Innovative Romanian Food Biotechnology *Vol. 2 No. 2 Issue of September 25, 2008*

Received December 23, 2008 / Accepted December 30, 2008

REVIEW ARTICLE

BIOINFORMATICS APPLIED IN BIOREMEDIATION

Professor Dr. M.H. Fulekar[*] & Jaya Sharma

University of Mumbai, Environmental Biotechnology Laboratory, Department of Life Sciences, Santacruz (E), Mumbai-400 098

Abstract

Bioinformatics is the combination of biology and information technology which focuses on cellular and molecular levels for application in modern biotechnology. Bioremediation is the recent technology which explores the microbial potentiality for biodegradation of xenobiotics compounds. Microorganisms display a remarkable range of contaminant degradation ability that can efficiently and effectively restore natural environmental conditions. Attempts have been made to interpret some areas of genomics and proteomics which have been employed in bioremediation studies .Bioinformatics requires the study of microbial genomics, proteomics, systems biology, computational biology, phylogenetic trees, data mining and application of major bioinformatics tools for determining the structures and biodegradative pathways of xenobiotic compounds. This paper highlights the significance of bioinformatics concepts applied in the bioremediation fields.

Key words: Bioremediation, Bioinformatics, Proteomics, Genomics.

Introduction

Environmental pollutants have become a major global concern, given their undesirable recalcitrant and xenobiotic compounds. A variety of polycyclic aromatic hydrocarbons (PAHs), xenobiotics, chlorinated and nitro-aromatic compounds were depicted to be highly toxic, mutagenic and carcinogenic for living organisms. Nevertheless, as a result of their diversity, versatility and adaptability, a number of microorganisms are considered to be the best candidates among all living organisms to remediate most of the environmental contaminants into the natural biogeochemical cycle. These microorganisms display a remarkable range of contaminant degradable ability that can efficiently restore natural environmental conditions. However, a variety of contaminants have been shown to be unusually recalcitrant, i.e. microorganisms either do not metabolize or transform them into certain other metabolites that again accumulate in the

[*] *Corresponding authors: mhfulekar@yahoo.com*

environment. Therefore, it may be more productive to explore new catabolic pathways that might lead towards complete mineralization of these pollutants. One of the reasons, our knowledge of microbial degradation pathways is so incomplete is the immense complexity of microbial physiology that allows response and adaptability to various internal and external stimuli (Fulekar, 2007).

Bioremediation has a potential to restore contaminated environments inexpensively yet effectively, but a lack of information about the factors controlling the growth and metabolism of microorganisms in polluted environments often limits its implementation. Researchers now have the ability to cultivate microorganisms that are important in bioremediation and can evaluate their physiology using a combination of genome-enabled experimental and modeling techniques. In addition, new environmental genomic and proteomic techniques offer the possibility for similar studies (Nair, 2007). Bioinformatics is based on proteomics and genomics offer remarkable promise as tools to address longstanding questions regarding the molecular mechanisms involved in the control of mineralization pathways. During mineralization, transcript structures and their expression have been studied using high-throughput transcriptomics techniques with microarrays. Generally however, transcripts have no ability to operate any physiological response; rather, they must be translated into proteins with significant functional impact. These proteins can be identified by proteomic techniques using powerful two-dimensional polyacrylamide gel electrophoresis (2-DE). Towards the establishment of functional proteomics, the current advances in mass spectrometry (MS) and protein microarrays play a central role in the proteomics approach. Exploring the differential expression of a wide variety of proteins and screening the entire genome for proteins that interact with particular mineralization regulatory factors would help us to gain insights into bioremediation (Fulekar, 2008).

Overview of Bioremediation

Bioremediation is defined as a process by which microorganisms are stimulated to rapidly degrade hazardous organic pollutants to environmentally safe levels in soils, sediments, substances, materials and ground water. Recently, biological remediation process has also been devised to either precipitate effectively or immobilize inorganic pollutants such as heavy metals. Stimulation of microorganisms is achieved by the addition of growth substances such as nutrients, terminal electron acceptors/donors or some combination thereby resulting is an increase in organic pollutant degradation and biotransformation. The energy and carbon are obtained through the metabolism of organic compounds by the microbes involved in bioremediation processes (Fulekar, 2005).

Biodegradation is nature's way of recycling wastes, or breaking down organic matter into nutrients that can be used by other organisms. The degradation is carried out by the microorganisms: bacteria, fungi, insects, worms etc. by taking nutrients such as C,N,P from the contaminant which on long term acclimatization convert the toxic compound into environment friendly compound. By harnessing these natural forces of biodegradation, people can reduce wastes and clean up some types of environmental contaminants. Through composting, we accelerate natural biodegradation and convert organic wastes to a valuable resource. Wastewater treatment also accelerates natural forces of biodegradation, breaking down organic matter so that it will not cause pollution problems when the water is released into the environment. Through bioremediation, microorganisms are used to clean up oil spills and other types of organic pollution. Therefore, *in situ* bioremediation provides a technique for cleaning up pollution by enhancing the same biodegradation processes that occur in nature (safer, less expensive and treatment in place).

Bioremediation of a contaminated site typically works in one of two ways:

- To enhance the growth of whatever pollution-eating microbes might already be living at the contaminated site
- Specialized microbes are added to degrade the contaminants (less common).

The fields of Biodegradation and Bioremediation offer many interesting and unexplored possibilities from the bioinformatics point of view. They need to integrate a huge amount of data from different

 Received December 23, 2008 / Accepted December 30, 2008

sources: chemical structure and reactivity of the organic compounds; sequence, structure and function of proteins (enzymes); comparative genomics; environmental biology etc. Bioinformatics provides data base for microarrays, gene identification and microbial degradation pathways of compounds (Ellis et al. 2001).

Bioinformatics

Bioinformatics is the combination of biology and information technology. It is the branch of science that deals with the computer based analysis of large biological data sets. Bioinformatics incorporates the development to store and search data and of statistical tools and algorithms to analyze and determine relationships between biological data sets, such as macromolecular sequences, structures, expression profiles and biochemical pathways. Bioinformatics is the focus on cellular and molecular levels of biology. Biology and computers are becoming close cousins which are mutually respecting, helping and influencing each other and synergistically merging more than ever (Fulekar, 2008). The huge data from biology mainly in the form of DNA, RNA and protein sequences is putting heavy demand on computers and computational scientists. Bioinformatics has taken on a new glittering by entering in the field of Bioremediation. Bioinformatics is the application of computer sciences and related technology to the industries for using the huge available database for computational biology. Computational biologists are those who are specialized in using of computational tools and computer systems to solve the problems of biology in the area of bioinformatics (Westhead, 2003). The major branches of bioinformatics are genomics, proteomics, biological databases, data mining, molecular phylogenetics, microarray informatics and systems biology, which are playing a vital role in understanding bioinformatics and its applications. The roles of bioinformatics related tools are described for bioremediation of hazardous wastes to develop environmental clean up technology.

1. Proteomics

The terms 'proteomics' and 'proteome' were introduced in 1995 , which is a key post genomic feature that emerged from the growth of large and complex genome sequencing datasets. Proteomic analysis is particularly vital because the observed phenotype is a direct result of the action of the proteins rather than the genome sequence. Traditionally, this technology is based on highly efficient methods of separation using two-dimensional polyacrylamide gel electrophoresis (2-DE) and modern tools of bioinformatics in conjunction with mass spectrometry (MS). However, 2-DE has been considered to be a limited approach for very basic and hydrophobic membrane proteins in compartmental proteomics. In bioremediation, the proteome of the membrane proteins is of high interest, specifically in Polycyclic Aromatic Hydrocarbon biodegradation. The improvements in 2-DE for use in compartmental proteomics have been made by introducing an alternative approach for multidimensional protein identification technology (MudPIT) (Santos, 2004).

1.1. Bioremediation using Proteomics

The cellular expression of proteins in an organism varies with environmental conditions. The changes in physiological response may occur due to the organism's adaptive responses to different external stimuli, such as the presence of toxic chemicals in the environment. The advent of proteomics has allowed an extensive examination of global changes in the composition or abundance of proteins, as well as identification of key proteins involved in the response of microorganisms in a given physiological state. A number of reports have described sets of proteins that are up- or down-regulated in response to the presence of specific pollutants. PAHs, ubiquitous environmental pollutants are extremely important to remove from the environment. In situ and ex situ bioremediation of PAHs has been partially achieved using natural and genetically engineered microorganisms. Using a proteomics approach, the physiological changes in an organism during bioremediation provide further insight into bioremediation-related genes and their regulation. An 81-kDa protein similar to catalase–peroxidase

that expressed in response to pyrene exposure was recovered using 2-DE from *Mycobacterium* sp. strain PYR-1. Later, six major proteins were significantly induced and overexpressed on 2-DE when *Mycobacterium* sp. strain PYR-1 was exposed to phenantherene, dibenzothiophene and pyrene. Several pyrene-specific polypeptides were identified by N-terminal and internal peptide sequencing as putative enzymes. Furthermore, the induction of two ring-hydroxylating dioxygenases, i.e. Pdo1 and Pdo2, in response to pyrene was proposed during pyrene catabolism by *Mycobacterium* sp. strain 6PY1. A composite profile for 20 PAH-induced proteins was presented when organism *Mycobacterium vanbaabenii* PYR-1was grown in the presence of high-molecular-weight PAHs. Progress has been made towards identification of unknown genes and proteins during anaerobic biodegradation of toluene and ethylbenzene. A global expression analysis (DNA microarray and proteomics) was performed using denitrifying bacterium strain EBN1 adapted to anaerobic growth with benzoate, toluene, ethylbenzene and a mixture of toluene and ethylbenzene. Besides various differentially expressed genes and related proteins, the expression of two toluene-related operons (bss and bbs) was specifically induced in toluene-adapted cells. In agreement with the sequential regulation of the ethylbenzene pathway, Ebd proteins were reported to be formed in ethylbenzene-adapted cells but not in acetophenon-adapted cells, while Apc proteins were found to be formed under both conditions. The recent combined approaches of transcriptomics and proteomics have revealed new pathways for aerobic and anaerobic biodegradation of toxic wastes that will certainly pave the way for further identification of new signature proteins (Chen, 2009).

2. Genomics

Genomic is a powerful computer technology used to understand the structure and function of all genes in an organism based on knowing the organism's entire DNA sequence. The field includes intensive efforts to determine the entire DNA sequence of organisms and fine-scale genetic mapping efforts. The field also includes studies of intragenomic phenomena such as heterosis, epistasis, pleiotropy and other interactions between loci and alleles within the genome. In contrast, the investigation of single genes, their functions and roles, something very common in today's medical and biological research, and a primary focus of molecular biology, does not fall into the definition of genomics, unless the aim of this genetic, pathway, and functional information analysis is to elucidate its effect on, place in, and response to the entire genome's networks.

2.1. Bioremediation using Genomics

Non-molecular techniques: at present, most applied microbiological investigations of bioremediation processes make use of the 'treatability study' in which samples of the contaminated environment are incubated in the laboratory and the rates of contaminant degradation or immobilization are documented. Such studies provide an estimate of the potential metabolic activity of the microbial community, but give little insight into the microorganisms that are responsible for bioremediation, or why particular amendments that can be evaluated for engineered bioremediation applications do or do not stimulate activity.

When bioremediation processes are researched in more detail, attempts are generally made to isolate the organisms responsible. The isolation and characterization of pure cultures has been and will continue to be crucial for the development and interpretation of molecular analysis. The recovery of isolates that are representative of the microorganisms responsible for bioremediation processes can be invaluable because, as outlined below, studying these isolates provides the opportunity to investigate not only their biodegradation reactions , but also other aspects of their physiology that are likely to control their growth and activity in contaminated environments. However, before the application of molecular techniques to bioremediation, it was uncertain whether the isolated organisms were important in bioremediation in situ, or whether they were weeds that grew rapidly in the laboratory but were not the primary organisms responsible for the reaction of interest in the environment (Fulekar, 2005).

Evolutionary approaches are extremely useful for optimization of an entire biodegradation pathway comparing to step by- step modifications offered by

 Received December 23, 2008 / Accepted December 30, 2008

rational design. This was recently demonstrated by the modification of an arsenic resistance operon using DNA shuffling. Cells expressing the optimized operon grew in up to 0.5 M arsenate, a 40-fold increase in resistance. Moreover, a 12-fold increase in the activity of one of the gene products (arsC) was observed in the absence of any physical modification to the gene itself. The authors speculate that modifications to other genes in the operon effect the function of the *arsC* gene product. Such unexpected but exciting results are more likely to be realized using irrational approaches. This strategy is particularly attractive since the ultimate goal of many remediation approaches is for complete mineralization of the pollutants, and the concurrent optimization of an entire pathway will allow the efficient search for the correct coordination between a complex set of biodegradation reactions. Along the same line, recent advances in genome shuffling between species, which allow the exchange and recombination of diverse pathways into a single species, will further accelerate the discovery of novel microbes that are useful for the remediation of even a complex mixture of pollutants (Wilfred, 2005).

3. Genomics and Proteomics in Bioremediation

The growing demands of genomics and proteomics for the analysis of gene and protein function from a global bioremediation perspective are enhancing the need for microarray-based assays enormously. In the past, protein microarray technology has been successfully implicated for the identification, quantification and functional analysis of protein in basic and applied proteome research. Other than the DNA chip, a large variety of protein-microarray based approaches have already been verified that this technology is capable of filling the gap between transcriptomics and proteomics (Singh et al. 2006).

The availability of bacterial genomes relevant to biodegradation in recent years has allowed the feasibility to study the complex interactions between cellular reactions from a genomic and proteomic level. A quantitative understanding of how cells function requires every gene and protein to be placed in their dynamic context, which entails the integrated consideration of many interacting components. From this perspective, a system biology approach is necessary to predict the functioning of an organism in a complex environment and to describe the outcome of the thousands of individual reactions that are simultaneously taking place in a microbial cell. So far, such prokaryotic models have been limited primarily to *E. coli* and a few pathogens. However, similar modeling approaches should be able to predict contaminant bioremediation by microorganisms that are known to predominate in polluted environments. Recently, de Lorenzo et al. (2003) presented a pioneering study on the characteristics of the "global biodegradation network", in which they considered the global pool of known chemical reactions implicated in biodegradation regardless of their microbial hosts. The characteristics of this network support an evolutionary scenario in which the reactions evolved from the central metabolism toward more diversified reactions, allowing us to understand the evolution of new pathways for the degradation of xenobiotics and provide the basis for predicting the abilities of chemicals to undergo biological degradation, and for quantifying the evolutionary rate for their elimination in the future. This type of analysis, when coupled with the predictive approach for microbial catabolism using the University of Minnesota Biocatalysis/Biodegradation Database (UM-BBD) as a knowledge base and various sets of heuristic rules, 33 will lead to untapped and improved strategies for bioremediation. This represents an excellent opportunity for chemical engineers who are already involved with system biology, and will undoubtedly evolve into an important research direction within the next 5 years (Wilfred, 2005).

4. Systems Biology

The rise of genomic technologies and systems biology provide fresh approaches to currently untactable biological processes that are at the root of serious environmental problems. One formidable challenge in this respect is the biological fate of the nearly 8 operons, etc. implicated in this process. The biodegradation database of the University of Minnesota documented new chemical compounds (~40 000 predominant) which are common in modern Organic and Industrial Chemistry. A large

number of microbial strains are able to grow on environmental pollutants (about 800 today). Bioremediation was studied from a molecular biology point of view, characterizing the chemical reactions, genes; University of Minnesota has made a pioneering effort in putting together nearly every aspect of our current knowledge on biodegradation pathways and in developing systems for dealing with that data e.g. to learn rules for predicting biodegradative features. Yet, most information available in the literature of microbial biodegradation of xenobiotics and recalcitrant chemicals deals with duos consisting of one pollutant versus one strain and thus, lacks essential aspects of the natural scenarios, like the interchange of genes between bacteria or their metabolic cooperation. This study of genomes and 'functionomes' from a community point of view (in contrast to organism point of view) is leading, for example, to the sequencing of 'genomes' of communities and ecosystems, instead of single organisms. These circumstances expose the need to qualify and to represent the information available in biodegradation databases in a fashion in which the entire known biodegradative potential of the microbial world can be crossed with the whole collection of compounds known to be partially or totally degraded through (mostly) bacterial action (Kitano, 2002).

5. Computational biology

A computational biology is a sub discipline within bioinformatics concerned with computation-based research devoted to understanding basic biological processes. It encompasses the fields of:

- Bioinformatics, which applies algorithms and statistical techniques to the interpretation, classification and understanding of biological datasets. Datasets typically consist of large numbers of DNA, RNA, or protein sequences. Sequence alignment is used to assemble the datasets for analysis. Comparisons of homologous sequences, gene finding, and prediction of gene expression are the most common techniques used on assembled datasets; however, analysis of such datasets have many applications throughout all fields of biology.
- Computational biomodeling, a field within biocybernetics concerned with building computational models of biological systems.
- Computational genomics, a field within genomics which studies the genomes of cells and organisms. High-throughput genome sequencing produces lots of data, which requires extensive post-processing (genome assembly) and uses DNA microarray technologies to perform statistical analyses on the genes expressed in individual cell types. This can help find genes of interests for certain diseases or conditions. This field also studies the mathematical foundations of sequencing.
- Molecular modeling, which consists of modelling the behaviour of molecules of biological importance.
- Systems biology, which uses systems theory to model large-scale biological interaction networks (also known as the interactome).
- Protein structure prediction and structural genomics, which attempt to systematically produce accurate structural models for three-dimensional protein structures that have not been determined experimentally.
- Computational biochemistry and biophysics, which make extensive use of structural modeling and simulation methods such as molecular dynamics and Monte Carlo method-inspired Boltzmann sampling methods in an attempt to elucidate the kinetics and thermodynamics of protein functions.(Nair, 2007)

6. Phylogenetic trees

A phylogenetic tree or evolutionary tree is a tree showing the evolutionary relationships among various biological species or other entities that are believed to have a common ancestor. In a phylogenetic tree, each node with descendants represents the most recent common ancestor of the descendants, and the edge lengths in some trees correspond to time estimates. Each node is called a taxonomic unit. Internal nodes are generally called hypothetical taxonomic units (HTUs) as they cannot be directly observed.

Received December 23, 2008 / Accepted December 30, 2008

Types

A rooted phylogenetic tree is a directed tree with an unique node corresponding to the (usually imputed) most recent common ancestor of all the entities at the leaves of the tree. The most common method for rooting trees is the uses of an uncontroversial outgroup — close enough to allow inference from sequence or trait data, but far enough to be a clear outgroup.

Unrooted trees illustrate the relatedness of the leaf nodes without making assumptions about common ancestry. While unrooted trees can always be generated from rooted ones by simply omitting the root, a root cannot be inferred from an unrooted tree without some means of identifying ancestry; this is normally done by including an outgroup in the input data or introducing additional assumptions about the relative rates of evolution on each branch, such as an application of the molecular clock hypothesis.

A dendogram is a broad term for the diagrammatic representation of a phylogenetic tree.

A cladogram is a tree formed using cladistic methods. This type of tree only represents a branching pattern, i.e., its branch lengths do not represent time.

A phylogram is a phylogenetic tree that explicitly represents number of character changes through its branch lengths.

An ultrametric tree or chronogram is a phylogenetic tree that explicitly represents evolutionary time through its branch lengths.

7. Data Mining:

Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases (Cipolla et al. 1995).

- Non-traditional Feature Selection
  - ❖ When the number of attributes >> number of samples?
  - ❖ Highly imbalanced
- Explainable and Accurate Data Mining Methods
- NN, SVM-> Rules?
- Transfer Learning
  - ❖ Can knowledge learned from one set of samples help data mining on another sample?
- Exploiting the network structure
  - ❖ Individual i.i.d type of classification vs social networks?
- Current methods, such as SVMs, discriminant analysis, neural networks, are '**black box**' models.
- The learned knowledge is hard to understand by biologists.
- Some potential solutions
  - ❖ **Logic based method**, e.g., decision trees and variants may be better in giving the 'IF-THEN' like rules that explicitly define the epigenetic logics in cancer and stem cell development.

DNA methylation rules can be learned by using SVM based recursive feature elimination and fuzzy logics.

Major categories of Bioinformatics Tools

There are both standard and customized products to meet the requirements of particular projects. There are data-mining software that retrieves data from genomic sequence databases and also visualization tools to analyze and retrieve information from proteomic databases. These can be classified as homology and similarity tools, protein functional analysis tools, sequence analysis tools and miscellaneous tools.

Here is a brief description of a few of these. Everyday bioinformatics is done with sequence search programs like BLAST, sequence analysis

programs, like the EMBOSS and Staden packages, structure prediction programs like THREADER or PHD or molecular imaging/modelling programs like RasMol and WHATIF.

Homology and Similarity Tools

Homologous sequences are sequences that are related by divergence from a common ancestor. Thus the degree of similarity between two sequences can be measured while their homology is a case of being either true of false. This set of tools can be used to identify similarities between novel query sequences of unknown structure and function and database sequences whose structure and function have been elucidated.

Protein Function Analysis

This group of programs allows one to compare a certain protein sequence to the secondary (or derived) protein databases that contain information on motifs, signatures and protein domains. Highly significant hits against these different pattern databases allow one to approximate the biochemical function of the query protein.

Structural Analysis

This set of tools allows one to compare structures with the known structure databases. The function of a protein is more directly a consequence of its structure rather than its sequence with structural homologs tending to share functions. The determination of a protein's 2D/3D structure is crucial in the study of its function.

Sequence Analysis

This set of tools allows one to carry out further, more detailed analysis on the query sequence including evolutionary analysis, identification of mutations, hydropathy regions, CpG islands and compositional biases. The identification of these and other biological properties are all clues that aid the search to elucidate the specific function of the sequence.

Bioinformatics in Bioremediation-MetaRouter

MetaRouter is a system for maintaining heterogeneous information related to Biodegradation in a framework that allows its administration and mining (application of methods for extracting new data). It is an application intended for laboratories working in this area which need to maintain public and private data, linked internally and with external databases, and to extract new information from it.

The system has an open and modular architecture adaptable to different customers. This multy-platform program, implemented in Postgre SQL (standard language for relational databases) and using SRS as an indexing system (used to connect and query Molecular Biology databases), works using a client/server architecture that allows the program to run on the user station or on the company server, so it can be accessed from any place in a secure way just by having a web browser.

The University of Minnesota Biocatalysts/Biodegradation Database (http://www.labmed.umn.edu/umbbd) begins its fifth year having met its initial goals. It contains approximately 100 pathways for microbial catabolic metabolism of primarily xenobiotic organic compounds, including information on approximately 650 reactions, 600 compounds and 400 enzymes, and containing approximately 250 microorganism entries. It includes information on most known microbial catabolic reaction types and the organic functional groups they transform. Having reached its first goals, it is ready to move beyond them. It is poised to grow in many different ways, including mirror sites; fold prediction for its sequenced enzymes; closer ties to genome and microbial strain databases; and the prediction of biodegradation pathways for compounds it does not contain (Ellis et al. 2000).

Conclusion

Bioinformatics technology has been developed to identify and analyze various components of cells such as gene and protein functions, interactions, metabolic and regulatory pathways. Bioinformatics analysis will facilitate and quicken the analysis of

Received December 23, 2008 / Accepted December 30, 2008

cellular process to understand the cellular mechanism to treat and control microbial cells as factories. The next decade will belong to understanding molecular mechanism and cellular manipulation using the integration of bioinformatics. Bioinformatics has wide application in bioremediation for the structure determination and pathways of biodegradation of xenobiotics.

References

1. Chen X., ShaoPing H., ChaoFeng S., ChangMing D., JiYan S., Chen Y. (2009). Interaction of *Pseudomonas putida* CZ1 with clays and ability of the composite to immobilize copper and zinc from solution. *Bioresource Technology*., 100.330–337.
2. Cipolla and Emil T. (1995). Data Mining: Techniques to Gain Insight into Your Data *Enterprise Systems Journal* .64.18-24.
3. Ellis L.B., Hershberger C.D., Bryan M.B., Wackett L.P.(2001). The University of Minnesota Biocatalysis/Biodegradation database: microorganisms, genomics and prediction. *Nucleic Acids Res*. 29(1):340-343.
4. Ellis L.B., Hershberger C.D., Wackett L.P.(2000). The University of Minnesota Biocatalysis / Biodegradation database: microorganisms, genomics and prediction. *Nucleic Acids Res*. 28(1):377-9.
5. Fulekar M.H. (2005). *Environmental Biotechnology*. Oxford & IBH publication, New Delhi.
6. Fulekar M.H. (2008). *Bioinformatics – Application in Life* & **Environment Sciences** Capital & Springer publication. Germany.
7. Fulekar M.H. (2008). *Environmental Biotechnology*.(In press).Science publisher.USA.
8. Fulekar M.H. (2007) Bioremediation technologies for environment. *Indian journal of environmental protection*. 27, no. 3, 264-271
9. Fulekar M.H. (2005) Bioremediation technologies for environment. *Indian journal of environmental protection*.25, no. 4: 358-364.
12. Guijas D. and Florencio P., Alma bioinformatics. Bioinformatics in Bioremediation MetaRouter**.Book**
13. Kim S.I., Kim J.Y., Yun S.-H., Kim J.H., Leem S.H. and Lee C. (2004). Proteome analysis of *Pseudomonas* sp. K82 biodegradation pathways, *Proteomics.* , 4, 3610–3621.
14. Kitano H. (2002). Systems Biology: a brief overview, *Science,* 295:1662-1664.
15. Lovely D. (2003). Cleaning up with Genomics-Applying Molecular Biology to Bioremediation. *Nature* .1. 35-43.
16. Nair, A.S. (2007). Computational Biology & Bioinformatics- A Gentle Overview. *CSI Communications.* ,30,7-12.
17. Santos P.M., Benndorf D., Sa-Correia I. (2004). Insights into *Pseudomonas putida* KT2440 response to phenol-induced stress by quantitative proteomics. *Proteomics*; 4**:**2640–52
18. Singh O. V. and Nagathihalli S. N.(2006). Transcriptomics, proteomics and interactomics: unique approaches to track the insights of bioremediation. Briefings In Functional Genomics And *Proteomics*.4. 355-362.
19. Westhead D.R., Parish R.M., Twyman. (2003). Bioinformatics-Instant Notes. Bios *Sceintific Publisher.*UK.
20. Wilfred Chen, Ashok Mulchandani, and Marc A. Deshusses. (2005).Environmental Biotechnology: Challenges and Opportunities for Chemical Engineers. *AIChE Journal*. 51.693.
21. Yang Q.(2007). Computer Science and Engineering, HKUST. Data Mining and Bioinformatics: Some Challenges.
22. http://www.geocities.com/bioinformaticsweb/tools.html
23. http://en.wikipedia.org/wiki
24. http://aleph0.clarku.edu/~djoyce/java/Phyltree/cover.html