The human genome sequence: a triumph of chemistry

If the history of molecular biology is written in the future, the first chapter should be devoted to the chemists who did the ground work • by Gáspár Jékely

In 2001 when the public consortium and Celera, its private competitor, both published their version of the sequence of the human genome, the then US President William J. Clinton and British Prime Minister Tony Blair heralded their efforts at a joint press conference as one of the most significant scientific projects of all time. This success was partly due to the fact that the Human Genome Project was—due to the finite length of human DNA—a project that could indeed be finished. Biology rarely has such projects. On the other hand, it was also obvious that the Human Genome Project was a major scientific achievement that biology and medicine will greatly benefit from for many years to come.

It is indeed tempting to think of these marvellous achievements as the crowning moment and a triumph of 20th century genetics. The often encountered, popular history goes like this: genetics was launched by Mendel, elaborated on by Morgan and later materialised itself in the form of the double helix. Molecular biology supplied new impetus and the science of genetics eventually culminated in the elucidation of the sequence of the human genome. It is rarely argued, however, that this sequence could also be viewed as the most magnificent feat of the science of chemistry.

In 1944, when Erwin Schrödinger's *What Is Life?* was published (Schrödinger, 1944), chemistry was already a mature science—in contrast to biochemical or physiological genetics, the molecular branch of genetics at that time. Its scope encompassed the quantum theory of the chemical bond, Albrecht Kossel's *Baustein* hypothesis—the idea that large polymers can be built from repeated units of similar nature—and Phoebus Levene's repetitive tetranucleotide model of DNA. Not long before, chemists had also generally accepted the theory of macromolecules, according to which large polymers with terrifying, unheard of molecular weights can be held together by covalent bonds. However,

with Schrödinger's prophetic suggestion of an 'aperiodic crystal' as the basis of hereditary information, the field of biochemistry and crystallography of biological macromolecules received new momentum. This had a phenomenal impact: many eminent scientists were stimulated to solve the problem of the autocatalytic hereditary material, the transforming principle, or the 'Riddle of Life', as Max Delbrück put it (Hunter, 2000).

The same year that Schrödinger's book was published, Oswald Avery and his co-workers proved that the transforming principle—the chemical substance that is able to transform a non-virulent strain of *Pneumococcus* into a virulent form—was equivalent to DNA. Although not properly recognised at that time, this was a major breakthrough. It meant that rather than exciting proteins being Schrödinger's aperiodic hereditary crystals as was generally believed, the harbinger of genetic

genetics began with Avery. Although the learning process was slow and weary, Chargaff himself immediately realised the importance of Avery's work and concentrated his efforts on DNA. His detailed chemical analysis of its composition disproved the tetranucleotide model. He also showed that the molecule is an aperiodic polymer, thereby removing the last obstacle to its acceptance as the hereditary material. As this notion became more established, it also became obvious that solving the riddle of life meant solving the three-dimensional chemical structure of DNA.

This task was not easy and again, chemistry supplied the key insights. As early as 1940, Linus Pauling, the leading structural chemist at the time, already explicitly stated the idea of chemical complementarity as the basis of gene replication. His statement 'the conditions under which complementarity and identity might coincide'

information was DNA, this seemingly boring chromosomal scaffolding molecule.

Indeed, Erwin Chargaff (1979) later remarked that the chemical education of

anticipated the anti-parallel, base-paired structure of DNA. In addition, the α -helical configuration of proteins, also elucidated by Pauling, provided further clues for

those determined to unravel the structure of nucleic acids.

James Watson was well aware of the need for a chemical approach, as exemplified in his complaint about the futility of phage work in *The Double Helix* (Watson,

sequencing polymerases and capillary gel electrophoresis. Ultimately, these improvements, again fertilised by organic chemistry, enabled the sequencing of the 3.2 billion base pairs that represent the human genome. Strong demands obviously promote new

a bid to ensure its continuation. One can stress the political aspect and the various agendas underlying the project, or concentrate on the scientific and economic needs and rationales that prompted it. But when the history of the Human Genome Project is written from a scientific point of view, recalling the original concepts behind it and the various advances and discoveries made on the way, chemists are the main players of this vital first stage. In the next chapter, it will fall to the geneticists, or *genomicists* if the former sounds too obsolete, to make sense of all the information and solve the myriads of secrets that genomes harbour.

Watson probably never became a chemist; nevertheless, his and Francis Crick's work on the structure of DNA was a masterpiece of stereochemistry

1980). 'It was obvious that I had not done anything which was going to tell us what a gene was or how it reproduced. And unless I became a chemist, I could not see how I would,' he wrote. Watson probably never became a chemist; nevertheless, his and Francis Crick's work on the double-helical structure of DNA was a masterpiece of stereochemistry. 'It is a fortunate fact that amateurs often are better in advancing science than are professionals,' Chargaff commented (Chargaff, 1979).

After the genetic code was finally cracked and it was revealed how nucleic acids could be translated into proteins, DNA became central to biological research. Researchers grew more and more interested in working out the precise nature of its aperiodicity—the sequence of genes. Alan Maxam and Walter Gilbert provided the first method for doing so by chemically cleaving DNA into fragments that could be analysed to infer the nucleotide sequence. This was later replaced by Frederick Sanger's chain-terminating dideoxy method that made ingenious use of dideoxynucleotides, oligonucleotide primers and DNA polymerase. Once the logic and basic technique of DNA sequencing had been established, there was ample space for technological improvements, which have since included the development of fluorescent chain-terminators that enabled four-colour sequence detection, design of new

technologies; on the other hand, technologies incite new demands. In either case, the importance of the massive input from chemistry, here again, cannot be underestimated.

In retrospect, it appears that arriving at the complete sequence of the human genome only required the application of these ideas in conjunction with the

The uncovering of the complete structure of the giant DNA molecule has been, in essence, heroic and wondrous chemistry

auxiliary concept of genetic mapping and an exponential amplification of technical resources and expertise. The sequence of the human genome, as well as animal, plant and bacterial genomes, represents the largest, non-repetitive chemical molecules ever deciphered by science. To understand how this vast amount of genetic information is set into motion is a task for biology. But the uncovering of the complete structure of these giant molecules has been, in essence, chemistry—heroic and wondrous chemistry.

History can be told in many ways, particularly the story of huge research projects, such as the Human Genome Project. One might give special emphasis to the agencies that funded the research in

References

- Chargaff, E. (1979) How genetics got a chemical education. *Ann. N. Y. Acad. Sci.*, **325**, 345–360.
Hunter, G.K. (2000) *Vital Forces; The Discovery of the Molecular Basis of Life*. Academic Press, London, UK.
Schrödinger, E. (1944) *What is Life?* Cambridge University Press, Cambridge, UK.
Watson, J.D. (1980) *The Double Helix*. W.W. Norton, New York, NY.

Gáspár Jékely is a post doctoral researcher at the European Molecular Biology Laboratory in Heidelberg, Germany, and a molecular biologist. E-mail: jekely@embl-heidelberg.de

DOI: 10.1093/embo-reports/kvf145