

Teaching the ABCs of Bioinformatics



Jingchu Luo

Centre of Bioinformatics, Peking University, EMBnet China node, China

It has been ten years since we started the semester course Computer Application to Molecular Biology in 2000. We now use Applied Bioinformatics Course, or ABC as a simple acronym for this course. Apparently, as its name indicates, ABC is an entry level introductory course, rather than an advanced one. We run the course in a training room (Fig. 1). Each student has a PC with both Linux and Windows installed.

The course is designed for biological graduate students to solve practical problems. After a brief introduction to the three most popular bioinformatics resources, the National Center for Biotechnology Information (NCBI), the European Bioinformatics Institute (EBI) and the Expert of Protein Analysis System (ExpASY), the students are happy to start with hands-on practice, to retrieve the alpha subunit of human, mouse and rat hemoglobin from Swiss-Prot, then make comparison of the amino acid sequences between each other [Swiss-Prot: HBA_HUMAN, HBA_MOUSE, HBA_RAT]. Couple of online web servers can be found from the links at ExpASY, such as the built-in



Figure 1. The students are doing hands-on practice in the training room of the ABC course.

programs in EBI SRS, and the NCBI tool box. We use the bioinformatics platform [WebLab](http://weblab.cbi.pku.edu.cn/)¹ we developed locally.

Surprisingly, the output of the Needle global alignment is out of expect, the sequence identity between mouse and rat is less than that of mouse and human. This triggers the interests of the students with enthusiastic discussion. Finally, they find the answer to this question by retrieving and comparing the nucleotide coding sequence (CDS) of the hemoglobin alpha gene of these three mammals. Indeed, the identity of CDS of mouse and rat is higher than that of mouse and human, which tells us that mouse and rat are close relatives based on the analysis of molecular biology data.

A dozen of simple exercises like this are designed. The students can get familiar with various bioinformatics tools by doing these exercises. For example, by comparing the 39kb genomic sequence of a Fugu cosmid [GenBank: AF164138] to itself using dottup, the dot plot program embedded in EMBOSS, we can easily spot out the tandem duplication of the multi-drug resistance gene in this cosmid. By comparing the predicted results of the trans-membrane helices of a post-floral protein [Swiss-Prot: Q9FY06] identified from a cDNA library of short-day grown G2 pea tissue, using several tools including TMAP, THHMM, TMHMM, TMPred, the students are well aware that different tools may have different output, a common scenario of doing dry lab experiments.

Running BLAST seems an easy job for most of the students. However, running a Good BLAST is not that easy. Literature search reveals that neuroglobin [Swiss-Prot: NGB_HUMAN] is a member of the human globin family to which nine hemoglobin subunits (alpha, beta, gamma, etc.) as well as myoglobin and cytoglobin belong. Take this as an example for BLAST search, we then ask, "can we find neuroglobin through BLAST search using the alpha subunit [Swiss-Prot: HBA_HUMAN] as a query"? The answer is "No" if we use the default parameters to run BlastP through the NCBI BLAST server. Nevertheless, we can obtain a good match using PSI-BLAST, by choosing Swiss-Prot as the preferred database, selecting Human as the organism to search, and setting E-value to 0.001. The 12 hits obtained by the above search are ready to retrieve for further analysis for this human globing family, such as making multiple

1 <http://weblab.cbi.pku.edu.cn/>

sequence alignment, drawing a nice sequence logo, building up a phylogeny tree. By doing so, the students are convinced that it is critical to know the general principles and biological background behind the programs such as BLAST.

Several projects are implemented throughout the course. One of them is the analysis of the sequence, structure and function of the bar-headed goose hemoglobin. As we all know, hemoglobin is one of the most well studied proteins in the last century. More than 700 entries can be found in the Swiss-Prot database. Three-dimensional structures of wild type and mutants from dozens of species have been solved. This provides us a good opportunity to study the relationship among sequence, structure and function of hemoglobin molecules of human and other species. Bar-headed goose is a migration bird. They live in the Qinghai lake during summer time and fly to India all way along over the Tibetan plateau in autumn and come back again in spring. Interestingly, close relative of bar-headed goose, the graylag goose lives in the low land all year around. Sequence alignment of bar-headed goose hemoglobin [PDB: 1A4F] with that of graylag goose [PDB: 1FAW] shows only 4 substitutions. One of them is Pro 119 in the alpha subunit of graylag goose and Ala 119 of bar-headed goose. This residue is located in the surface of the alpha/beta interface. In 1983, Perutz proposed that this substitution reduces the contact between the alpha and beta subunit and increases the oxygen affinity, due to the relation of the tension status in the deoxy form. During the past decade, a research group at Peking University has solved the crystal structure of both deoxy and oxy form of the bar-headed goose, as well as the oxy form of the graylag goose hemoglobin. Using the powerful free software [Swiss-PdbViewer](http://www.vital-it.ch/spdbv/)², the students make a Magic Fit to superpose the alpha/beta heterodimer of the two goose hemoglobins on each other. They are excited to see the conformation difference caused by one amino acid residue substitution Pro119Ala between the two goose hemoglobin molecules (Fig. 2) by measuring the distance of the side chain atoms between Ala/Pro 119 and Leu 55 in the beta subunit. Furthermore, they make a mutation by replacing Leu 55 in the beta subunit with Ala 55 and propose that this mutant may further increase the oxygen affinity.

2 <http://spdbv.vital-it.ch/>

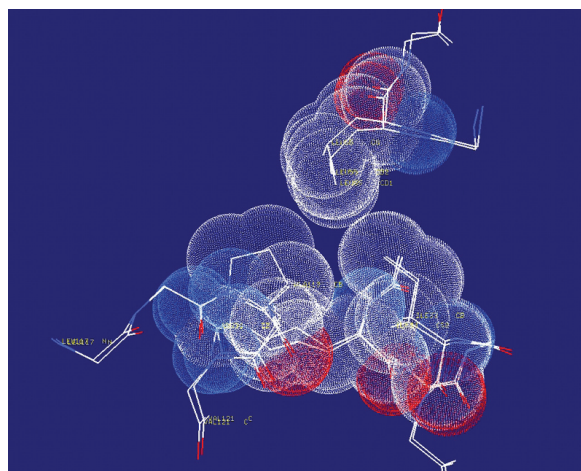


Figure 2. The superposed structure of bar-headed goose [PDB: 1A4F] and graylag goose [PDB: 1FAW] hemoglobin to show the interface of alpha and beta subunit. Pro 119 of alpha subunit (1FAW) has closer contact with Leu 55 of beta subunit than that of Ala 119 (1A4F).

In addition to the pre-defined projects, the students are also encouraged to bring their own projects. They are divided into groups, four students each, and discuss the project outside the class as homework assignments, work together to solve the problems. At the end of the course, a workshop is organized. Speakers chosen by group members make presentation on behalf of each group to summarize what they have learned during the course, and what they plan to continue to learn in the future.

Indeed, the ABC course is aimed to teach ABCs of bioinformatics. We hope that, by learning the course, the students will be well convinced that *"half day on the web, saves you half month in the lab!"*