) “There is a lack of A/G to G/A covariation, which, except for the G(N2) position, would be entirely isosteric.” and “they are not conserved when there is no tertiary or neighbor interaction”. It is also important to note that A:U, U:A, G:C and C:G W:WC are isosteric in terms of formation of double helical structure but their isostericity is lost when they interact with any other molecule. A:U and U:A are still symmetric from the sugar edge (minor groove) side but even these two base pairs are different from each other when a third component interacts with them through their Hoogsteen edge (Major groove side). More so is dissimilarity between G:C and C:G.

Given their structural dependence on their base identities, classifying multiplets based simply on geometric families of constituent base pairs, will not necessarily facilitate their structural and functional annotation. In the current context the scenario becomes even more complex when the topology varies (linear/star/cyclic etc.). As explained in the manuscript, this approach would also lead to counting identical multiplets multiple times and the total count will require further processing. This is avoided by using alphabetical order of base identities and base pair geometries, for tie breaking, while numbering the nucleotides in a multiplet. Of course, once the
A multiplet list is generated, it should be easily possible to appropriately annotate/tag each multiplet, to explain multiplet conservation across homologous organisms.

**Figure SF1: Different topologies of (A) base-triplets, (B) quartets and (C) pentets.** Numbers are assigned to the nodes for better illustration. Note that, the assigned numbers are only one out of many possible base and pairing geometry combinations. The final assignment of numbers to the nodes is decided by the identity of the nodes, the geometry of base-pairing interactions and the corresponding residue numbers as illustrated in Figure SF2.
Figure SF2: Demonstration of how the ambiguity in assigning the vertex (node) numbers is resolved in the proposed nomenclature scheme.

| How to resolve the ambiguity in assigning the vertex number? |
|-----------------------------------------------------------|
| **Rules** | **Unambiguous assignment** | **Unresolved ambiguity** |
| Rule 1: Prioritize on the basis of the order of the vertex | ![Unambiguous assignment](image) | ![Unresolved ambiguity](image) |
| Rule 2: Prioritize on the basis of alphabetic order of the residues (A > C > G > U) | ![Unambiguous assignment](image) | ![Unresolved ambiguity](image) |
| Rule 3: Prioritize on the basis of the alphabetic order of the neighbouring residues (A > C > G > U)* | ![Unambiguous assignment](image) | ![Unresolved ambiguity](image) |
| Rule 4: Prioritize on the basis of the alphabetic order of base pairing geometries (HHC > HHT > HSC > HST > HW C > HWT > S H C > SHT > S SC > SST > SW C, SWT > WH C > WHT > WSC > WST > WWC > WWT) | ![Unambiguous assignment](image) | ![Unresolved ambiguity](image) |
| Rule 5: Prioritize on the basis of the nucleotide number (in ascending order) | ![Unambiguous assignment](image) | ![Unresolved ambiguity](image) |

*If more than one neighbouring residues are present, the one which comes first in alphabetic order is selected for decision making.
Section S2:
Detailed description of nomenclature convention for cyclic-3 and cyclic-4 quartet topologies

Q3: Cyclic-3 quartet

Description:
1. Graph of four nodes (N1, N2, N3, and N4).
2. One central node (N1) with degree = 3.
3. Two peripheral nodes (N2 and N3), each having degree = 2. These nodes are part of a 3 membered cycle along with the central node (N1).
4. One terminal node (N4) having degree = 1.

Numbering convention:
1. In Cyclic-3 quartets, numbering starts from the central base (node having degree = 3).
2. The two peripheral bases with degree = 2 come next in numbering, according to their alphabetical order, followed by the terminal base with degree = 1.
3. If the two bases with degree = 2 are identical, then base-pair geometry between N1:N2 and N1:N3 are considered and priority order mentioned in generic priority rule (ii) is followed.
4. If base names, and base-pair geometry of N1:N2 and N1:N3, all are identical then lower nucleotide number (between N2 and N3) is given priority.

Nomenclature convention:
1. Here nodes with different degrees are segregated by using square brackets and angular brackets. The central base is written outside the square bracket and other bases are written within the square bracket, which is similar to the nomenclature convention used in star topology. Then inside the square bracket, bases having degree = 2 are grouped by using an angular bracket.
2. Interaction geometries between the central base and the other three peripheral/terminal bases are written respectively separated by a "/" delimiters, followed by the interaction geometry between two peripheral bases with degree = 2, i.e., N2:N3 separated by a ",".
**Representation scheme:**

**Name:** Q3 or Cyclic-3 N1[<N2 N3 >N4]; **Geometry:** Bp12/Bp13/Bp14, Bp23

**Q4: Cyclic-4 quartet**

**Description:**

1. Graph of four nodes (N1, N2, N3, and N4).
2. All the four nodes (N1, N2, N3, and N4) have degree = 2.

**Numbering convention:**

1. In Cyclic-4 quartets, as all four constituent bases save same degree (degree = 2), numbering of the bases starts from the base, which comes first in the alphabetical order.
2. If more than one base with same identity, which have the potential for being the starting base (according to the alphabetical order), then lower nucleotide number in the RNA chain is given priority.
3. The next base in numbering the sequence, and hence the naming direction, is decided by recursively applying alphabetical order of base name, base-pair geometry and the nucleotide number rules, respectively, as explained above.
4. The names of the remaining bases then come sequentially as the cycle is completed.

**Nomenclature convention:**

1. Interaction geometry of connected bases will be written serially, separated by "-" delimiters. To describe the cycle a "->" symbol is added after the interaction geometry between N4 and N1.

**Representation scheme:**

**Name:** Q4 or Cyclic-4 N1-N2-N3-N4 ->; **Geometry:** Bp12-Bp23-Bp34-Bp41->
Section S3:
Priority rules for nomenclature of higher order structural elements having one or more branch points but no cyclic components:

1. For starting the name of a multiplet, priority will be given to the nucleotide (node) having highest degree.
2. In case, there is more than one node with the same (highest) degrees, priority will be given to the node, which have child nodes (directly connected nodes) having greater degrees.
3. In the next level, priority will be given to alphabetical order.
4. At the last level, priority will be given to the nucleotide numbers.
5. In order to represent the connectivities between nucleotides (nodes) nested brackets (square bracket “[]”) are used.
6. The above mentioned priority rules (rules 1-4) are implemented for writing every node in a sequential order.

Pentet topology P2 is the example of a higher order structural element having open structure (have no cyclic components) with one branching point. Nomenclature of P2 pentet topology described in SF2 (shown below) will provide better understanding about the implementation of the priority rules.

Rules for nomenclature of higher order structural elements having cyclic components:

1. If higher order structural elements contain 4 member or larger cyclic component, then cyclic components will be written similar to the cyclic 4 quartet topology. Check nomenclature for P12 pentet topology.
2. In case there is any additional interaction between any two members of cyclic the cyclic component or branching observed in any point of the cycle then the additional interactions will be written separately delimited by “,”.
3. If the structure contains one or more three member cycle, then nomenclature rules for cyclic-3 quartets will be extended for those cases. Check nomenclature of P3 pentet topology for example.
Figure SF3: Nomenclature rules for observed Pentet topologies.

| Topology type | Quintet or Pentet nomenclature | Geometry               |
|---------------|--------------------------------|------------------------|
| P1 topology   | \( N_1-N_2-N_3-N_4-N_5 \)    | \( Bp_{12}-Bp_{23}-Bp_{34}-Bp_{45} \) |
| P2 topology   | \( N_1[N_2[N_3 \ N_4 \ N_5]] \) | \( [Bp_{12}-Bp_{23}]/Bp_{14}/Bp_{15} \) |
| P3 topology   | \( N_1[\langle N_4 \ N_5 \rangle \ N_4 \ N_5] \) | \( [Bp_{12}-Bp_{23}]/Bp_{14}/Bp_{15}, Bp_{45} \) |
| P8 topology   | \( N_1[N_2 \ N_3 \ N_4 \ N_5] \) | \( Bp_{12}/Bp_{13}/Bp_{14}/Bp_{15} \) |
| P12 topology  | \( N_1-N_2-N_3-N_4-N_5-> \)   | \( Bp_{12}-Bp_{23}-Bp_{34}-Bp_{45}-Bp_{51}-> \) |
Section S4: Comparison between 3DNA and BPFind: Why we have chosen BPFind output to identify multiplets?

Many base-pair detection tools are available which implement different algorithms, based on either the Sanger scheme or the LW scheme, to identify base-pairs and annotate them. Here we have compared two such widely used programs, namely “find_pair” (Lu et al. 2010) and “BPFind” (Das et al. 2006). ‘Find_pair’ is a major component of 3DNA software suite (Lu and Olson 2008; Zheng et al. 2009), which follows purely geometry based criteria for identification of base pairs of all types - canonical, non-canonical and base-pairs having modified bases, irrespective of their protonation and tautomeric state. This program identifies base pairs having nearly planar geometry and at least one hydrogen bond, based on a set of stringent parameters -(i) vertical distance between two base planes, (ii) angle between two base normal vectors and (iii) potential donor-acceptor distance. DSSR software tool (Lu et al. 2015) implements the same criteria to identify higher order base associations (multiplets) present in a pdb files. It simply searches horizontally in a base-pair plane to determine whether there is any further hydrogen bond associated with that base pair. Thus, DSSR identifies coplanar base associations linked by at least one hydrogen bond.

BPFind, on the other hand, implements a hypothesis driven approach, where for each possible type of base pairing geometries (as per LW scheme) the program predefines a set of donor acceptor combinations along with precursor atoms for each donor and acceptor atoms. Then based on a set of geometric criteria (which is illustrated later in the method section in the main manuscript) BPFind identifies and annotates all varieties of base pairs, stabilized by at least two hydrogen bonds, which are present in a given pdb file.

This difference in base-pairing criteria, results in certain base-pairs being detected by one of the tools and not by the other. On the one hand, because of the stringent criteria used, ‘find_pair’ may miss out certain interactions due to minor distortions. Simultaneously, on the other hand, consideration of single hydrogen bonded base-pair may lead to the detection of a large number of higher order structural elements, several of which may not be structurally and functionally relevant. Whereas, in this sense the BPFind software is likely to yield stabler (having at least two hydrogen bonds in every base-pair units), and not so planar but relevant, higher order structural elements.

In order to compare the “find_pair” and the “BPFind” output, as a test case the RNA chains present in large ribosomal subunit of H. marismortui (pdb:1vqo.pdb, chain 0: 23S rRNA and chain 9: 5S rRNA) are considered. A brief summary of the comparison is discussed below.
Base pairs identified by 3DNA
Total base-pairs identified by 3DNA: 1175
Single hydrogen bonded base-pairs: 65
Base pairs having two or more hydrogen bonds: 1110
Annotation: The details of interacting nucleotides (residue number, chain and residue names) are mentioned. Interaction geometry of the base-pairs are not reported.

Base pairs identified by BPFind
Total base-pairs identified by 3DNA: 1249
Single hydrogen bonded base-pairs: 0
Base pairs having two or more hydrogen bonds: 1249
Annotation: The details of interacting nucleotides (residue number, chain and residue names) are mentioned. Base-pairs are also annotated as per the Leontis-Westhof nomenclature scheme.
Except single hydrogen bonded base pairs, which are detected by 3DNA software, 1022 instances of normal base pairs are identified commonly by both the tools, BPFind and 3DNA. However, each of these to software detects some base pairs uniquely. The list of base pairs (only residue number of interacting nucleotides), which are commonly identified by both the tools and which are exclusively detected by any one of the tools are listed in attached spread sheet named “supplemental-table-BPFind-3DNA-comp.xls”.

After manual investigation, we have considered BPFind output for retrieving information about base-pair units for the following reasons.

1. BPFind identifies a greater number of base-pairs having at least two hydrogen bonds.
2. BPFind identifies and properly annotate base-pairs, including class-I protonated base-pairs according to the Leontis-Westhof classification and nomenclature scheme of base-pairs.
3. BPFind performs a quality check of the base-pairs by calculating E_value, hence this software reports only those base-pairs having significantly good geometry. Therefore, the multiplets which are detected by using base-pair information from BPFind output, mostly show significantly strong connectivities and good geometry.
4. Because of the stringent criteria used in find_pair algorithm of 3DNA software, it misses those base pairs which are not exactly planar and also misses bifurcated base-pairs. Therefore, many of the important higher order structural elements containing bifurcated pairing and not so planar base-pair units will not be identified if base-pair information from 3DNA output is considered.
5. 3DNA identifies single hydrogen bonded base-pairs. If single hydrogen bonded base pairs are considered, then the number of instances of higher order structural elements will
increase significantly. However, many of those may not have significantly stable structure. Moreover, single hydrogen bonded base-pairs are difficult to annotate as per the LW scheme. Therefore, to avoid ambiguity in annotation, single hydrogen bonded interactions are not considered in the present study.
Figure SF4: Sextet present in different structural contexts of ribosomal RNAs. (A) Conserved sextet elements in *H. marismortui* and *T. thermophilus* 23S rRNA (domain-I). They appear in similar structural contexts. (B) Conserved sextet element in 5' domain of 16S rRNAs of *E. coli* and *T. thermophilus*, which brings close the two distantly placed internal loop regions (backbone colors are different for two loops) and stabilize the folded structure. (C) Sextet element observed in 3'M domain of *T. thermophilus* 16S rRNA. Backbones of distantly placed regions are colored differently.
Figure SF5: Septet present in different structural contexts of ribosomal RNAs. (A) Recurrently observed septet elements in 5’ domain of *T. thermophilus* 16S rRNA having a linear topology. This septet is found in a multi-helix junction. (B) Recurrently observed septet element in an internal loop region of domain-I of *E. coli* 23S rRNA. In this septet topology, a branching is observed at nucleotide number 480 A.
**Table ST1:** Pentets present in different structural contexts of *T. thermophilus* and *S. cerevisiae* rRNAs (based on data available in HDRNAS non-redundant dataset only). P1, P2 and P3 are pentet (pentet) topologies as shown in Figure 2 of main manuscript.

| Pentets present in different RNA types | Frequency in multihelix junction | Frequency in mediating long distance interaction |
|---------------------------------------|----------------------------------|-----------------------------------------------|
|                                       | P1  | P2  | P3  | P1  | P2  | P3  |
| *T. thermophilus* 16S rRNA (Unique pentet instances total=7) | 2   | 0   | 1   | 3   | 1   | 0   |
| *T. thermophilus* 23S rRNA (Unique pentet instances total=17) | 7   | 3   | 0   | 7   | 0   | 0   |
| *S. cerevisiae* 18S rRNA (Unique pentet instances total=3) | 0   | 0   | 0   | 3   | 0   | 0   |
| *S. cerevisiae* 25S rRNA (Unique pentet instances total=9) | 4   | 1   | 0   | 2   | 0   | 0   |

The data shown in the table is based on pentets available in HDRNAS non-redundant dataset only. Pentets are detected in multiple pdb files, which corresponding to *T. thermophilus* ribosomes. Frequency data reported in this table are calculated after removing all repetitive instances. The list of pdb files corresponding to each of the above mentioned RNA types are given below:

* T. thermophilus* 16S rRNA: 2vqe, 3knh, 3knl, 3knn, 3pyn, 3pyu, 3tvf, 3uz6, 3dh9
* T. thermophilus* 23S rRNA: 2xg0, 3kir, 3kit, 3uye, 3uz8, 3v2d
* S. cerevisiae* 18S rRNA: 3u5b
* S. cerevisiae* 25S rRNA: 3u5d
Section S5:
Name of all PDB files analyzed

List of PDB files in HDRNAS non-redundant dataset

List of PDB files in NDB non-redundant dataset

Name of all PDB files analyzed

Section S5:
List of PDB files in Large crystal (X-ray) structure dataset from RCSB-PDB

2pxj, 3hga, 30k4, 4e4t, 1rxa, 3zw2, 4ang, 3hu2, 1a1d, 1mad, 1aj1, 1osu, 2c4z, 2c50, 2c51, 1p79, 2f2z, 4a4f, 4alp, 2xnr, 4qvc, 4iq3, 4ijy, 4n2s, 333d, 3u2e, 2x1f, 1h2c, 1eqq, 4io0, 4oi1, 4ohy, 4n2q, 3qjp, 27r8, 2r7v, 4yoe, 4q6u, 4lmz, 2r7s, 308c, 3m0j, 4r3i, 4qvd, 3hsb, 4ms9, 4hoz, 2r7r, 2r7t, 283d, 3p4f, 4rcm, 3qsu, 3qsp, 3rtj, 2vop, 1g2j, 3bou, 3nj6, 4dwa, 3q9g, 3q0q, 3q0r, 3qas, 3af6, 4s2y, 5bud, 2a1r, 4jzv, 4jvh, 3sk5, 3sk5, 3k5z, 3k6i, 3k6i, 3k6d, 4tu0, 3tn, 3p5f, 1hdr, 3erc, 3t0o, 3x9e, 1b0g, 4i6k, 2sx7, 4rby, 4rb, 4rc0, 4rvk, 3qgb, 3qcc, 4s2x, 1lj3, 4jzu, 1kq2, 4hot, 4hos, 3avt, 4nk4, 4ij0, 2asb, 4j7m, 4qu7, 3mna, 2g91, 4u3p, 4xk0, 1n1h, 4rj1, 413d, 308r, 3v7l, 5bte, 3v6y, 2xs2, 3bsb, 4j7l, 40au, 1nb7, 4hor, 1fxl, 2bbv, 1si3, 4nl3, 3nma, 1i5l, 1pgl, 3nmr, 4j4af, 255d, 2q66, 2r7w, 40av, 3t5n, 3v74, 3li2, 47yw, 4r2o, 4z76, 1uvk, 4rvc, 1jge, 4kre, 3nnn, 1wrg, 4d26, 4d25, 1uv1, 3ice, 3gib, 2c0b, 2p01, 4ola, 40lb, 3pey, 2q1r, 3g9y, 2vnu, 4am3, 3mdg, 3mdl, 4e59, 4r3x, 2v7r, 4u34, 4r37, 259d, 402d, 4u3i, 3u03, 4u3r, 4u47, 7u4b, 4cs1, 3iee, 1wpu, 3sqw, 3sqx, 1uvj, 2eaa, 3csw, 4fte, 4fts, 3pew, 4fyy, 4kxt, 2cwp, 4wrk, 4e78, 2x5s, 4c8y, 49g2, 3r1e, 4a1l, 1rxb, 3nd3, 3nd4, 3sq0, 3i61, 3i62, 3i5x, 3i5y, 1wmq, 1av6, 3osi, 2a8v, 4qnm, 4i99, 4mdx, 2xqj, 4krf, 4alc, 48z6, 4cus, 4z6l, 2vuo, 2aqp, 3p4c, 3p4d, 44c0, 4fn1, 306e, 4h8t, 2xzo, 2b2d, 2r7y, 3jrx, 4g3v, 4v6e, 3oin, 4a6l, 2xli, 1kdd3, 1kdd4, 1kdd5, 438d, 1kfo, 1e1v, 1umv, 3qom, 3sqp, 2bx2, 1ddl, 3o7y, 2ju, 4r7v, 2q0a, 3b3x, 2qgb, 3aeav, 3e9r, 2von, 1n38, 4wtk, 4wtl, 4w4m, 4w4t, 4wtj, 4gv9, 3kmq, 2x1j, 2f8k, 3knk, 4o8j, 3do4, 3rer, 1uvl, 5c0y, 1pvo, 4nng, 4aba2, 4jvy, 4f5s, 2v0d, 12ze, 14qz, 1zdj, 4e58, 2gx8, 3gvm, 421d, 479d, 4msr, 2r7x, 3q0n, 1bmx, 3bnt, 1bmv, 4n6h, 3go9, 4h7t, 4n4g8, 3bx2, 4dze, 4bsg9, 4fth, 3klv, 3kms, 3kna, 3nlo, 4222, 47d2, 3q3x, 1ef0, 1dfq, 1dhq, 4jrd, 377d, 3ibk, 3qol, 1m8y, 4nbg, 4nrc, 4ngd, 3qrr, 3bsx, 4nha, 4ft3, 2bs1, 3ul3, 3vnu, 2g9f, 2g9u, 2g8u, 2g8w, 2gun, 165d, 2b8y, 439d, 48hh, 3e0t, 4g92, 1gq4, 1fiz, 404d, 2ykg, 205d, 2r21, 2r22, 4hn3, 2yj2, 2yj1, 2xp0, 1xpr, 1xpu, 1knz, 3hft, 4wta, 4wct, 4wtf, 4wtg, 2zin, 2iz8, 2c4q, 2cy4, 2wy4, 2tvd, 4w8y, 4zd2, 3avl, 26v6, 4fuv, 3s12, 2dqo, 4dpb, 4wpb, 3z6d, 4j39, 1l8q, 3pkm, 4jko, 2jlv, 4w5n, 3sqj, 3qjl, 4x2b, 1yty, 1zh5, 3gpr, 3uad, 1l8b, 18x, 4jgn, 1j6s, 4h6v, 4j5v, 2x3j, 2dss, 2atw, 3twh, 1n35, 2gb8, 1uvn, 1ytu, 4nmg, 3env, 4nxh, 4nlf, 3fs0, 1nt, 3vyy, 157d, 3cjj, 4jrt, 4knq, 2jyj, 4ed5, 2x1l, 4wzq, 354d, 3avv, 4e6b, 3dw4, 3dw6, 480d, 2ppj, 1pjq, 1pjo, 21s, 2r20, 433d, 4pei, 4f02, 2bs0, 4qpx, 3bso, 2aba4, 3zd7, 430d, 1uf1, 3bov, 2py9, 3m58, 3qt2, 3bnn, 3cqs, 3cpp, 3cgg, 3cgr, 3cgs, 3s5f, 437d, 1rna, 4kky, 3klv, 2ije, 3dd2, 3m7n, 3nhr, 2aun, 2e9r, 4wzm, 3hs5, 3hy5, 1a34, 4c9d, 4w5r, 357c, 385u, 3dvz, 3dw5, 3dw7, 483d, 3adi, 112x, 3q51, 4h5p, 1b7f, 2e9z, 387d, 4tv0, 3xmn, 5amr, 3rd2, 1q9a, 4ay2, 1msy, 353d, 3csl, 2a4s, 213m, 434d, 436d, 450d, 35q0, 4pmw, 18f8, 3izm, 1k8w, 3aww, 1mhk, 3q9d, 4gha, 2val, 1m8w, 4peh, 3rc2, 2anr, 1l2f, 1zhd, 4oq8, 3avx, 2bpg, 3ncu, 3lfr, 3k49, 3avvy, 2jos, 7msf, 3ho1, 3tmi, 2xsl, 4pcj, 2w89, 430d, 3ogq8, 1jdz, 1jzv, 4w5t, 2iz9, 4ato, 4jki, 3hm9, 3sn2, 2h8k,
List of PDB files in Large NMR structure dataset from RCSB-PDB

1t2r, 176d, 2xc7, 2oj7, 2oj8, 2l5d, 2mxj, 4b8t, 2kfy, 2kg1, 2l41, 2mz1, 2ixz, 1hs1, 1hs3, 1hs4, 1hs8, 1wwd, 1wwg, 1wwf, 176d, 2xc7, 2oj7, 2oj8, 2l5d, 2mxj, 4b8t, 2kfy, 2kg1, 2l41, 2mz1, 2ixz, 1hs1, 1hs3, 1hs4, 1hs8, 1wwd, 1wwg, 1wwf, 2k2q, 2k3g, 2k4l, 2k5m, 2k6n, 2k7o, 2k8p, 2k9q, 2ka0, 2kb1, 2kc2, 2kd3, 2ke4, 2kf5, 2kg6, 2kh7, 2ki8, 2kj9, 2kk0, 2kl1, 2km2, 2kn3, 2kp4, 2kq5, 2kr6, 2ks7, 2kt8, 2kv9, 2kw0, 2kx1, 2ky2, 2kz3, 2la4, 2lb5, 2lc6, 2ld7, 2le8, 2lf9, 2lg0, 2lh1, 2li2, 2lj3, 2lk4, 2ll5, 2lm6, 2ln7, 2lo8, 2lp9, 2lq0, 2lr1, 2ls2, 2lt3, 2lu4, 2lv5, 2lw6, 2lx7, 2ly8, 2lz9, 2ma0, 2mb1, 2mc2, 2md3, 2me4, 2mf5, 2mg6, 2mh7, 2mi8, 2mj9, 2mk0, 2ml1, 2mm2, 2mn3, 2mo4, 2mp5, 2mq6, 2mr7, 2ms8, 2mt9, 2mu0, 2mv1, 2mw2, 2mx3, 2my4, 2mz5, 2na6, 2nb7, 2nc8, 2nd9, 2ne10, 2nf11, 2ng12, 2nh13, 2ni14, 2nj15, 2nk16, 2nl17, 2nm18, 2nn19, 2no20, 2np21, 2nq22, 2nr23, 2ns24, 2nt25, 2nu26, 2nv27, 2nw28, 2nx29, 2ny30, 2nz31, 2oa32, 2ob33, 2oc34, 2od35, 2oe36, 2of37, 2og38, 2oh39, 2oi40, 2oj41, 2ok42, 2ol43, 2om44, 2on45, 2oo46, 2op47, 2or48, 2os49, 2ot50, 2ou51, 2ov52, 2ow53, 2ox54, 2oy55, 2oz56, 1a057, 1a158, 1a259, 1a360, 1a461, 1a562, 1a663, 1a764, 1a865, 1a966, 1aa67, 1ab68, 1ac69, 1ad70, 1ae71, 1af72, 1ag73, 1ah74, 1ai75, 1aj76, 1ak77, 1al78, 1am79, 1an80, 1ao81, 1ap82, 1aq83, 1ar84, 1as85, 1at86, 1au87, 1av88, 1aw89, 1ax90, 1ay91, 1az92, 1ba93, 1bb94, 1bc95, 1bd96, 1be97, 1bf98, 1bg99, 1bh100, 1bi101, 1bj102, 1bk103, 1bl104, 1bm105, 1bn106, 1bo107, 1bp108, 1bq109, 1br110, 1bs111, 1bt112, 1bu113, 1bv114, 1bw115, 1bx116, 1by117, 1bz118.
