Molecular Medicine

Anthony Cerami Award in Translational Medicine

A Journey in Science: Early Lessons from the Hemoglobin Field

David J Weatherall

Weatherall Institute of Molecular Medicine, Oxford University, Oxford, United Kingdom

Real innovations in medicine and science are historic and singular; the stories behind each occurrence are precious. At *Molecular Medicine* we have established the Anthony Cerami Award in Translational Medicine to document and preserve these histories. The monographs recount the seminal events as told in the voice of the original investigators who provided the crucial early insight. These essays capture the essence of discovery, chronicling the birth of ideas that created new fields of research; and launched trajectories that persisted and ultimately influenced how disease is prevented, diagnosed, and treated. In this volume, the Cerami Award Monograph is by David J Weatherall, Founder, Weatherall Institute of Molecular Medicine, Oxford University, John Radcliffe Hospital. A visionary in the field of hemoglobin, this is the story of Professor Weatherall's scientific journey.

INTRODUCTION

This short review describes the author's experience in the early days of study of the inherited disorders of hemoglobin and how many of the lessons learned from this field still have important implications for the further development of molecular medicine. It also offers some advice on the basis of the author's experience for young medical or basic science students who are contemplating a career in this rapidly expanding field.

EARLY DAYS

During my 6 years as a medical student at Liverpool University in the early 1950s, genetics was rarely mentioned and med students' exposure to hematology was extremely scanty, a situation that changed dramatically during the first years of my internship. My mentor during this period was Cyril Clarke, a remarkable general internist with a special interest in asthma. His lifelong hobby was collecting butterflies, and he and his wife had developed the unusual skill of being able to hand-mate them. Together with Philip Sheppard, an Oxford geneticist who later became the first Professor of Genetics in Liverpool, they carried out seminal studies on the evolution of mimicry in the swallowtail butterfly *Papilio machaon* (1). During the early 1950s, Clarke began to wonder if his work on genetics was applicable to human disease, and Sheppard suggested to him that, to move in this direction, it would be best to start with human blood groups, since they were by far the easiest genetic phenotypes to follow. While I was working for Clarke, he became deeply involved in studies relating particular blood groups and secretor status to several diseases, notably duodenal ulcer. Later his interests turned to the Rhesus (Rh) blood group system, and his team was the first to discover a highly successful approach to the prevention of Rh hemolytic disease of the newborn. A similar discovery was made quite independently by John Gorman's group in New York at the same time (2). I was extremely fortunate to be trained by Clarke, who left me in no doubt of the future importance of genetics in medical research and practice at a time when this field was still in its infancy.

In the late 1950s, newly qualified doctors in the UK who had completed their internships still had to do a compulsory 2 years of military service. Because the

army had mislaid my address, I had spent a second year after internship obtaining further clinical experience. This scenario led to my obtaining a higher degree in internal medicine, which, I was told, meant I would carry out my military service in an army hospital. Being terrified of bullets, snakes, and, in particular, flying, I volunteered to serve in a hospital in England. Three weeks later, I was on a troopship bound for Singapore. Between 1948 and 1960, during the Emergency—the war in Malaya (as it was then called) between the Chinese communists and the Commonwealth Forces—the Alexandra Hospital in Singapore acted as the main center for casualties who could not be managed in the smaller hospitals in Malaya and for the families of the Commonwealth Forces. Because the only pediatrician in the army was delayed in Berlin, I was asked to look after the pediatric ward. One of the first patients I encountered was a young Nepalese girl named Jaspir, whose father was a sergeant in a Gurkha regiment and had brought his family down from Nepal when he was posted to Malaya during the emergency. I have told the full story of this remarkable child elsewhere (3). In short, while still very young, Jaspir had arrived with her family in Malaya and was found to be profoundly anemic and had to be maintained on regular blood transfusions; no diagnosis had been made. Later, when her father was posted down to Singapore, she was looked after at the Alexandra Hospital. I was very puzzled by this child's clinical condition because it did not seem to fit in with any of the standard causes of profound anemia in childhood.

At that time, I spent some of my spare hours in Singapore visiting the medical staff at the University Hospital, and it was there that I met a biochemist named Frank Vella. Vella had developed an interest in hemoglobin electrophoresis to try to discover hemoglobin variants and related disorders. A method for defining the levels of different hemoglobin fractions using starch-block electrophoresis

had just been published in the United States (4). Together, we spent some time, both covered in starch, developing this approach, and when we studied the hemoglobin pattern of Jaspir and that of her parents, we found that they were compatible with the diagnosis of thalassemia. This discovery was extremely exciting because, as evidenced by its curious name, which is based on the Greek roots for blood and the sea, it was thought at the time to be largely confined to the Mediterranean region. For this reason, we published an account of this family in the *British Medical Journal* in 1960 (5).

Seeing one's first paper in press is an exciting experience, often followed by various delusions, including the possibility of a telephone call from Stockholm. But when the phone rang, it was not from Stockholm, but from the office of the Director General of the Southeast Asian Land Forces. When I entered his office, he had my paper spread out before him on his desk. "Did you write this paper?" he asked. Since my name appeared as the first author, denial seemed unwise. "Did you get permission from the War House?" he asked. Not knowing what the War House was, and only finding out later that it was the War Office in London, I said no. "Did you know that you could be court-martialed for publishing information about our military staff without permission? Do not do it again." And he went on to add that it was bad form to tell the world that members of one of our greatest regiments carried bad genes. This was not, I thought, the most promising start to a research career.

I spent the second half of my military service in Taiping in North Malaya looking after the general medical and pediatric beds. At that time, the Commonwealth Forces were engaged in the final months of the emergency and in fighting with the remaining Chinese communists who were trying to flee over the border into Thailand. This period gave me wide experience in the management of different tropical diseases and, in my spare

time, I looked for more cases of patients with thalassemia or abnormal hemoglobins in the Malay population. There were no facilities for hemoglobin electrophoresis in the military hospital, and I had to construct my own device by using a couple of car batteries and some filter paper. When I found something abnormal, I posted a sample to Herman Lehmann in London, one of the pioneers in the discovery of abnormal hemoglobins.

Shortly before I finished my military service, I wrote to my mentor Cyril Clarke and said that, when I returned to England, I wanted to find a job where I could get experience with genetics, hematology and protein chemistry. He replied that if I tried to find something like this in the UK, I would be immediately referred for psychiatric assessment and that I had better go to the United States. He kindly assisted me to find a postdoctoral position at Johns Hopkins Hospital in Baltimore, Maryland. On the way through London on my return to the UK, I visited Herman Lehmann to thank him for his help. His advice to me was that there was nothing left to do in the hemoglobin field and I had better turn to red cell enzymes for my research program. Fortunately, I ignored this suggestion.

THE YEARS AT JOHNS HOPKINS UNIVERSITY AND THE BEGINNINGS OF MOLECULAR MEDICINE

In Baltimore, I was extremely fortunate to work under the care of several remarkable mentors including geneticist Victor McKusick, hematologist Lockard Conley and biophysicist Howard Dintzis. My first work was carried out in the biochemistry laboratory in the Moore Clinic with Samuel Boyer. By 1960, it was known that human hemoglobins are characterized by different pairs of globin subunits. Adult hemoglobin has two α chains and two β chains ($\alpha_2\beta_2$), while fetal hemoglobin has α chains and γ chains ($\alpha_2\gamma_2$). What was not known was whether the α chains of adult and fetal hemoglobin are regulated by the same locus or whether there are distinct fetal and adult α chain loci. This question was

clearly important because, in the former case, α chain variants would give rise to abnormal hemoglobins in both fetal and adult lives, whereas in the latter case, the variants would be restricted either to fetal or adult life only. Answering this question entailed a long search round the Baltimore hospitals for babies with abnormal α chain variants and appropriate studies of the hemoglobin constitution of their families. Within a few months, I was fortunate enough to find several affected families. By developing a simple hybridization technique to determine whether the variants involved the α or β chains, I soon found that the same set of α globin genes regulate α chain synthesis in adult and fetal lives (6).

During my analysis of cord bloods in search of fetal hemoglobin variants, I noticed that quite a high proportion of babies born in Baltimore had a rapidly migrating hemoglobin. I later identified this hemoglobin as Hb Bart's, a variant that had been discovered a few years earlier by Herman Lehmann in London and that consisted of a tetramer of γ chains (γ_4). Since this structure must reflect a decreased level of α chains, there may be a common mild form of α thalassemia in individuals of African origin. But no abnormal tetramers occurred in their parents, although many of them showed mild morphological changes of their red cells compatible with a very mild form of α thalassemia (7). At that time, it was not possible to study the structure of the α genes. While I was carrying out this work, and shortly before my return to the UK at the end of my first period in Baltimore, I helped Conley to further define the fascinating disorder, hereditary persistence of fetal hemoglobin, and its interactions with sickle cell anemia (8).

By 1962, my first fellowship at Johns Hopkins was complete, and I returned to the UK and put together the research that I had been doing. I presented this research as a doctoral thesis for my MD qualification. The external examiner was an extremely distinguished neurologist. It was clear from his questions that he had not understood a word of my thesis,

and he ended up my interrogation by saying that the work was all very well but that a bright chap like me should move into psychiatry as soon as possible. Since I had already been invited back to Baltimore to continue my research, I was happy to turn down yet another piece of career advice.

On returning to Baltimore, I was convinced that little progress would be made in the thalassemia field unless it were possible to analyze hemoglobin synthesis in these diseases by an *in vitro* system. At that time, several groups were studying hemoglobin synthesis by using isotopic labeling of hemoglobin in the reticulocytes of rabbits that had been made anemic by injections of phenylhydrazine. However, even in the days when ethics committees were much less active than they are today, it would not have been possible to inject humans with phenylhydrazine. I therefore had to carry out the labeling experiments on blood samples obtained from patients with a hemolytic anemia. I spent a long time working out the conditions in which I could follow linear synthesis of hemoglobin in this way, so that my assay would not suffer from artifacts due to dying red cells. The next problem was how to separate the globin chains of hemoglobin. I tried several methods and, although I obtained unequal labeling of the α and β chains, none of them were quantitative.

I then had a lucky break. Because the hematology department at Johns Hopkins did not have equipment for work on protein synthesis or radiolabeling experiments, I had to move some of my activities to the Biophysics Department, which had recently developed at Hopkins under the direction of Howard Dintzis. Dintzis had just done some beautiful work on hemoglobin synthesis in rabbits showing how globin chains are assembled on messenger RNA templates (9). While working in his department, I met two Englishmen, Michael Naughton and John Clegg, who had been trained in Fred Sanger's laboratory in Cambridge. At that time, Clegg was rather unsuccessfully attempting to study the synthe-

sis of insulin. Over coffee one morning, as we were commiserating with each other about our research problems, Clegg told me that, as part of his PhD thesis work in Cambridge, he had learned to separate the subunits of another protein by subjecting them to 8 molar urea and preventing aggregation with the addition of the evil smelling agent, mercaptoethanol. Together, we tried this approach to separate the α and β globin chains of some of my labeled hemoglobin samples, and it worked beautifully; we achieved complete separation with almost 100% recovery of our radioactive fractions. Soon the penny dropped. In β thalassemia, there were excess α chains synthesized, while in α thalassemia, excess β chains were produced. Thalassemia was not so much a disease of hemoglobin synthesis as of unbalanced globin-chain synthesis (10,11).

As well as providing an approach to a better understanding of the pathophysiology of thalassemia by relating it to the effects of the globin chains, which were produced in excess, the development of the *in vitro* system for studying hemoglobin synthesis was soon to have a valuable place in the control of β thalassemia by prenatal diagnosis. By using this method, several groups had found that although the main hemoglobin that is produced in fetal life is Hb F, a low level of the β chains of adult hemoglobin are synthesized from the midtrimester onward (12). Because fetal blood sampling was developed at about this time, it became possible to apply our *in vitro* technique to the prenatal diagnosis of β thalassemia and sickle cell anemia (12). By 1989, over 13,000 prenatal diagnoses had been carried out this way in many different countries with remarkably low error rates (13).

After returning to Liverpool in 1965 and establishing a small hemoglobin research group, I was joined by John Clegg who, after returning from the US to work in another field in Cambridge, had decided that further work on thalassemia was what he really wanted. Together, we applied our *in vitro* labeling method to the study of the patterns of globin chain

synthesis in β thalassemia and found that there was no fault in the rates of initiation, patterns of elongation and rates of termination of globin chain production. These findings implied that, in at least some forms of β thalassemia in which β globin production was reduced but not absent, a primary defect is reflected in the production of otherwise normal β globin messenger RNA (14).

In the early 1970s, we were able to further anticipate the direction of the hemoglobin field by defining the molecular basis for two different forms of α thalassemia. In the first case, as so often happened in the hemoglobin field, the first clue came from the identification of an unusual phenotype. Paul Milner, who worked in Jamaica, sent us blood samples from members of a family with a moderately severe form of α thalassemia, Hb H disease. Some of the family members carried an electrophoretic variant of hemoglobin that was present at about 2% of the total hemoglobin and that had not been identified previously. After purifying this variant, we found that it contained an α chain that was extended by 31 amino acids from the usual C-terminal arginine (codon 141). The next amino acid along was glutamine, suggesting that the normal chain termination codon UAA had changed to CAA and then a length of messenger RNA that was not normally translated had been translated until another in-phase stop codon was reached, thus resulting in the elongated α chain. It was found later that this variant was produced at an extremely low rate, probably because of the instability of the elongated mRNA. We named this variant Hb Constant Spring after the part of Jamaica in which the family lived (15). Similar variants were later found in the Mediterranean region and in Thailand; in the latter case, up to 4% of the population carried the variant in certain regions (16). An analysis of the amino acid composition of the elongated portion of the α chain enabled us to anticipate the structure of this end of the messenger RNA for α globin, a finding that caused some surprise to scientists in Cambridge, who

were struggling to sequence the globin messenger RNAs at that time.

The second success story in defining the molecular pathology of α thalassemia resulted from our studies of a condition that was, by the early 1970s, known as the Hb Bart's hydrops fetalis syndrome. These babies were either stillborn or died shortly after birth; their hemoglobin pattern showed no normal fetal or adult hemoglobin and consisted largely of γ_4 molecules, or Hb Bart's, suggesting that they had a severe defect in α chain production. In 1970, in collaboration with a team from Singapore, we were able to show by *in vitro* hemoglobin synthesis that these babies had a complete absence of α chain synthesis (17). This result raised the possibility that this group might have a deletion of both linked pairs of α globin genes. Around this time, thoughts were already turning to the possibility of exploring the notion that some forms of thalassemia could be caused by gene deletions by using cDNA/DNA hybridization to probe for the presence or absence of globin genes. Having demonstrated an absence of α chain synthesis in babies with Bart's hydrops, we approached the group of John Paul in Glasgow, who had some experience of this type of molecular hybridization, to discuss whether such an experiment might be feasible by using blood and tissues from an infant with this syndrome. Working in collaboration with a group in Thailand, we obtained a fetal liver from an affected infant and isolated DNA, which was then hybridized with specific α and β cDNA probes complimentary to human α and β globin mRNA (DNA) sequences. It was found that the hydropic fetal liver red cell precursors contained normal amounts of β globin DNA but a more or less complete absence of α globin DNA. A control sample from the tissues of normal infants showed the expected amounts of α and β DNA sequences (18). A similar study was carried out independently by Yuet Wai Kan's group in San Francisco at the same time (19).

Hence, by the mid 1970s, the genetic regulation and synthesis of the human

hemoglobins had been clearly defined and some hints had been obtained about the likely heterogeneity of the molecular basis for several forms of thalassemia. The scene was set for further rapid progress toward an understanding of the molecular pathology of these diseases and of many other single gene disorders.

FURTHER DEVELOPMENTS IN HEMOGLOBIN GENETICS AND THE EVOLUTION OF MOLECULAR MEDICINE

From the late 1970s onward, the further development of analytical methods including Southern blotting and cloning and sequencing were developed, and rapid progress was made toward determining the molecular basis for both the α and β thalassemias. At the same time, these approaches were also starting to be applied successfully for the analysis of the molecular basis for other monogenic diseases.

By use of the Southern blotting technique, progress was made toward the elucidation of the molecular basis of different forms of α thalassemia (20–23). By then, it was clear that the normal α globin genes are duplicated, with the genotype αα/αα. It was found that there are two common forms of α thalassemia that were designated α^+ and α^0 thalassemia. The α^+ thalassemias were found to result from deletions of one of the linked α globin genes (−α/αα) or from point mutations like the previously described hemoglobin Constant Spring ($\alpha\alpha^{CS}/\alpha\alpha$). The α^0 thalassemias result from the deletion of both α globin genes (−−/αα). It soon became apparent that many different-sized deletions are responsible for both α^+ and α^0 thalassemia; similarly, the non-deletion forms also turned out to be extremely heterogeneous. The coinheritance of α^+ and α^0 thalassemia leads to a condition called "Hb H disease," a hemolytic anemia of variable severity, whereas, as described earlier, the homozygous state for α^0 thalassemia causes stillbirths.

Equally rapid progress was made in cloning and sequencing genes for the β thalassemias (24–28). This work was

aided by the earlier observation that globin genes of normal individuals have a limited number of polymorphic restriction enzyme sites (that is, haplotypes). This observation intimated that the various types of thalassemia mutations were likely to be associated with particular haplotypes and that these could be used within families to mark individual globin genes and hence provide an approach to sequencing the appropriate gene (27). Using this method, the discovery of new β thalassemia mutations moved extremely quickly, and by the end of the 1980s, almost 100 had been identified.

There is little doubt that work during this period and over subsequent years has played a major role in developing the field of molecular medicine. As pointed out in a recent review (23), it is difficult to find a single example in which what has been learned from the study of the mutations involved in the hemoglobin disorders (involving enhancers, locus control regions, boundary elements, promoters, RNA processing and many other aspects of protein translation, structure and function) has been limited to the globin genes rather than helping to establish the principles of genome organization and how this relates to gene expression. Similarly, the actual mechanisms by which these mutations arose, including chromosomal rearrangements, telomere truncations, homologous and illegitimate recombination, gene conversion, copy number variation and the involvement of antisense RNAs have now been established as common to many human diseases and yet were first recognized in the hemoglobin disorders.

It is not surprising that the speed of development of molecular medicine during this period presented a number of training problems for young scientists. In countries such as the United Kingdom, there were few places where young clinicians could be trained to apply this technology to clinical problems. Similarly, it was difficult to find an environment in which young PhD researchers who wished to apply the technology to clinical problems could be trained. For this

reason, in 1989, at Oxford University, we developed the Institute of Molecular Medicine as an environment for bringing together young clinical and basic science research workers in this field. Today, the Institute houses over 400 scientists of wide-ranging backgrounds working in a variety of fields of molecular medicine and seems to provide a valuable environment for interplay among research in the increasingly diverse areas of molecular medicine.

POPULATION GENETICS AND EVOLUTIONARY BIOLOGY

Although opinions to the contrary have been expressed, there seems to be little doubt that it was Haldane who, in 1949, first suggested that the high frequency of thalassemia in the Mediterranean region might reflect heterozygote resistance to malaria (29). A few years later, and quite independently, Allison provided the first evidence that carriers of the sickle cell trait in Africa show significant resistance to malarial infection (30). Later, some evidence was obtained that the β thalassemia trait might also provide malaria resistance, but the findings were not confirmed in some studies.

Because there were reliable data on the level of malaria transmission in the southwest Pacific region, and because we had good connections in that region and some of our PhD students wished to gain experience working in the tropics, we established a research program in that region in the 1980s. It was found that there was a remarkable decline in the frequency of α⁺ thalassemia across this region, ranging from a carrier rate of almost 70% in Papua New Guinea and falling gradually in a southerly direction down to New Caledonia, where the frequency was approximately 6%. This decline in the frequency of α⁺ thalassemia was associated with a similar decline in the frequency of malaria transmission over this region (31). While this was suggestive evidence for the protective effect of α⁺ thalassemia against malaria, this could of course have simply resulted from α⁺ thalassemia being introduced

into this region from the mainland of Southeast Asia and then been gradually diluted as the populations moved south.

However, our team found that the molecular variety of α⁺ thalassemia and its haplotype in the island populations was quite different to that on the mainland. It appeared, therefore, that it had arisen locally and then been gradually magnified in frequency because of selection against malaria (31). This suggestion was proven to be correct beyond any doubt by further studies from our group, who carried out the first case-control analysis of thalassemia on the north coast of Papua New Guinea (32). This process entailed studying the frequencies of α thalassemia in patients admitted to the hospital with severe malaria compared with age-matched and randomly chosen control subjects from the same village. This work showed an undoubted protective effect of α thalassemia, a finding that was repeated in several different studies in sub-Saharan Africa many years later. During these studies, we also obtained evidence that α thalassemia (32,33), and later other forms of thalassemia, were more susceptible to infection with *Plasmodium vivax* malaria. It is likely that this is because the red cell receptor for this parasite is the Duffy blood group, and this receptor is expressed more actively in young red cell populations, which are common to these different forms of thalassemia.

It is now clear that, as well as α thalassemia and sickle cell traits, protection against malaria is also found in individuals with Hb C (34), and there is also evidence (although not strong) that Hb E and β thalassemia carriers also show relative protection against *Plasmodium falciparum*. The complex issue of the likely mechanisms of protection of these variants against different types of malaria, which have still not been completely worked out, is reviewed elsewhere (35).

Quite recently, there has been another important twist to the story of the relationship between the hemoglobin disorders and their interaction with malarial infection. Work carried out by the Oxford group in Africa has shown that, while

those who are heterozygous for the sickle cell gene or for α^+ thalassemia are protected against *P. falciparum* malaria (36,37), in individuals who inherit both of these mutations, the protective effect is completely lost and they are equally prone to malaria as individuals who have neither trait (38). As well as providing further evidence about the mechanisms of protection, these findings have important implications for providing further information about the regional variation in the frequency of the different hemoglobin mutations in different populations (39). There seems little doubt that epistatic interactions of this type will not be restricted to the hemoglobin field and may have important implications for the distribution and pathophysiology of many monogenic diseases.

As well as their impact on the better understanding of the evolutionary biology of common monogenic diseases, these findings have important practical applications. From several of our studies and those of others it has been found that the distribution of these common genetic variants is extremely heterogeneous in most high-frequency countries. Whether this fact reflects heterogeneity of malaria transmission is currently under study. But from a practical point of view, it means that research directed at attempting to define the frequency of these diseases as the basis for determining the number of births of severely affected homozygotes or compound heterozygotes as the basis for the control and management of these diseases requires micromapping of many different areas in high incidence countries to provide accurate data for public health developments (40).

PHENOMICS AND GENOMICS

One of the most remarkable features of the hemoglobin disorders is their extreme phenotypic variability. The clinical picture may vary enormously to similar or even identical mutations. These findings have important implications for the better understanding of all monogenic diseases and, in particular, of current genomic approaches to defining the genetic compo-

nent of common acquired conditions.

One of the best examples of the remarkable phenotypic diversity of the hemoglobin disorders is Hb E β thalassemia, which, globally, accounts for approximately 50% of all severe cases of β thalassemia and which reaches its highest frequency in the eastern half of south Asia and throughout southeast Asia (41). It results from compound heterozygosity for Hb E and a variety of different β thalassemia mutations. Because Hb E is a β chain variant that is synthesized at a slightly reduced rate, it acts as a mild form of β thalassemia. In a survey of over 200 patients with this condition in Sri Lanka over the last 15 years, and by a detailed analysis of the clinical phenotypes of these patients at least four times a year, it has been possible to make some progress toward an understanding of the reasons for the remarkable heterogeneity of the condition. It ranges from a severe form of thalassemia similar to β thalassemia major, requiring lifelong transfusion, to a disorder characterized by relatively normal growth and development and quality of life (albeit at a relatively low hemoglobin level, which is not much different than in individuals with the severe phenotype) (41,42).

On the basis of the multilayered complexity that appears to underlie the phenotypic diversity of Hb E β thalassemia, it has been necessary to describe the genetic component of this phenomenon under three classes of modifiers (43). The primary modifiers are the β thalassemia mutations of varying severity. Modifiers of this type turned out to be irrelevant in Sri Lanka because, of over 40 different β thalassemia mutations that have been identified, all but one are of the extremely severe variety (44). The secondary modifiers are those that involve the other globin gene loci. The coinheritance of α^+ thalassemia, which occurs commonly in Sri Lanka, has a remarkable effect in reducing the phenotypic severity of the disease, whereas the much rarer triplicated or quadruplicated form of the α globin genes has the opposite effect (42). Several different polymorphisms as-

sociated with an increased level of Hb F production have been identified in other populations (45), and in Sri Lanka, they also have a modifying effect on the phenotype by a significant elevation of Hb F production. The tertiary modifiers are those that affect the complications of the disease. For example, a significant number of patients with Hb E β thalassemia in this population have profound jaundice and an extremely high frequency of gallstones; it turns out that 25% of the population has the mutation that underlies Gilbert syndrome, and when this is coinherited with Hb E β thalassemia, they are affected in this way (46). Currently, approximately 40% of the phenotypic heterogeneity can be ascribed to these genetic modifiers.

Other factors are also of considerable importance in understanding the phenotype of this condition. Adaptation to anemia has been studied in relation to the oxygen affinity of the hemoglobins, and it has been found that patients with Hb E β thalassemia are better able to right-shift their oxygen dissociation curve in response to anemia, probably because of a relatively lower level of Hb F that is produced compared with other forms of severe β thalassemia (47). Environmental factors are also extremely important (notably infection due to malaria and other organisms). And a variety of other factors including social conditions, familial attitude to inherited disease, compliance with treatment, and many others, also have a remarkable effect on the course of the illness and phenotypic findings.

These observations underlie the importance of careful and regular phenotypic analysis in relating the latter to underlying genetic and environmental variability. If this is the case for a so-called "simple" monogenic disease, it underlies the critical importance of accurate phenotypic description of complex multigenic disorders.

LESSONS FROM THE HEMOGLOBIN DISORDERS

Throughout this short review, the critical importance of accurate phenotyping

as the basis for the study of genetic disease has been emphasized. While this is critical for the better understanding of monogenic disease, it is even more important for the investigation of the genetic component of many common acquired disorders. One of the current problems with the use of genome-wide association studies for the analysis of the latter conditions is the rather loose phenotypic definition of many of these conditions because of the very large numbers that are required for research of this type. In the future, it will be increasingly important that clinicians and molecular geneticists form closer associations so that the genotypes of patients who are involved in these huge studies are more clearly defined. Another important lesson from the hemoglobin disorders is the extreme value of analysis of the rare phenotype. Almost all the progress toward an understanding of the effect of unusual mutations and, in particular, the definition of the major regulatory regions involved in the control of hemoglobin synthesis and the variation in different types of hemoglobin production during different stages of development has come from a detailed study of the unusual phenotype. This approach is equally valuable in the case of common multigenic disease.

Although this brief review has focused mainly on thalassemia, many of the issues discussed, particularly those that relate to phenotypic diversity, are equally relevant to sickle cell anemia. As discussed earlier, there is clear evidence that the frequency of the severe hemoglobin disorders is likely to increase over the next half century. It has been estimated that in sub-Saharan Africa alone, the use of neonatal screening, followed by simple prophylactic antibiotics to prevent early deaths that occur from infection, will result in populations of many millions of patients with this condition in the future (48,49). The international health agencies and governments of countries where this disease is common cannot continue to neglect these common genetic disorders.

CAREERS IN MOLECULAR MEDICINE

Are there any lessons for young people contemplating a career in molecular medicine to be derived from this short account of work in a disease that helped to lead to the foundation of this field? Any success that the author of this review has had in the field is based on developing a stable team of a few senior scientists with backgrounds both in clinical medicine or molecular biology, together with a constant throughput of excellent young people wishing to be trained in the field. In addition, because the hemoglobin disorders are particularly common in some of the poorer tropical countries, it has been equally important to develop partnerships with these countries, not just for carrying out research, but also to help with capacity building for improving programs for the prevention and treatment of these conditions. For young medical students wishing to work in molecular medicine, our experience has suggested that it is wise to spend several years after qualification working toward their particular specialty and then breaking off their clinical education to spend several years working in an appropriate laboratory to be trained in the various aspects of molecular biology as applied to medicine. Young PhD graduates in the biological sciences require a further period of training in an environment in which the research is directed at various aspects of molecular medicine. In some parts of the world, the so-called MD/PhD program is another possibility for the medical student. The only disadvantage to this approach is that students are carrying out their early research before they really know which branch of medicine is going to interest them at a later stage in their career.

In the earlier sections of this review, I described the numerous pieces of advice that I received to the effect that I should certainly not continue working in the hemoglobin field. Although now that this field and molecular medicine in general have achieved a much higher level of respectability, I doubt if this problem will face many students in the future. But it is still important for young people to spend some time after receiving their first degree to gain further experience before deciding which particular aspect of the field they wish to pursue and, once they have decided, to follow their line of research intensely and with single-minded enthusiasm, and not to be put off by ill-directed advice by ill-informed mentors.

SUMMARY

The central message of this short review has been the critical importance of the value of well-defined or rare phenotypes as a background to the remarkable advances in the application of molecular biology to clinical medicine that is now possible. This step will require the input of clinical practice of the very highest level, which has tended to decline over recent years, even in the richer countries. It is vital that this trend is corrected to take advantage of the remarkable developments in the examination of disease at the molecular level that seem almost certain to occur over the next half-century.

ACKNOWLEDGMENTS

DISCLOSURE

The author declares that he have no competing interests as defined by *Molecular Medicine*, or other interests that might be perceived to influence the results and discussion reported in this paper.

REFERENCES

1. Weatherall D. (2002) Sir Cyril Astley Clark, C.B.E., 22 August 1907–21 November 2000. *Biogr. Mem. Fellows R. Soc*. 48 71–85.
2. Weatherall DJ. (2012) Cyril Clarke and the prevention of rhesus haemolytic disease of the newborn. *Br. J. Haematol*. 157:41–6.
3. Weatherall DJ. (2010) *Thalassaemia: The Biography*. Oxford: Oxford University Press.
4. Kunkel HG, Ceppellini R, Müller-Eberhard U, Wolf J. (1957) Observations on the minor basic hemoglobin component in blood of normal indi-

viduals and patients with thalassemia. *J. Clin. Invest.* 36:1615–25.

5. Weatherall DJ, Vella F. (1960) Thalassaemia in a Gurkha family. *BMJ*. i:1711–13.

6. Weatherall DJ, Boyer SH. (1961) The genetic control of the α chains of human hemoglobins. *Trans. Assoc. Am. Physicians*. 74:89–99.

7. Weatherall DJ. (1963) Abnormal haemoglobins in the neonatal period and their relationship to thalassaemia. *Br. J. Haematol*. 9:265–77.

8. Conley CL, Weatherall DJ, Richardson SN, Shepard MK, Charache S. (1963) Hereditary persistence of fetal hemoglobin: a study of 79 affected persons in 15 Negro families in Baltimore. *Blood*. 21:261.

9. Dintzis HM. (1961) Assembly of the peptide chains of hemoglobin. *Proc. Natl. Acad. Sci. U. S. A.* 47:247–50.

10. Clegg JB, Weatherall DJ. (1967) Haemoglobin synthesis in alpha-thalassaemia (haemoglobin H disease). *Nature*. 215:1241–3.

11. Weatherall DJ, Clegg JB, Naughton MA. (1965) Globin synthesis in thalassemia: an *in vitro* study. *Nature*. 208:1061–5.

12. Weatherall DJ, Clegg JB. (2001) *The Thalassaemia Syndromes*. 4th Ed. Oxford: Blackwell Science.

13. Alter BP. (1990) Antenatal diagnosis. Summary of results. *Ann. N. Y. Acad. Sci*. 612:237–50.

14. Clegg JB, Weatherall DJ, Eunson CE. (1971) The distribution of nascent globin chains on human reticulocyte polysomes. *Biochim. Biophys. Acta*. 247:109.

15. Clegg JB, Weatherall DJ, Milner PF. (1971) Haemoglobin Constant Spring: a chain termination mutant? *Nature*. 234:337–40.

16. Fessas P, *et al.* (1972) Identification of slow-moving haemoglobins in haemoglobin H disease from different racial groups. *Lancet*. i:1308.

17. Weatherall DJ, Clegg JB, Boon WH. (1970) The haemoglobin constitution of infants with the haemoglobin Bart's hydrops foetalis syndrome. *Br. J. Haemat*. 18:357–67.

18. Ottolenghi S, *et al.* (1974) The severe form of α thalassaemia is caused by a haemoglobin gene deletion. *Nature*. 251:389–92.

19. Taylor JM, *et al.* (1974) Genetic lesion in homozygous α-thalassaemia (hydrops foetalis). *Nature*. 251:392–93.

20. Embury SH, Lebo RV, Dozy AM, Kan YW. (1979) Organization of the α-globin genes in the Chinese α-thalassemia syndromes. *J. Clin. Invest*. 63:1307–10.

21. Embury SH, *et al.* (1980) Two different molecular organizations account for the single α-globin gene of the α-thalassaemia-2 genotype. *J. Clin. Invest*. 66:1319–25.

22. Orkin SH, *et al.* (1979) The molecular basis of α-thalassemias: frequent occurrence of dysfunctional α loci among non-Asians with Hb H disease. *Cell*. 17:33–43.

23. Higgs DR. (2012) The alpha thalassemias. *Cold Spring Harb. Perspect. Med*. 2:a011718.

24. Orkin SH, Old JM, Weatherall DJ, Nathan DG. (1979) Partial deletion of β-globin gene DNA in certain patients with β^0-thalassemia. *Proc. Natl. Acad. Sci. U. S. A.* 76:2400–4.

25. Spritz RA, *et al.* (1981) Base substitution in an intervening sequence of a β$^+$ thalassemic human globin gene. *Proc. Natl. Acad. Sci. U. S. A.* 78:2455–9.

26. Westaway D, Williamson R. (1981) An intron nucleotide sequence variant in a cloned β$^+$ thalassaemia globin gene. *Nucl. Acids Res.* 9:1777–88.

27. Orkin SH, *et al.* (1982) Linkage of β-thalassaemia mutations and β-globin gene polymorphisms with DNA polymorphisms in human β-globin gene cluster. *Nature*. 296:627–31.

28. Thein SL, Wood WG. (2009) The Molecular Basis of β Thalassemia, δβ Thalassemia, and Hereditary Persistence of Fetal Hemoglobin. In: *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management*. 2nd ed. Steinberg MH, Forget BG, Higgs DR, Weatherall DJ (eds.). Cambridge: Cambridge University Press; pp. 323–56.

29. Haldane JBS. (1949) The rate of mutation of human genes. *Hereditas*. 35:267–73.

30. Allison AC. (1954) Protection afforded by sickle-cell trait against subtertian malarial infection. *BMJ*. i:290–4.

31. Flint J, *et al.* (1986) High frequencies of α thalassaemia are the result of natural selection by malaria. *Nature*. 321:744–9.

32. Allen SJ, *et al.* (1997) α+-Thalassemia protects children against disease due to malaria and other infections. *Proc. Natl. Acad. Sci. U. S. A.* 94:14736–41.

33. Williams TN, *et al.* (1996) High incidence of malaria in α-thalassaemic children. *Nature*. 383:522–5.

34. Modiano D, *et al.* (2001) Haemoglobin C protects against clinical *Plasmodium falciparum* malaria. *Nature*. 414:305–8.

35. Williams TN, Weatherall DJ. (2012) World distribution, population genetics, and health burden of the hemoglobinopathies. *Cold Spring Harb. Perspect. Med*. 2:a011692.

36. Williams TN, *et al.* (2005) An immune basis for malaria protection by the sickle cell trait. *PLoS Med*. 2:e128.

37. Williams TN, *et al.* (2005) Both heterozygous and homozygous alpha+ thalassemias protect against severe and fatal Plasmodium falciparum malaria on the coast of Kenya. *Blood*. 106:368–71.

38. Williams TN, *et al.* (2005) Negative epistasis between the malaria-protective effects of alpha+-thalassemia and the sickle cell trait. *Nat. Genet*. 37:1253–7.

39. Penman BS, Pybus OG, Weatherall DJ, Gupta S. (2009) Epistatic interactions between genetic disorders of hemoglobin can explain why the sickle-cell gene is uncommon in the Mediterranean. *Proc. Natl. Acad. Sci. U. S. A.* 106:21242–6.

40. Weatherall DJ. (2010) The importance of micromapping the gene frequencies for the common inherited disorders of haemoglobin. *Br. J. Haematol*. 149:635–7.

41. Fucharoen S, Weatherall DJ. (2012) The hemoglobin E thalassemias. *Cold Spring Harb. Perspect. Med*. 2:a011734.

42. Olivieri NF, *et al.* (2008) Studies in haemoglobin E beta-thalassaemia. *Br. J. Haematol*. 141:388–97.

43. Weatherall DJ. (2001) Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. *Nat. Rev. Genet*. 2:245–55.

44. Fisher CA, *et al.* (2003) The molecular basis for the thalassaemias in Sri Lanka. *Br. J. Haematol*. 121:662–71.

45. Thein SL, Menzel S. (2009) Discovering the genetics underlying foetal haemoglobin production in adults. *Br. J. Haematol*. 145:455–67.

46. Premawardhena A, *et al.* (2001) Genetic determinants of jaundice and gallstones in haemoglobin E beta thalassaemia. *Lancet*. 357:1945–6.

47. Allen A, *et al.* (2010) Adaptation to anemia in hemoglobin E-beta thalassaemia. *Blood*. 116:5368–70.

48. Piel FB, Hay SI, Gupta S, Weatherall DJ, Williams TN. (2013) Global burden of sickle cell anaemia in children under five, 2010–2050: modelling based on demographics, excess mortality, and interventions. *PLoS Med*. 10:e1001484.

49. Piel FB, *et al.* (2010) Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. *Nat. Commun.* 1:104.