Computational Biology & Bioinformatics: A Gentle Overview

Article · January 2007

1 author:

Achuthsankar S Nair
University of Kerala
162 PUBLICATIONS   **688** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Compressing genomic data View project

Nutrimetabolomics View project

Computational Biology & Bioinformatics: A Gentle Overview

Achuthsankar S Nair

Extracts from my Guest Editorial of Communications of Computer Society of India, Jan 2007.

*Bio*informatics ? Biology and Computers ? What do they have to do with each other? I suppose that this question could have been raised even in 19th century when technologies of computers and biology were just emerging. At one city in France the great Louis Pasteur (1822–1895) was studying how fermentation of alcohol was linked to the existence of a specific microorganism. In another city in England, equally great Charles Babbage (1791–1871) was oiling his Analytical Engine in which Ada Lovelace, a mathematician who understood Babbage's vision, was trying to calculate the Bernoulli numbers. These gentlemen are today hailed as father of biotechnology and father of computers respectively. Did Pasteur and Babbage ever meet ? They had about 25 years to do so, and were less than 1000 Km apart. We do not know if they ever met, but had they met, they possibly would not have talked to each other ! If I may be pardoned for a politically incorrect pun, remember that Pasteur was French and Babbage was British !. Anyway, what do they have in common to talk, other than the weather? What is there in common between the gear wheels that were turning away in an attempt to crunch numbers and the microbes playing mysterious role in fermenting alcohol ?

Is this true today ? *Not a bit, not even as much as a bacteria*. It seems imminent, if not already true, that Biology and Computers are becoming close cousins which are mutually respecting, helping and influencing each other and synergistically merging, more than ever. The flood of data from Biology, mainly in the form of DNA, RNA and Protein sequences, is putting heavy demand on computers and computational scientists.   At the same time, it is demanding a transformation of basic ethos of biological sciences.   A common misconception is that bio-informatics is about creating and managing bio-data bases. Nothing would be farther from the truth. Fine analytical and engineering skills are in great demand in the area, as seen by vigorous attempts of machine-learning on the protein folding and gene-finding problems. The great Donald Kunth, renowned Stanford computer science professor, is quoted often for pointing out that biology has 500 years of exciting problems to work on. He feels that biology is "so digital, and incredibly complicated, but incredibly useful"(Computer Literacy Interview with Donald Knuth by Dan Doernberg, December 1993). However, there are still some spokes in the wheel for the grand union between two great sciences and their offshoot technologies. Due to the estrangement which existed for many decades, professionals from both the fields have a lot to do in terms of fine tuning their communication. Skepticism from puritans in both fields towards the claim of Bioinformatics as an independent field also needs convincing answers.

Many universities world over have started teaching and research in the area. Journals are plenty and so are conferences and professional meetings. As the disciplines of bioinformatics and computational biology are gaining prominence day by day, an industry is also emerging fast on their shoulders, estimated at $1.82 billion in 2007. Bioinformatics has taken on a new glitter by entering the field of drug discovery in a big way. Bioinformatics has taken on a new glitter by entering the field of drug discovery in a big way. This is one area that seems to be becoming the single largest. bioinformatics application, from an Industry view point. In India, it has a special relevance in the context of the recent patent amendment that has brought in product patents.

There has been a green-shift in all prominent technology publications. IEEE has prominently adopted such a shift. I did a quick check. If you use the key word "biology" and search the IEEE Digital Library limiting the year of search, you get the following hits for the years indicated in brackets: 13 (1975), 40(1985), 3484 (1990), 9617 (1995), 16233 (2000) and 27526 (2006). I did this on 26 November 2006, among the 14,32,467 documents in the data base. About 2% documents have been greened! One of the latest additions to the prestigious IEEE Transactions series is IEEE & ACM Transactions on Computational Biology and Bioinformatics. It may be noted that biological motivation has a long history in the computer field, in the form of artificial neural networks, genetic algorithms, to the recent ant-colony optimization techniques. Applications of computers in biology were mostly in the bio-medical field, in early days. One new facet that has emerged with Bioinformatics, is the focus on sub-cellular and molecular levels of Biology. Systems biology promises great growth in modeling cellular life, using conventional engineering approach, as already pointed to by projects such as e-Cell.

1. Introduction.

I will attempt to give the *big picture* of Computational Biology and Bioinformatics by presenting basic ideas in minimal technical vocabulary, aimed specifically at IT community. I do not have anything against life scientists attempting to read this and I think it could be useful in patches to them also. They are however likely to be uncomfortable with my bio-wisdom.

2. What is Bioinformatics/Computational Biology ?

Computational Biology/Bioinformatics is the application of computer sciences and allied technologies to answer the questions of Biologists, about the mysteries of life. A mere application of computers to solve any problem of a biologist would not merit a separate discipline. It looks as if Computational Biology and Bioinformatics are mainly concerned with problems involving data emerging from within cells of living beings. It might be appropriate to say that Computational Biology and Bioinformatics deal with application of computers in solving problems of molecular biology, in this context.

What are these data emerging from a cell ?. Though not exhaustive, at the risk of oversimplifying I will list 4 important data: DNA, RNA and Protein sequences and Micro array images. Surprisingly, first 3 of them are mere text data (strings, more formally) that can be opened with a text editor. The last one is a digital image. See Fig 1. We can now list some computer applications as Computational Biology/Bioinformatics and some as not:

- Analysing DNA sequence data to locate genes √
- Analysing RNA sequence data to predict their structure√
- Analysing protein sequence data to predict their location inside cell √
- Developing medicinal plant data base ×
- Analysing gene expression images √
- Using computers to identify finger prints ×
- Using computers in process control in bio-technology industries ×
- Identifying new Drug Molecules √
- Using computers to analyse ECG signals ×

Is DNA Computing & Bioinformatics related ? No, they are not. While bioinformatics deals with analysis of information represented by DNA, DNA computing is about creating bio-computers using DNA and enzymes (a class of proteins) to do mathematical calculations. The field got fame due to experiments which where done by Adleman in early 90s. He succeeded in solving the traveling sales-man problem by making strands of DNA to represent each city and the path between cities. Mixing many copies of each strand in a test tube, he went on to produce the correct answer as a strand left in the test tube. This is obviously a whole lot of biology than informatics. **Is Bioinformatics & Biometry related ?** Again, no. Biometry is all about uniquely recognizing humans based upon intrinsic physical traits such as finger prints, eye retinas and irises, facial patterns and hand geometries. However, let us note that a DNA of a person could be the best such unique trait for identifying people.

(a) DNA Data (4 letter strings)

(b) RNA Data (4 letter strings)

(c) Protein Data (20 letter strings)

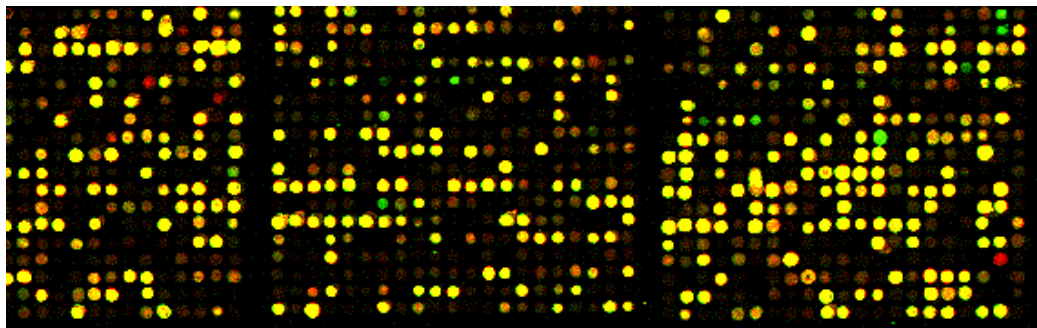(d) Micro Array Image Data (traditional Digital Images)

Fig. 1: Four kinds of data required by analyzed in Bioinformatics/Computational Biology

What is difference between Bioinformatics and Computational Biology ? This is a bit tricky. Both are "Computers + Biology". Difference is subtle but important. Bioinformatics = Biology + Computers whereas Computational Biology = Computers + Biology. In other words, biologists who specialize in use of computational tools and systems to answer problems of biology are bioinformaticians. Computer scientists, Mathematicians, Statisticians, and Engineers who specialize in developing theories, algorithms and techniques for such tools and systems are computational biologists. Arguably, there will be overlaps, but one can also identify some clear demarcations. I am yet to find a biologist who is at absolute ease in understanding, let alone developing a hidden Markov model, which is a machine learning paradigm used extensively in Bioinformatics.

3. A 5-minute primer on Biology

Biology looks at the wonderful and complex phenomena of life at many levels (organisms, organs, cells etc). Our interest is at the level of cells. This would approximately correspond to Molecular Biology or Cell Biology. At this level, the following is the minimum essential vocabulary list:

- Eukaryotic, Prokaryotic
- Cell
- Nucleus, Chromosomes, DNA, DNA bases A, G, C, and T.
- Genome, Gene
- RNA
- Proteins, Amino acids

I am now giving a very simplified explanation of these terms. If you are a biologist you are likely to hate me for trivializing things !:

3.1 Eukaryotic, Prokaryotic

Eukaryote is a developed organism like a human being or a tree. Prokaryotes are lower forms of life like bacteria. The problems of analyzing their information are also different. If you are a beginner, you might mix up these words. The *Pro* of Prokaryotes rhymes somewhat with *Pradhama,* meaning first is *Sanskrit*, Remember, bacteria existed before human beings appeared on the face of earth, they are *pradhama* organisms.

Prokaryote   Eukaryote !   Eukaryote !
Fig 2. Examples of prokaryotes and eukaryotes

3.2 Cell

If you scratch the skin on your hand right now, thousands of cells would fall down. Every living organism is made of cells, though some are just made up of a single cell too. Cells are most complex, wonderful and mysterious machines which are always a buzzing with activity. There are many complex things to know about a cell, but in our simplified view, 3 things are key: DNA, RNA and Proteins.
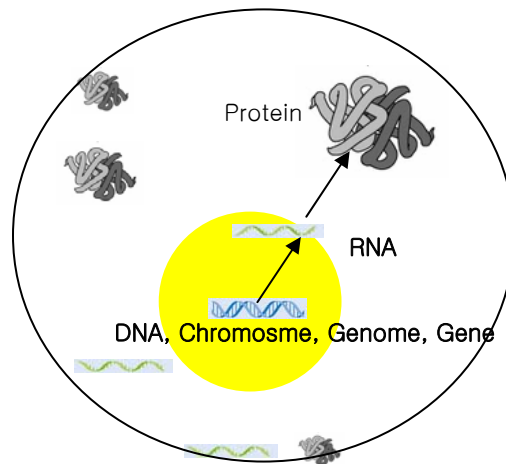
Fig 3. A schematic of the Cell

3.3 Nucleus, Chromosomes, DNA, A, G, C, T

Cells have a central core called nucleus, which is storehouse of an important molecule known as **DNA** (we work without full-forms, it is not my business). They are packaged in units known as **chromosomes**. DNA is a chain of 4 types of molecules, A, G, C and T. They are *double stranded* molecules as shown in figure 4, but informationally, we read the DNA from one strand alone, as the other side can be predicted. A G C and T always hook up in a predictable manner on the left and right strands: A always links with T, and C with G. If one drop of your blood is made available, advanced Biotech laboratories are able to isolate a cell, "cut" open the nucleus, "pull out" the genome, "read" it using machinery known as sequencing machinery and finally give you, in about 5 CDs, text files totaling an uncompressed size of 3200, 000, 000 bytes (3.2 GB). These files could be opened by any text editor and would look equally uninteresting on any of them, running into long and seemingly nonsensical sequences of A, G, C and T: **TCCTGAT AAGTCAG TGTCTCCT GAGTCTA GCTTCTG TCCATGC TGATCAT GTCCATG TTCTAGT CATGATA GTTGATTC TAGTGTCC TGATTAG CCTTGA ATCTTCT AGTTCT GTCCAT TATCCAT.** But it is the complete blue print of your life, including indication of what diseases you are susceptible to and may be even predict your infidelity. More interestingly, it also has the whole history book of evolution of life on earth, if only we could read (Are you looking back at the cells you scratched off ?). Every cell of your body has this information and cells are simply great in copying them with astonishingly small error rates into newer cells, when they divide.
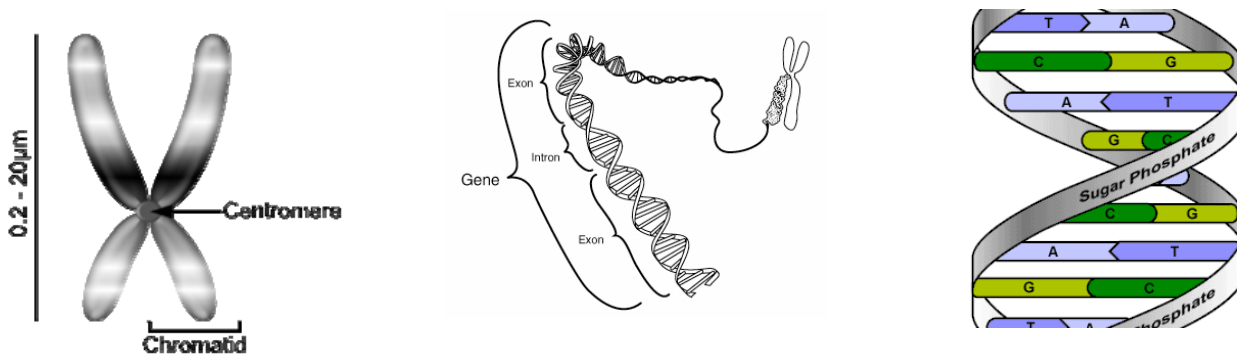

Fig. 4. The Chromosome & DNA

3.4 Genomes, Genes

Recall that **DNA** is packaged into units known as **chromosomes.** Humans have 23 pairs of it. They are together known as the **genome,** and today is known to be the blue-print of life. (The word –

ome, of late, is very popular in biology. If a modern biologist describes the collection of all students studying in various universities in India, they would call it the *studentome*. They may raise their voice against the *corruptome* that is prevalent in the Government ! Beware of meeting more – *omes* soon). **Genes** are specific regions of the genomes (about 1%) spread throughout the genome, sometimes contiguous, many times non-contiguous. The study of the genome is known as genomics. When this word is used by life scientists, it encompasses bio/chemical studies also. IT personnel possibly confine to 'computational genomics', the computational part of the study. A word about the human genome which was completely sequenced in 2003: Only 0.2% of human genome differs between individuals. Black or white, Hindu or Muslim, we are all 99.8% the same.

3.5 RNA

RNAs are similar to DNA informationally, their major purpose is to copy information from DNA selectively and to bring it out of the nucleus to use it where it is designated to be. However there are other varieties of RNA which do different sort of things. RNA contains, like the DNA, 4 kinds of molecules – A G C and U, the last one replacing the T in DNA. An RNA sequence may run like this: **UCCUGAU AAGUCAG UGUCUCCU GAGUCUA GCUUCUG UCCAUGC UGAUCAU GUCCAUG UUCUAGU CAUGAUA GUUGAUUC UAGUGUCC UGAUUAG CCUUGA AUCUUCU AGUUCU GUCCAU UAUCCAU.** There are different kinds of RNA and biologists have lot of questions to ask about RNAs after they give you a text file of their sequence. The RNA is single stranded unlike the DNA and can also assume certain unique shapes.

3.6 Proteins and Amino Acids

People who are far removed from Biology have this "healthy" notion that proteins are something good to eat: milk, egg, yoghurt, meat, fish, beans, lentils, peas, peanuts … From this very moment, let us go beyond that innocent notion. Proteins are the most important molecules in life. In a way, you can say your body is just a protein factory, capable of producing 100,000 vivid proteins. When they are produced in the right time, at the right place, in the right quantity, we are healthy. To shake off the conventional notion about proteins, let me tell you that silk sarees are made out of a protein produced by silk worms, spider webs are proteins (which are five times stronger, on a weight-to-strength basis, than steel) produced by spiders. And snake venom, is a concoction of proteins. And now, tell me, would you add any of this to your healthy food list ?!. Let me add on, that our hair and nails are made with help of a protein known as Keratin. Proteins are large (macro) molecules continuously manufactured by the cells and the instructions to produce them are stored in the DNA.
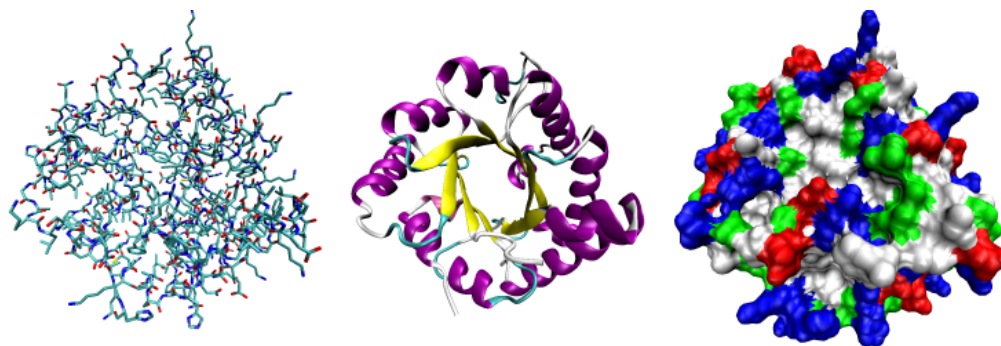

Fig 5. Three representations of the protein *triose phosphate isomerase*

Proteins are made of amino acids, which are twenty in count (researchers are debating on increasing this count, as couple of new ones are claimed to be identified). The amino acid list starts like this: Alanine, Arginine, Asparagine, Aspartic Acid, Cysteine, Glutamic Acid, Glutamine ⋯ Happily, they have single letter codes – A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y. Easier way to remember them is to note that they have all English letters except B, J, O, U, X and Z. Adult humans can produce within their bodies, 12 amino acids. The other eight have to be eaten through protein rich food, and these proteins will be chopped back to amino acids by the liver, so that cells can use them to build the proteins that body requires. (My student Amjesh asks this question: In this case would not eating human flesh be a good idea, so that the amino acids will be fully recyclable? No, he is not a cannibal, as far as I know).

A protein sequence will look like:CFPUQEGHILDCLKSTFEWCUWECFPWRDTCEDUSTTW EGHILDNDTEGHTWUWWESPUSTPPUQWRDCCLKSWCUWMFCQEDTWRWEGHILKMFPUSTWYZEGN DTWRDCFPUQEGHILDCLKSTMFEWCUWESTHCFPWRDT. Protein sequences are shorter than most DNA sequences and are mostly in 100s of characters, whereas DNA sequences easily run to 10000s of characters.

Proteins are not linear chains of amino acids. They are famous for their shapes. They turn, twist, and fold into very unique shapes. These shapes determine what they do. These shapes are studied at 4 levels – primary, secondary, tertiary and quaternary. One big question that biologists want computational biologists to answer runs like this – "given this protein sequence (say, in a 500KB text file), tell me the exact structure that this protein will fold into, by specifying the coordinate of every atom in it". This is considered the biggest open problem in science. Machine learning approaches have reached slightly above 75% accuracy in answering this problem.

The entire ensemble of proteins in an organism of interest, is known, not surprisingly, as the **proteome** and the field of its study, as proteomics.

3.7. The "Central Dogma of Molecular Biology"
The gene regions of the DNA in the nucleus of the cell is copied (*transcribed*) into the RNA and RNA travels to protein production sites and is *translated* into proteins. In short, **DNA→ RNA → Proteins**, is the Central Dogma of Molecular Biology. Imagine, there are trillions of cells in your body, the DNA of each of them is churning out thousands of RNAs which in turn cause thousands of proteins to be produced, every moment. One of them is making your hair strong, another giving the glitter in your eyes, another one carrying oxygen to different parts, and yet another one helping in the making of proteins themselves ! No wonder that famous life scientist Russel Doolittle exclaimed: **"We are our proteins"**

4. On some of the branches of Bioinformatics
Arguably, following could be some of the major branches of Bioinformatics: Genomics, Proteomics, (in strict sense, should be used with the prefix *Computational*), Computer–Aided Drug Design, Bio Data Bases & Data Mining, Molecular Phylogenetics, Microarray Informatics and Systems Biology. We will briefly touch upon their scope in the ensuing paragraphs.

Genomics & Proteomics are both big fields, encompassing various studies of the genome and the proteome. Computationally, both start with sequence data, and attempt to answer questions like

this: **Genomics:** Given a DNA sequence, where are the genes ? (Gene Finding); How similar is the given sequence with another one ? (Pair-wise Sequence Alignment); How similar are a set of given sequences ? (Multiple Sequence Alignment); Where on this sequence does another given bio-molecule bind ? (Transcription factor binding site identification); How can we compress this sequence ? How can we visualize this sequence insightfully ? (genome browsing); **Proteomics:** Given a protein sequence data, how similar it is with another one, or how similar are a set of protein sequences (pair-wise and multiple sequence alignment); What is the primary, secondary or tertiary structure of the molecule ? (the great protein folding problem); Which part is most chemically active ? (Active site determination problem); How would it interact with another protein ? (protein-protein interaction problem); To which cell compartment is this protein belonging to ? (protein sub-cellular localization or protein sorting problem).

The technique of **sequence alignment** which is widely applied in both genomics and proteomics, deserves a special mention. It is all about writing two bio-sequences (DNA/RNA/Protein), one below the other, to highlight their similarity to the maximum extent possible. You can do this on English strings also. Consider the strings "Gates like cheese" and "Grated cheese". If you write one below the other and compare letter for letter, you find only 2 letters matching, indicated by |.

```
Gates likes cheese          G-ates likes cheese
|         |                 | |||        ||||||
Grated cheese               Grated ------cheese
```

As soon as you stretch the sequences to highlight similarity by inserting gaps, we find it more truthfully highlights similarity with 10 matches. Consider doing this on DNA sequences millions of letters long ! **BLAST** is a software which can do this using dynamic programming, as fast as Google searches for your keywords, considering the length of query words of bio-sequences. In addition it uses very sophisticated scoring mechanisms (**PAM, BLOSUM** scoring matrices) to overlook 'pardonable' mismatches of characters, like that of 's' and 'z' in English. When this is done on more than two sequences at a time, we have a hard nut to crack. Software such as ClustalX does this, sub-optimally as in Fig. 6.
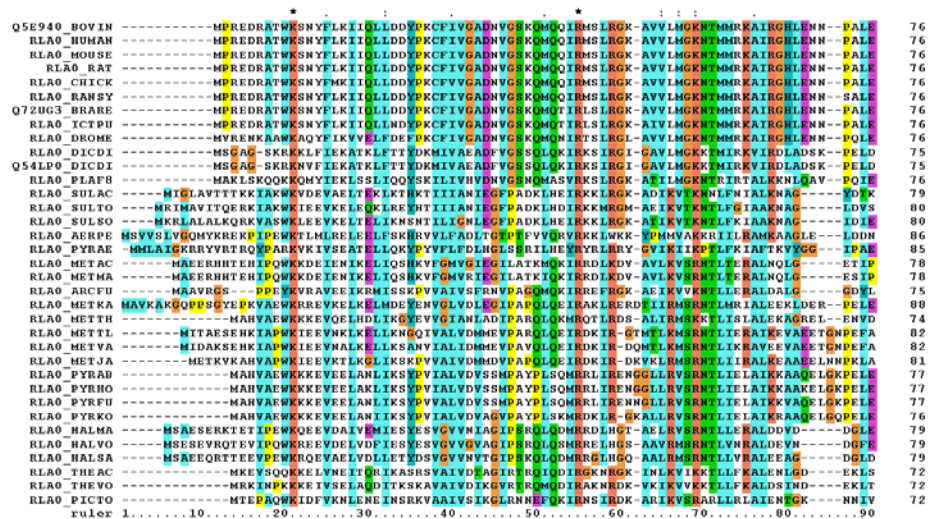

Fig 6. A multiple sequence alignment

9

Computer aided drug design is the use of computational techniques to cut down the search for drug molecules. A large class of diseases arise out of an unwelcome molecule, possibly a protein produced from the gene of a pathogen, an intruder organism, like a virus. A simplified picture of diseases could be given based on "good" and "bad" proteins. The human body can be assumed to be producing proteins P1, P2, P3 ··· that are useful and required for the human body. When a pathogen, a virus or a bacteria, enters the human body, it could produce its own protein, say X, which is possibly harmful. How exactly is it harmful? X could interact and form a complex, in which two molecules are bound together into a new one, with one of the good proteins, say P1, thereby inhibiting it from its routine activities and causing the onset of a disease. The strategy to combat the disease is to introduce a new molecule, say Y, into the body such that X is more attracted to Y than to P1, thereby freeing P1 to get back to routine work. It must be noted that all diseases do not fit into this model. Sometimes, our own protein-making machinery can go wrong and produce P1' instead of P1, causing disease. Identifying a disease and bringing out an effective drug into the market could take anywhere from 10–15 years, cost up to US$800 million, and involve testing of up to 30,000 candidate molecules. The economic significance of the activity thus needs no special emphasis. This costly, time-consuming activity has been traditionally based on a blind search for molecules, rightly termed as serendipitous discovery. Computer aided drug design or rational drug design has cut the cost and time of drug discovery with great effect. Today computationally it is possible to select candidate drug molecules from huge available databases and check whether it can bind to the active site of the troublesome molecule using computational *docking* procedures. **Docking software** such as Hex, Argus Lab, and Autodock are capable of docking the small molecules to selected *active sites* of target molecules and give a relative score for the binding. The small number of (a few dozen) of molecules thus predicted computationally is then passed on to the wet lab for synthesis and clinical trials.

Molecular Phylogenetics is where biologists have been ahead of computer scientists for long. Computer programmers started talking about "classes" after they got their hands sticky with spaghetti code for close to 2 decades. Biologists have been known to have put their house in order in 1750s itself, trying to make classes, superclasses and abstract classes of all organisms. A phylogenetic tree is a pictorial representation of such classifications. These are starting points of studies on **evolution**.
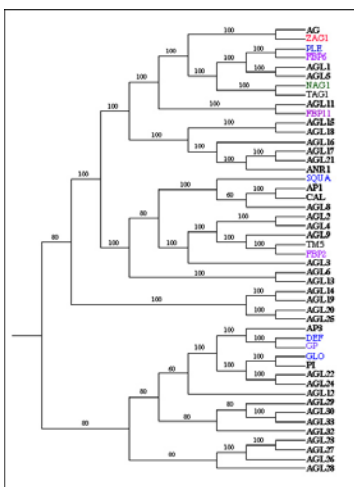

Fig 7. A Phylogenetic tree

Biologists were till recently not having quantitative criteria to do this. But today they do have: the dna/protein sequences. Based on their comparison we are today re assured that chimpanzees are our close cousins. This requires quite involved computation, including facing some intractable problems. Most phylogenetic trees are computed based on a multiple-sequence alignment of protein sequences of the set of organisms that we want to classify.

Bio data bases are huge data bases of mostly sequence data pouring in from many genome sequencing projects going on all over the world. The primary data bases include European Molecular Biology Laboratory DNA database **(EMBL)**, **GenBank** at National Center for Biotechnology information, Bethesda and DNA Data Bank Japan **(DDBJ)**, and Protein databases at **SWISS-PROT** (Protein sequence database at Swiss Institute of Bioinformatics, Geneva) and **PDB –** the Protein 3D structure databases. As the databases continue to grow, mining them offers newer challenges.

I will now describe **Microarray Bioinfromatics.** Micro arrays are tiny chips that are used to study a phenomena called gene expression. All the genes in all the cells are not active all the time. Which of them are "expressed" at any given time/situation is the question that microarrays help to answer. Microarray chips have fragments of "normal" human DNA stuck to tiny spots such that if you sprinkle appropriately modified DNA fragments of yours, they will stick to each other where they match. If you sprinkle two sets taken at two different states, identified by fluorescent coloring, then a digital image can be derived out of the micro array, which looks like a set of fluorescent spots in green, yellow and red. Biologists need answers to many questions about gene expression based on these images. This is the scope of microarray bioinformatics. Before answering the questions of the biologists, there are some basic image processing to be done: gridding the image, normalizing intensities etc. After this, a lot of statistical analysis is called for, mainly clustering of the data taken at different intervals or states. K-means clustering, principal component analysis and other popular statistical tools are in great use in this area.
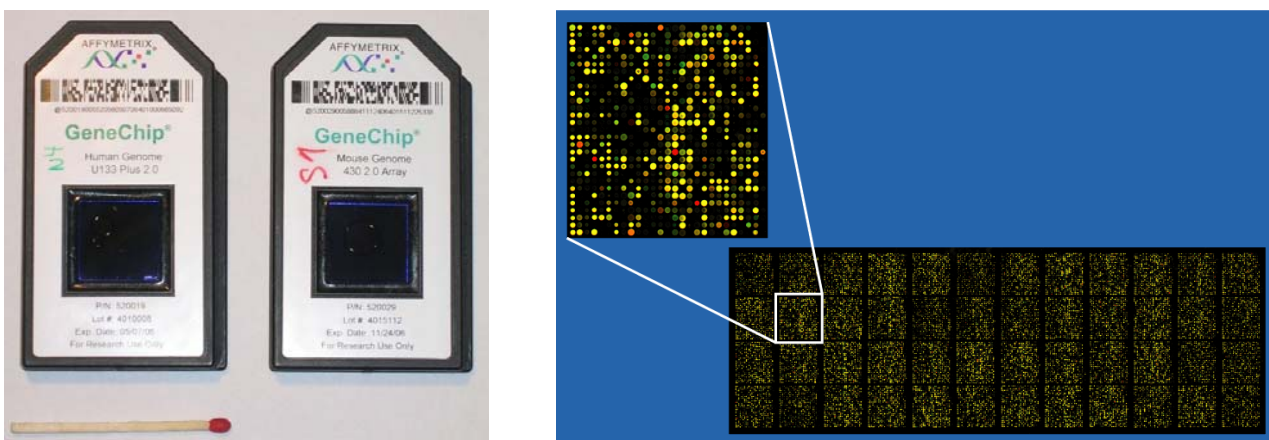

Fig 8. Left: DNA Microarray Chips, Right: Microarray Image

Systems Biology is where engineers might be turned on. Engineers can claim great success in modeling some of the very complex creations of his/her own – huge power systems, towering multi-storeyed structures, amazing kilometer-long bridges, and miniature silicon chips. However,

they are yet to face the grand challenge of modeling the engineering of life at the cellular level. In a power system, the electrical engineer is able to predict with required level of accuracy, what the effects of a particular loading would be, at every spot of interest in the power network. However, ask the biologist, if the pH in a cell compartment is increased, what would happen at every important spot in the cell after an hour (cell machinery is mostly sluggish, but lightning fast at times). She does not have a model to predict. The field of systems biology attempts exactly this, to identify the basic components, parameters, variables and networks and to model them with differential/integral equations to the extent that the previous question can be answered. The Japanes project named **e-Cell** is a great beginning towards this. This is an international research project aiming to model and reconstruct biological phenomena *in- silico,* and developing necessary theoretical supports, technologies and software platforms to allow precise whole cell simulation. The latest version of their cell simulation software is available at www.e-cell.org.

5. Concluding Remarks
As a recent convert into the field, I am still amazed and excited by the beauty, complexity and challenge of analyzing information that is exploding from biological systems. I hope I have been successful to a good measure in transferring my excitement to you. I also hope that the atlas of the domain of Bioinformatics and Computational Biology that I have sketched has given you enough exposure to help you make your own judgment about the depth and breadth of the field and also decide on your destinations to frequent.

6. Acknowledgements

7. To Probe further
I shall limit my suggestions to just two contrasting books. **Bioinformatics for Beginners** (actually *for Dummies,* but renamed for Indian text book market !) by Jean-Michel **Claverie** and Cedric **Notredame**, presents an amusing writing with web-based experiments explained cleanly. **Introduction to Computational Molecular Biology** by **Setubal** and **Meidanis** presents the computational challenges in the field, aimed at hard-core computer scientists. For Journals in the filed, see a compilation below.

Journals in Bioinformatics

Compiled in December, 2006 by Betsy Sheena Cherian

	PLoS Computational Biology is an open-access, peer-reviewed journal published monthly by the Public Library of Science (PLoS) in association with the International Society for Computational Biology (ISCB). *PLoS Computational Biology* features works of exceptional significance that further our understanding of living systems at all scales through the application of computational methods. www.compbiol.plosjournals.org
	Bioinformatics, from Oxford University Press, publishes the highest quality scientific papers and review articles of interest to academic and industrial researchers. Its main focus is on new developments in genome bioinformatics and computational biology. Some articles and archives are open access. Impact factor: 6.019, www.bioinformatics.oxfordjournals.org
 IEEE/ACM TRANSACTIONS	**IEEE/ACM Transactions on Computational Biology and Bioinformatics** is a quarterly publishing archival research results related to the algorithmic, mathematical, statistical, and computational methods that are central in bioinformatics and computational biology; the development and testing of effective computer programs in bioinformatics; the development and optimization of biological databases; and important biological results that are obtained from the use of these methods, programs, and databases. www.computer.org/tcbb
	In silico Biology (ISB) is an international a peer-reviewed, open access journal on computational molecular biology. It focuses on biologically significant computational methods and results and aims at providing essential contributions to Systems Biology. It is issued online (Germany) & print (IOS Press, New zealand) www.bioinfo.de/isb/index.html
	BMC Bioinformatics is an open access journal publishing original peer-reviewed research articles in all aspects of computational methods used in the analysis and annotation of sequences and structures, as well as all other areas of computational biology. The journal is published by BioMed Central Ltd, UK. Impact factor is 4.96 www.biomedcentral.com/bmcbioinformatics/
	EURASIP Journal on Bioinformatics and Systems Biology publishes research results related to signal processing and bioinformatics theories and techniques relevant to a wide area of applications into the core new disciplines of genomics, proteomics, and systems biology. http://www.hindawi.com/journals/bsb/
	Algorithms for Molecular Biology is an open access, peer-reviewed online journal that encompasses all aspects of algorithms and software tools for molecular biology and genomics. Areas of interest include algorithms for RNA and protein structure analysis, gene prediction and genome analysis, comparative sequence analysis and alignment, phylogeny, gene expression, machine learning, and combinatorial algorithms. www.almob.org
	Briefings in Bioinformatics publishes reviews for the users of databases and analytical tools of contemporary genetics and molecular biology and provides practical help and guidance to the non-specialist. //bib.oxfordjournals.org/
Other Journals	**Nucleic Acids Research** : www.nar.oxfordjournals.org, **Protein Science:** www.proteinscience.org, **DNA Research:** www. dnaresearch.oxfordjournals.org, **Online Journal of Bioinformatics:** www.cpb.ouhsc.edu/ojvr/bioinfo.htm, **IEEE Transactions on nano Bioscience**: www.ieee.org, **Bioinformation**: www.bioinformation.net

13