Meaningful Learning at the Interface of Computing and Molecular Biology

Mohammad Ashraf Iqbal
Yasser Hashmi
Muhammad Aurangzeb
Lahore University of Management Sciences (LUMS), Lahore, Pakistan
aiqbal@lums.edu.pk
yasser@lums.edu.pk
05030087@lums.edu.pk

Abstract - **Meaningful Learning requires that the learner must consciously choose to relate the concepts and propositions in some substantive way to what relevant past knowledge already exists in the learner's cognitive structure. In fact the richness of meaning for a concept increases exponentially with the number of meaningful propositions that relate that concept to other concepts in the concept map. Concept formation and concept assimilation in a given domain advances learning in a natural way but meaningful learning at the interface of different domains poses new challenges. The language, terminology, and the nomenclature of Mathematics & Computing are very different from that of Molecular Biology. Thus making direct propositional links between concepts belonging to such diverse domains may be an uphill task if not impossible. We need to design and insert new concepts at the interface of the two domains; we call such concepts as Bridging Concepts as they bridge or link concepts belonging to Computing and Molecular Biology. The major challenge is to assimilate these Bridging Concepts in both domains. In this research, we shall describe our initial experiences of teaching and learning in a standard course on Computational Biology offered to students majoring in Computer Science and Computer Engineering at the Lahore University of Management Sciences, Lahore, Pakistan.**

Index Terms - Bridging Concepts, Concept maps, Class Room Learning, Molecular Biology, Computer Science, and Hybridization of Probes.

INTRODUCTION

Molecular Biology had progressed significantly in the first twenty years after the discovery of the double helix in the mid fifties. Then came a number of technological breakthroughs in the sixties, seventies, and eighties that made it possible to read, understand and manipulate the very basic code of life that is the DNA. This revolution in biology in the last forty years has not only created new sub fields like Bioinformatics or Computational Biology, it has also impacted the science and engineering of algorithms. Molecular Biology has in fact become the largest source of new algorithmic problems and because of the immense size of the data involved there is a growing need to design better algorithms which are efficient in terms of time and space [1]-[3].

In this research, we shall be closely looking at the problems of learning at the interface of computing and molecular biology. Our world view of the interface follows the route number 3 in Fig. 1; this starts from the concepts of molecular biology, creates a high level abstraction or a model from a problem in molecular biology and then maps the model on a number of known problems in algorithms. The model is in fact a bridging concept between computing and molecular biology without which it will be almost impossible to use tools and techniques from computing to solve problems in biology. We claim that learning at the interface of the two domains is significantly facilitated provided we spend enough time in understanding the respective bridging concepts.

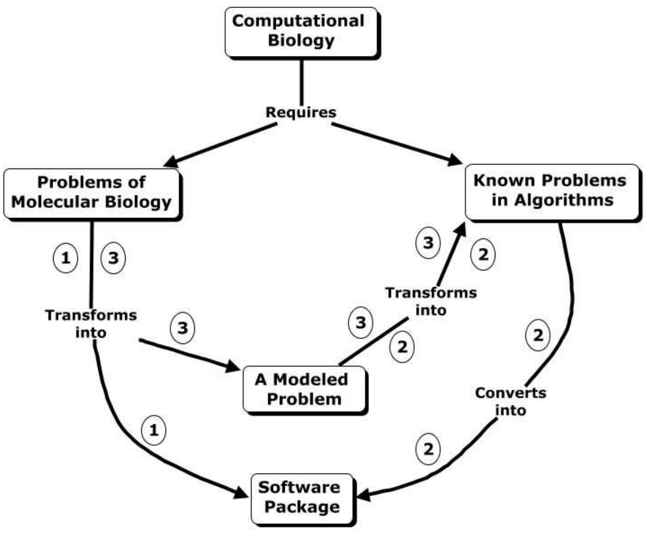

FIGURE 1
DIFFERENT WORLDVIEWS REGARDING THE INTERFACE OF COMPUTING AND BIOLOGY. THE ROUTE NUMBER 1 IS FOLLOWED BY TRADITIONAL BIOLOGISTS; NUMBER 2 IS COMMON AMONG TRADITIONAL COMPUTER SCIENTISTS, WHILE WE INTEND TO FOLLOW ROUTE NUMBER 3.

It is interesting to note that the traditional biologist has a different worldview of the interface. This has been depicted in route number 1; he or she does not worry about the abstractions or the algorithms and simply concentrates on using a software package provided one such package exists.

The traditional computer scientist's point of view is depicted by route number 2. It starts from an abstraction (assuming that someone else has done the intellectual work of creating an abstraction), selects a suitable algorithmic solution and then designs and engineers appropriate software. It is quite evident that the real challenge lies in route number 3 where one starts with the understanding of a problem in (the language of) molecular biology, creates a suitable abstraction, and then maps it into one of the known problems in (the language of) algorithms. This obviously covers the whole spectrum of problem solving in a domain of science using the tools, techniques, and abstractions of another domain [3].

The interface between computing and molecular biology produces exciting challenges both for computer scientists and biologists. Computer scientists have to study molecular biology and biologists have to study computing, in general, and algorithms, in particular. There is a growing tendency in which many biology major students are taking some sort of fundamental course in algorithms while curious computer science majors have started taking interest in courses in molecular biology. This pattern has become quite common in graduate studies and research [1] & [2].

In this paper, we shall describe our initial experiences of teaching and learning in a standard course "Computational Biology" offered to students majoring in Computer Science and Computer Engineering at the Lahore University of Management Sciences, Lahore, Pakistan. There was a mix of graduate and undergraduate students and some of them had earlier taken a basic course in biology at the undergraduate level. For the course we have used the standard text book by Jones and Pavezner, An Introduction to Bioinformatics Algorithms [1]. This paper is organized in the following five sections. The next section describes the exciting opportunities offered by the interface of computing and molecular biology. Section 3 describes the problems of learning and understanding at the interface. We describe our experiences in Section 4. The paper concludes in Section 5.

THE INTERFACE OF COMPUTING & MOLECULAR BIOLOGY

Grand problems posed by nature do not respect artificial boundaries drawn between conventional disciplines; thus in order to solve challenging problems, practitioners belonging to different disciplines should learn to work on problems that are obviously shared and also on problems that are not obviously shared. Sometimes it requires a great mind to break the ice on such cross disciplinary problems; Erwin Schrödinger was a renowned physicist and was famous for his major contributions in wave theory. His book "*What is Life?*", published in 1944, has influenced an entire generation of scientists to work at the interface of biology, physics and chemistry. Both James Watson, who wanted to become a naturalist, and Francis Crick, a physicist, switched their fields and started working in the newly emerging DNA science after reading Schrodinger's book. It is interesting to note that James Watson did not know much about organic or physical chemistry before he started his pioneering work on the double helix while Francis Crick knew very little biology [2].

Tools and techniques coming from other disciplines may provide perspectives that are not available otherwise [3]. These perspectives sometimes lead to a new integrative explanation for the intricacies of a phenomenon; sometimes they open an entirely new subfield. The discovery of the double helix was not possible without a sound knowledge of crystallography; it started a new subfield of molecular biology which borrowed tools and techniques from physics and chemistry, and has changed the face of conventional biology. In recent times, it was the sequencing of complete genomes that brought computing to the forefront of biology and has also identified the need of understanding algorithms in order to tackle biological problems [2].

Advances in computer hardware, computational tools, and models have fueled the progress of Molecular Biology in the last decade of the 20th century. Modeling and simulation of biological systems or subsystems such as cells have encouraged computer scientists, mathematicians, and molecular biologists to work together on activities as diverse as pharmaceutical design to environmental analysis. Indeed the interaction between computing and molecular biology has the potential to impact strongly on human health and society; this includes improved disease treatment, better and more robust food production, and efficient solutions to environmental cleanup [2]. Nowadays, computing has become an integral part of molecular and cell biology and has enabled researchers to expand their scope of inquiry from gene sequence analysis to broader investigations of biological complexity. As biological problems become bigger and more complex there is a growing need of a formal study of computer science, in general, and algorithms, in particular. It is interesting to note that computing not only provides sophisticated tools or models; it can also provide intellectual abstractions in order to get a better insight into biological phenomena and also a (language) platform for studying such phenomena.

PROBLEMS OF LEARNING AT THE INTERFACE

Novak claims that organized knowledge is essential for effective learning, and it consists of concepts which are perceived regularities between events and objects. Concepts are in turn connected by propositions which form the cognitive structure of the learner. Novak has devised a novel tool, known as *concept map*, to model and represent organized knowledge in a cognitive structure [4]. Concept maps consist of concepts usually enclosed by circles while propositions connect two or more concepts using linking words or phrases in a meaningful sentence (see Fig. 2).

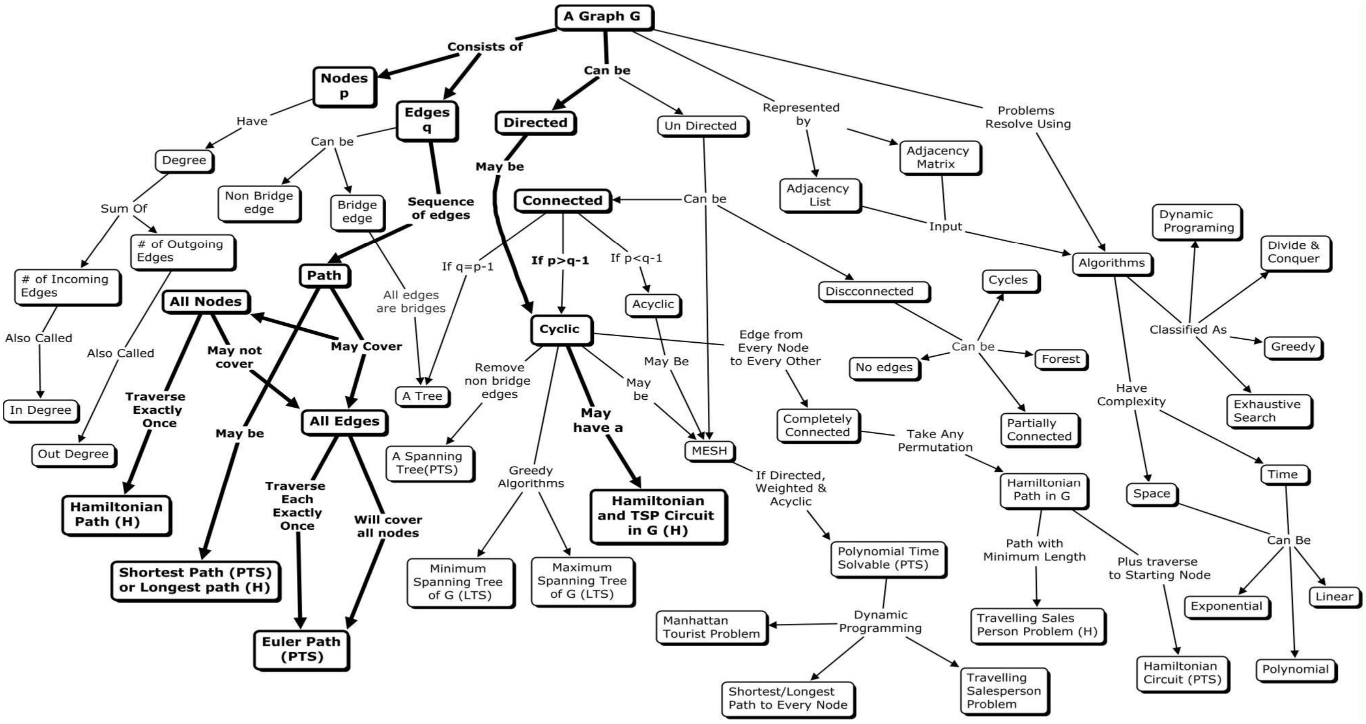

FIGURE 2

A CONCEPT MAP SHOWING DIFFERENT CONCEPTS IN GRAPH ALGORITHMS. THE MAP SHOWS CONCEPTS FROM GRAPH ALGORITHMS, IN GENERAL, AND CONCEPTS NEEDED FOR SOLVING THE PROBLEM OF SEQUENCING BY HYBRIDIZATION, IN PARTICULAR (SHOWN IN BOLD).

Instructional strategies that emphasize relating new knowledge to the learner's existing knowledge encourage meaningful learning. Evaluation strategies that encourage learners to relate concepts they possess with new concepts also encourage meaningful learning. According to Ausubel's cognitive psychology learning takes place by assimilation of new concepts and propositions into the existing cognitive structure [5]. This means that the amount of meaningful learning of a learner depends upon the number of propositional links that he or she creates between old and new concepts in his or her concept map.

Sequencing by Hybridization

(From Molecular Biology to Graph Algorithms)

In terms of Molecular Biology: Given a probe consisting of short single-stranded synthetic DNA fragment, and a single-stranded target DNA fragment, the target will hybridize, that is it will form hydrogen bonds with the probe provided the probe is a substring of the target's Watson-Crick complement. If we have a list of all possible probes of a fixed length that hybridize with the target sequence then the problem is to reconstruct the sequence of the target DNA fragment.

In terms of Graph Algorithm 1: Find a Hamiltonian path in a directed and possibly cyclic graph.

In terms of Graph Algorithm 2: Find an Euler path in a directed and possibly cyclic graph.

Box No. 1

An important feature or facility of the concept map is the inclusion of cross-links. These links represent relationships between concepts in different domains of knowledge. Cross-links encourage us to appreciate or discover how a concept in one domain of knowledge is related to a concept in another domain. Creation of knowledge at the interface of two domains is essentially discovering cross-links between concepts belonging to different domains; such cross-links sometimes represent creative leaps on the part of the knowledge creator. The challenge is to find cross-links between two domains where scientists speak different languages, use very different terminologies, nomenclatures, and even have different intellectual styles. Computer scientists and mathematicians may not be able to appreciate the complexity inherent in biological systems while the biologists may be skeptical about the power of the mathematical models and abstractions used by computing people in their routine work. Making a model which is simple enough to be manipulated but which can capture the essential details of a biological problem creates a lot of intellectual tension between computer scientists and biologists [3].

The Problem of Sequencing by Hybridization

We shall use a specific example in molecular biology, that is, sequencing by hybridization, to analyze problems of learning at the interface of computing and molecular biology. We know that learning (at this interface) is equivalent to finding appropriate cross-links between two concept maps, one belonging to computing while the other belonging to molecular biology. We present a concept map showing a hierarchy of algorithms in Fig. 2.

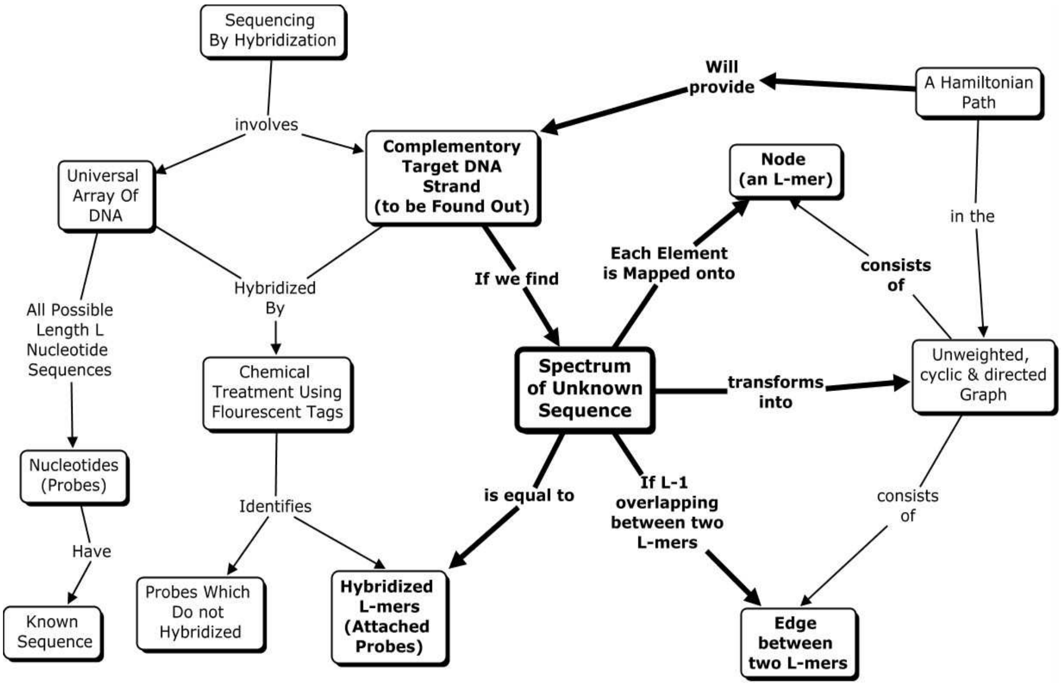

FIGURE 3

Concept map showing Bridging Concepts needed to understand the concept of sequencing by hybridization. The concepts in bold are the bridging concepts while the concepts shown on the left belong to molecular biology while the concepts on the right belong to algorithms. There may be a grey area in which a concept may be a bridging concept while it belongs to either biology or algorithms as well. The Spectrum of a Sequence is a true bridging concept between the two different domains.

It is quite evident that making direct propositional links between concepts belonging to such diverse domains is not possible (see Box No. 1 and Fig. 2 & Fig. 3). We need to design and insert new concepts at the interface of the two domains; we call such concepts as Bridging Concepts as they bridge or link concepts belonging to Computing and Molecular Biology. The Bridging Concepts for the problem of sequencing by hybridization are shown in bold in Fig. 3. The major challenge, as far as learning is concerned, is to assimilate these Bridging Concepts in both domains [6] & [7].

Returning back to our specific problem, the concept Spectrum (*s*, *l*) of a sequence *s* of letters is a bridging concept between molecular biology and computing, and is equal to all strings of length *l* that the sequence *s* contains. For example if *s* = BIOLOGY then Spectrum(*s*, 3) = {BIO, IOL, OLO, LOG, OGY}. The elements in the spectrum may not appear in the same order, (in fact, they may not have any order?) and the challenge is to find the string *s*, given its spectrum. The sequencing by hybridization problem in molecular biology directly transforms into the spectrum problem while the spectrum problem can be transformed either into a Hamiltonian path problem or a Euler path problem in directed graphs depending upon whether we map every element of the spectrum into a node (see Fig. 3) or an edge of a directed graph (not shown in Fig. 3) respectively.

AN EXAMPLE LEARNING EXPERIENCE

The above concepts of sequencing by hybridization were taught as part of a standard course "Computational Biology", which was offered for senior undergraduate and graduate students in the Department of Computer Science at the Lahore University of Management Sciences, Lahore, Pakistan. In order to maintain a reasonable amount of rigor in Computing as well as Molecular Biology, this course was jointly offered by a professor of Computer Science and a professor of Molecular Biology. After the completion of the said course, a number of students, who passed this course, were randomly selected and then participated in an interviewing process based on the problem of sequencing by hybridization. The interview was based on four modules: In the "Computing Basics" module, the students were encouraged to recall relevant concepts in Computing (or Algorithms). In the "Molecular Biology Basics", concepts relevant to sequencing by hybridization were discussed and then tested. In the "Direct Problem" module, the problem of constructing a DNA sequence from a list of all possible probes that hybridize with an unknown target sequence is addressed without the help of an intermediate or bridging concept. If a subject fails to provide any reasonable solution to the problem then "Intermediate Stage" module is presented where the same problem is addressed again with the help of bridging concepts.

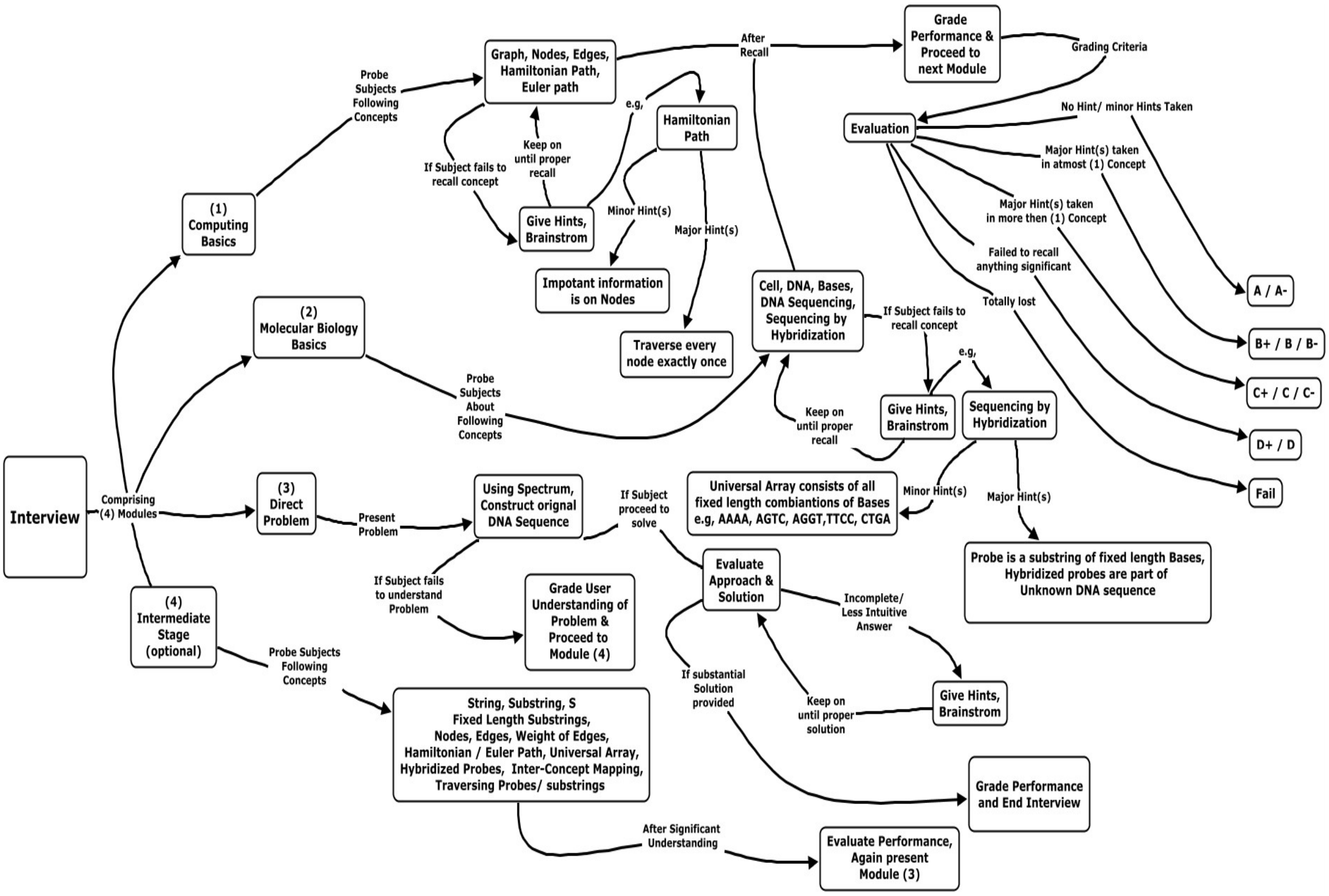

FIGURE 4
CONCEPT MAP SHOWING OUR SURVEY APPROACH INCLUDING OUR INTERVIEWING PATTERN, WHAT TYPE OF QUESTIONS ASKED AND HOW WE EVALUATED AN INDIVIDUAL'S PERFORMANCE.

The concept map showing our survey approach is shown in Fig. 4. It illustrates our interviewing strategy as well as our evaluation policy. A summary of the survey is tabulated in Table 1. The average performance in the "Computing Basics" was the highest while the lowest performance was witnessed in the "Direct Problem" module. The average performance increased considerably when a number of intermediate or bridging concepts were introduced. This observation is further highlighted in Fig. 5 where individual performances of all 20 subjects are plotted with and without the help of bridging concepts.

CONCLUSIONS

We have addressed a problem of learning at the interface of computing and molecular biology. We have faced this and other such problems [8]-[17] while teaching a course "Computational Biology" at senior undergraduate and graduate level. This problem is essentially discovering propositional links between concepts belonging to two very different domains. The problem is further compounded by the fact that people belonging to these domains speak different languages and use different terminologies. Thus transforming a problem of molecular biology into an algorithmic problem requires an intellectual effort which is far greater than is needed in general problem solving, which by itself is not a trivial exercise from the point of learning and cognition [11]-[14]. Our preliminary studies have shown that this problem transformation is considerably facilitated provided a number of bridging concepts are inserted between concepts belonging to molecular biology and algorithms [6] & [7]. It is interesting to note that the so called bridging concepts provide a domain independent platform for the conceptual traffic to move freely from one domain to the other. We need to conduct more experiments to find the effectiveness of varying levels of Bridging Concepts inserted at different locations. It will also be interesting to study the performance of Biology majors as compared to Computer Science or Computer Engineering students.

ACKNOWLEDGMENT

We wish to thank R. Khan, N. Malik, M. Arslan, R. Jaferi, S. Rahman, F. Lodhi, K. Fahd and A. R. Shakoori for their help and encouragement. We are extremely thankful to Sadaf Siddiqi for carefully reading the manuscript and pinpointing and correcting a number of grammatical mistakes. We also thank the Computer Science Department of the Lahore University of Management Sciences for supporting this research.

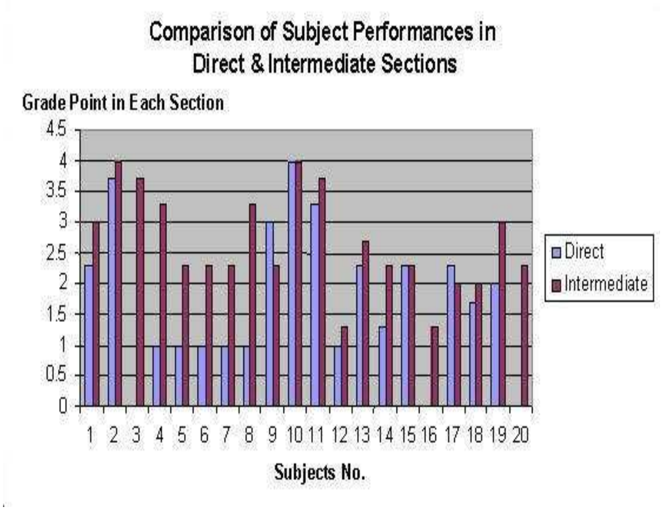

FIGURE 5
GRAPH COMPARING THE PERFORMANCE OF INDIVIDUAL SUBJECTS.

TABLE I
TABLE SHOWING THE GRADE AVERAGES (OUT OF 4.0) OF ALL TWENTY (20) SUBJECTS' PERFORMANCES IN INDIVIDUAL SECTIONS. ROW 5 SHOWS THE SUBJECTS WHO HAD SOLVED THE PROBLEM POSED IN MODULE 3 OF INTERVIEW COMPLETELY WITHOUT USING HINTS / BRIDGING CONCEPTS PROVIDED IN MODULE 4 OF THE INTERVIEW.

Module No.	Module Description	Average Grade / Score
1.	Computing Basics	B
2.	Molecular Biology Basics	C
3.	Direct Problem	C
4.	Intermediate Stage	B-
5.	Able to Solve Directly	5/20 (Subjects)

REFERENCES

[1] Jones, N.C. & Pevzner, P.A., *An Introduction to Bioinformatics Algorithms*, 2004, The MIT Press.

[2] Watson, J.,"DNA: The Secret of Life", 2004, Arrow Books.

[3] Wooley, J.C. and Lin, H.S., (Editors) "Catalyzing Inquiry at the Interface of Computing and Biology", Committee on Frontiers at the Interface of Computing and Biology, National Research Council, 2005, The National Academic Press.

[4] Novak, J.D., "Learning Creating, and Using Knowledge", 1998, Lawrence Erlbaum Associates, Publishers.

[5] Ausubel, D.P., et al., *Educational psychology: A cognitive view*, 1978, Holt, Reinhart, and Winston, New York.

[6] Clement, J., "Using Bridging analogies and anchoring intuitions to deal with student's preconceptions in Physics". *Journal of Research in Science Teaching 30(10)*, 1993, pp. 1241-1257.

[7] Iqbal, M. A. Hashmi, Y., "Bridging Over Problems of Learning in Finding Strongly Connected Components", *IADIS International Conference on Cognition and Exploratory Learning in Digital Age*, 2005, (CELDA 2005), Portugal.

[8] Anderson, J.R, *Cognitive Psychology & its Implications*, 1985, W.H. Freeman & Company, New York.

[9] Anderson, L.W., et al., A *Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, 2001, Addison Wesley Longman, Inc.

[10] Bransford, J.D., et al., *How People Learn: brain, mind, experience and school*, 2000, Committee on Development in the Science of Learning and Committee on Learning Research and Educational Practice, Commission on Behavioral and Social Sciences and Education, National Research Council, National Academy of Sciences,

[11] Iqbal, M. A. and Tahir, S., "Should We Teach Algorithms?", *IJECE, Vol.2, No.2*, 2003, Summer-Fall, pp. 134-140.

[12] Iqbal, M. A. and Alvi, A., "The Magic of Dynamic Programming", *Innovations 2004: World Innovations in Engineering Education and Research*, Int. Network for Eng. Ed. and Res. (iNEER), 2004, Arlington, VA, pp. 409-418.

[13] Iqbal, M. A., et al., "Teaching Science and Mathematics: Discovery Based Learning", *Innovations 2005: World Innovations in Engineering Education and Research*, Int. Network for Eng. Ed. and Res. (iNEER), 2005, Arlington, VA, pp. 327-346.

[14] Nadeem, N., et al.,"Alternate Models to Teach Probability", International Conference on Engineering Education & Research: Exploring Innovation in Education & Research, (iCEER-2005), 2005, Tainan, Taiwan.

[15] Mintzes, J.J., et al., "Teaching Science for Understanding: A Human Constructivist View", 1998, Academy Press.

[16] Mintzes, J.J., et al., 2000. Assessing Science Understanding: A Human Constructivist View. Academy Press.

[17] Treagust, D.F., et al., "Overview: Research on Students' Pre-instructional Conceptions-The Driving Force for Improving Teaching and Learning in Science & Mathematics: Chapter 1", Treagust, D.F., et al. (ed.), *Improving Teaching and Learning in Science and Mathematics*, Teachers College Press, pp. 1-16.