The Biotech-Bioinfo Interface in the Context of Education and Growth of the Biotechnology Industry in India Today

[written in Jan 2006]

Dr. Kshitish K Acharya *(kshitish@ibab.ac.in)*
Institute of Bioinformatics and Applied Biotechnology (**IBAB**), Bangalore
Originally written for *Biovistas 2006, Vol 1 (1): 7-13* -

Abstract: Biotechnology and bioinformatics have a lot in common. Apart from the fact that they have been incorrectly equated to the 'IT' in career aspects, the knowledge in either of these biological subjects can enhance the learning or contributions in the other. A careful attention to the common aspects of biotechnology and bioinformatics (the biotech-bioinfo interface) could improve the value of curricula in both the areas. This article identifies the curricular components which would be important not only for biotechnology and bioinformatics courses but also to other life science programs. A few examples of personal research projects have been discussed to illustrate the need to include/strengthen the biotech-bioinfo interface in our educational programs.

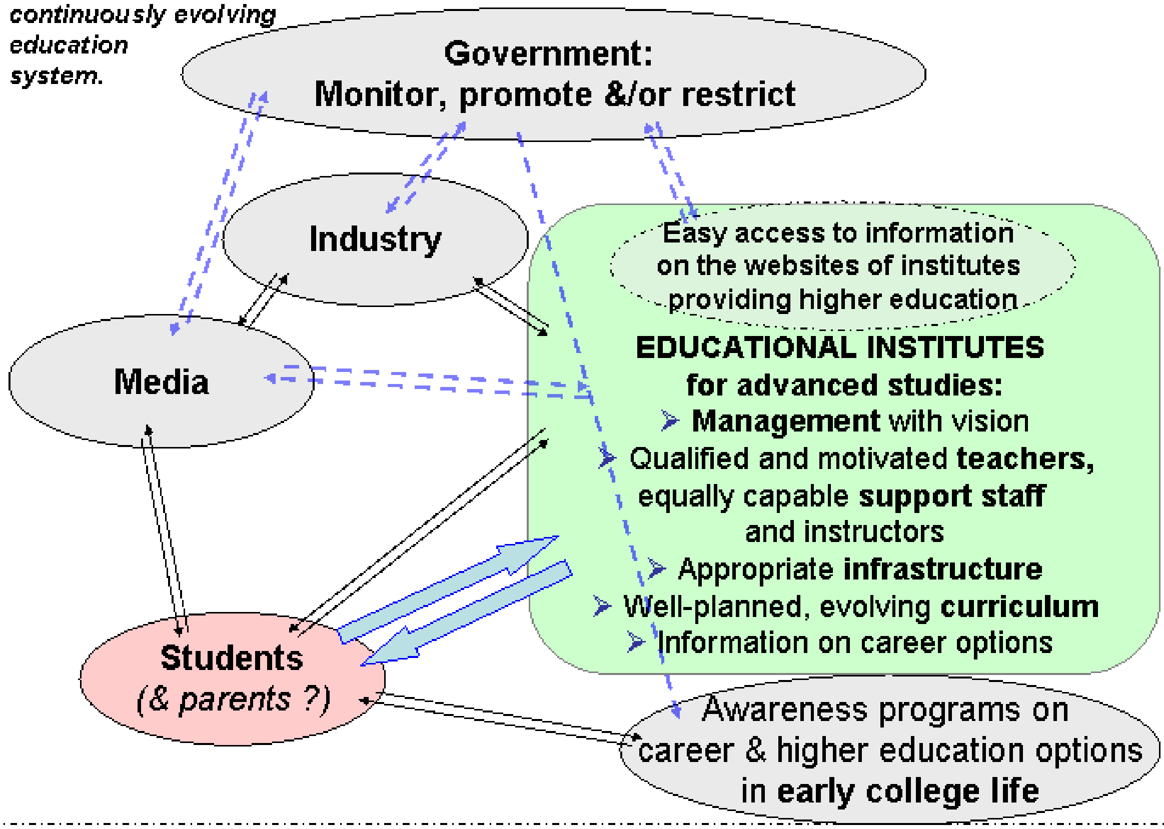
Standards in life science education have to be improved and we can start by evaluating one aspect at a time. A decade back, several colleges taught mainly botany and zoology as life science subjects without caring much for the requirements of biologists in the country. Even though the choice of subjects has broadened now, things do not seem to have significantly changed in terms of student employability. One of the reasons may be that we, the people in education system (see Fig. 1), haven't learned to keep pace with the changing times and demands. In fact, with less-informed students and parents, the situation seems to have worsened recently. Many colleges have opened post-graduate, and even graduate, programs in new subjects without proper teaching capacities. The lacunae include teachers' level of expertise and/or the infrastructure. This is not good for the growing Indian biotech-industry, which happens to be very diverse. The success of this sector depends a lot on the availability of top quality human resources. Hence, there is an urgent need to focus on specific aspects of educational programs and start improvising each of them. Certain aspects of life science curricula that involve parts of the following subjects will be discussed here.

1. Biotechnology: Broadly, the 'applied' and 'technical' aspects of science of microbial and bio-molecular processes.
2. Bioinformatics: In brief, the science of managing and/or studying the data in biology, particularly concerning the biomolecules, using computational, statistical and/or mathematical skills.

Why these subjects have gained prominence? These two subjects have gained importance in the recent decades not only from industrial point of view but also from basic research perspective. Biological research can be performed at different levels of cellular organization of life forms. But, molecular details are inevitable for a complete understanding of most of the biological phenomena. Molecular research also has a huge potential for therapeutic, diagnostic or other industrial applications. Since biotechnology and bioinformatics deal with biomolecules, they have become crucial subjects in today's context.

However, *not every student of life sciences need to study everything in biotechnology and bioinformatics; specific components in these areas need to be identified that can be and should be studied.*

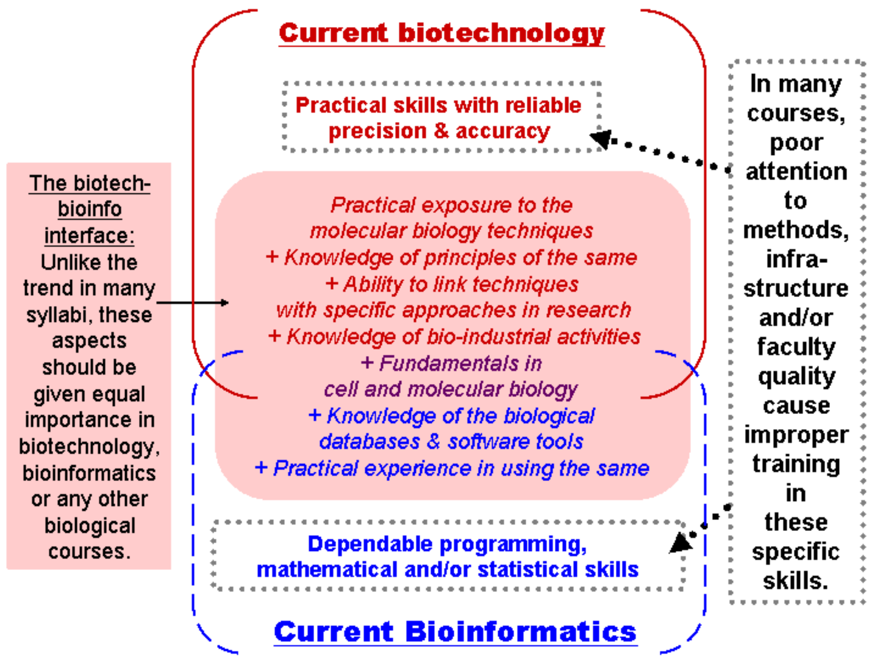
Figure 1: ***Ideal, actively interacting sectors in higher education, in general. Proper interactions would promote awareness and career planning in students and, ensure better and continuously evolving education system.***



Biotechnology integrates many subjects, including bioinformatics: As many recognize (e.g., *Vermette et al 2003*, *Rao 2005*) biotechnology is a multi-disciplinary subject. Proper integration of engineering and biological concepts is one of the important improvements required and this has not been easy even in some of the developed western countries (*Vermette et al, 2003*). It is a good sign that syllabi are being updated/modified (see *suggested syllabi from DBT*) in India.

Biotechnology engineering courses have been mainly focusing on the technical aspects, like fermentation, from an engineering perspective. But neglecting the fundamentals of Cell and Molecular Biology (CMB) seems to have reduced the value of such technical programs. In contrast, many of the post-graduate courses have almost failed to be relevant to the industry as they have largely ignored the industrial processes and requirements. There can be several improvements in both engineering and MSc biotech-programs. One of the areas where both types of courses need to improvise is the training in molecular biology techniques. In fact, the bio-technical training at the initial level, theoretical CMB and certain aspects of bioinformatics form the biotech-bioinfo interface (see figure 2) that should be part of any life science study today.

Fig. 2. Bioinformatics and biotechnology training have many common components.



The biotechnical training for beginners is best started with experiments in cell and molecular biology, biochemistry and microbiology. There is an unhealthy trend to cover as many techniques as possible and students also seem to have a big misconception that it is best 'touch upon' majority of the techniques. Unfortunately, not enough attention is paid to the following aspects of training: a) Safety, b) Calculations required to make solutions, c) the right way of doing basic things like pipeting, weighing, transferring, mixing etc, d) awareness of equipment and reagent details, e) record keeping, f) potential variation introducing factors (it is very important for students to repeat a few selected experiments, rather than doing more variety of experiments), g) designing experiments, h) observing during different steps of protocol and i) interpreting the results. These are best stressed at the graduate level of education. At IBAB, we have begun a post-graduate course (*LCBT-IBAB*; 6 months or 1 year) for laboratory biotechnical training and we end up spending a significant amount of the first 2 months for just filling in gaps in these basic aspects of training.

Bioinformatics is also multidisciplinary in nature: Like biotechnology, bioinformatics is an area where people have been working since several decades. In the recent past, however, bioinformatics has emerged as a major subject, particularly in post-graduate studies (*Ranganathan 2005*). The attention of industry and the media to the bioinformatics was mainly due to the launch and progress in the genome projects. Though the 'hype' seems to be fading, bioinformatics has remained an essential and growing subject due to
(a) increased research and industrial activities in the area of bioinformatics, and
(b) accumulation of enormous amount of data on biomolecules, particularly the sequence, structure, location and relative abundance of the macromolecules, and their interactions.

As with many other subjects, though there are lots of degree/diploma holders in bioinformatics, the quality of bioinformatics courses at many educational organizations in India have not been of good standards. The launch of the national quality assessment test (*BINC 2005*) for post graduates, by DBT, will hopefully prove beneficial for the quality of bioinformatics in India.

Bioinformaticians have roles to play in the biotechnology and pharmacology companies as well as other service-providing organizations in the health/life sciences area. Hence, even if the number of pure bioinformatics companies may not increase considerably in India or abroad in near future, bioinformaticians will most probably continue to be in demand. However, the type of bioinformaticians required will be different depending on the type of the organization and/or responsibilities of the job.

Bioinformatics can be identified into 4 main types:
It is convenient to consider a bioinformatician as someone with (or without) a specialization of the following types:

A. User-end bioinformatics. This simplest type of bioinformatics involves use of the existing biological databases and software programs. The activities in this category of bioinformatics require a general understanding of the process of research in life sciences and acquaintance with at least the most commonly used databases and tools. It is important to note that the use of such bioinformatics tools or databases can only be of limited value without the basic understanding of the algorithm/structure behind them.

Many articles have stressed the need of incorporating bioinformatics in different subjects at different levels (Altman 1998, Campbell 2002, Centeno et al 2002, Eberhardt et al 2003, Honts 2003, Pevzer 2004). Specifically, it is the user-end bioinformatics which biotechnologists, bioinformaticians and other biologists should be versatile in today's context.

The user-end bioinformatics is very basic and simple. However, this might have been taken too lightly so far in our teaching programs. Also, there seems to be a shortage of appropriate teaching faculty and facilities, and it is not an active component in most life-science courses. Even in the places where this section of bioinformatics is taught, students seem to lack practical experience with the databases and software tools. These disadvantages can be overcome by incorporating carefully designed assignments/exercises in the curriculum and teacher-training programs. With the number of databases and bio-softwares growing everyday, obviously we can't teach many of them to students. However, some of the major tools and databases available in public websites like the NCBI, EMBL, DDBJ, ExPasy, etc can be studied in detail.

B. Creating and managing databases as well as software tools. These areas would require good computational/programming skills. Both types of activities can be performed by a group with pure biologists and programmers. However, they are best carried out by a bioinformatician who actually specializes in this type of bioinformatics, or a team that has

(a) life scientist(s) who know basic concepts in algorithm development and programming
(b) expert programmer(s) familiar with fundamentals of biology.

.

Creating and managing databases also requires thorough biological knowledge. The expertise required in understanding experimentation and the research process are highest particularly when the database creation involves data curation. In fact, sometimes biologists with bio-technical training or experience in experimental research would be preferred by bioinformatics companies for the jobs that need extensive data curation. This is because data curation involves making 'sense' of the data and/or assessing the reliability of the same.

C. Developing new algorithms or computational methods for bioinformatics. These areas would need high level of mathematical and/or statistical skills along with the fundamental knowledge of CMB and other aspects in biology.

D. Biological research using the knowledge and skills of computational science and one or more of the following subjects: mathematics, statistics, physics and chemistry. This category includes computational structural biology, computational/mathematical modelling and other research involving data analysis of various types.

Many regard only the 'B, C and D' sections as true/core bioinformatics.

Bioinformatics-courses might usually begin with CMB and the user-end bioinformatics. However, it is a misconception that bioinformatics suits biologists better. In fact, biologists and non-biologists perform equally well and, sometimes, non-biologists can do better than biologists in bioinformatics learning and placements (experiences at IBAB and *Lyon et al 2004*). It is sad that many students of statistics, physics, mathematics and chemistry subjects are not aware of the opportunities in bioinformatics and/or good education centres where these subjects are best integrated.

In addition to knowing the user-end bioinformatics it will benefit biotechnologists if they have at least a broad understanding of these core bioinformatics types, apart from the knowledge and skills of user-end bioinformatics. Similarly, bioinformatics of any type would be more complete if the students know the basic biotechniques and understand how the data/information is obtained in molecular biology. In fact, the biotech-bioinfo interface should be part of the syllabi in graduation/post graduation programs in life sciences: bioinformatics, microbiology, biochemistry, zoology, botany, biotechnology, and even more applied subjects like agriculture, veterinary science, medicine and pharmacology.

The biotech-bioinfo interface is inevitable in research today: The following points and examples illustrate the importance of the biotech-bioinfo interface in certain types of research projects.

1. With the large amount of information accumulated on biomolecules, it has become almost impossible to initiate many biological research projects without the analysis of the existing data/information. Further, once a research project identifies/discovers a gene/protein sequence, it is difficult to draw valid conclusions without bioinformatics analysis. This is illustrated by the following research work: A study that attempted to elucidate transcriptional regulation of a specific gene, expressed in the post-meiotic male germ cells. This research project involved characterizing

    (a) promoter of the gene and
    (b) 3 putative transcription factors that bind to a critical region of this promoter (*Acharya et al, 2006*).

A small region was identified as the potential critical region of the promoter by sequence comparison of the genomes, in the vicinity of the specific gene of interest, of 3 species. BLASTn, alignment of two nucleotide sequences, PubMed, map viewer, genome sequences of NCBI or other similar tools and databases can be used for such comparisons. Wet-lab techniques like electrophoretic mobility shift assay, south western assay, two dimensional electrophoresis and expression library screening were then used. Eventually three proteins that bind to the critical region of the promoter were cloned and sequenced. Sequence analysis, with the use of different tools from NCBI and ExPasy, could help to:

i) identify other reported isoforms of the cloned sequences (proteins),
ii) discover that 2 of the cloned sequences code for proteins with 2 RNA-binding motifs each,
iii) conclude that none of the 3 discovered proteins or their isoforms were not earlier reported in mouse testis,
iv) identify the genomic location of the sequences discovered, and
v) postulate potential molecular interactions of the discovered proteins and their functions.

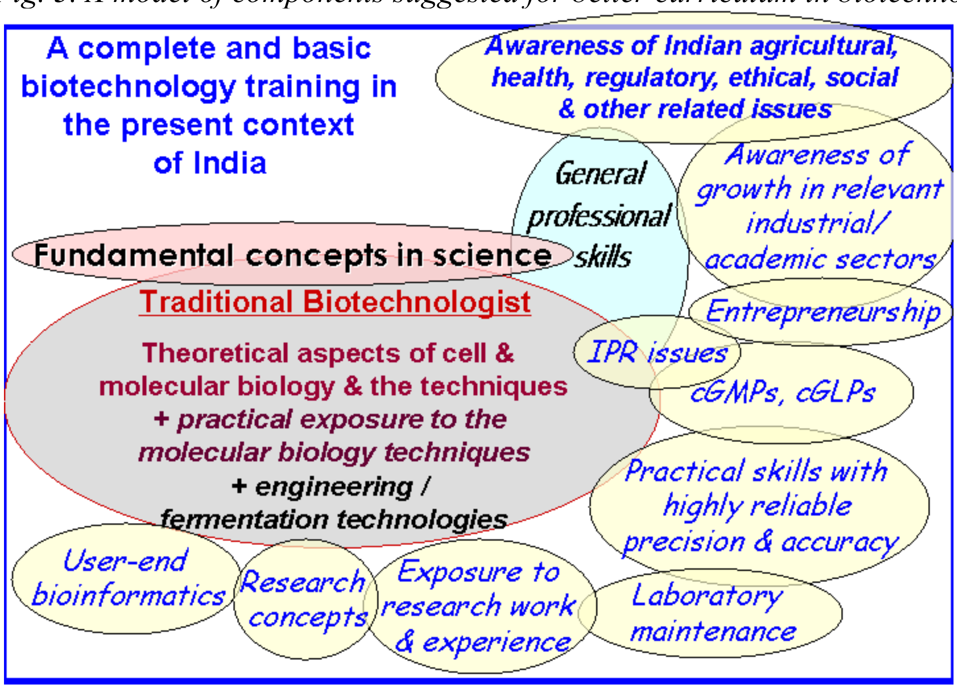
2. One of reasons for shortcomings in many software tools is the lack of proper integration of expertise from CMB and algorithm developing or programming skills. Some of the research projects at IBAB in the area of primer designing software development (*Sandhu and Acharya 2005*), promoter prediction and experimental validation of strategies involved in such computational programs, can be good examples of such integration. While there are several primer designing tools available online, the determination of melting temperature (Tm) is still vague and needs experimental analysis. In fact, relying on Tm determined by currently available mathematical approaches may not always be helpful as false product formation might occur during the polymerase chain reactions (*Acharya and Suswapna, unpublished*). A team at IBAB (Acharya KK, Reddy BK & Praveena K) is performing experimental validation of the primer/probe designing programs. Integrating the knowledge and skills related programming as well as CMB have proved very helpful in development of new primer designing programs.

There are also other important components often neglected in our curricula (see figure 3): Though this means a little straying from the central theme of this article, it is worth stressing the universal relevance of several other topics in our curricula, particularly in the context of the young and diverse Indian biotech-sector. Research and education in life sciences can no more be separated from the 'unconventional' topics like the Intellectual Property Rights (IPR), entrepreneurship, current Good Manufacturing & Laboratory Practices (cGMPs & GLPs) and verbal and written communication skills. Since there is an increasing demand for expertise in many of these areas, the graduation level education needs to create awareness in such aspects. It is crucial to sensitize the students to both academia and industry.

Further, there is a need to put any education/knowledge in the context of current problems of the nation. One cannot also forget that nothing would make sense in biotechnology/bioinformatics education unless students acquire very clear fundamental concepts in basic sciences, including biology. All these areas must be given appropriate attention while designing curricula in life sciences. In figure 3, these aspects are schematically represented with the example of the biotech-education, an area which has been increasingly in demand in many colleges, even at BSc level.

However, it may not be possible to integrate all novel ideas immediately into the syllabi that exist today. There is no point in simply increasing the volume of syllabi. There has to be a careful and gradual change in the course-structures with multi-dimensional efforts, which should involve the educational and industrial experts, government and the media (see figure 1). The feasibility of many modifications also depends on the infrastructure and the financial strength of educational institutes. To begin with, we could introduce the concepts of these unconventional topics with invited speakers, out-source short training programs, increase industry-academic tie ups and keep the students informed about the career options.

Fig. 3. A model of components suggested for better curriculum in biotechnology:



Conclusion:

Considering the requirements of growing Indian biotech sector, our educational programs need improvements in different aspects. A biotech-bioinfo interface has been identified in this article as an essential component of any life science (basic biosciences, bioinformatics, biotechnology, health and agricultural sciences) education. The components of this interface are:

1. Fundamental concepts in biology, particularly the cell and molecular biology (CMB)
2. User-end bioinformatics
3. Basic knowledge of the other 3 types of bioinformatics.
4. An understanding of the principles of molecular biology techniques and at least a preliminary level experience in these techniques.
5. Conceptual understanding of the research process and an exposure to some research projects in biology.

If students are well versed with the biotech-bioinfo interface, at least by the time they complete their post-graduation, they would be better equipped for biological research as well as other academic/industrial activities.

It is also crucial to recognize other lacunae in curricula; for example, IPR, entrepreneurship, career planning, communication skills, GMPs, ethics and awareness of national problems in health or agricultural areas.

References:

1. Vermette P, Proulx P and Marcos B (2003) Toward an integrated biotechnological engineering education program: a Canadian perspective. Nature Biotechnology 21 (12): 1525-1527.
2. Rao KC (2005) The State of Biotechnology Education. http://www.fbae.org/Channels/Biotech_education/state_of_biotechnology_education.htm
3. Suggested syllabi from DBT: Some syllabi are suggested on the DBT site (http://www.dbtindia.nic.in/scholarships/studenstsmain.html). This part of the DBT site was being updated and incomplete at the preparation of this article. More serious efforts like this with respect to the proper syllabi, required facilities and teaching faculty quality would help the students and parents would definitely help the students in choosing their future steps in education. These sites should be advertised more perhaps.
4. LCBT-IBAB: For details visit: http://www.ibab.ac.in/prog_biotechniques.html
5. Ranganathan S. (2005) Bioinformatics Education – Perspectives and Challenges. PLOS computational Biology. November 1: 6 e52.
6. BINC 2005: BINC is abbreviation of the Bioinformatics National Certification test launched last year by the Department of Biotechnology (DBT), Govt. of India. More details are expected to be at: http://bioinfo.ernet.in In the first test, last year, none of the candidates qualified! The preamble summary from the website: In India, Universities and Institutions, Government and Private, are regularly training Bioinformaticians. Trained Bioinformaticians are not of the similar standards though they hold similar certificates across India. Hence Department of Biotechnology (DBT), Government of India, has decided to evaluate the standard of these trained Bioinformaticians by conducting a national level test at regular intervals. This is expected to help the industry in selecting required manpower in Bioinformatics area in the country. The eligibility for the course:

Minimum Bachelor's degree in Science, Pharmacy, Engineering, Agriculture, Veterinary and Medical Sciences, Ayurveda, Homeopathy from recognised Indian Universities and Institutions. Post graduates and Ph.D. in above mentioned disciplines are also eligible.

7. Altman R (1998) Editorial: A curriculum for bioinformatics: The time is ripe. Bioinformatics 14 (7): 549-550
8. Campbell MA (2002) Meeting report: Genomics in the undergraduate curriculum-Rocket science or basic science? Cell Biology Education 1: 70-72.
9. Centeno NB, Villa-Freixa J and Oliva B (2003) Teaching structural bioinformatics at the undergraduate level. Biochemistry and molecular biology education 31 (6): 386-391.
10. Eberhardt ES, Hansen J, Riservato L, Cole M, Smaglo B and Szaniawski P (2003) Preparing undergraduates to participate in the post-genome era. Biochemistry and molecular biology education 31 (6): 402-409.
11. Lyon J, Giuse NB, Williams A, Koonce T and Walden R. (2004). A model for training the bioinformationist. J Med Libr Assoc 92 (2): 188-195.
12. Honts JE (2003) Evolving strategies for the incorporation of bioinformatics within the undergraduate cell biology curriculum. Cell Biology Education 2 p. 233-247.
13. Pevzner PA (2004) Editorial: Educating biologists in the 21st century: bioinformatics scientists versus bioinformatics technicians. Bioinformatics 20 (14) p. 2159-2161.
14. Sandhu KS and Acharya KK (2005) ExPrimer: To design primers from exon-exon junctions. *Bioinformatics* 21(9)2091-2.
15. Acharya KK, Govind CK, Shore AN, Stoler MH, Reddi PP (2006) *cis*-Requirement for the Maintenance of Round Spermatid-Specific *Dev Biol. Jul 15;295(2):781-90*
16. Acharya KK and Suswapna P (unpublished) The work carried out at IBAB, Bangalore, showed that while in many cases determination of Tm is not that critical to the PCR experiments, some of the existing Tm-determination tools may not be reliable as they sometimes yield false positives.

Other suggested readings:

17. Delivering Bioinformatics Training: Bridging the gaps between computer science and biomedicine. Proceedings of the AMIA 2002, Annual Symposium, p. 220. Dubay C, Brundege JM, Hersh W and Spackman K.
18. Bioinformatics Research and Education in Germany. In Silico Biology 2 (2002) 169-171. Schomoburg D and Vingron M.
19. Rapid development of bioinformatics education in China. J Biol Education (2003) 37 (2). Zhong Y, Zhang X, Ma J and Zhang L.
20. Guidelines for establishing undergraduate bioinformatics courses J Sc Education Tech. 2003. Cphen J.

The article was originally presented as a lecture in the Biovistas's inaugural function (A UGC sponsored National conference, 24-25 Feb 2006) at the St. Joseph College, Bangalore