The Energy Landscape Library
- a Platform for Generic Algorithms

Martin Mann and **Rolf Backofen**
Albert-Ludwigs-University Freiburg · Inst. of Computer Science · Chair for Bioinformatics
Georges-Köhler-Allee 106 · 79110 Freiburg · Germany
{mmann,backofen}@informatik.uni-freiburg.de
http://www.bioinf.uni-freiburg.de/sw/ell/

The Energy Landscape Library

Chair for Bioinformatics
University Freiburg

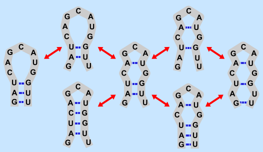
Introduction

The study of energy landscapes of biopolymers and their models is an important field in bioinformatics [1, 2]. For instance, the investigation of kinetics or folding simulations are done using methods that are based on sampling or exhaustive enumeration [3, 4]. Most of such algorithms are independent of the underlying landscape model. Therefore, frameworks for generic algorithms to investigate the landscape properties are needed.
Here, we present the **Energy Landscape Library** (ELL) that allows such a model-independent formulation of generic algorithms dealing with discrete states. The ELL is a completely object-oriented C++ library that is highly modular, easy to extend, and freely available online. It can be used for a fast and easy implementation of new generic algorithms (possibly based on the provided basic method pool) or as a framework to test their properties for different landscape models, which can be formulated straightforward.

Energy Landscapes

An **Energy Landscape model** consists of the following

- A **set of elements**
  e.g. all structures of an RNA-sequence
- An **energy for each** element
  e.g. the minimum free energy
- A **neighborhood definition**
  e.g. addition or removal of one bond



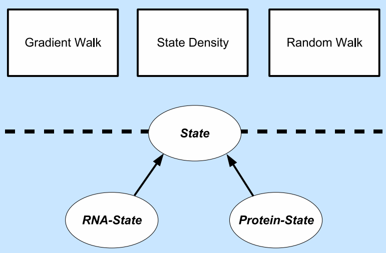
RNA structure neighborhood extract

Based on that several features can be defined like local minimum (all neighbors have higher energy), saddle points (minimal barrier height between two minima), etc.

To provide the independence between energy landscape model creation and the development of algorithms a clear and slim interface is necessary. This is realized by an abstract `State` class in the ELL.

The Core - an Abstract State

The core element of the ELL is an **abstract `State` class** representing an element of the energy landscape. All generic algorithms deal only with objects and available methods of this class to perform the computations. For a concrete landscape model, a subclass of `State` has to be derived combined with the implementation of the abstract methods. A visual sketch of the concept is given below.



The figure shows the partition of the algorithms (rectangles) and the concrete landscape models (elipses) that are all derived from the interfacing `State` class.

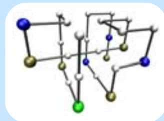
Features a state has to provide:

- Its energy/fitness
- Access to its neighbors in sequential or random order
- Conversion between compressed and decompressed version

Features of the ELL

The fundamental features of the ELL are

- Strictly object oriented C++ library following STL coding
- Highly modular by strict partition of algorithms and models
- Development of both layers independent
- Complete API documentation of all classes etc.
- Freely available



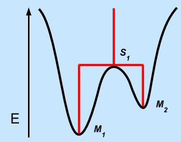
HPNX - lattice protein

Current energy landscape models

- RNA structures for a given sequence
- Lattice protein structures for various contact energies and lattices
- Compressed state representations to reduce memory usage for storing high numbers of states

Available algorithm pool

- Random, gradient, adaptive, ... walks
- Barrier tree generation [5] and sampling
- Density of states estimation [3]



Barrier tree (red)

Future Work

The ELL is currently under development and we are focusing our work on

- More and new energy landscape models from different fields of research
- Implementation of further basic and extended algorithms
- Improving the current ELL in terms of runtime and memory consumption
- ...

Conclusion

The ELL is a C++ programming platform for the study of energy landscapes with generic methods. The partition of the algorithms and the underlying landscape model, interfaced by a state abstraction, allows an independent development of both layers. The ELL supports the user with a pool of common methods and some landscape models (e.g. RNA) to provide a fast and efficient base for further investigations. It is highly modular, object-oriented and easy to extend. The ELL is available for free under

http://www.bioinf.uni-freiburg.de/sw/ell/

References

[1] Peter F. Stadler. Fitness landscapes. *Biological Evolution and Statistical Physics*, pages 187–207, 2002.

[2] Yu Xia and Michael Levitt. Simulating protein evolution in sequence and structure space. *Current Opinion in Structural Biology*, 14(2):202–7, 2004.

[3] F. Wang and D. P. Landau. Determining the density of states for classical statistical models: a random walk algorithm to produce a flat histogram. *Phys Rev E Stat Nonlin Soft Matter Phys*, 64, 2001.

[4] Michael T. Wolfinger, Sebastian Will, Ivo L. Hofacker, Rolf Backofen, and Peter F. Stadler. Exploring the lower part of discrete polymer model energy landscapes. *Europhysics Letters*, 74(4):725–732, 2006.

[5] Christoph Flamm, Ivo L. Hofacker, Peter F. Stadler, and Michael T. Wolfinger. Barrier trees of degenerate landscapes. *Z.Phys.Chem*, 216:155–173, 2002.

Acknowledgments

Martin Mann is supported by the EU project EMBIO (EC contract number 012835).