

# Structure and evolution of self-assembly modules in RNA folding

Eric Westhof

Institut de Biologie Moléculaire et Cellulaire du CNRS,  
Bioinformatique, modélisation et simulations des acides nucléiques,  
UPR 9002 'Architecture et Réactivité de l'ARN',  
Université Louis Pasteur, 15 rue René Descartes  
67084 Strasbourg Cedex, France  
<http://www-ibmc.u-strasbg.fr/upr9002/westhof/>

RNA architecture is visualized as the hierarchical assembly of preformed double-stranded helices defined by Watson-Crick base pairs and RNA modules maintained by non-Watson-Crick base pairs. This architectural hierarchy is coupled with an electrostatic hierarchy whereby RNA folding occurs first with an electrostatic collapse to compact states with most of the secondary structure elements induced by non-specific ion binding. Later, there is a cooperative transition to native states with all tertiary contacts induced by specific ion binding, especially magnesium ions. We have assembled several RNAs on the basis of previously identified RNA-RNA contacts and sequence analysis. Crystal structures, published after modelling, are now available for comparisons. RMS values ranging between 3.7 and 8.5 Å are obtained depending on the system being compared. Comparisons between NMR and X-ray structures give rms values around 5 Å for two RNA structures. Under the hypothesis that homologous sequences fold into similar 3D structures, a systematic analysis of sequences in the light of X-ray structures allows to derive covariation rules for non-Watson-Crick base pairs using a geometric classification of non-Watson-Crick pairs. First, we search the sequence space available to a given 3D motif present in the structural database. Secondly, this knowledge is used to identify regions in sequences compatible with a given 3D motif. The process goes from structure space to sequence space and then back from a sequence to the prediction of a fold. Along those lines, the occurrences of two recurrent motifs in ribosomal RNA sequences, the Kink-turn and the C-loop, have been examined in crystal structures and systematically compared with sequence alignments of rRNAs from the three kingdoms of life in order to identify the range of the structural and sequence variations. Similarly, the three-way junctions contained in X-ray structures of folded RNAs have been compiled and analysed. Three-way junctions with two helices approximately co-axially stacked can be divided into three main families depending on the relative lengths of the segments linking the three Watson-Crick helices. Each family has topological characteristics with some conservation in the non-Watson-Crick pairs within the linking segments as well as in the types of contacts between the segments and the helices. The most populated family presents tertiary interactions between two helices as well as extensive shallow/minor groove contacts between a linking segment and the third helix.

Leontis NB, Westhof E. 2003. Analysis of RNA motifs. *Curr Opin Struct Biol* 13:300-308.

Lescoute A, Leontis NB, Massire C, Westhof E. 2005. Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. *Nucleic Acids Res* 33:2395-2409.