We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Introductory Chapter: From Hard to Soft Biology

Payam Behzadi

1. Experimentation and computation

On the evening of Monday, April 13, 2020, during "the Wuhan-China virus (COVID-19) Home-Self Quarantine Era," I was drinking coffee and simultaneously searching on Google Scholar to find some valuable papers regarding "computational biology."

Among a mass of article links, an article entitled "laptop biology" [1] attracted me. I began to read this paper carefully and found some valuable terms including "hard science" and "soft science." Indeed, the term "soft science" was used for "experimentation, classification, observation and intuition," while the "hard science" term depicted "mathematics and algorithms" [1].

Then I checked https://www.thefreedictionary.com/ and searched for the terms "hard science" and "soft science." The results were as follows:

- "Hard Science: one of the natural or physical sciences, such as physics, chemistry, biology, geology, or astronomy, any of the natural or physical sciences, in which hypotheses are rigorously tested through observation and experimentation" (https://www.thefreedictionary.com/hard+science).

- "Soft Science: a science, such as sociology or anthropology, that deals with humans as its principle subject matter, and is therefore not generally considered to be based on rigorous experimentation, any of the scientific disciplines, as those which study human behavior or institutions, in which strictly measurable criteria are difficult to obtain" (https://www.thefreedictionary.com/Soft+science).

Although the meanings of these terms were very different from what I thought, I liked them. Due to this fact I found that it is better to use my terminology talent to represent new terms "hard biology" and "soft biology."

I used the term "hard biology" for experimental (in vitro, in vivo, and in situ investigations) biology. This term has a direct deal with experimentation and work bench within the wet labs [2].

In contrast to "hard biology," I used the term "soft biology" based on in silico or desktop work [2] and laptop biology [1].

I hope that these terms are useful for the readers of this book and other scientists around the world. With this background I begin the main text of the introductory chapter.

2. Biology and computer

Undoubtedly the famous physical chemist from the USA, Margaret Dayhoff (1925–1983), the mother and father of bioinformatics, was the key scientist who employed computers and the related software tools in biochemistry [3, 4].

Indeed, it was Dayhoff who understood the importance of computers and computational methods not only in biology but also in medicine [4].

In 1960, Dayhoff as the Associate Director of the National Biomedical Research Foundation began her collaboration with her physicist colleague, Robert S. Ledley, who, like Dayhoff, was interested in employing computers in biomedical sciences [3, 5, 6].

The outcome of their scientific collaboration (Dayhoff-Ledley) in the period of 1958–1962 led to a computer program COMPROTEIN (coded FORTRAN) which was designed for IBM 7090. The software COMPROTEIN (a de novo sequence assembler) was able to determine the primary structure of protein throughout the Edman peptide sequencing data [3, 7].

The amino acid one-letter coding system was founded by Dayhoff [3, 8]. Dayhoff and Eck continued their scientific activities by publishing the first edition of the invaluable book entitled *Atlas of Protein Sequence and Structure* in 1965 which involved 65 protein sequences [3, 4, 9]. The fourth edition of *Atlas of Protein Sequence and Structure* which was published in 1969 included more than 300 protein sequences. So, this atlas established the first database of biological sequence [3, 4]. Interestingly, the sequence alignment of biopolymers was started by proteins not DNA molecules. This claim is proven by representing a 12-sequenced DNA fragment in 1971 [4].

3. Molecular biology and computer

During the golden decade between the years 1970 and 1980, the DNA language was decoded. However, the genetic codes of 64 codons were decoded in 1968 [3, 10]. Sanger's sequencing method of DNA based on "plus and minus" strands was performed 25 years after the recognition of the first protein sequence [3, 11, 12]. 1979 is the historical year for using the first software for Sanger's DNA sequencing method. In the paper published by Rodger Staden in 1979 via the journal of *Nucleic Acids Research*, the applied programs including OVRLAP, XMATCH, and FILINS (coded FORTAN) were described [3, 13].

During the years of 1980–1990, the application of computational sciences significantly increased. In 1983, the polymerase chain reaction (PCR) was invented by Kary B. Mullis (1944–2019 (https://www.nytimes.com/2019/08/15/science/kary-b-mullis-dead.html)). Kary Banks Mullis as an American biochemist invented a valuable molecular method which was based on in vitro synthesis of DNA [14–16]. So, by the discovery of DNA molecules in the 1950s, and in consequence the early application of pro-computers, invention of molecular and sequencing methods, and utilizing Internet services within a short duration, it seems that several revolution- ized features have happened in molecular biology [17].

Although computers and the related software tools were employed since the 1960s in biology, it was in the limited scales. I believe that by the invention of PCR as an in silico-in vitro (dry lab-wet lab) technology and its flying speed as a general molecular biology approach changed the traditional methodologies. By global generalization of PCR, the use of Internet services, computers, and software tools got significant acceleration. In this regard, designing different primers in large and global scales led to progression of in silico studies and appearance of dry labs within the wet labs.

Due to this fact, during a very short time, a mass of raw data was obtained by scientists around the world, and these data got stored within different biological databases like the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/), European Molecular Biology Laboratory (EMBL) (https://www.embl.org/), Kyoto Encyclopedia of Genes and Genomes (KEGG) (https://www.genome.jp/kegg/), and DNA Data Bank of Japan (DDBJ) (https://www.ddbj.nig.ac.jp/index-e.html) [18].

At the same time, these giant databases began to give more free software tools, information, and other services. These features have led to establishing 1637 free online databases (http://www.oxfordjournals.org/nar/database/c/) up to now [19].

Today, some databases including The Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) (https://www.rcsb.org/) serve its global users for free. 3D macromolecular structure data is one of the most popular products which scientists and researchers use for free around the world [20].

Since 1971 when the US data center of RCSB PDB was founded, it provided digital data in biology for its users with open-access policy [20], an opportunity which all the consumers should be grateful for.

All in all, soft biology was founded in the 1960s with a low speed, but it accelerated by the invention of PCR in the 1980s. The PCR invention softened the science of biology throughout Internet services, an occurrence which resulted in progressive in silico studies, dry labs, and software tools.

Today, a biologist is recognized by her/his laptop, Internet connection, and filled USB flash drives with different bioinformatics software tools!

Hence, hard biology got softened by establishing the science of bioinformatics and is continued to be more softened by computational biology!

But the important question is:

"How does biology get more softened in the future?"

Conflict of interest

The authors declare no conflicts of interest.

Author details

Payam Behzadi
Department of Microbiology, College of Basic Sciences, Shahr-e-Qods Branch, Islamic Azad University, Tehran, Iran

*Address all correspondence to: behzadipayam@yahoo.com

IntechOpen

References

[1] Hunter P. Laptop biology. EMBO Reports. 2005;**6**(3):208-210

[2] Penders B, Horstman K, Vos R. Walking the line between lab and computation: The "moist" zone. BioScience. 2008;**58**(8):747-755

[3] Gauthier J, Vincent AT, Charette SJ, Derome N. A brief history of bioinformatics. Briefings in Bioinformatics. 2019;**20**(6):1981-1996

[4] Moody G. Digital Code of Life: How Bioinformatics Is Revolutionizing Science, Medicine, and Business. London: John Wiley & Sons; 2004

[5] Ledley RS. Digital electronic computers in biomedical science. Science. 1959;**130**(3384):1225-1234

[6] November JA. Early biomedical computing and the roots of evidence-based medicine. IEEE Annals of the History of Computing. 2011;**33**(2):9-23

[7] Dayhoff MO, Ledley RS. Comprotein: A computer program to aid primary protein structure determination. In: Proceedings of the December 4-6, 1962, Fall Joint Computer Conference. New York, NY: ACM (Association for Computing Machinery); 1962

[8] International Union of Pure and Applied Chemistry (IUPAC)- the International Union of Biochemistry (IUB) Commission on biochemical nomenclature (CBN). A one-letter notation for amino acid sequences. Tentative rules. European Journal of Biochemistry. 1968;**5**:151-153

[9] Dayhoff MO. Atlas of Protein Sequence and Structure. National Biomedical Research Foundation. Washington, D.C.: Georgetown University Medical Center; 1972

[10] Crick FH. The origin of the genetic code. Journal of Molecular Biology. 1968;**38**(3):367-379

[11] Sanger F, Thompson E. The amino-acid sequence in the glycyl chain of insulin. 1. The identification of lower peptides from partial hydrolysates. Biochemical Journal. 1953;**53**(3):353

[12] Sanger F, Thompson E. The amino-acid sequence in the glycyl chain of insulin. 2. The investigation of peptides from enzymic hydrolysates. Biochemical Journal. 1953;**53**(3):366

[13] Staden R. A strategy of DNA sequencing employing computer programs. Nucleic Acids Research. 1979;**6**(7):2601-2610

[14] Kadri K. Polymerase Chain Reaction (PCR): Principle and Applications. Perspectives on Polymerase Chain Reaction. Croatia: IntechOpen; 2019

[15] Mullis KB, Faloona FA. Specific Synthesis of DNA in Vitro Via a Polymerase-Catalyzed Chain Reaction. Recombinant DNA Methodology. San Diego: Academic Press, Elsevier; 1989. pp. 189-204

[16] Pai-Dhungat J. Kary Mullis—Inventor of PCR. Journal of the Association of Physicians of India. 2019;**67**:96

[17] Bartlett JM, Stirling D. A Short History of the Polymerase Chain Reaction. PCR Protocols. Totowa New Jersy: Humana Press, Springer; 2003. pp. 3-6

[18] Zou D, Ma L, Yu J, Zhang Z. Biological databases for human research. Genomics, Proteomics & Bioinformatics. 2015;**13**(1):55-63

[19] Rigden DJ, Fernández XM. The 27th annual Nucleic Acids Research database

issue and molecular biology database collection. Nucleic Acids Research. 2020;**48**(D1):D1-D8

[20] Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Di Costanzo L, et al. RCSB Protein Data Bank: Biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. Nucleic Acids Research. 2019;**47**(D1):D464-DD74