MOLECULAR MEDICINE; THE ROAD TO THE BETTER INTEGRATION OF THE MEDICAL SCIENCES IN THE TWENTY-FIRST CENTURY

by

D. J. WEAVERALL*

Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DS, UK

The history of the evolution of medical research is characterized by a long period of division between the basic biological sciences and the health sciences, reflecting the seventeenth-century trends towards the experimental and the empirical. It was not until the middle of the twentieth century that, by their ability to straddle both worlds, the work of small groups of basic scientists in the USA and Europe led to closer integration between the medical sciences. This change in direction is well exemplified by the work of the Cambridge scientists Max Perutz, Vernon Ingram and Herman Lehmann, from 1950 onwards. Their research, and its later development by others, was to lay the basis for what became known as ‘molecular medicine’, and at the same time set the scene for a more integrated approach to medical research that continued into the new millennium.

Keywords: medical science; haemoglobin; genetic disorders of haemoglobin; molecular medicine

Although three of the dozen natural philosophers who met regularly to discuss their research in Oxford in the mid seventeenth century and later formed the nucleus of the Royal Society were physicians, the Society’s involvement in the health sciences has gone through a long period of evolution. In 1897 it established the Buchanan Medal, given quinquennially in recognition of distinction in the broad area of medical research, but it was not until 1997 that a Sectional Committee for the election of Fellows was established with the title ‘Health and Human Sciences’.

The history of this divide between the basic biological and health sciences is almost as long as that of the Society itself. Although in part it may reflect what Peter Medawar described as that ‘terribly, terribly English’ class distinction between pure and applied science,1 the reasons go much deeper, reflecting the changing patterns of medical education and practice over many centuries. Shortly after the end of the English Civil War, Thomas Sydenham, who was later to have an enormous influence on the practice of medicine and who became known as the English Hippocrates, arrived in Oxford to study medicine. At that time medical education in Oxford was dominated by some of the medical polymaths who were among the founders and early Fellows of the Royal Society,
including Thomas Willis, Richard Lower and John Locke. Sydenham was not impressed with the Oxford scene, and one of his contemporaries records that he thought it better to send a man to Oxford to learn shoemaking than medicine! He believed that little could be learnt by animal experimentation, dissection of the human body or the use of the microscope. Rather, he emphasized the primacy of the doctor–patient relationship, stressed the acquisition of observational skills at the bedside and, above all, regarded doctoring as a pragmatic art; his ghost still walks the corridors of our medical schools today.

These four distinguished Oxonians, among the most outstanding in the history of British medicine, were at the vanguard of the conflicting seventeenth-century trends towards the experimental and the empirical. It has been suggested that it was Locke’s medical training during his later association with Sydenham that influenced him in his formulation of the empirical movement of modern philosophy in his *Essay Concerning Human Understanding*, later described by Gilbert Ryle as ‘not a theory of knowledge but a theory of the sciences’.

The current organization of medical science and education is still broadly based on a report written by the American educationalist Abraham Flexner in 1910 at the request of the Carnegie Foundation for the Advancement of Teaching. Flexner had travelled widely in Germany in the latter part of the nineteenth century and had been influenced by the strong basic science departments that were evolving and by the integration of medical science and education into the university system. He was not impressed with what he saw in the USA and recommended that a similar pattern to that in continental Europe be adopted. In short, medical education should be fully integrated into a university system with strong basic biological science departments in which students spent their first few years, followed by clinical instruction in departments led by full-time professors in which there was a critical and research-based approach to clinical practice. He later conducted a similar review in Great Britain, but apart from Cambridge, where he was enthusiastic about the strengths of its evolving science departments, he was equally unimpressed. However, in 1913, and despite strong opposition from the medical establishment, a Royal Commission chaired by Lord Haldane ruled that Flexner’s principles should be applied in this country. But as this pattern evolved there was little integration or mutual respect between the basic and health sciences. These tensions, and some of the complex reasons for them, are well illustrated in the history of the development of basic medical research and practice at Cambridge University throughout the nineteenth and early twentieth centuries. And they were not confined to the older universities: on returning from the USA to Liverpool University in 1964 I was asked by the distinguished Professor of Physiology R. A. Gregory FRS to give a series of lectures on the blood because his department was rather light in that area. I subsequently discovered that the first of the series was taped so that Gregory could reassure himself that it was safe to let a clinician loose on his students. When I moved to Oxford 10 years later it was clear from the attitude of the heads of the biological science departments that they shared Gregory’s concerns.

After the middle of the twentieth century, and the movement of physicists such as Erwin Schrödinger and Francis Crick to the biological sciences, there was also a gradual change in the relationship between the latter and the health sciences. A good example of this trend is reflected in the slow evolution of what became rather optimistically called ‘molecular medicine’. In this essay I shall briefly discuss the work of three distinguished Fellows of the Royal Society—Max Perutz, Vernon Ingram and Herman Lehmann—whose
remarkable ability to straddle the basic and health-based aspects of research had a major impact on the later development of the field and, at the same time, on the closer integration of the medical sciences.

**THE ORIGINS OF MOLECULAR MEDICINE**

The origins of molecular medicine can be traced back to the work of Archibald Garrod at the beginning of the twentieth century.\(^6\) In 1899 and 1901 he published accounts of a disease called alkaptonuria, characterized by joint swelling and the passage of dark pigment in the urine, pointing out that it occurred in siblings and that there was a high frequency of consanguinity in their parents. Although he did not realize the significance of these findings at first, William Bateson, one of the leading supporters of Mendel’s work at the time, pointed out to him that they were just what might be expected in a disease that had a recessive form of Mendelian inheritance. Garrod continued to study rare disorders of this type and in June 1908 presented this work in his Croonian Lectures to the Royal College of Physicians under the title ‘Inborn errors of metabolism’.\(^7\) In a later and long forgotten work, ‘Inborn factors in disease’, he emphasized human biochemical individuality and its interplay with the environment as the basis for disease.\(^8\) His work was largely disregarded and medical genetics only became part of medical practice after World War II.

In the years after Garrod’s seminal work, several diseases were discovered that seemed to be inherited. In 1910, the American physician James Herrick described a young student from the West Indies with chronic anaemia and curiously elongated and sickle-shaped red blood cells, which he named ‘sickle-cell anemia’.\(^9\) And in 1925 Thomas Cooley in America and Fernando Rietti in Italy quite independently described a severe form of anaemia in patients of Mediterranean origin that later was called ‘thalassaemia’, a word derived from the Greek roots for ‘sea’ and ‘blood’.\(^10\) It was found later that both these conditions are inherited in a Mendelian recessive fashion and that they are by far the commonest monogenic diseases and not restricted to those of African or Mediterranean origin.\(^10\)

The discovery of the cause of sickle-cell anaemia in 1949 was the result of a chance conversation between Linus Pauling and the Boston haematologist William Castle. Castle told Pauling that he had been examining the blood of patients with sickle-cell disease and had noticed that when their red cells were deprived of oxygen and formed a sickle shape they showed unusual properties when examined under polarized light. Pauling realized that this could reflect a form of molecular reorganization and therefore that it might indicate that the haemoglobin of patients with sickle-cell disease was in some way abnormal. Together with his students, he examined the pattern of haemoglobin of patients with sickle-cell disease and compared it with that of normal persons by moving-boundary electrophoresis with a Tiselius apparatus. They found that the haemoglobin from those with sickle-cell disease migrated at a different rate from that of normal persons and, even more importantly, that the red cells of healthy parents (or carriers) of the sickle-cell gene contained both normal and abnormal haemoglobin. The discovery that sickle-cell anaemia results from an inherited abnormality of the structure of haemoglobin was published in *Science* in November 1949 under the title ‘Sickle-cell anemia, a molecular disease’.\(^11\)
The story now moves to Cambridge and the work of three scientists, all of whom had emigrated to England from Germany before World War II. Max Perutz (figure 1) was born in Reichenau; in 1936 he moved to the Cavendish Laboratory where, apart from a period of internment in Canada as a ‘friendly alien’, he was to spend the rest of his career, later becoming head of the Molecular Biology Unit of the Medical Research Council and, from 1962 to 1979, Chairman of the Council’s Laboratory of Molecular Biology. Vernon Ingram (figure 2) was born in Breslau and after moving with his family to London he obtained his early training at Birkbeck College. After postgraduate experience in the USA, in 1952 he returned to England to join Max Perutz at the Cavendish Laboratory. Herman Lehmann (figure 3) was born in Halle and after obtaining a medical degree and studying with Otto Meyerhof in Heidelberg he moved to Cambridge, where he obtained a PhD working with Gowland Hopkins and first met Max Perutz. At the beginning of the war he shared the pleasures of internment at Huyton near Liverpool with Perutz; however, unlike the latter he was not sent to Canada but was recruited by the British army. He spent part of his service in Africa, where he first became aware of sickle-cell anaemia. After the war, and a short spell working in London, he moved to Cambridge, where he developed a close collaboration with Perutz and became the university’s first Professor of Clinical Biochemistry in 1967.

In a series of papers published from the 1950s onwards, Perutz and his colleagues produced beautifully elegant X-ray crystallographic evidence of the three-dimensional structure of haemoglobin, work that had already occupied him for more than 30 years and for which in 1962 he received the Nobel Prize in Chemistry. Over the same period simpler laboratory methods for haemoglobin electrophoresis were developed, and workers
Figure 2. Vernon Ingram (1912–2009).

Figure 3. Herman Lehmann (1910–85).
all over the world began to examine blood samples to find new variants. Lehmann, who had pioneered some of these methods, became a leader in this field and travelled widely in search of other haemoglobins. Many of the variants were harmless and had no effect on the protein’s function or stability. In rare cases, however, a mutation had profound effects on its oxygen-carrying properties or stability. During the daunting task of trying to relate the structure of haemoglobin to its function, Perutz became very interested in these rare variants and, like Lehmann, developed several clinical collaborations. He was later to recall how valuable these ‘experiments of nature’ were in the final elucidation of the structural details of the haemoglobin molecule that underlie its stability and allosteric properties as an oxygen carrier. This work was summarized in 1968 in a classical paper by Perutz and Lehmann on the molecular pathology of haemoglobin.13

When Ingram joined Perutz in 1952 his first project was to help to place a ‘heavy atom’ in a specific location in Perutz’s haemoglobin crystals. By then it was clear that haemoglobin consists of two identical half-molecules, αβ, and hence its structure can be represented as α2β2. Each α and β peptide chain encloses haem, a complex porphyrin ring structure containing iron. Ingram’s next task was to study the interactions of particular amino acids of the globin chains with haem. However, as he wrote later, a moment of serendipity arrived.14 A previous worker in the laboratory had left a sample of sickle-cell haemoglobin behind him and, at the suggestion of Perutz and Crick, he set out to determine its structure by adapting some of the methods that Frederick Sanger had used to study the structure of insulin. Of course this presented considerable problems because the αβ half-molecule of haemoglobin is 10 times the size of either peptide chain of insulin. Ingram therefore decided to take a slightly different approach. In short, he digested the haemoglobin into smaller peptide fragments with trypsin and then separated them in two dimensions by electrophoresis and chromatography, a technique that he called ‘fingerprinting’. The results showed that the ‘fingerprint’ of normal haemoglobin, HbA, differed in the position of only one peptide from that of sickle-cell haemoglobin (HbS). Subsequent amino acid analysis demonstrated that this reflected a single amino acid substitution: a glutamic residue in HbA is replaced by valine in HbS.15 The substitution occurs in the β chain of haemoglobin, suggesting that it must result from a mutation in the globin gene that regulates the synthesis and structure of the β-globin chain. The significance of this work, which never received the recognition that it warranted, went considerably further than the first identification of the molecular basis for a genetic disease. In 1958, Beadle and Tatum received the Nobel Prize in Physiology or Medicine for demonstrating from their much earlier work on Neurospora that there is a direct relationship between a gene and an enzyme, the ‘one gene – one enzyme’ principle;16 now Ingram had extended this concept to the ‘one gene – one peptide chain’ level. In short, sickle-cell haemoglobin results from a single amino acid substitution in the β chain of haemoglobin, reflecting a mutation of the β-globin gene.

Together with his students and with the help of Lehmann, Ingram analysed a variety of other abnormal human haemoglobins and was able to demonstrate the different amino acid substitutions that were involved. At the same time he became interested in the problem of whether thalassaemia might also result from defective haemoglobin synthesis. By the 1950s, although it was clear that this was a severe form of anaemia that was inherited in a Mendelian fashion, nothing was known about its underlying cause. However, at about this time reports appeared of patients who had inherited a gene for thalassaemia from one parent and the sickle-cell gene from the other, a condition that was
called ‘sickle-cell thalassaemia’. Remarkably, unlike those who were carriers of the sickle-cell mutation, who produced considerably more HbA than HbS, those with sickle-cell \( \beta \) thalassaemia showed a reversal in this ratio, having significantly more HbS than HbA. Furthermore, in matings between those with sickle-cell \( \beta \) thalassaemia and normal persons, the offspring showed either the carrier state for HbS or \( \beta \) thalassaemia; they never were normal or had sickle-cell \( \beta \) thalassaemia. These findings suggested that the \( \beta \) thalassaemia and sickle-cell mutations are alleles—that is, they occur on homologous pairs of chromosomes—and that the disease might be due to defective \( \beta \)-chain production. However, patients were also encountered with forms of sickle-cell thalassaemia in which this change of ratio did not occur; that is, the amount of HbA and HbS appeared in the usual carrier ratio. Could this be a form of thalassaemia that affected \( \alpha \)-chain synthesis and would therefore affect the output of both \( \beta^A \) and \( \beta^S \) chains equally? These observations were discussed at an important meeting held in 1959, and at the end of 1959 Ingram and one of his students, Tony Stretton, wrote a theoretical review of the thalassaemia field in *Nature* entitled ‘Genetic basis of the thalassaemia diseases’. They reviewed the evidence which suggested that the thalassaemias might result from a defective rate of production of the \( \alpha \)-globin and \( \beta \)-globin chains of haemoglobin and proposed some mechanisms whereby these defects might be mediated. Although this synthesis reflected the ideas of many workers in this rapidly moving field, including Lehmann, a fact that was freely acknowledged at the end of their paper, it undoubtedly set the scene for the rapid progress that was made towards our understanding of the molecular basis of the thalassaemias over subsequent years. Shortly after completing this work, Ingram returned to the USA and moved into other fields of interest.

Although others in the USA and Europe determined the structure of the globin chains that constitute the different human haemoglobins that are found in fetal and embryonic life, and made important contributions to the beginnings of an understanding of the genetics of haemoglobin production, there is no doubt that the work of these three scientists in Cambridge, particularly that conducted between 1950 and 1960, played a central role in the evolution of the human haemoglobin field. In each case they were able to focus their skills in the basic biological sciences towards their potential clinical application and hence to act as an inspiration to the next generation of scientists, both clinical and basic, to move in the same direction. With similar trends in the USA it was the beginning of a much closer integration between the basic and health sciences.

**FURTHER PROGRESS, 1960–2000**

From 1960 onwards the haemoglobin field evolved slowly on the back of technical developments in protein chemistry and molecular biology. It was driven largely by a new breed of clinical scientists who had trained in these fields or who at least had become sufficiently acquainted with them to pose appropriate questions to their basic science colleagues.

In 1965 methods were developed for studying the synthesis of haemoglobin in systems *in vitro* and it was found that the thalassaemias are characterized by defective \( \alpha \)-globin or \( \beta \)-globin chain synthesis, and that rather than being disorders of haemoglobin production they are characterized by unbalanced synthesis of globin chains: in \( \beta \) thalassaemia, as a result of defective \( \beta \) chain production excess, \( \alpha \) chains are produced; the converse occurs in \( \alpha \) thalassaemia. It became apparent that the severe defects in red cell production and
survival are mediated by damage caused by excess α or β chains on red cell precursors or mature red cells. It was soon clear that there are many different forms of thalassaemia and in two cases it was possible to predict the likely molecular defects on the basis of the structure of haemoglobin variants that were associated with some of these conditions.10

By 1970 methods had become available for isolating messenger RNA from red cell precursors and later for synthesizing complementary DNA (cDNA) probes by using messenger RNA as a template. In 1974 groups in Great Britain and the USA, using cDNA as a probe, were able to demonstrate that the most severe form of α thalassaemia results from a deletion of the α-globin genes.22,23 In both cases these experiments were designed by clinical scientists and performed with the technical help of molecular biologists with expertise in these new hybridization techniques. In the mid 1970s the technology for cloning and sequencing individual genes was developed, and by the early 1980s it had become possible to begin to characterize the specific mutations that underlie many different forms of thalassaemia. Within 10 years more than 100 different mutations had been defined and a picture quickly emerged of the remarkable heterogeneity of genetic lesions that underlie diseases with a Mendelian pattern of inheritance.10

It was of course no accident that the haemoglobin field led the way to the characterization of disease at the molecular level. Because haemoglobin constitutes more than 98% of the protein of red cells it was a natural target for early studies of the structure of proteins and, similarly, because red cell precursors contain relatively large amounts of globin messenger RNA they were an obvious target for the growing technology of molecular biology. However, the situation was much more complex for other monogenic diseases. As early as 1948, J. B. S. Haldane had suggested that an approach to defining the location of genes of this type might be to study affected families in which there were other genetic markers available, for example blood groups.24 If the gene for the particular disease was inherited together with a marker of this type, and the location of the marker was known, the location of the defective gene could be established. The problem was that very few markers of this type were available. After the discovery of restriction enzymes—bacterial enzymes that cleave DNA at specific sequences and, in particular, enable a simple approach to the analysis of the size of the resultant DNA fragments25—it was found that human DNA is remarkably heterogeneous with respect to its structure. In short, it contains many sites at which there is variation in base structure, thus producing new cleavage sites or removing pre-existing ones, namely restriction fragment length polymorphisms (commonly known as RFLPs). In 1978 it was found that there is always a polymorphism of this type close to the β-globin gene that carries the sickle-cell mutation.26 Therefore a search began to find linkage markers of this type in other monogenic diseases. Once the chromosomal location of a particular gene had been defined in this way, the next problem was to isolate the gene itself. Although the necessary technique, which became known as ‘positional cloning’, took more than 10 years to develop, the gene for cystic fibrosis was defined in 1988 and genes for many other monogenic disorders were discovered during the next 20 years.24

Over this period, with each advance in an understanding of the molecular pathology of disease there were new clinical applications. For example, within 10 years of the discovery of a method for measuring globin chain synthesis it was applied to the prenatal detection of different forms of thalassaemia.27 Similarly, as these diseases became amenable to identification at the DNA level, within just a few years this technology was applied for the prenatal detection of the thalassaemias and sickle-cell disease very much
earlier in pregnancy.\textsuperscript{28,29} Similar approaches were also developed for the more effective
counselling and prenatal detection of other monogenic diseases.

Not surprisingly, there was also a growing interest in the applications of molecular
biology to many other conditions, ranging from infectious disease to cancer. The
advances in the latter field are a particularly good example of the increasing interplay
between the medical sciences towards the end of the twentieth century. They reflected
work in basic molecular and cell biology, clinical genetics, clinical oncology, population
genetics and epidemiology that saw the beginnings of an understanding of the complex
interplay between the genome and the environment that underlies the genesis of
malignant disease.\textsuperscript{30} Thoughts were also turning to whether it might be possible, by using
linkage technology, to define the genetic components of other common conditions such as
heart disease, stroke and diabetes.

All this activity posed several problems, not the least of which was where young
clinicians who wished to be trained in the technology of molecular biology could find an
environment that encouraged its clinical application. Similarly, there was nowhere that
young molecular biologists who wished to apply their skills in a clinical field could be
trained. In an attempt to solve this problem, in the mid 1980s it was suggested that an
Institute of Molecular Medicine be established in Oxford where young clinicians who
wished to be trained in molecular biology could work in different areas of medical
research and where young molecular biologists could work alongside them, directing their
skills towards clinically based problems. Particularly through the vision and support of
James Gowans, who at that time was Secretary of the Medical Research Council, this
Institute was opened in 1989. Since then, similar developments have occurred in many
medical schools throughout the UK and elsewhere.

\textbf{THE RECENT PAST AND FUTURE}

The culmination of this remarkable period in the evolution of the biological and medical
sciences was undoubtedly the announcement of the completion of the human genome
project in 2003. This was followed by predictions that medical practice would be
completely changed within the next 20 years. The causes of many of our common killers
would be rapidly disclosed, therapeutics would become personalized on the basis of our
genetic make-up, and our lifespans would be further increased by knowledge gained from
the molecular and cellular mechanisms of ageing. But it was never likely that things
would be as simple as this. Even a disease such as thalassaemia, which can be ascribed to
the action of a single defective gene, shows wide variation in its clinical severity between
patients with an identical genetic defect, resulting from the interaction of layers of
modifier genes, individual differences in ability to adapt to profound anaemia, and
complex interplay with the environment.\textsuperscript{10} It soon became apparent that many of our
common killers reflect the action of many different genes, each with a relatively small
role, combined with the effects of the environment.\textsuperscript{31} In addition, because we have far
fewer genes than was originally predicted, there was now the daunting task of trying to
understand the extreme complexities of how they are regulated, both in health and in disease.

Over the whole of this period, the Royal Society has taken an increasing interest in these
new developments in the health sciences and related fields. As judged by its policy reports,
statements and responses since 1995, a very broad field has been covered, ranging from the
ethical issues raised by recent advances in molecular and cell biology to the hopes and realities of personalized medicine. Its further interactions with the Academy of Medical Sciences undoubtedly have the potential to increase its influence on the health sciences, mirroring the highly successful partnership between the National Academy of Sciences and the Institute of Medicine in the USA.

It therefore seems that there has been a steady integration of the basic biological sciences with the health sciences over the past half-century. Very recently, and probably reflecting the rather over-optimistic predictions of the timing of the medical applications that would follow the Human Genome Project, there has been a mood of disillusionment over the value of research in molecular and cell biology for the betterment of health. This has resulted in a call from some quarters for the funding of medical research to be diverted from the field of molecular medicine to the study of more immediate problems in the clinic. Overnight, the term ‘translational research’ has become a prerequisite for approaches for financial support for medical research.

It would be a pity if this impatience with the rate of progress in the more fundamental aspects of the medical sciences, with such enormous potential, were to set back the advances that have been made in their better integration over recent years, particularly because our knowledge of fundamental biological processes, let alone their breakdown in disease, is still so flimsy as to make it difficult to define what is ‘translatable’. After the first landing on the Moon the then President of the USA, perhaps conscious of an impending election and the large African-American population with sickle-cell disease, demanded that a large amount of money be directed at finding a cure. Although something was learnt about the sickling process, none of it was of any value whatever to patients; it was simply too early in the evolution of knowledge about the complex biology of the disease and of the regulation of haemoglobin production in general to yield results of direct advantage to patients. Surely we have now outlived the divide between the basic biological sciences and the health sciences. As George Porter, a former President of the Royal Society, put it, ‘there are only two kinds of science, applied and not yet applied’.32

ACKNOWLEDGEMENTS

I thank Jeanne Packer and Liz Rose for their help in preparing this paper, and Beth Ingram and Lisa Steiner for the photograph of Vernon Ingram.

NOTES


