Automated Geometric Search for Structural Motifs in RNA 3D Structures

Jesse Stombaugh, Bowling Green State University

Find RNA 3D (**FR3D**), suite of programs for exhaustively searching RNA 3D atomic-resolution structures for recurrent RNA motifs, will be described. Recurrent RNA 3D motifs are sets of RNA nucleotides with similar spatial arrangements. They can be local or composite. Local motifs comprise nucleotides that occur in the same hairpin or internal loop. Composite motifs comprise nucleotides belonging to three or more different RNA strand segments or molecules. A base-centered approach was implemented to construct efficient, yet exhaustive search procedures using geometric, symbolic, or mixed representations of RNA structure. Each base is represented geometrically by the position of its geometric center in 3D space and by the rotation matrix that describes its orientation with respect to a common frame. Base--pairing and base--stacking interactions are calculated from the base geometries and are represented symbolically according to the Leontis/Westhof basepairing classification, extended to include base--stacking. These data are stored and used to organize motif searches. For geometric searches, the user supplies the 3D structure of a query motif which **FR3D** uses to systematically find and score geometrically similar candidate motifs, without regard to the sequential position of their nucleotides in the RNA chain. To score and rank candidate motifs, FR3D calculates a geometric discrepancy by rigidly rotating candidates to align optimally with the query motif and then comparing the relative orientations of the corresponding bases in the query and candidate motifs. Given the growing size of the RNA structure database, it is impossible to explicitly compute the discrepancy for all conceivable candidate motifs, even for motifs with less than ten nucleotides. The screening algorithm that we describe finds all candidate motifs whose geometric discrepancy with respect to the query motif falls below a user-specified cutoff discrepancy. Candidate motifs identified geometrically may be further screened symbolically to identify those that contain particular basepair types or base--stacking arrangements or that conform to sequence continuity or nucleotide identity constraints. Purely symbolic searches for motifs containing user-defined sequence, continuity and interaction constraints have also been implemented. We demonstrate that **FR3D** finds all occurrences of sarcin/ricin and kink--turn motifs in the 23S and 5S ribosomal RNA 3D structures of the *H. marismortui* 50S ribosomal subunit and assigns the lowest discrepancy scores to bona fide examples of these motifs. The search algorithms have been optimized for speed to allow users to search the non--redundant RNA 3D structure database on a personal computer in a matter of seconds or minutes.