

Does biotechnology need a new database treaty?

Suresh Chandran

Biotechnology now is perceived to be capable of providing solutions to many of the ills afflicting mankind. The number of databases in biological sciences have mushroomed. What do databases mean for biotechnology?

MODERN biotechnology is perceived to be capable of providing solutions to many of the ills afflicting mankind. Today, the life scientist has the tools to hunt down a defective gene and replace it precisely with its normal version, to synthesize new therapeutic molecules, to re-engineer naturally-occurring molecules for greater efficacy or to modulate individual immune responses to acceptable performance levels. Recent advances in various fields of biology such as genomics, computational biology, structural and molecular biology, robotics, systematized ethno-pharmacobotany and combinatorial chemistry have catalysed a geometric progression in the corpus of information generated. The number of databases (dbs) in biological sciences has mushroomed. Nearly 210 million basepairs of sequence data is added to databases annually and the size of a database content is doubling every fourteen months. The number of solved protein structures is doubling every two years¹. As a Chief Executive Officer of a drug discovery firm remarked – ‘As much biological or drug discovery information is now generated in 18 months as in the entire preceding period of world history.’²

What do databases mean for biotechnology?

The evolution of virtually seamless disciplines that borrow heavily from each other, has become the characteristic of modern biotechnology. Data generated by one discipline invariably has to be cross-referenced to reinforce research findings in other disciplines too. The need for information sharing perhaps is highest in the area of modern biological sciences. An example would illustrate the interdependency. A scientist working on plant products refers to bibliographic and patent dbs to search for information available in the literature. He would refer to ecological dbs to learn about the plant/source of the product. He may refer to protein dbs to look for analogous proteins with similar functions that would aid him in characterizing the compound of his interest. He would refer to structural dbs to guide him in determining the molecular configuration of the product he isolates. He may

also refer to nucleic acid dbs to determine the genes encoding that particular product, etc.

Molecular biology is the interaction between gene sequences, amino acid sequences, their 3D structure and function. The biotechnologist can activate/inhibit a cell function by mimicking the 3D structure of either the signal molecule or the target receptor. The data available in structure and sequence dbs are critical tools for molecular drug design. Similarly, the science of genomics also hinges around sequence dbs. Comparison of nucleotide sequences between species is always informational because it gives hints towards the possible function of proteins across species. Once a functional motif has been discovered, very often the same building blocks are found in other species as well. The study of genomes of lower species helps in identifying their equivalents in human genomes too. The list of reference dbs accessed often by life scientists is many. These include agricultural dbs, ethnopharmacobotany, diagnostics, environment, food, medical, microbiological, user-group or professional research groups, instrumentation, government grants, publications and many more. However, structural and sequence information are of critical importance to biologists. Public dbs have become the major medium for publishing macromolecular structural data, sequencing and genome mapping data. The range of interests within biology often precludes construction of a single db which can satisfy all the potential demands. Specific requirements have led to the development of ancillary dbs or more derived dbs from general dbs. Interoperability amongst dbs assumes priority in such a context. Biotech dbs must necessarily have features such as timeliness (data should be accessible soon after submission), annotations, support of primary experimental data and integrability (cross-references to the same or related material in other dbs).

A point that needs repeated emphasis is that any structural or sequence information is of value only in context and not in isolation. Information required by a researcher in modern biology is scattered amongst hundreds of dbs and it is pointless to confine his access to one or two dbs. The evolution of search engines and sequence alignment software that can work in multiple db environments of varying formats is attributable to this overriding concern³. Also, it is not prudent to ration

Suresh Chandran is at the National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi 110 067, India.

information access, as serendipity or research breakthroughs cannot come from a quota system of information allocation.

Indian researchers discovering any new amino acid or nucleic acid sequence submit it to a db before publication to a journal. As researchers, we do contribute substantially to many public dbs. The growing number of genome research group and pharmaceutical company partnerships approximating to the tune of over \$1020 millions⁴ only underscores the importance of db-supported genomic research and therapeutic development. A number of private companies overseas have invested in developing their own Established Sequence Tags (EST) dbs. Is there a difference whether the information goes to a public or a private db? Three strategies of private db companies are in evidence: (a) Provide exclusive license for accessing sequence information [Human Genome Sciences (HGS) to SmithKline Beecham for \$125 m]; (b) Non-exclusive license for access to their db [Incyte to Pfizer, Upjohn, Hoechst, Johnson & Johnson, etc. for over \$100 m] and (c) dedication to public domain [Merck sponsoring Washington University, St Louis' sequencing efforts]. The motive behind the last arrangement is unclear. Perhaps Merck gets more value-added information by larger participation for a fraction of the cost borne by Incyte or HGS. Private db companies thus exploit their dbs by regulating access to them by contracts or trade secret agreements.

Existing protection mechanisms for databases

Section (3) of the Berne Convention and Article 10(2) of the TRIPs Agreement which reads that compilations of data or other material which by reason of selection/arrangement of the contents constitute an intellectual creation, shall be protected as such, circumscribes dbs. It would not protect the data itself and protection will not be prejudicial to copyright subsisting in the data or material itself. The Indian Copyright Act, 1957 also through Sections 2 (o) and 13 (a) brings databases within the ambit of literary works.

The philosophy of copyright itself is to protect the form of expression while concomitantly releasing the idea for wider exploitation. If creativity is explicit in the selection, arrangement and presentation of the contents of a db, the db is accorded protection. Data elements from dbs can be extracted and reutilized, honouring the same spirit of knowledge enrichment embodied in copyrights. It is obvious then that there are several dbs that would be out of the purview of copyright protection. Global events in recent years present evidence that this spirit of balancing financial reward for authors (read compilers of dbs) with societal gains has been abandoned in favour of protecting investments and commercial interests of the db creator.

Genesis of a policy shift

Copyright jurisprudence has witnessed a remarkable shift with the advance of digital technology. It has become a facile exercise today to shift around voluminous data, modify the contents of a db or to create a new db altogether from data elements scavenged from a host of other dbs making a mockery of the very tenets providing legitimacy to protection of dbs. Contents of a db need not be in close physical proximity either. The fear of misappropriation of data and their reuse in preparing competing dbs has influenced the present stance of db owners. Another major influence is the US Supreme Court decision in *Feist Publications vs Rural Telephones Service Company Inc*⁵. Justice O'Connor had to decide whether copyright protection can exist in compilations of data. The data in this case comprised of names and telephone numbers from a rural telephone directory. The US statute specifically says that compilations of data 'if selected, coordinated or arranged in such a way that the resulting work as a whole constitutes an original work of authorship, would be eligible for protection. The judge observed that the creativity involved in alphabetically arranging addresses as in the Rural Telephone case was very low and that method of organizing the data so commonplace as to exclude originality. Courts have subsequently focused on the manner in which the data is selected and arranged in a db and judged that copyright protection can exist if the selection and arrangement of data exhibited a minimum degree of creativity. The Feist judgement seems to legitimize selective extraction and reutilization of data, reformatting it into a different arrangement and creation of a new db therefrom, without the fear of having infringed copyright privileges. Db owners fear that the law unwittingly acquiesces with digital pirates to extract 'insubstantial portions' of dbs with minimal investment, creativity, or value addition, have pressurized their respective governments to incorporate stronger db protection provisions into the intellectual property framework. The costs involved in obtaining, selecting, verifying, arranging data and operating db electronically were emphasized to be a signal investment warranting protection. The responses to these concerns are reflected in the European Union Directive⁶ adopted by the Ministers of the European Council on 11 March, 1996, the pending US Bill H.R. 3531 'Database Investment and Intellectual Property Antipiracy Act' and in the draft mooted at the WIPO Diplomatic Conference for *sui generis* protection of dbs at Geneva in December 1996.

Contours of the proposed db treaty

The EU database directive

The definition of db to include electronic dbs, all collections of works, data, other material arranged in a systematic way and capable of being accessed by elec-

tronic or other means, appears too broad. Two distinct rights have been vested, one a copyright for those dbs that by virtue of selection, arrangement of the content constitute the author's own intellectual creation and, an extraction and/or reutilization right for dbs in respect of which there has been qualitatively and/or quantitatively substantial investment in either obtaining, verification or presentation of the contents. The treaty provides for member states to make 'fair use' exceptions. Even temporary reproduction is tantamount to copyright violation. What is critical, however, is that extraction and/or reutilization right gives the right-holder the right to control permanent or temporary transfer of all or substantial part of the contents of a db to another medium by any means or in any form and to control any form of making available to the public, all or substantial part of the contents of a db by the distribution of copies, by renting, by on-line transmission or by other forms of transmission. The right-holder has, however, the right to control the extraction or reutilization of insubstantial parts of a db and cannot impose such control by contract. Both qualitative and quantitative factors are considered to determine what is substantial. The fair use clause can be established under national law but would only be available for extraction right and not reutilization. The extraction and reutilization rights have a duration of fifteen years.

H.R. 3531 'Database Investment and Intellectual Property Antipiracy Act' (USA)

The US Bill was influenced by the EU Directive. However, it takes a few more steps to ensure a more exclusive monopoly on information. A creator of a db can prevent extraction of the contents of his db, and the use or reutilization of the data therein on the pretext of having invested substantially in the creation of that db. The bill recommends a 25 year protection unlike its European counterpart. Addition or modification of the db with 'substantial' investment further extends the protection by another term. The db-maker's rights are elaborate in that any act prejudicial to his actual or potential commercial interests, are forbidden. Thus, there is no scope for installing a 'fair use' provision in the proposed Bill. A second-comer even if he invests substantially cannot access the data freely from an existing db to build an independent, incrementally value added db. Making use of protection-defeating mechanisms or performing such services is also declared illegal.

WIPO database treaty draft

The WIPO draft has many provisions common to the EU Directive and the H.R. 3531 Bill. The *sui generis* system it advocates for db-creators is irrespective of whether

the db is protected by copyright. The protection provided must be without prejudice to any other right in or obligations with respect to a db or its contents including copyright. The draft also seeks to protect dbs that represent substantial investment defined both qualitatively and quantitatively. A significant feature is the absence of a fair use provision in this treaty. The maker of the db has the right to authorize the extraction or utilization of any portion of the db. It leaves to national legislations to determine the protection granted to government-created dbs. Another significant departure is the duration of the protection afforded to db makers. 25 or 15 years of protection was initially mooted. However, if any substantial change is made to the db that involves substantive investment, the term of protection starts again for the changed db from the time of such investment. Another difference as compared to the EU Directive is that this does not demand reciprocity unlike the latter. The European directive would protect db contents of foreign dbs only if those nations have adopted equivalent laws to the European db law. The WIPO draft also allows db developers to enter into contracts with users about giving up their rights to take insubstantial parts of the db contents. Understandably, this treaty has not received the welcome the Americans and Europeans hoped it would. The WIPO draft treaty on intellectual property in respect of databases was deferred at the November 1996 Diplomatic Conference in Geneva.

Opening Pandora's box?

Going by the trends in which db protection law is evolving, developing countries have reasons to worry. They normally are db-users and are information providers to dbs worldwide. The WIPO Draft transfers all the rights to the investor in a db, thereby precluding data providers to the db from accessing the data they generated after it is put into a db, which is disconcerting. Examination of the pros and cons of adopting such a protection strategy appears prudent.

Firstly, the definition of what constitutes a db itself is very broad. A 'collection of independent works, data or other material arranged in a systematic way and capable of being individually accessed by electronic or other means' brings a wide range of database services into the protected category. Non-electronic dbs thus come into the ambit. A published document already in prior art can be digitized and converted into an electronic db and protected. Some of the dbs containing structures of small molecules have been built this way and have become unaffordable to publicly-funded researchers. It may also be argued that search engines or sequence alignment software also are a part of the db and may get protected from the now freely accessing research community. Many of such tools provide for inter-operability

Box 1. A representative list of databases and their contents

Area	Database	Remarks
Protein sequence collection	OWL, SWISS-PROT, PIR	Public domain
Protein sequence motif	PROSITE, PDB	Public domain
Protein structural alignments	FSSP	Public domain
Predicted protein structures	HSSP	Public domain
Proteins grouped for structural similarities	SCOP	Public
Enzymes	ENZYME	Public
Restriction enzymes	REBASE	Public
Nucleotide sequences	GenBank, EBI, DDBJ	Public
Expressed sequence tags	dbEST	Public, Several private companies have dbs.
Promoters	TFD	Public domain
Genomic maps	GDB	Public domain
Genes (immunological)	Kabat	Public domain
Human genomics	Integrated Genomic Database	Public domain
Drosophila	FlyBase	Public domain
Mouse	MGD	Public domain
Molecular biology software	BioCatalog	Public domain
Patents	DERWENT world patent index	Private
	INPADOC	Private
Fellowships in bioinformatics	TechResources	Public
Agriculture	AGRICOLA, CAB Abstracts	Private
Pharmaceuticals	Pharmaceutical news index	Private
Biotechnology	Derwent biotech, Abstracts	Private
	Current biotechnology abstracts	Private

between dbs and enable searching and locating similar information in otherwise incompatible dbs. Depriving the scientific community of such essential tools would handicap research capabilities.

The philosophy of this proposed treaty is undoubtedly to protect investment and not to further knowledge. While one concedes the point that db creation and maintenance is a costly exercise, it is hard to relegate the investment in time, money and skills made by research teams generating the sequence and structural information that enters the dbs. The essence of copyright is to reward intellectual efforts and simultaneously release the knowledge for public gain. This balance has been upset in favour of financial investment of the db owner at the cost of societal benefit.

While the db owner controls the right to authorize the extraction and reutilization of 'a substantial part' of the contents of his db, the extraction of a 'substantial' part is defined elsewhere as any portion of db, including an accumulation of small portions, i.e. of qualitative and quantitative significance to the value of the db. This practically precludes db users from accessing any portion of a db without express permission of the owner. In sequence dbs where often the function of these sequences is yet undeciphered, it is unreasonable to expect categorization of the db content into what is 'significant and what is not'. It would thus be left to db owners to dictate what is qualitatively substantial in their dbs, and lawful users will be deprived even of, what according to them (users) comprises 'insubstantial parts'. Most dbs in biotechnology are de-

veloped by incremental increases. The WIPO Draft effectively kills the creation of such dbs.

The research community would be affected the most by the absence of 'fair use' provisions in the WIPO draft. Even when a palliative is suggested in the form of national legislation for determining protection to be granted to dbs created by Government initiatives, the earlier provisions defining 'substantial', and the general language of the document leads one to believe that conflict of interest with db owners may be discouraged. It must be noted that even where the EU Directive makes a fair use exception, namely, substantial extraction for non-commercial purposes (research, review, education, etc.) it is allowed only for non-electronic databases. The modifications even outlaw temporary reproduction, which means that leave alone downloading, even onscreen analysis will not meet approval.

The requirement of obtaining permission of the db-owners to extract more than 'insubstantial' data will be counterproductive to the development of the db industry as a second-comer would be effectively discouraged from selecting and placing together data from different dbs to create a new db. Such a service is essential in the field of biotechnology. Were a researcher to collect material on his field of interest, say a colon cancer gene, he would be prevented from creating one by denying him free access to genome/sequence db, protein and structural dbs of the gene product, etc. As mentioned, the nature of biological research itself makes it farcical to restrict a researcher's access to one or two dbs. Ser-

endipity, which heralds most of the scientific breakthroughs, cannot be expected in an environment where information is rationed. In paying for accessing dbs, research costs would escalate, knowledge generation will be impeded and submission of data to public dbs will eventually diminish. The greater the access to different dbs, the better would be the quality of new knowledge generated. It is also critical that more scientists can access a given db, because it would be a very cost-effective way of checking quality of the data in that db. If the data is not verified, one could end up paying for erroneous data.

The duration of protection mooted is another contentious issue. It is immaterial now if a 15 or 25 years' protection is implemented because another provision effectively provides db owners perpetual protection. If any substantial change is made to the content of a db, (as it would be with many dbs) that constitute a substantive investment, the 15 or 25-year clock is set again. This can go on *ad infinitum*. Who would judge that substantive investment was made in updating and/or adding new data, are questions unresolved. Where would litigation lead if an evaluation mechanism was indeed created?

A major loophole the treaty would open up is in the area of patent law. Many countries have listed therapeutic, surgical and diagnostic inventions as not patent eligible. Even the TRIPs document allows for such exceptions. In case a therapeutic protein or the sequence encoding a genetic defect is discovered and is not granted a patent in that country, the inventor could easily submit the sequence into a db and seek protection, thereby preventing research on that protein by other research groups. Thus sequences that are not patentable, may achieve the objective *via* copyright protection and thus stifle competition.

Today when a public R&D researcher publishes a gene sequence of a therapeutically or commercially important protein, it is possible that private EST db owners may scan their own dbs for full length ESTs that show similarity (and naturally are undisclosed) to the sequence and file patent applications on those sequences. Such a strategy would undermine public interest, as by patenting all related genes, the right of a gene-finder will be hedged in⁷.

The draft provides for nations to legislate on the fate of Government-created dbs. However, one must realize that there has been and effectively continues, massive anonymous ftp downloads of sequence or structural data by companies⁸. Most of Government-generated data is already in private hands!

It thus becomes evident that legitimate users would suffer for the wrong-doings of data pirates. Modern biotechnology which thrives on the interaction between complementary disciplines would suffer a setback with the information censorship brought about by this form of

protection. While the information free-riders must be stopped, (especially if they provide nothing new but exploit the pirated contents of a db without any financial investment) a distinction necessarily must be made between them and fair-users of the data. Else, it would be throwing away the baby with the bath water. In the interest of the research community, a rethink is mandated.

In quest of an alternative

The WIPO Draft fuses the idea-expression dichotomy. Feedback from biological scientists has only reinforced the view that the proposed database treaty is not in the interest of the sciences. It is customary to submit sequence or structural data to international public dbs (some journals specify submission to a given db). Many Indian researchers have also contributed sequence matching programmes to public domain dbs. It would be unfair, even scientifically unethical to relinquish openness and information-sharing for the benefit of a few private individuals. The government will have to restore the balance sought out in copyright law where societal benefit was not subjugated to private reward. India must oppose the move to ratify the treaty in its present form. What can be the palliative measures?

Introduction of the 'fair-use' exception as in the EU Directive, with one variation – the permission to access electronic dbs, is necessary. Db-users amongst the research community may be asked to sign a non-distribution agreement for the contents of the accessed dbs as is done in many standard material transfer agreements, with liability provisions. If concessional access rates can be worked out for the research community engaged in non-commercial research, they should be pursued. Raw data must not be rationed. If necessary, an organization such as the CCC may be established to work out concessional access rates for academic and non-commercial R&D use of dbs.

The duration of protection must be reduced to 5 or 7 years as this time is sufficient to generate a research lead. The idea behind the proposed treaty is to protect investment. If db-makers have to be recompensed for their investment, one method is to give them a lead period in the niche market for say 5 years, after which competitors may be allowed to grow and compete. This works like the patent system. Perpetuation of information monopoly by continuous modifications will be self-defeating as this can bring about long-term damage to R&D and innovative capabilities. It would also harm the db industry a great deal. Even the quality of unchallenged data may be poor.

Formal registration may be desirable as this can obviate future disputes. A board akin to the Copyright Board may adjudicate the quality and quantum of modification

and approve of any further extension of protection period. Protection must be provided only for the additional data appended and not renewed for the entire db whenever a 'substantive' addition/modification is claimed.

What counter-strategies can India make? Our approach must stem from our strengths and the knowledge of our weaknesses. You need strong laws only if you have a strong industry. There are not many indigenous biotech dbs we have at present. Perhaps in the coming years we may create some ecological/traditional medicine dbs. Isn't it prudent to wait till we have something to protect instead of jumping to ratify the treaty? Most of the data today is generated with Government funding. Can this be stopped from getting into private dbs? Aren't the db-makers themselves free-riders when they ftp and download data from public domain dbs? If the scientific community in public R&D stands up in unison to resist privatization of dbs and their economic returns, the damage can be undone. Ideally dbs must be restricted to copyright protection. Protection for dbs kept in secret should be resisted. There are adequate provisions such as anti-trust laws and trade-secrets to address those concerns and new legislation is not called for.

Public domain dbs should be strengthened and researchers in public R&D organizations encouraged only to contribute to public dbs. If the quality of public dbs improves and if a researcher finds that he can access the same data contained in a privately-owned dbs in a public db, he would prefer the latter, as it would not hamper him with obligations. This way one would be able to offset partially the exploitative hold of the private db owner.

The last is a more presumptuous one. Sequence and structural dbs may be alienated from the purview of the db treaty. It has been traditionally argued that DNA sequences, etc. are not copyrightable hence they already existed in nature and that their author cannot be defined. It also stands to reason, however, that the intellectual input of the research scientist made it possible to deci-

pher the information without whose effort, the very existence of the sequence would have remained unknown. Parallels may be drawn between the genetic code and the binary code used in computer programs. A string of 0s and 1s can be copyrighted. By extension of the same logic, genetic code and the sequence should be eligible for copyright protection.

One could, perhaps, even argue that sequences are not even eligible for copyright protection. Would a string of alphabets qualify for copyright protection? Amino acid or nucleic acid sequences are represented by alphabets akin to the binary code. Some case laws of the software industry may be used in argument. The alphabets and the sequence in which they appear, together give meaning to the information. In other words, to be functional, it is essential that a sequence is represented precisely by a particular selection of alphabets in a particular order. Reading the judgements of the *Kenrick vs Lawrence*, 1890 which stated that if there is only one way of expressing an idea, then that way is not the subject of copyright and that of *Keaton in Lotus Development Corporation vs Paperback Software International*, which states that where the expression of an idea is inseparable from its function, it forms part of the idea and is not entitled to copyright protection. Exploring the legal nuances of these judgements may perhaps provide a solution to deliver structural and sequence dbs to their rightful owners – the global research community.

1. Marchington, T., *Biotechnol.*, 1995, 13, 239–242.
2. Marchington, T., *Biotechnol.*, 1995, 13, 239–242.
3. Bains, W., *TIBTECH*, 1996, 14, 312–317.
4. Freidrich, G. A., *Nature Biotechnol.*, 1996, 14, 1234–1238.
5. USPQ2d 1275, US Supreme Court in *Feist Publications US Rural Telephones Service Company Inc.*, 1991.
6. CRNR/DC/6, Basic Proposal for the substantive provisions of the treaty on intellectual property in respect of databases to be considered by the diplomatic conference at Geneva, Dec. 1996.
7. Eisenberg, R. S., *TIBTECH*, 1996, 14, 302–307.
8. Samuelson, P., *Commun. ACM*, 1996.