RNA motif database catalogued by kink parameters

Satoshi Fujii
School of Pharmaceutical Sciences, University of Shizuoka, 52-1 Yada, Shizuoka 422-8526, Japan

ABSTRACT
In the large RNA molecules, the helical stems close to each other with specific arrangement. These specific interactions are mostly identified by the similar mode and can be classified to the several patterns. Several motifs are well characterized at an atomic resolution. The database, structural information about the assemblage and the space distribution of base moieties, can be a useful tool. Kink parameter affords the structural description between the adjacent groups in any irregular geometry. The relative rotational and positional features may be a preferable information to cite the curved or bend phenomenon in not only the successive strand but also the inter strands. The geometrical parameters about hydrogen bond and base stacking would provide the tolerant aspect for multiple strand packing. The tentative database including these geometrical parameters has been constructed. The user can obtain the selected list with the several descriptors for the structure definition and sequence properties. In addition to this nucleic acid system, RNA-protein system would be included.

INTRODUCTION
With the specific three-dimensional structure, RNA molecules express the specific structural and catalytic functions. RNA molecules form complex structures containing A-form helices and non-helical regions, such as loops and bulges. These structures have been classified as secondary structure diagrams. A series of the crystal structures of ribozyme and ribosome has offered the huge amount of structural information for RNA structure. In the large RNA molecules, ribozymes and ribosome(1-4), the helical stems close to each other with specific arrangement. These specific interactions are mostly identified by the similar mode and can be classified to several patterns. Several motifs are well characterized at atomic resolution. In addition to U-turn, S-turn A-platforms and tetraloop, several motifs, such as kink-turn and A-minor, have been recently identified(5,6).

For the formation of these motifs, there are specific interactions to the adjacent RNA segments and/or protein beside the specific base sequences. A kind of kink parameter affords the structural description between the adjacent groups in any irregular geometry (7). The relative rotational and positional features are represented and can be converted to the other coordinate system. Essential angular components of base stacking, tilt, roll and twist angles may be a preferable information to cite the curved or bend phenomenon in not only the successive strand but also the inter strands. In addition to these parameters, the geometrical parameters about hydrogen bond and base stacking would provide the tolerant aspect for multiple helical packing. The solvent accessible surface area becomes a good indicator to estimate the packing force. The tentative database including these geometrical parameters has been constructed.
METHODS, RESULTS AND DISCUSSION

Coordinates. The initial coordinates have been obtained from the Protein Data Bank. The database, assemblage and distribution of structural information become an essential tool. The NDB (Nucleic Acid Database) provides the structural data about nucleic acid. This project also makes the software available to produce reports describing any of the stored properties of any subset of the structures in the database. CIF is an acronym for the X-ray crystallographic information file and the CIF format is suitable for archiving, in any order, all types of text and numerical data. Database, which shows the search menu and allows a variety of kink parameter combination, is written using CIFtbx2 software package.

Geometrical parameters. The coordinates are converted by translation and rotation matrices to arrange the atomic positions in the normalized coordinate system on the basis of the standard base-pair. Geometrical parameters. The geometrical parameters, location and skew matrix (3x3) for adjacent function groups are calculated. The calculated kink parameters are written in mmCIF format using data description language. The description, such as item.name, item.category_id, item.man datory_code and so forth, are referred to the mmCIF Dictionary. The program BIOCON also lists the hydrogen bonding between base moieties. The number of hydrogen bonding interactions between bases that are remote in the secondary structure is about 100. In the case of non-canonical base pairing, the hydrogen bonding mode, such as atom name of hydrogen donor or acceptor, is written in mmCIF format. The solvent accessible surface area of the base is calculated with a 1.4 Å radius probe, using program O and VOIDOO.

The adenosine moiety plays a key role in helical packing and/or loop and/or bulge interactions. It is well known as A-minor motif. It involves adenines inserted into the minor grooves of RNA helices. In the A-minor motif, the hydrophobic patch is the essential packing force. There are several hydrophobic faces such as C2 atom and sugar moiety. The formation power (number) of hydrogen bonding also affords a key discriminator. These stereo specific information are noted in the mmCIF format with the competent word.

Multi frameworks are created and stabilized by two kinds of interaction, base stacking and specific hydrogen bonding. In general, stacking force filed extends to three even to four bases. There are several combination patterns with the base kind, number of the embedded bases, direction and there are several stacking patterns by the stacking steps, rotation and translation. There are many difficulties, such as classification of pairing and visualizations. The performance in our procedure is in progress.

ACKNOWLEDGEMENTS
This work was supported in part by Grants-in-Aid from the Ministry of Education, Science, Sports and Culture, Japan.

REFERENCES