Role of Bioinformatics in Biotechnology

M.K.Jhala, C.G.Joshi, T.J.Purohit, N.P.Patel and J.G.Sarvaiya

Information Technology Centre, GAU, Anand.

INTRODUCTION

The term 'bioinformatics' is the short form of 'biological informatics', just as biotechnology is the short form of 'biological technology'. Anthony Kerlavage, of the Celera Genomics, defined bioinformatics as 'Any application of computation to the field of biology, including data management, algorithm development, and data mining'. Clearly, a number of divergent areas, many of them outside biotechnology, come under bioinformatics.

Bioinformatics is the field of science in which biology, computer science, and information technology merge to form a single discipline. The ultimate goal of the field is to enable the discovery of new biological insights as well as to create a global perspective from which unifying principles in biology can be discerned.

Initial interest in Bioinformatics was propelled by the necessity to create databases of biological sequences. The first database was created within a short period after the Insulin protein sequence was made available in 1956.

The sequence information generated by the human genome research, initiated in 1988 has now been stored as a primary information source for future applications in medicine. The available data is so huge that if compiled in books, the data would run into 200 volumes of 1000 pages each and reading alone (ignoring understanding factor) would require 26 years working around the clock. For the population of about 5 billion human beings with two individuals differing in three million bases, the genomic sequence difference database would have about 15,000,000 billion entries. The present challenge to handle such a huge volume of data is to improve database design, develop software for database access and manipulation, and device data-entry procedures to compensate for the varied computer procedures and systems used in different laboratories. A single experiment can now yield data on the transcription level of 100,000 different mRNA species from a given tissue (Winzeler et al., 1998).

APPLICATIONS OF BIOINFORMATICS:

GENOMICS:

Estimating the number of genes in an organism basing on the number of nucleotide base pairs was not reliable, due to the presence of high numbers of redundant copies of many genes. Genomics has corrected this situation. Useful genes can be selected from a gene library thus constructed and inserted into other organisms for improvement or harmful genes can be silenced. In the areas of structural genomics, functional genomics and nutritional genomics, bioinformatics plays a vital role.

PROTEOMICS:

Proteomics involves the sequencing of amino acids in a protein, determining its three-dimensional structure and relating it to the function of the protein. Before computer processing comes into the picture, extensive data, particularly through crystallography and NMR, are required for this kind of a study. With such data on known proteins, the structure and its relationship to function of newly discovered proteins can be understood in a very short time. In such areas, bioinformatics has an enormous analytical and predictive potential. Protein folding alone of the most significant and fundamental problem in biological science realizing this, IBM in Dec 1999, had built a supercomputer, which is 2 million times faster than the today's fastest desktop PC. This new computer nicknamed "Blue Gene" by IBM researchers will be capable of performing more than one quadrillion operations per second. Better understanding of how proteins fold will give scientists and doctors better insight into diseases and ways to combat them.

CHEMINFORMATICS AND DRUG DESIGN:

It is now possible, through computer algorithm based bioinformatic procedures, to identify and structurally modify a natural product, to design a drug with the desired properties and to assess its therapeutic effects, theoretically. Such procedures, similar to an architect's on board plan before construction, are described as *in silico* (in the computer, based on silicon chip technology)*,* as opposed to the earlier *in vitro* (in experimental models) and *in vivo* (in clinical trials*)* methods. The risk involved in the earlier random processes of drug discovery methods is largely removed by bioinformatics.

Cheminformatics involves organisation of chemical data in a logical form to facilitate the process of understanding chemical properties, their relationship to structures and making inferences.

MOLECULAR PHYLOGENIES:

Phylogeny is the origin and evolution of organisms. Biologists have constructed very elegant systems of classifications for the known organisms, though problems persist. Extensive work was carried out this way, comparing a very large number of organisms of plants and animals. Amino acid sequences and characteristics of proteins are also used in systematics.

DRUG MODIFICATION:

Several synthetic products are quite useful but cannot be used by one and all for certain side effects in some people. For example, aspartame (marketed under different trade names) is a dipeptide of aspartic acid and phenylalanine, and is 300 times sweeter than cane sugar. Aspartame is widely used as an alternate sweetener by diabetics and others who cannot take sweeteners loaded with calories. Unfortunately, pregnant women and people suffering from phenylketonuria, a disorder due to an impaired metabolism of phenylalanine, should not use aspartame. It would be useful if phenylalanine were substituted by some other amino acid without affecting its sweetness, to remove the restriction on its use.

PARTNERSHIP IN BIOINFORMATICS

a) **Data Gatherers:** Enormous amounts of basic data from biomolecular chemistry and related areas, very painstakingly gathered over long years by experimental and analytical scientists, are the body and substance of bioinformatics.

b) **Data Processors:** The second party use skills of complex software, to serve the needs of the data gathered.

c) **Process Product Users:** End users of products.

CONCLUSION

Today it is possible to perform (using heuristic algorithms) 80% accurate searches perhaps 90 - 95% accuracy from the leading software systems. Sensitive algorithms which improve the search accuracy, such as hidden Markov models and Smith-Waterman algorithm, are also available but take more time to execute the search. Now, to handle these demanding needs, computers are being designed around the biologists.

Leading bioinformatics companies are developing software systems which permit research scientists to integrate their diverse data and tools under Common Graphical User Interfaces (GUIs). It also permits scientists to share information and provides powerful solution to archive data. The whole area of biology can immensely benefit from the bioinformatic approach. Bioinformatics tools for efficient research will have significant implications in life sciences and betterment of human lives.

REFERENCES

Anthony Kerlavage, Vivien Bonazzi, Matteo di Tommaso, Charles Lawrence, Peter Li, Frank Mayberry, Richard Mural, Marc Nodell, Mark Yandell, Jinghui Zhang and Paul Thomas (2002). The Celera Discovery System. *Nucleic Acids Res.*, **30**(1): 129-136.

Barioch A. and Apweiler R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acid Res.,* **28**: 45-48.

Chris Sander (2003). Growth in Bioinformatics. *Bioinformatics* **19**(1): 1-1.

Dicks-J; Anderson-M; Cardle-L; Cartinhour-S; Couchman-M; Davenport-G; Dickson-J; Gale-M; Marshall-D; May-S; McWilliam-H; O'-Malia-A; Ougham-H; Trick-M; Walsh-S; Waugh-R. (2000). UK CropNet: a collection of databases and bioinformatics resources for crop plant genomics. *Nucleic Acid Res.,* **28** (1): 104-107.

http://www.fbae.org/Channels/bioinformatics/BIOINFORMATICS IN.htm

http://www.mrc-lmb.cam.ac.uk/genomes/madanm/articles/bioinfo.htm

Tramontano, Damien Devos, Christian Blaschke, Alfonso Valencia, David Brett, David Martin, Christophe Leroy, Isidore Rigoutsos, Chris Sander, Christos A. Ouzounis (2003). Evaluation of annotation strategies using an entire genome sequence. *Bioinformatics* **19**(6): 717-726.

Winzeler, E.A., Richards,D.R., Conway,A.R., Goldstein,A.L., Kalman,S., McCullough,M.J, Mc-Cusker,J.H., Stevens,D.A., Wodicka,L., Lockhart,D.J. and Davis,R.W (1998). Direct Allelic variation scanning of the Yeast Genome. *Science*, **281**: 1194-1197.