Locomotif - a Graphical Programming System for RNA Motif Search

Janina Reeder Peter Steffen Robert Giegerich

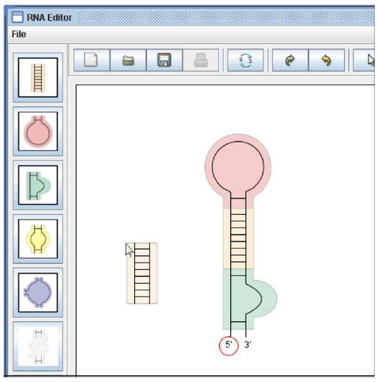
Faculty of Technology, University Bielefeld, Bielefeld, Germany
Email: {janina,psteffen,robert}@techfak.uni-bielefeld.de

We describe a graphical programming system for a domain specific language in biosequence analysis [2]. It supports the development of programs for RNA structure prediction and motif search, created by biologists with little or no programming skills. The system combines several programming paradigms in a productive way. It has a client - server architecture, with a transport layer in XML. The graphical frontend is implemented in the object-oriented paradigm (using Java). Graphics are compiled into a declarative domain-specific language for dynamic programming (ADP) that is embedded in Haskell. Finally, motif search programs expressed in ADP are compiled to imperative code in C, a step which includes substantial domain-specific optimization.

RNA structural motifs RNA is the active form of genetic information. It is transcribed from DNA as a chain of A, C, G and U, where U denotes Uracil, the RNA representative of Thymine. Possible base pairs in RNA are G–C, A–U and also G–U. RNA is typically single stranded, and by folding back onto itself, it forms structure. Structure formation is driven by the forces of hydrogen bonding between base pairs, and energetically favorable stacking of base pairs. Typical elements found in RNA secondary structure are stacking regions (or helices), bulge loops, internal loops, hairpin loops and multiple loops.

An RNA structural motif is independent of a concrete RNA sequence. It can be seen as a parameterized structure that specifies a particular arrangement of structural features. This arrangement may be annotated with additional information to make the motif more specific.

Locomotif In Locomotif, RNA motifs are composed of a limited set of "building blocks", namely stems, bulges, internal loops, hairpin loops, multiloops, and single stranded regions. See the figure for a screenshot of the Locomotif system. These building blocks are combined by the user to a complete RNA motif. When the design process is finished, the user can choose to execute the prediction program for this motif. This comprises several steps. First, the motif is converted into an XML representation, which is then further translated into a program in the ADP language [1]. This ADP program is then sent to a webserver, which compiles it into executable C code using the ADP compiler [3]. The generated C program is then compiled on the webserver and made available to the user via a web interface. Alternatively, the user can also choose to receive the C source code of the generated program to compile it on his own machine.



References

[1] R. Giegerich, C. Meyer, and P. Steffen. A discipline of dynamic programming over sequence data. *Science of Computer Programming*, 51(3):215–263, 2004.

[2] J. Reeder. *Locomotif - a Graphical Programming System for RNA Motif Search*. PhD thesis, Bielefeld University, 2006.

[3] P. Steffen. *Compiling a Domain Specific Language for Dynamic Programming*. PhD thesis, Bielefeld University, 2006.