The Vienna RNA Package 2.0

Stephan Bernhart
Ivo Hofacker
Christian Höner zu Siederdissen
Ronny Lorenz

October 22, 2009

Vienna RNA Package 2.0

Extended nearest neighbor model (Turner '04):

- tri-, hexa-loop parameters
- mismatch energies for exterior and multibranch loops
- 1xN-, 2x3-interior loops
- enthalpies for all energy contributions
- updated energy contributions

New programs and API functions:

- RNA2Dfold
- direct path heuristic (`findpath`)
- energy parameter file conversion tool (Version 1.x $\rightarrow$ 2.x)

Vienna RNA Package 2.0

What was/is going on behind the curtain?

- partially rewritten code for almost all included programs
- changes in API to simplify code linked against `RNAlib`
- backward compatibility for code linked against Version 1.x
- `doxygen` documentation style for API functions
- no `Perl` bindings for new functions yet :(

RNAlib C/C++ API

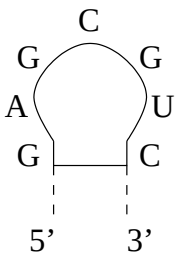
What has changed?

Reusable key functions:

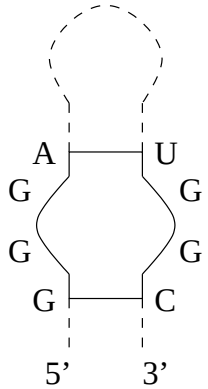
- common initiation procedures
- datastructures containing precalculated energy contributions
- loop energy evaluation (almost) independent from candidate generating algorithm
- One looptype $\Rightarrow$ one function
- loop energy functions include dangling end and mismatch contributions

Scheme for all present and future ViennaRNA programs

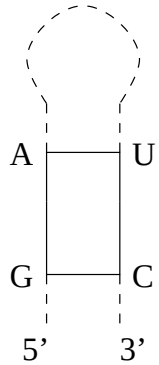
Loop Types in RNA Folding



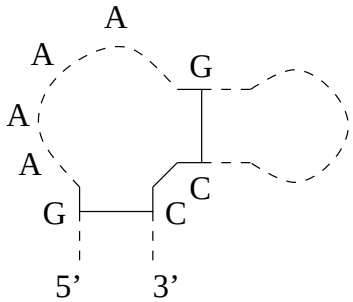
Hairpin loop



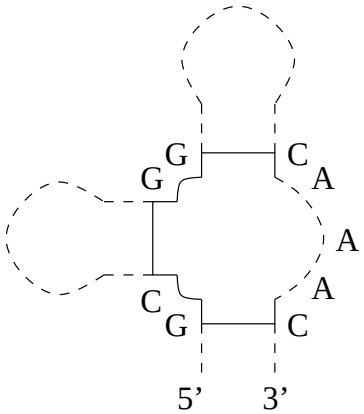
Interior loop



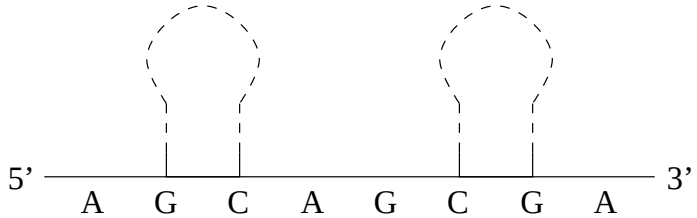
Stacking pair



Bulge loop

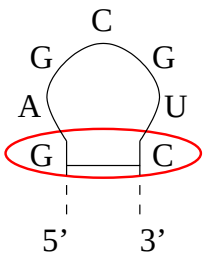


Multibranch loop

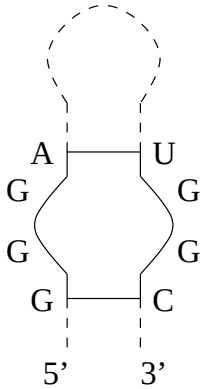


Exterior loop

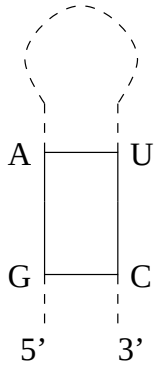
Necessary Loop Attributes for Energy Contribution



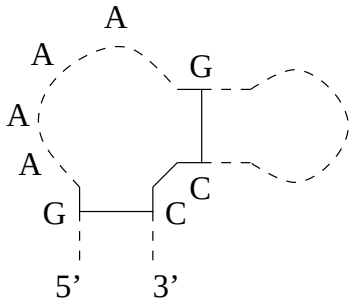
Hairpin loop



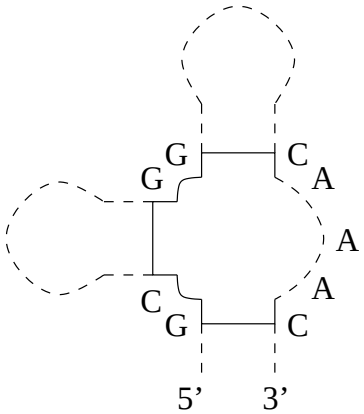
Interior loop



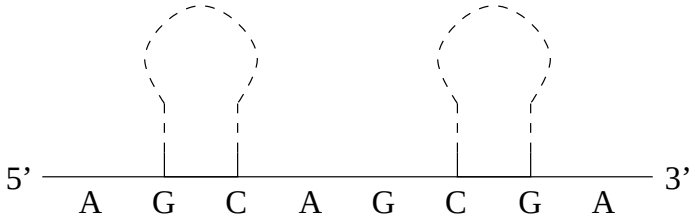
Stacking pair



Bulge loop

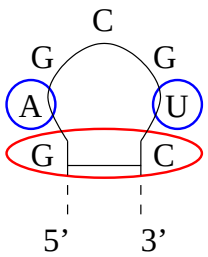


Multibranch loop

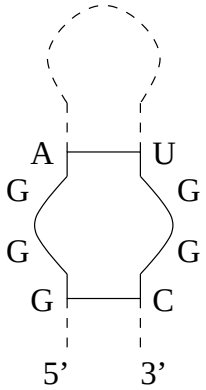


Exterior loop

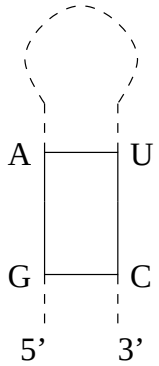
Necessary Loop Attributes for Energy Contribution



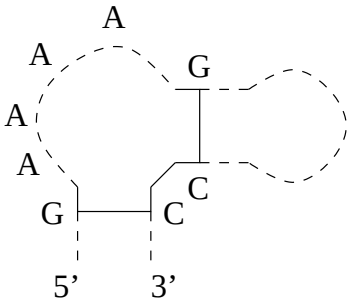
Hairpin loop



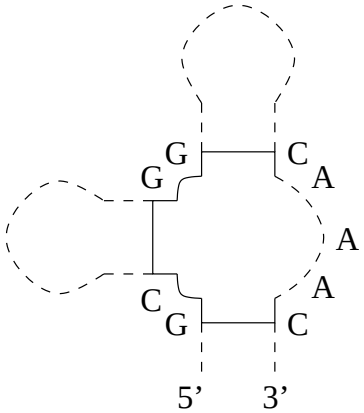
Interior loop



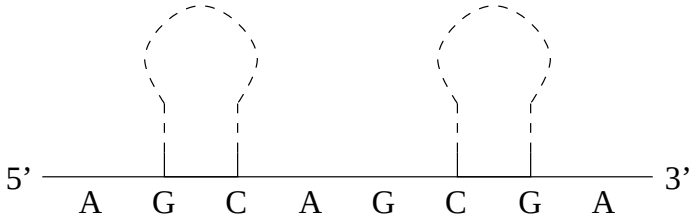
Stacking pair



Bulge loop

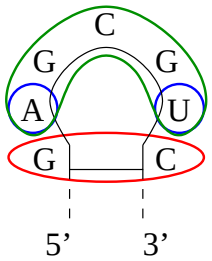


Multibranch loop

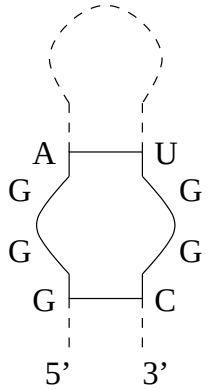


Exterior loop

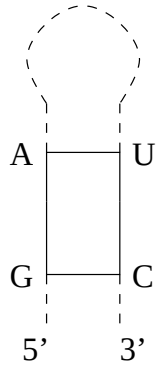
Necessary Loop Attributes for Energy Contribution



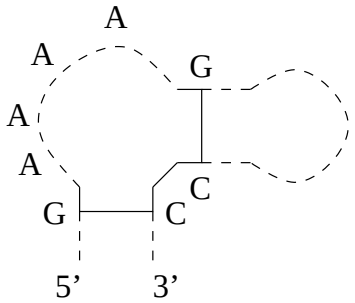
Hairpin loop



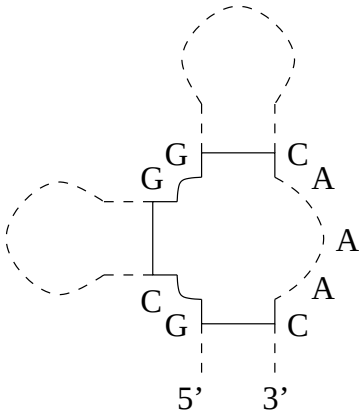
Interior loop



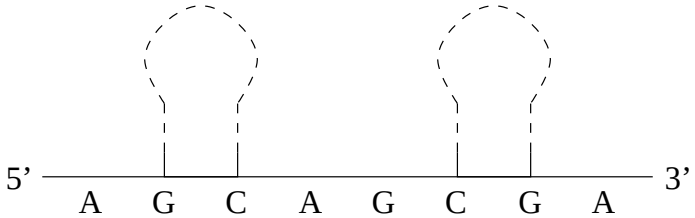
Stacking pair



Bulge loop

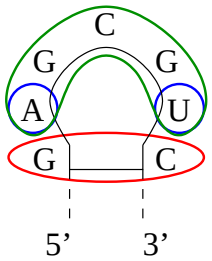


Multibranch loop

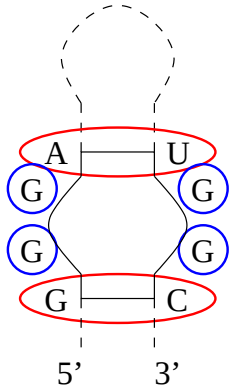


Exterior loop

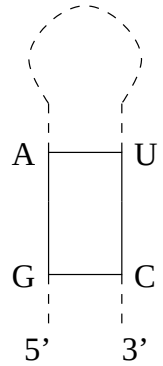
Necessary Loop Attributes for Energy Contribution



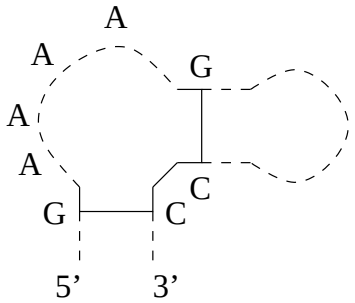
Hairpin loop



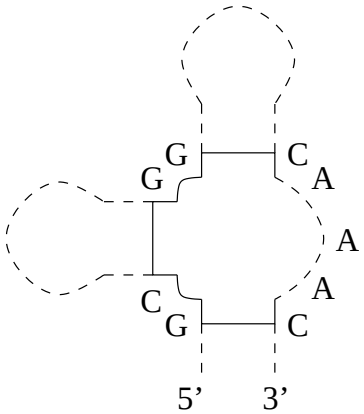
Interior loop



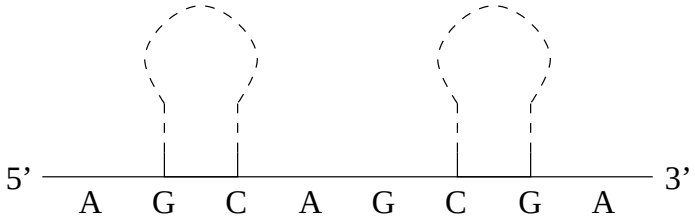
Stacking pair



Bulge loop

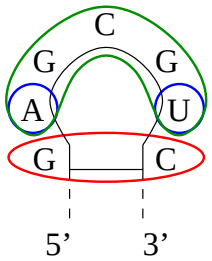


Multibranch loop

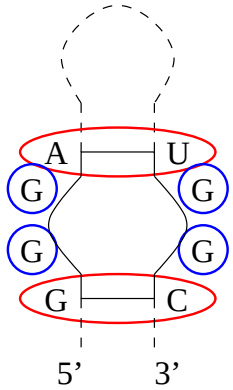


Exterior loop

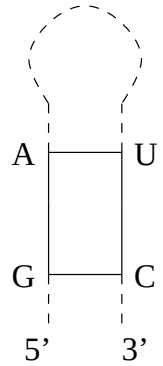
Necessary Loop Attributes for Energy Contribution



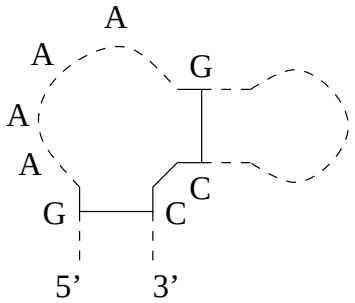
Hairpin loop



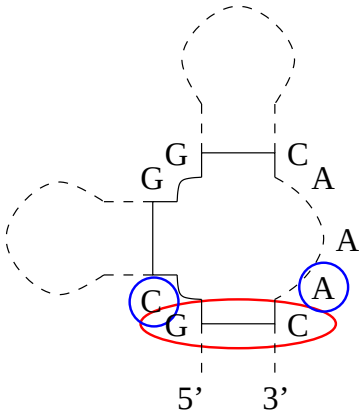
Interior loop



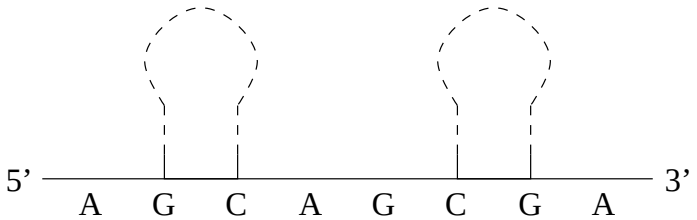
Stacking pair



Bulge loop

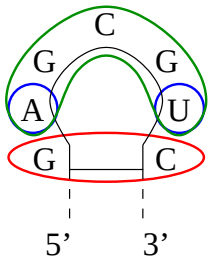


Multibranch loop

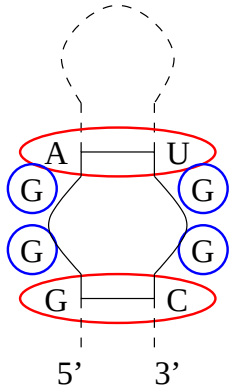


Exterior loop

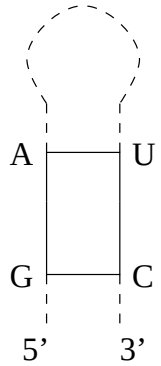
Necessary Loop Attributes for Energy Contribution



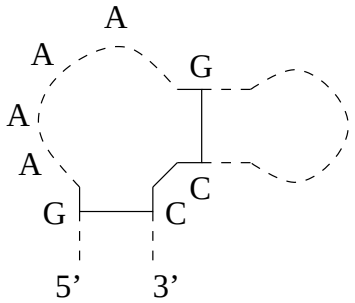
Hairpin loop



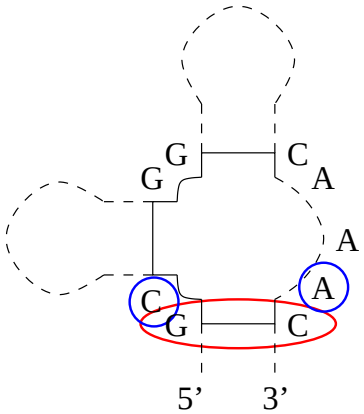
Interior loop



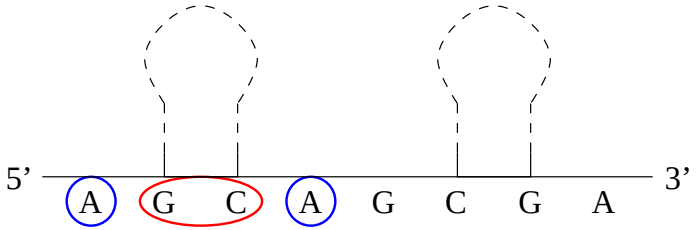
Stacking pair



Bulge loop

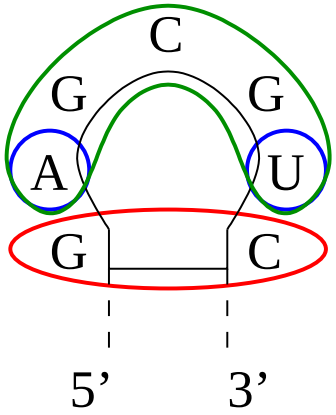


Multibranch loop



Exterior loop

Loop Type Specific Energy Contributions

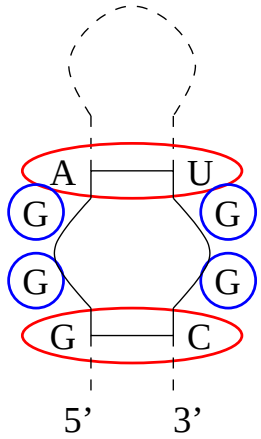


Hairpin loop

`E_Hairpin(length, type, mm1, mm2, sequence, energies)`

$$E_{\text{hairpin}} = \begin{cases} \gamma_1(\text{sequence}) & \text{short tabulated loopsequence} \\ \gamma_2(\text{length}) + \gamma_3(\text{type}, \text{mm}_1, \text{mm}_2) & \text{otherwise} \end{cases}$$

Loop Type Specific Energy Contributions

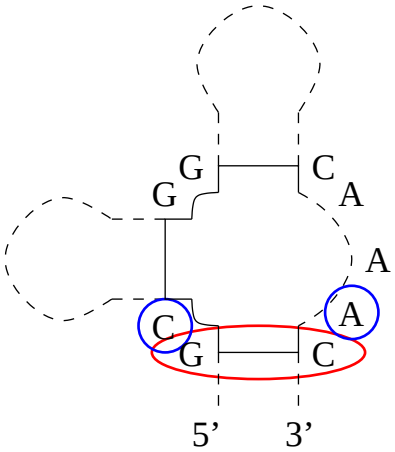


Interior loop

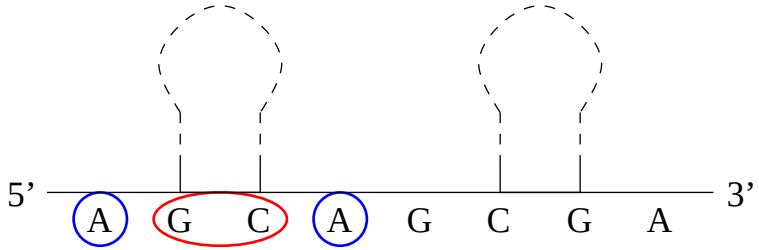
```
E_IntLoop(L1, L2, type, type2, mm1, mm2, mm3, mm4, energies)
```

$$E_{\text{intloop}} = \begin{cases} \gamma_4(\text{type}, \text{type}_2) & L_1 = L_2 = 0 \\ \gamma_5(\max(L_1, L_2)) & L_1 = 0 \text{ or } L_2 = 0 \\ \gamma_6(\text{type}, \text{type}_2, \text{mm}_1, \text{mm}_2, \text{mm}_3, \text{mm}_4) & \text{special intloop} \\ \gamma_7(L_1 + L_2) + \gamma_8(\text{type}, \text{mm}_1, \text{mm}_2) + \gamma_9(\text{type}_2, \text{mm}_3, \text{mm}_4) & \text{otherwise} \end{cases}$$

Loop Type Specific Energy Contributions



Multibranch loop



Exterior loop

```
E_MLstem(type, mm1, mm2, energies)
E_ExtLoop(type, mm1, mm2, energies)
```

$$E_{\text{stem}} = \begin{cases} \gamma_{10}(\text{type}) + d_5(\text{type}, \text{mm}_1) & \text{mm}_2 \text{ not available} \\ \gamma_{10}(\text{type}) + d_3(\text{type}, \text{mm}_2) & \text{mm}_1 \text{ not available} \\ \gamma_{10}(\text{type}) + \gamma_{11}(\text{type}, \text{mm}_1, \text{mm}_2) & \text{otherwise} \end{cases}$$

Experimental Multibranch Folding Algorithm

ML Parameter h: helices, u: unpaired bases

Linear $\alpha + \beta h + \gamma u$

Affine $\alpha + \beta h + \sum_{u \in U}([\gamma_{\text{init}} | u > 0] + [\gamma_{\text{extend}}(u-1) | u > 1])$

Logarithmic $\alpha + \beta h + \sum_{u \in U}([\gamma_{\text{init}} | u > 0] + [\gamma_{\text{extend}} \log u | u > 0])$

(implemented in Haskell)

Performance Measures

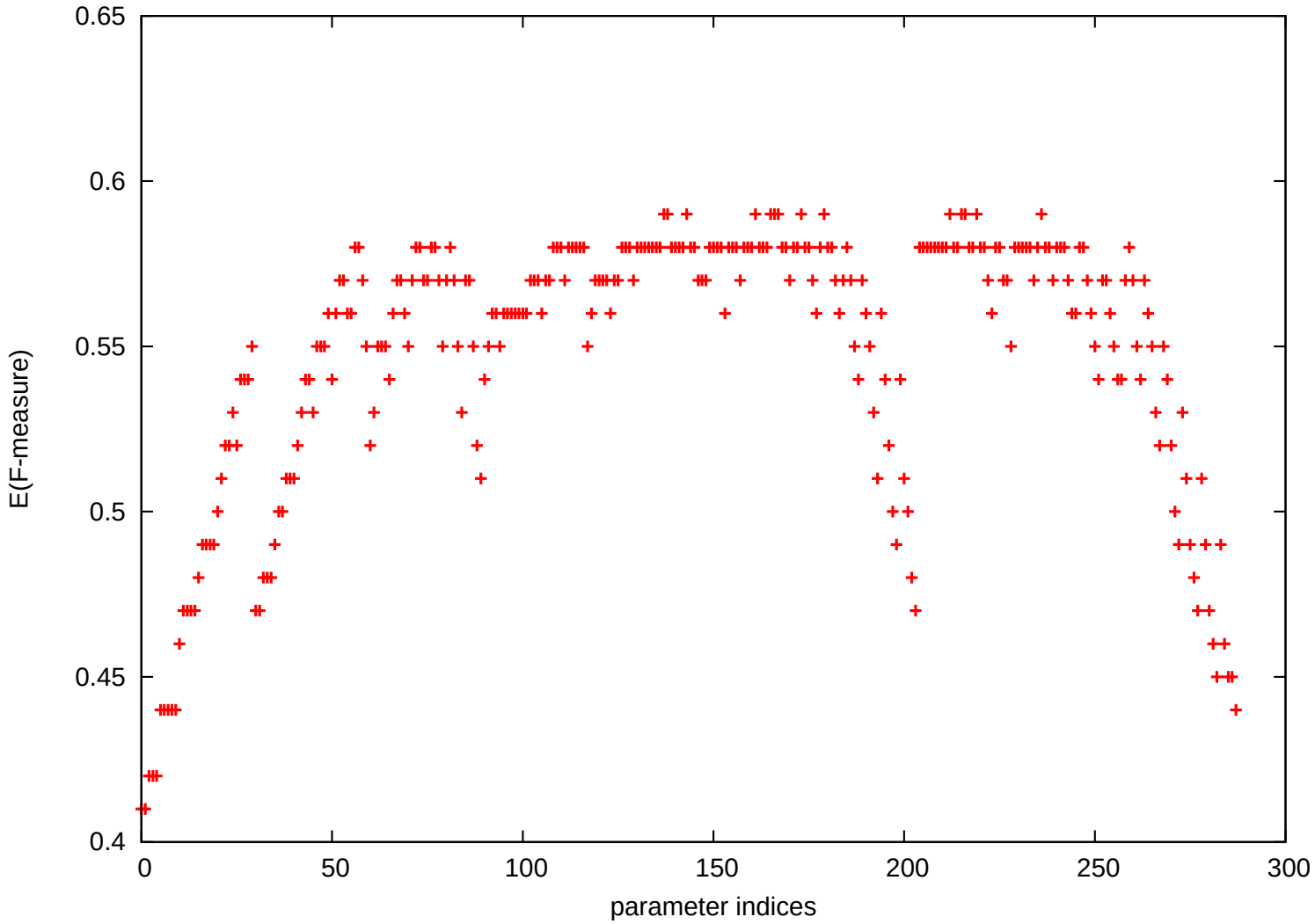
$$\text{Sensitivity} \quad \frac{\#\text{ correct pairs}}{\#\text{ reference pairs}}$$

$$\text{PPV} \quad \frac{\#\text{ correct pairs}}{\#\text{ predicted pairs}}$$

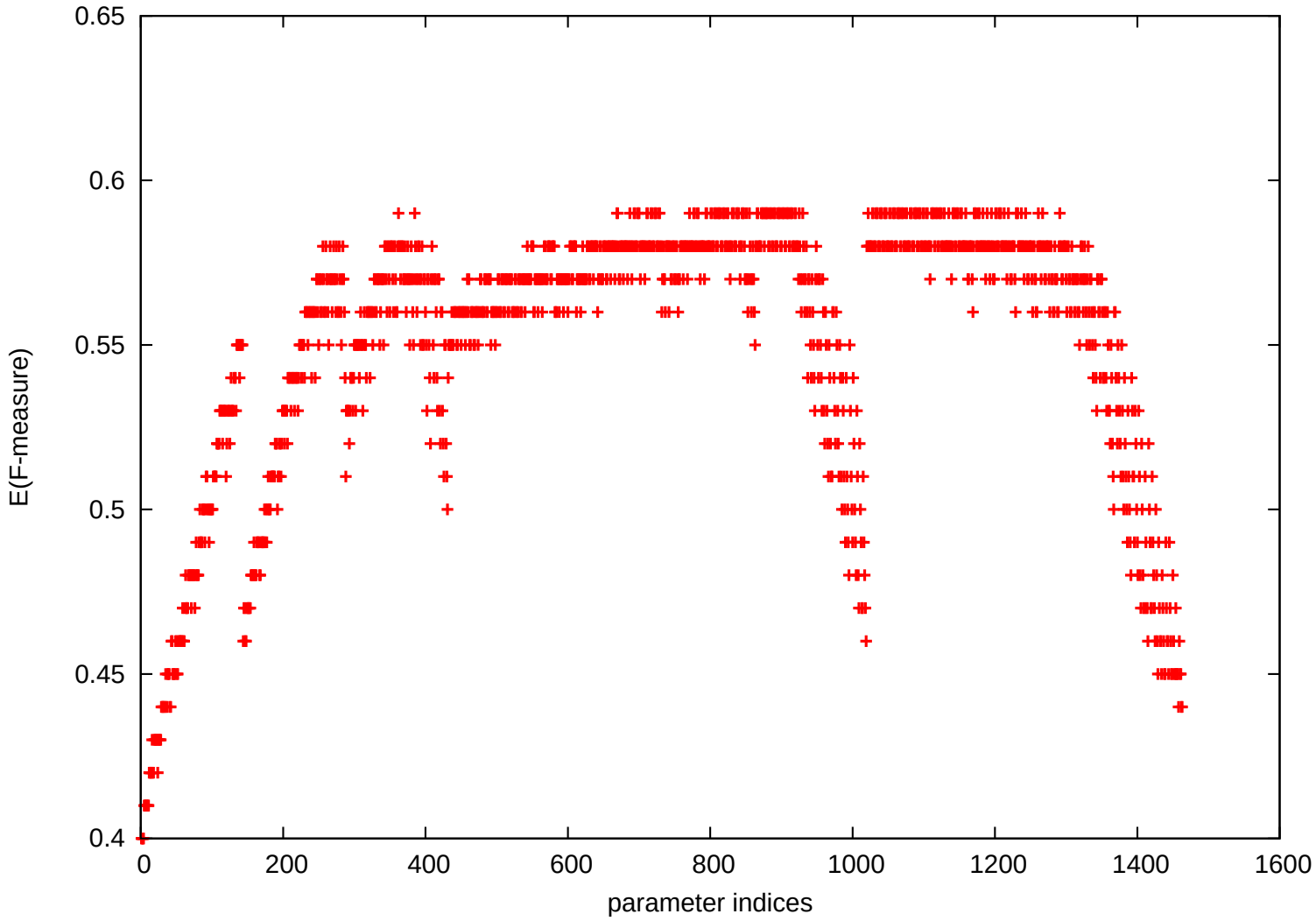
$$\text{F-measure} \quad \frac{2 \times \text{Sensitivity} \times \text{PPV}}{\text{Sensitivity} + \text{PPV}}$$

cf. Andronescu et al

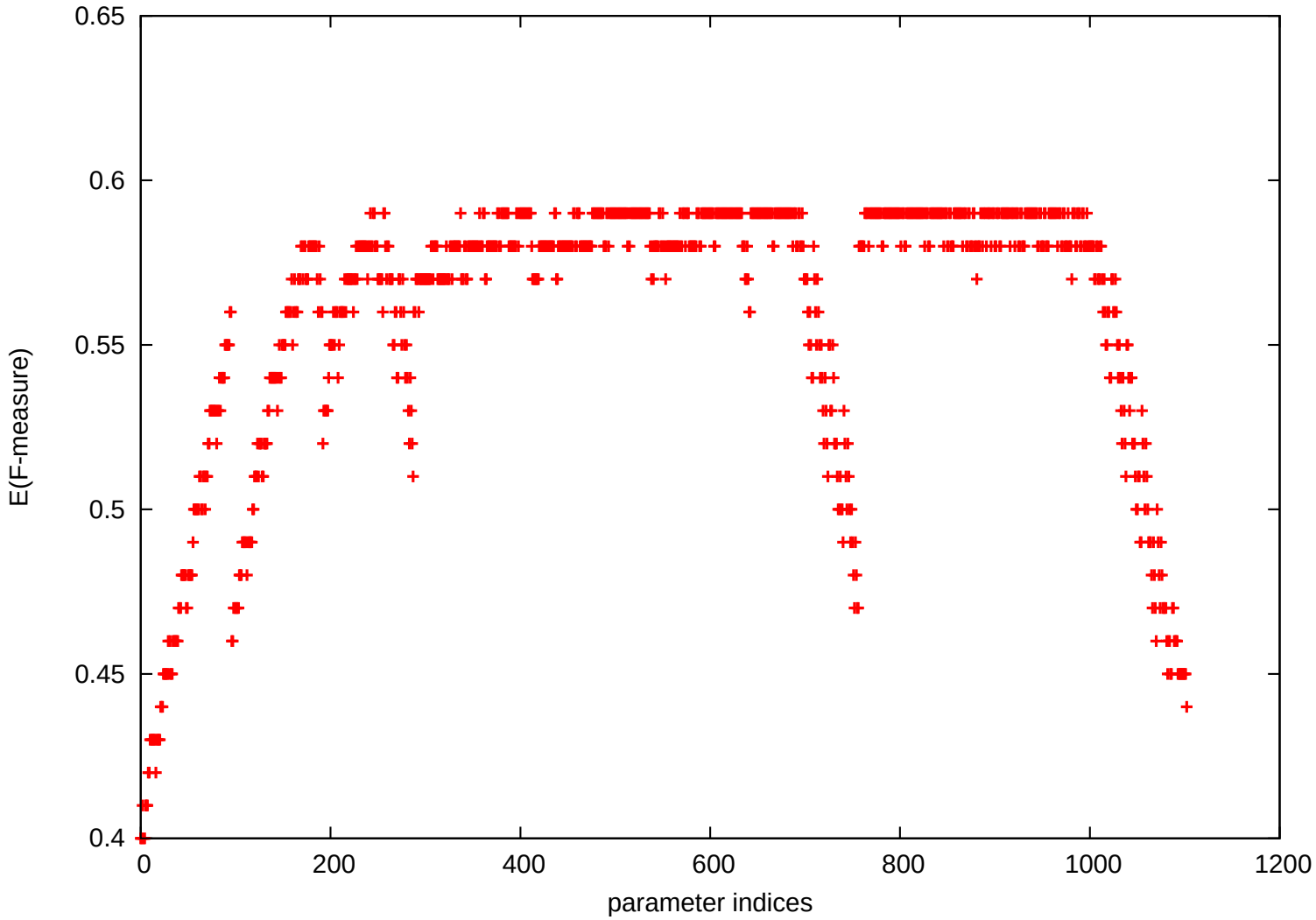
Linear Multibranch Energy Function



Affine Multibranch Energy Function



Logarithmic Multibranch Energy Function



Best Results

Type	α	β	γ_{init}	γ_{extend}	F-measure	PPV
Turner '99	340	40		0	0.48	0.50
Andronescu '99	315	15		-2	0.51	0.55
Turner '04	930	-90		0	0.58	0.61
Linear	900	-80		-5	0.59	0.63
Affine	900	-90	-60	5	0.59	0.63
Log	950	-90	-60	-30	0.59	0.62

- $\leq$ 500 nc, no pseudoknots, 1222 structures, rnastrand 2.0 DB
- different training / test set than Andronescu et. al

Summary

- improved API:
  one structure $\longrightarrow$ one function
- new energy tables:
  specialized interior loops & mismatches
- more energy parameter sets:
  Turner '99 / '04 & Andronescu '99
- Andronescu '04 expected to arrive soon
- if you dare: Haskell'ized RNAfold with more options

Andronescu et al

`http://people.cs.ubc.ca/~androness/param-training/`