Abstract: Using the Fast Fourier Transform to accelerate the computational search for RNA conformational switches

Evan Senter[1], Saad Sheikh[2], Ivan Dotu[1], Yann Ponty[3], and Peter Clote[1]

[1] Biology Department, Boston College, Chestnut Hill MA 02467 USA
`clote@bc.edu,ivan.dotu@bc.edu,evansenter@gmail.com`
[2] Computer Science Department, University of Florida, Gainesville FL 32601 USA
`sheikh@cise.ufl.edu`
[3] Laboratoire d'Informatique, Ecole Polytechnique, F-91128 Palaiseau Cedex, France,
`yann.ponty@lix.polytechnique.fr`

Abstract. We describe the broad outline of a new thermodynamics-based algorithm, `FFTbor`, that uses the fast Fourier transform to perform polynomial interpolation to compute the Boltzmann probability that secondary structures differ by k base pairs from an arbitrary reference structure of a given RNA sequence. The algorithm, which runs in quartic time $O(n^4)$ and quadratic space $O(n^2)$, is used to determine the correlation between kinetic folding speed and the *ruggedness* of the energy landscape, and to predict the location of riboswitch expression platform candidates. The full paper appears in *PLoS ONE* (2012) 19 Dec 2012. A web server is available at `http://bioinformatics.bc.edu/clotelab/FFTbor/`.

Keywords: RNA secondary structure; partition function; fast Fourier transform; Lagrange interpolation

In [2], we developed a dynamic programming algorithm, `RNAbor`, which simultaneously computes for each integer k, the Boltzmann probability $p_k = \frac{Z_k}{Z}$ of the subensemble of structures whose base pair distance to a given *initial*, or *reference*, structure S^* is k.[4] `RNAbor` stores the value of the (partial) partition functions $Z_k(i,j)$ for all $1 \leq i \leq j \leq n$ and $0 \leq k \leq n$, each of which requires quadratic time to compute. Thus it follows that `RNAbor` runs in time $O(n^5)$ and space $O(n^3)$, which severely limits its applicability to genomic annotation. This restriction is somewhat mitigated by the fact that in [1], we showed how to use sampling to efficiently approximate `RNAbor` in cubic time $O(n^3)$ and quadratic space $O(n^2)$, *provided* that the starting structure S^* is the minimum free energy (MFE) structure. We expect that a more efficient version of `RNAbor` could be used in applications in genomics and synthetic biology, to detect potential conformational switches – RNA sequences containing two or more (distinct) metastable structures.

---

[4] Here Z denotes the partition function, defined as the sum of all Boltzmann factors $\exp(-E(S)/RT)$, over all secondary structures S of a given RNA sequence, R denotes the universal gas constant and T absolute temperature. Similarly Z_k denotes the sum of all Boltzmann factors of all structures S, whose base pair distance to the initial structure S^* is exactly k.

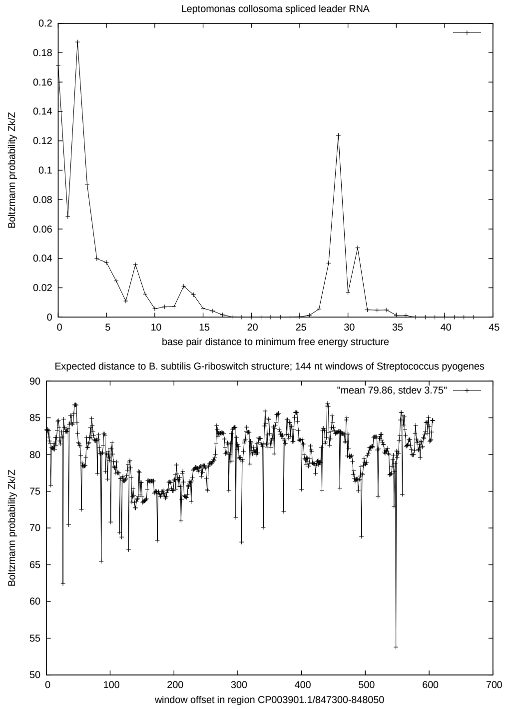

Fig. 1. *(Top)* Output of `FFTbor` on the 46 nt spliced leader conformational switch of *Leptomonas collosoma*, where reference structure S^* is taken to be the minimum free energy structure. *(Bottom)* Expected base pair distance $\sum_k k \cdot Z_k/Z$ from the reference structure of the guanine riboswitch of *Bacillus subtilis*, depicted in Figure 1A of [3]. `FFTbor` was run on all 144 nt windows of CP003901.1/847300-848050, comprising the 5′ untranslated region of the XPT gene (guanosine monophosphate reductase, with coding region at CP003901.1/848026-848607) of the unrelated organism *Streptococcus pyogenes A20*. `FFTbor` detects the guanine riboswitch at position 847848, where expected base pair distance to S^* achieves a minimum of 53.79, corresponding to a Z-score of This prediction corresponds well with the Rfam prediction at nearby position 847844.

In this abstract, we announce a radically different algorithm, `FFTbor`, that uses polynomial interpolation to compute the coefficients $p_0, \ldots, p_{n-1}$ of the polynomial $p(x) = p_0 + p_1 x + \cdots + p_{n-1} x^{n-1}$, where p_k is defined by $p_k = \frac{Z_k}{Z}$. Due to severe numerical instability issues in both the Lagrange interpolation formula and in Gaussian elimination, we employ the Fast Fourier Transform (FFT) to compute the inverse Discrete Fourier Transform (DFT) on values $y_0, \ldots, y_{n-1}$, where $y_k = p(\omega^k)$ and $\omega = e^{2\pi i/n}$ is the principal nth complex root of unity. This gives rise to an improved version of `RNAbor`, denoted `FFTbor`, which runs in time $O(n^4)$ and space $O(n^2)$ on a single processor, and in time $O(\frac{n^3}{m})$ on an m-core machine using `OpenMP`. Figure 1 (top) depicts the *rugged* energy landscape typical of a conformational switch, while Figure 1 (bottom) depicts expected base pair distance, of each size 144 window in the 5′-UTR of *S. pyogenes*, to the XPT riboswitch structure of *B. subtilis*.

References

1. P. Clote, F. Lou, and W.A. Lorenz. Maximum expected accuracy structural neighbors of an RNA secondary structure. *BMC Bioinformatics*, 13(Suppl 5):S6, April 2012.
2. E. Freyhult, V. Moulton, and P. Clote. Boltzmann probability of RNA structural neighbors and riboswitch detection. *Bioinformatics*, 23(16):2054–2062, August 2007.
3. A. Serganov, Y. R. Yuan, O. Pikovskaya, A. Polonskaia, L. Malinina, A. T. Phan, C. Hobartner, R. Micura, R. R. Breaker, and D. J. Patel. Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem. Biol.*, 11(12):1729–1741, December 2004.