Guidelines to map the notations used in RNABP COGEST and Non-canonical RNA Base-pair database (SINP database)

- In RNABP COGEST base pairs are described as for example, A:U W:H Cis, which means Watson Crick edge of Adenine interacts with Hoogsteen edge of Uracil in a Cis orientation of glycosidic bond with respect to each other.
  In SINP database also the representation is similar.

- A:U W:H Cis and U:A H:W Cis are considered as geometrically similar in both the database. So, if A:U W:H Cis is written in RNABP COGEST, the same entry in SINP database can be written as both A:U W:H Cis or U:A H:W Cis. Here, we have taken one particular example to explain, but this guidelines are valid for all the Cis and Trans base pairs in the database.

- In the table below the abbreviation of nucleotide names and edges, used in both database to represent a base pair identity and geometry are listed:

Abbreviations	RNABP COGEST	SINP database
A	Adenine	Adenine
U	Uracil	Uracil
G	Guanine	Guanine
C	Cytosine	Cytosine
rA/ rU/ rG/ rC	Bases with attached ribose sugar moiety	Not used
W	Watson Crick edge	Watson Crick edge (strong h-bonds* through Watson Crick edge >= 2)
H	Hoogsteen edge	Hoogsteen edge (strong h-bonds* through Hoogsteen edge >= 2)
S	Sugar edge	Sugar edge (strong h-bonds* through Sugar edge >= 2)
w	Not used (Hydrogen-bonds details are given separately)	Watson Crick edge (strong h-bonds* through Watson-Crick edge < 2, i.e., having C-H..O, C-H..N type of hydrogen bonds)
h	Not used (Hydrogen-bonds details are given separately)	Hoogsteen edge (strong h-bonds* through Hoogsteen edge < 2, i.e., having C-H..O, C-H..N type of hydrogen bonds)
s	Not used (Hydrogen-bonds details are given separately)	Sugar edge (strong h-bonds* through Sugar edge < 2, i.e., having C-H..O, C-H..N type of hydrogen bonds)
+	Not used (Protonated base is mentioned separately)	Protonation in Watson-Crick edge

z	Not used (Protonated base is mentioned separately)	Protonation in Sugar edge

* Strong h-bonds in RNA are those hydrogen bonds where donors are N-H or O-H groups. N-H..N, N-H..O, O-H..N, and O-H..O are four possible types of strong hydrogen bonds.

- The frequency value of a particular base pair may not be same in both the databases, as the non redundant dataset used for frequency calculation is not same.

- RNABP COGEST, QM calculation data for some single hydrogen bonded base pairs, for such base pairs no corresponding entries in the SINP database.

- For some base pairs which are designated as modelled in SINP database, RNABP COGEST says that as observed. Because the information given in SINP database are for those base pairs which are with at least two good hydrogen bonds. For QM calculation of different base pairs some times more rigorous searching has been implemented to find out a crystal structure occurrence of a particular base-pair, even in a larger crystal structure database.
  For such rigorous searching, some base-pair instances have been observed, which were initially designated as modelled base pair.

- Example base pairs are different in two databases. In RNABP COGEST, the example base pair given in the first table in the detailed base pair information page is the instance which was chosen for further QM calculation.