RNA StrAT: RNA Secondary Structure Analysis Toolkit

Valentin Guignon[1], Cedric Chauve[2] and Sylvie Hamel[1]

1. DIRO, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, QC, H3C 3J7, Canada.
2. Department of Mathematics, Simon Fraser University, 8888 University Drive, Burnaby, BC, V5A 1S6, Canada. CGL and LaCIM, Université du Québec à Montréal, Montréal, QC, Canada.

Université de Montréal

UQÀM

SUMMARY

RNAStrAT is a **web server** dedicated to the comparison of sets of RNA secondary structures. This server offers tools to **align pairs of RNA secondary structures** and to **search for structural homologs in a database of RNA secondary structures** (based on the RFAM). The alignment and search are based on an **edit distance algorithm** that considers a wide range of edit operations defined in [Jiang et al., 2002]. Tools for the **vizualisation of secondary structures and structures alignments** are also available.
Up to date RNA StrAT is the only server offering all these features (general RNA edit model, RFAM database search, rendering) together.

Availability: **http://www-lbit.iro.umontreal.ca/rnastrat/**

Contact: **guignonv@yahoo.fr**
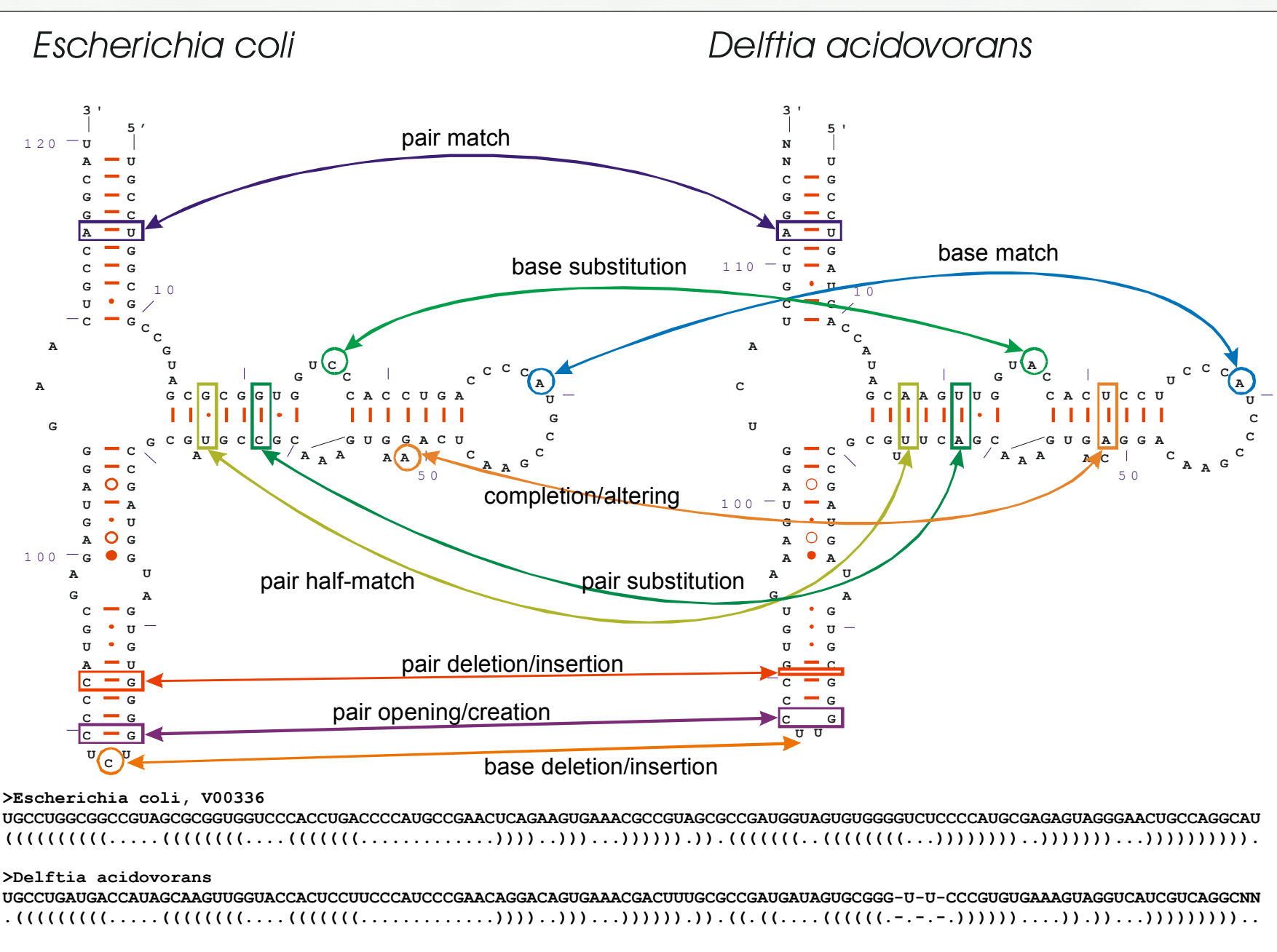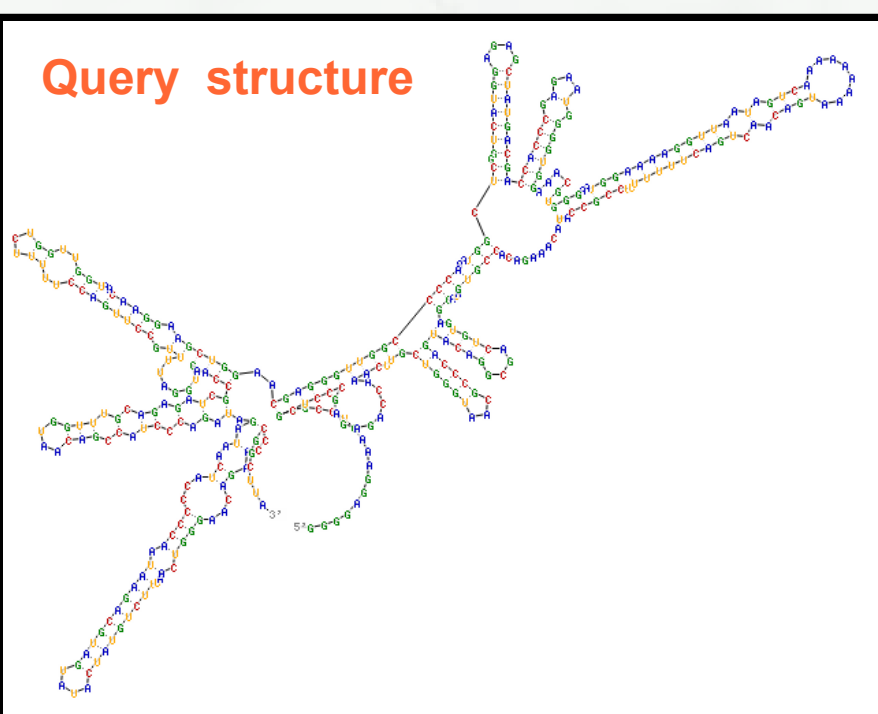
ALIGNMENT OF SECONDARY STRUCTURES

Figure 1. Structural alignment between two 5S rRNA. Some edit operations are displayed on the structures using arrows.

SEARCH FOR STRUCTURAL HOMOLOGS

For example, structured RNA gene detected by a high-throughput computational analysis.
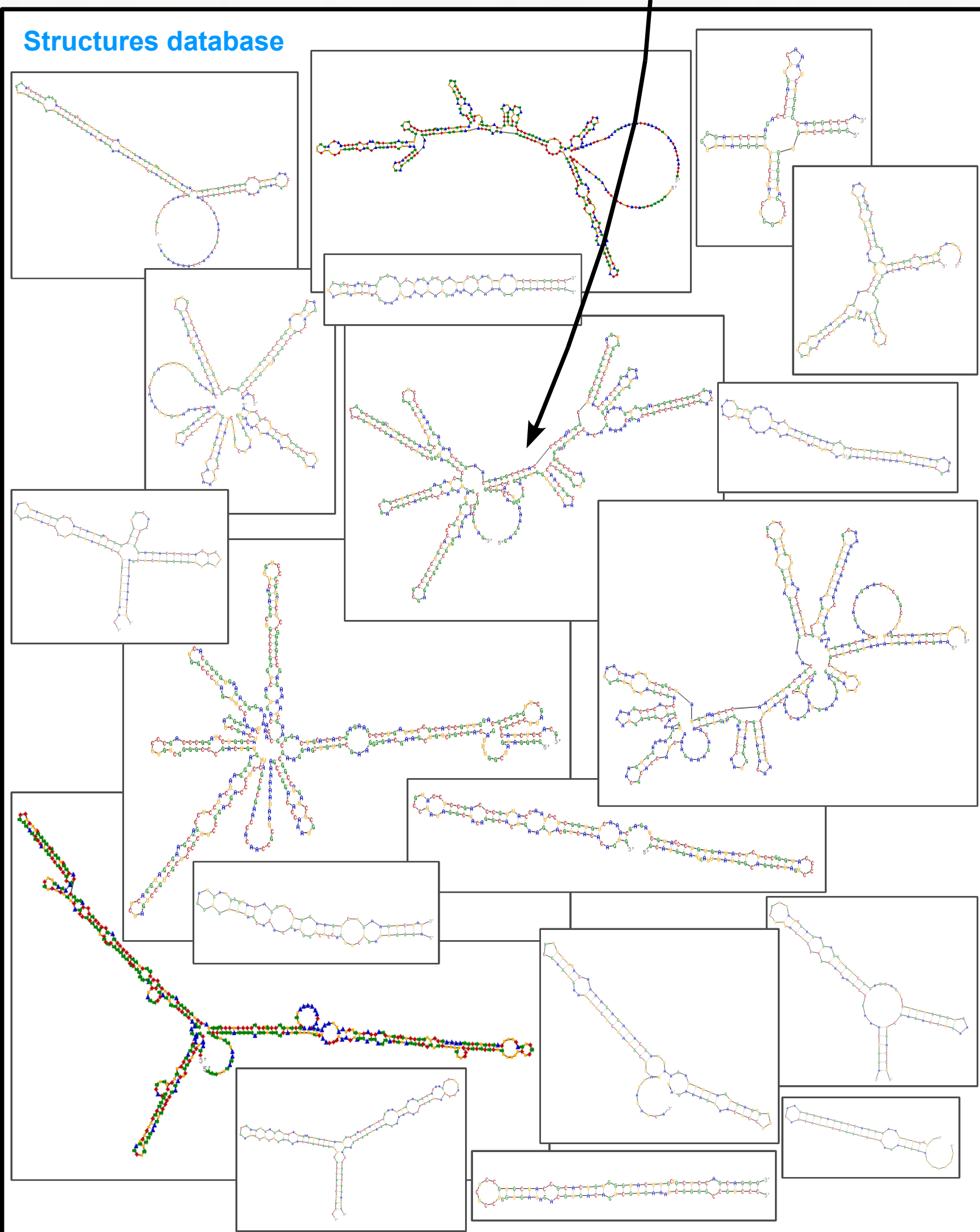

Figure 2. Given a query structure and a database of structures, structural homologs can be found using structural similarity scores.

ALGORITHMIC MODEL EDIT DISTANCE

The edit distance algorithm is the following problem:
- given
  • two RNA secondary structures,
  • a set of allowed **edit operations** and
  • a **cost** for each possible operation,
- compute
  • an **alignment of minimum cost** between the two structures.

RNAStrAT uses the edit distance model defined in [Jiang et al. 2002] that comports:
• single base edit operations (substitution/insertion/deletion),
• base pairs operations such as insertion/deletion, creation/opening or alteration of a hydrogene bond.
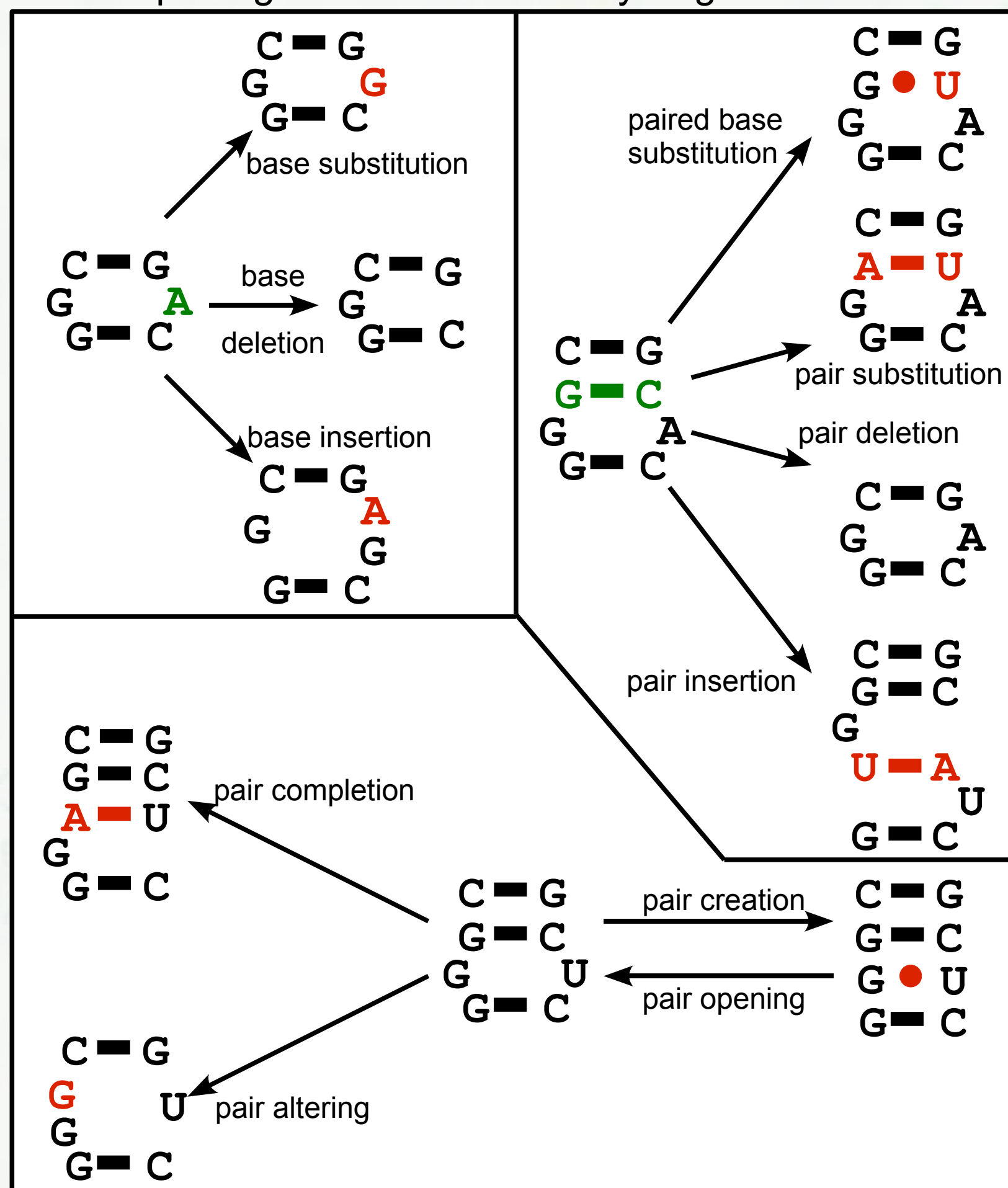

Figure 3. Edit operations.

This model was introduced by [Jiang et al., 2002]. Computing the distance is an NP-hard problem, but several less general version of this problem can be solved exactly and are used in widely in RNA secondary strcutures comparison tools such as RNAForrester [Höchsmann et al., 2004].

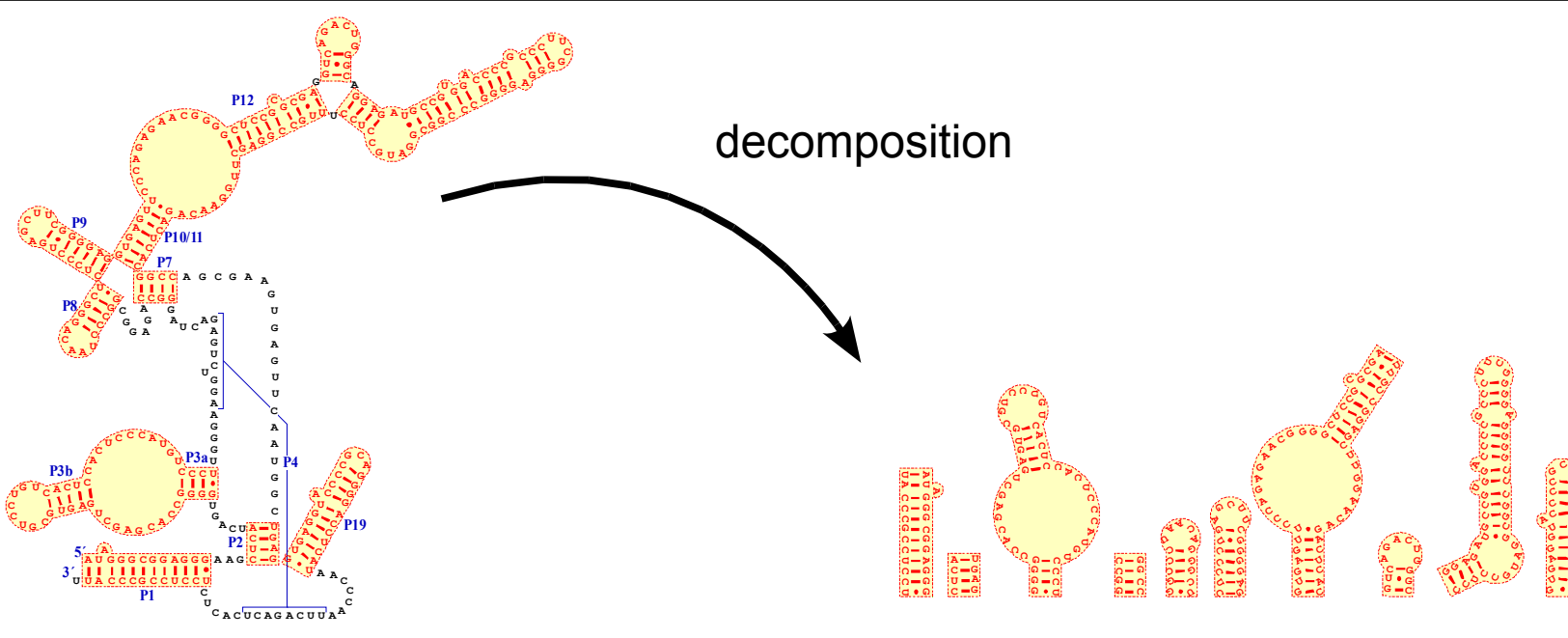## A STEM/STEM-LOOP DECOMPOSITION BASED HEURISTIC

Figure 4. Stem/stem-loop decomposition of an Rnase P structure.

RNA secondary structures are decomposed into stems and stem-loops, that are then compared using an exact algorithm for the conservative edit distance [Guignon et al., 2005], a variant of the general distance of [Jiang et al., 2002].
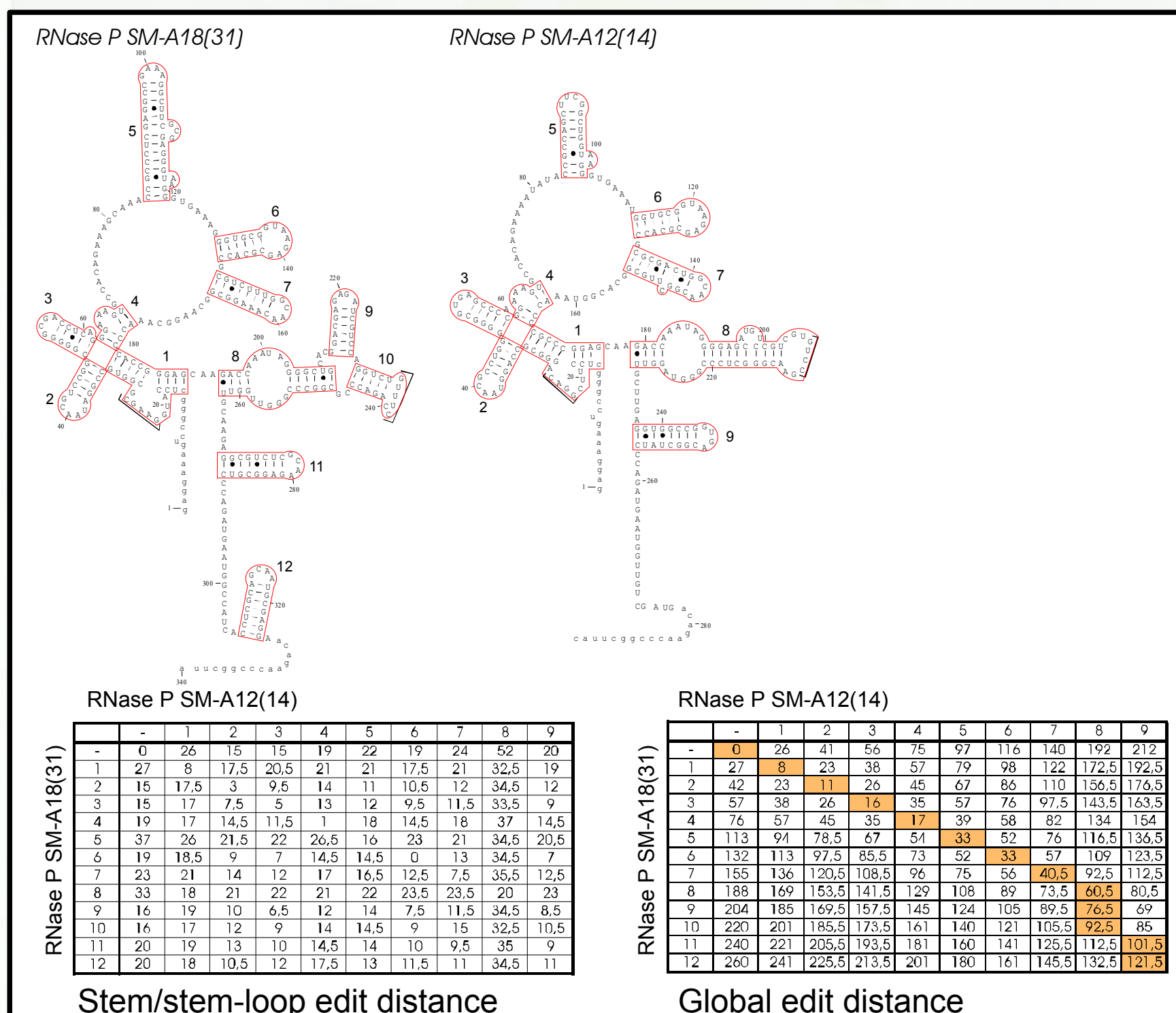

Figure 5. Edit distance computation between 2 Rnase P structures.

Then, these hairpins pairwise comparison are used in a Smith-Waterman based heuristic to produce a distance and alignment between the two complete structures.

ADDITIONAL FEATURES AND DEVELOPMENT

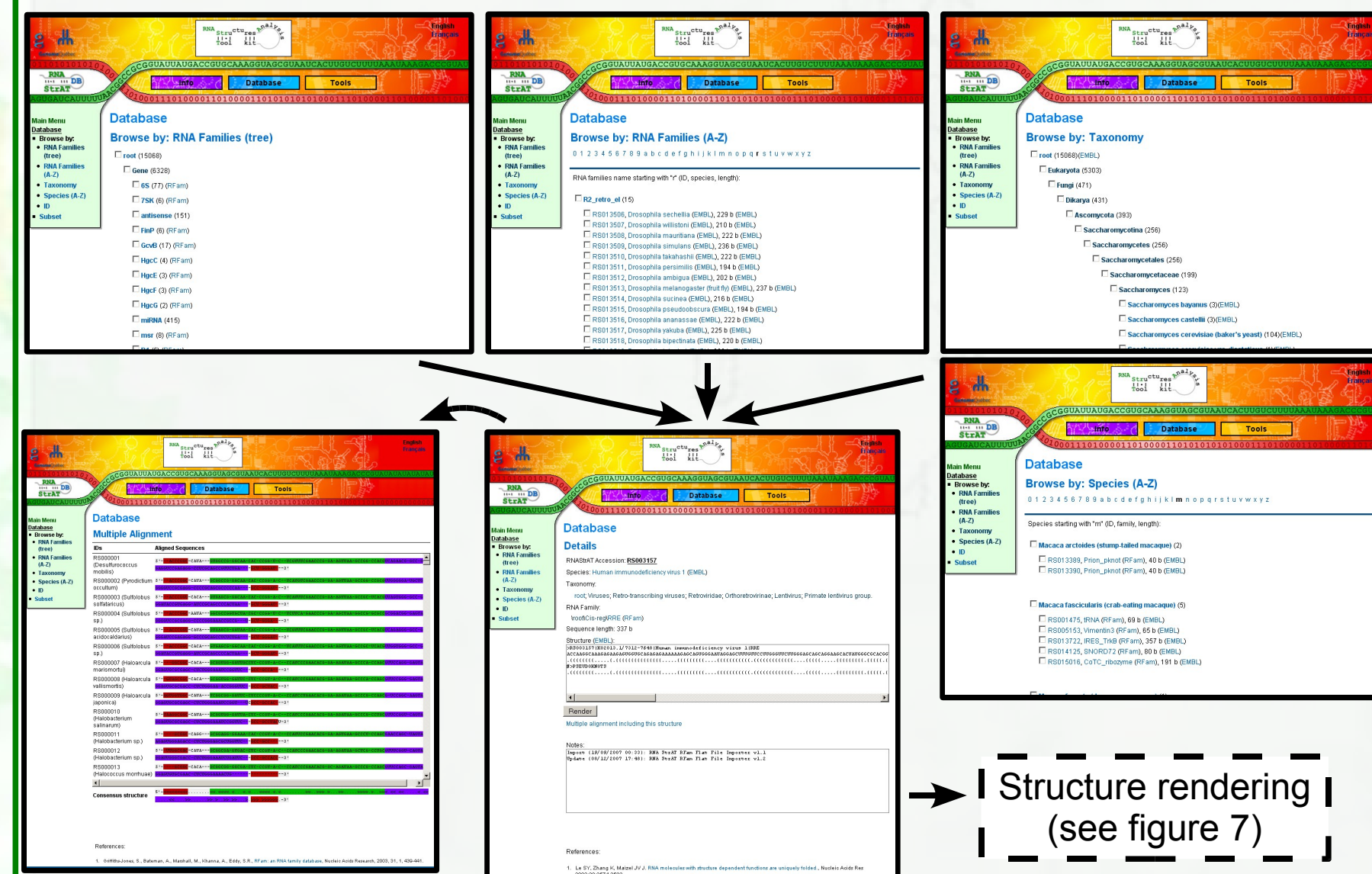### The database of RNA secondary structures

Figure 6. Database browsing features.

Users can access to structure information including links to its RNA family (in the Rfam classification), its organism taxonomy (EMBL), its sequence (EMBL). Structures stored in our database are extracted from the Rfam seed alignments and for each RNA gene, its specific secondary structure is obtained from both its sequence and the family consensus structure.

Database search improvements

In order to speed-up the database search the search engines first analyze the structural characteristics of the query structure to select a group of candidates in the database that share similar characteristics close to the query ones, eliminating at the same time irrelevant structures. Then, the query structure is compared to these candidates to find which ones have the best similarity scores. The user can modify the parameters that define the candidates.
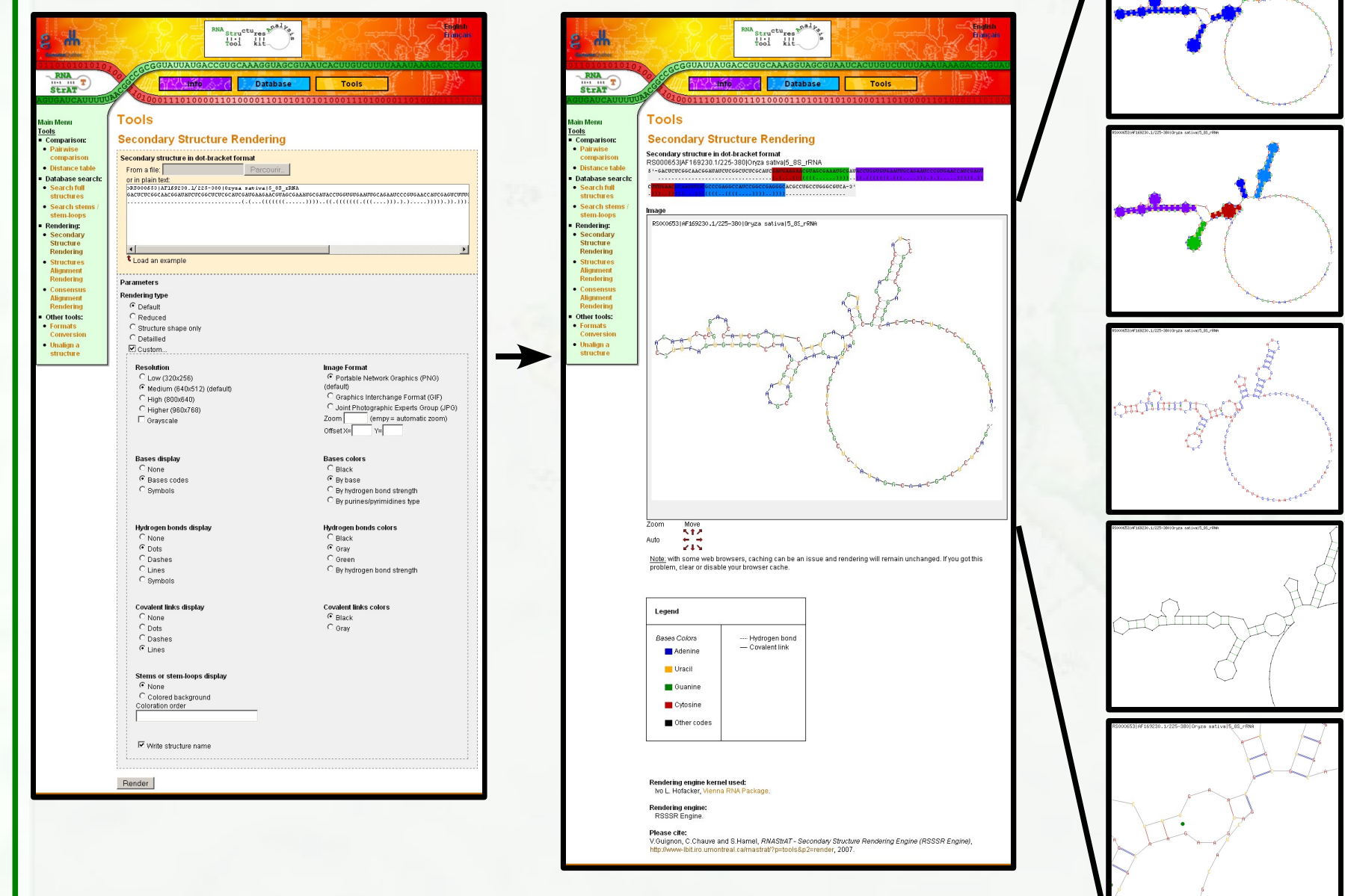
Structures and alignment rendering

Figure 7. Structure online rendering (base on Vienna RNA Package): the rendering form (on the left) enables the user to display structures in various ways (on the right) and browse the structure (middle).
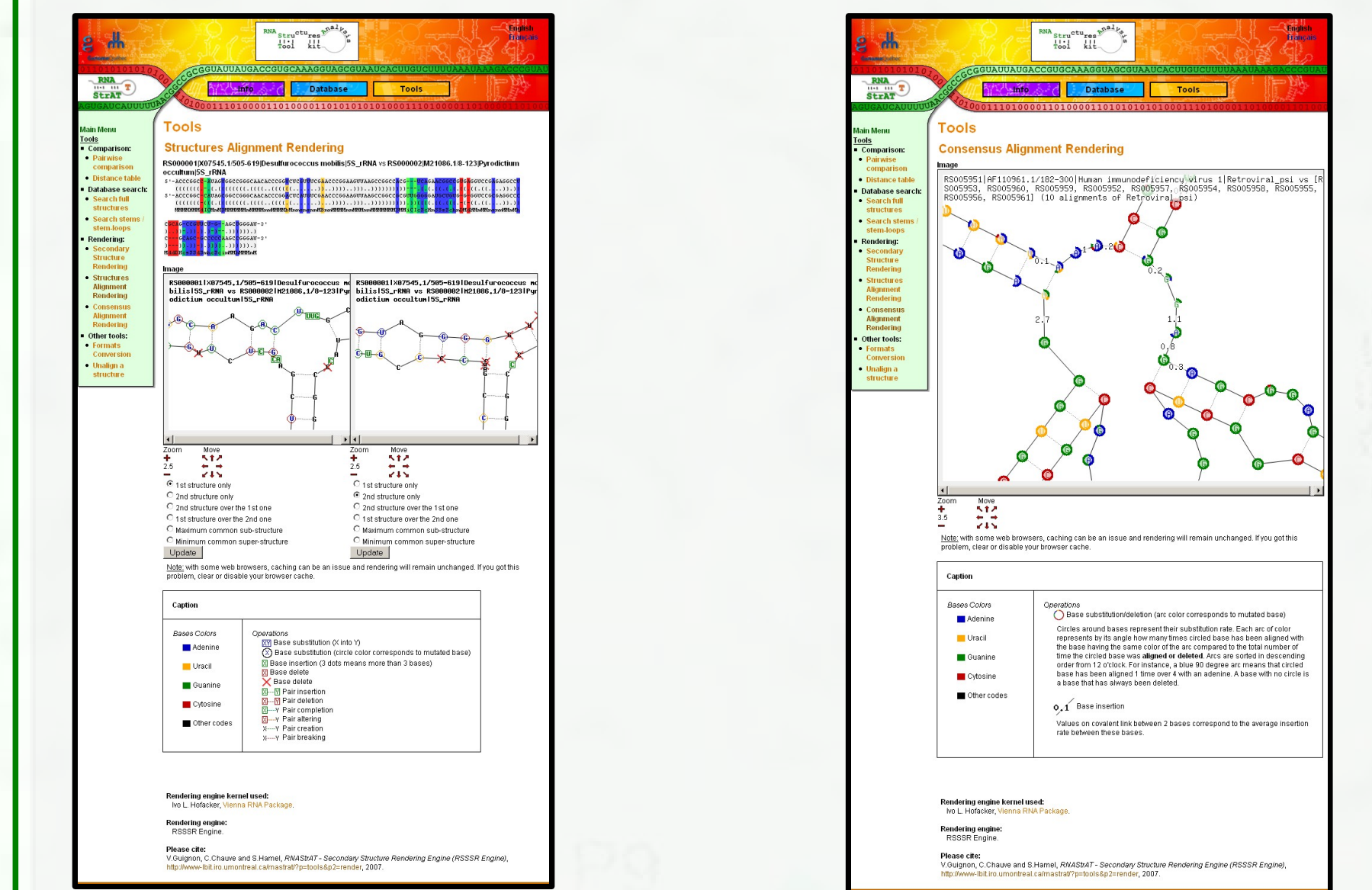

Figure 8. The rendering engine enables to compare two aligned structures (on the left) and see which bases changed from a structure to the other. The characteristics of a structure compared to a set of structures can also be rendered (on the right). Each base of the query structure is displayed with a pie chart that shows how often the base has been kept, replaced by an other one or deleted.

REFERENCES

Griffiths-Jones S., Bateman A., Marshall M., Khanna A., Eddy S.R. RFam: an RNA family database, Nucleic Acids Research, 2003, 31, 1, 439-441.

Jiang T., Lin G., Ma B., Zhang K. A general edit distance between RNA structures. J. Comput. Biol., 9(2):371–388. 2002.

Guignon V., Chauve C., Hamel S. Distance d'édition entre tige-boucles, JOBIM 2005, 2005, poster 82.

Smith T.F., Waterman M.S. Identification of Common Molecular Subsequences, Journal of Molecular Biology 147: 195–197. doi:10.1016/0022-2836(81)90087-5, 1981.

Höchsmann M., Voss B., Giegerich R. Pure Multiple RNA Secondary Structure Alignments: A Progressive Profile Approach in IEEE/ACM Transactions on Computational Biology and Bioinformatics, 1(1), 2004, pp53-62.

Vienna RNA Package, http://www.tbi.univie.ac.at/~ivo/RNA