Available online at www.jobpcr.com

ISSN-2394:3076
CODEN(USA) : JBPCBK

Journal of Biological Pharmaceutical And Chemical Research , 2015,2(4):1-4

(http://www.jobpcr.com/arhcive.php)

BIOINFORMATICS: ENRICHING THE MATERIAL SCIENCE

M. K. Sharma[1], R. N. Yadav[1], S. K. Chakrabarti[2]
[1]Department of Mathematics and [2]Department of Physics
Tribhuvan University, MMAM Campus, Biratnagar, Nepal

ABSTRACT
Material science is an emerging field of science. It is an interdisciplinary science comprising of physics, chemistry, biology, earth science, computer science and technology as well as materials engineering. Due to the advancement of knowledge now-a-days each of these branches of science has different sub-branches. There are several interrelations among these branches and sub-branches. Bioinformatics is such an emerging interdisciplinary science. In this paper we have discussed that day by day as the material science is gradually being enriched, bioinformatics has also an important role in it.

Key words: Biophysics, Biochemistry, Computer science, DNA, Protein.

INTRODUCTION

It is undeniable that in material science biology played a vital role in the twentieth century. That role is likely to acquire further importance in the years to come [1]. In the wake of the work of Watson and Crick and the sequencing of human genome far reaching discoveries are constantly happening [2]. One major factor promoting the importance of biology is its relationship with medicine. Fundamental progress in medicine depends upon elucidating some of the mysteries that occur in biological sciences. Biology depended upon chemistry to make major strides and this led to the development of biochemistry. Similarly, the need to explain biological phenomena at the atomic level led to biophysics. The enormous amount of data gathered by biologists and the need to interpret them requires tools which are in the realm of computer science. Thus through bioinformatics both chemistry and physics have been benefited from the symbiotic work done by the biologists. The collaborative work functions as a source of inspiration for novel pursuits in the material science.

MATERIALS AND METHOD

EMERGING OF BIOINFORMATICS

A common problem with the maturation of an interdisciplinary subject is that, inevitably, the forerunner disciplines call for differing perspectives. It is observed that differences in working with

the biological scientists gave rise to an area called computational biology. Computational biologists take justified pride in the formal aspects of their work. Those often involve proofs of algorithmic correctness, complexity estimates and other themes which are central to theoretical computer science. Nevertheless, the biologists' needs are so pressing and broad that many other aspects related to computer science have to be explored. For example, biologists need software that is reliable and can deal with huge amount of data as well as interfaces which facilitate the human machine interaction.

This computational biology has now emerged as what is called the bioinformatics. A distinctive aspect of bioinformatics is its widespread use of the web. The immense databases containing DNA sequences and 3D protein structures are available to almost any researcher. Furthermore, the community interested in bioinformatics has developed a myriad of application programmes accessible through the internet. Some of these programmes, e.g. BLAST, have taken years for development and finely tuned. The vast number of daily visits to some of the NIH sites containing genomic databases is comparable to that of widely used search engines or active software downloading sites [3]. This explains the great interest that bioinformaticians have in script languages such as Perl, Python, VBScript, JScript that allow the automatic examination and gathering of information from websites [4].

In computer science we favour generality and abstractions. Our approach is often top-down as if we were developing a programme or writing a manual. In contrast, biologists often favour a bottom-up approach. This is understandable because the minutiae are so important and biologists are often involved with time consuming experiments that may yield ambiguous results which in turn have to be resolved by further tests. The work of synthesis eventually has to take place but since in biology most of the rules have exceptions, biologists are wary of generalisations.

In the past decades biologists have gathered information about the cell characteristics of many species. That information can be extrapolated to other species with the help of evolutionary principle. Since understanding the human cell is a primary concern in medicine, one usually wishes to infer human cell behaviour from that of other species. However, most available data are fragmented, incomplete and noisy. So, if one had to characterise bioinformatics in logical terms, it would be: reasoning with incomplete information. That includes providing ancillary tools allowing researchers to compare carefully the relationship between new data and data that have been validated by experiments.

ANALOGY WITH COMPUTER SCIENCE

In the universal Turing Machine (TM) model of computing one does not distinguish between programme and data. They co-exist in the tape of the machine and it is the TM interpreter that is commanded to start computations at a given state examining a given element of the tape [5].

Let us introduce the notion of interpretation in our simplified description of a single biological cell. Both DNA and proteins are components of our model but the interactions that take place between DNA and other components, i.e. existing proteins, result in producing new proteins each of which has a specific function needed for cell survival—growth, metabolism, replication and others [6].

Let a gene *G* in the DNA component be responsible for producing a protein *P*. Interpreter *I* capable of processing any gene may well utilise *P* as one of its components. This implies that if *P* has not been assembled into the machinery of *I*, no interpretation takes place.

Another instance in which *P* cannot be produced is due to the fact that another protein *P′* may position itself at the beginning of gene *G* and temporarily prevent the transcription.

The interpreter in the biological case is either one that already exists in a given cell (prior to cell replication) or else it can be assembled from proteins and RNA generated by specific genes e.g. ribosomal genes. In biology the interpreter can be viewed as a mechanical gadget that is made of

moving parts which produce new components based on given templates (DNA or RNA). The construction of new components is made by subcomponents that happen to be in the vicinity. If they are not, interpretation cannot proceed.

One can imagine a similar situation when interpreting computer programmes. Assume that the components of *I* are first generated on the fly and once *I* is assembled as data, control is transferred to the execution of *I* as a programme.

RELEVANCE WITH OTHERS

It is undeniable that probability and statistics play an influential role in bioinformatics. This is not surprising since the data available are huge, varied and noisy. Recent articles on interpreting micro-array experiments utilise statistical approaches such as SVM and Bayesian networks [7].

Hidden Markov Model (HMM) is also machine learning technique. In this approach one starts by specifying a topology of finite states representing the structure one believes as applicable. Based on the learning set the probabilities are computed. Given a new sequence, then we can use DP (the Viterbi algorithm) to determine the most likely succession of states corresponding to the said sequence. The method amounts to the generation of probabilistic grammar from a learning set. The topology of states in HMM is generalised to correspond to the presumed grammar rules whose frequency one wishes to estimate. Therefore, the method using probabilistic grammar is expected to have a salient place in bioinformatics.

Computational geometry also plays a key role in analyzing 3D structures. An example is 3D pattern matching in proteins. In this case the 'pattern' is a portion of the backbone of a protein and the 'text' corresponds to all the proteins in the PDB. One would want to determine the set of proteins which exhibit that pattern. As in the case of alignments we would like to tolerate small discrepancies between the pattern and elements in the text.

The example of phylogenetic tree construction using data compression illustrates the importance of information technology in analysing massively long sequence of symbols.

Graphics and graphical interfaces are of course a necessity for displaying biological data. As in other CS applications the knowledge of biology and the capacity to interact with biologists are vital to successful software development in bioinformatics.

CONCLUSION

Searls rightly pointed out that many current problems remain as challenging tasks. His list includes: protein structure prediction, homology search, multiple alignment and phylogeny construction, genomic sequence analysis and gene finding. The most recent developments in biology point in the direction of functional genomics research.

The accomplishments made in molecular biology in the past half century have been remarkable. Nevertheless, they pale in comparison to the wondrous tasks that lie ahead. We are still quite ignorant to answer to the questions like: (1) How do brain cells establish linkage among themselves while an embryo is being formed? (2) Is it possible to understand better the origin of language and the nature-nurture paradigm? (3) How does Darwinian evolutionary theory operate at the molecular level? For answering to these types of question the scientists will have to go through vigorous research on the concerned sectors of physics, chemistry, biology, earth science and computer science—all of which have collectively given rise to the emerging field of material science. Thus bioinformatics is also enriching the material science.

REFERENCES

[1]. P Hogeweg; DB Searls. PLS Comp. Biol., **2011,** 7, 2021.
[2]. RD Fleischmann; MD Adams; O White; RA Clayton; EF Kirkness; AR Kerlavage; CJ Bult; JF Tomb; BA Dougherty; JM Merrick. Science, **1995,** 269, 496.
[3]. N Cristianini; M Hahn. Introduction to Computational Genomics, Cambridge University Press, Cambridge, **2006.**
[4]. D Gilbert. Brief. Bioinform., **2004,** 5, 300.
[5]. P Baldi; S Brunak. Bioinformatics: The Machine Learning Approach, MIT Press, Massachusetts, **2001.**
[6]. AD Baxevanis; BFF Ouellette. Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins, Wiley, New Jersey, **2005.**
[7]. L Pachter; B Sturmfels. Algebraic Statistics for Computational Biology, Cambridge University Press, Cambridge, **2005.**