Introducing BioPhysConnectoR

Kay Hamacher[1], Franziska Hoffgaard, Phillip Weil
Technische Universität Darmstadt
AG Bioinformatics and Theoretical Biology, Fachbereich Biologie
Schnittspahnstr. 10, 64287 Darmstadt, Germany

The biggest challenge for systems biology and bioinformatics in the post-genome area is the integration of countless experimental data such as sequence information, gene expression data, physio-chemical values, phylogenetic relationships, or physiological data.

With R researchers command over an efficient framework for statistical modelling and accordingly R became – with the event of Bioconductor at the latest – the major platform for analyzing biostatistical data. Up to now much effort has been invested in the statistical modelling and subsequent implementation of *information-driven* packages, and protocols. This allowed tremendous progress in understanding *information* contained within e.g. biological sequence data. Experiments are nowadays guided to a large extend by the knowledge gained from such protocols.

The *information* contained within biological sequences reflects the whole evolutionary history of the organism under investigation (including external selective pressure such as drugs and resistance development). The selection step of every evolutionary process is, however, an event in the *physical realm* as selection tests the physiochemical properties of molecules involved in relevant processes.

Therefore, to construct molecular interaction networks [1], there is a pressing need to connect *information* (the evolutionary memory) with *the physical realm*, its forces, the molecular dynamics and mechanics (the selective „horizon“).

We achieve this with our ongoing efforts [2] in integrating standard sequence/statistical-model-driven methodologies with new reduced-molecular-models derived from biophysical interaction theories [3,4], eventually bridging the gap between bioinformatics and molecular dynamics simulations/molecular biophysics. We developed an R-package (BioPhysConnectoR) to this end. With this package we connect the information space and the physical space – thus allowing for functional annotation of sequence data and systematic *in silico* experiments. Additional useful functions for dealing with sequences and matrices are provided within the package.

We integrated C-code with R-routines and found that regarding the run-time efficiency our packages compares perfectly with our original code in C/FORTRAN. Due to the abstraction offered by R and leveraging the power of the packages Rmpi and papply, we were able to implement the package in a massively parallelized fashion.

As it is possible in R to interactively examine the results of the computations, this allows for both large-scale screening and high-throughput-scans on the one hand and online, interactive method development and hypothesis testing on the other. We discuss future research directions.

References

[1] K. Hamacher, J. Trylska, J.A. McCammon. *Dependency Map of Proteins in the Small Ribosomal Subunit.* **PLoS Comput. Biol.** 2(2): e10, 2006
[2] K. Hamacher. *Information Theoretical Measures to Analyze Trajectories in Rational Molecular Design*, **J.Comp.Chem.**, 28(16):2576-2580, 2007.
[3] K. Hamacher, J.A. McCammon. *Computing the amino acid specificity of fluctuations in biomolecular systems*, **J.Chem.Theo.Comp.** 2:873, 2006.
[4] K. Hamacher, *Relating Sequence Evolution of HIV1-Protease to Its Underlying Molecular Mechanics,* Gene, accepted, 2008

[1] *corresponding author, hamacher(AT)bio.tu-darmstadt.de, http://www.kay-hamacher.de*