HAEMOGLOBIN: MODEL STRATEGIES FOR MOLECULAR DISEASE

Privileged Communication.

Proceedings IXth Global Impaacts of Applied Microbiology and Biotechnology,
UNESCO, Valletta, Malta. September 1991. A.E. Felice & E. DaSilva, editors.

HAEMOGLOBIN: MODEL STRATEGIES FOR MOLECULAR DISEASE

C.A. Scerri and A.E. Felice
Comprehensive Genetics Programm,
Dept. of Biomedical Sciences,
University of Malta, Msida, Malta,
and Department of Health, Valletta, Malta.

Haemoglobin occupies a special position among the many proteins of man and other animals. It is among the most abundant of proteins and one of the best studied. Traditionally, haemoglobin has occupied the leading edge in the development of our understanding on the structure, function and biosynthesis of proteins and more recently in molecular biology and medicine. Approximately fifteen years since α-thalassaemia was documented as the first molecular disease, we take this opportunity to review some of the most salient features of haemoglobin molecular biology and relate this body of knowledge to the development of strategies that contribute in understanding other problems in molecular medicine and improve patient care. Haemoglobin R & D programmes offer tremendous opportunities to enhance the education and training of young scientists in developing countries and help to develop high quality services.

The term molecular disease is relatively new in medical nomenclature. It refers to the study of a range of conditions which are due to abnormalities in the sequence or organisation of the nuclear or mitochondrial nucleic acids. These encompass genetics, oncology and infectious disease, typically viral infections.

At the level of protein structure, the haemoglobin molecule is a tetramer consisting of two pairs of dissimilar globin polypeptides. Each one of these is associated with a haeme prosthetic group; haeme being the ligand or oxygen binding site. As in the case of other animals, the bulk haemoglobin solution isolated from peripheral erythrocytes of man, is heterogenous. Its components which differ in subunit composition and developmental expression can be resolved by advanced biochemical techniques. Electrophoresis has been one of the most versatile techniques for haemoglobin analysis. Iso-electric focusing is a modern modification which permits fine resolution of abnormal haemoglobin and is suitable for clinical investigation and population testing. A typical separation of the adult Hb A from that of the foetal Hb F as well as some of the more common abnormal haemoglobins such as Hb S ($\alpha_2\beta_2{}^{6\ Val}$), common among peoples of African descent and Hb F Malta I (or $\alpha_2\gamma_2{}^{117\ Arg}$), a variant present in about 2% of newborn in Malta is shown in Fig. 1.

Another approach in haemoglobin analysis, employs one of the many possible chromatographic procedures, especially high performance liquid chromatography with ion exchange or reverse phase matrices. They are extensively used in the investigation of normal and abnormal haemoglobins as well as in providing services for patients. Ion exchange chromatography has developed into one of the most effective steps in downstream processing of biotechnology products. It is interesting to note that many of these analytical and preparative techniques have been developed in different laboratories utilising haemoglobin as a model protein.

The succession of different haemoglobins during the course of human development is characterised by at least two genetic switches and perhaps even higher levels of complexity. During the first six week of gestation the haemoglobin appearing in the yolk sac, is composed of Hb Gowers I or $\zeta_2\varepsilon_2$(1) and II or $\alpha_2\varepsilon_2$(1) and Hb Portland or $\zeta_2\gamma_2$(2,3) which are assembled from the embryonic ε and ζ globin chains. These are then replaced by the permanent α globin chains and the two types of γ globin chains, $^{G}\gamma$ and $^{A}\gamma$(4), that differ from each other at position 131 of the polypeptide chain which is glycine in $^{G}\gamma$ and alanine in $^{A}\gamma$. The relative level of expression of the two γ chains is genetically controlled. Around the time of birth, a second genetic switch shuts down the production of the γ chains and turns on the production of the β chains of Hb A ($\alpha_2\beta_2$) and the minor δ component of Hb A_2 ($\alpha_2\delta_2$) (Fig. 2).

Each of these seven polypeptides is represented at least once in nuclear DNA by corresponding structural genes. The α and the γ globin genes are duplicated. The z and α genes occur on chromosome #16 while the ε, γ, δ and β genes are on chromosome #11. Both sets of genes occur on their respective chromosomes in the order in which they are developmentally expressed.

The perinatal shutdown of Hb F, is incomplete. In fact, small quantities of Hb F persist in normal adults at levels below 1% of the total haemoglobin. However, this is restricted to a subpopulation of erythrocytes designated F-cells. This is a typical example of tissue heterogeneity in gene expression and undoubtedly reflects complex gene control which is only now being unravelled(see 5 for references). It reflects similar levels of possible complexity and heterogeneity of gene expression in other tissues such as skin and gastro-intestinal tissue.

Following transcription, the primary RNA transcript undergoes a series of processes wherein the intron sequencies are spliced and the resulting RNA matured through capping and the addition of adenyl residues at the 3'tail end.

The final mRNA is exported into the cytoplasm for translation into globin monomers and subsequent assembly into dimers and tetramers. Hereditary abnormalities can affect any one of these biochemical steps and can thus impair gene expression as they commonly do in thalassaemia. Many abnormalities of splicing and mRNA processing have been identified in globin(6) as well as other genes.

At a higher level of resolution, the structural organisation of a typical globin gene can be understood in terms of regions of DNA that are critical for the control of gene activity. These include the locus control region (LCR), the promoter region with its RNA polymerase binding sites, the cap site, the initiation codon, the coding regions or exons, intervening sequencies or introns, with their invariant dinucleotides at the intron-exon junction, the termination and polyadenylation signals.

DNA abnormalities can be considered to fall in two broad categories with some overlap. One class is due to base pair substitutions that alter the coding sequence in the exons and hence in the amino acid sequence of the protein product and result in the production of haemoglobin variants. Another class is due to deletion or base pair substitution that impairs gene expression and results in globin deficiencies typical of the thalassaemias(6).

Approximately 600 abnormal haemoglobins each differing in the amino acid sequence of one of the polypeptide globin subunits are known thus far (An updated listing is published at regular intervals by the International Haemoglobin Information Centre in Hemoglobin (7)). Many are useful in the study of the structure-function relationship in proteins, while most contribute to population genetics. A few, such as Hb S, Hb C and Hb E are the cause of significant health problems in certain populations. The polymerisation of deoxygenated Hb S molecules results in the formation of irreversibly sickled cells and the clinical syndrome of sickle cell disease. Although much has been learnt about the molecular and cellular pathogenesis of sickle cell disease, management remains very unsatisfactory. In the early 1980's, large newborn testing programmes in New York, Georgia and California in the USA, showed that if Hb S homozygotes could be identified early in life and recruited into treatment programmes with pneumococcal vaccination and prophylactic penicillin, a significant decrease in infant morbidity and mortality could be achieved(8).

Newborn genetics testing programmes such as these, have become an important health tool for the control of genetic disorders. They serve to identify and recruit patients into ad hoc clinical care programmes; to identify couples at risk of

having children with genetic disease and who could benefit from counselling; to establish databases for public health planning; and to enhance resources for education, training and research. They were initially developed for the care of metabolic disease such as Phenylketonuria but they have also contributed significantly to improve the care of haemoglobin disorders in certain populations. Large programmes also exist for the identification of neonatal hypothyroidism. A review of metabolic disease and haemoglobin testing programmes serves to establish criteria and protocols for the appropriate design of programmes that may be applicable in other areas of medical genetics.

One of the outcomes of Newborn Haemoglobin Testing Programmes, is the ability to collect data on family members, including many heterozygotes and identify previously undectected homozygotes. Considerable heterogeneity in the phenotypic expression of disease has been uncovered. It is often observed that the heterozygous parents of Hb S homozygotes, have quantities of Hb S significantly below the expected value of 50%.

In fact, the proportion of Hb S in a large number of Hb S heterozygotes, falls in categories, depending on the number of active α globin genes inherited by the carrier. The lower the number of α globin genes present, the lower the proportion of Hb S in cell lysates (Fig. 3). This observation has been explained by a lower affinity of the abnormal β^S chains for normal α chains as compared to that of the normal β^A chains. Consequently, whenever α chains are deficient, the post-translational assembly of Hb A is favoured over that of Hb S resulting in a lower proportion of Hb S in heterozygotes(9).

This type of interaction could be representative of a more general type of interaction between protein subunits at the level of post-translational assembly. Consequently the relative quantities of two heteropolymers which share a common subunit and which differ in their rates of assembly can be influenced by alteration in the levels of the shared polypeptides.

Possible examples of this type of interaction could be found in the production of the hybrid oncogene products such as jun and fos(10); or of polymorphic HLA heterodimers which may be in linkage with conditions such as Diabetes Mellitus Type I(see 11 for review); or among the different collagens present in various types of connective tissues and which are assembled from the products of at least nine different collagen genes(12). Many enzymes of metabolism are also heteropolymers. Thus this type of interaction between different genes could explain heterogeneity in other gene systems in addition to haemoglobin.

Another outcome of Newborn Haemoglobin Testing is the ability to document prospectively the developmental changes in phenotypic features such as the haematological values or other age dependent genetic effects. As many as one third of Hb S homozygotes also inherit chromosomes with α globin gene deletions - a reminder that even the typical single gene disease can in fact be multigenic in origin.

Up to the age of around 7 years, Hb S homozygotes who differ in their number of α globin genes have very similar total haemoglobin values. After this age however, those with four α globin genes have worse haemolytic disease and lower Hb levels, whereas those with two α globin genes have less severe haemolysis, higher Hb levels and higher viscosity with rheologic complications(13). Although the age dependent interaction between Hb S homozygosity and α-thalassaemia is difficult to explain it could serve as a model to account for genetic heterogeneity perhaps for acquired disease with strong genetic components. Multiple gene defects could interact at different levels during development and aging, though this is somewhat speculative. It raises questions that could be profitably pursued in many communities.

The α globin gene deletions are very common in many ethnic groups. They are readily identified with straightforward gene mapping techniques using α gene probes (Fig 4). The normal fragment is the 16Kb fragment carrying two α globin genes. The 12Kb fragment has only one α globin gene due to a 4Kb deletion inclusive of the other α globin gene and is associated with α+ thalassaemia.

Larger deletions are also known to occur among α globin genes. These lead to more severe types of α-thalassaemia. In general, the size of the deletions is about the same, and their ends map quite closely in DNA. Perhaps, the deletion events are associated with a physical looping of the DNA or the ends share sequences which predispose this DNA to deletion.

DNA deletions are a frequent cause of genetic disease. For instance, Haemophilia deletions are spread all over the very large Factor VIII gene. In haemophilia B it was initially thought that the ocurrence of deletion could distinguish between patients with or without immunological inhibitors. In Muscular Dystrophy, the site of deletion in the dystrophin gene most times correlates with clinical phenotype although patients with the same deletion and different phenotypical picture are also known. The co-inheritance of defects at multiple loci, as in the case of Hb S and α globin gene deletions should be considered in pleiotropic disease.

On the other hand, deletions are rarely the cause of β-thalassaemia though they have been documented. More often, β-thalassaemia is associated with mutations in a critical region of the β-globin gene. These have been difficult and laborious to document until recently.

The approach for the identification of base pair substitutions in thalassaemia and other genetic disease has been radically changed by the introduction of the Polymerase Chain Reaction or PCR(14). In this technique, substantial quantities of target globin DNA can be reproduced in vitro from exceedingly small quantities of starting material. The products can then be sequenced or analysed with relatively simple biochemical techniques. A common approach is to probe the products with synthetic allele specific oligonucleotides (ASO) which detect specific normal or abnormal sequences. We have used this approach to identify the molecular types of β- thalassaemia in the Maltese population(15).

In addition to providing information about gene control, the documentation of these mutations is clinically useful because it has predictive value on the future course of the disease in individual patients. As therapeutic alternatives improve, management may be made to fit the needs of the patients.

Over 110 different mutations accounting for β-thalassaemia have been identified and their distribution in different populations studied (see 6 for current listing). Across the Mediterranean there is a gradient of mutations. The prevalance of the codon-39 mutation decreases from a maximum of 95% in the Western Mediterranean to zero in the Middle East. The opposite is observed for the IVSI-110 mutation. In the case of the Maltese population it is remarkable that around 60% of the thalassaemiacs thus far identified can be accounted for by one mutation i.e. the IVSI nt6 T C. Data have also been obtained that objectively quantify the level of expression of the mutant β+ Thal(IVSI-6 T C)(15).

Similar data have been obtained on another common hereditary disease - cystic fibrosis (CF) (see 16 for details). Most commonly CF is due to a three base deletion designated DF508. Recent studies in some Mediterranean regions showed a lower percentage for this mutation as compared with Northern European and American populations. In the case of the Maltese population, out of 6 chromosomes thus far studied, 3 have resulted from the 508 deletion.

Molecular biology has contributed in a remarkable way to our understanding of haemoglobin and its abnormalities, whilst biotechnology has equally gained from haemoglobin. Our understanding of haemoglobin, provides us with

models, on the basis of which, testable hypotheses on other disorders can be formulated. Furthermore, strategies to offer high quality services and training opportunities in the context of Comprehensive Genetics Programmes can be developed.

1. Huehns, E.R., Flynn, F.V., Butler, E.A. and Beaven, G.H. (1961) Two new haemoglobin variants in a very young human embryo. Nature, 189, 496.

2. Capp, G.L., Rigas, D.A. and Jones, R.T. (1967) Haemoglobin Portland I - a new human haemoglobin unique in structure. Science, 157,65.

3. Capp, G.L., Rigas, D.A. and Jones, R.T. (1970) Evidence for a new haemoglobin chain (ζ chain). Nature, 228,278.

4. Schroeder, W.A., Shelton J.R., Shelton J.B., Cormack, J. and Jones, R.T. (1963) The amino acid sequence of the γ chain of human fetal haemoglobin. Biochemistry, 2, 992.

5. Ley T.J. (1991) The Pharmacology of Haemoglobin switching: of Mice and Men. Blood, 77(6), 1146.

6. Huisman T.H.J.(ed). β-thalassaemia reference. (1990) Hemoglobin, 14(6), 661

7. IHIC variants list. (1991) Hemoglobin 15,3.

8. Serjeant G.R. (1985) Sickle cell disease. Oxford Medical Publications.

9. Felice A.E., Altay C., Milner P.F. and Huisman T.H.J. (1981) The occurrence and identification of α-thalassaemia-2 among Hb S heterozygotes. Am. J. Clin. Pathol. 76,70

10. McKnight S.L. (1991) Molecular zippers in gene regulation. Scientific American 264(4),32

11 Thomson G., Robison W.P., Kuhner M.K., Joe S., MacDonald M.J., Gottshall J.L., Barbosa J., Rich S.S., Bertrams J., Baur M.P., Partanen J., Tait B.D., Schober E., Mayr W.R., Ludvigsson J., Lidblom B., Farid N.R., Thompson C. and Deschamps I. (1988) Genetic heterogeinity, modes of inheritance, and risk estimates for a joint study of caucasians with insulin-dependent diabetes mellitus. Am. J. Hum. Genet. 43, 799.

12. Vurio E. and de Crombrugghe B. (1990) The family of collagen genes. Annual Review of Biochemistry 59,837.

13. Felice A.E., Cleek M.P., Marino M., Butler A., McKie K.M. and McKie V.C. (1987) Effects of α-thalassaemia-2 on the Developmental changes of haematological values in children with sickle Cell Disease from Georgia. Am. J. Heamatol. 25, 389.

14. Mullis K., Faloona F., Scharf S., Saiki R., Horn G. and Erlich H. (1986) Specific enzymatic amplification of dna *in vitro*: the polymerase chain reaction. Cold Spring Harbor Symposia on Quantitative Biology 51(1), 263.

15. Felice A.E., Abela W., Bartolo C., Borg I., Camilleri R., Grech J.L., Pizzutto M. and Sciberi C.A. (1991) Quantification of *in vivo* expression of the β IVS-I-nt#6 β^+ thalassaemia in heterozygotes and homozygotes. Proceedings of the 4th International Conference on Thalassaemia nad Haemoglobinopathies. p. 101.

16. Human Genetics 85,(4),404, 1990.

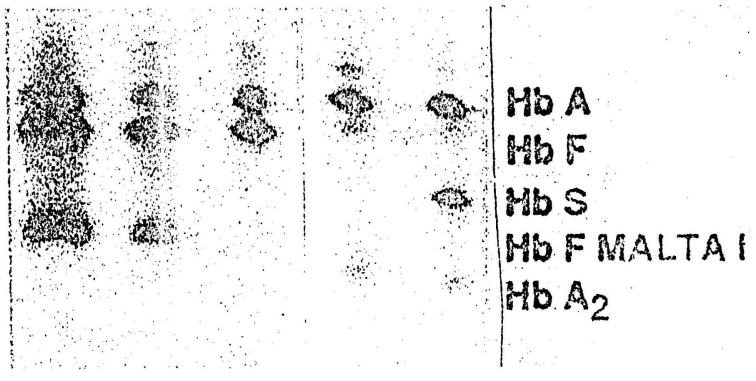

Fig. 1. Isolectric focusing of haemoglobin showing a typical separation of the adult Hb A from that of the foetal Hb F as well as some of the more common abnormal haemoglobins such as Hb S ($\alpha_2\beta_2{}^{6\ Val}$), common among peoples of African descent and Hb F Malta I (or $\alpha_2\gamma_2{}^{117\ Arg}$), a variant present in about 2% of newborn in Malta.

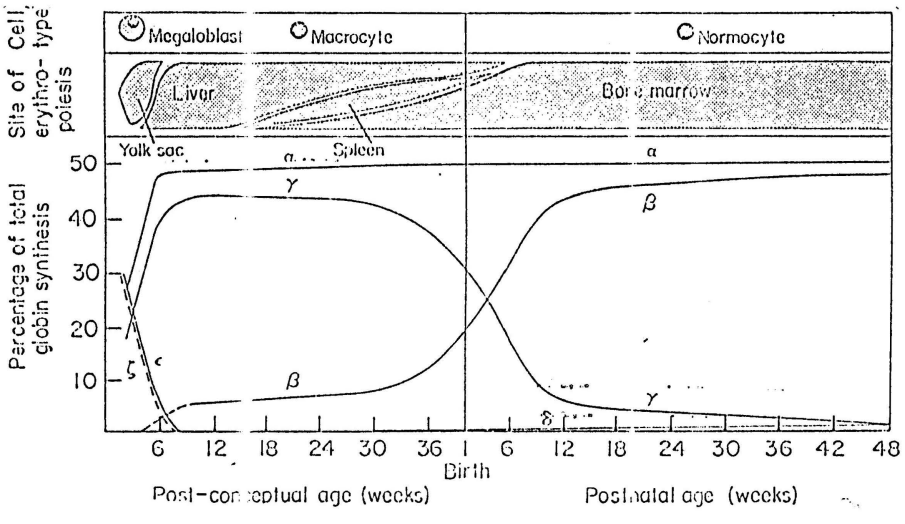

Fig. 2. Developmental changes of the alpha-like and non-alpha like globins during prenatal and postnatal period.

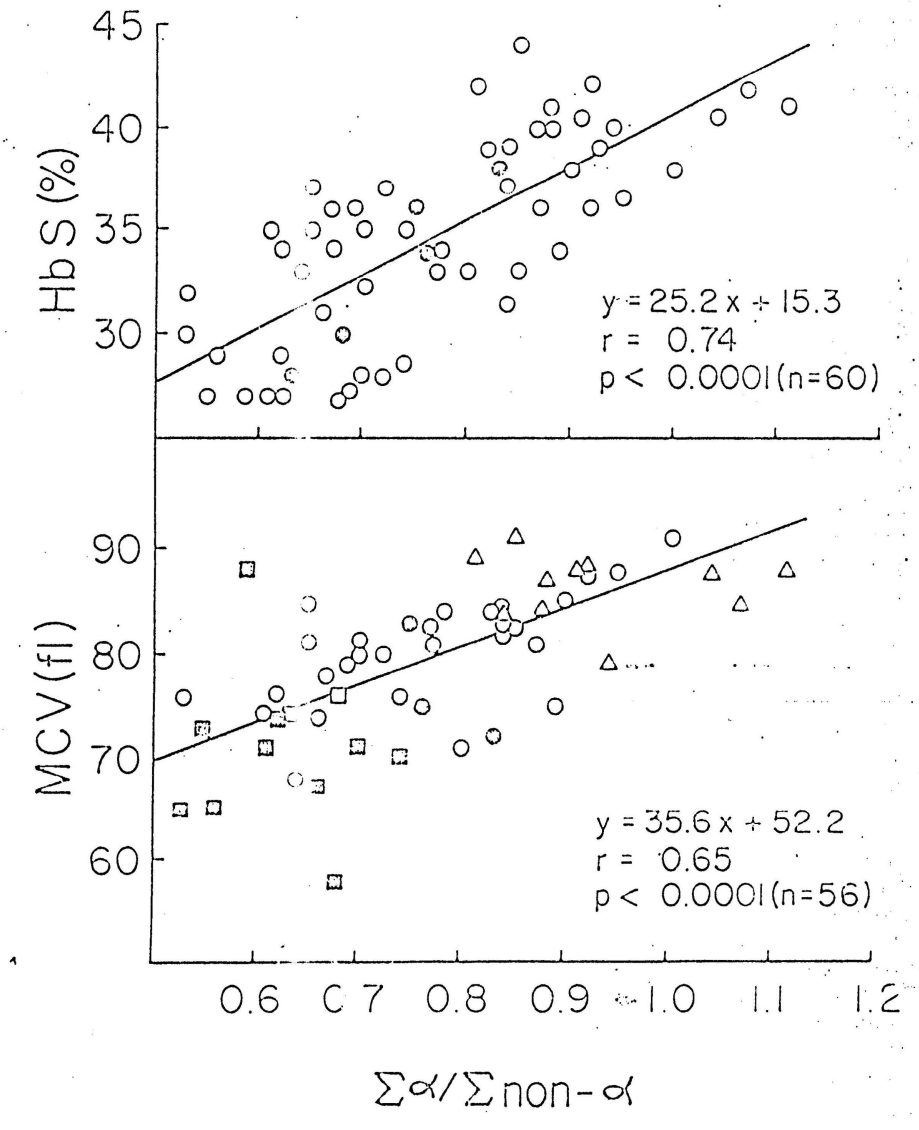

Fig.3. The relationship between a) MCV and b)%HbS and number of α-genes in Hb S Heterozygotes.

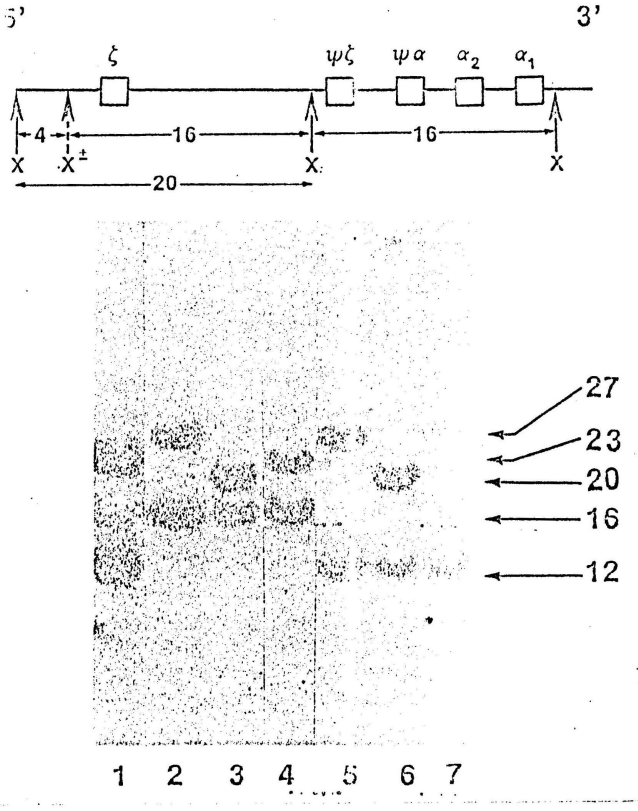

Fig. 4. Diversity of normal and abnormal α and ζ globin gene specific fragments obtained with Xba I and an α globin gene probe. The same fragments hybridize with a ζ globin gene probe (not shown). The 12 Kb fragment is due to α-thalassaemia-2 (or -α/); the 16Kb fragment is normal (or αα/); the 20Kb fragment is due to α globin gene triplication (ααα/); the 23kb and 27kb are new fragments which are considered variants of the same or very similar ζ globin gene deletions (or $-\zeta^{*}\alpha\alpha/$) occurring on chromosomes in the absence (X^{-}) or the presence (X^{+}) of the polymorphic Xba I site 5' to the ζ globin gene.