HEMoglobin S Gelation and Sickle Cell Disease

By William A. Eaton and James Hofrichter

The fundamental cause of sickle cell disease is the decreased deformability of the sickled red cell produced by gelation of hemoglobin S. Partial inhibition of gelation should therefore reduce clinical severity, while complete inhibition should result in a "cure." These basic ideas have stimulated an enormous effort to understand the gelation process in detail and to relate the results of these studies to the pathophysiology of sickle cell disease. Discoveries concerning gelation have also led to new lines of research on a specific therapy. The early finding that fetal hemoglobin inhibits gelation, has ultimately led to the development of methods to increase the production of F cells in the bone marrow of sickle cell patients, while the discovery of the enormous sensitivity of the rate of gelation to hemoglobin concentration has stimulated studies on the reduction of intracellular hemoglobin concentration as a means of therapy. Studies on the structure of the hemoglobin S polymer, moreover, have guided the development of agents designed to inhibit gelation by interfering with the formation of intermolecular contacts in the polymer.

The purpose of this article is to review recent developments in the relation between hemoglobin S gelation and sickle cell disease. We present our current understanding of the major features of the gelation process. Since gelation is a physical rather than a chemical process, its description necessarily requires more physical detail than that of most biological processes. From these studies we are able to develop a more rigorous and comprehensive description of the relation between gelation and the pathophysiology than has been possible up to now. By combining the gelation studies with work on the rheology of sickle cells and blood flow in the microvasculature, a clearer picture emerges of the outstanding issues in understanding the mechanism of vasoocclusion in patients and the resulting cardiovascular response. Finally we discuss the variation in clinical severity and analyze the problem of inhibiting gelation in patients. Throughout this discussion we shall see that the kinetics of gelation is a dominant factor in understanding gelation both in vitro and in vivo, and it will become clear that discussions of the pathophysiology that do not include a kinetic analysis are inadequate.

A broader treatment of sickle cell disease, including genetic and clinical aspects, has recently appeared in two excellent books. Also the structure, physical chemistry, and rheology of hemoglobin S gelation in solution and in red cells is discussed much more extensively in an article that is being published elsewhere.

GELATION AT EQUILIBRIUM

To understand gelation we first must describe a gel at equilibrium. As shown in Fig. 1, a gel can be separated into two phases, a solution phase that contains free hemoglobin molecules and a polymer phase. The structure of the individual polymers has now been determined in considerable detail.

It is a fiber made up of 14 intertwined helical strands of hemoglobin S molecules. The fiber can alternatively be described in terms of seven intertwined double strands of molecules in which the double strands have a structure that, except for a slight helical twist, is nearly identical to the double strands that form the fundamental unit of the deoxyhemoglobin S crystal. In each molecule one of the two b valines of the tetramer is involved in an intermolecular contact with its neighbor in the double strand. The structure of the deoxyhemoglobin S crystal is known to atomic resolution, so that there is a very detailed picture of the intermolecular contacts within the double strand that must be very similar to what occurs in the polymer.

A gel at equilibrium behaves very much like a suspension of microscopic protein crystals suspended in a saturated protein solution. The concentration of hemoglobin in the solution phase, which is called the solubility, is a measure of the stability of the polymer phase. The solubility is determined experimentally by measuring the hemoglobin concentration in the supernatant obtained after high-speed sedimentation of the polymers. Because the concentration of hemoglobin in the polymer phase appears to be relatively constant, the fraction of the total hemoglobin that is polymerized can be calculated from the solubility using a simple mass conservation relation. There have now been systematic investigations of the solubility under a wide variety of solution conditions. These include the dependence on temperature, pH, salts, 2,3-DPG, carbon monoxide, oxygen, and non-S hemoglobins.

The role of non-S hemoglobins in the gelation of hemoglobin mixtures has been the focus of a large number of studies, since early investigations showed that the presence of hemoglobins A, C, and F reduces sickling and is accompanied by decreased clinical severity. For mixtures of hemoglobins S and F (and S + A), a detailed analysis of solubility data, including the large contribution of nonideality arising from excluded volume effects, indicates that over the physiologic range of compositions there is little or no copolymerization of either the homotetramers, or the hybrid tetramers, or the non-S hemoglobins. The low probabilities for copolymerization of these molecules can be rationalized as resulting from destabilizing effects on the intermolecular contacts of the double strand that accompany specific
Fiber consists of 1-4 strands and that it can be determined from measurements of the solubility as a function of solution phase saturation. Also, because there is no aggregation in the solution phase of the gel and the β6 mutation has no effect on the intrinsic affinity of the hemoglobin molecule, the binding curve for the solution phase hemoglobin S molecules is normal. The major problem has been to obtain the polymer binding curve, which was accomplished using an optical technique called linear dichroism.

The principal experimental results are shown in Fig 2. The solubility increases slowly at low oxygen saturation, then increases sharply at high saturations (Fig 2b). The most interesting finding from the binding studies is that the polymer binds oxygen noncooperatively, as evidenced by a slope of unity in a Hill plot. The two-state allosteric model, which has provided an excellent framework for interpreting a wide variety of experiments on hemoglobin, provides a simple molecular interpretation of these results. According to this model a hemoglobin molecule free in solution exists in one of two affinity states at all stages of oxygenation. The low-affinity state, called T, has the quaternary structure of completely deoxygenated hemoglobin, while the high-affinity state, called R, has the quaternary structure of the fully oxygenated molecule. Binding to either quaternary structure is noncooperative. Cooperativeness arises from the continuous conversion of low-affinity T-state molecules to high-affinity R-state molecules as the saturation increases (Fig 2a). The simplest extension of this model to the gelation of hemoglobin S is to postulate that all T-state molecules polymerize with equal probability independent of the number of oxygen molecules bound and that there is no polymerization of R-state molecules.

The model is based on the idea that R-state molecules do not polymerize because their structure is sufficiently different from T-state molecules that they cannot fit into the polymer lattice. Analysis of the structure of the double strand of the deoxyhemoglobin S crystal does indeed show that it is impossible to replace the T-state molecule with an R-state molecule. Since the polymer contains only T-state molecules, the model predicts that it will bind oxygen noncooperatively, exactly as observed. The model is not quantitatively perfect, however, because the affinity of the polymer is slightly lower than that of solution T-state molecules (Fig 2a). Evidence for this small difference is also found in the solubility data (Fig 2b), indicating that T-state molecules with oxygen bound are partially discriminated against by polymers. The small conformational changes that are known to take place within the T quaternary structure upon oxygen binding could explain this result.

The picture that emerges, then, is that the simplest extension of the two-state allosteric model provides an excellent description of the effect of oxygen on gelation.

The results of these experiments can be used to explain data on polymerization in sickle red cells. The first step is to consider further the oxygen binding curve of a gel. As mentioned above, at a given oxygen pressure the saturation of the gel is a weighted average of the saturations of the solution and polymer phases. Figure 2c shows gel binding curves under near physiologic conditions that have been constructed from solution and polymer phase binding curves (Fig 2a) and from the solubility curve (Fig 2b). As the

![Diagram of a gel of hemoglobin S at partial saturation with oxygen. The 64 kD molecule (ie, the tetramer) is represented as a circle. A gel of hemoglobin S contains large polymers, often called fibers, and a concentrated solution of free molecules with one or more oxygen molecules bound. There are relatively fewer filled circles in the polymer because it has a lower oxygen affinity than the solution. The structure was determined by Edelstein and coworkers using electron microscopy and image reconstruction techniques.](image-url)
Fig 2. Effect of oxygen on gelation in solution and in sickle cells. (A) Solution, polymer, and theoretical R- and T-state binding curves. The solution binding curve is the binding curve for normal blood. The polymer binding curve is calculated from the data in phosphate buffer of Sunshine et al. The R- and T-state binding curves are theoretical and were obtained by fitting to the solution binding curve with the two-state allosteric saturation function. The dotted curve shows the theoretical solubility for the hypothetical case in which all T-state molecules polymerize with equal probability, independent of the number of oxygen molecules bound. (C) Gel-binding curves at different total hemoglobin concentrations calculated from the results in (A) and (B). (D) Fraction polymerized as a function of total fractional saturation calculated from the results in (A) and (B). (E) In vivo oxygen binding curve calculated for a population of cells having the distribution of intracellular hemoglobin S concentrations in (G). The binding curve calculated in the absence of polymer (long-dashed curve) and the binding curve calculated when polymer is present at equilibrium (solid curve) are shown for reference. The data points are for SS blood from Winslow. The oxygen unloading curve under in vivo conditions (dotted curve) was calculated by requiring that only the densest 18% of the cell population contain polymer at the average venous PO2 found in SS patients (46 torr) denoted by the arrow. These conditions were simulated by requiring that each cell be sufficiently supersaturated for polymerization to occur within about 200 ms at each PO2. (F) Fraction of total hemoglobin S that is polymerized. The fraction is calculated under equilibrium conditions (curve) and under in vivo conditions (dotted curve). (G) The average distribution of intracellular concentrations from a study of 43 patients by Fabry et al used in the calculations in panels (E) and (F). The probability density, P, in dL/g is plotted v the intracellular hemoglobin concentration, C, in g/dL. The blackened area shows the subpopulation of cells that contain polymer at PO2 of 48 torr. The equations and parameters used in the calculation of all of the above panels are given by Eaton and Hofrichter, which are derived from the work of Sunshine et al.

Oxygen pressure increases, not only do the saturations of both phases increase but the solubility also increases, decreasing the contribution of the low affinity polymer phase to the total binding curve. As a result the gel binding curves appear to have lower-than-normal affinity with higher-than-normal cooperativeness. At sufficiently high oxygen pressures the solubility exceeds the total hemoglobin concentration, and the binding curve superimposes on the normal solution binding curve (Fig 2 c). As the total hemoglobin concentration increases, the affinity of the gel decreases owing to the increased fraction of the low affinity polymer (Fig 2 d), and the oxygen pressure at which the polymer disappears also increases.

The oxygen-binding curve of a single red cell should be identical to the binding curve of a gel having the same composition (total hemoglobin S concentration, fraction fetal
hemoglobin, pH, 2,3-DPG concentration, etc). Oxygen-binding curves of sickle blood are an average of the gel-binding curves for the individual cells. As in gels, the binding curve for sickle blood is "right shifted" compared to normal blood. Although 2,3-DPG levels are elevated in sickle cell blood, the formation of the low-affinity polymer is the major cause of the right shift in the blood-binding curve.\textsuperscript{77, 79}

The major difference between the binding curves for cell suspensions (Fig 2 e) and for gels (Fig 2 c) results from the wide distribution of hemoglobin S concentrations, which varies from about 20 g/dL in F cells to almost 50 g/dL.\textsuperscript{78, 81} This distribution produces a wide range of median affinities within the red cells from a given patient and hence smears the characteristic features of the gel-binding curve (compare Figs 2 c and 2 e). To calculate blood-binding curves it is necessary to utilize the results of recent investigations that have characterized the distribution of total hemoglobin concentrations from density measurements.\textsuperscript{62, 75, 76, 82, 88} The density distributions for SS cells are broader and more variable than those for normal individuals. Since the distribution of intracellular hemoglobin S and hemoglobin F concentrations was not determined for the cells employed in the oxygen-binding measurements, it is only possible to make qualitative comparisons between the observed binding curves and those calculated from solution data. Figure 2 e compares a whole blood oxygen-binding curve with the binding curve calculated from the solution data using the average concentration distribution from a study of 43 patients.\textsuperscript{26, 75} The p50s for the curves calculated from concentration distributions for individual patients vary from 37 torr to 46 torr, compared to the 33 torr to 45 torr observed in a study of 14 patients.\textsuperscript{77, 79} This comparison shows that the patient-to-patient variability in oxygen binding curves can be readily accounted for by the variability in intracellular concentration distributions. A more direct comparison on a relatively homogeneous cell population obtained by density fractionation, in which the 2,3-DPG and hemoglobin F levels were also measured, gives very good agreement between the cell and solution data.\textsuperscript{26, 79}

Gelation in isolated solutions and cells has also been compared using nuclear magnetic resonance techniques to measure the average fraction of polymerized hemoglobin as a function of the total saturation of the cells.\textsuperscript{76, 80, 81} The nuclear magnetic resonance measurements take advantage of the fact that the polymerized molecules do not rotate freely, making it possible to selectively measure the spectra of the polymerized and unpolymerized molecules.\textsuperscript{79} Measurements on a cell population of known concentration distribution are in good agreement with the curve calculated from the solution data.\textsuperscript{78} Agreement is also obtained in a comparison of density fractionated cells, although the experimental uncertainties are much larger.\textsuperscript{76}

At this point we should emphasize that the oxygen binding curves and polymer fraction curves that we have discussed are equilibrium or near-equilibrium curves and, as we shall see later, are very different from the in vivo situation in which most cells are very far from equilibrium because of the large kinetic effect of the delay time (Fig 2 e and 2 f).

**KINETICS AND MECHANISM OF GEL FORMATION**

The most unusual and interesting aspect of the gelation process is the kinetics and mechanism of gel formation. The simplest kinetic experiment takes advantage of the characteristic property that a hemoglobin S solution gels upon heating. A completely deoxygenated solution, having a concentration significantly less than the solubility at 0°C (< 30 g/dL), is heated to some temperature where the concentration exceeds the solubility. Polymer formation can be detected by a variety of techniques, including linear birefrin-
cence,6,43,45 turbidity,36,37,43,44,57,95-112 light scattering (G.W. Christoph and R.W. Briehl, unpublished results),113-115 viscosity,99,112,116-122 water proton magnetic resonance line-widths,124-128 and transverse relaxation times.124-128 and electron paramagnetic resonance.125 All of the techniques show the same type of time course. There is an apparent delay period during which there is no evidence for any aggregation, followed by the explosive appearance of

Upon lowering the temperature of a preformed gel, depolymerization proceeds much more rapidly and without a delay.6,94,95 The most striking finding from these studies is that the delay time is enormously sensitive to solution conditions, in particular to the hemoglobin S concentration. The inverse of the delay time is found to be proportional to the 30th to 50th power of the initial hemoglobin concentration.6,36,37 This is the highest known concentration dependence for a process taking place in solution. The delay time is also found to be directly proportional to the 30th to 45th power of the solubility, independent of the manner in which the solubility is altered.97 For example, in temperature-jump experiments, the delay time for a solution that is 10% saturated with carbon monoxide is increased by a factor of about 10 relative to deoxyhemoglobin S at the same total hemoglobin concentration, and at a saturation of 40% the delay time is increased by a factor of about 10.

The temperature-jump technique is limited to measuring delay times longer than about 100 seconds. To extend the kinetic measurements to physiologic times and hemoglobin concentrations required the development of a laser photolysis technique that could be used to prepare a completely deoxygenated hemoglobin S solution in less than a few milliseconds.131-133 In this technique the carbon monoxide complex of hemoglobin S, which is soluble up to at least 48 g/dL,134 can be converted to deoxyhemoglobin S by photodissociation under continuous laser illumination. The laser also serves as a source for monitoring gel formation from the change in light scattering. When the laser is turned off, the carbon monoxide recombines, the polymer disassembles to form a solution of monomers (ie, 64 kD hemoglobin S tetramers), and the experiment can be repeated indefinitely. Because the volumes of observation are as small as 10^-12 cc, the laser photolysis technique can also be used for investigating gelation in single red cells.135-137

A combination of the temperature jump and laser photolysis techniques has been used to examine the kinetics of gelation over a wide range of concentrations, temperatures, and times.132 Figure 3 shows that as the concentration decreases, the delay time increases from about 10 milliseconds at 40 g/dL to about 100,000 seconds at 20 g/dL. An important clue to the mechanism by which gelation occurs comes from a very unusual result, discovered in the course of the laser photolysis experiments, which is described in Fig 3. Highly reproducible delay times are observed for solutions with delay times of a few hundred milliseconds or less. When the delay times become longer than a few seconds, however, the delay times become very irreproducible, despite the fact that the progress curves have very similar shapes once polymerization begins (Fig 3 b to d).131-133 An important companion observation is that only a single birefringent domain of polymers forms when there are large fluctuations

![Homogeneous Nucleation](image)

Homogeneous Nucleation

- Fig 4. The double nucleation mechanism (Ferrone et al136). The two pathways for nucleation of polymers are shown. In the homogeneous pathway nuclei form in the solution, while in the heterogeneous pathway nuclei form on the surface of existing polymers. As more polymers form the increased surface area results in a continuously increasing rate of heterogeneous nucleation. This autocatalytic formation of polymers via the heterogeneous nucleation pathway is responsible for the appearance of a delay period prior to the observation of polymer. For both nucleation pathways there are competing thermodynamic forces. Initially aggregation is unfavorable because entropic forces tend to keep molecules apart. As the nuclei become larger, however, there is an increased number of bonds per monomer, 1/2 for a dimer, 3/3 for a trimer, 6/4 for a tetramer, up to 4.1 in the infinite polymer. As the aggregates grow this increase in the stability from more bonds per monomer finally overcomes the unfavorable entropic forces. The aggregate for which addition of monomer finally becomes favorable is called the critical nucleus.

![Heterogeneous Nucleation](image)

Heterogeneous Nucleation
All of these kinetic observations can be quantitatively explained by the double nucleation mechanism shown schematically in Fig 4. According to this mechanism gelation is initiated by the nucleation of a single polymer. This process is called homogeneous nucleation because it takes place in the bulk solution, and no surfaces are involved. By nucleation we mean that small aggregates of hemoglobin S molecules are unstable relative to monomers, and addition of a monomer to the aggregate produces a less stable aggregate. Once a certain size, called the critical nucleus, is reached, however, addition of a monomer produces a more stable aggregate, and monomers add endlessly to form a very large polymer. Nucleation results from competition between two thermodynamic forces, an increased freedom of motion (ie, increased entropy), which tends to keep molecules apart, and the favorable free energy of intermolecular, noncovalent bond formation that makes the molecules associate. Initially the entropy dominates, but once a sufficient number of intermolecular bonds per monomer are formed, aggregation becomes favorable (Fig 4). Although homogeneous nucleation by itself can explain the very high concentration dependence of the rate, it cannot explain the existence of a pronounced delay period. The delay period is produced by the second pathway for nucleation. In this pathway nucleation takes place on the surface of preexisting polymers and is therefore called heterogeneous nucleation. As more hemoglobin is polymerized, the surface area on which new polymers can be nucleated continuously increases, resulting in an autocatalytic polymerization for the initial stages of the gelation process.

Mathematical analysis of the double nucleation mechanism shows that the delay period is a manifestation of the autocatalytic formation of polymer via heterogeneous nucleation. For slowly gelling samples the model predicts that throughout the early portion of the measurable progress curve incorporation of monomers into polymers is exponential (region II in Fig 5 b). A consequence of this exponential polymerization is that there is an apparent delay, the length of which depends on the sensitivity of the measurement. No new or different process is occurring during the delay period. The rates of nucleation and growth of polymers are simply less than they are when polymers first become detectable. This has been verified in high-sensitivity-light-scattering measurements (G.W. Christoph and R.W. Briehl, unpublished results) 

According to the mechanism, both homogeneous and heterogeneous nucleation rates are proportional to the initial monomer concentration raised to a power that is the size of the critical nucleus. The enormous concentration dependence of the delay time can thus be readily explained as resulting from a large nucleus. As the hemoglobin S concentration increases, aggregation becomes more probable. As a result both homogeneous and heterogeneous nuclei become smaller, and the concentration dependence of the delay time decreases. Because the nuclei are in equilibrium with monomers, the large sizes for the critical nuclei also explain the high dependence of the delay time on the solubility. In concentrated phosphate buffer, gels form with a solubility of only 2 g/dL, suggesting stronger intermolecular bonds and smaller nuclei, and the concentration dependence is also much lower. Finally, the mechanism explains the irreproducibility of the delay time when single polymer domains are formed in small volumes as resulting from stochastic fluctuations in the time at which single homogeneous nuclei appear. Under these conditions a single polymer initiates the formation of a domain. The remainder of the polymers, which fill the entire observed volume, are formed by heterogeneous nucleation. This "amplification" of the homogeneous nucleation event allows the stochastic fluctuations to be observed.

We now turn to the important question of whether gelation
inside sickle cells proceeds at the same rates and by the same mechanism as in purified solutions. Several results indicate that the answer is yes. First, studies on the addition of red cell membrane components to deoxyhemoglobin S solutions show little or no effect on the delay time. Second, the laser photolysis technique has permitted the measurement of the kinetics of gelation in single red cells, yielding results that are in qualitative agreement with those predicted from the solution studies. The shapes of the kinetic progress curves are very similar to those observed in solution, and for slowly polymerizing cells there are the expected stochastic fluctuations in the delay times. Figure 6 shows that the distribution of observed delay times, which range from a few milliseconds to over 100 seconds, is almost exactly what is predicted from the solution studies and the known concentration heterogeneity. A more detailed comparison of solution and cell data can be made by calculating the intracellular hemoglobin S concentration distribution from the delay time distribution using the solution delay times. The calculated distribution in Figure 6 is qualitatively the same as the measured distributions (Figure 2g), showing that there are no major differences in the rates of gelation in solution and in cells.

SICKLING AS AN INDICATOR OF INTRACELLULAR GELATION

While it has long been accepted that the deformation of SS red cells upon complete deoxygenation is caused by intracellular polymerization, the detailed relation between gelation and cell deformation has remained somewhat ambiguous. In this section we address two questions. The first is whether or not there is a well-defined relationship between cell shape and intracellular polymerization. The second is whether the wide variety of observed cell shapes can be rationalized in terms of what we now know about the kinetics and thermodynamics of gelation. The answers to these questions are important, since a direct link between cell morphology and polymerization would permit a variety of experiments to be performed by using morphological criteria in place of more complex and difficult physical measure-
The relationship between cell deformation and intracellular polymerization has recently been studied with a variation on the laser photolysis technique. Gelatin in partially saturated single cells was investigated by using the kinetics of gelation after complete photodissociation as a probe for the presence of polymer (Fig 7). The delay time provides a very sensitive probe for polymer because even vanishingly small amounts of polymerized hemoglobin (< 0.05%) drastically reduce or eliminate the delay period. These experiments demonstrate clearly that sickling accurately reflects the onset of gelation and that unsickling indicates the complete disappearance of polymer. As a result, curves that describe the fraction of cells containing polymer as a function of saturation or oxygen pressure may be designated sickling and unsickling curves. Figure 7 shows the fraction of sickled cells, ie, the cells that contain polymerized hemoglobin S, as a function of the oxygen pressure, calculated from the measured saturation with carbon monoxide. There is a very large hysteresis between the sickling and unsickling curves. The oxygen pressure at which polymers are first observed in deoxygenation experiments in an initially polymer-free cell is always much lower than the pressure at which polymers disappear in reoxygenation experiments. The hysteresis occurs because there is a delay period before polymer can be detected upon deoxygenation, but in reoxygenation experiments depolymerization occurs without any delay period. Thus the unsickling curve is very close to an equilibrium curve, while the sickling curve depends on the rate of deoxygenation. The sickling curve in Fig 7 was measured by lowering the saturation over a period of one minute. In the microcirculation, deoxygenation occurs in about one second; sickling curves have not yet been measured on this time scale. However, the theoretical one-second sickling curve shown in Fig 7 is seen to be extremely left shifted, in qualitative accord with the results of kinetic studies that show that about 50% of cells sickle after about one second at zero oxygen pressure. Experiments in which cellular deformation occurs much more rapidly have also been performed by deoxygenating cells in a mixer. These data are consistent with the results on intracellular gelation using the laser photolysis technique, suggesting that cell deformation is a reliable indicator of intracellular gelation on the second and subsecond time scale as well.

Essentially identical conclusions have been reached from experiments in which rheologic techniques are used to examine individual sickle cells. The most direct data have come from measurements as a function of oxygen pressure using micropipette techniques. Both the static and dynamic rigidities of the cells can be measured. The static rigidity is characterized by the change in the length of the "tongue" aspirated into the pipette with a change in negative pressure, while the dynamic rigidity is characterized by the half-time required to achieve the final tongue length after initiating the pressure change. For oxygenated cells there are only small increases in these quantities, with the largest increases for the irreversibly sickled cells and the densest cells. As the oxygen pressure is decreased, cells are observed to undergo a variety of morphological changes, and cells having a given set of morphological characteristics can be examined. Cells that are "spiculated" or have a "granular" surface show markedly altered rheology. In contrast, discocytes that maintain a "smooth surface" show the same static and dynamic rigidities at all oxygen pressures as normal cells, in agreement with the conclusion from the kinetic studies. In cells showing morphologic evidence of gelation, both the static rigidity and the half-time for tongue growth increase with decreasing oxygen pressure. At the lowest oxygen pressures the static rigidity increases by up to a factor of 100, and the half-time for tongue growth increases by a factor of 150 to 1000 relative to normal cells and oxygenated sickle discocytes. Since both parameters are much greater for all cells containing polymerized hemoglobin than for polymer-free cells, it appears that the presence or absence of intracellular polymer is much more important in determining cellular rigidity than the extent of intracellular polymerization.

A closely related problem is to understand the enormous variety of cell shapes that are observed in a population of sickled cells. It has been known for over 45 years that slow deoxygenization results in elongated, birefringent cells, while rapid deoxygenization produces a much less distorted cell, originally called a granular form. An unexpected bonus provided by the double nucleation mechanism is that it suggests an explanation for these observations. When cells are rapidly deoxygenated the solubility is suddenly decreased to a low value. The resulting high supersaturation (the ratio of the total concentration to the solubility) causes a high rate of homogeneous nucleation, and the resulting gel contains a very large number of small polymer domains or randomly oriented polymers that could give the cell a granular appearance (Fig 3). In contrast, when deoxygenization is slow, the rate of homogeneous nucleation is reduced to the point that only one homogeneous nucleation event takes place in the cell, and a single polymer domain forms. It if were not for the limited amount of hemoglobin in the cell, this domain would grow to a much larger size than the cell. The cell membrane presumably restricts domain growth to one general direction, resulting in an elongated cell with approximately parallel polymers, the classic "sickle" form (Fig 3). At intermediate rates of deoxygenation, cells may contain a countable number of domains, where, for example, each domain could produce one of the projections of a so-called "holly-leaf" shape (Fig 3).

The appearance of a wide range of morphological forms at a fixed rate of deoxygenation might also be explained as resulting from different rates of polymerization. Remember that the solubility of hemoglobin \( S \) decreases very rapidly at high fractional saturations and much more slowly at low fractional saturations (Fig 2 b). Consequently at a fixed deoxygenization rate dense cells become much more super-saturated and consequently polymerize with much shorter delay times than the light cells. As a result dense cells would be expected to contain many more polymer domains and to have a lumpy, granular appearance, as opposed to a classic sickle shape. Cell morphology is therefore expected to be highly correlated with intracellular concentration and cell...
density. The results of morphological studies on cells polymerized at different rates and on density fractionated cells are consistent with this explanation. The observation of smooth discocytes containing polymer in time-resolved electron microscope studies may represent the initial phase in the formation of a granular form.

**GELATION IN VIVO AND VASO-OCCCLUSION**

We now turn to the question of obstruction of blood flow in the microcirculation resulting from intracellular gelation. Vaso-occlusion is believed to be the cause of pain crises and of the widespread organ damage that contributes substantially to the morbidity and mortality of the disease. Because of the enormous complexity of this problem, the discussion must, of necessity, become much more qualitative and speculative than that which has been presented up to now. We shall see that there are surprisingly little hard data on some of the most basic questions about vaso-occlusion. Nevertheless we believe that a critical examination of this problem is necessary at this point to clarify the important issues and to point to areas where research is most needed. We first discuss gelation and vaso-occlusion, and in the next section we consider the response of the circulatory system to this abnormality.

To gain some perspective on the problem it is instructive to consider the various types of events that have been postulated to occur as a red cell travels through the circulation of an SS patient. In describing these events we shall equate sickling with intracellular gelation. Figure 8 shows a schematic summary. Cells containing no polymerized hemoglobin in the arterial circulation may pass through the microcirculation and return to the lungs without sickling, they may sickle in the veins, or they may sickle in the capillaries. The probability for each of these events will be determined by the delay time for intracellular gelation relative to the appropriate transit time. If it is thermodynamically impossible for gelation to take place (i.e., the intracellular concentration is always lower than the solubility so that at equilibrium no polymer can form) or if the delay time at venous oxygen pressures is longer than about 15 seconds, then sickling will not occur. If the delay time is between about one and 15 seconds, then the cell will sickle in the veins, and, if it is less than about one second, the cell will sickle within the capillaries. For cells that sickle within the capillaries a number of possibilities exist, ranging from no effect on its transit time to transient occlusion of the capillary or a more permanent blockage that ultimately results in destruction of the cell. For some cells the intracellular hemoglobin S concentration may be so high that the solubility is exceeded even at arterial oxygen pressures. These cells will still contain polymerized hemoglobin after oxygenation in the lungs. Upon deoxygenation further gelation will occur rapidly and without a delay time because nucleation of polymers is already complete. Such cells could become stuck in the arterioles or capillaries or could experience a normal transit time through the microcirculation in spite of the decreased deformability.

Figure 8 points out one fundamental problem in describing the pathophysiology of sickle cell disease is to determine the relative probabilities for each of these events. These probabilities will depend on a number of factors, including the total intracellular hemoglobin concentration, the composition of the intracellular hemoglobin, the rate and extent of deoxygenation, and the various transit times involved. For unsickled cells entering the microcirculation, a long capillary transit time will increase the probability of the potentially vaso-occlusive events depicted in Fig 8 in two ways. First, it will permit increased oxygen extraction, which will shorten the delay time. Second, it will increase the probability that a cell with a given delay time will sickle within the capillary. For cells that either enter the microcirculation already sickled or become sickled in the microcirculation, there is a finite probability for occlusion of the small vessels. The duration of an occlusion may be sufficiently long to compromise the oxygen supply to the surrounding tissues and hence may alter the probabilities for sickling and consequent vaso-occlusion in nearby microvessels. This is a somewhat refined version of the "vicious cycle." It is important to recognize that vaso-occlusion is a dynamic process in which the fraction of capillaries that are occluded depends on both the rates of occlusion and the rate of capillary reopening. Thus factors that influence the transit times and the duration of occlusions also play a critical role in the pathophysiology.

With this brief heuristic description as a framework for subsequent discussion, we can now proceed to examine experimental results that help to establish the probabilities.
for the various events depicted in Fig 8. The most straightforward problem is the determination of the fraction of cells that are sickled in the arteries and the fraction that sickle as a result of deoxygenation in the microcirculation. Morphological examination of cells sampled from arterial blood suggests that the average fraction of sickled cells is about 10%.\textsuperscript{78,135-136} This number is, unfortunately, only a rather crude estimate because it is possible that deformed cells such as granular discocytes have not been counted as sickled in some studies; moreover, some irreversibly sickled cells, which may frequently be a major contributor to this count,\textsuperscript{136} may contain no polymer.\textsuperscript{140} The values for reversibly sickled cells in different patients range from 1% to 16%, while total sickled cell counts range from 9% to 30%.\textsuperscript{136} This variation presumably results from differences in the distribution of intracellular hemoglobin composition and concentration, as well as from differences in arterial oxygen saturation.

It would appear, then, that an average of about 90% of cells entering the peripheral circulation contain no polymer and hence would undergo gelation with a delay period if sufficiently deoxygenated. The morphological data suggest that about 20% of cells are sickled in the mixed venous return,\textsuperscript{78,135-136} indicating that an additional 10% of cells have sickled as a result of passing through the microcirculation.* Any analysis based on studies of mixed arterial and venous blood is clearly somewhat oversimplified because oxygen extraction in the microcirculation of some tissues, such as in the coronary and hepatic circulations,\textsuperscript{155} is considerably greater than average. As a result the cells in these organs will have much shorter delay times leading to a higher number of sickled cells in the veins.

These findings are consistent with existing information on in vitro delay times. Although the ideal in vitro experiment in which gelation and degelation are continuously monitored in individual red cells at physiologic rates of deoxygenation and oxygenation is not yet possible, the laser photolysis experiment\textsuperscript{136,137} affords an informative preview of the results expected from such experiments. The unsickling curves obtained in these experiments (Fig 7) show that at equilibrium only about 5% of the cells contain polymerized hemoglobin S at an oxygen pressure of 85 torr, which is the average arterial value found in homozygous SS patients,\textsuperscript{9} while over 90% of cells remain sickled at the average mixed venous pressure of 45 torr. In contrast, the sickling curves show that only about 5% of cells (overlapping significantly with the cells that were found to be sickled at 85 torr at equilibrium) are sickled after deoxygenation to venous oxygen pressures on physiologic time scales. These results indicate that the delay time is preventing more than 80% of cells from sickling in vivo in this homozygous SS patient. That is, for over 80% of cells gelation would occur if equilibrium were achieved, but the delay times are so long that these cells return to the lungs and are reoxygenated before any significant amount of polymer has formed. The fact that about 10% of cells already contain polymer in the arterial circulation does not substantially affect the fraction of sickled cells in the microcirculation, since if polymer were not present the delay times for the large majority of these cells would be less than the capillary transit time.† The enormous difference between the unsickling curve and the sickling curve in these experiments (Fig 7) graphically demonstrates the significance of the delay time for gelation in vivo and simultaneously shows that equilibrium data or data obtained in slow deoxygenation experiments are not at all representative of the in vivo situation in which the relevant time scale is seconds. Similar large differences are expected for oxygen binding and polymer fraction curves (Fig 2 e, 2 f and discussion below).

This analysis points to the critical need for obtaining much more data relevant to sickling in vivo. An accurate description would require direct measurements of the distribution of delay times at physiologic rates and extents of deoxygenation. It would also be desirable to have more precise data on the extent of sickling in arterial and venous samples using experimental methods that take into account the kinetics of sickling as well as the recently acquired information on the relation between cellular deformation and intracellular gelation.\textsuperscript{136,137,140} Most of the data on morphological sickling were obtained before the gelation kinetics were described, and since then very little attention has been given to designing accurate morphological experiments on venous and arterial samples. Such experiments would require rapid fixation of cells and careful examination by scanning electron microscopy or high resolution optical microscopy. It would, of course, be preferable to develop rapid sampling techniques that could assess the extent of intracellular gelation, or at least the presence or absence of polymer, in individual cells from arterial and venous blood.

We next consider the question of occlusion of the microcirculation by sickled cells, which is clearly a central problem in understanding the pathogenesis. The results and calculations described above as well as data on red cell survival suggest

\textsuperscript{*}The estimate of 10% additional sickling in the microcirculation is only a very approximate number. Differences in this number are observed for different venous returns of the same patient and for the same venous return in different patients, but the tissue-to-tissue variation is generally smaller than the differences observed between patients.\textsuperscript{78,135-136} Since cells are deoxygenated in the microcirculation within one to two seconds, significant variations may result from additional sickling during the time required to sample the cells from the veins and to fix them with glutaraldehyde and possibly some additional unsickling for cells sampled from the arteries.\textsuperscript{137,138}

\textsuperscript{†}Shearing forces are known to decrease the delay time by breaking polymers, producing new ends, and thereby increasing the rate of nucleation.\textsuperscript{121,129,130,139} In this discussion we have assumed that there is no significant effect of shear on the in vivo delay time. While no direct experimental information on this point exists, two considerations suggest that the effect of shear on intracellular polymerization in vivo will be small. First, cells flowing in small tubes concentrate in the low-shear region near the center of the tube, while the high shear regions near the walls are preferentially occupied by plasma.\textsuperscript{161,162} The shear field to which cells are exposed is thus very much smaller than the average field in the microvessels. Second, the high internal viscosity of even unpolymerized cells makes coupling of the external shear field to the inside of the cell inefficient, as evidenced by the absence of tank-treading behavior in low viscosity media.\textsuperscript{163}
that the densest cells are primarily responsible for occlusion of the microcirculation. These cells are predicted to have a much greater probability for the events in Fig 8 that could lead to vaso-occlusion. Because of the high intracellular hemoglobin S concentration, they have shorter delay times, or no delay times and therefore are more likely to sickle in the capillaries when deoxygenated. In addition, if gelation occurs at any given oxygen pressure, the higher intracellular concentrations of polymerized hemoglobin (Fig 2d) produce stiffer cells (ie, cells with lower static and dynamic deformabilities). The half-life of dense cells (~ two days) is significantly shorter than that of randomly labeled cells from the same patient population (~ five days) or the average half-life of cohort-labeled cells from a wider patient population (~ 17 days). Experiments in which the density distribution of a population of labeled reticulocytes was followed over the cell life span also show that the dense cells are the last to appear in the labeled population. These results show clearly that as cells age their density increases and that once they become dense they are removed from the circulation rather quickly. The mechanism by which cells concentrate has been the subject of much recent work. The major contribution appears to result from the loss of cell water associated with potassium loss, but there may also be some contribution from the loss of membrane surface area caused by the sickling-unsickling cycle. If it is assumed that all cells must become dense cells before they are removed from the circulation, then the fraction of dense cells is predicted to be equal to the ratio of the half-life of the dense cells to the mean half-life for new cells. The measured half-life of the dense cell fraction is roughly consistent with this prediction, since the densest 10% to 15% of the cell population has a half-life that is 0.1 to 0.2 times that of labeled reticulocytes. Cells that have an abnormally low probability of sickling, such as low density F cells, are expected to have extremely long delay times even at venous saturations. Since these cells can only be sickled upon stasis or passage through tissues where deoxygenation is extreme, they will presumably concentrate much more slowly than cells having a higher probability of sickling. This protection of F cells results in a longer life span and therefore an increased concentration of F cells in the fractions of intermediate density, which represent the oldest cells in the population, and a reduced concentration in the densest cell fraction.

Of the vaso-occlusive events depicted in Fig 8, it is only possible to make even a crude quantitative estimate for the probability of occlusion followed by destruction of the cell. Based on a mean cell lifetime of 17 days and a circulation time of 15 seconds, an “average” red cell makes about 100,000 trips through the microcirculation before being removed. If we use the fraction of sickled cells as an estimate of the fraction of dense cells (~ 20%), the above argument suggests that if dense cells were removed from the circulation only by vaso-occlusive events, they would be trapped and destroyed once in about 20,000 trips through the microcirculation. However, hemolysis data indicate that about 60% to 70% of sickle cells are destroyed in the reticuloendothelial system, so the probability for destructive vaso-occlusion is reduced to about one in 60,000 trips. While this probability appears, at first glance, to be extremely low, it is large enough to result in the steady-state blockage of a significant fraction of the total number of capillaries, since there are approximately 10^13 circulating red cells and only about 10^10 capillaries. The exact fraction blocked is highly dependent upon the duration of the destructive blockade, which, by this estimate, occurs in the average capillary about once every hour. If, for example, each blocked vessel remains occluded for one hour, the fraction blocked is 50%, but this value is reduced to 2% if the vessel is blocked for only one minute. This result clearly suggests that factors that influence the duration of a capillary blockade could play a critical role in determining the extent to which tissue oxygen supply is compromised.

It has been very difficult to obtain quantitative information on the frequency of the other events depicted in Fig 8. Recently a series of studies using an excited rat mesococum preparation has begun to provide some interesting results. This preparation permits control over the tissue oxygen pressure and the perfusing pressure, allowing vascular resistance (defined as the ratio of the arteriovenous pressure difference divided by the venous outflow) changes to be measured quantitatively in both denervated and innervated vascular beds. Alternatively, trapping of cells can be measured by first perfusing the innervated bed and subsequently washing out the trapped cells by increasing the perfusion pressure or denervating the prepared bed. These experiments make two important points. The first is that at venous PO_2 the fraction of capillaries blocked at steady state in this preparation can be as large as 80%. The second is that the ability of cells to block the microcirculation is correlated with their density, a finding that is consistent with the description of Fig 8.

These investigations provide the best opportunity to simulate the events occurring in the microcirculation of SS patients. In addition to measuring vascular resistance and cell trapping, cinematographic observations of these preparations permit determination of the sites at which blockage occurs. The limited information obtained so far has not established the relative importance of precapillary and intracapillary sites as the principal sites of occlusion. It would be important to extend studies of the microvasculature to use preparations that permit deoxygenation in the tissue to more closely simulate the in vivo situation. It will also be important to use these preparations to determine the factors that affect both the frequency and duration of occlusions, since they are equally important in determining the steady-state fraction of blocked capillaries.

GELATION AND OXYGEN DELIVERY

An important aspect of the pathophysiology of sickle cell disease is to understand how the circulatory system maintains adequate oxygen delivery in the face of anemia and vascular obstruction. In other severe anemias, for example, those arising from blood loss or iron deficiency, there are two primary compensation mechanisms. One is an increased blood flow through the tissues resulting from an expansion of the muscular arterioles. The increased blood flow from
this decreased peripheral resistance increases filling of the right atrium, thereby increasing the cardiac output. The second is an increase in intracellular 2,3-DPG concentration, which lowers the oxygen affinity and facilitates oxygen unloading in the tissues. As a result there is an abnormally slow venous oxygen saturation at or near a normal oxygen pressure. The increased blood flow through the lungs may also result in a decreased oxygen pressure and saturation in arterial blood from the decreased efficiency of gas exchange, making oxygen delivery less efficient.

In sickle cell disease the abnormalities are somewhat different. The cardiac output is higher than in anemias of comparable severity, the oxygen pressure and saturation of arterial blood are lower, and the oxygen pressure and saturation of venous blood are higher. This decrease in arteriovenous saturation difference, particularly when compared to other anemic states, means that there is significant impairment of oxygen unloading to the tissues. These differences between sickle cell disease and other anemias presumably result from the intracellular gelation and vascular obstruction that are unique to sickle cell disease.

The very low oxygen affinity of polymerized hemoglobin S explains the lower arterial oxygen saturation in sickle cell disease, compared with other anemias, and even the lower arterial oxygen pressure. Infarctive damage to lung tissue of SS patients could also decrease the efficiency of oxygen loading in the lungs, but this does not appear to be a major factor because a comparable degree of arterial unsaturation is observed in children in whom there is no other evidence of impaired lung function. We have estimated earlier that about 20% of cells entering the lungs contain substantial amounts of polymerized hemoglobin, which would be expected to decrease both the rate and extent of oxygen binding to the sickled cells in the alveoli. The few in vitro experiments support this contention. Slow depolymerization of the sickled cells after they leave the lungs could also contribute to the lower arterial oxygen tension by scavenging oxygen from the plasma and from the cells that contain no polymer.

In considering oxygen unloading in the tissues, the absence of polymerized hemoglobin in most cells because of the long delay times is an important consideration. The traditional approach has been to ignore this fact and to utilize oxygen dissociation curves measured in vitro over periods of minutes or longer where intracellular gelation is much more extensive than in the in vivo situation. Use of the in vitro curve has led to the conclusion that the large right shift in the equilibrium or quasi-equilibrium dissociation curves substantially compensates for the anemia. This conclusion is misleading, since oxygen binding to approximately 80% of the cells should be similar to that found for other states of comparable anemia. The calculation in Fig 2e shows that the in vivo oxygen unloading curve is predicted to be significantly less right shifted than the in vitro equilibrium curve. This smaller right shift is consistent with the observation that the fractional saturation of venous blood from sickle cell patients has near normal values, while the oxygen pressure is higher than in normal blood by about 5 torr.

Perhaps the most puzzling circulatory abnormality is an increase in cardiac output that is larger than is found in anemias of comparable severity. The absence of a significant increase in arterial blood pressure requires that the vascular resistance be decreased in direct proportion to the increase in cardiac output. While decreased vascular resistance may appear paradoxical in the presence of capillary blockage, this finding can be rationalized by the fact that the bulk of the peripheral resistance arises from the muscular arterioles and not the capillaries. In severely anemic states the arterioles open to increase blood flow through the tissues, thereby compensating for the low hematocrit. A similar response also appears to be the primary mechanism of compensation for the reduced hematocrit in SS disease, but the peripheral resistance is decreased to a significantly greater extent than in other anemias of comparable severity. The opening of the muscular arterioles must therefore increase the fraction of capillaries in the tissue bed that are perfused to above normal, even in the presence of blockage, if there is sufficient capillary reserve.

In addition to increasing the number of perfused capillaries, there is evidence that opening of the muscular arterioles increases the pressure drop across the capillaries, thereby increasing the rate at which red cells traverse the capillaries. A recent study in which laser Doppler velocimetry was used to measure capillary flow in the forearm skin of SS patients showed that the average rate of red cell flow was close to normal. This result implies that the ~40% decrease in hematocrit is almost exactly compensated by a combination of an increase in the number of perfused capillaries and an increase in capillary flow rate. By decreasing the time available for the equilibration of the red cell with the oxygen tension of the capillary wall, an increased capillary flow rate would be expected to decrease oxygen unloading. In spite of this effect a decreased capillary transit time could benefit the patient by decreasing the rate of capillary blockage. The decreased transit time would not only increase the delay time by increasing the final fractional saturation of hemoglobin but it would also decrease the time during which a cell is at risk from sickling within the microcirculation. Such a mechanism provides an attractive explanation for the decreased arteriovenous oxygen saturation difference in SS patients. A frequently invoked explanation for the low arteriovenous saturation difference is shunting through large vessels.

There is no direct measurement of the fraction of occluded capillaries in any tissue in sickle cell disease. However, a tentative estimate for muscle can be extracted from data on the exercise tolerance of sickle cell patients. In these studies patients were subjected to increasing work loads, and the lactic acid level in the blood was monitored. The work load at which lactate began to increase is defined as the anaerobic threshold. An extension of the Krohn model for oxygen delivery to tissues predicts that the work output at this point is nearly directly proportional to the density of perfused capillaries. If it is assumed that the muscle is maximally perfused and that capillary densities in the muscle of SS patients are normal, then the fraction of blocked capillaries can be estimated from the anaerobic threshold to be about 0.4. This fraction decreases to about 0.1 to 0.2 when the fraction of SS cells is decreased to about 50% by exchange transfusion.
While this mechanism might account for the unexpectedly high venous saturation in specific tissues, it cannot explain the fact that similar results are found for vessels such as the femoral vein.\(^7\) The femoral vein primarily drains muscle beds in which there is anatomic evidence that significant shunting is impossible.\(^9\)

Finally, we should point out that there has been only one attempt to quantitatively evaluate the response of the complete circulatory system to the altered properties of SS blood.\(^9\) This study used an established model for microcirculatory control\(^9\) to calculate the changes in peripheral resistance blood flow, and capillary oxygenation.\(^9\) This model incorporates approximate but realistic descriptions of oxygen supply and consumption in the tissues as well as local feedback control of both the arteriolar resistance and capillary density to regulate the tissue oxygen pressure. When anemia is simulated by reducing the hematocrit, the model predicts a compensatory decrease in peripheral resistance and increased blood flow. If, however, the quasi-equilibrium increase in viscosity and reduced equilibrium affinity of SS blood are also introduced, the model predicts a capillary resistance that is about 1.4 times normal and blood flow that is about 80% of normal. The calculated effects are in striking contrast to the observed decrease in peripheral resistance and increase in blood flow.\(^7,8,11\) The discrepancy presumably results in part from the incorrect assumption of equilibrium oxygen unloading and viscosity changes made in carrying out these calculations. If the kinetics of intracellular gelation were to be incorporated into this model (Fig 2 e), the predicted effects would be closer to the observed.

VARIATIONS IN CLINICAL SEVERITY

It is now widely recognized that there are large differences in clinical severity among patients with homozygous SS disease, some patients having only the mild symptoms associated with a chronic hemolytic anemia, others suffering from repeated painful episodes and severe organ damage.\(^24,25,98,199\) The reasons for this broad spectrum of clinical manifestations are not at all clear, and it is one of the major areas of current research. To discuss the role of gelation in producing differences in clinical severity among homozygous SS patients, it is useful to briefly summarize the most important results of the preceding discussion.

The picture that emerges is a dynamic one in which a balance between the rate of obstruction and reopening of capillaries results in a steady state in which a certain fraction of capillaries is blocked in each tissue. This balance may be very delicate, with small changes in either the rate of obstruction or reopening capable of significantly altering the fraction of occluded capillaries. Any increase in this fraction, particularly in tissues with inadequate capillary reserve, could result in irreversible hypoxic damage and may be the cause of pain crises. While almost nothing is known about the opening of occluded capillaries, we are beginning to understand the mechanisms that control the rate of capillary obstruction. This rate must depend, at least in part, on the fraction of sickled cells in the microcirculation, which is determined by the times required for intracellular gelation relative to the transit times.\(^6,7\) Factors that favor gelation can decrease the steady state number of obstructed capillaries by decreasing the delay time, thereby increasing the fraction of cells that sickle within the microcirculation and the rate of obstruction. An increased extent of polymerization in a sickled cell could also increase the probability of an occlusion because of the decrease in deformability.\(^140\) Because the delay time is so much more sensitive to changes in physiologic variables than the extent of polymerization (compare, eg, Figs 2 d and 3), it is most probably the dominant factor in determining changes in the rate of capillary obstruction. Simultaneous small changes in a number of physiologic variables could result in a sufficient change in the distribution of delay times to produce the fluctuation in the fraction of blocked capillaries that precipitates a pain crisis. In this way the sensitivity of the delay time could account for the episodic nature of crises.

Another mechanism for increasing the rate of obstruction is to increase the transit time in the microcirculation, which increases the probability of sickling. In this way factors that slow down cells can also affect the rate of obstruction. The only such factor that has been identified so far is the adherence of cells to the vascular endothelium.\(^201-204\) In addition to capillary blockage, other events influence these probabilities by altering the characteristics of the cell population. For example, cells that normally would return to the lungs may sickle in the venous return, particularly in tissues in which the residence times in the veins are long, resulting in an increase in intracellular concentration and therefore a decreased delay time in subsequent trips through the microcirculation.

This picture immediately raises the question of how much of the variation in clinical severity in homozygous SS disease can be explained by variations in intracellular gelation and how much must be attributed to variations in circulatory dynamics. Under the category of intracellular gelation are included the effects of intracellular hemoglobin concentration and composition, arising either from genetic variations or from cell aging. To begin, let us consider the relation between gelation times and clinical severity among the various sickling disorders, where there are easily measurable differences in both gelation and standard hematologic parameters. Figure 9 shows the effect of hemoglobins A, C, and F on the delay time and a comparison of the distribution of delay times at zero saturation for the three most common syndromes: homozygous SS disease, SC disease, and sickle trait. SC disease is generally a much milder sickling disorder than is SS disease, while sickle trait is totally benign.\(^24,37\) The delay times for SC cells are considerably longer than those for SS cells, indicating that many fewer cells sickle in vivo. For sickle trait cells the delay times at zero saturation are all longer than about one second, indicating that even under totally anoxic conditions cells would escape the microcirculation before polymerization has begun. With the possible exceptions of the hypertonic renal medulla, it would appear that sickle trait cells never sickle in vivo, explaining the lack of any clinical manifestations.

The reasons for the increased sickling of SC cells compared to sickle trait cells are quite interesting. Little or no difference is observed in the gelling properties of hemoglobin S + C mixtures and S + A mixtures. A careful comparison has shown that there are no significant differences in either
the delay times\textsuperscript{62} or solubilities\textsuperscript{62,63}. The principal reason for the increased sickling of SC cells is that they contain a higher hemoglobin S concentration than sickle trait cells. This increase results from two effects\textsuperscript{62}. First, there is a greater fraction of hemoglobin S in SC cells (50/50 S/C) than in sickle trait cells (40/60 to 30/70 S/A). The reduction in the fraction of hemoglobin S in sickle trait cells is caused by a decreased rate of association of \(\alpha\) chains to \(\beta\) chains relative to \(\beta^A\) chains during the tetramer assembly process\textsuperscript{205,206}. When the concentration of \(\alpha\) chains is reduced because of coexisting \(\alpha\) thalassemia, this competition is enhanced and a disproportionately larger fraction of \(\beta^A\)-containing tetramers are formed. Second, the total intracellular hemoglobin concentration is higher in SC cells\textsuperscript{62,64}. Since reticulocytes have nearly the same density distribution as the average cell population, the red cells must emerge from the marrow more concentrated. The reasons for this are not yet completely understood, but it has been suggested that the binding of hemoglobin C to the red cell membrane induces a potassium and water efflux\textsuperscript{207}.

Hemoglobin F also has a marked effect on gelation. This is clinically most evident in the uncommon double heterozygous condition of hemoglobin S with pancellular hereditary persistence of fetal hemoglobin, which may be asymptomatic. In this condition hemoglobin F is more evenly distributed, and most cells contain a substantial amount of hemoglobin F (up to 35\%).\textsuperscript{34} This mixture has gelling properties in vitro that are similar to the 40/60 Hb S + A mixture found in sickle cell trait (Fig 9)\textsuperscript{37,37} and would therefore be predicted to have a very mild or asymptomatic clinical course. In homozygous SS disease there is a variable increase in hemoglobin F that results from two factors: an increased production of F reticulocytes and preferential survival of F cells.\textsuperscript{61} At hemoglobin F levels above 20\%, corresponding to about 60\% F cells, there may be some amelioration of the disease, but below 20\% there appears to be no significant effect.\textsuperscript{201,206}

Because of the tremendous sensitivity of the delay time to the total hemoglobin concentration, it has been suggested that the clinical severity of homozygous SS disease may be improved by a small dilution of the intracellular hemoglobin.\textsuperscript{6,7} The increase in the delay time resulting from a decrease in the intracellular concentration would allow more cells to escape the microcirculation before gelation has begun.\textsuperscript{6,7} To estimate the effect of concentration on the delay time we use a 15th power inverse concentration dependence, since this is the concentration dependence found for delay times of about one second (Fig 3). The decrease in MCHC from 32 g/dL to 30 g/dL associated with the coexistence of \(\alpha\) thalassemia, in which two of the four \(\alpha\) genes are deleted \((-\alpha/\alpha)\)\textsuperscript{309} produces an almost three-fold increase in the delay time for the “average cell.” The result is increased red cell survival\textsuperscript{67,70} and an indication of fewer episodes of the acute chest syndrome and leg ulceration.\textsuperscript{211,212} Also, in SS disease there may be an increased frequency of the \(\alpha\) gene deletion with age, suggesting that a decreased total intracellular hemoglobin concentration is associated with a longer life expectancy.\textsuperscript{213} In HbS-\(beta^A\)-thalassemia there is a similar decrease in MCHC, and the clinical course relative to SS disease is “milder in many features.”\textsuperscript{124}

Thus far we have seen that for genetically different sickling disorders there is a good correlation between intracellular gelation in vitro, for solutions having compositions of the average cell, and disease severity for the average patient.\$ To investigate the role of clinical diversity one would ideally want to know at least the distribution of intracellular delay times for patients from a clinically well-characterized population. No such data are yet available. An efficient but limited method of examining distributions of intracellular gelation is to measure density distributions, since the density is proportional to the total intracellular hemoglobin concentrations. Differences in intracellular solvent conditions of pH, 2,3-DPG concentration, etc, are expected to have a much smaller effect on gelation than differences in intracellular hemoglobin concentrations. Consequently the distribution is expected to reflect the distribution of intracellular delay times, except for the effect of F cells.

The only study carried out so far is one in which cell density distributions were compared with the incidence of painful crisis.\textsuperscript{714} No correlation was found between the fraction of cells in the highest density range and crisis frequency. This result was interpreted as evidence that the greater probability of intravascular sickling is not the principal cause of increased crisis frequency, but that variations in the anatomy and dynamic properties of the microcirculation are responsible for differences among patients. As pointed out earlier, one factor that could be important in determining transit times in the microcirculation is adherence to the vascular endothelium. A strong correlation has in fact been

\section*{Footnotes}

\$Correlations between gelation and both overall clinical severity and degree of anemia have also been obtained using the in vitro fraction polymerized \textit{at equilibrium} as a measure of gelation in vivo.\textsuperscript{209} Although the equilibrium fraction polymerized in vitro is not relevant to the in vivo situation as discussed earlier, these correlations give a very similar result\textsuperscript{209} because of the close correlation between the kinetic and equilibrium properties of gelation.\textsuperscript{28}
found between overall clinical severity and the tendency of the red cells to adhere to vascular endothelium in in vitro experiments.201-203 The severity score used in this study included evidence for organ damage resulting from microvascular occlusions as well as the frequency of pain crises. In this same study there was no correlation between severity and hemoglobin F levels or irreversibly sickled cells, which are known to correlate with the fraction of dense cells.82

THE PROBLEM OF INHIBITING GELATION IN PATIENTS

The strong correlation between gelation and severity for the "average" patient with the various sickling disorders clearly indicates that inhibition of gelation should result in amelioration of the disease. The data on hemoglobin mixtures shows that it will not be necessary to completely inhibit gelation (ie, increase the solubility such that it equals or exceeds the total intracellular hemoglobin concentrations at all oxygen pressures) but that a therapeutic effect should result from sufficiently increasing the delay time to allow more cells to escape the microcirculation and be reoxygenated in the lungs before gelation has begun.6,7,57,98 In this way there should be a reduction in the rate of production of dehydrated, rapidly polymerizing cells, which have been generally assumed to be the subpopulation of cells most responsible for initiating vaso-occlusion.6 To give this concept a quantitative basis we may ask: how much must gelation be inhibited to obtain a specified therapeutic effect in patients? An approximate answer to this question can be obtained from the correlation between in vitro delay times or solubilities in solutions of deoxyhemoglobin mixtures having the compositions found in various sickling disorders and their "average" clinical course.75

The data for this comparison are found in Fig 9, and Table 1 shows the increase in delay time and solubility relative to pure deoxyhemoglobin S for solutions having the hemoglobin composition found in sickle-β⁺-thalassemia, sickle cell disease with hereditary persistence of fetal hemoglobin, and sickle trait. The results in Table 1 establish a set of criteria for obtaining a specified therapeutic effect. They suggest that the threshold for obtaining a therapeutic effect in SS disease would result from a method that produces an increase in the in vitro delay time of about a factor of 100 (corresponding to a solubility ratio of about 1.2), which is the increase found for solutions having the hemoglobin composition of sickle-β⁺-thalassemia; an increase of about 2 to 10³ (solubility ratio of about 1.3) should produce a major therapeutic effect; and a 10⁴ to 10⁶-fold increase in the in vitro delay time (solubility ratio of about 1.5 to 1.6), found for solutions with the composition of sickle trait cells, is predicted to result in a "cure."

With these estimates we can examine the potential utility of the various strategies that have been proposed to inhibit gelation in patients. Four different approaches have been explored or considered in some detail: (1) blocking intermolecular contact formation in the polymer, (2) raising the oxygen affinity, (3) decreasing the total intracellular hemoglobin concentration, and (4) promoting fetal hemoglobin production. The oldest idea is to develop a competitive or covalent inhibitor that would bind stereospecifically to hemoglobin S and interfere with polymer formation. Two general types of mechanisms have been considered. In one the "drug" acts directly by binding to an intermolecular contact site in the polymer, thereby competitively inhibiting polymerization. In the other the "drug" inhibits polymerization indirectly by changing the conformation at the intermolecular contact site so that it no longer "fits" into the polymer. The direct approach to inhibiting gelation poses a number of problems. Unlike an enzyme, where a substrate analogue can be a powerful inhibitor of catalysis by binding to the active site, none of the known intermolecular contact sites provide such a target. There are no clefs, grooves, or other obvious structural features that can be used to design molecules with complementary structures that might bind to hemoglobin S with high specificity. Examination of the intermolecular contacts also gives no real clues. This result might have been anticipated because the interactions between molecules in the polymer are weak. One approach would be to determine the structure of hemoglobin-antibody complexes in which polymer contacts are the antigenic determinant. Since hapt-en-antibody interactions are generally much stronger, the

| Table 1. Clinical Course, Gelation Delay Time and Requirements for Therapy |
|---|---|---|---|
| Disorder | S/β⁺-Thalassemia | S/HPHF | A/S Trait |
| Clinical course relative to S/S disease | Less severe | Much less severe | No disease |
| Red cell composition* | | | |
| % Hb A | 20-30 | 0 | 60-75 |
| % Hb F | 0 | 20-35 | 0 |
| % Hb S | 80-70 | 80-65 | 40-25 |
| Log delay-time ratio† | 1.5-2.5 | 2.5-5.0 | 6.0-8.0 |
| Solubility ratio‡ | 1.1-1.2 | 1.2-1.35 | 1.45-1.65 |
| Therapy requirements | | | |
| Percent saturation of inhibitory site§ | 20-40 | 40-55 | ≥65 |
| Decrease in intracellular concentration (g/dL) | 3-5 | 5-9 | ≥11 |

*The data are from Serjeant.24 For S/β⁺-thalassemia this is the composition of the non-F cells.
†This is the ratio of the delay time for the mixture to the delay time for pure deoxyhemoglobin S at the same total hemoglobin concentration and is obtained from the data in Fig 9. These ratios are for subphysiologic concentrations using the temperature-jump technique for measuring delay times. For physiologic concentrations where the dependence of the delay time on supersaturation is smaller, these ratios are expected to be smaller, as is found with intact cells (see Fig 9b to d).
‡This is the ratio of the solubility for the mixture to the solubility for pure deoxyhemoglobin S in the limit of no polymerized hemoglobin (from Eaton and Hofrichter24).§This is the fractional saturation of an ideal inhibitory site, ie, one that completely prevents polymerization, required to produce the delay time increase for pure deoxyhemoglobin S.§This is the required decrease in intracellular concentration, assuming an intracellular hemoglobin S concentration of 34 g/dL in the cells entering the circulation in S/S disease. Adapted from Sunshine et al.87
antibody-binding site would be expected to have the structural features of a very effective inhibitor and hence could serve as a model for the ambitious organic chemist attempting to construct molecules that cover the contact sites.

A natural target on the hemoglobin molecule for attack by the indirect mechanism is the pocket between the $\beta$ subunits, which constitutes a specific, relatively high-affinity binding site for 2,3-DPG. For example, bifunctional aspirin derivatives have been described that crosslink the $\beta$ subunits by covalently binding to opposite $\beta$2 lysines. Analysis of the three-dimensional structure of the complex by x-ray crystallography shows that this modification causes a shift in residues of the F-helix that are part of the acceptor site for the $\beta6$ contact region, explaining the very large increase in solubility (solubility ratios up 1.5). Although these particular inhibitors may not turn out to be therapeutically useful, this and other recent studies demonstrate the power and feasibility of using x-ray crystallography to understand the mechanism of action of inhibitors and to design more effective ones. Most studies of inhibitors of gelation have not taken such a "rational" approach. Nevertheless a number of effective inhibitors have been found, although none has been developed to the point of being a serious candidate for use in patients.

A second, more speculative strategy for inhibiting gelation is to increase oxygen affinity by shifting the allosteric equilibrium toward the R structure. At any given oxygen pressure there will be a lower concentration of molecules in the T quaternary structure and therefore a decreased tendency to polymerize. Calculations based on the effect of saturation on gelation suggest that therapeutically useful effects might result, although homeostatic responses that maintain oxygen delivery could buffer the inhibitory effect. One interesting way of shifting the allosteric equilibrium toward R and one that would require much lower doses of a drug than directly attacking the hemoglobin molecule would be to inhibit 2,3-DPG synthesis. An additional beneficial effect would result from the fact that 2,3-DPG promotes gelation of T-state molecules. It will be important to evaluate the effect of an increase in oxygen affinity in some detail because many inhibitors of deoxyhemoglobin S gelation also increase oxygen affinity.

The third strategy is to decrease the intracellular hemoglobin concentration, an idea directly generated from the kinetic studies. This approach takes advantage of the enormous concentration dependence of the delay time. There are two obvious ways that could, in principle, be used to decrease the total intracellular hemoglobin concentration. One is to permanently increase the red cell volume, and the other is to reduce hemoglobin biosynthesis without a decrease in red cell volume, for example by slowly introducing iron deficiency. There are some clinical data to suggest that concomitant iron deficiency is in fact beneficial. The idea of swelling red cells has been tested in a preliminary way. A combination of sodium restriction, high fluid intake, and the use of an antidiuretic reduced the serum sodium to 120 to 125 mg/dL, which resulted in a 2 to 3 g/dL decrease in the MCHC. Both the frequency and duration of painful crises appeared to be reduced. Although this study was quite limited, involving only three patients who served as their own controls, it suggests that small reductions in intracellular hemoglobin concentration may indeed have a therapeutic effect, as predicted from the in vitro gelation studies. Another approach to swelling red cells in patients has been to alter the ion transport properties of the red cell membranes so as to affect a net water influx. Several agents have been described. Of these the most extensively studied is cetylid, which may be effective in directly retarding the dehydration that produces the rapidly polymerizing dense cells as a result of sickling-unsickling cycles. In a placebo-controlled, double blind study cetylid had some effect in reducing the severity and duration of pain crises, but there is yet no information on its effectiveness in decreasing crisis frequency or organ damage.

The fourth strategy for inhibiting gelation in patients is to stimulate the production of $\gamma$ globin. As discussed earlier, the inhibitory effect results from the inability of the $\alpha_2\gamma_2$, or $\alpha_2\gamma_6$ tetramers to copolymerize with $\alpha_2\beta_2$. If $\gamma$ chains are exchanged for $\beta$ chains in all cells, then some therapeutic effect is expected with hemoglobin F levels of about 10% to 15% (Fig 9 and Table 1). If hemoglobin F is heterogeneously distributed, clinical data from Saudi Arabians, where sickle cell disease is milder, suggest that amelioration would result if the percentage of F reticulocytes exceeds 20%, which results in a steady-state level of about 60% F cells and 20% hemoglobin F. Data on American blacks suggest that at hemoglobin F levels above 10% there is a decreased probability of major organ failure, while the threshold for a decrease in crisis frequency is about 20%. Although the molecular mechanism is not at all well understood, significant stimulation of F reticulocyte production has been achieved in SS patients with two drugs: 5-azacytidine and hydroxyurea. With 5-azacytidine hemoglobin F levels of 12% and 20% were achieved in two patients treated for more than 100 days, and there was a concomitant decrease in pain crises. The preceding analysis indicates that there is cause for optimism, as there are several totally independent and viable approaches to the therapy of sickle cell disease. Too frequently a single approach has been criticized as not being useful because by itself it does not produce a dramatic effect in patients. There is, of course, no reason why a specific treatment for sickle cell disease could not consist of the use of several drugs simultaneously, each inhibiting gelation by a different mechanism and at nontoxic doses that would produce only a small effect if given alone.

ACKNOWLEDGMENT

We thank H. Franklin Bunn for many helpful discussions and criticisms.

REFERENCES

1. Singer K, Fisher B: Studies on abnormal hemoglobins. V. The distribution of type S (sickle cell) hemoglobin and type F (alkali resistant) hemoglobin within the red cell population in sickle cell anemia. Blood 7:1216, 1952

From www.bloodjournal.org by guest on October 23, 2017. For personal use only.
HEMOGLOBIN S GELATION AND SICKLE CELL DISEASE


From www.bloodjournal.org by guest on October 23, 2017. For personal use only.
67. Williams RC: Concerted formation of the gel of hemoglobin S. Proc Natl Acad Sci USA 70:1506, 1973
90. Noguchi CT, Torchia DA, Schechter AN: 13C NMR quanti-
HEMOGLOBIN S GELATION AND SICKLE CELL DISEASE


35. Steinhardt J, Malfa R: Anomalous viscosity properties of solutions of deoxy sickle-cell hemoglobin at temperatures below...


150. Sherman JI: The sickling phenomenon, with special reference to the differentiation of sickle cell anemia from the sickle cell trait. Bull Johns Hopkins Hosp 67:309, 1940


165. McCurdy PR, Sherman AS: Irreversibly sickled cells and red cell survival in sickle cell anemia. A study with both DF'2P and DF'3P. J Lab Clin Invest 47:1731, 1968


HEMOGLOBIN S GELATION AND SICKLE CELL DISEASE

175. Little RC: Physiology of the heart and circulation. Chicago, Year Book, 1965
191. Krouth A: The number and distribution of capillaries in muscle with calculation of the oxygen pressure head necessary for supplying the tissues. J Physiol (Lond) 52:409, 1919
Hemoglobin S gelation and sickle cell disease

WA Eaton and J Hofrichter