

## Statistics of RNA structures

Evgeny Baulin<sup>1</sup>, Dmitriy Ivankov<sup>2</sup>, Mikhail Roytberg<sup>3</sup>

<sup>1</sup>Higher School of Economics, Russian Federation

<sup>2</sup>Technical University of Munich, Germany

<sup>3</sup>Institute of Mathematical Problems in Biology RAS, Russian Federation, [mroytberg@lpm.org.ru](mailto:mroytberg@lpm.org.ru)

We present first results of investigation of experimentally obtained RNA structures. The aim of the study is to learn more about special type of bonds between RNA nucleotides (“links”, see below).

The initial data were taken from Protein Data Bank [1] and Nucleic Acid Database [2]; only documents containing only one model we have considered. That gave us 10300 documents describing structures with RNA, the documents divided into 4 categories: RNA; RNA-Protein complex; RNA-DNA complex; RNA-DNA-Protein complex. The option “find\_pair” from the X3DNA [3] were used to create the files containing the information on bonds between nucleotides and on the number of helices, that constitute these bondings. At this stage 52 original files were temporary put aside, because they included modified elements that cannot be handled with X3DNA program. The results of the “find\_pair” operation were converted to tables, which became the subject of the further processing. At the current stage of investigation we are interested in the RNA structures themselves; therefore we have temporary put aside 2621 files containing several RNA bound with each other.

We use the following terminology: D-helix – Helix defined by X3DNA; Isolate – Bonding defined by X3DNA as isolated; Standard helix (Helix) – Non-extendable sequence of X3DNA-bondings of form:  $(x, y), (x+1, y-1), \dots, (x+s, y-s)$ , here  $(i, j)$  denote the bond between  $i$ -th and  $j$ -th nucleotides of the chain. Link – Bond that is not a part of the standard helix of length 2 or more.

We say that bonds  $(a, b)$  and  $(c, d)$  are *correlated* if the fragments  $(a, b)$  and  $(c, d)$  do not intersect each other or one of them is a part of another. Otherwise we say that the bonds are *conflicting*. Since a nucleotide cannot form more than one bond a link conflicts with a bond belonging to a standard helix iff it conflicts with all bonds of the standard helix. Therefore we say “a link conflicts with a helix”, “two helices are conflicting”, and so on. The links have been divided in three classes: Internal (non-conflict with spirals or other links), Connected (with conflicts only with other links), Free (with conflicts with helices). At this stage 900 files containing conflicts between the helices were temporary put aside.

The analysis of the data have resulted in tables of several types (table of helices, table of unpaired fragments (“loops”), table of links, etc.). Our study for example shows that ~ 36% of considered structures contain links. About 30% of links have conflicts with helices, about 55% of links has no conflicts at all and only 15% are “connected” links, i.e. links having conflicts with other links but not with helices. Interestingly, positions of nucleotides forming a link are close to ends of helices. As to our knowledge the role of links in RNA structure was not studied so far. At the next stages of investigation we plan to learn more about types of loops related to links, relation between links and RNA-protein interaction and so on.

We thank S.A.Spirin for useful discussions.

1. <http://www.pdb.org/pdb/home/home.do>
2. <http://ndbserver.rutgers.edu>
3. <http://rutchem.rutgers.edu/~xiangjun/3DNA/>