Bioinformatics Tools as Power Hypothetical Predictors

Mohammad Najafi*

Biochemistry Department, Cellular and Molecular Research Center, Razi Drug Research Center, Tehran University of Medical Sciences, Tehran, Iran

With the development and completion of genome projects, the genome data not only are available on the search of primary databases, but also the structural and functional prediction powers are increased on the growth of secondary databases. Although, data retrieval tools suggested by the database updating centers are relatively simple for use of cellular and molecular investigators, but many prediction tools are based on understanding algorithms. This is the evidence, that the data management is related to mathematics and computer sciences in bioinformatics. There is a gap on the interest of biologists and biochemists to the use of prediction tools, since a little information is observed on designing algorithms. Computational biology is a scope more considered by investigators; however, the mathematics schemes are also used in their flowcharts.

It is obvious that the progressions in bioinformatics are rate-limited; it depends on the application of new matrices in algorithms and elevation of mathematical instruments. The basic question is that, "what should molecular and cellular investigators know about algorithms and their updated versions?"

On the excellent idea to know the details of bioinformatics algorithms, biologists and biochemists should spend the long and not-gracious educational terms on mathematical aspects of computer and information sciences. In addition, they should refresh your information on update of tools.

I thought these are the most important causes of investigators' unwilling to apply bioinformatics tools in their studies. My opinion is that a biologist or a biochemist should know only common descriptions and updates about algorithms designed for tools, in different branches of bioinformatics including sequence homology, genomics, transcriptomics and structural modeling.

Algorithm

Bioinformatics algorithms are step-by-step procedures for predicting an event. There are a set of mathematical and biological rules that support a prediction on calculating, processing and reasoning. The principles of each algorithm are commonly based on the two models; homology and *ab initio*. The models are updated on the development of mathematical schemes, whose details are not essential for a biologist to know. Although, the precision of updates can be considered on checking the outputs with quality control factors, but the link of tools within layers of a neural network point out that there are needs to know the descriptive characteristics of updated versions. The most important changes are primarily considered in the scoring matrices of structural and functional units (amino acid, base, rotamer), and then, the programs that try to relate each unit to its flanking sections.

Matrices

The primary scoring matrices are obtained on the molecular distance/similarity and conservation. Other soft matrices such as Position-Specific Scoring Matrix (PSSM) are also created and corrected on the primary data and score weights [1].

Distance/Similarity

DNA matrices were not complicated on the base substitutions; however, there were constitutional tendencies on the transition and transversion changes. PAM (Point Accepted Mutation) amino acid matrix [2], developed by Margaret Dayhoff was based on the evolutionary model, while BLOSUM matrix (BLOcks of Amino Acid SUbstitution Matrix), developed by Henikoff and Henikoff [3] was created on the BLOCK identity.

Sequence conservation

The scores to predict the secondary structures were created by Chou-Fasman, namely the Chou-Fasman method [4], and were developed by Gor method [5]. Furthermore, some scores were suggested on the backbone-dependent rotamers [6], and fold [7] recognitions for protein 3D prediction.

Development of Programs

The next generations of algorithms were developed on the Dynamic Programming and Hidden Markov Models (HMM). Furthermore, they were assimilated with other predicted documents on the homology methods [8,9].

Conclusion

In conclusion, I thought the biological hypotheses can be supported using bioinformatics tools, without the consideration of mathematical schemes, so that the prediction results may convert the two-way hypotheses into the one-way ones, and confirm the results obtained of experimental studies.

References

1. Lim SJ, Tong JC, Chew FT, Tammi MT (2008) The value of position-specific scoring matrices for assessment of protein allegenicity. BMC Bioinformatics 9: S21.

2. Mount DW (2008) Using PAM Matrices in Sequence Alignments. CSH Protoc 2008.

3. Henikoff S, Henikoff JG (1993) Performance evaluation of amino acid substitution matrices. Proteins 17: 49-61.

4. Chen H, Gu F, Huang Z (2006) Improved Chou-Fasman method for protein secondary structure prediction. BMC Bioinformatics 7: S14.

5. Sen TZ, Cheng H, Kloczkowski A, Jernigan RL (2006) A Consensus Data Mining secondary structure prediction by combining GOR V and Fragment Database Mining. Protein Sci 15: 2499-2506.

6. Dunbrack RL Jr, Karplus M (1994) Conformational analysis of the backbone-

*Corresponding author: Mohammad Najafi, Biochemistry Department, Cellular and Molecular Research Center, Razi Drug Research Center, Tehran University of Medical Sciences, Tehran, Iran, E-mail: nbsmmsbn@tums.ac.ir

Citation: Najafi M (2012) Bioinformatics Tools as Power Hypothetical Predictors. Biochem Physiol 1:e113. oi:10.4172/2168-9652.1000e113

Page 2 of 2

dependent rotamer preferences of protein sidechains. Nat Struct Biol 1: 334-340.

7. Shen HB, Chou KC (2009) Predicting protein fold pattern with functional domain and sequential evolution information. J Theor Biol 256: 441-446.

8. Giegerich R (2000) A systematic approach to dynamic programming in bioinformatics. Bioinformatics 16: 665-677.

9. Beerenwinkel N, Siebourg J (2012) Probability, statistics, and computational science. Methods Mol Biol 855: 77-110.