

# The Architectural Modules of RNA

Les modules architecturaux de l'ARN

Eric Westhof

---

Architecture et Réactivité de l'ARN, Université de Strasbourg,  
Institut de Biologie Moléculaire et Cellulaire, CNRS,  
15 rue René Descartes, 67084 Strasbourg, France  
E-mail: [e.westhof@ibmc-cnrs.unistra.fr](mailto:e.westhof@ibmc-cnrs.unistra.fr)  
<http://www-ibmc.u-strasbg.fr/arn/Westhof/>

RNA molecules are characterized by the formation of hydrogen bonded pairs between the nucleotide bases along the polymer. All base-base interactions present in nucleic acids can be classified in twelve families where each family is a 4x4 matrix of the usual bases. The common Watson-Crick pairs belong to one of these families and the other eleven families gather the non-Watson-Crick pairs. The Watson-Crick pairs form the secondary structure and all the other families are critical for the tertiary structure. RNA architecture can then be viewed as the hierarchical assembly of preformed doublestranded helices defined by Watson-Crick base pairs and RNA modules maintained by non-Watson-Crick base pairs. RNA modules are recurrent ensemble of ordered non-Watson-Crick base pairs. The geometrical constraints attached to each base pairing family explain the surprising molecular neutrality observed in sequences and structures during biological evolution. Through systematic comparisons between homologous sequences and x-ray structures, followed by automatic clustering, the whole range of sequence diversity in recurrent RNA modules has been characterized. These data permitted the construction of a computational pipeline for identifying known 3D structural modules in single and multiple RNA sequences in the absence of any other information.

Any module can in principle be searched, but four can be searched automatically: the Gbulged loop, the Kink-turn, the C-loop and the tandem GA loop. The present pipeline can be used for RNA 2D structure refinement, 3D model assembly, and for searching and annotating structured RNAs in genomic data.