Journal Homepage: - www.journalijar.com
INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)
Article DOI:10.21474/IJAR01/4371
DOI URL: http://dx.doi.org/10.21474/IJAR01/4371

RESEARCH ARTICLE

PATENTING BIOINFORMATICS INNOVATIONS: EMERGING TRENDS AND CHALLENGES IN THE UNITED STATES.

TuhinChatterjee.
Ph.D. candidate, The West Bengal National University of Juridical Sciences (WB NUJS) and Indian Patent Attorney.

Manuscript Info

Manuscript History

Received: 26 March 2017
Final Accepted: 20 April 2017
Published: May 2017

Key words:-
Bioinformatics, Computational biology, *In Silico* screening, Patent,Structural genomics, USPTO

Abstract

Bioinformatics tools and techniques are useful not only to manage and analyze vast amount of raw biological data generated from various genomics research but also to understand the phenomena of biological system at the macromolecular level. The development of bioinformatics has come a long way from DNA sequencing tools of the Human Genome Project (HGP) era to DNA circuits and programmable synthetic biological devices in the twenty first century. The present article attempts to analyze and reveal the emerging trends in bioinformatics and computational biology research and innovation and challenges in patenting them under the current US patent regime.

Introduction:-

Rapid development of high-throughput techniques in molecular biology and computational methods transformed biology into a data-rich science (Beck 2010). Genome projects are not only changing the understanding of biology but also generating mountains of *omics* (Barh, Zambare, and Azevedo 2013) data. For example, the Human Genome Project alone has generated vast amount of nucleotide data containing 3 billion base pairs (bp) residing in the 23 pairs of human chromosomes ("Human Genome Project Completion: Frequently Asked Questions" 2010). Analyzing, storing, organizing and retrieving raw biological data significantly propelled biological research. Bioinformatics and computational biology techniques manage the data interpretation and analytical tasks in a very efficient manner and offer useful information about how biological systems work and evolve over time. The nucleotide and amino acid sequence information are frequently used in conventional biological research. Besides that, sequencing new genes and assigning their functions, discovering single nucleotide polymorphisms (SNP), modelling three-dimensional (3D) structures of proteins etc. added a whole new dimension to modern biological research and development.

Although many bioinformatics tools and databases are publicly available in the internet, e.g. BLAST (Altschul et al. 1990), GenBank, EMBL, DDBJ, PIR, SWISS-PROT etc. (Kanehisa and Bork 2003), however, protecting bioinformatics tools and services as platform technology has been increased worldwide. Intellectual property (IP) protection of bioinformatics is inherently difficult. One of the main reasons is that it is multidisciplinary in nature. There are several ways in which bioinformatics IP can be protected (Harrison 2003) and patent is the most effective form of IP protection among them in which most of the components of bioinformatics innovations can be covered.

Corresponding Author:-TuhinChatterjee.
Address:-Ph.D. candidate, The West Bengal National University of Juridical Sciences (WB NUJS) and Indian Patent Attorney.

Material and Method:-

The main source of the present study is patent literatures collected from the United States Patent and Trademark Office (USPTO). Patent literatures and its analysis serve an important role in assessment of technology, its development and forecasting (Narin 1998, Oppenheim 2000). Bibliometric data can be used to assess and forecast progress in the technological field (Martin 1995, Watts 1997, Daim 2006). Further, within the available bibliometric data sources, patent data have been extensively used to gauge innovative activity in a particular area (Pavitt 1985, Narin 1996). Moreover, changes in patenting activity are commonly used to assess the development stage of various technological fields (Andersen 1999).

The patent search and evaluation framework of the present study integrates IPC classification (Table 2), bibliographic, citation (Garfield 1955), network(Albert 2002) and statistical analysis. The title-abstract or title-claim keyword search is widely used to collect relevant patent data (Yan 2009). The principle motivating factor of using IPC classification in the present patent search framework is that the accuracy of patent categorization technique adopted by the World Intellectual Property Organisation (Fall 2003) to classify a large variety of bioinformatics innovations (Table 2).

The IPC guided framework has been applied to bioinformatics and computational biology patent publications in the US Patent and Trademark Office (USPTO). The present study has considered all patent applicants and assignees i.e., private companies and R&Ds, public R&Ds, academic institutions and individual applicants focusing on the patenting activity during the year 2012-2016.

The Growth Of Bioinformatics And Computational Biology Research And Innovation:-

Bioinformatics tools and other advanced computational biology applications manipulate biological data in a variety of meaningful ways. Bioinformatics algorithms and software tools are generally involved in analyzing molecular biological data, particularly, DNA and protein sequences. However, its more advanced applications perform various complex tasks, e.g. mapping DNA and protein sequences, gene prediction (Wong 2016), modeling three-dimensional structure of proteins (Joyce et al. 2015) and drug discovery, modeling evolution (Liò and Goldman 1998) and cell division, simulating biomolecular interactions (Spiga, Degiacomi, and Dal Peraro 2014) etc.

Current trends in bioinformatics patenting:-

The patenting trend in bioinformatics area is not as aggressive as seen in other fields of molecular biology or genomics innovations. However, a steady growth has been noticed in 2012-2016 time-frame. A sharp increase in patent-filing has been observed in last couple of years besides a noticeable increase in the number of issued patents since early 2013 (Figure 1).

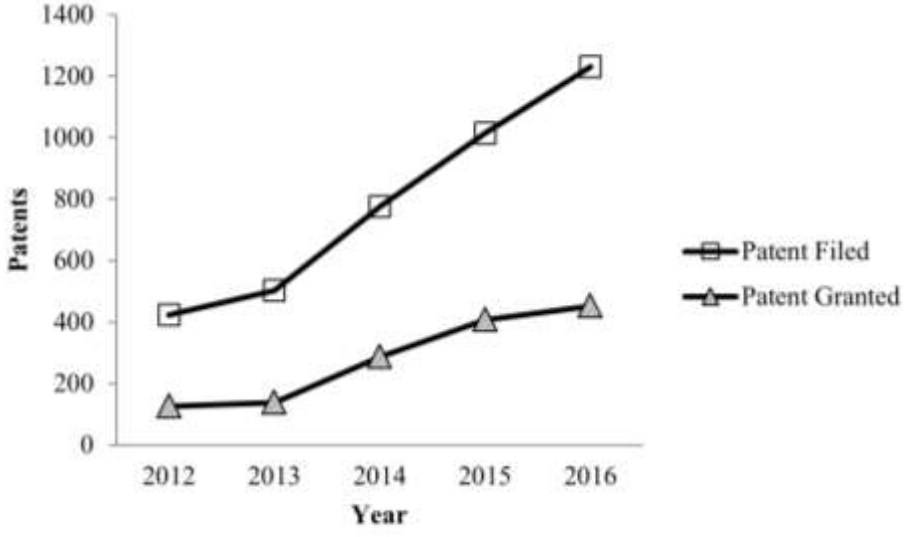

Figure 1. Annual distribution of patents in the area of bioinformatics and computational biology invention

Most of the patenting activities in last five years were mainly concentrated on three areas of bioinformatics and computational biology innovations. Machine learning and data mining secured the 1st position as the most successful area of technology with regard to number of patents granted during 2012-2016, followed by functional genomics or proteomics and sequence comparison involving nucleotides or amino acids (Figure 2). Although most of the patent-filing activities have been seen in the field of nucleotide or amino acid sequence comparison, one of the oldest areas of computational biology (Kanehisa and Bork 2003), however, this area has emerged as the third most successful area of technology in the list with regard to issued patents. Phylogeny or evolution has appeared as the least developed area with regard to bioinformatics innovation.

On the other hand, patenting activity in the field of programming tools or database system was average in last five years. Also, no major patenting activity has been seen in the data visualization area and placed at the 2nd last position in the list (Figure 2). However, the success rate in this area is highest in comparison to other areas. More than 55% out of the total patent applications filed in the data visualization area were finally granted by the USPTO.

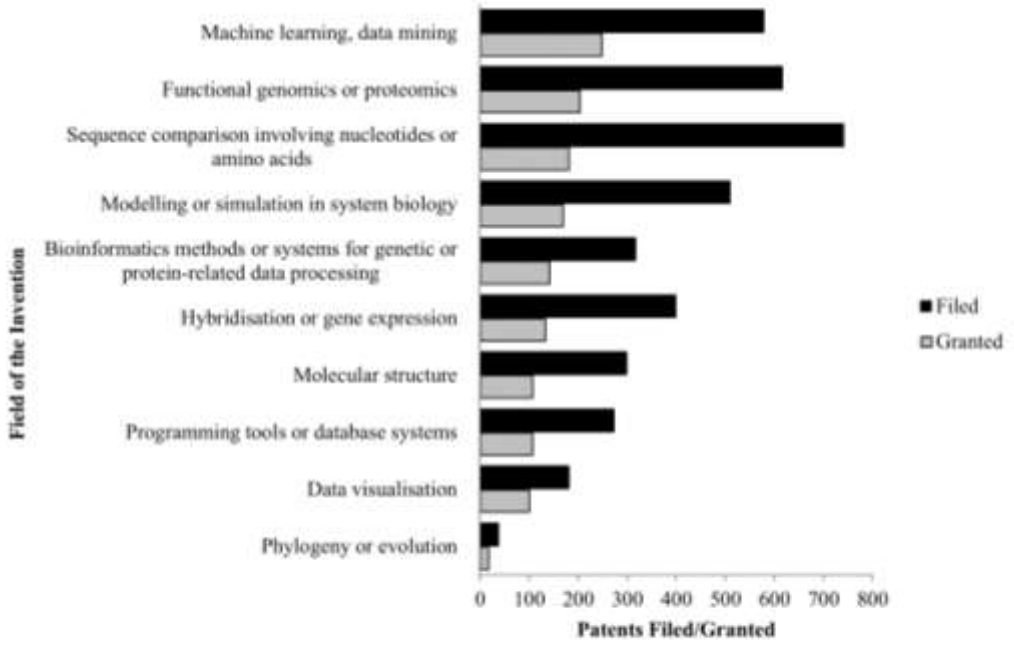

Figure 2. The Growth of research and innovation in different areas of bioinformatics and computational biology. The Patent search was conducted for ten IPC subclasses relating to bioinformatics and computational biology patent applications as categorized in International Patent Classification version v7.0e - 15.12.2016. To showcase the most successful areas of bioinformatics research and innovation in 2012-2016, each IPC subclasses (representing each technological areas) have been arranged according to their counts of issued patents.

Contribution of academic institutions and private companies towards the growth of bioinformatics research and innovation:-

The present patent search & analysis observed that privately owned companies played a key role to the development of bioinformatics and computational biology innovations (Figure 3). The analysis of patent documents revealed that a large pool of different patent applicants/assignees have been involved in the innovation activity. However, in the present study, only top twenty patent applicants/assignees have been listed (Figure 3) according to their patenting performance during 2012-2016. The International Business Machines Corporation (IBM) has significantly contributed to the growth of bioinformatics and computational biology and placed at the top position of the list. The most active area of its bioinformatics research and innovation was sequence comparison involving nucleotides and amino acids followed by machine learning and functional genomics.

On the other hand, academic institutions have also played some active entrepreneurial roles (Etzkowitz, Webster, and Healey 1998) to the development of bioinformatics and computational biology. Present analysis noted that only three academic institutions were majorly involved in the patenting activity (Figure 3). The University of California

has emerged as the major player followed by the Chinese University of Hong Kong and the University Leland Stanford Junior. Patenting activity of the present study revealed that the University of California has put most of its inventive effort in system biology followed by functional genomics and machine learning. The present analysis also observes that the research and inventive effort in the area of functional genomics and proteomics is common to all three academic institutions and top three privately owned companies. Whereas, both the top private company and the academic institution, i.e. IBM and the University of California, were extensively involved in machine learning and data mining innovations.

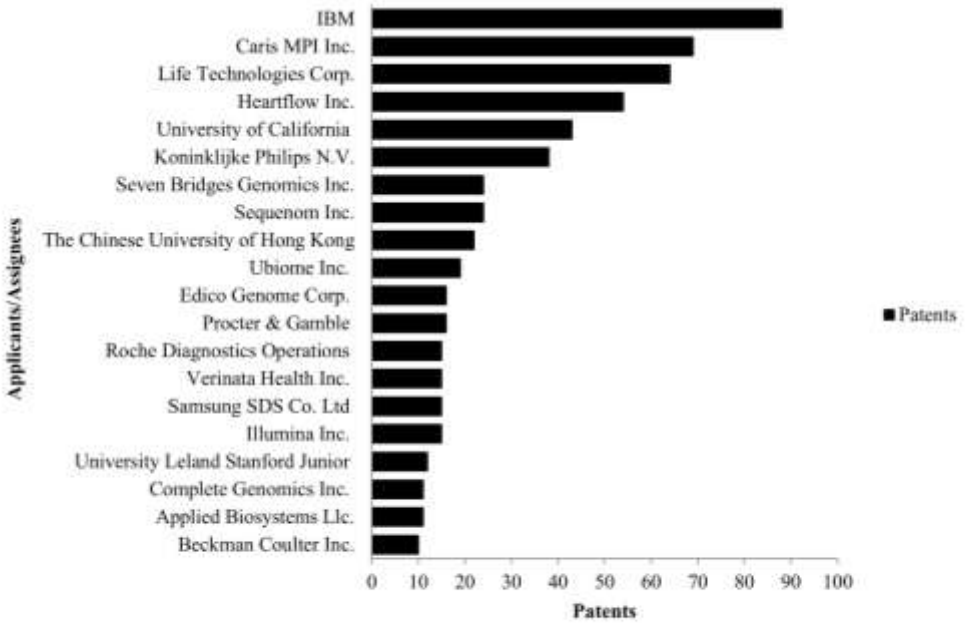

Figure 3. List of top twenty bioinformatics patent applicants/assignees. Twenty patent applicants/assignees have been arranged according to their patenting activities in the field of bioinformatics and computational biology research and innovation in 2012-2016.

Patenting Bioinformatics And Computational Biology Innovations:-

According to the American patent system, "any new and useful art, machine, manufacture, or composition of matter, or any new or useful improvement thereof" –are patentable. However, in relation to biotechnology innovations, the US patent regime has been witnessed several refinements in its patent-eligible subject matter jurisprudence starting from "anything under the sun made by man" (U.S. Supreme Court 1980) doctrine to *natural product/phenomena* exceptions applicable for genes or DNA sequences (SUPREME COURT OF THE UNITED STATES 2013).

Bioinformatics and computational biology are relatively new fields in the technology domain. It combines molecular biology, mathematics, statistics and computer technology as main components (Hagen 2000), (Luscombe, Greenbaum, and Gerstein 2001). Patenting bioinformatics inventions is relatively difficult not only because it is interdisciplinary in nature but also for its prophetic applications.

According to general practice of the United States Patent and Trademark Office (USPTO), innovations related to data *per se* are not patent-eligible (United States Court of Appeals 2014). The examination guidelines (United States Patent and Trademark Office 1996) for computer-implemented inventions made a clear distinction between "functional descriptive materials", e.g. data structures and computer programs which impart functionality when encoded on a computer-readable medium, and "non-functional descriptive material", e.g. music, literary works and a composition or mere arrangement of data which is not structurally and functionally interpreted to the medium but is merely carried by the medium. According to a report on the comparative study (Trilateral Patent Office 2002) conducted by trilateral patent offices[i], inventions that claim protein three-dimensional structural coordinates fall under the category of "information contents" and innovations related to these subject areas are not patent-eligible under §101 since they are nothing more than "mere presentations of information or abstract ideas".

Claiming strategies in Bioinformatics and Computational biology patents:-

Patent claims are considered as the most vital part of patent specification (Daniel Richards 2016), written description of invention, for which protection is sought before the patent granting authority. Claiming patterns of promising technological areas, e.g. gene expression, functional genomics or proteomics, modeling in system biology etc., have become highly complex with the increase in understanding of these subject areas.

The major claiming strategies in the field of bioinformatics and computational biology inventions are given in Table 1 to present a clear view about what innovators intend to claim at the strategy level and the breadth of protections those hypothetical claims encompass when translated into actual patent claims.

Table :-. Major bioinformatics claim types, hypothetical claiming patterns and their corresponding actual claims found in patent applications.

Claim Type	Hypothetical	Actual Patent	
		Publication No.	**Claims**
Type-I Computer model and data array claims	Example-1: A computer model of protein P generated with the atomic coordinates listed in Fig.1.	US5453937	Claim-1: A method in a computer system for modeling a three-dimensional structure of a model protein…… the method comprising the computer-implemented steps of:…….. the template protein has an amino acid aligned with the amino acid of the model protein, establishing the position of each backbone atom of the amino acid…….
	Example-2: A data array comprising the atomic coordinates of protein P as set forth in Fig.1 which, when acted upon by protein modeling algorithm, yields a representation of the 3-D structure of protein P.	US20060141600	Claim-6: A data array comprising the atomic coordinates of an Argonaute protein as set forth in Table 3. Claim-7: An electronic representation of a crystal structure of an Argonaute protein or a portion thereof.
Type-II Claims directed to database with coordinate data	Example-1: A database encoded with data comprising names and structures of compounds identified by a method of identifying compounds which can bind protein P by comparing the 3-D structure of candidate compounds with the 3-D molecular model shown in Fig.1 which comprises the steps of…….(1), (2), (3), (n).	US20060141600	Claim-26: A computer-readable storage medium encoded with the Argonaute atomic coordinates of claim 6.
Type-III Pharmacophore claims	Example-1: A pharmacophore having a spatial arrangement of atoms within a molecule defined by the following formula: Formula-1…….wherein (A) represents an electron donor atom…… and (C) represents a carbon atom.	US20080305041	Claim-2 : A compound that modulates PF4 activity comprising functional groups I, II, III, IV, VII, IX and X wherein the distance between the functional groups in three-dimensions is about: 2.25±0.05Å between group I and II; 6.03±1.37Å between groups I and III; 6.92±1.60 Å between groups I and IV;

			……; Wherein functional group I corresponds to the OD1 atom of the amino acid side chain Asp7, functional group II corresponds to the OD2 atom of the amino acid side chain Asp7,……in the PF4 sequence set forth in FIG. 1C (SEQ ID NO.1).. .
Type-IV 3D structure of protein	Example-1: An isolated and purified protein having the structure defined by the structural coordinates as shown in Fig.1.	US20070048853	Claim-1: An isolated protein having the structure defined by the structural coordinates as set forth in FIG. 4
Type-V Protein Crystal	Example-1: A crystalline form of protein P having unit cell dimensions of a=4.0nm, b= 7.8nm, and c= 11.0nm.	US6403330	Claim-1: A crystalline form of mammalian TRAP (tartrate-resistant and purple acid phosphate)……, and wherein the crystalline form of the mammalian TRAP has a crystal structure with atomic structural coordinates as given in Table 2, or with coordinates having a root mean square deviation therefrom, with respect to conserved backbone atoms of the listed amino acid sequence, of not more than 1.5 Å.
Type-VI Protein domain	Example-1: An isolated and purified molecule comprising a binding pocket of protein P defined by structural coordinates of amino acids residues 223, 224,295,343,366,370,378 and 384 according to Fig.1. Example-2: An isolated and purified polypeptide consisting of a portion of protein P starting at one of amino acids 214 to 218 and ending at one of amino acids 394 to 401 of protein P as set forth in SEQ ID NO:1.	US7700340	Claim-1: A crystal comprising an unphosphorylated Polo-Like Kinase 3 (PLK3) catalytic kinase domain polypeptide in complex with adenosine, wherein said PLK3 catalytic kinase domain polypeptide consists of amino acids 48-332 of SEQ ID NO: 1, and wherein said crystal is in space group C2 and has unit cell dimensions of a=145.95 Å, b=58.82 Å, c=47.10 Å, α=γ=90° and β=94.9°.
Type-VII *In silico* screening methods directed to a specific protein	Example-1: A method of identifying compounds that can bind to protein P, comprising the steps of: applying a 3-D molecular modeling algorithm to the atomic coordinates of protein P to determine the spatial coordinates of the binding pocket of protein P; and electronically screening……to identify compounds that can bind to protein P.	US19995856116	Claim-1: A method for identifying a potential inhibitor for an interleukin-1β converting enzyme, comprising the steps of: (a) Using a three-dimensional structure of said enzyme as defined by atomic coordinates of interleukin-1β converting enzyme according to FIG.5; (b) Employing said three-

	Example-2: A method of identifying compounds that bind to protein P by using the atomic coordinates of protein P shown in Fig.1 in a method of rational drug design.		dimensional structure to design or select said potential inhibitor; (c) ……..; (d) Contacting said potential inhibitor with said enzyme in the presence of a substrate to determine the ability of said potential inhibitor to inhibit said enzyme.
Type-VIII Compounds/leads generated by *in silico* method	Example-1: A compound which can bind to protein P generated by a process of comparing the 3-D structure of candidate compounds with the 3-D molecular model shown in Fig.1 which comprise the following steps: (1), (2), (3)……..(n).	US20026490588	Claim-5: A ligand compound for a biopolymer which is retrieved by a method according to claim 1. [Claim-1: A method of selecting one or more ligand compound to target biopolymer from a three-dimensional structure database, which comprises the step of: (i) the first step……..; (ii) the second step…….; (iii) the third step of estimating the most stable docking structure through structural optimization, wherein……..; (iv) The fourth step……..; (v) The fifth step of repeating the step (iii) and the step (iv) for all of the trial compounds.]
Source: The Trilateral Co-operation, *URL: www.**trilateral**.net/projects/biotechnology/WM4.pdf (*last visited on 20th October, 2016), USPTO Patent Application Full-Text and Image Database (AppFt), URL: http://appft.uspto.gov/netahtml/PTO/search-adv.html, USPTO Patent Full-Text and Image Database (PatFT). URL: http://patft.uspto.gov/netahtml/PTO/search-bool.html (last visited on 21st October, 2016).			

There are eight major categories of claiming patterns (Table 1) abundantly found in ten different fields (Group 2010) of bioinformatics patents (Table 2). Inventions directed to type I, II and III as described in Table-1, are rarely considered as patent-eligible under §101 since they either claim an array of data or computer model or database encoded with data comprising names and structure.

Table 2:- Different fields of bioinformatics inventions as classified in International Patent Classification (IPC), version v7.0e - 15.12.2016.[ii]

Sl. No.	Field of the Invention	IPC
1	Bioinformatics methods or systems for genetic or protein-related data processing in computational molecular biology including bioinformatics methods or systems where digital data processing is inherent or implicit, but not explicitly mentioned.	G06F 19/10
2	Modelling or simulation in system biology, e.g. probabilistic or dynamic models, gene-regulatory networks, protein interaction networks or metabolic networks.	G06F 19/12
3	Phylogeny or evolution, e.g. evolutionarily conserved regions determination or phylogenetic tree construction.	G06F 19/14
4	Molecular structure, e.g. structure alignment, structural or functional relations, protein	G06F 19/16

	folding, domain topologies, drug targeting using structure data, involving two-dimensional or three-dimensional structures.	
5	Functional genomics or proteomics, e.g. genotype-phenotype associations, linkage disequilibrium, population genetics, binding site identification, mutagenesis, genotyping or genome annotation, protein-protein interactions or protein-nucleic acid interactions.	G06F 19/18
6	Hybridisation or gene expression, e.g. microarrays, sequencing by hybridisation, normalisation, profiling, noise correction models, expression ratio estimation, probe design or probe optimization.	G06F 19/20
7	Sequence comparison involving nucleotides or amino acids, e.g. homology search, motif or Single-Nucleotide Polymorphism [SNP] discovery or sequence alignment.	G06F 19/22
8	Machine learning, data mining or biostatistics, e.g. pattern finding, knowledge discovery, rule extraction, correlation, clustering or classification.	G06F 19/24
9	Data visualisation, e.g. graphics generation, display of maps or networks or other visual representations.	G06F 19/26
10	Programming tools or database systems, e.g. ontologies, heterogeneous data integration, data warehousing or computing architectures.	G06F 19/28

Structural genomics and drug designing innovations:-

Computer assisted methods are extremely important in structural genomics (Goldsmith-Fischman and Honig 2003) and they are frequently used for Structure-Based-Drug-Design (SBDD). Newly evolving areas of biotechnology heavily rely on computer modeling and screening algorithms to data that describe a protein by its three-dimensional structure in order to design potential biopharmaceuticals. Protein three-dimensional structures represented by spatial arrangements of atoms or structural coordinate data are considered to have technical effect as long as they are used in an *in silico* or bioinformatics screening method to search for compounds.

There are three categories of inventions generally found in structural genomics patents. Inventions based on *information contents*, inventions directed to *in silico screening methods* that use structural information of proteins and inventions based on the end products resulting from *in silico* or bioinformatics predictive methods, e.g. compounds and pharmacophores(Langer and Hoffmann 2006).

Inventions directed to pharmacophore[iii], as exemplified in type-III, are considered as *mere* presentation of information or abstract ideas. Because, pharmacophores can neither be considered as a compound nor article of manufacture and they lack any immediate application to practical problems. Thus, such inventions are not patent-eligible (United States Patent and Trademark Office 1996) within the meaning of §101 of title 35 U.S.C.

A clear picture regarding the scope of protection of non-issued patent applications has been presented in Table-3.

Table 3:- Representative list of non-issued US patent applications related to computer model and 3D structural coordinate information of protein, database of protein 3D structures and co-ordinate data and pharmacophores.

Invention	Publication Number	Scope of Protection
Three-dimensional Structure Of DNA Recombination/repair Protein And Use Thereof.	US20070031849	DNA recombination/repair protein complex having a three-dimensional structure substantially defined by the atomic coordinates.
Electronic Database Of Enzyme Substrate And Enzyme Inhibitor Structures	US20020161599	An electronic database comprising a plurality of enzyme substrate structures
Three Dimensional Coordinates Of Melanocortin-4 Receptors	US20050171000	A G-protein-coupled receptor having three-dimensional structure obtained by computer-processing of atomic coordinates and a method
TripartitleRaftophilicStrutures And Their Use	US20080317767	A compound comprising a tripartite structure of C-B-A or

		C′-B′-A′
Drug Discovery Methods	US20110269732	Methods for assaying compounds for activity as Aurora kinase inhibitors and compounds having the features of the pharmacophore.
Pf4 Pharmacophores And Their Uses	US20080305041	A novel PF4 pharmacophore that is useful, inter alia, for identifying peptidomimetics and other compounds capable of modulating PF4 activity
Source: USPTO Patent Application Full-Text and Image Database (AppFt), URL: http://appft.uspto.gov/netahtml/PTO/search-adv.html(last visited on 23 October, 2016)		

The "isolated" and "purified" protein molecules (Type-IV, Table-1) having practical applicability but defined by their tertiary structural information are considered as patentable subject matter (United States Patent and Trademark Office 1996) under the US patents law.[iv] In summary, the isolated and purified proteins represented either in the form of standard amino acid sequence or in the form of their three-dimensional structures or combination thereof, are patentable as long as they have credible utility, even if that utility is computationally asserted (Trilateral Patent Office 2002).

On the other hand, the crystalline form of protein (Type-V, Table-1) is considered as "composition of matter". Inventions directed to similar subject area are patent-eligible on the condition that it meets other vital criteria of the US patents law, e.g. credible utility, novelty etc (United States Patent and Trademark Office 1996).

The specific region or domain of protein molecules, e.g. active sites or binding pockets etc., plays an important role in receptor-ligand interaction. Inventions directed to such type of protein domains (Type-VI, Table-1) represented either in the form of standard amino acid sequence information or three-dimensional coordinate data are considered as "composition of matter". According to USPTO's general patent practice, inventions relating to similar subject area are patent-eligible within the meaning of §101 of U.S.C.35. Table 4 shows the scope of protection of issued patent related to various fields of structural genomics innovations.

Table 4:- Representative list of granted US patents directed to protein 3D structures involved in computational methods and crystalline form of proteins represented by 3D structural coordinate data.

Invention	Patent Number	Scope of Protection
Ligand Identification And Matching Software Tools	US 8468001	Screening method for generating leads/ligand for treatment of a disease
Three-dimensional Structure Of Complement Receptor Type 2 And Uses Thereof	US 6820011	Method of structure-based identification of candidate compounds
Annotating Descriptions Of Chemical Compounds	US 8468002	A computer-implemented method for screening a chemical compound to identify a lead for treating a disease
Three Dimensional Structures And Models Of Fc Receptors And Uses Thereof	US 6675105	A model of a Fc receptor (FcR) protein represents a three-dimensional structure.
Quantitative, High-throughput Screening Method For Protein Stability	US 7148071	A method of detecting a binding event involving a protein with a ligand
Rat CathespinDipeptidyl Peptidase I (dppi) Crystal Structure And Its Uses	US 7736875	An isolated crystalline form of a dipeptidyl peptidase I-like protein
Three Dimensional Structure Of A Zap Tyrosine Protein Kinase Fragment And Modeling Methods	US 6251620	A method for determining three-dimensional structure of protein-ligand complex
Human Il-18 Crystal Structure	US 7253260	Human IL-18 protein in a crystalline form represented three-dimensional structural co-

		ordinates
Three Dimensional Coordinates Of Hptp beta	US 7769575	A computer-implemented method of identifying a drug candidate compound
Crystallized N-terminal Domain Of Influenza Virus Matrix Protein M1 And Method Of Determining And Using Same	US 6090609	A crystallized N-terminal domain of the M1 protein of influenza virus
Structure-based Identification Of Candidate Compounds Using Three Dimensional Structures And Models Of Fc Receptors	US 6675105	A method of structure-based identification of candidate compounds for binding to Fc receptor (FcR) proteins
Three-dimensional Structure Of Complement Receptor Type 2 And Uses Thereof	US 6820011	A method of structure-based identification of candidate compounds for binding to complement receptor type 2 (CR2) proteins or to a complex of CR2 and its ligand
*Source:*USPTO Patent Full-Text and Image Database (PatFT). URL: http://patft.uspto.gov/netahtml/PTO/search-bool.html (last visited on 28 October, 2016)		

Bioinformatics tools and database related innovations:-

Although computer programs are patentable[v](United States Patent and Trademark Office 1996) with appropriate limitations (USPTO 2007), however, data array and computer-readable storage medium encoded with atomic coordinates of protein are not patent-eligible under the present US patent regime. Moreover, the data array or the information storage medium encoded with protein three-dimensional structural coordinate data (*see* Type-I and II, Table 1) do not take part in the functional interaction with computer hardware or computing process. Because of these reasons, inventions related to those subject areas do not qualify as patent-eligible subject matter under §101.

Bioinformatics method and system related innovations:-

The *in silico* or bioinformatics screening methods and systems that search for compounds using three-dimensional structural information of proteins are patent-eligible under §101 as they generate useful, concrete and tangible results (U.S. Court of Appeals Federal Circuit 1998). The novelty, obviousness and utility assessment for inventions relating to those technological areas are comparatively rigorous. The utility of an *in silico* screening method depends on the utility of the candidate compound it generates (*see* step (d) of claim-1, patent no. US5856116, Type-VII, Table-1) (United States Patent and Trademark Office 1996). Moreover, the 'useful-result' criteria for such bioinformatics screening methods are distinct from that of the criteria of usual utility test. It is a further inquiry of the compound and its practical application always requires to be significantly functional in character (U.S. Court of Appeals Federal Circuit 1992). For example, a screening method is considered to be patent-eligible if its resulting compound is able to activate or inhibit certain key protein molecules to reduce blood pressure (United States Patent and Trademark Office 1996). The current patent analysis observes that reach-through claiming patterns are common in majority of the computational biology patent applications. However, it is likely that the USPTO is also aware about this fact and dealing effectively under the current US patent regime.

Conclusion:-

With the advent of various genomic projects, a range of high-throughput techniques have been developed in last couple of decades. These molecular biological techniques have been instrumental not only to understand the biological phenomena with greater details but also generated vast amount of raw data. Various bioinformatics tools and computational biology applications have been played significant roles since the Human Genome Project (HGP) era, and now they have become an integral part of almost every fields of biological research. The diversity of molecular biological data has also been increased significantly in past decades alongside the increase in volume of raw biological data. Advanced mechanisms have been evolved to handle this highly diversified data which further proved to be extremely useful in understanding the underlying complex mechanisms of biological system. Machine learning applications and bioinformatics data mining have emerged as the fastest growing fields of computational biology. Importance of data visualization tools and techniques in biological data analysis cannot be ignored, though patent filing activity in this area was not impressive in last five years. Patenting activity in the area of functional genomics and macromolecular data analysis still remained as the most focused areas for academic institutions and private companies. The present analysis shows that the patenting activity in the field of bioinformatics and computational biology encompasses a wide variety of subject areas which include modelling or simulation in system

biology, sequence comparison and discovery of single nucleotide polymorphisms (SNP), phylogeny, hybridization or gene expression, programming tools and database systems etc.
The challenges in patenting innovations have also been seen besides the success of these promising fields of computational biology. Although computer programs are patentable, however, data array or bioinformatics database systems, e.g. database containing atomic coordinate data of protein molecules, are not allowable subject matter under the current US patent regime. The overall patentability scenario in the field of structural genomics and drug discovery is encouraging. Pharmacophores are not acceptable at the USPTO since they do not qualify the doctrine of article of manufacture. However, chemical compounds generated with the aid of computational methods are allowable on the condition that the compounds have credible utility with regard to their technical abilities.

In summary, it can be said that the research and innovation scenario in the emerging fields of bioinformatics and computational biology was encouraging in last five years. Some of the vital reasons behind this success include logical and less cumbersome patent examination strategies of the USPTO. Hence, a persistent growth in these fields is expected in days to come till any new patentability norms are introduced in contrary to current patent practice.

Endnotes:-

[i] Trilateral Co-operation between EPO, JPO and USPTO was set up in 1983 with the objectives including improvement of the quality of patent examination process, improving quality of incoming applications, solving common problems related to IPR protection, harmonization in practice between three patent office etc.
*See www.**trilateral**.net/projects/biotechnology/WM4.pdf* (last accessed 20th October 2016)

[ii]International Patent Classification, International Patent Classification, *See* http://web2.wipo.int/classifications/ipc/ipcpub7?notion=scheme&version=20170101&symbol=G06F0019100000&menulang=en&lang=en&viewmode=p&fipcpc=no&showdeleted=yes&indexes=no&headings=yes¬es=yes&direction=o2n&initial=A&cwid=none&tree=no (last accessed Jan 1, 2017).
[iii]According to the first definition offered by Paul Ehrlich in the early 1900, pharmacophore is "a molecular framework that carries (*phoros*) the essential features responsible for a drug's (*pharmacon*) biological activity". According to another well accepted definition, pharmacophore is "a set of structural features in a molecule that is recognized at a receptor site and is responsible for that molecule's biological activity". –*See* PHARMACOPHORE PERCEPTION, DEVELOPMENT AND USE IN DRUG DESIGN**,** edited by Osman F. Guner. However, IUPAC offered more specific definition in 1998: *"A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or block) its biological response*". (*See* Langer and Hoffmann 2006).
[iv]*See* Comments of USPTO on trilateral comparative study of *"Protein 3D Structure Related Claims"; paragraph A-2 at page 65. URL: www.**trilateral**.net/projects/biotechnology/annex3w.pdf* (last visited on 22 October 2016)

[v] According to USPTO's examination guidelines on Computer related inventions, 1996, computer programs alongwith or having functional relationship with computer processing means are patent eligible. For example: functional data structure that is capable of increasing efficiency of computer processing is patent eligible. However, mere data arrangements recorded onto a computer storage medium (e.g. a CD) is considered as *mere* "information content" which does not have any functional corelation with computer processing means and thus are not patentable. – *See* Examination Guidelines for Computer-Related Inventions (1996), https://www.uspto.gov/web/offices/com/sol/og/con/files/cons093.htm. (last accessed 20th December 2016).

References:-

1. Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. 1990. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215 (3): 403–10. doi:10.1016/S0022-2836(05)80360-2.
2. Andersen, B. 1999. The hunt for S-shaped growth paths in technological innovation: A patent study. Journal of Evolutionary Economics, 9, 487–526.
3. Albert R, Barabasi AL (2002) Statistical mechanics of complex networks. Rev Mod Phys 74:47–97.
4. Barh, Debmalya., Vasudeo. Zambare, and Vasco. Azevedo. 2013. *Omics : Applications in Biomedical, Agricultural, and Environmental Sciences*. CRC Press/Taylor & Francis.
5. Beck, Hans Christian. 2010. "Mass Spectrometry in Epigenetic Research." In *Methods in Molecular Biology (Clifton, N.J.)*, edited by Rune Matthiesen, 593:263–82. Methods in Molecular Biology. Totowa, NJ: Humana Press. doi:10.1007/978-1-60327-194-3_13.
6. C.J. Fall 2003, Computer-Assisted Categorization of Patent Documents in the International Patent Classification, Proceedings of the International Chemical Information Conference, http://www.wipo.int/ipc/itos4ipc/ITSupport_and_download_area/Documentation/presentations/categ_wipo_icic03.pdf (last visited 20/01/2017).
7. Daniel Richards. 2016. "Supreme Court Upholds Patent Office's Method Of Claim Construction." *The Columbia Science and Technology Law Review*. http://stlr.org/2016/12/07/supreme-court-upholds-patent-offices-method-of-claim-construction/.
8. Daim, T. U., Rueda, G., Martin, H., & Gerdsri, P. 2006. Forecasting emerging technologies: Use of bibliometrics and patent analysis. Technological Forecasting and Social Change, 73, 981–1012
9. Etzkowitz, Henry, Andrew. Webster, and Peter Healey. 1998. *Capitalizing Knowledge : New Intersections of Industry and Academia*. *SUNY Series, Frontiers in Education*. State University of New York Press.
10. Goldsmith-Fischman, Sharon, and Barry Honig. 2003. "Structural Genomics: Computational Methods for Structure Analysis." *Protein Science : A Publication of the Protein Society* 12 (9). Wiley-Blackwell: 1813–21. doi:10.1110/ps.0242903.
11. Group, IPC Revision Working. 2010. "Special Union for the International Patent Classfication (Ipc Union) Ipc Revision Working Group." *#WIPO*. Geneva. http://www.wipo.int/edocs/mdocs/classifications/en/ipc_wg_23/ipc_wg_23_2.pdf.
12. Garfield, E 1955. Citation indexes for science: a new dimension in documentation through association of ideas. Science 122: 108–111.
13. Hagen, Joel B. 2000. "The Origins of Bioinformatics Joel." *Nature Reviews Genetics* 1 (3): 231–36. doi:10.1038/35042090.
14. Harrison, Robert. 2003. "Protecting Innovation in Bioinformatics and in-Silico Biology." *BioDrugs* 17 (4): 227–31. doi:10.2165/00063030-200317040-00001.
15. "Human Genome Project Completion: Frequently Asked Questions." 2010. *National Human Genome Research Institute (NHGRI)*. https://www.genome.gov/11006943/human-genome-project-completion-frequently-asked-questions/. (last accessed 2nd January 2017)
16. Joyce, A. P., C. Zhang, P. Bradley, and J. J. Havranek. 2015. "Structure-Based Modeling of Protein: DNA Specificity." *Briefings in Functional Genomics* 14 (1). Oxford University Press: 39–49. doi:10.1093/bfgp/elu044.
17. Kanehisa, Minoru, and Peer Bork. 2003. "Bioinformatics in the Post-Sequence Era." *Nature Genetics* 33 (3s): 305–10. doi:10.1038/ng1109.
18. Langer, Thierry., and Rémy D. Hoffmann. 2006. *Pharmacophores and Pharmacophore Searches*. *Pharmacophores and Pharmacophore Searches*. Wiley-VCH. doi:10.1002/3527609164.
19. Liò, Pietro, and Nick Goldman. 1998. "Models of Molecular Evolution and Phylogeny." *Genome Research* 8 (12). Cold Spring Harbor Laboratory Press: 1233–44. doi:10.1101/gr.8.12.1233.
20. Luscombe, N M, D Greenbaum, and M Gerstein. 2001. "What Is Bioinformatics? A Proposed Definition and Overview of the Field." *Methods of Information in Medicine* 40 (4): 346–58. doi:10.1053/j.ro.2009.03.010.
21. Narin F. 1998. Patents and publicly funded research: assessing the value of research in the chemical sciences. National Academy Press, Washington, pp 59–72.
22. Narin F. and Hamilton K. S. 1996. Bibliometric performance measures. Scientometrics, 36(3), 293–310.
23. Martin, B. R. 1995. Foresight in science and technology, Technology Analysis & Strategic Management, 7(2), 139–168.
24. Oppenheim C. 2000. Do patent citations count? The Web of Knowledge (Information Today Inc.), Medford.
25. Pavitt, K. 1985. Patent statistics as indicators of innovative activities: Possibilities and problems.

Scientometrics, 7(1–2), 77–99.
26. Spiga, Enrico, Matteo Thomas Degiacomi, and Matteo Dal Peraro. 2014. "New Strategies for Integrative Dynamic Modeling of Macromolecular Assembly." In *Advances in Protein Chemistry and Structural Biology*, 96:77–111. doi:10.1016/bs.apcsb.2014.06.008.
27. SUPREME COURT OF THE UNITED STATES. 2013. "Association for Molecular Pathology v. Myriad Genetics, No. 12-398 (569 U.S. ___ June 13, 2013)."
28. Trilateral Patent Office. 2002. "Trilateral Project WM4 Comparative Studies in New Technologies (Biotechnology , Business Methods , Etc .) Report on Comparative Study on Protein 3-Dimensional (3-D) Structure Related Claims United States Patent and Trademark Office." Vienna. http://www.trilateral.net/projects/biotechnology/WM4.pdf. (last accessed 12th November 2016)
29. U.S. Court of Appeals Federal Circuit. 1992. "Arrhythmia Research v. Corazonix." http://digital-law-online.info/cases/22PQ2D1033.htm. (last accessed 10th December, 2016)
30. ———. 1998. State Street Bank v. Signature Financial.
31. U.S. Supreme Court. 1980. Diamond v. Chakrabarty : 447 U.S. 303 (1980) :: U.S. Supreme Court.
32. United States Court of Appeals. 2014. Digitech Image Tech., LLC v. Electronics for Imaging, Inc.
33. United States Patent and Trademark Office. 1996. "Examination Guidelines for Computer Related Inventions." https://www.uspto.gov/web/offices/com/sol/og/con/files/cons093.htm.
34. USPTO. 2007. "Patent Subject Matter Eligibility [R-9]." https://www.uspto.gov/web/offices/pac/mpep/s2106.html (last accessed 15th November, 2016).
35. Watts, R. J., & Porter, A. L. 1997. Innovation forecasting, Technological Forecasting and Social Change, 56(1), XIV–XIV47.
36. Wong, Ka-Chun. 2016. *Computational Biology and Bioinformatics : Gene Regulation*. CRC Press. https://www.crcpress.com/Computational-Biology-and-Bioinformatics-Gene-Regulation/Wong/p/book/9781498724975 (last accessed 20th November, 2016).
37. Yan Dang et. al 2009. Trends in worldwide nanotechnology patent applications: 1991 to 2008, J Nanopart Res DOI 10.1007/s11051-009-9831-7.

Competing Interest: -

The author declares that he is the sole author and no other author has any kind of interest in the manuscript.