Protein Structure, Function and Disease

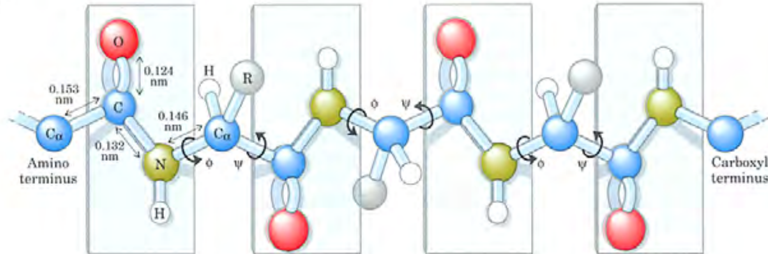
Molecular basis of diseases

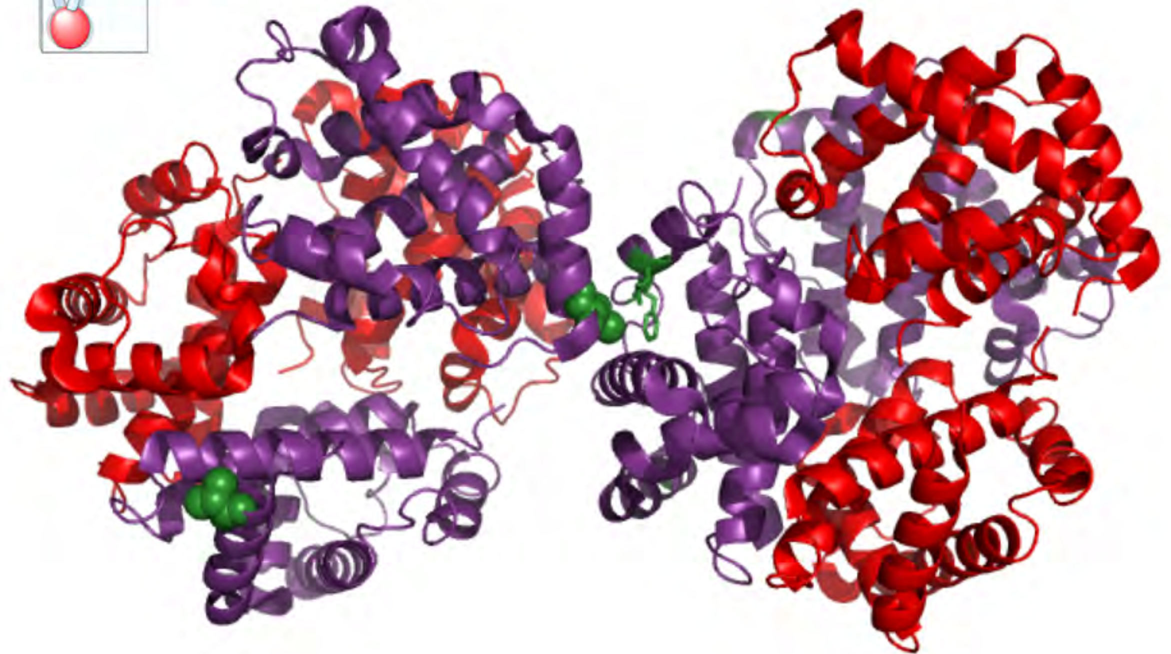
(Partially adopted from Prof. John Baenziger's former lectures)

Jyh-Yeuan (Eric) Lee, Assistant Professor, BMI

What are proteins?

Proteins are linear chains (polymers) of covalently linked amino acids...





... that fold to adopt complexed three dimensional (3-D) structures.

What are the roles proteins in cells?

Enzymes (biological catalysts):
catalyse millions of chemical reactions that occur in cells.

For example, the firefly contains an enzyme, luciferase, that catalyses the oxidation of its chemical substrate, luciferin, releasing green light.

Transport: transport proteins move molecules from one area of the body to another, across cell membranes, etc.

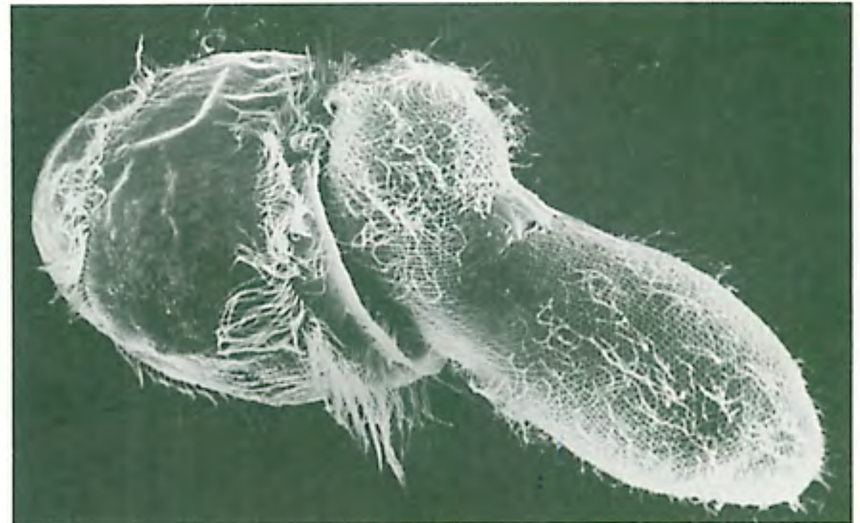
For example, Hb binds O_2 and transports it from lungs to peripheral tissues, lipoproteins in blood carry lipids from the liver to tissues.

What are the roles proteins in cells?

Nutrient and Storage: the seeds of many plants store nutrients as protein. Ovalbumin, the major protein of egg white, and casein, the major protein of milk are other nutrient proteins

(c)

Mobility/Contractility: proteins allow cells to contract, change shape, and swim. Actin and myosin form the contractile systems of skeletal muscle.



(d)

What are the roles proteins in cells?

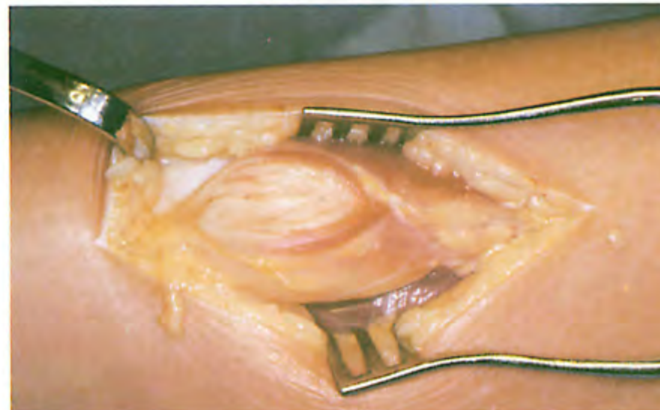
Structural: many proteins serve as supporting filaments to give biological structures. E.g. fibroin is major component of spider webs, collagen forms tendons and cartilage

(e)

(f)

Defense: proteins defend organisms against invasion. The castor bean (left) produces a protein called ricin that is cytotoxic to animals. Humans use antibodies to defend themselves from foreign invaders

Regulatory: many proteins regulate cellular activity. Mutations in proteins that control cell growth can lead to tumour formation



What are the roles proteins in cells?

Proteins are the molecular machines that make life possible!

If we are to understand life, we must understand protein function.

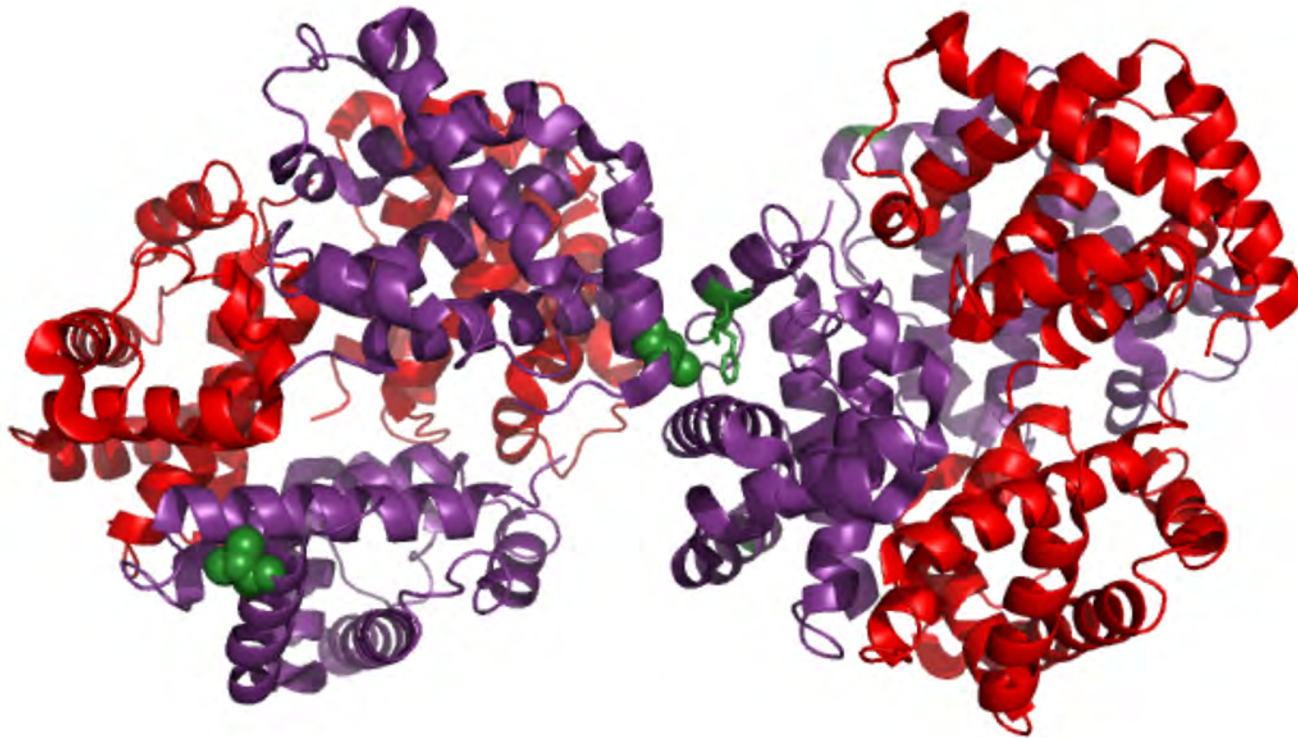
Mutations in proteins often lead to human disease

If we are to understand and treat human disease, we must understand cell function at the molecular – i.e. the protein level.

Proteins are the targets of pharmaceuticals!

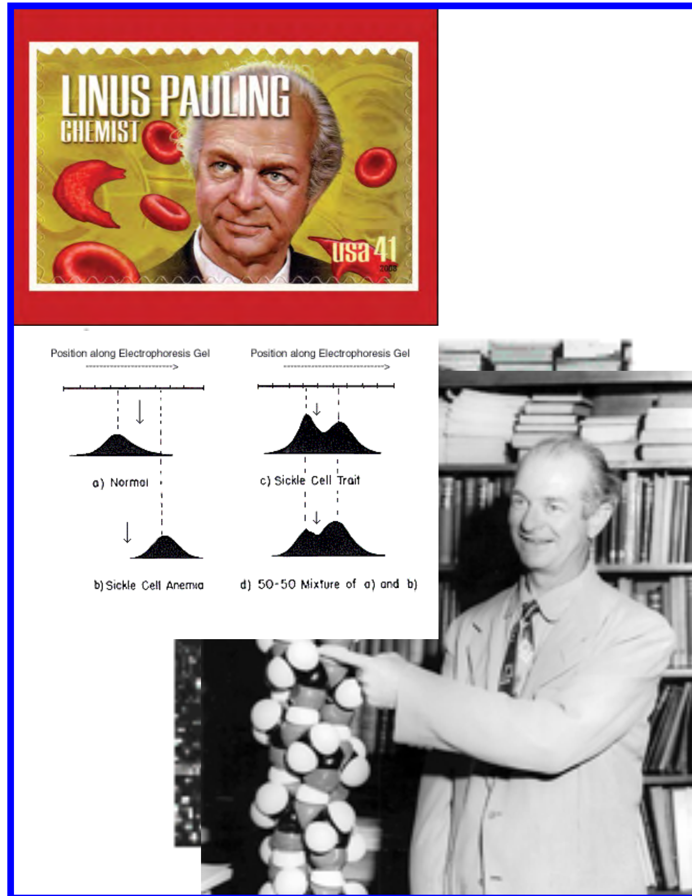
If we understand how proteins work, and what is not working properly in diseased states, we can design pharmaceuticals that correct the abnormalities associated with human disease.

Proteins: the molecular basis of disease



Proteins are the molecular machines of life. To understand life, we must understand protein function. Understanding protein function also serves as a basis for understanding and ultimately treating human disease. In today's class, we focus on sickle cell anemia. Biochemical studies focused on sickle cell anemia showed for the first time that human disease can result from chemical changes to a protein – hence the term "molecular basis of disease".

The story of sickle cell anemia



(*Science* 110, 543 (1949))

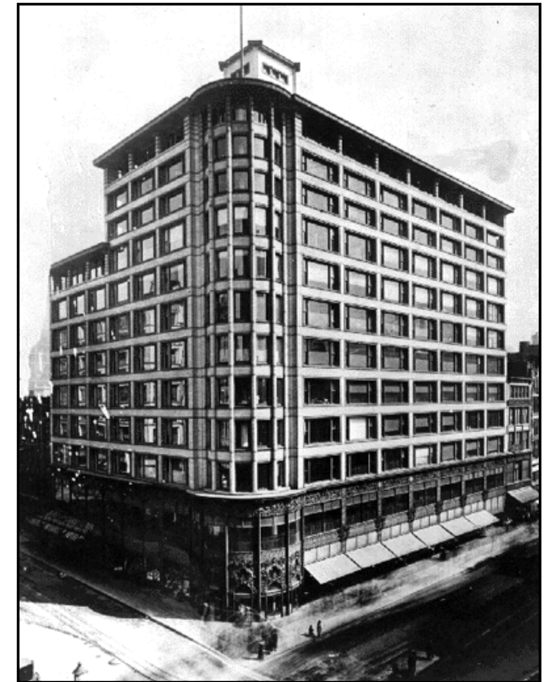
Starring: Walter Noel, James Harrick, Linus Pauling, Vernon Ingram, Max Perutz, and many others…

refs: Science (1999) *286*, 1488-1490
Pediatr. Radiol. (2001) 31, 889-890

Walter Clement Noel (Grenada, West Indies)

IN LOVING MEMORY OF
WALTER CLEMENT NOEL
WHO DIED 1ST MAY 1916
AGED 32 YEARS

Walter Noel arrived in Chicago 1904...

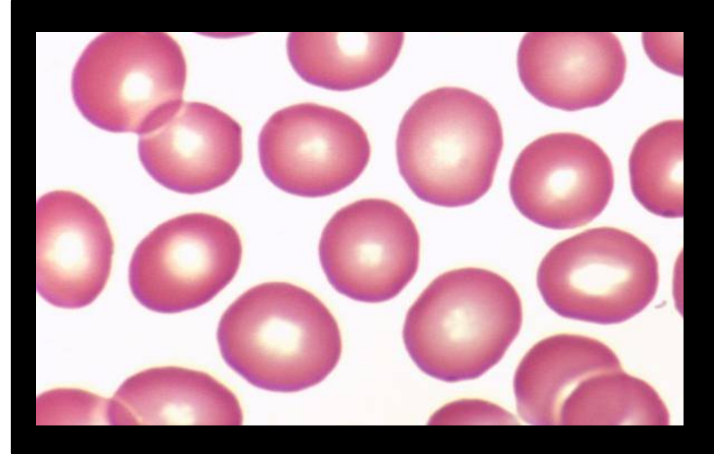


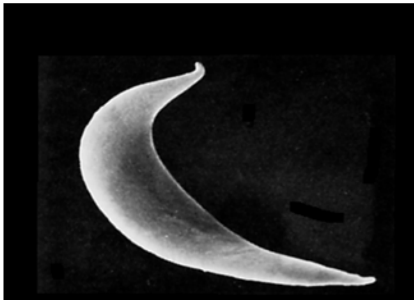
On September 14, 1904, a young black man from Grenada, named Walter Clement Noel, arrived in New York. He took the train to Chicago and began studies in dentistry at the Chicago College of Dental Surgery. Three months after arriving, he was admitted to Chicago's Presbyterian Hospital suffering from severe respiratory distress. Dr. Ernest E. Irons was the intern on duty, and the attending physician was Dr. James B. Herrick. Dr. Irons took Mr. Noel's medical history and did a routine blood and urine analysis. The blood smear showed that…

Walter Noel was anaemic

Blood smear from patient

Normal blood smear





… the patient was anaemic (~1/2 the normal number of red blood cells) and many of these had an elongated, sickle shape.

First clinical report on “sickle cell anaemia (SCA)

Drs. Irons and Herrick followed Mr. Noel’s medical saga for 2.5 years. He was admitted several times with respiratory problems, muscle and joint pain, etc. The two physicians consulted with other physicians, with current literature, etc., but were unable to make a diagnosis. They postulated many causes ranging from hookworm disease to malaria to parasites, etc. In 1910, Dr. Herrick published his observations in an article entitled “*Peculiar elongated and sickle shaped red blood corpuscles in a case of severe anemia*” in Arch. Int. Med. (1910) 6:517-521, an article that was reprinted in 2001.

YALE JOURNAL OF BIOLOGY AND MEDICINE 74 (2001), pp. 179-184
Copyright © 2001. All rights reserved.

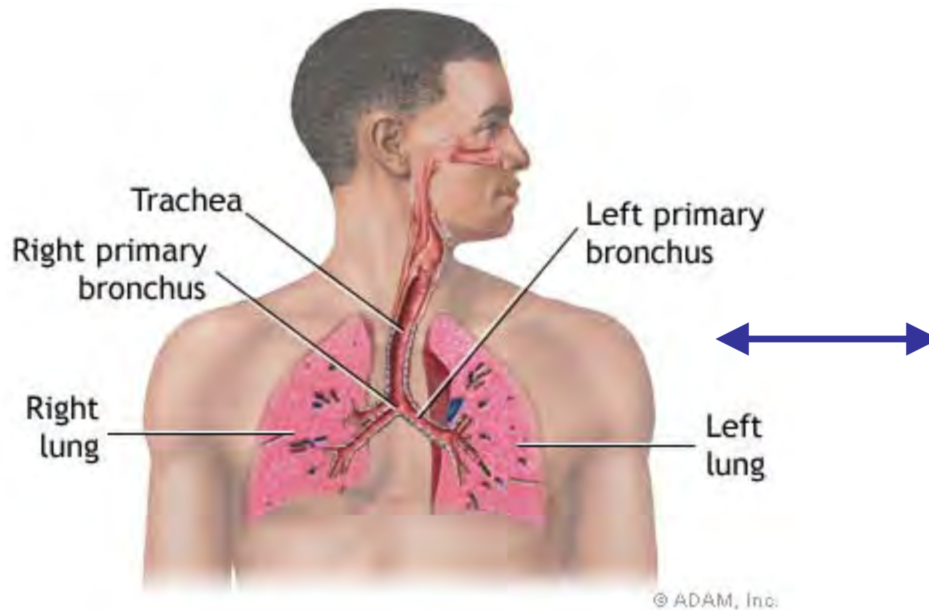
CLASSICS OF BIOLOGY AND MEDICINE

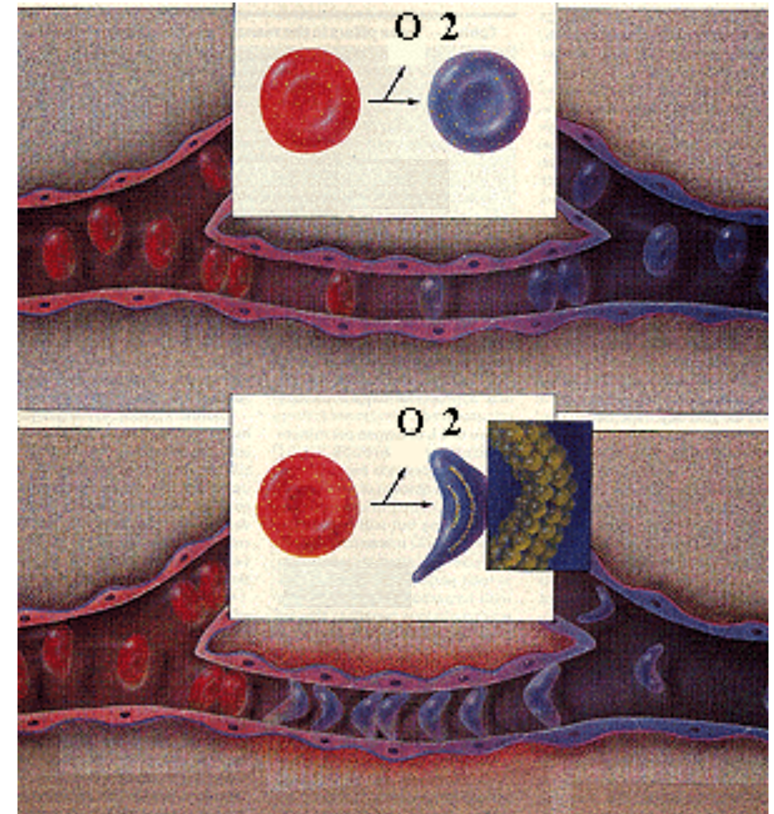
Peculiar Elongated and Sickle-shaped Red Blood Corpuscles in a Case of Severe Anemia[a]

James B. Herrick, M.D.

1013 State Street, Chicago, Illinois

What is the cause of SCA?

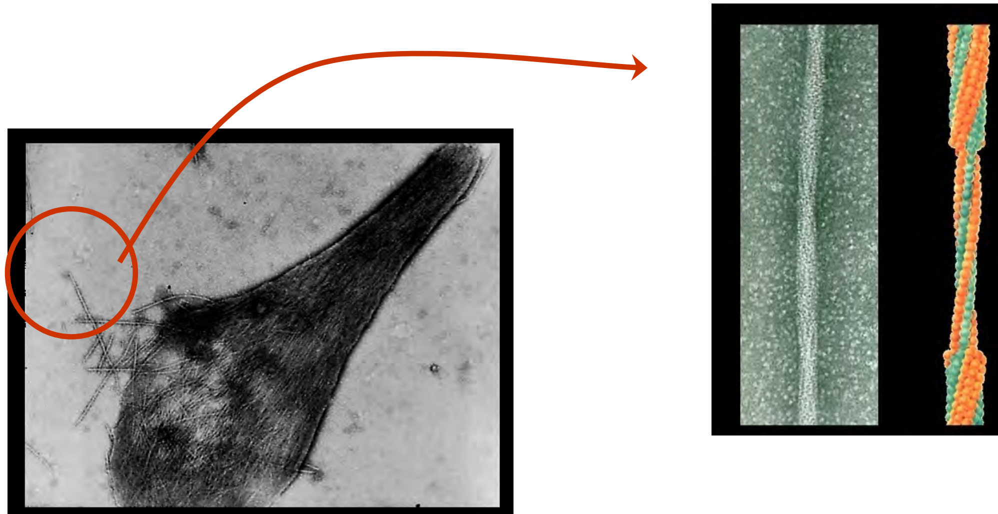




At low $[O_2]$, blood cells adopt a sickle shape that does not pass easily through capillaries thus impairing circulation leading to peripheral tissue damage. *But why do red blood cells adopt a sickle shape?*

Why do cells adopt a sickle shape?

The cells become quite fragile at *low [O_2]* because of the formation of fibers inside the cells, leading to the sickle cell shape. In fact, after prolonged exposure to low [O_2], the cells often rupture.



Given that the most abundant protein in the RBC is hemoglobin (Hb), and that sickling only occurs at low [O_2] this led to the hypothesis:

"Hb from sickle cell anemic individuals (HbS) is different from Hb from normal individuals (HbA), and polymerizes at low [O_2]"

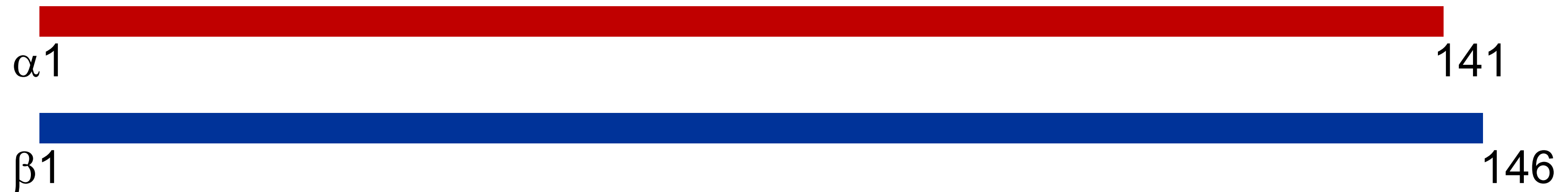
HbS is less negative than HbA

In the 1940s, Linus Pauling hypothesized that Hb in SCA patients (HbS) has a different "chemical structure". Pauling was unable to detect any difference between normal adult HbA and SCA HbS until he tried a new technique called electrophoresis:

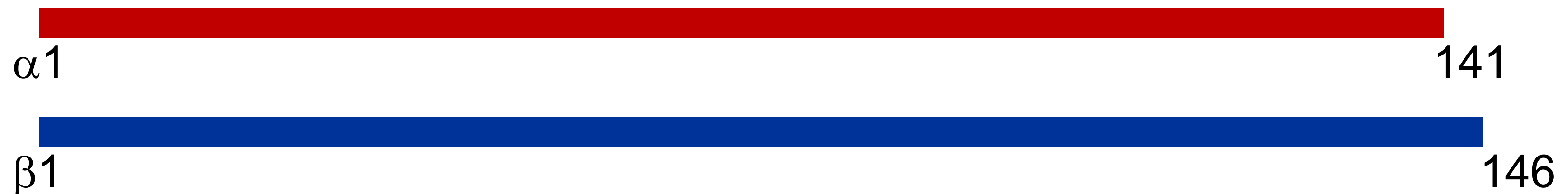
Electrophoresis showed that HbS has 2 - 4 fewer negative charges than HbA (Pauling et al. Science (1949) **110** page 543). *But what is the chemical basis of this change in charge? How do we figure this out?*

How do we ID differences between HbA vs HbS?

HbA is a tetramer of two α and two β chains:



HbS is a tetramer of two α and two β chains:



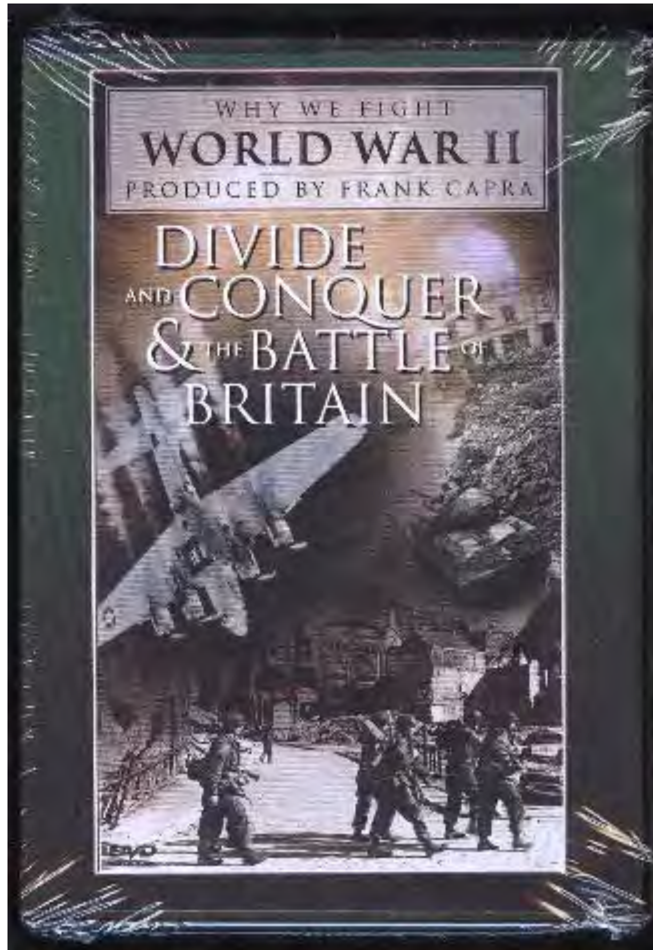
One possibility is to determine the sequence of amino acids from the N-terminus to the C-terminus of both the α and β chains in HbA versus HbS.

How do we ID differences between HbA vs HbS?

But where is mutation, and how many mutations are there?

One must sequence the entire polypeptide chain to be certain that one identifies all chemical changes. Unfortunately, sequencing long polypeptide chains was almost impossible and took a long, long time. *Is there a better way?*

In 1957, V. Ingram devised a better way

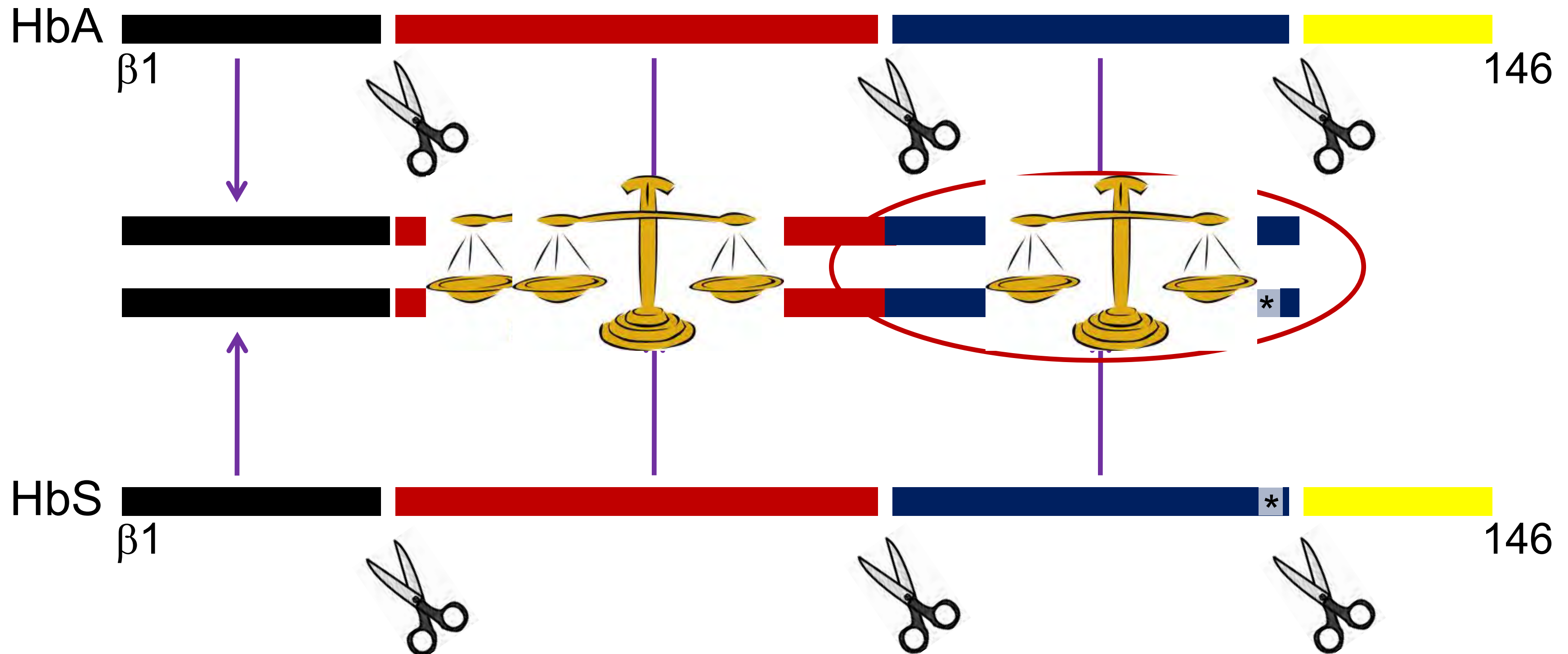


Vernon Ingram

Vernon Ingram devised a brilliant but simple approach to solve this problem. He used the trial-and-error method – divide and conquer! (V. M. Ingram, *Nature* **180**, 326 (1957)):

Instead of sequencing the entire HbA and HbS...

V. Ingram divided each chain into small pieces



Separately, cut the β chains of both HbA and HbS into small peptides and then compare each of the fragments from HbA and HbS to ID the one(s) that contains the mutation - then sequence these peptides to ID the difference between HbA and HbS! (*refers to the mutation)

Peptide Map of HbA vs HbS

In the mid 1950s Vernon Ingram developed a technique called protein fingerprinting (peptide mapping) to identify the mutation in HbS:

1) purify both normal (HbA) and sickle cell anemic (HbS) hemoglobin

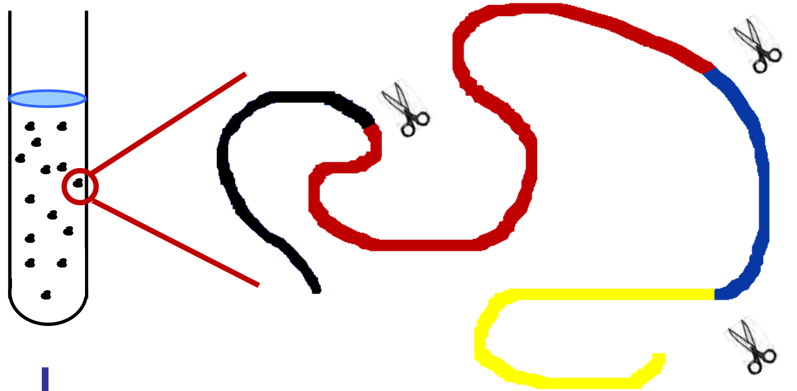
2) treat both samples with a protease, such as trypsin or chymotrypsin, which preferentially cleave after either large positive (Lys, Arg) or aromatic (Phe, Tyr, Trp) residues, respectively. Given the natural prevalence of these amino acids in proteins, one would expect each to cut a 100 residue protein ~ 10 times leading to ~11 short peptides varying in length from 5 -15 residues. Each peptide corresponds to a different region of the polypeptide sequence

3) separate out the individual peptides and those with chemical differences by either HPLC or, in the old days, a 2D separation involving electrophoresis and chromatography

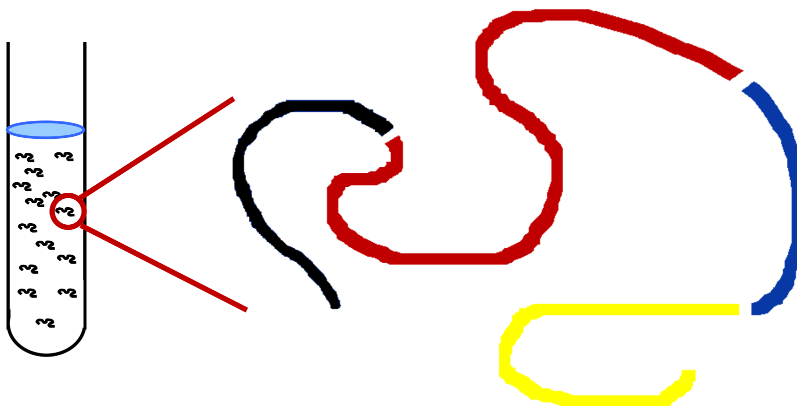
4) sequence the identified short peptide to identify chemical modification

How do we cut the polypeptide chain?

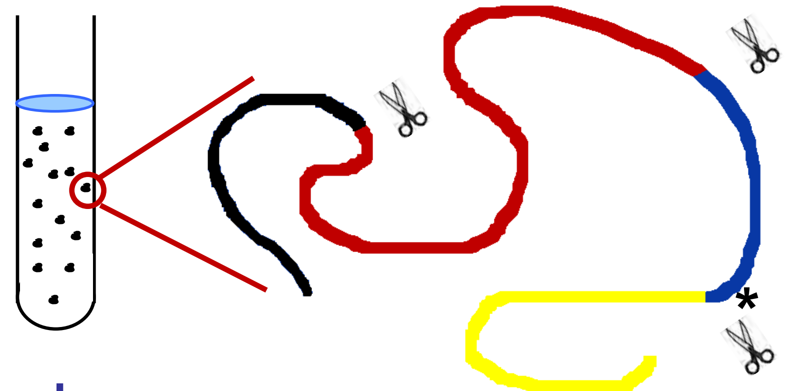
HbA



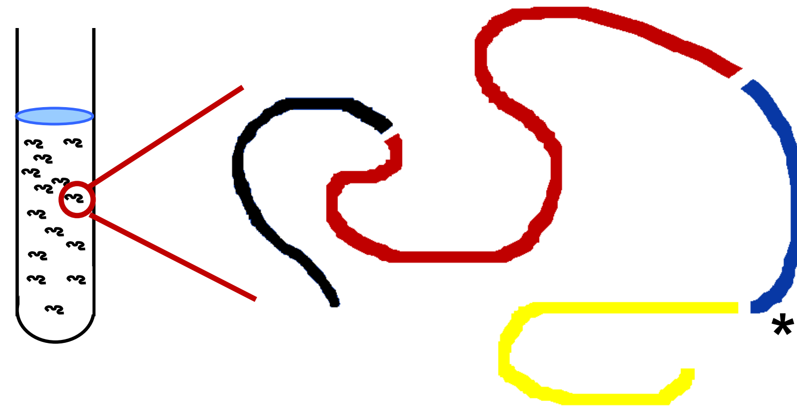
Chymotrypsin
(cuts after aromatic residues)



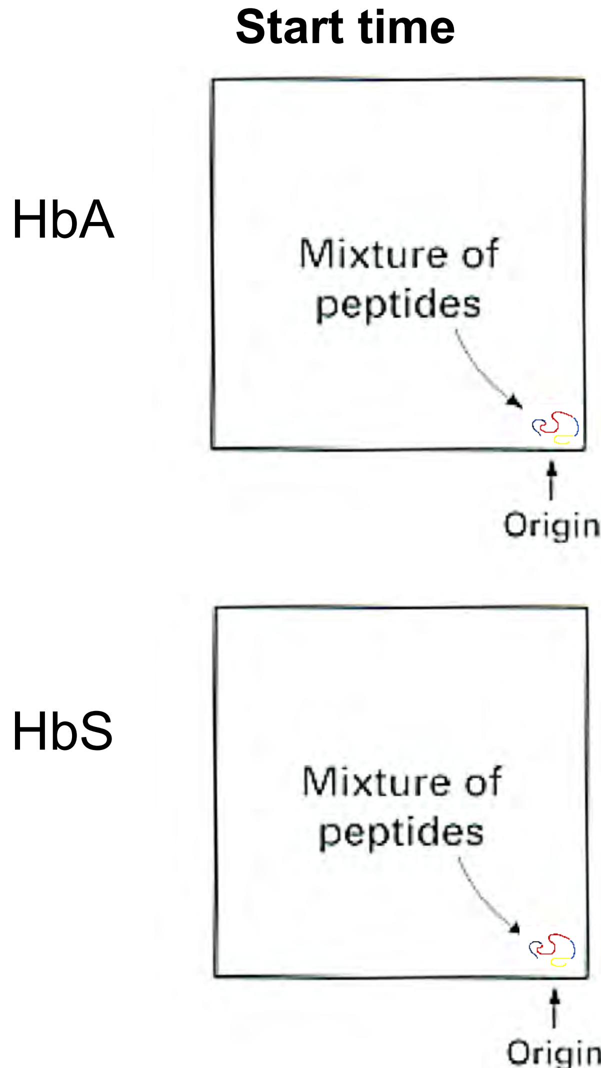
HbS



Chymotrypsin
(cuts after aromatic residues)

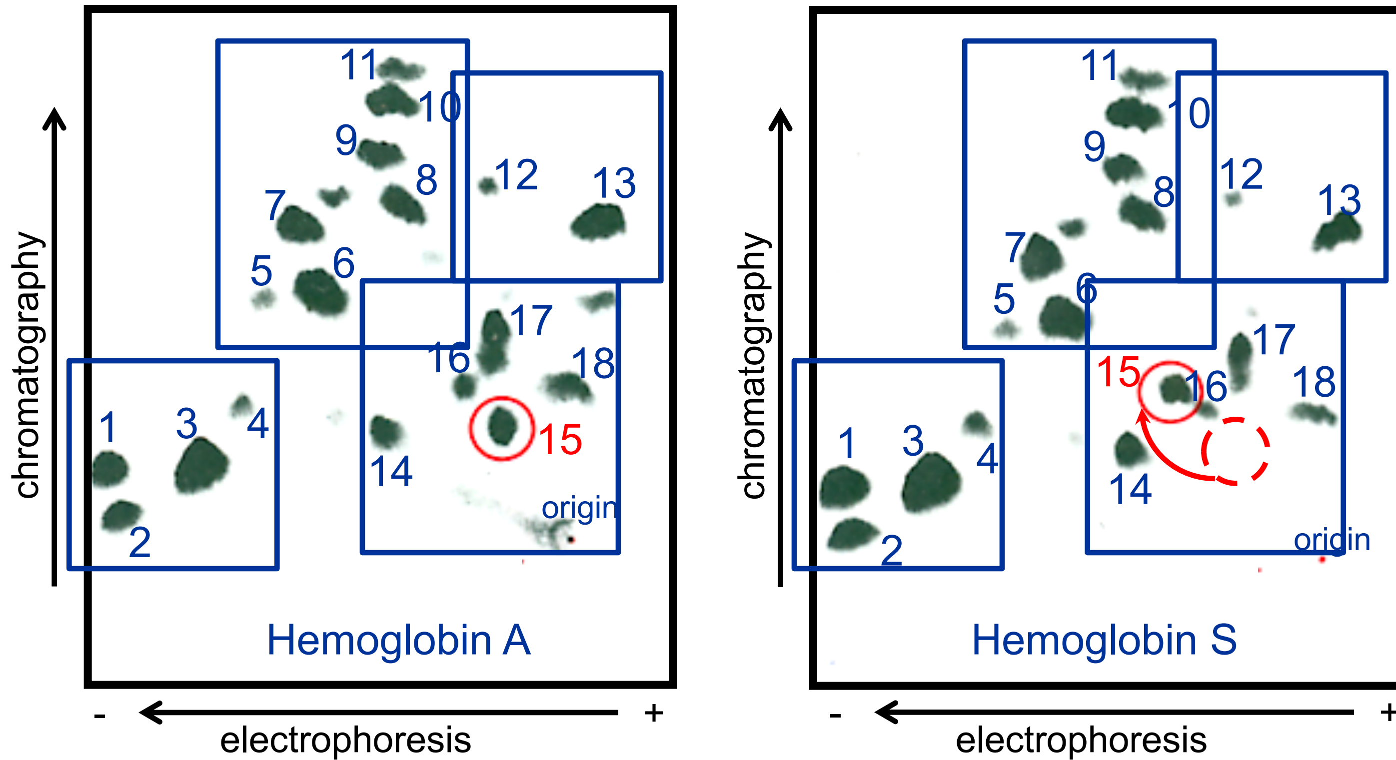


Compare peptides by 2D electrophoresis/chromatography



Identify the short polypeptide that contains the mutation – then extract and sequence the peptide to identify the mutation.

Peptide maps of HbA and HbS



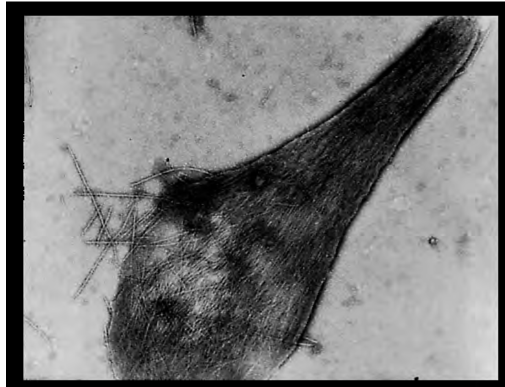
Hb_A: **Val-His-Leu-Thr-Pro-Glu-Glu-Lys**

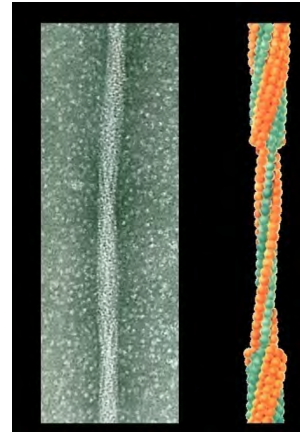
Hb_S: **Val-His-Leu-Thr-Pro-Val-Glu-Lys***

* The mutation is at position 6 in the β chain

Why does mutation lead to the formation of HbS fibers?

Ingram showed that HbS has a Glu (negative charge) to Val (hydrophobic) mutation at position 6 in the β subunit (note that Hb is a tetramer of two α and two β subunits, $\therefore$HbS has two less negative charges than HbA).

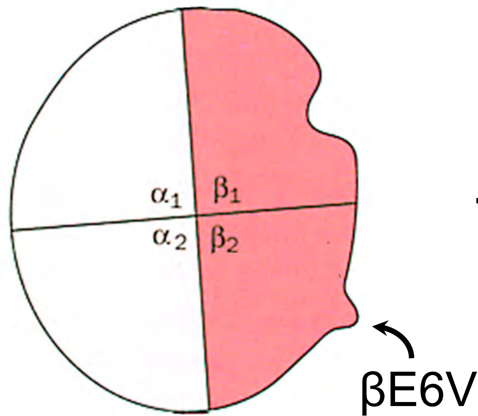


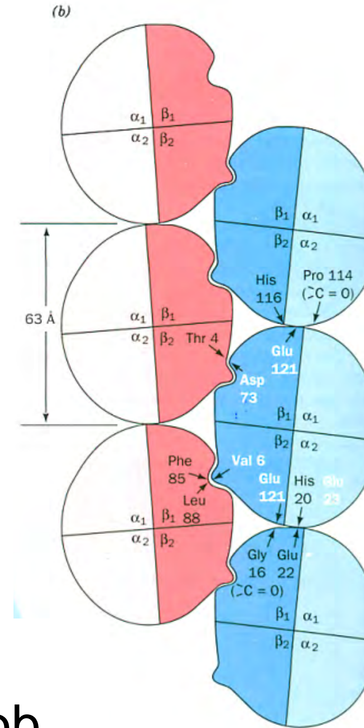


But why/how does this mutation lead to polymerization of HbS into long polymers?

Why does mutation lead to the formation of HbS fibers?

1) *How does this mutation lead to polymerization of HbS into long polymers?*





1) *Hypothesis:* the hydrophobic βGlu6Val mutation is on the surface creating sticky knob that binds to a hydrophobic pocket on an adjacent molecule of HbS.

2) *But why does polymerization only occur at low [O_2]?*

Crystal structure of myoglobin and hemoglobin

1962 Nobel Prize (Chemistry)

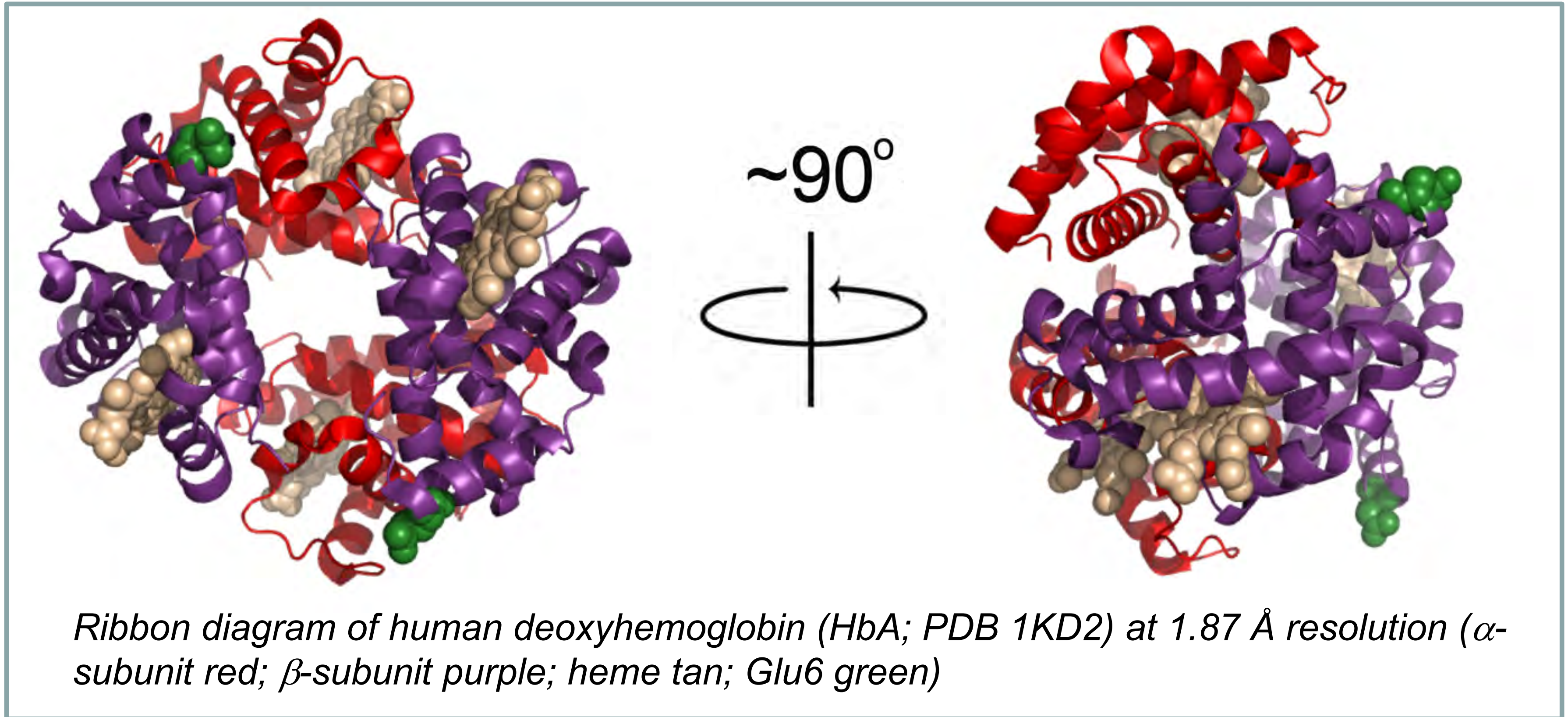
Max Perutz
Sir John Kendrew

“for their studies of the soluble structures of proteins

In the 1940s and 1950s, John Kendrew and Max Perutz developed methods for determining the 3D structures of proteins at atomic resolution. These methods were first applied to myoglobin and then to hemoglobin – including hemoglobin from sickle cell anemic patients.

Atomic Resolution Structure of Mb - 1958

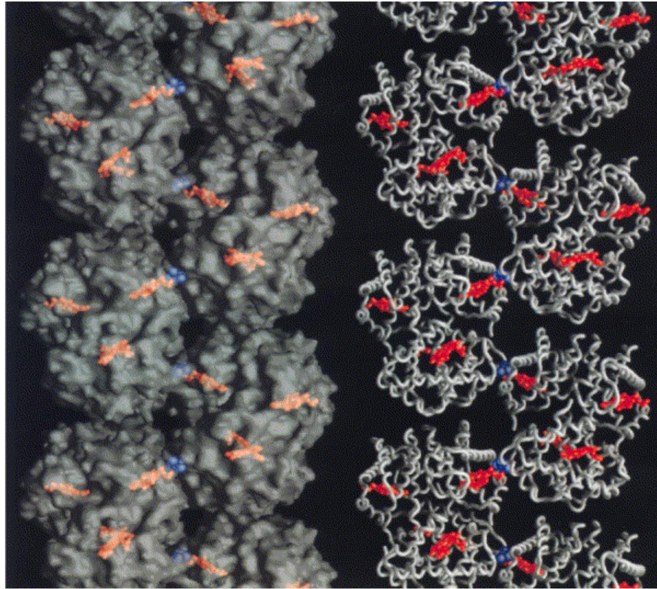
Structure shows Glu6 is on the surface of HbA



Ribbon diagram of human deoxyhemoglobin (HbA; PDB 1KD2) at 1.87 Å resolution (α-subunit red; β-subunit purple; heme tan; Glu6 green)

Crystal structure of Hb shows that the βE6V mutation is on the protein surface thus supporting the hypothesis that the hydrophobic effect drives polymerization, but we still need to know *why the βE6V mutation leads to long polymers and why polymerization only occurs at low [O_2].*

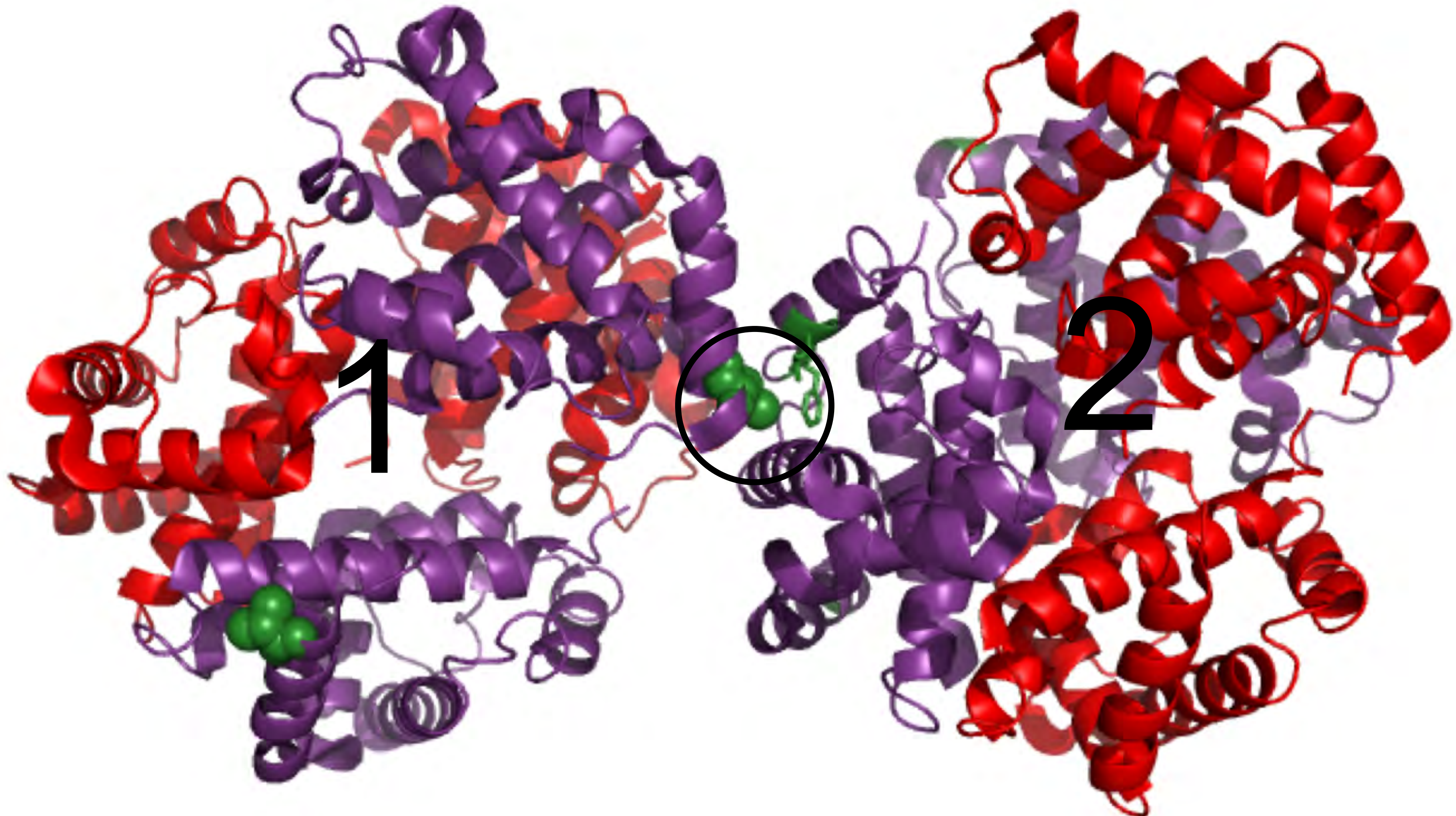
Structure of HbS at low [O_2]



Crystal structure of HbS polymers at low [O_2] shows the hydrophobic contacts.

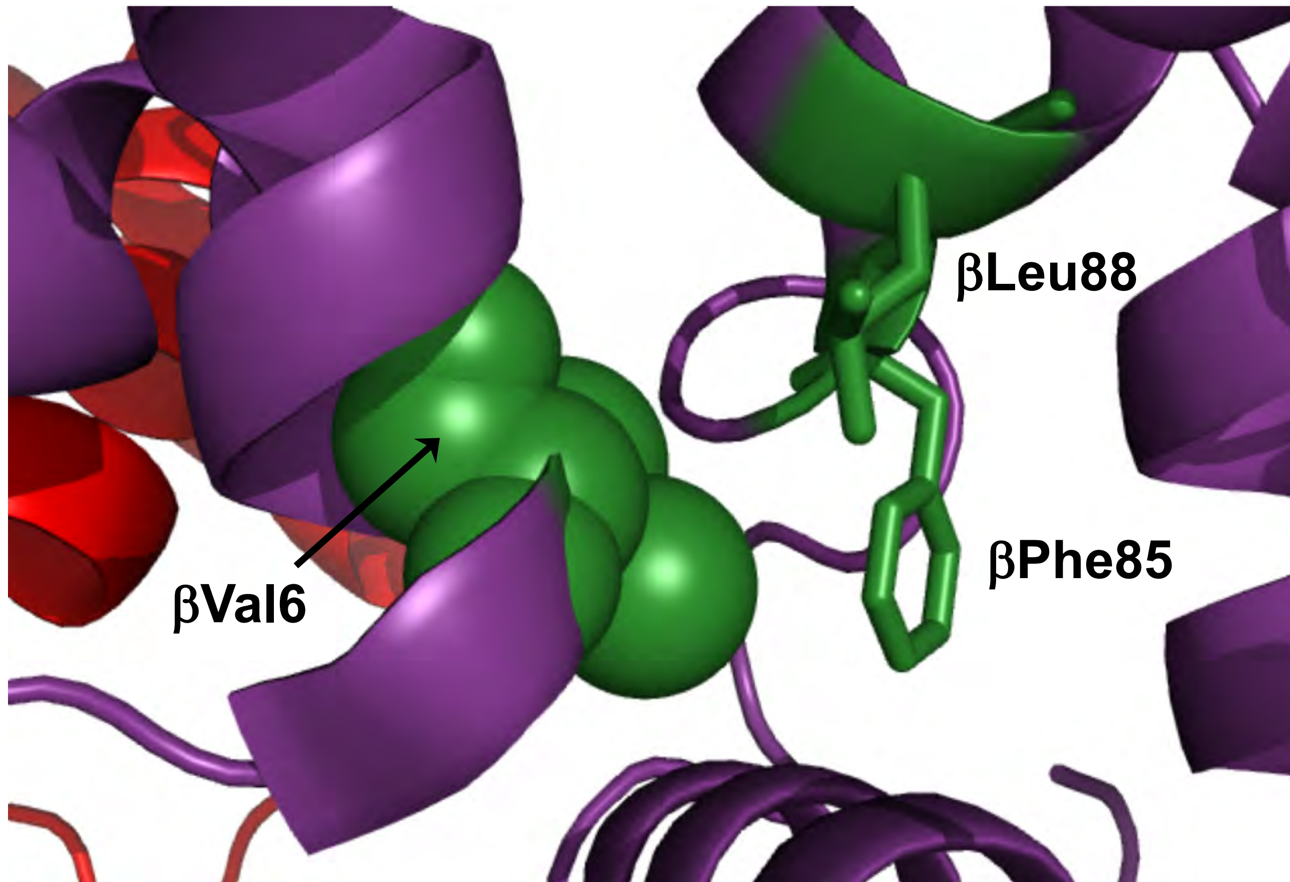
Harrington et al (1997) *J. Mol. Biol.* **272** *pp.* 398 (PDB 2HBS)

Glu to Val substitution leads to hydrophobic contacts



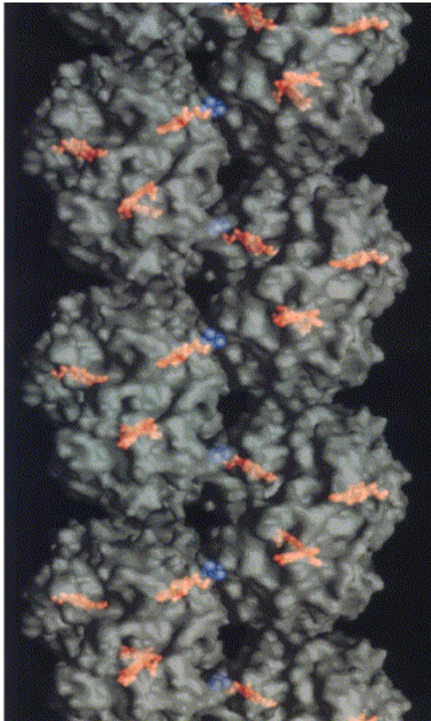
An HbS dimer of tetramers at low [O_2]. βVal6 (green spheres) from HbS tetramer 1 fits into a hydrophobic pocket on HbS tetramer 2 (pocket formed by βPhe85 and βLeu88 and shown as green sticks).

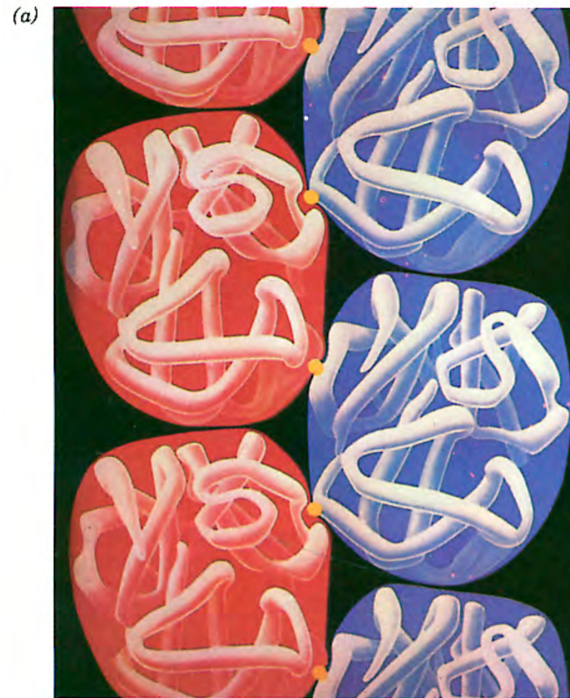
Close up of the Val interactions with Leu88 and Phe85

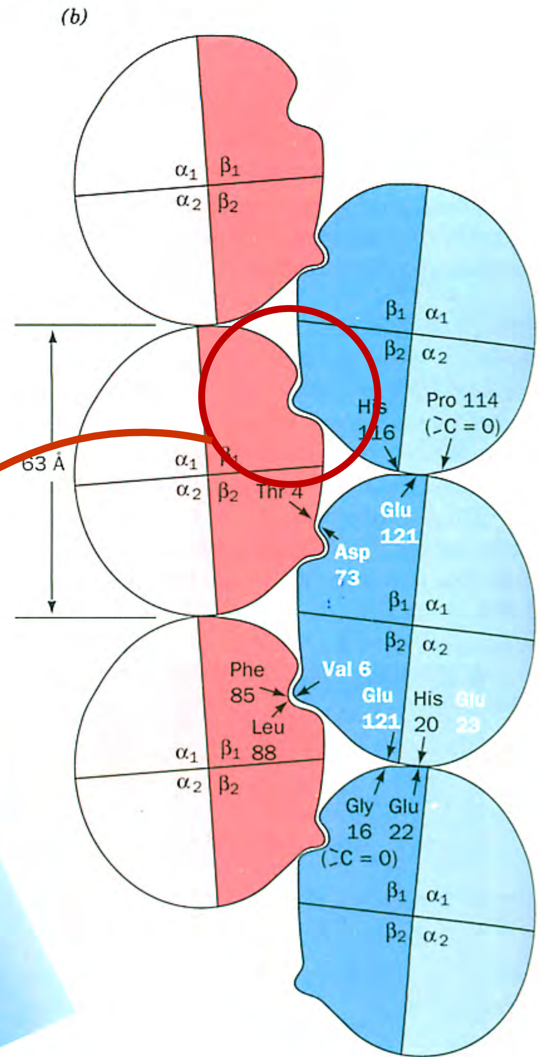


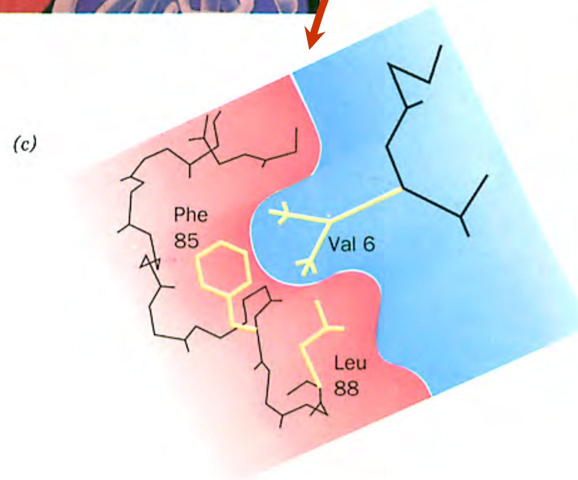
But why do we get long polymers of HbS?

Structure of HbS at low [O_2]



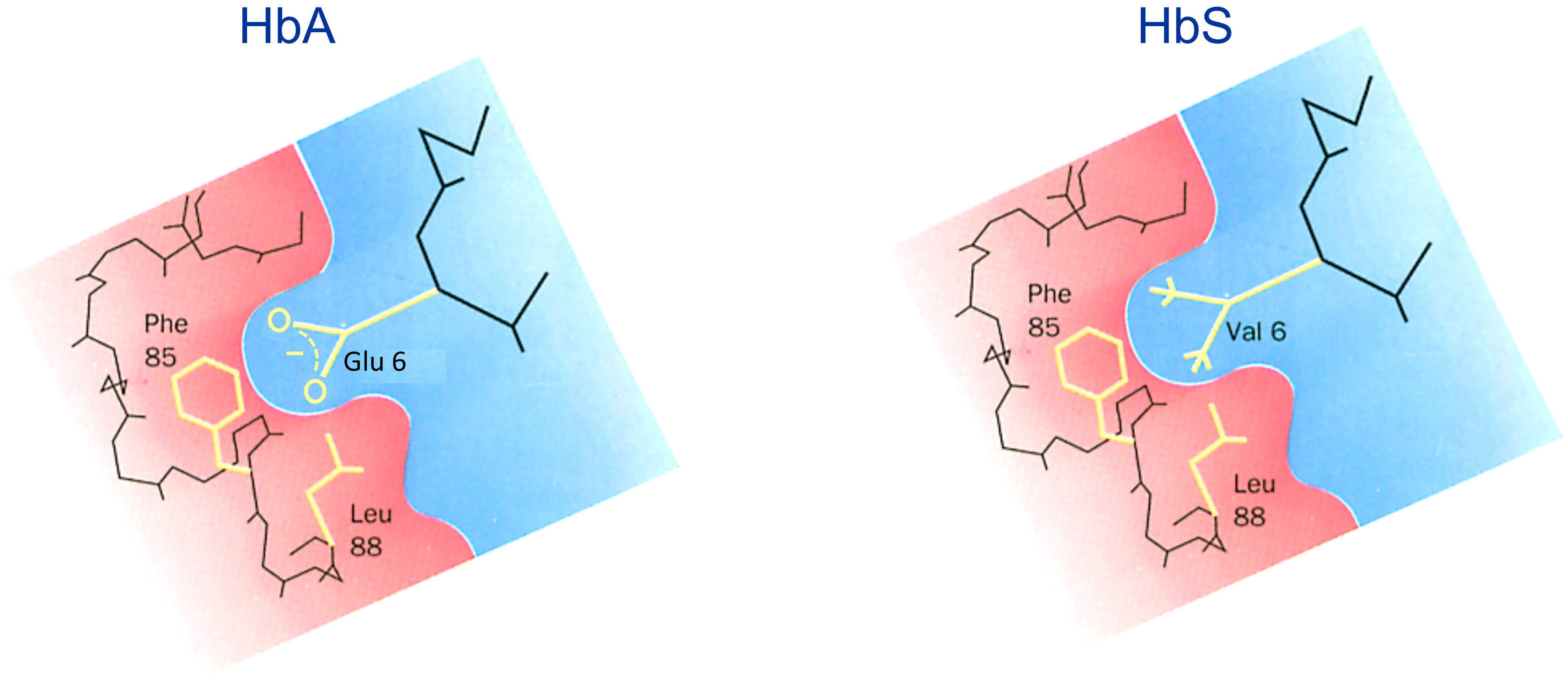






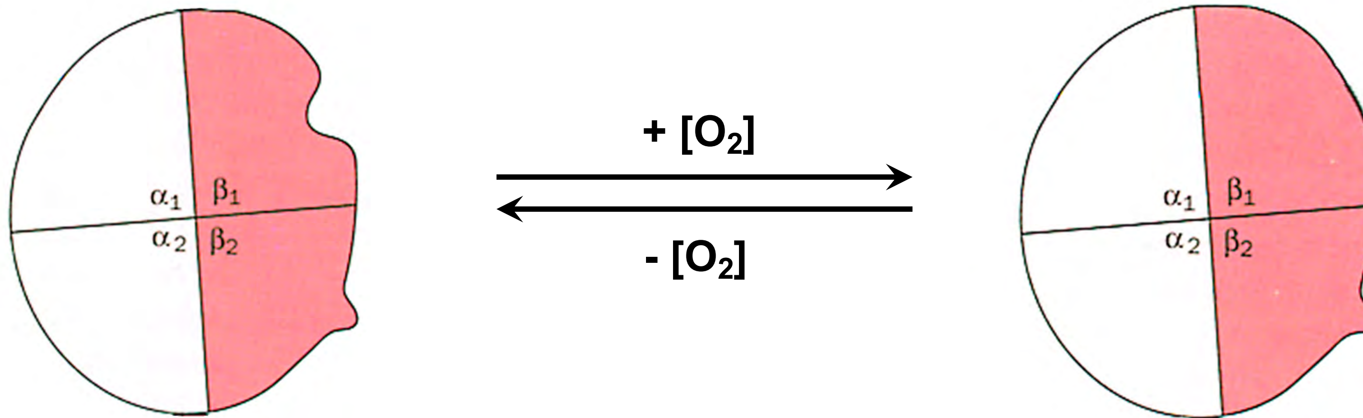
By a fluke of nature, each end of an HbS α/β dimer can interact with another dimer leading to long polymers.

Structure of HbA vs HbS at low [O_2]



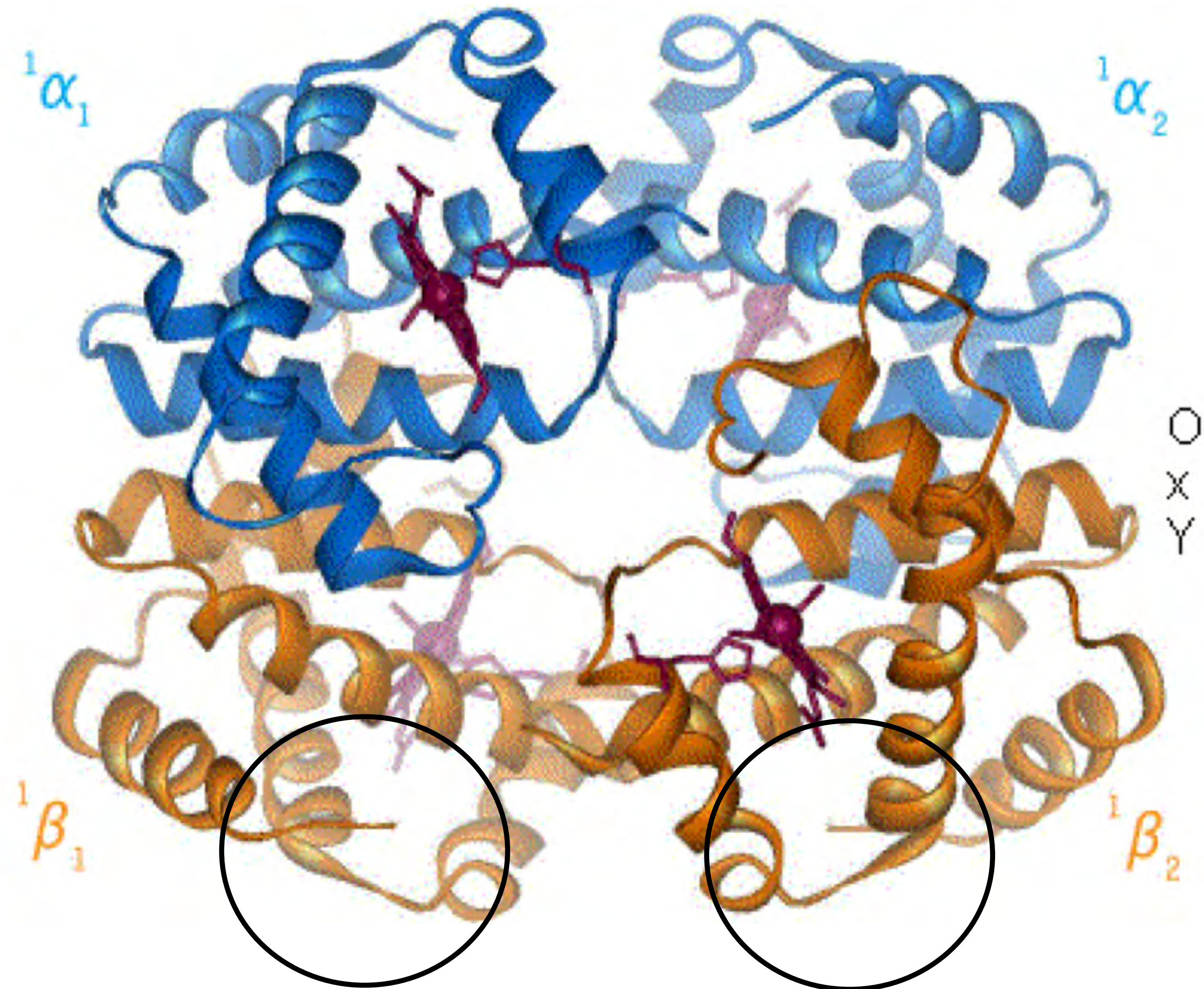
In HbA, the interactions between Glu6 and Phe85/Leu88 are not favorable enough for dimerization/polymerization of HbA to occur – i.e. the Glu6 residue would have to interact with the hydrophobic pocket (*left*). In HbS, the hydrophobic effect drives Val to interact with Phe/Leu and thus polymerization (*right*). *But why does polymerization occur only at low [O_2]?*

Why does HbS only polymerize at low $[O_2]$?



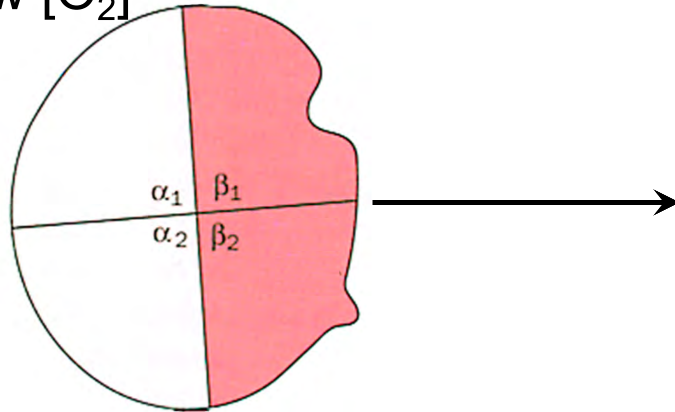
Structural biology shows that there is a change in shape of HbS upon the binding of O_2. The change in shape eliminates the binding pocket formed by Phe 85 and Leu 88 so that at high [O2] the binding pocket is no longer available to interact with Val 6.

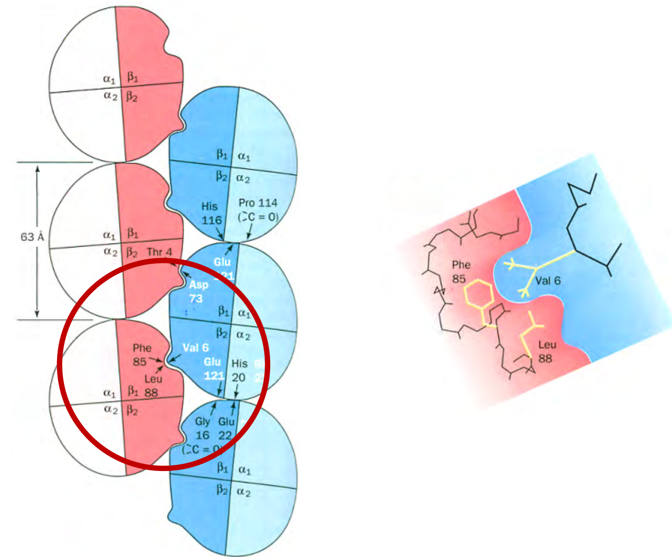
Formation of hydrophobic binding pocket at low [O_2]



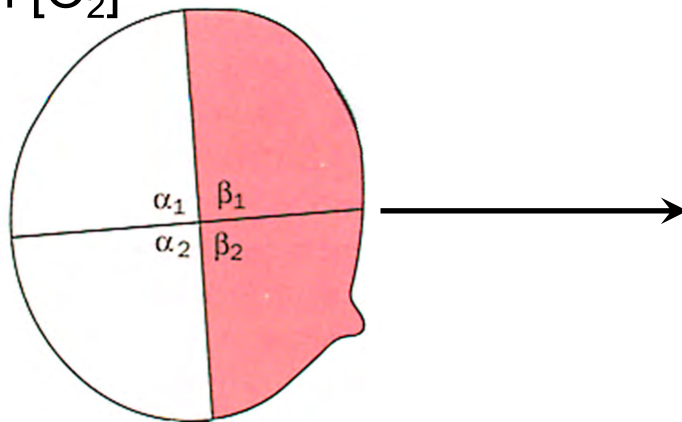
Hydrophobic pocket at low [O_2] allows polymerization

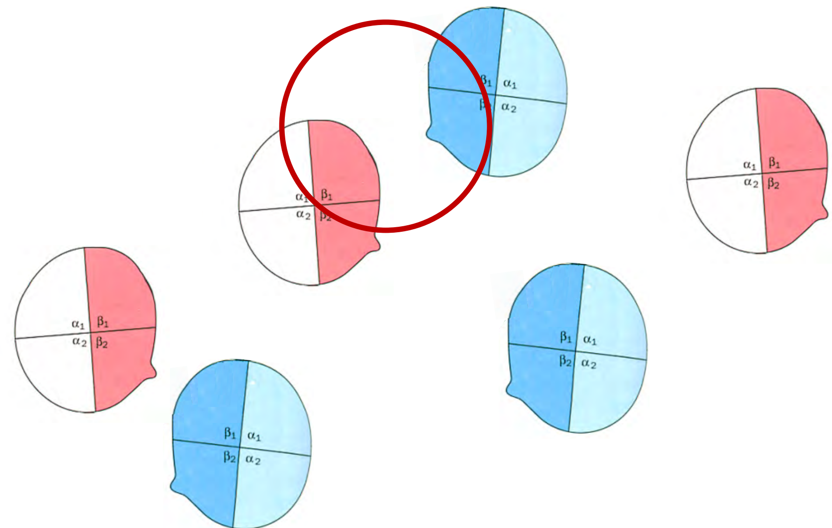
Low [O_2]



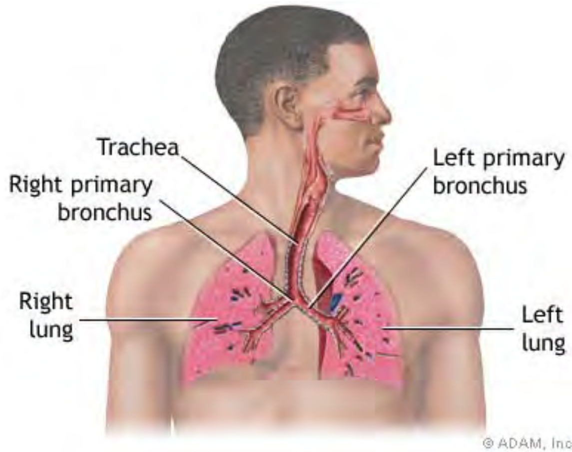


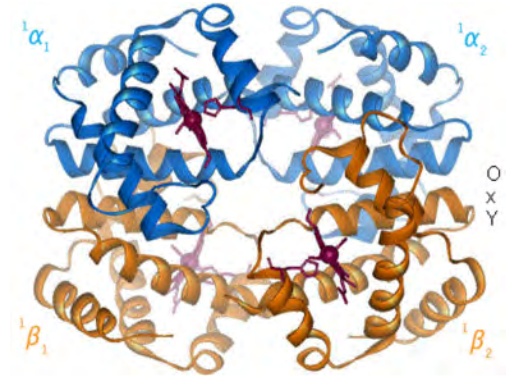
High [O_2]



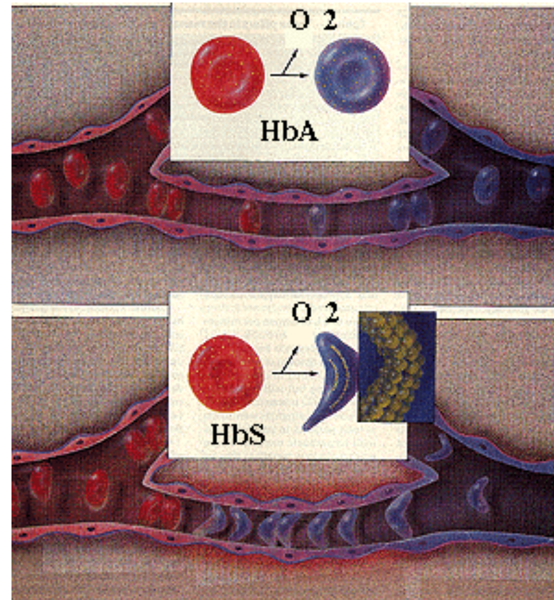


Molecular basis of Sickle cell anemia





βGlu6 to βVal6 mutation



Conclusions:

Linus Pauling's studies of sickle cell anaemia were significant for many reasons, but most importantly they demonstrated for the first time that:

- Human disease can be caused by a change in the chemistry (i.e. the amino acid sequence) of a protein. Human disease thus has a chemical or "biochemical" basis

  $\therefore$ if we can identify mutations in proteins and understand how mutations influence function, we may learn how to treat human disease

Conclusions:

The studies described today demonstrate fundamental principles of biochemistry:

- even though our interest is to understand human biology and disease, this understanding inevitably requires a finer and finer “microscope” to ultimately study cellular function at the biochemical level – the level of proteins, DNA, RNA, etc.
- protein structure dictates function
- Innovation in experimental approaches are required for advances in biology and medicine – we have to understand technology
- non-covalent interactions are fundamental to protein structure and function and thus play a central role in all of biology (and disease)